



# Mid-Atlantic Crossroads

Advanced Regional Internetworking for  
Higher Education and Research

Office: 8400 Baltimore Avenue  
Suite 102  
College Park, Maryland 20740  
301.405.6666

August 31, 2007

## Welcome

Welcome to the August 2007 edition of the MAX Newsletter. In these updates we highlight current technical activities and policy initiatives amongst our staff, our participants, and the Regional Optical Network and national networking communities. Please let us know your thoughts, and especially your suggestions.

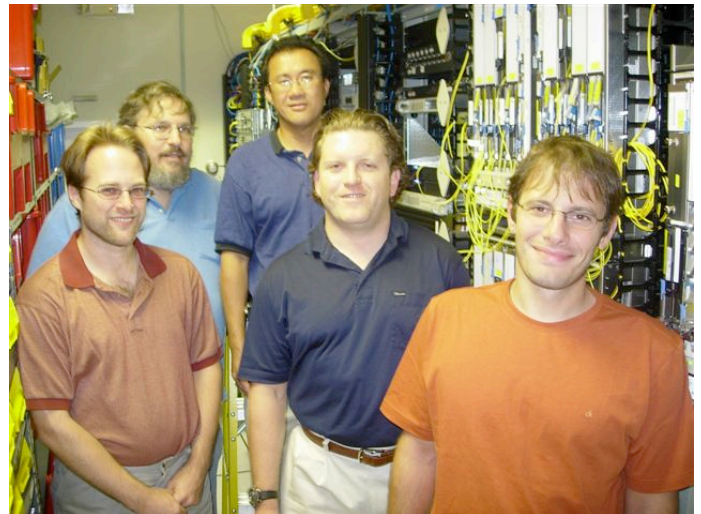
## Executive Director Message

This has really been an exciting month as facility work for space and power is now complete in our McLean, College Park, and Baltimore Points of Presence (PoPs), our engineering staff spent a week being trained, and the new Fujitsu optical equipment delivered to our lab. The boxes were opened quickly and within a couple days chassis installed, modules installed, boxes configured, and fiber interconnects completed. The Sun server was also installed along with the Oracle database and Network Management System software.

A group from NASA Goddard has already taken a tour of the lab and received an up close look at the new DWDM equipment.

Look for a detailed installation schedule to be sent out in about a week. As stated last month we are really excited about this effort and the new service capabilities it will enable for all of you, our participants, this fall.

- Peter O'Neil



Here's a photo of Dave Diller, Dan Magorian, Quang Bach (up on the ladder), Matt Siniscal, and Chris Tsonis all standing proudly alongside the Fujitsu's (with all the yellow fiber wires hanging off) after getting the equipment up and running.



# Mid-Atlantic Crossroads

Advanced Regional Internetworking for  
Higher Education and Research

Office: 8400 Baltimore Avenue  
Suite 102  
College Park, Maryland 20740  
301.405.6666

August 31, 2007

## New Tools for September Back to School

SRI and Georgia Tech have been working on a pretty cool new tool that quickly locates bot traffic inside a network. A government/military version of this software has been in use successfully and a public version was recently made available. BotHunter introduces a new kind of passive network perimeter monitoring scheme, designed to recognize the intrusion and coordination dialog that occurs during a successful malware infection. It employs a novel dialog-based correlation engine (patent pending), which recognizes the communication patterns of malware-infected computers within your network perimeter. BotHunter is available for download at <http://www.cyber-ta.org/BotHunter/> and runs under Linux Fedora, SuSE, and Debian distributions.

There is also a highly interactive honeynet using BotHunter run by SRI. The URL is <http://www.cyber-ta.org/releases/malware-analysis/public/>. dozens of new infections are being detected each day and this site is helpful in understanding the behavior of the received malware. Also, it generates a list of potentially evil IP addresses and DNS queries.

## Information Security Economics

by: Ross Anderson and Tyler Moore

Abstract. The economics of information security has recently become a thriving and fast-moving discipline. As distributed systems are assembled from machines belonging to principals with divergent interests, incentives are becoming as important to dependability as technical design. The new field provides valuable insights not just into 'security' topics such as privacy, bugs, spam, and phishing, but into

more general areas such as system dependability (the design of peer-to-peer systems and the optimal balance of effort by programmers and testers), and policy (particularly digital rights management). This research program has been starting to spill over into more general security questions (such as law-enforcement strategy), and into the interface between security and sociology. Most recently it has started to interact with psychology, both through the psychology-and-economics tradition and in response to phishing. The promise of this research program is a novel framework for analyzing information security problems – one that is both principled and effective.

