PERSPECTIVE

# Climate Data Challenges in the 21st Century

Jonathan T. Overpeck,[1]* Gerald A. Meehl,[2] Sandrine Bony,[3] David R. Easterling[4]

**Climate data are dramatically increasing in volume and complexity, just as the users of these data in the scientific community and the public are rapidly increasing in number. A new paradigm of more open, user-friendly data access is needed to ensure that society can reduce vulnerability to climate variability and change, while at the same time exploiting opportunities that will occur.**

Climate variability and change, both natural and anthropogenic, exert considerable influences on human and natural systems. These influences drive the scientific quest for an understanding of how climate behaved in the past and will behave in the future. This understanding is critical for supporting the needs of an ever-broadening spectrum of society's decision-makers as they strive to deal with the influences of Earth's climate at global to local scales. Our understanding of how the climate system functions is built on a foundation of climate data, both observed and simulated (Fig. 1). Although research scientists have been the main users of these data, an increasing number of resource managers (working in fields such as water, public lands, health, and marine resources) need and are seeking access to climate data to inform their decisions, just as a growing range of policy-makers rely on climate data to develop climate change strategies. Quite literally, climate data provide the backbone for billion-dollar decisions. With this gravity comes the responsibility to curate climate data and share it more freely, usefully, and readily than ever before.

## The Exploding Volume of Climate Data

Documenting the past behavior of the climate system, as well as detecting changes and their causes, requires the use of data from instrumental, paleoclimatic, satellite, and model-based sources. The earliest instrumental (thermometer and barometer) records stretch back to the mid- to late 1600s, although widespread land- and ship-based observations were not initiated until the early to mid-1800s, mostly in support of weather forecasting and analysis. Changes in observations through time, due to shifts in observing practices, instrumentation, and land use, have made it necessary to develop and apply advanced data-processing algorithms in order to describe the time evolution of climate. Inevitably, there are uncertainties in the observational records that need to be translated into the degree of confidence associated with our understanding of how the climate system behaves.

In addition to the already large body of digital instrumental data available in diverse holdings around the globe, a substantial number of critical observations, such as many early temperature observations, are not yet widely available as digital records. It is important to create and maintain central repositories of these data in a manner that firmly defines the origin and nature of the data and also ensures that they are freely available (*1*, *2*). In addition, an increasing array of paleoclimatic proxy records from human and natural archives, such as historical documents, trees, sediments, caves, corals, and ice cores, are being generated. These records are particularly helpful in understanding climate variability before the period of instrumental data,

