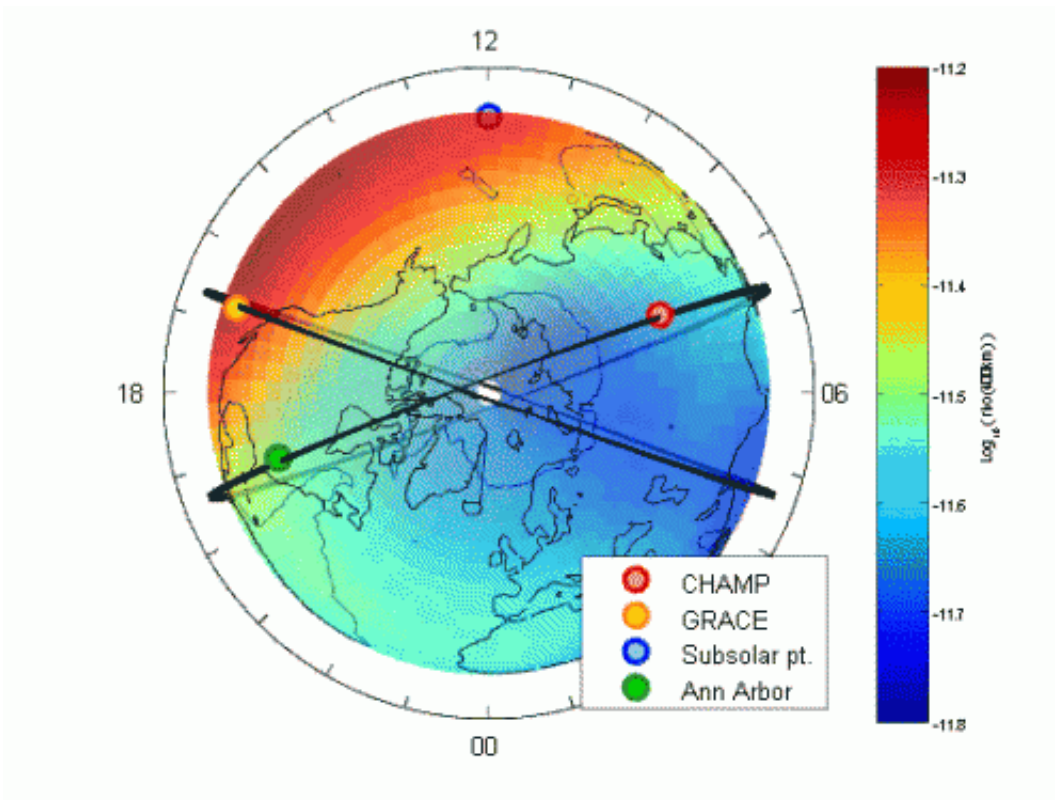


# Atmospheric researchers develop data assimilation tools used throughout science

November 10 2014, by Aaron Dubrow



Solar radiation in the thermosphere (the layer of the Earth's atmosphere directly above the mesosphere and directly below the exosphere) significantly affects the drag experienced by objects like satellites and spacecraft in low-Earth orbit. The fact that the drag changes depending on several factors leads to uncertainty in the position of objects in orbit, which could result in the loss of a spacecraft. Here, we see animation of orbiting satellites responding to the atmosphere. The DART data assimilation reduced up to 70 percent of the bias from the Global Ionosphere-Thermosphere Model. Credit: Alexey Morozov, University of Michigan

A hurricane rages off the coast of Florida. Planes fly into the eye of the storm, capturing details about the speed and structure of the hurricane and beaming this data back to headquarters.

On the coast, sensors draw in data on wave heights. Satellites image inundated neighborhoods. Twitter sentiment analysis tracks growing unease centered on a highway bottleneck. All of this information pours into a central control system and helps guide forecasts and shape evacuation plans.

This is the potential power of data-driven science.

We're not there yet, but increasingly access to streams of observational data is transforming many branches of science. From sequencers and satellites to telescopes and tele-operated drone swarms, massive amounts of data are being collected in ever-new ways. This is offering a means to test and improve existing models that had been developed over decades.

Scientists use predictive models because they cannot experiment with the future climate or a city's highways in a laboratory, the way they do with chemical reactions or cell cultures. But these models are not without errors and a certain degree of uncertainty. Increasingly, researchers are finding that incorporating real-time data into the [model](#) can improve the predictions that models give, in a process called [data assimilation](#).