[http://www.cl.cam.ac.uk/~rja14/Papers/econ\\_crypto.pdf](http://www.cl.cam.ac.uk/~rja14/Papers/econ_crypto.pdf)

## Corning Announces New Fiber Technology

Corning's breakthrough is based on a nanoStructures™ optical fiber design that allows the cabled fiber to be bent around very tight corners with virtually no signal loss. These improved attributes will enable telecommunications carriers to economically offer true high-speed Internet, voice and HDTV services to virtually all commercial and residential (apartment and condominium) buildings. Current optical fiber installations lose signal strength and effectiveness when bent around corners and routed through a building, making it difficult and expensive to run fiber all the way to customers' homes.

"This is a game-changing technology for telecommunications applications," said Peter F. Volanakis, president and chief operating officer at Corning. "We have developed an optical fiber cable that is as rugged as copper cable but with all of the bandwidth



# Mid-Atlantic Crossroads

Advanced Regional Internetworking for  
Higher Education and Research

Office: 8400 Baltimore Avenue  
Suite 102  
College Park, Maryland 20740  
301.405.6666

August 31, 2007

benefits of fiber. By making fundamental changes in the way light travels in the fiber, we were able to create a new optical fiber that is over 100 times more bendable than standard fibers." Corning's newest fiber technology achieves this while maintaining compatibility with industry performance standards, existing manufacturing processes and installation procedures. "So, customers don't have to sacrifice one benefit to get another," he said.

[http://www.corning.com/media\\_center/press\\_releases/2007/2007072301.aspx](http://www.corning.com/media_center/press_releases/2007/2007072301.aspx)

## myGrid and Taverna Services

myGrid is a suite of components designed to support in silico science, encompassing workflow design and execution, data and metadata management and provenance collection.

The wealth of bio-informatics resources and data available in the public domain provides scientists with great opportunities for research, but the distribution and heterogeneity of the data means that integrating resources is challenging.

myGrid addresses this problem by providing a mechanism of interoperating and integrating between these resources using Taverna workflows. The Taverna project aims to provide a language and software tools to facilitate easy use of workflow and distributed compute technology within the eScience community. Taverna workflows connect distributed web services and other services. Most of these services are not developed or maintained by myGrid, they are provided by third parties, for example, the major bio-informatics service providers – EBI, NCBI, DDBJ, and KEGG. In most cases, they are not even developed for Taverna, yet simply exploit the Web Service

technologies being adopted by the community. Taverna can now access over 3000 services in the bio-informatics domain and others in the fields of astronomy, chemo-informatics, health informatics and many more. <http://www.mygrid.org.uk/>  
<http://taverna.sourceforge.net/>

## ISOC Study

The Internet Society (ISOC) Board of Trustees is currently engaged in a discovery process to define a long term Major Strategic Initiative to ensure that the Internet of the future remains accessible to everyone. The Board believes that Trust is an essential component of all successful relationships and that an erosion of Trust: in individuals, networks, or computing platforms, will undermine the continued health and success of the Internet.

The Board will meet in special session the first week October of 2007 for intensive study focused on the subject of trust within the context of network enabled relationships. As part of this process, the Board issued a call for subject experts to participate in the two day discussion on topics including: the changing nature of trust, security, privacy, control and protection of personal data, methods for establishing authenticity and providing assurance, management of threats, and dealing with unwanted traffic. A study report is expected later this fall.

<http://www.isoc.org/>

## DOD Mandates Encryption

The Defense Department has tightened its rules for protecting sensitive but unclassified information. In what likely is the first time in government, DOD's chief information officer, John Grimes, is requiring



# Mid-Atlantic Crossroads

Advanced Regional Internetworking for  
Higher Education and Research

Office: 8400 Baltimore Avenue  
Suite 102  
College Park, Maryland 20740  
301.405.6666

August 31, 2007

DOD to encrypt all sensitive but unclassified data stored on mobile devices.