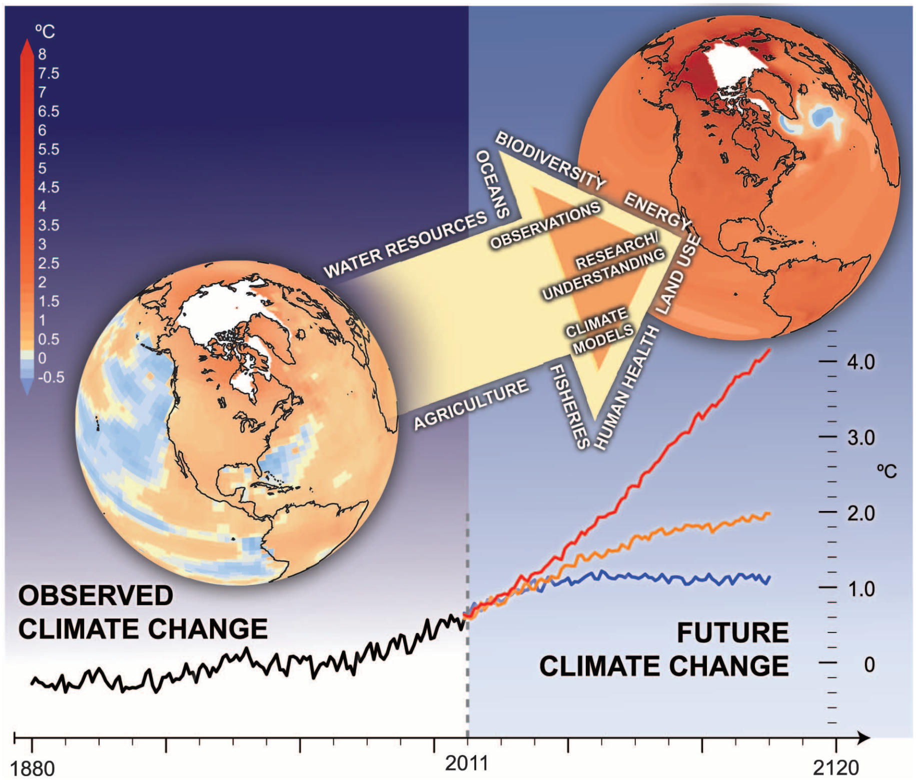


**Fig. 1.** Climate data from observations and climate model simulations are critical for understanding the past and predicting the future. Increasingly, the climate data enterprise must serve both scientist and nonscientist equally well in term of observed (left) and future (right) climate variability and change. Observations, models, research, and understanding are all underpinned by climate data, and all in turn inform uses in society such as those shown surrounding the arrow. The globe at left shows observed annual mean surface temperature anomalies (2006–2010) from the 1951–1980 base period average [NASA data are described in (*20*)]. Arctic sea-ice extent is the 5-year (2006–2010) June-July-August average, excluding sea-ice concentrations less than 10% [NOAA data are described in (*21*)]. The globe at right depicts projected surface temperature anomalies for an example five-member annual mean ensemble average from a climate model (CCSM4), 2081–2100 minus the 1986–2005 average, for the future greenhouse gas and aerosol emission scenario RCP8.5 (*22*). The sea-ice extent from the model is a 5-year (2096–2100) June-July-August ensemble average, excluding sea-ice concentrations less than 10%. The left side of the time series at bottom is annual mean observed globally averaged surface temperatures (*20*), and the right side depicts future projections for five-member ensemble averages from CCSM4 for three emission scenarios (RCP2.6, RCP4.5, and RCP8.5). The magnitude of future climate change depends on what society decides to do now in terms of emissions reductions. Taking little action produces the greatest warming as reflected by the RCP8.5 trajectory, whereas aggressive reductions as represented by RCP2.6 result in stabilized warming at a much lower level.

[1]Institute of the Environment, 845 North Park Avenue, Suite 532, University of Arizona, Tucson, AZ 85721, USA. [2]National Center for Atmospheric Research, Boulder, CO, USA. [3]CNRS, Laboratoire de Météorologie Dynamique, Institut Pierre-Simon Laplace, Université Pierre et Marie Curie, Paris, France. [4]National Oceanic and Atmospheric Administration (NOAA)/ National Climatic Data Center, Asheville, NC, USA.

*To whom correspondence should be addressed. E-mail: jto@email.arizona.edu

over century to millennial time scales, through periods of past abrupt climate change, and during times when climate forcing was substantially different from that of today (*3*)—all critical for understanding what the climate of the future is likely to be. Some of these records have been centrally archived (*4*), but many have not or are described only in isolated references.

Another key source of climate data is spaceborne instruments. The development of long-term, high-quality climate observations from satellites is more difficult than from surface-based instrumental data, because individual satellites and their instruments have short life spans (typically a few years), over which their orbits and sensitivities can change. These problems require the use of advanced data-processing techniques, and the resulting data are prone to being reprocessed as previously unknown problems are discovered over time. In addition, gaps in the records and systematic errors between satellites (or a lack of overlapping calibration periods) make the increasingly important construction of coherent climate data records more of a challenge (*5*).

A third broad type of data is model-based "reanalyses": hybrid model-observational data sets created by assimilating observations into a global or regional forecast model for a given time period (such as 1958 to the present). These provide physically consistent and expanded depictions of the observed time-evolving climate system and have become indispensable in climate system research. The future of reanalysis rests in the establishment of dedicated efforts that include frozen model versions and allow reprocessing of all observational data fields as models and input data sets improve. Future reanalysis methods will include more diverse observational data types (such as atmospheric chemistry, biospheric, oceanographic, and cryospheric data) and longer time scales (including paleoclimatic time scales).

Finally, there has been an explosion in data from numerical climate model simulations, which have increased greatly in complexity and size. Data from these models are expected to become the largest and the fastest-growing segment of the global archive (Fig. 2). The archiving and sharing of output from climate models, particularly those run with a common experimental framework, began in the mid-1990s, starting with output from the early global coupled atmosphere-ocean general circulation models (AOGCMs) used for making future climate change projections (*6*). This led to the Coupled Model Intercomparison Project (CMIP), organized by the World Climate Research Program (WCRP), inviting all the modeling groups to make increasingly realistic simulations of 20th-century and possible future 21st-century climates (*7*–*9*). Recently, CMIP3 involved 16 international modeling groups from 11 countries, using 23 models and submitting 36 terabytes of model data, all archived by the Program for Climate Model Diagnosis and Intercomparison (PCMDI), signaling a "new era in climate change research" (*10*). This activity has made it possible for anyone to openly access these state-of-the-art climate model outputs for analysis and research.

