

Cloud Computing Technologies Facilitate Earth Research

NASA Technology

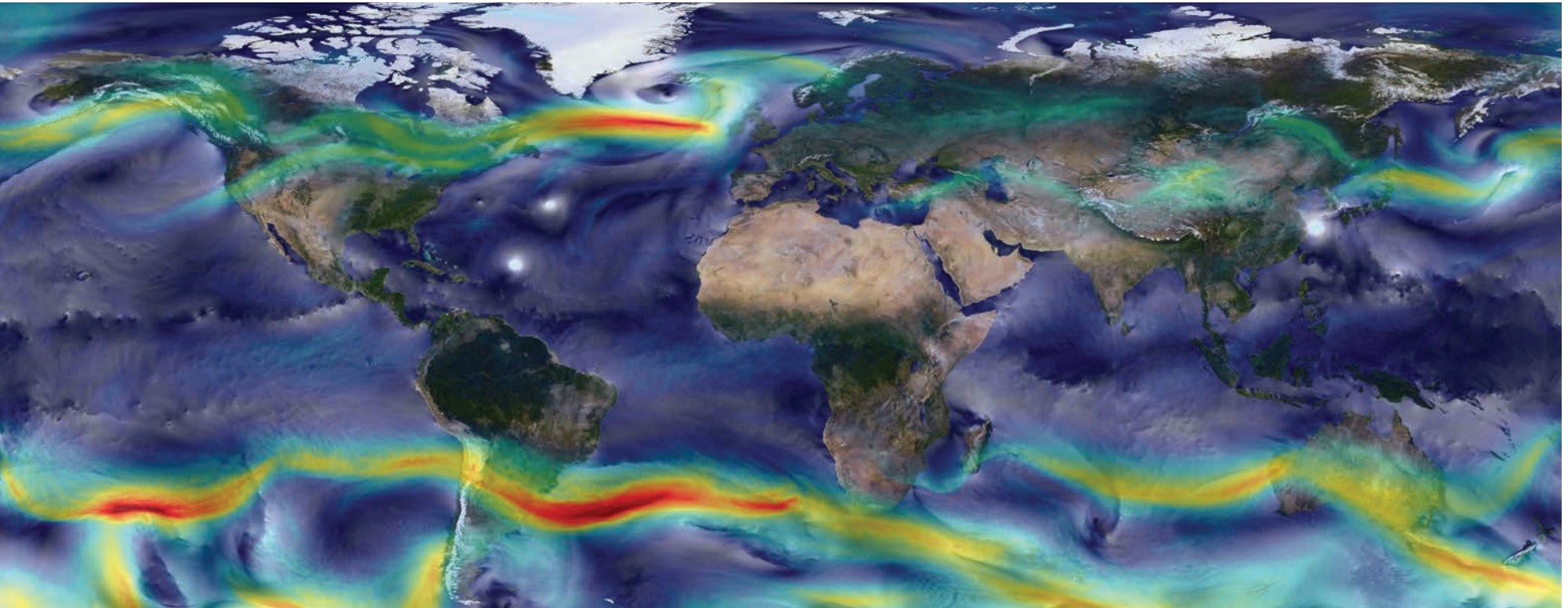
Throughout the years, NASA has done more than any agency to explore what lies beyond our world. It has sent rovers to traverse the Martian landscape and launched spacecraft to monitor the sun. Another satellite, Voyager, is even leaving the solar system for the uncharted territory of interstellar space. But equally important are its satellites whose lenses are trained back on Earth. Since 1972, Landsat satellites operated by NASA have been providing high-resolution topographical

land data, and the Moderate Resolution Imaging Spectroradiometer (MODIS), launched on Terra in 1999 and Aqua in 2002, offers a fresh perspective of Earth's surface every one to two days.

By using satellite data, NASA scientists are able to investigate an array of environmental issues, such as deforestation, aerosol accumulation, carbon cycles, and glacier recession, to name just a few. The data are expansive enough for tracking worldwide carbon emissions, yet refined enough for monitoring vacant Californian farmland.