"This is a common problem, almost a generic problem in science," said Jeffrey Anderson, a senior scientist at the National Center for Atmospheric Research (NCAR) in Boulder, Colorado. "You observe a physical system and then you try to model it, and to do that, somehow you have to relate your model with your observations.

"At the end of the day, the scientific method is about prediction, and data assimilation is this core piece of the scientific method that sort of got ignored for a long time. We really view data assimilation as the tools for confronting models with observations."

## **Data assimilation for all**

Typically, data is analyzed to determine the fundamental way a system—like ocean currents or tornadoes—operates and again to establish the initial conditions—the starting point from which a simulation begins. This assures that the models and simulations accurately reflect the best understanding of the science and the known conditions on the ground.

But such models have typically used static data. Today it's possible to incorporate dynamic, real-time data that offers even more assurance that a model or forecast is realistic.

Data assimilation had been a part of weather prediction since the 1970s, but it was difficult to develop and implement, and laborious to change. As a result, it was only used in the most important and intensive simulations—like official global weather predictions.

In the early 2000s, shortly after arriving at NCAR, Anderson—formerly a climate model developer—began thinking about ways to improve data assimilation and make it accessible to all scientists.

More data was becoming available every day. The scientific community needed a way of using it. But the only data assimilation methods available were entangled and inseparable from the codes that were used at the numerical weather prediction centers.

"I came to NCAR with the idea that we were at a time where, with

proper software engineering techniques and the proper data assimilation algorithms, we could actually build a data assimilation system that could be used with any number of models and any number of observations," Anderson recalled.

He started a data assimilation research section at NCAR called [DARES](#)—a small, data-savvy team that helps Earth scientists incorporate data into their research.

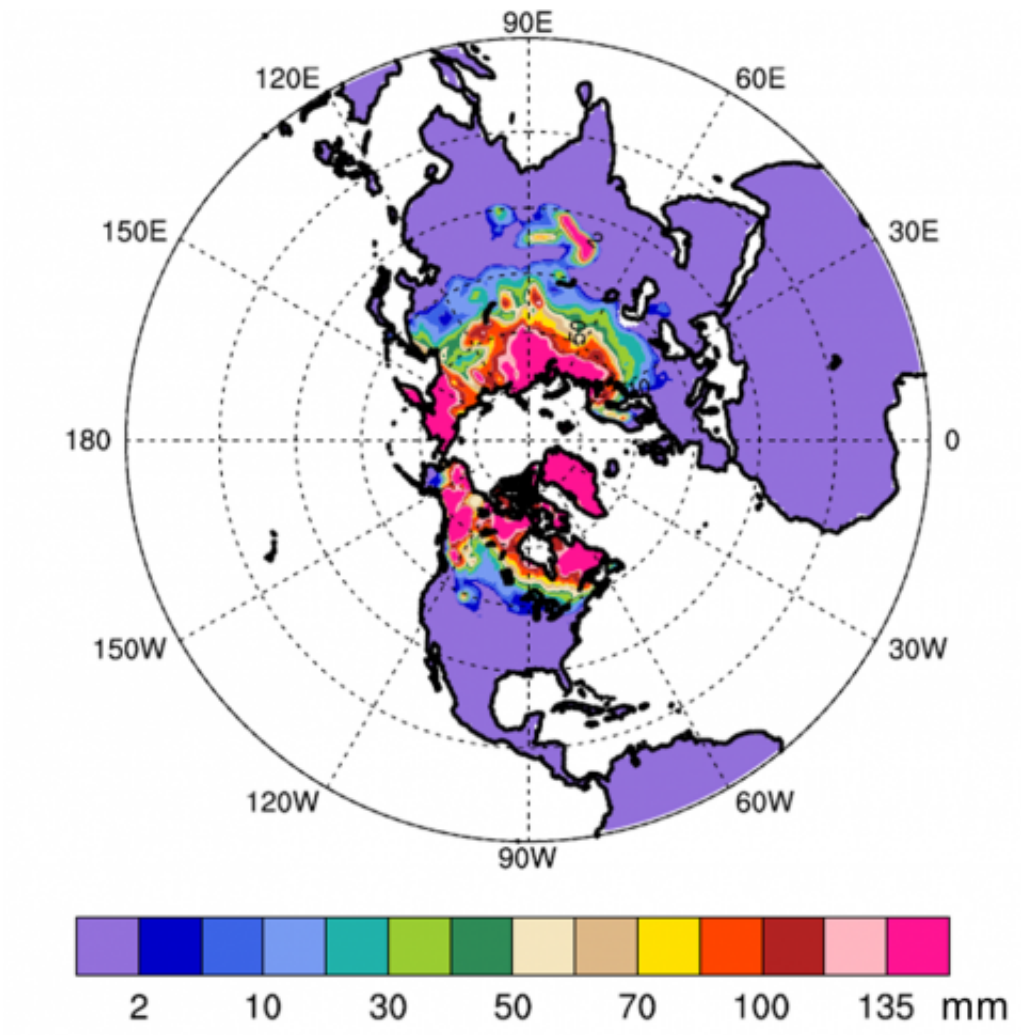
"We really see ourselves as the unsexy member of this triumvirate of models, observations, and the data assimilation that puts the two together," Anderson said.