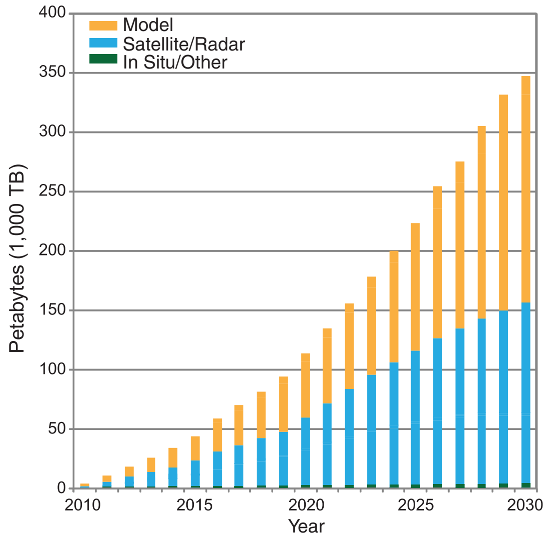


**Fig. 2.** The volume of worldwide climate data is expanding rapidly, creating challenges for both physical archiving and sharing, as well as for ease of access and finding what's needed, particularly if you are not a climate scientist. The figure shows the projected increase in global climate data holdings for climate models, remotely sensed data, and in situ instrumental/proxy data.

Climate model data have been archived and accessed, exchanged, and shared primarily within the physical climate science research community, although there has been growing interest in the use of these climate model data by other communities of researchers. CMIP was designed to provide this broader access to climate model output for researchers from a wide range of communities. The Intergovernmental Panel on Climate Change (IPCC) was also able to use CMIP multimodel data sets to provide state-of-the-art assessments of what the models as a group indicate about possible future climate change (*10*). Now climate models are beginning to be used for much more than climate research. In particular, they are expected to inform decisions that society must take at global to local scales to adapt to natural climate variations as well as to anthropogenic climate change, and to guide the implementation of possible mitigation measures. This puts new demands on the variety, scale, and availability of observational data needed for model evaluation and development, and expands, yet again, the volume of climate data that must be shared openly and efficiently (Fig. 2).

An illustration of the challenges and possibilities posed by the future interaction of fine-scale observational data with more complex models is the evaluation of clouds and the hydrologic cycle. These processes are critical for simulating atmospheric dynamics and regional precipitation, as well as predicting natural climate variability and how much Earth, and local parts of it, could warm for a given amount of greenhouse gas forcing. New high-resolution active remote sensing observations from satellite instruments (such as CALIPSO lidar or CloudSat radar) are revealing the vertical distribution of clouds for the first time. However, to facilitate the comparison of model outputs with these complex new observations effectively, it has been necessary to develop and distribute new diagnostic tools (referred to as "observation simulators") visualizing what these satellites would see if they were flying above the simulated atmosphere of a model (*11*, *12*). Thanks to these developments, it will soon be possible to rigorously assess the realism of cloud simulations in the latest generation of models; for the price of an additional 6% [160 terabytes (TB)] of CMIP5-related climate data that must be shared.

Climate change modeling has evolved in just 5 years from running a few AOGCM experiments with a single category of model, to running many more experiments with a much larger profusion of models of increasing resolution and complexity. First-generation Earth system models (ESMs) are now being run as part of the current CMIP5 exercise (*13*, *14*). ESMs include at least an interactive carbon cycle coupled to the traditional AOGCMs, which have atmosphere, ocean, land, and sea-ice components. Also, high-resolution climate models (such as those with 20-km grid spacing) are run for time slices, past and future, for integrations of a decade or two in order to obtain a better quantification of regional climate change and smaller-scale phenomena such as hurricanes [for example, see (*15*)]. The net result is a huge increase in data volume (Fig. 2). Early phases of the CMIP project involved less than 1 TB of model data, whereas CMIP3 archived 36 TB, and CMIP5 is expected to make available 2.5 petabytes (PB). New capabilities of the Earth System Grid portal will provide distributed access to a large part of this new model output (*16*), making it possible for modeling groups to share data from distributed local servers with Web-based access tools. Model data thus do not need to be centrally archived but can be accessed in a distributed fashion. Clearly, this is an example to be followed more broadly, with the caveat that the safety and reliability of long-term archives of these data must not be jeopardized.