Grimes' July 3 memo mandates that such data stored on mobile devices must be encrypted in compliance with the National Institute of Standards and Technology's Federal Information Processing Standard 140-2. The term mobile devices describes laptop PCs, personal digital assistants and removable storage media, such as thumb drives and compact discs.

Paul Kurtz, chief operating officer at Good Harbor Consulting, said the new policy is "a watershed development within the federal government that has not received a lot of attention."

"DOD is making an important step forward here to ensure that all data, except that approved for public release, is encrypted," he said. "It's watershed because, frankly, the rest of the federal government should operate the same way."

Kurtz said government information, even if it is unclassified, can be used for criminal purposes if it falls into the wrong hands.

The FIPS 140-2 specification, approved in 2001, grew from Federal Standard 1027, General Security Requirements for Equipment, which used the now-outdated Data Encryption Standard. NIST is now working on the next iteration, FIPS 140-3.

<http://www.fcw.com/article103467-08-13-07-Print>

## SciVee

"The National Science Foundation, Public Library of Science and the San Diego Supercomputing Center have partnered to set up what can best be described

as a "YouTube for scientists", SciVee. Scientists can upload their research papers, accompanied by a video where they describe the work in the form of a short lecture, accompanied by a presentation. The formulaic, technical style of scientific writing, the heavy jargonization and the need for careful elaboration often renders reading papers a laborious effort. SciVee's creators hope that that the appeal of a video or audio explanation of paper will make it easier for others to more quickly grasp the concepts of a paper and make it more digestible both to colleagues and to the general public."

<http://www.scivee.tv/>

## Head of TSA Interviewed

Next time you plan to fly and prepare to pack the items you wish to take onboard, perhaps this interview between Bruce Schneier and Kip Hawley, head of the Transportation Security Administration (TSA) will help you better understand the little dance we all go through in navigating the security lines.

<http://www.schneier.com/interview-hawley.html>

## Merger Update

The NLR-Internet2 Merger Planning Team (MPT) issued its report this month. It was used to develop a definitive merger agreement and a new set of by-laws. The two Boards are still in the process of reviewing and discussing further modifications to the documents.



# Mid-Atlantic Crossroads

Advanced Regional Internetworking for  
Higher Education and Research

Office: 8400 Baltimore Avenue  
Suite 102  
College Park, Maryland 20740  
301.405.6666

August 31, 2007

## Member Spotlight

### NASA Goddard's Computational and Information Sciences and Technology Office

#### MERRA Project to Reconstruct Last 30 Years of Earth's Climate and Weather

By Jarrett Cohen

MERRA, the largest application ever hosted by the NASA Center for Computational Sciences (NCCS), will soon be running on an augmented "Discover" Linux cluster. The Modern Era Retrospective-analysis for Research and Applications, as MERRA is formally known, will consume 544 processors of the NCCS's Linux Network Custom Supersystem for 18 to 24 months.

As its name implies, a retrospective analysis, or re-analysis, starts in the past and marches forward in time to computationally reconstruct atmospheric conditions. Reanalyses are commonly undertaken by climate and weather centers, which use the same data assimilation systems as they apply to their forward-looking forecasts.

MERRA is an endeavor of the Global Modeling and Assimilation Office (GMAO) at Goddard Space Flight Center. One of the GMAO's primary functions is to support NASA Earth Observing System (EOS) satellite instrument teams and field experiments with assimilated data products in near real-time. The GMAO also uses models and assimilation systems to document and understand climate variability and predictability.

Traditionally, weather centers have an archive of analyses built up from their real-time systems, which are continually changing as upgrades occur. In contrast, the GMAO "can go back with a reanalysis and have a consistent system over time for doing the processing," said Michael Bosilovich, GMAO mete-

orologist and MERRA co-principal investigator with fellow GMAO meteorologist Siegfried Schubert.

With MERRA, GMAO scientists particularly hope to gain new insights into Earth's water cycle and understand how it changes with underlying climate variability. Research community access to the resulting data will expand the prospects for advances.

Data assimilation systems pair a computer model with an analysis that ingests, calibrates, and quality controls observations and then feeds them to the model. "Observational products are discontinuous in space and time," Bosilovich said. "Where we don't have data, the model fills in the gaps." Together, he explained, model and analysis provide globally gridded meteorological information that is uniform and reliable.

