

Tom Maxwell/NASA-NCCS, UVCDAT

Better Tools to Build Better Climate Models

By Dean N. Williams

A Department of Energy collaboration aims to make climate model development faster and more efficient by creating a prototype of a system for testing model components.

Developing, maintaining, and enhancing a predictive climate model demand enormous human and computing resources. Decades' worth of observational data must be compiled, vetted, and integrated into a database. Parameters and variables must be identified and built into algorithms that simulate physical processes. Massive calculations can then convert past observations into predictions of the future.

To determine the accuracy of predictions, results are validated by comparing them to present-day observations. As new data are fed to the model and scientific understanding of climate systems evolves, new information gets built into the model, and the testing and validation continue.

One of the most resource-intensive aspects of climate modeling is the creation of a system for calibrating climate models, where model simulations are used to validate model output against observational data sets that span the globe. We call this system a "climate model test bed." Such test bed environments typically evaluate each component of the model in isolation, using a skeleton framework that makes the module behave as if it were functioning within the larger program.

To calibrate the model against regional observational data sets, uncertainty quantification techniques assess the accuracy of predictions, given the limitations inherent in the input information.

If model developers could compare test bed output to observational measurements as the output was being generated, the comparison could facilitate aligning the model with the observed data. This capability could eliminate some of the more tedious activities associated with model development and evaluation.

Researchers from five Department of Energy (DOE) laboratories are currently developing this real-time compari-

son capability. If successful, the capability could accelerate the development of climate submodel components, such as atmosphere, land, ocean, and sea ice. It could also improve the process by which the submodels are integrated with one another to form the resulting coupled Earth system climate model.

Leveraging Tools and Building Collaborations

For this effort, which began in mid-2011, the test bed developers fed representative observational data sets—for example, satellite data from NASA's Atmospheric Infrared Sounder (AIRS) and Clouds and the Earth's Radiant Energy System (CERES)—into the specialized model testing and verification platform that they developed. This prototype platform allows for the rapid evaluation of model components and algorithms. A broad goal is to enhance predictive capabilities through the DOE's Biological and Environmental Research (BER) Climate Science for a Sustainable Energy Future (CSSEF) project, which was the sponsor of this work.

Over the past several years, CSSEF team members have collaborated extensively with national and international institutions, universities, and private companies that specialize in data-intensive science and exascale computing to advance scientific model development and evaluation by leveraging state-of-the-art tools. CSSEF's Testbed and Data Infrastructure (TDI) subteam has also worked closely with climate scientists to develop and refine the tools for evaluating model components.

To build the test bed prototype, the CSSEF team has employed DOE's high-performance computing resources to make use of several open-source software projects that are steadily gaining recognition and usage in their respective research communities. In particular, the test bed prototype uses the distributed data archival and retrieval system established under the Earth System Grid Federation [Williams

A three-dimensional view of relative humidity from the surface to an altitude of about 3 kilometers, made using data from the Community Atmosphere Model (CAM). Red represents air parcels at maximum humidity—more water added to the system would cause precipitation. Blue means that an air parcel is 75% of the maximum. Values below 75% are transparent. Height above the Earth is expanded for visibility.