DARES became that community facility he had envisioned, with software, tools and documentation, plus people offering dedicated, hands-on support.

"We take very seriously NCAR's mission of supporting university scientists, providing them with the tools they need to move their research ahead," he said.

As part of their community-development work, they created a data assimilation tool, called [DART](#) (Data Assimilation Research Testbed), used by more than three-dozen large community codes and hundred of scientists in areas ranging from space debris prediction to ocean currents. Released in 2004, DART continues to evolve and grow.

In several recent journal papers, one can see the impact that DART and data assimilation in general are having on climate, weather and ocean modeling and diverse other research areas.



The Community Land Model version 4 (CLM4) results are used to plot the figure. It shows the amount of water derived from snow pack in Spring 2003 predicted by CLM4, using snow-pack data from 2002. Its spatial pattern is comparable to those observed by satellites. Credit: Yong-Fei Zhang, The University of Texas at Austin

Below are a few "snapshots" of findings enabled by data assimilation.

## Cosmic rays and soil moisture

Scientists are always coming up with new ways of sensing the environment and of adapting those sensors to perform useful functions in society. One such example is the use of cosmic-ray sensors—among them the NSF-funded [COsmic-ray Soil Moisture Observing System](#) (COSMOS) project led by the University of Arizona—to measure soil moisture dynamics at an unprecedented scale.

The sensor measures the number of neutrons at a particular energy level (called "fast neutrons") whose absorption is directly related to the amount of hydrogen in the soil. By removing the effect of additional sources of hydrogen, the sensor can measure soil moisture at between 12 and 76 centimeters (or up to two and a half feet), depending on the water content.

Rafael Rosolem, a lecturer in Water and Environment Engineering at the University of Bristol, has been assimilating measurements from the COSMOS network to improve the performance of land surface models.

"Our collaborative work with NCAR and the University of Arizona showed the benefits of employing data assimilation techniques, such as the suite of algorithms provided by the DART software, to improve simulations of soil moisture using novel technology such as the cosmic-ray sensors available from the COSMOS network," Rosolem said.

The work has implications for future efforts to improve the quality of weather and climate predictions, agriculture monitoring, flood forecasts and drought monitoring. The research was recently published in *Hydrology and Earth System Science*.

## **Snow water resources**

Snow is an important, but not well-understood factor in global climate due to the lack of high-quality datasets.

Zong-Liang Yang, a professor of geoscience at The University of Texas at Austin, and his graduate student Yong-Fei Zhang have been using DART to improve the representation of snow in the land component of the Community Earth Systems Model—an earth system model composed of coupled atmosphere, ocean, land surface, sea ice, land ice and other models, used by the wider climate research community.

The work is part of a multi-institution effort led by UT-Austin, along with NCAR and NASA, focused on developing a global-scale multi-sensor snow data assimilation system.

"DART fits my group's goal of developing a flexible and extensible land data assimilation system," said Yang. "Besides our prototype snow data assimilation, DART is useful for data assimilation involving other variables, such as [soil moisture](#), skin temperature, and leaf area index from various satellite sources and ground observations."