The GMAO is currently doing production-scale testing of the next version of its Goddard Earth Observing System-5 Data Assimilation System (GEOS-5 DAS), planned for MERRA. GEOS-5 DAS incorporates the GEOS-5 atmospheric general circulation model and the Gridpoint Statistical Interpolation (GSI) analysis. GSI was developed by the National Weather Service's National Centers for Environmental Prediction (NCEP) with GMAO contributions.



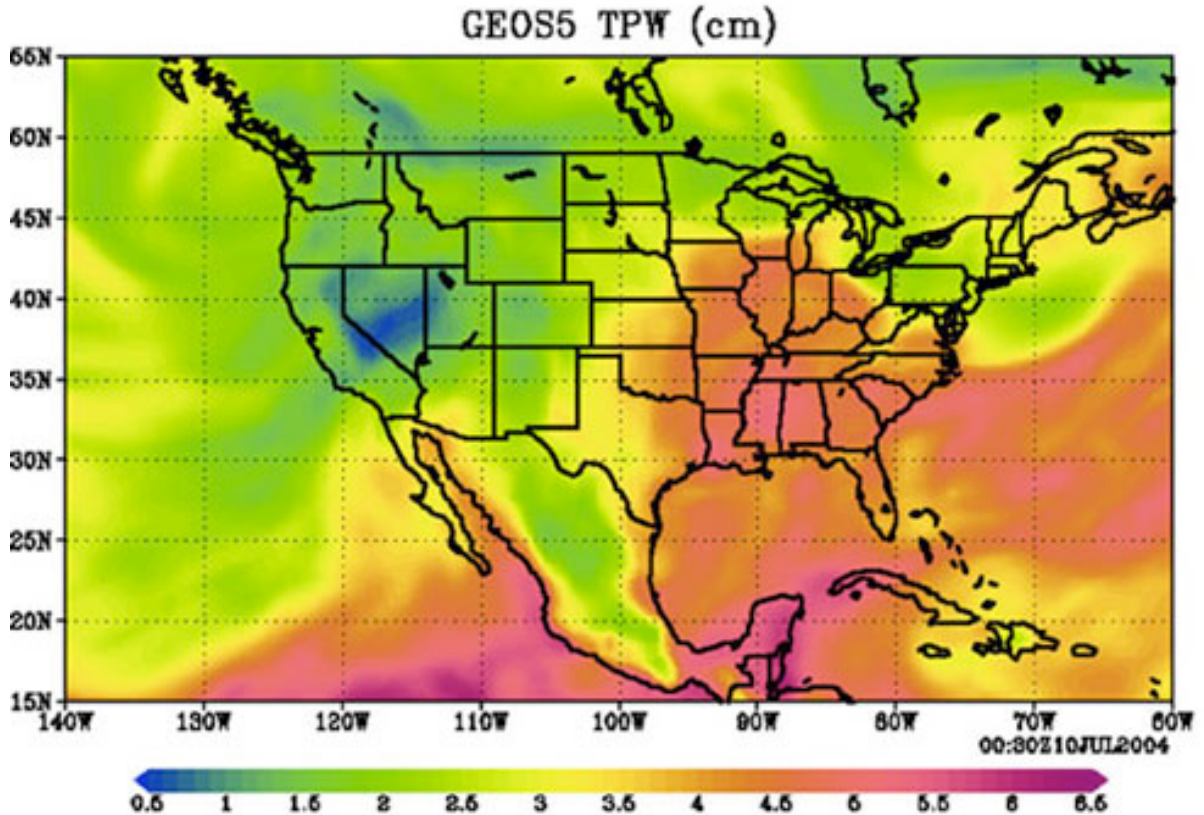
# Mid-Atlantic Crossroads

Advanced Regional Internetworking for  
Higher Education and Research

Office: 8400 Baltimore Avenue  
Suite 102  
College Park, Maryland 20740  
301.405.6666

August 31, 2007

This visualization shows Total Precipitable Water (TPW) in centimeters over the United States on July 10, 2004. The data come from an experimental reanalysis of Summer 2004 running the GEOS-5 Data Assimilation System (DAS) at the 1/2-degree resolution being used for MERRA.



## Reanalysis at a new scale

By using the GEOS-5 DAS, MERRA is leveraging the GMAO's most capable analysis tool to date, as well as the researchers developing it.