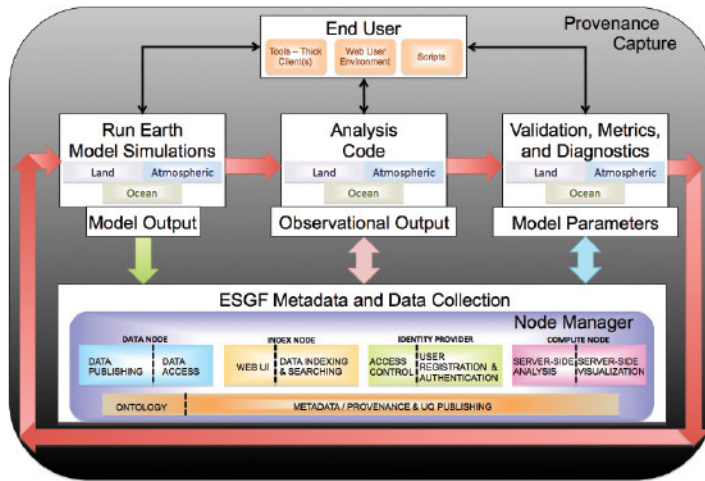


Fig. 1. A high-level conceptual view of the U.S. Department of Energy's Climate Science for a Sustainable Energy Future (CSSEF) test bed architecture and workflow with provenance capture. The red arrows show the baseline ensemble loop in which model simulations are conducted using a variety of input parameters generated by metrics and ensemble drivers. At any stage, data can be collected and stored in the Earth System Grid Federation (ESGF) distributed archive. The black arrows show where desktop clients, Web browsers, or scripts gain access and control the test bed. UI is user interface, and UQ is uncertainty quantification.

et al., 2016] and the Ultrascale Visualization Climate Data Analysis Tools framework [Williams, 2014]. Existing exploratory analysis tools for these databases are also accessible from a Web browser [Steed et al., 2013], allowing the test bed to easily handle incoming data.

The tools and experience resulting from these DOE-sponsored projects provide the foundation for the prototype test bed's infrastructure. Now, through the integration of existing technologies, open standards, and community expertise, the CSSEF team has unveiled a unique and flexible prototype that it hopes will accelerate the development of future climate models.

Incorporating Powerful Provenance Capability

The prototype includes integrated workflows and capabilities for evaluating model provenance, running diagnostics, and examining data analysis and visualization. It also includes automated testing and evaluation.

Provenance, in this context, is the details concerning the setup, execution, and analysis of the model. The test bed prototype captures and archives this information. It also standardizes metadata creation and annotation, and it provides forums for group discussion and sharing. Provenance is of particular interest because it increases scientific and experimental reproducibility, repeatability, productivity, and credibility of collaboration.


CSSEF uses the Provenance Environment (ProvEn) framework for the test bed prototype [Stephan et al., 2013]. Once it is fully imple-

The team has unveiled a unique and flexible prototype that it hopes will accelerate the development of future climate models.

mented within the test bed, ProvEn will provide comprehensive services for the collection and storage of processing provenance including published metadata.

ProvEn correlates computational provenance with knowledge provenance—i.e., newly formed understandings gained from the integration of disparate information—to help scientists browse data, as well as infer and question conclusions. ProvEn will also help scientists mix simulations with observa-

Vienna University of Technology



Vienna Doctoral Programme on Water Resource Systems

The Centre for Water Resource Systems at the Vienna University of Technology announces competition for the third intake of doctoral candidates for the Doctoral Programme on Water Resource Systems. The programme is anticipated to host a total of 70 doctoral students over a period of 12 years. This is a dedicated programme of the Austrian Science Fund (FWF) that promotes doctoral research and education at the highest standards and provides excellent opportunities for cross-disciplinary research. International networking is facilitated by a mobility programme with a spectrum of attractive international partner institutions and a comprehensive guest scientist programme.

Seven PhD student positions are available in the following research themes related to Water Resource Systems:

- Flood-hydrology
- Aquatic microbiology
- Micro-meteorology
- Socio-hydrology
- Environmental economics
- Environmental engineering
- Soil moisture remote sensing
- Mechanics of structures