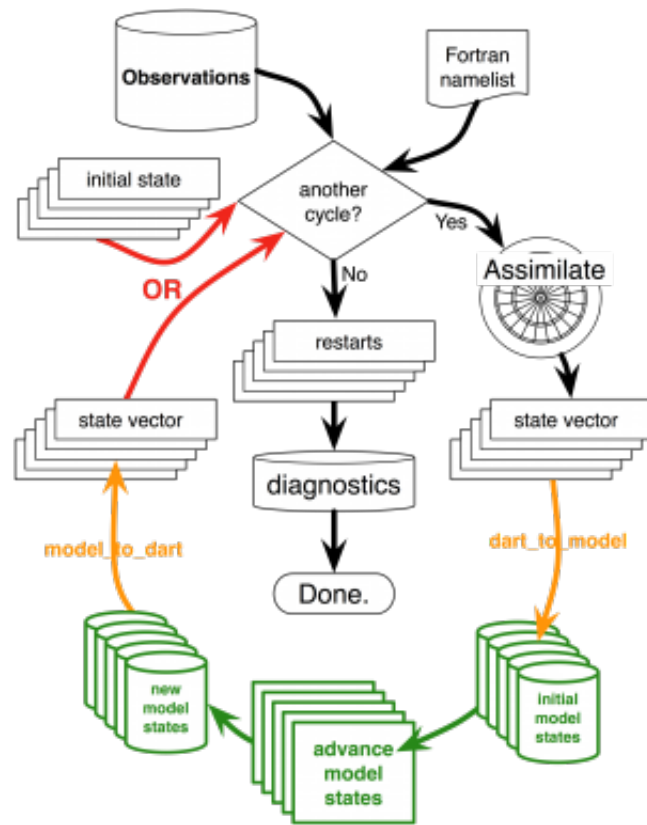
The results of Yang and his team's data assimilation effort were published recently in the *Journal of Geophysics Research: Atmospheres*.

Said Yang, "Such a truly multi-mission, multi-platform, multi-sensor, and multi-scale data assimilation system with DART will, ultimately, help constrain earth system models using all kinds of observations to improve their prediction skills."

## Space debris

Solar radiation in the thermosphere (the layer of the Earth's atmosphere directly above the mesosphere and directly below the exosphere) significantly affects the drag experienced by objects like satellites and spacecraft in low-Earth orbit. The fact that the drag changes depending on several factors leads to uncertainty in the position of objects in orbit, which could result in the loss of a spacecraft.





This is the DART view of ensemble data assimilation for models that run as separate executables. Starting at the top and working clockwise: Everything is driven by a Fortran namelist and the presence or absence of observations. A Fortran executable named 'filter' reads a namelist, an initial state for the ensemble, and a file containing observations and goes to work. Given the observations and an initial state, 'filter' assimilates the observations and then determines how far to advance the model (using information from the namelist and the observation file). 'filter' forks a shell script to the system and it is this shell script that is responsible for three things: 1) for converting the DART state vectors and 'advance\_to\_time' to the format required by the underlying model, 2) advancing the model, and 3) converting the model output into a form suitable for 'filter'. [The script is responsible for the lower portion of the diagram.] The model advances each ensemble member (either in turn or all-at-once) and the model output is converted to the input format expected by 'filter'. The shell script finishes and signals 'filter' to continue. We are now back at the beginning and the cycle continues as long as there are observations to assimilate or until the control information in the Fortran namelist is met. When that happens, a set of



restart files is written (suitable to continue an experiment with more observations) and diagnostic files are written. These diagnostic files allow for the exploration of the assimilation before and after each assimilation step and for exploration of the assimilation in 'observation space'; each real observation is paired with the estimates of the observation from all of the ensemble members (if desired). Minimally, the ensemble mean estimate of the observation and the ensemble spread of the estimates is recorded. Credit: NCAR

One way of decreasing this uncertainty is by obtaining more precise estimates about the neutral density of the atmosphere from thermospheric models. And an effective way of improving the accuracy of these models is via data assimilation.

In a recent paper published in the *Journal of Atmospheric and Solar-Terrestrial Physics*, Alexey Morozov and colleagues from the University of Michigan showed that DART was able to improve the accuracy of the Global Ionosphere-Thermosphere Model (GITM) by assimilating measurements data from CHAMP (Challenging Minisatellite Payload), a German satellite used for atmospheric research.

In their experiments, Morozov and his team used DART's data assimilation and machine learning capabilities to fix holes in GITM and to eliminate a bias they were finding in some simulations.