MERRA will produce a comprehensive record of Earth's climate and weather from 1979, the beginning of the Earth-observing satellite era, up to the day it finishes sometime in 2009. MERRA will cover twice as much time as the first GEOS reanalysis, which the predecessor Data Assimilation Office conducted with the GEOS-1 assimilation system back in 1993. In addition, MERRA will have a horizontal resolution of 1/2-degree—roughly one model grid

point every 55 kilometers—and 72 vertical levels. That compares to the 1-degree resolution used by the most recent large-scale reanalysis, the Japanese 25-year Reanalysis Project that ended in March 2006. Bosilovich said that finer resolution better locates observational inputs geographically, improving precipitation and other details in the analysis.

The NCCS computational offerings drove the GMAO's decision to alter their original plans for 1-degree resolution and 36 levels. "Since the computational capability is there, we are running MERRA at this higher resolution," said Gi-Kong Kim, GMAO production group lead. "Five years ago, such high spatial resolution was inconceivable."



# Mid-Atlantic Crossroads

Advanced Regional Internetworking for  
Higher Education and Research

Office: 8400 Baltimore Avenue  
Suite 102  
College Park, Maryland 20740  
301.405.6666

August 31, 2007

MERRA also will have fine temporal resolution, with much of the diagnostic data produced every hour. Other diagnostics will be released at 3-hour intervals, which is typical for recent reanalyses. Among MERRA's more than 300 diagnostic variables are temperature, moisture, wind, surface pressure, and—unique with reanalyses—fields for the chemistry transport community. Hourly diagnostics will resolve the diurnal (daily) cycle of minimum and maximum values more precisely than existing 3- and 6-hourly reanalysis data products. They also will allow close study of individual weather events that can pop up within a few hours.

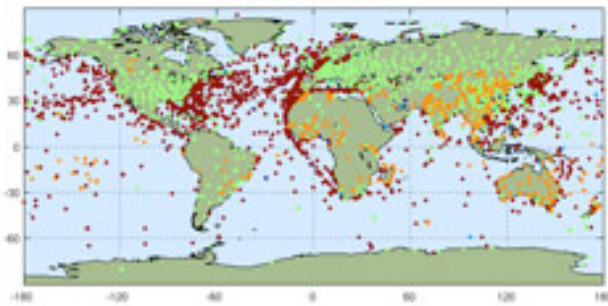
“Since it appears early in production, I am especially interested to see the Presidents’ Day Storm of February 1979,” Bosilovich said. This storm spread record-breaking snow amounts from the Ohio Valley to the Mid-Atlantic, with snowfall rates greater than 4 inches per hour around Washington, DC. At the time, weather models failed to predict the storm’s

intensity. Bosilovich also mentioned the 1988 Central U.S. droughts and floods, the 1993 Midwest floods, and several major El Niños as events of interest.

Besides increased spatial and temporal resolution, the variety and number of observations that MERRA must assimilate is much greater than in the previous reanalysis. Observation sources include satellites, ground stations, weather balloons, ships, and aircraft. “GSI can handle a lot more observations per day than in the past,” Bosilovich said. He said this capability is especially important for assimilating 151 channels from the Atmospheric Infrared Sounder (AIRS) instrument on board the EOS Aqua satellite. AIRS represents 40 percent of the observations to be assimilated since Aqua’s 2002 launch. For the Aqua period, the GSI will read in 8 million observations per day for potential inclusion in the assimilation. Through quality control and other thinning, 5 million of those finally get assimilated into GEOS-5.

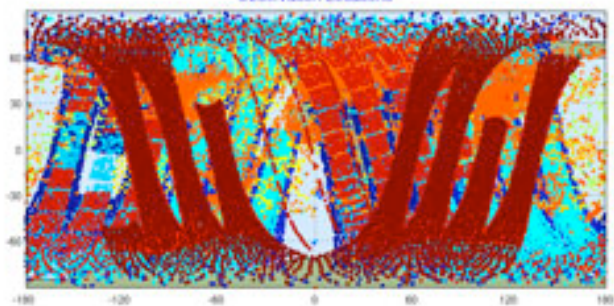
07-Jan-1973 12UTC All data: 77,098 observations