In the late 1990s, NASA made all of its satellite data freely available to the public. Outside researchers could, and still do, make requests through NASA data centers, which upload requested information via File Transfer Protocol applications. But over the years, as datasets have grown larger in volume and are covering longer periods of time, it has become more time-consuming for data to be shared that way.

According to Rama Nemani, senior Earth scientist at the Advanced Supercomputing Division at Ames Research Center, "For somebody to do a large-scale



NASA's Goddard Earth Observing System Model, known as the GEOS-5, uses satellite data to simulate worldwide weather at resolutions as fine as 3.5 kilometers. Here it displays a snapshot of global winds with a 10-kilometer resolution. Slower surface winds, shown in white, trace features such as storm systems. Meanwhile, faster-moving upper-level winds are colored red.

continental or global analysis, things are difficult because you have to download all this data from the various centers, which takes weeks to months,” he says, “and then you have to write the code on your own to analyze the data instead of reusing the codes that other scientists already wrote.”

To help researchers—specifically those receiving NASA funding—more quickly access and analyze satellite data, in 2010 Nemani led the development of NASA Earth Exchange, or NEX. Through NEX, researchers not only have access to datasets; they are also able to tap into Ames’ Pleiades supercomputer—one of the world’s most powerful. The platform also facilitates collaboration and information sharing between participating scientists. “In this way, it’s like Facebook,” Nemani says. “We share results, algorithms and things like that, and in the forums, people suggest new ways of doing things.”

In November 2013, NASA and AWS [Amazon Web Services] announced the availability of climate and Earth science satellite data to researchers and educational users through the AWS cloud.

In the few years since its founding, the platform has been utilized in several studies and investigations, including the NASA Earth Exchange Downscaled Climate Projections, or NEX-DCP30. Scientists at Ames, the Climate Analytics Group in nearby Palo Alto, and California State University, Monterey Bay, collaborated on the dataset, which provides a view of future continental U.S. temperature and precipitation patterns based on four different greenhouse gas emissions scenarios, spanning the period from 1950 to 2099. What makes the projections even more intriguing is their scale: While most climate change forecasts cover large geographic areas on the order of 100-250 kilometers, the downscaled dataset can provide half-mile resolution analysis. “Folks can look at this data and see how their neighborhoods are

expected to change in the next 100 years,” notes Nemani. “That’s how fine the scale is.”

While NEX has done much to improve access to NASA satellite data and supercomputing services and to promote scientific collaboration, after two years, Nemani says, he and his colleagues concluded that more could be done to simplify access and encourage innovation. A few issues were apparent. First, because outside researchers were accessing NASA’s computer network, they were waiting as long as six to eight months to receive security clearance. Another bottleneck was that the large number of requests to use Pleiades often resulted in weeks of wait time. Lastly, only NASA-funded researchers were allowed access. “Given these constraints, we started looking for yet other ways of engaging the community,” Nemani says.

At about the same time, in 2012, the Obama administration announced its Open Data Executive Order, the aim being to make government data more accessible to the public. One way the White House was pushing to make that happen, says Tsengdar Lee, program manager for High-End Computing at NASA headquarters, was by encouraging more public-private partnerships. “That’s how we got into a conversation with Amazon.”

Technology Transfer

Amazon.com Inc., headquartered in Seattle, started as a retail business in the mid-1990s. As the Internet began to play an increasing role in commerce, the company developed the infrastructure to meet demand. Its datacenters are now located all over the world and have the capacity to simultaneously handle millions of purchases and also stream multimedia to its global consumer base.

As it happens, the technology developed for this intricate computer network would also benefit researchers and companies that need access to both large amounts of data storage and supercomputing capabilities. As a result, in 2006, the company began offering those options through Amazon Web Services (AWS), which operates through a worldwide set of datacenters separate from its retail ones.


Some of AWS’s tools include Amazon S3, or Simple Storage Services, for keeping digital objects for website hosting, and EC2, or Amazon Elastic Compute Cloud, which hosts database servers and server infrastructure

for custom-developed applications, including large-scale supercomputing. Among its noted clients are the television show and movie provider Netflix, software company Adobe Systems, and Thomson Reuters news service.