"We had to get our hands wet at seeing if DART can do a simple thing—push a lever in the right direction to increase the density to match the CHAMP data," said Morozov, who now works at InvenSense, an intelligent sensor company. They found that in some cases, using DART reduced up to 70 percent of the bias from the model.

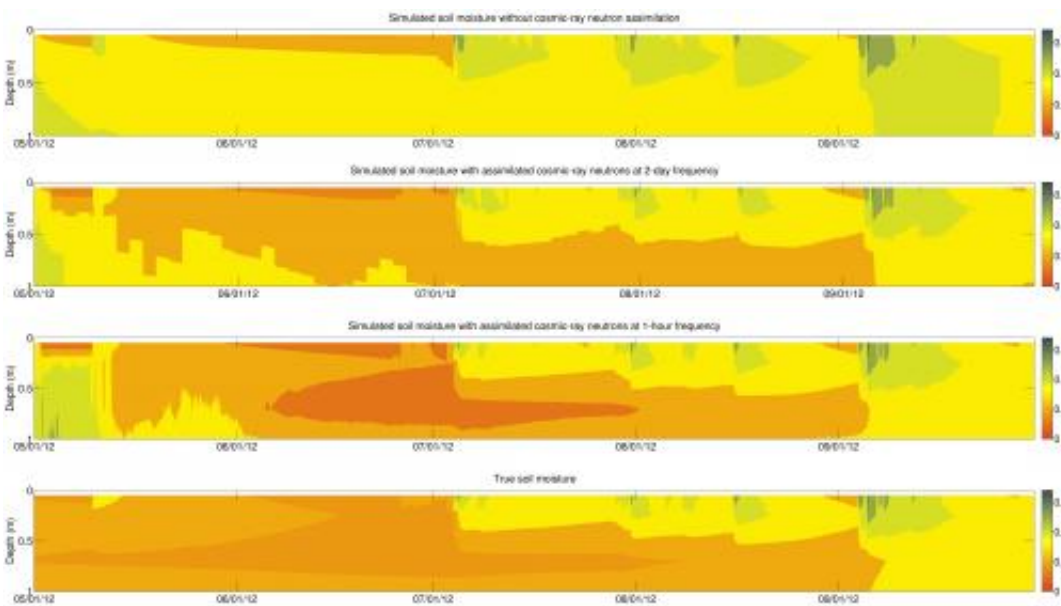
"The space weather research is one of a number of applications where we've let people do science where no one had been able to confront the

models with observations before," said Anderson.

## Assimilating the assimilators

Typically, when a researcher wants to add DART to their code, Anderson invites them to the NCAR campus for a week. There, he and his team work to understand how their code operates and determine how to incorporate DART so that the model—now using dynamic data—produces more accurate results.

There is a downside to data assimilation, however. Assimilating data into a simulation in a statistically-accurate way requires one to run a simulation many times (sometimes up to 60)—a process called ensemble forecasting.



Soil moisture simulations for the Park Falls Ameriflux site. The results show the benefits of assimilating aboveground cosmic-ray neutrons to better constrain soil moisture profile in hydrometeorological models. Assimilating aboveground cosmic-ray neutrons show remarkable improvement in simulated soil moisture