Observation Locations



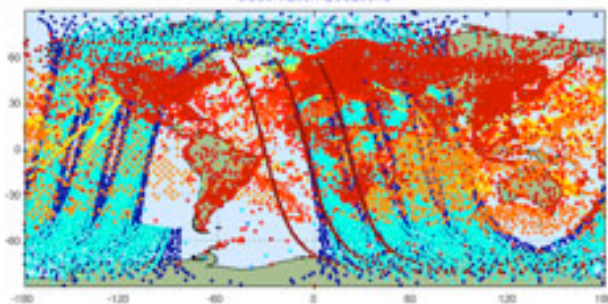
02-Aug-1987 12UTC All data: 550,602 observations

Observation Locations



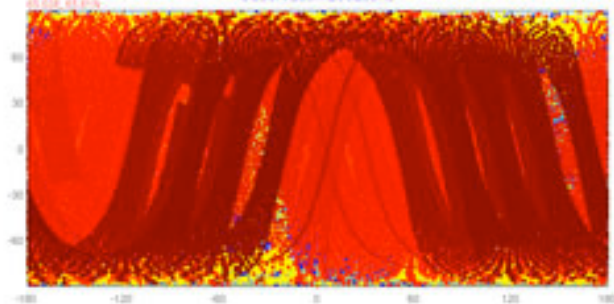
07-Jan-1979 12UTC All data: 325,765 observations

Observation Locations



07-Jan-2006 12UTC All data: 4,217,655 observations

Observation Locations





# Mid-Atlantic Crossroads

Advanced Regional Internetworking for  
Higher Education and Research

Office: 8400 Baltimore Avenue  
Suite 102  
College Park, Maryland 20740  
301.405.6666

August 31, 2007

The number of Earth observations to be assimilated has increased dramatically over the last few decades. The panels demonstrate the evolution of observing systems from 1973 (pre-satellite) to 1979 (TIROS Operational Vertical Sounder [TOVS]) to 1987 (add Special Sensor Microwave Imager [SSM/I] and several TOVS) to 2006 (add Atmospheric Infrared Sounder [AIRS] and several each of TOVS and SSM/I). The headers list the number of observation points for an example 6-hour period during each year.

## Production strategies

These new scale factors translate into substantial computing and storage requirements during MERRA's production phase. For example, doubling the spatial resolution increases processing needs eight-fold. Add the factors together, and requirements grow exponentially.

As the MERRA design expanded, it became clear that no current supercomputer could analyze all 30 years in sequence and finish within the 18-to-24-month window. To meet that timetable, MERRA must analyze about 30 "data days" per wall-clock day. Consequently, the GMAO divided the reanalysis into three concurrent streams of 10 years each. By running on 128 Discover processors, each stream can analyze 10 to 11 data days per day. An additional 128 processors will run supporting experiments.

All MERRA output will reside in the NCCS mass storage system. With an estimate of 15 gigabytes of output per data day, expected total data volume is 165 terabytes.

"The NCCS has done a very good job in having its production systems ready," Kim said. "The GMAO and NCCS have been working together very closely to keep the computing requirements updated."

The original plan was to run MERRA on "Explore," a 1,152-processor SGI Altix 3700 system. After test runs showed GEOS-5 performing faster on Dis-

cover, the GMAO decided to take advantage of the cluster's expansion to 2,560 processors this summer (see "Discover Cluster Expands with New Processors and Visualization Capabilities" in this issue). They also wanted to ensure completing the reanalysis on one supercomputer. "Given the length of the run, Explore's service contract might run out," Bosilovich said. Citing a large weather center that had to stop a reanalysis because the computer went away, he stressed the need for "continuity in computing systems and a stable platform."

The Discover expansion includes not only 1,024 new processors but also an additional 120 terabytes of online disk, which will better support MERRA production assessment and post-processing. Specialized Visualization Nodes enable close monitoring of the output streams. To speed data transfer between the mass storage system and Discover, the NCCS technical staff increased the network pipes to 10 gigabits per second. Staff also reconfigured the mass storage system to accommodate MERRA input/output, which is particularly write-intensive, explained Harper Pryor, CISTO Programs Development Manager (SAIC).

For the GMAO, Kim said one of the most challenging aspects about preparing for MERRA production has been collecting, identifying, and preparing all the input observations. Over several months, the GMAO made a "data sweeper" run of the reanalysis at 2-degree resolution using Explore. "The purpose was not so much to validate the scientific product but to ensure that the input observations would not give us surprises while we do production," Kim said. "If we encountered GEOS-5 system crashes due to problems related to input observations, we investigated and then designed solutions for production." To finish as quickly as possible, the sweeper also used multiple streams.