AWS’s experience hosting data also extended to government. For example, earlier that year the company collaborated with the National Institutes of Health (NIH) to host the world’s largest set of data on human genetic variation, which was collected as part of the 1000 Genomes Project. Organizations and university researchers interested in the data now have easy access; in addition, they can pay to utilize Amazon’s EC2 services, giving them the necessary supercomputing power to pursue scientific objectives.

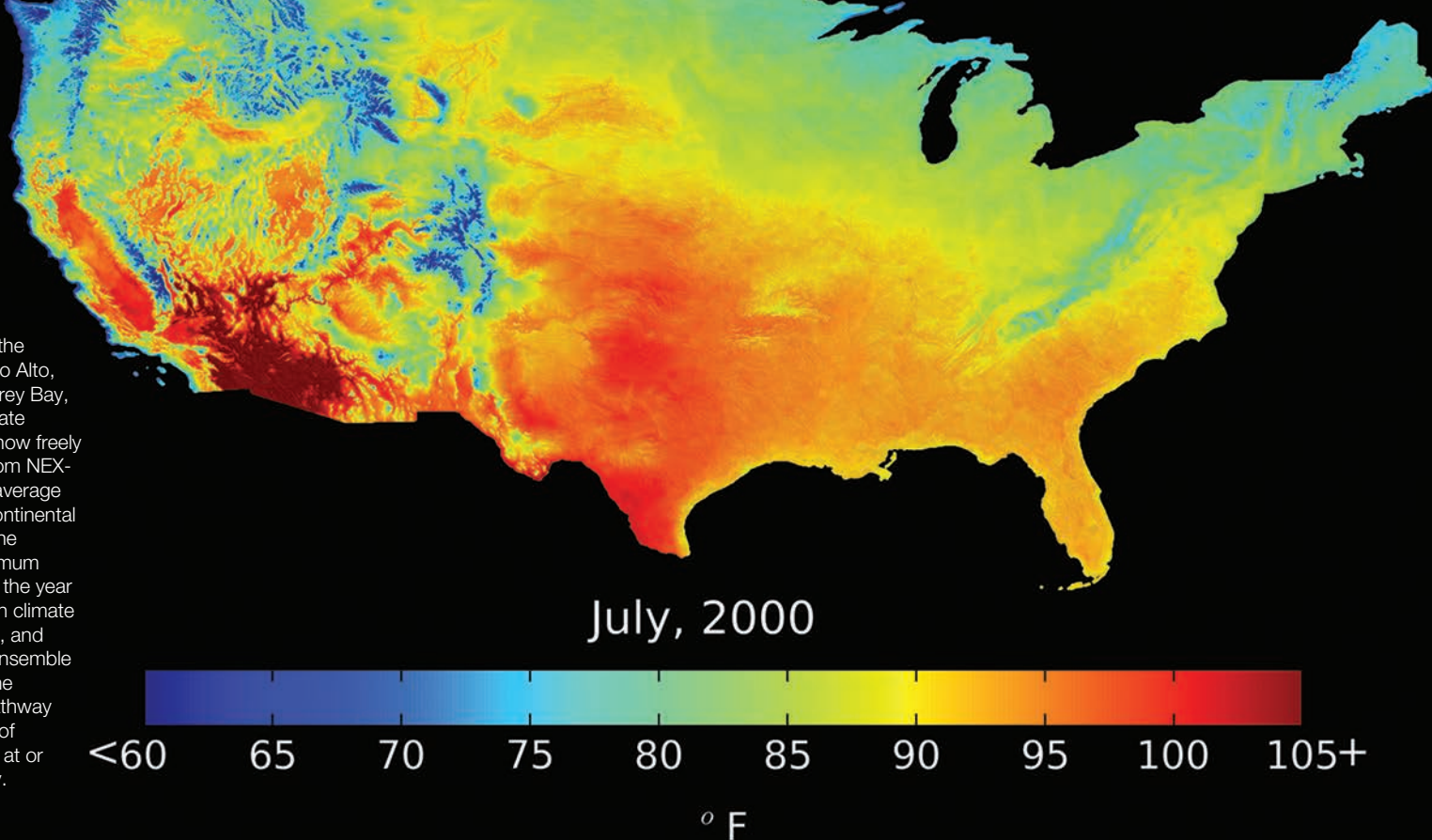
The successful NIH-AWS collaboration gave NASA the impetus to reach a similar agreement with the company because it would improve on NEX in all the ways that Nemani and his team had hoped. For one, NASA-funded researchers would be able to access the datasets directly, bypassing the time-consuming security clearance procedures for accessing the agency’s network. Work would also get done faster because there would be less of a backlog of requests to use Pleiades. Finally, the greater scientific community would also now have access to both NASA datasets and AWS’s supercomputing services.