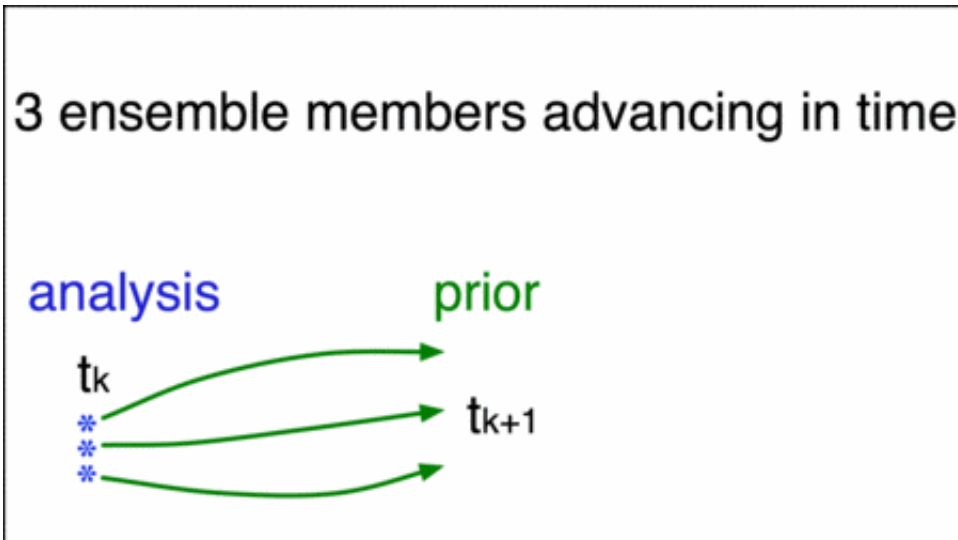
(second and third panels) compared to the case where these measurements are not assimilated (top panel). Furthermore, the high-frequency measurement commonly used in cosmic-ray neutron sensors (i.e., 1-hour frequency; third panel) suggests better agreement with true soil moisture profile compared to a lower-frequency case (i.e., every 2 days) which is common for satellite remote sensing products. Credit: Rafael Rosolem (University of Bristol)

"What we're trying to do is sample from a distribution of these ensemble and then make a forecast," Anderson said. This requires additional computing power, which, for already compute-hungry simulations, can be a challenge to find.

However, these extra runs don't only correct errors in the models. They also provide new information and allow scientists to ask different types of questions.

"Ensemble forecasting offers an opportunity to study the sensitivity of forecasts, for instance, to correlate bad weather in Oklahoma City with winds over New Mexico," he said, citing a recent study his team was involved in.

Another thing data assimilation can do is identify errors in a model or an observing device. When researchers realize that their data isn't matching the model, it provides an opportunity to find systematic problems—say, an out-of-alignment observing satellite or a bug in a forecasting code.



DART is a community facility for ensemble data assimilation developed and maintained by the Data Assimilation Research Section (DAReS) at the National Center for Atmospheric Research (NCAR). DART provides modelers, observational scientists, and geophysicists with powerful, flexible data assimilation tools that are easy to implement and use and can be customized to support efficient operational DA applications. Credit: Jeffrey Anderson

Simulation and modeling are often referred to as the third pillar of science, after theory and experimentation. But some have suggested that data-driven approaches, like the projects powered by DART, are fast becoming a fourth pillar.

With sensors and computer processing getting cheaper and more ubiquitous every year, it's not hard to imagine a world where data is available to an even greater degree than today. With it will come a need to use this data to calibrate models and improve predictions, and, as shown by the recent papers, DART is one effective way to do so.

"We do all this complicated statistics that puts these pieces together to make forecasts, and that's really been hard for the community as a whole

to convey the importance of," Anderson said. "But the rest of these pieces don't fly without good assimilation in the center."

**More information:** "Translating aboveground cosmic-ray neutron intensity to high-frequency soil moisture profiles at sub-kilometer scale." *Hydrol. Earth Syst. Sci.*, 18, 4363-4379, 2014 [DOI: 10.5194/hess-18-4363-2014](https://doi.org/10.5194/hess-18-4363-2014)

Zhang, Y.-F., T. J. Hoar, Z.-L. Yang, J. L. Anderson, A. M. Toure, and M. Rodell (2014), "Assimilation of MODIS snow cover through the Data Assimilation Research Testbed and the Community Land Model version 4," *J. Geophys. Res. Atmos.*, 119, 7091–7103, [DOI: 10.1002/2013JD021329](https://doi.org/10.1002/2013JD021329).

Alexey V. Morozov, Aaron J. Ridley, Dennis S. Bernstein, Nancy Collins, Timothy J. Hoar, Jeffrey L. Anderson, "Data assimilation and driver estimation for the Global Ionosphere–Thermosphere Model using the Ensemble Adjustment Kalman Filter," *Journal of Atmospheric and Solar-Terrestrial Physics*, Volume 104, November 2013, Pages 126-136, ISSN 1364-6826, [dx.doi.org/10.1016/j.jastp.2013.08.016](https://doi.org/10.1016/j.jastp.2013.08.016).

Provided by National Science Foundation

Citation: Atmospheric researchers develop data assimilation tools used throughout science (2014, November 10) retrieved 17 April 2024 from <https://phys.org/news/2014-11-atmospheric-assimilation-tools-science.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.