The final preparatory step is validation studies, which began running on Discover and Explore during July. "We are making sure the science has been



# Mid-Atlantic Crossroads

Advanced Regional Internetworking for  
Higher Education and Research

Office: 8400 Baltimore Avenue  
Suite 102  
College Park, Maryland 20740  
301.405.6666

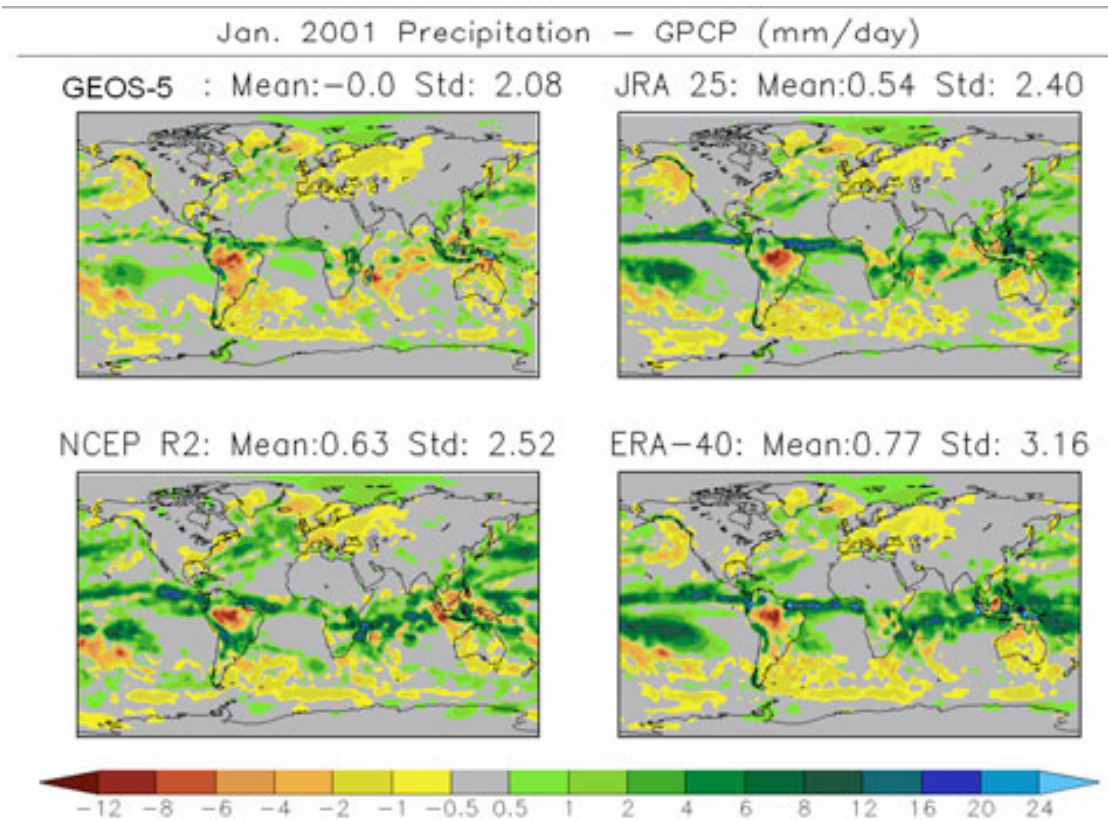
August 31, 2007

applied properly,” Bosilovich said. With the water cycle receiving prime attention, one of the comparative data sets is from the Goddard Laboratory for Atmospheres’ Global Precipitation Climatology Project. Additional validation data come from remotely sensed data products (including MODIS, the Moderate Resolution Imaging Spectrometer on Terra and Aqua), other reanalyses, and ground stations. Such variety will enable verifying correct large-scale and regional circulation in the atmosphere.

Review of validation results is the responsibility of an External Users Group of scientists from universities, other climate and weather centers, and NASA field centers. “Rather than determining on our own when we are or are not ready to start production, we assembled an External Users Group to evaluate community needs,” Bosilovich said.