### Meeting the Needs of a Wide Range of Users

The burgeoning types and volume of climate data alone constitute a major challenge to the climate research community and its funding bodies.

Institutional capacity must exist to produce, format, document, and share all these data, while, at the same time, a much larger community of diverse users clamors to access, understand, and use climate data. These include an ever-increasing range of scientists (ecologists, hydrologists, social scientists, etc.) and decision-makers in society who have real money, livelihoods, and even lives at stake (resource managers, farmers, public health officials, and others). Key users also include those with public responsibilities, as well as their constituents in the general public who must support and understand decisions being made on their behalf. As a result, climate scientists must not only share data among themselves, but they must also meet a growing obligation to facilitate access to data for those outside their community and, in doing so, respond to this broader user community to ensure that the data are as useful as possible.

In addition to the latest IPCC assessment, various ongoing national climate assessment and climate services activities are being initiated that will need to access and use climate data, just as a growing number of climate adaptation and mitigation efforts around the globe will need to be better informed by climate data. These efforts will succeed only if climate data are made readily accessible in forms useful to scientists and nonscientists alike.

### The Future of Climate Data: An Emerging Paradigm

Thus, two major challenges for climate science revolve around data: ensuring that the ever-expanding volumes of data (Fig. 2) are easily and freely available to enable new scientific research, and making sure that these data and the results that depend on them are useful to and understandable by a broad interdisciplinary audience. A new paradigm that joins traditional climate research with research on climate adaptation, services, assessment, and applications will require strengthened funding for the development and analysis of climate models, as well as for the broader climate data enterprise. Increased support from the funding agencies is needed to enhance data access, manipulation, and modeling tools; improve climate system understanding; articulate model limitations; and ensure that the observations necessary to underpin it all are made. Otherwise, climate science will suffer, and the climate information needed by society—climate assessment, services, and adaptation capability—will not only fall short of its potential to reduce the vulnerability of human and natural systems to climate variability and change, but will also cause society to miss out on opportunities that will inevitably arise in the face of changing conditions.

At present, about half of the international modeling groups are restricted from sharing digital climate model data beyond the research community because of governmental interest in the sale of intellectual property for commercial applications; the same holds true for some observational data. Open and free availability of model data, observations, and the software used for processing is crucial to all aspects of the new paradigm. Governments that currently restrict either model output or observed data distribution must be convinced that it is in the best interests of everyone that all climate data be made openly available to all users, including those engaged in research and applications. International agreements must eliminate data restrictions, just as journals and funding agencies should require easy access to all data associated with the papers they publish and the work they fund.

The optimal use of climate data requires a more effective interdisciplinary communication of data limitations with regard to, for example, spatial and temporal sampling uncertainties; instrument changes; quality-control procedures; and, in particular, what model-based climate predictions or projections do well and not so well. The first step is to increase the accessibility of observations and high-resolution simulations to a wide range of users, either via a few centralized portals (such as PCMDI) with broader responsibilities, or using a more decentralized approach. A second step is to develop an international depository site for model diagnostic tools (such as satellite simulators) and evaluation metrics that would help users assess the reliability of specific aspects of model simulations (such as sea ice, El Niño–Southern Oscillation, or monsoons, droughts, and other climate extremes). The key is that new data-sharing systems have to be evaluated and improved until all types of interdisciplinary users are able to be effective partners in the use of climate data.

An increasingly daunting aspect of having tens and eventually hundreds of petabytes of climate data openly available for analysis (Fig. 2) is how to actually look at and use the data, all the while understanding uncertainties. More resources need to be dedicated to the development of sophisticated software tools for sifting through, accessing, and visualizing the many model versions, experiments, and model fields (temperature, precipitation, etc.), as well as all of the observed data that is online. In parallel, it is becoming increasingly important to understand complex model results through a hierarchy of models, including simple or conceptual models (*17*). Without this step, it will be extremely difficult to make sense of such huge archived climate data sets and to assess the robustness of the model results and the confidence that may be put in them. Again, fulfilling the needs of all types of interdisciplinary users needs to be the metric of success.