With those goals in mind, following the signing of a NASA Nonreimbursable Space Act Agreement in November 2013, NASA and AWS announced an initial one-year partnership called OpenNEX, whereby select



Background image: Produced with vegetation indices gathered from the Moderate Resolution Imaging Spectroradiometer (MODIS) on NASA’s Terra satellite, the map displays a multi-year average of leaf surface area to ground surface area in the Amazon rainforest. The darker the shade of green, the thicker the tree canopy. Through NASA’s partnership with Amazon Web Services, researchers will also have access to MODIS data.

Scientists at Ames Research Center, the Climate Analytics Group in nearby Palo Alto, and California State University, Monterey Bay, collaborated on the Downscaled Climate Projections (NEX-DCP30), which are now freely available through OpenNEX. Taken from NEX-DCP30, the image on the left shows average maximum temperatures across the continental US in July 2000, while the image on the right displays projected average maximum temperatures over the same region in the year 2100. Data for 1950-2000 is based on climate simulations driven with observed data, and data for 2010-2100 is based on the ensemble average of all climate models under the highest greenhouse gas emissions pathway evaluated. At present, the global rate of greenhouse gas emissions is tracking at or above this highest emissions pathway.



agency satellite data and climate change datasets would be freely available through the AWS cloud. Researchers could either download the datasets directly to their computers and run their own analysis, or they could access AWS's computing as a service (a pay-as-you-go approach), which comes with free data storage. In addition, through the OpenNEX platform, tens of thousands of researchers—climatologists and geophysicists, among many others—from around the world will have a means to collaborate and share information in order to address the world's most pressing environmental problems.

Data and models hosted on NEX and OpenNEX will aid researchers in their investigations for a plethora of