MERRA validation studies compare merged satellite and in situ observations from the Global Precipitation Climatology Project (GPCP) to reanalysis data. The panels show the differences in

monthly mean precipitation (in millimeters per day) between GPCP and data from the GEOS-5 DAS and three completed reanalyses: the National Centers for Environmental Prediction Reanalysis-2 (NCEP R2), the European Centre for Medium-Range Weather Forecasts 40 Year Reanalysis (ERA-40), and the Japanese 25-year Reanalysis Project (JRA 25). GEOS-5 has a mean of zero and the smallest standard deviation, indicating the best agreement with GPCP. Of particular note is GEOS-5’s improved tropical precipitation bias compared to the reanalyses.





# Mid-Atlantic Crossroads

Advanced Regional Internetworking for  
Higher Education and Research

Office: 8400 Baltimore Avenue  
Suite 102  
College Park, Maryland 20740  
301.405.6666

August 31, 2007

## Sharing the results

The broader Earth science community will have on-line access to a subset of MERRA data through the Goddard Earth Sciences Data and Information Services Center (GES DISC), formerly the Goddard Distributed Active Archive Center (DAAC). The GMAO estimates the available data will equal 55 terabytes, including the chemistry transport fields

The NCCS will funnel the data to GES DISC before archiving to tape, so scientists will have access to provisional data while MERRA is running. In collaboration with the GMAO, GES DISC is building a disk-based data storage and distribution system with a Web user interface. Kim said that planned capabilities include search using a variety of parameters, on-the-fly subsetting, and visualization to aid data search and comparison.

MERRA has a range of potential applications. For one, a researcher may pick case studies from the historical record. Because MERRA spans several generations of Earth-observing platforms, the GMAO will be able to look at climate variability and predictability from a uniform perspective. This research is part of their contribution to the U.S. Climate Change Science Program. Modeling groups could also use the data to drive mesoscale/regional models or global chemistry models. "Transferring a scientific product to the community has been our goal since writing the proposal," Bosilovich said. "That will be exciting to see in action."

MERRA funding comes from NASA's Research, Education, Applications Solutions Network (REA-SoN) and Modeling, Analysis, and Prediction (MAP) programs.

<http://gmao.gsfc.nasa.gov/research/merra>

<http://disc.gsfc.nasa.gov>



# Mid-Atlantic Crossroads

Advanced Regional Internetworking for  
Higher Education and Research

Office: 8400 Baltimore Avenue  
Suite 102  
College Park, Maryland 20740  
301.405.6666

August 31, 2007

## MAX Participants

### Federal Labs and Agencies:

- ATDnet - NRL, LTS, DISA
- D.C. Government
- Energy Sciences Network (ESNet)
- Laboratory for Telecommunications Sciences
- Library of Congress
- NASA / GSFC
- National Archives and Records Administration (NARA)
- National Institutes of Health (NIH)
- National Institute of Standards and Technology
- National Library of Medicine (NLM)
- National Oceanic and Atmospheric Administration (NOAA)
- National Science Foundation (NSF)
- U.S. Census
- USDA, Beltsville Agricultural Research Center
- U.S. Department of Health and Human Services
- U.S. Department of State (through GWU)
- U.S. Geological Survey
- U.S. Holocaust Memorial Museum

### Higher Education:

- Baltimore Education & Research Network
- Catholic University
- GEANT
- Georgetown University
- George Mason University
- George Washington University
- Johns Hopkins University

- Johns Hopkins University - Applied Physics Laboratory (JHU-APL)
- Montgomery College
- National Consortium for Supercomputing Applications / ACCESS
- Network Virginia
- Smithsonian Institution
- Southern Universities Research Association (SURA)
- University of California, D.C. campus
- University Consortium for Advanced Internet Development (UCAID / Internet2)
- University of Maryland, College Park
- University of Maryland, Baltimore
- University of Maryland, Baltimore Co.
- Univ. System of Maryland Network
- University of Southern California, Information Sciences Institute / East
- Washington Research Library Consortium

### Corporate and Non-profit:

- Columbia Telecommunications Corporation (CTC)
- Howard Hughes Med. Institute
- Fujitsu Labs of America
- Inter-American Development Bank (IADB)
- Northrop Grumman Corporation
- The Institute for Genomic Research
- Windber Professional Services, Inc.
- World Bank
- The Venter Institute