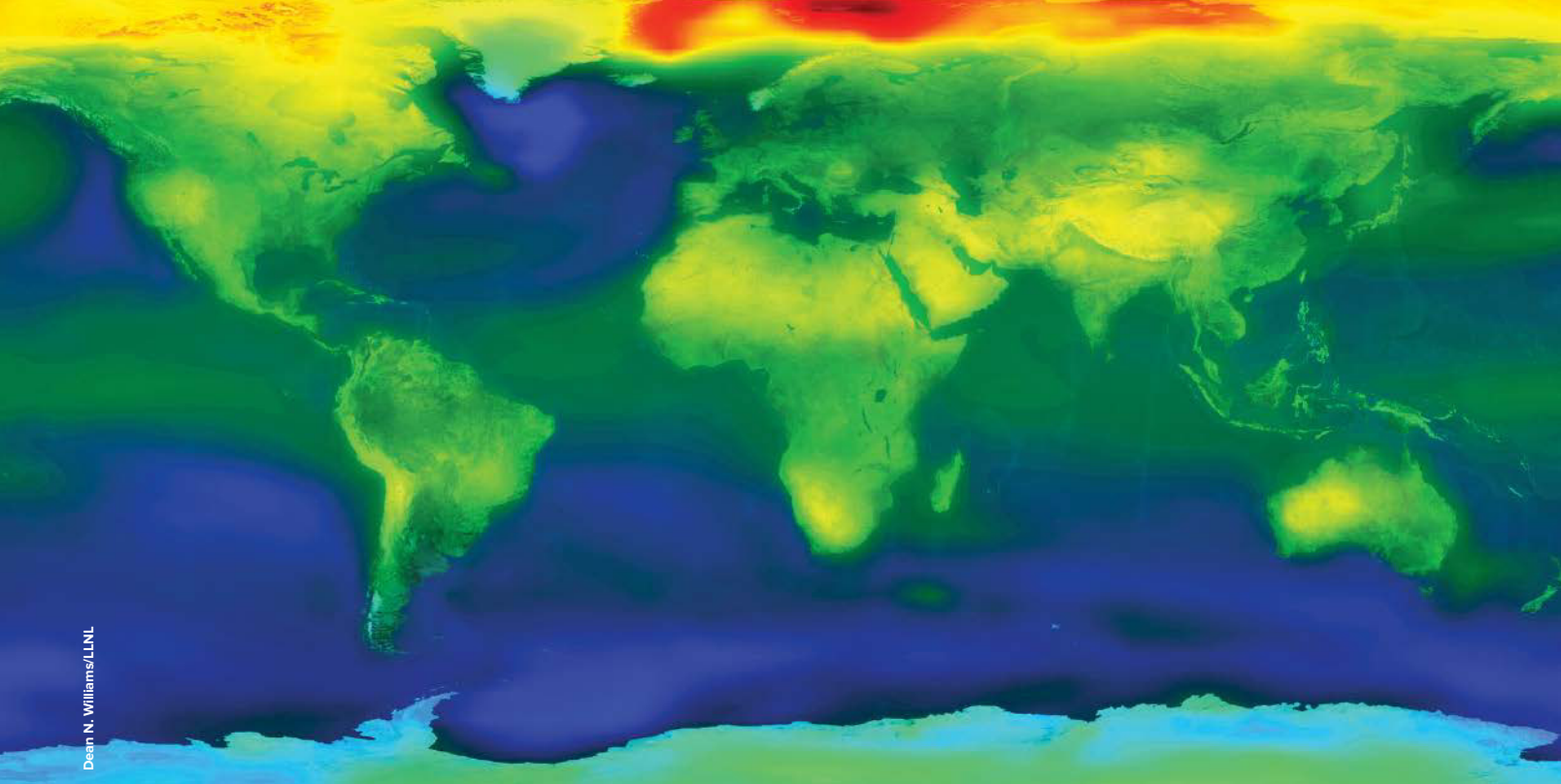
Applicants must have a Master's degree (or equivalent), preferably in a subject related to water resource systems. The working language of the programme is English. Students are expected to work across disciplines and in cooperation with others. A capacity and willingness to integrate and collaborate is essential.

The Programme provides a salary according to the FWF scheme (approx. EUR 20000/year net), together with health and social security benefits. There is also significant allowance for travel and research support. TU Wien is an equal opportunities employer. The preferred starting date is Oct. 1, 2016.

Candidates should send a letter of application, a statement of research interests, a Curriculum Vitae, and copies of education certificates including transcripts of grades as a single .pdf file to: office@waterresources.at.