Increasingly, climate scientists and other types of scientists who work effectively at the interface between research and applications are working closely together, and even "coproducing" knowledge, with climate stakeholders in society (*18*, *19*). These stakeholders, along with the interdisciplinary science community that supports them, are the users that must drive the climate data enterprise of the future.

### References and Notes


1. C. K. Folland *et al.*, in *Temperature Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences*, T. R. Karl, S. J. Hassol, C. D. Miller, W. L. Murray, Eds. (U.S. Climate Change Science Program and the Subcommittee on Global Change Research, Washington, DC, 2006), pp. 119–127.
2. D. R. Easterling *et al.*, in *Weather and Climate Extremes in a Changing Climate. Regions of Focus: North America, Hawii, Carribbean, and U.S. Pacific Islands*, T. R. Karl *et al.*, Eds. (U.S. Climate Change Science Program and the Subcommittee on Global Change Research, Washington, DC, 2008), pp. 117–126.
3. E. Jansen *et al.*, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon *et al.*, Eds. (Cambridge Univ. Press, Cambridge, 2007), pp. 433–497.
4. Archived at the World Data Center for Paleoclimatology, www.ngdc.noaa.gov/wdc/usa/paleo.html.
5. National Research Council, *Climate Data Records from Environmental Satellites* (National Academy Press, Washington, DC, 2004).
6. G. A. Meehl, G. J. Boer, C. Covey, M. Latif, R. J. Stouffer, *Eos* **78**, 445 (1997).
7. G. A. Meehl, G. J. Boer, C. Covey, M. Latif, R. J. Stouffer, *Bull. Am. Meteorol. Soc.* **81**, 313 (2000).
8. G. A. Meehl, C. Covey, B. McAvaney, M. Latif, R. J. Stouffer, *Bull. Am. Meteorol. Soc.* **86**, 89 (2005).
9. C. Covey *et al.*, *Global Planet. Change* **37**, 103 (2003).
10. G. A. Meehl *et al.*, *Bull. Am. Meteorol. Soc.* **88**, 1383 (2007).
11. J. M. Haynes, R. T. Marchand, Z. Luo, A. Bodas-Salcedo, G. L. Stephens, *Quickbeam. Bull. Am. Meteorol. Soc.* **88**, 1723 (2007).
12. H. Chepfer *et al.*, *Geophys. Res. Lett.* **35**, L15704 (2008).
13. K. A. Hibbard, G. A. Meehl, P. Cox, P. Friedlingstein, *Eos* **88**, 217 (2007).
14. K. E. Taylor, R. J. Stouffer, G. A. Meehl, A summary of the CMIP5 Experimental Design (2009); www-pcmdi.llnl.gov/.
15. K. Oouchi *et al.*, *J. Meteorol. Soc. Jpn.* **84**, 259 (2006).
16. D. N. Williams *et al.*, *Bull. Am. Meteorol. Soc.* **90**, 195 (2009).
17. I. M. Held, *Bull. Am. Meteorol. Soc.* **86**, 1609 (2005).
18. M. C. Lemos, B. J. Morehouse, *Global Environ. Change Hum. Policy Dimensions* **15**, 57 (2005).
19. R. S. Pulwarty, C. Simpson, C. R. Nierenberg, in *Integrated Regional Assessment of Global Climate Change*, C. G. Knight, J. Jäger, Eds. (Cambridge Univ. Press, Cambridge, 2009), pp. 367–393.
20. J. Hansen, R. Ruedy, M. Sato, K. Lo, *Rev. Geophys.* **48**, RG4004 (2010).
21. R. W. Reynolds, N. A. Rayner, T. M. Smith, D. C. Stokes, W. Wang, *J. Clim.* **15**, 1609 (2002).
22. R. H. Moss *et al.*, *Nature* **463**, 747 (2010).


23. The authors thank G. Strand [National Center for Atmospheric Research (NCAR)] and S. Veasey (National Climatic Data Center) for their contributions. NOAA supported this work (J.T.O.) through its Regional Integrated Sciences and Assessments Program. Portions of this study were also supported (G.A.M.) by the Office of Science (B.E.R.), U.S. Department of Energy, Cooperative Agreement no. DE-FC02-97ER62402; and NSF. NCAR (G.A.M.) is sponsored by NSF.

10.1126/science.1197869

**Climate Data Challenges in the 21st Century**
Jonathan T. Overpeck, Gerald A. Meehl, Sandrine Bony and David R. Easterling (February 10, 2011)
*Science* **331** (6018), 700-702. [doi: 10.1126/science.1197869]