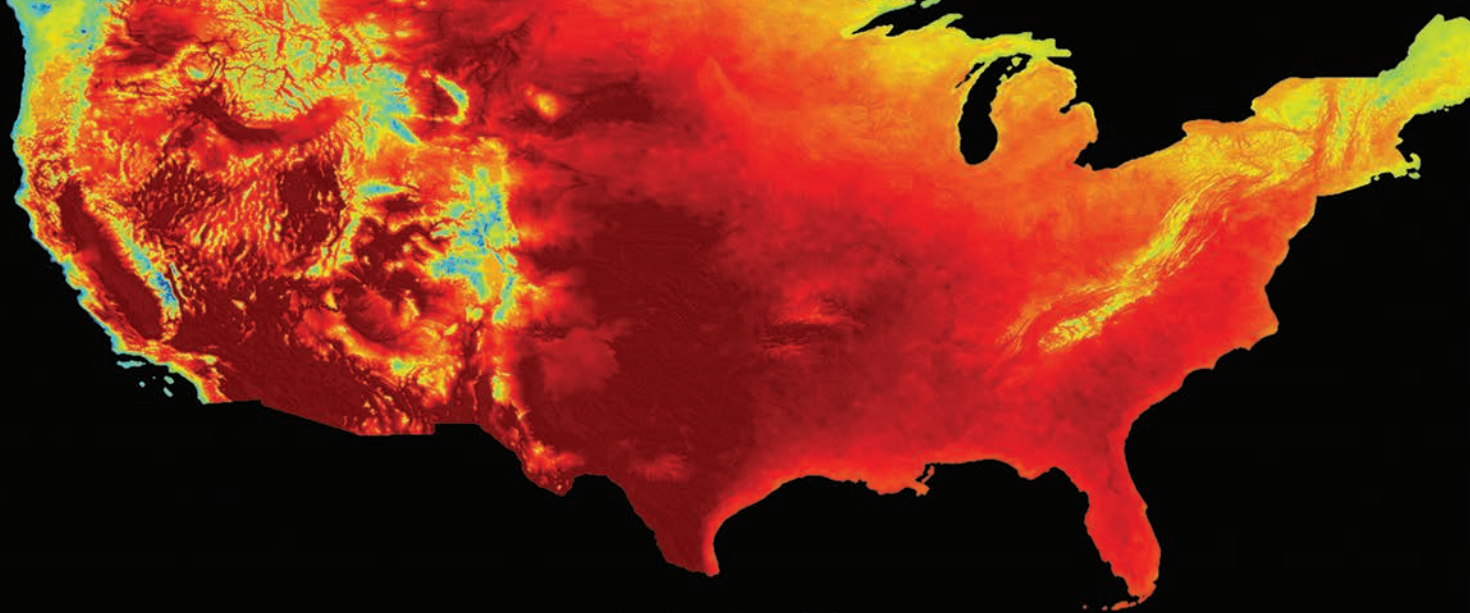
activities, including terrestrial ecology, land use and land cover, carbon cycle science, ecological forecasting, biodiversity, data mining, climate change impacts, and climate change mitigation strategies. Besides the datasets in NEX, the researchers have access to NASA Earth science models, such as TOPS (Terrestrial Observation and Prediction System), the Goddard Earth Observing System Model, Version 5 (GEOS-5) and the NASA-Unified Weather Research and Forecasting (NU-WRF) model.

Benefits

Through the AWS platform, users can currently work with portions of three sizeable datasets: Landsat

Global Land Survey information from the 1970s to 2005, MODIS vegetation indices, and the NEX Downscaled Climate Projections. After the one-year agreement passes, the agency will review the feedback it has received from the scientific community. If the reception is positive, NASA plans on extending the agreement with AWS and will make available additional satellite datasets as well as provide regular updates of those that are currently available.

In exchange for freely hosting the data, AWS benefits from users having the option of purchasing the company's computing services. Jamie Kinney, principal solution architect at AWS, notes that there are two major benefits



July, 2100



Average Daily Max Temperature (° F)
Highest Greenhouse Gas Concentration Pathway

of using AWS for computing: simplicity of use and cost-effectiveness.

“Now anybody who has access to the Amazon cloud, which is publicly available, can easily create an account and, within a few minutes, provision a cluster,” says Kinney. “For a few dollars or maybe 10 or 20 dollars an hour, they can provision a very powerful 25–30-teraflop cluster on Amazon and quickly analyze and visualize that data using the same exact software that NASA researchers are using on internal facilities.”

Moving forward, Kinney says that Amazon is looking forward to continuing its partnership with NASA, not only for increasing business but also because the company

is excited to help bring about future discoveries. “We make investments in public datasets and these types of grants so that we can really help develop the next generation of technologies that will be used by all of our customers down the road, both public sector and commercial,” he says. “So it’s a very natural partnership for us.”

In its short run, Nemani says reception to OpenNEX has been very positive. In June, the agency instituted a series of contests on OpenNEX that challenges citizen scientists to use the available datasets for developing applications and algorithms that promote climate resilience, or our ability to adapt to climate change. The contest falls in line with the White House Big Data and Climate

Data Initiatives, which encourage government agencies to improve their ability to extract knowledge and insight from their digital data collections, as well as President Obama’s calls for developing tools to fight climate change.

“We already have over 400 scientists signed up to be part of the challenges,” Nemani said in July, “so we’re pretty excited to see what they’ll come up with. But keep in mind that this is just the beginning; in a few years we hope to have a large section of the scientific community from around the planet using OpenNEX on a daily basis for climate change research. That’s our big goal.” ❖