Application deadline is April 30, 2016. Short listed candidates will be invited to a selection seminar. Financial support towards travel expenses is available on request.

Information about the Doctoral Programme on Water Resource Systems may be viewed at <http://waterresources.at>



Dean N. Williams/LNL

An ensemble of temperature data from 23 climate models from phase 3 of the Coupled Model Intercomparison Project (CMIP3) archive, superimposed over a snippet of the standard world satellite map. Colors represent the average temperature change for the year 2080 compared to current conditions, with dark blue representing about a 1°C change and dark red representing about a 10°C change. Images like these are automatically generated by playing Lawrence Livermore National Laboratory's Interactive Energy & Climate Simulation, a product that undergoes constant evaluation and validation.

tions—through ProvEn, these very different data sets can be compared and harmonized so that the result can better refine models and calculations.

Adaptable Architecture

The CSSEF test bed architecture (Figure 1) allows users to run individual or groups of model components in isolation. The team designed the prototype test bed's infrastructure so that it could be easily customized to users' specific requirements, for example, to test models of ocean dynamics or land cover changes.

The prototype test bed analyzes climate model output and verifies it against observed data sets. Its user interface allows investigators to search and discover scientific data from the entire system (observations, model input, model output), browse data collection hierarchies, download and organize data collection files individually or in bulk, run model components, track deep storage file download requests, and access user profile information.

The CSSEF scientific community is studying the use of the Web browsers and client analysis tools in the prototype test bed, but the CSSEF team has noted the limitations of creating repeatable processes and provenance capturing. For example, multiple sharing of Web and other remote resources often slows the manipulation of data and the sharing of visualization results. Therefore, for Web browser interfaces and remote client analysis tools to work properly, workflow scripts must be well-defined, repetitive computational tasks that integrate existing applications according to a set of rules.

Development Continues Under New Banner

The prototype test bed team is now under the banner of the newly formed Accelerated Climate Modeling for Energy (ACME) project, under the auspices of the U.S. Department of Energy's Office of Science. Under ACME, the team will continue its efforts to deliver an advanced model development, testing, and execution workflow and data infrastructure production test bed for DOE climate and energy research needs. We anticipate rolling out the test bed by the end of 2016 for ACME use.

Acknowledgments

The CSSEF test bed prototype was developed by many scientists from several institutions. They are Ian Foster, Rachana Ananthakrishnan, Eric Blau, and Lukasz Lacinski from Argonne National Laboratory; Renata McCoy, Jeff Painter, Elo Leung, Carla Hardy, Matthew Harris, Charles Doutriaux, and Tony Hoang from Lawrence Livermore National Laboratory; Galen Shipman, John Harney, Chad Steed, Brian Smith, Benjamin Mayer, Marcia Branstetter, and John Quigley from Oak Ridge National Laboratory; Kerstin Kleese-Van Dam, Z e Guillen, Eric Stephan, and Carina Lansing from Pacific Northwest National Laboratory; and Cosmin Safta from Sandia National Laboratory. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344.

References

- Steed, C. A., et al. (2013). Big data visual analytics for exploratory Earth system simulation analysis. *Comput. Geosci.*, 61, 71–82.
- Stephan, E. G., et al. (2013). Bridging the gap between scientific data producers and consumers: A provenance approach, in *Data Intensive Science*, pp. 279–299, CRC Press, Boca Raton, Fla., doi:10.1201/b14935-16.
- Williams, D. N. (2014). Visualization and analysis tools for ultrascale climate data, *Eos Trans. AGU*, 95(42), 377–378, doi:10.1002/2014EO420002.
- Williams, D. N., et al. (2016). A global repository for planet-sized experiments and observations, *Bull. Am. Meteorol. Soc.*, doi:10.1175/BAMS-D-15-00132.1, in press.

Author Information

Dean N. Williams, Computation Division, Lawrence Livermore National Laboratory, Livermore, Calif.; email: williams13@llnl.gov