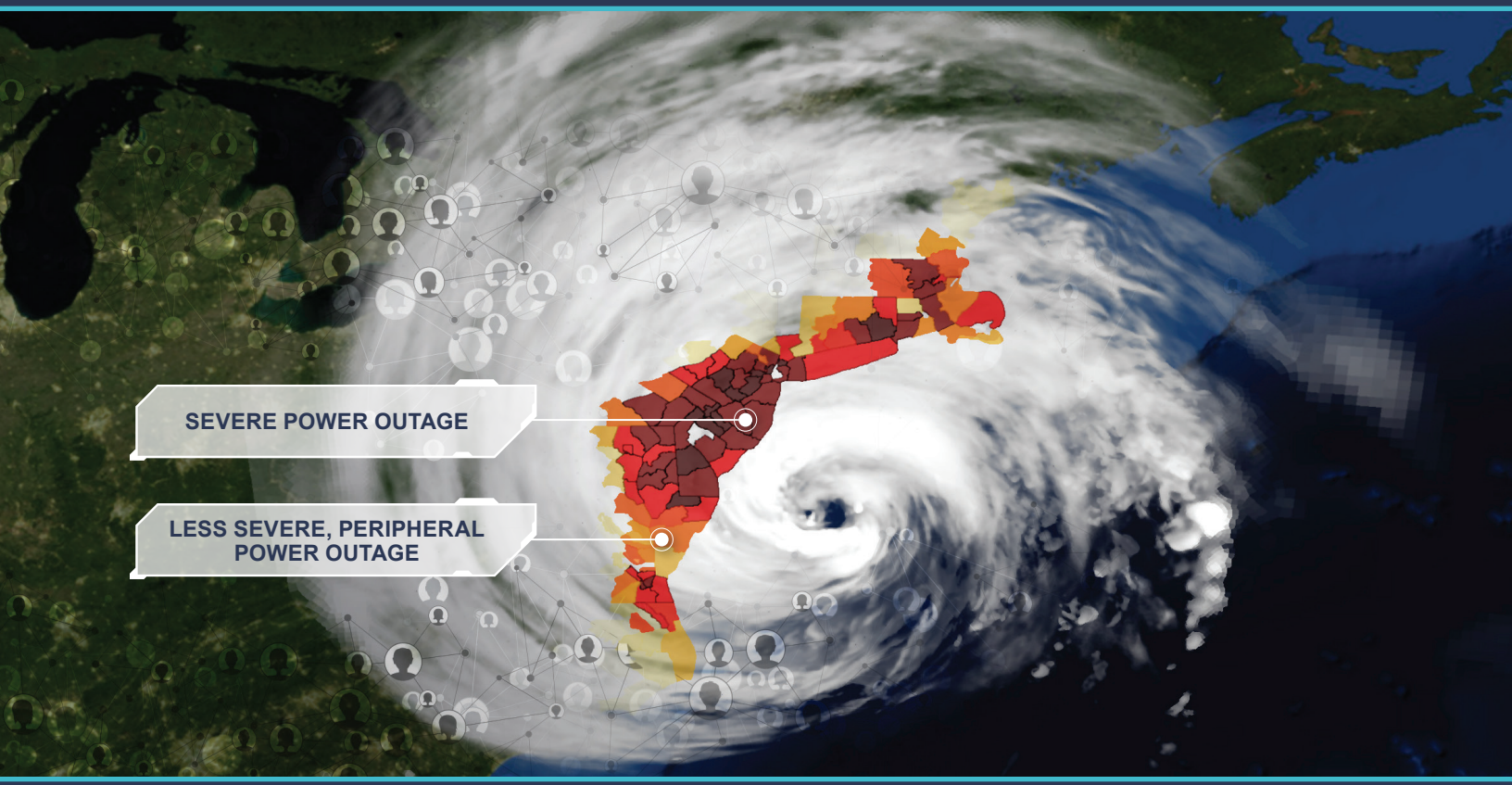
DOE/SC-2021

Understanding Decision-Relevant Regional Climate Data Products

WORKSHOP REPORT | EXECUTIVE SUMMARY



U.S. Department of ENERGY | Office of Science

Biological and Environmental Research Program

October 2024

Executive Summary

A broad community of climate adaptation practitioners, stakeholders and policymakers rely on historical reconstructions and future projections of local to regional climate. To be of value to these users, climate data must be credible, salient, and authoritative (Cash et al. 2002). Namely, data must be consistent with our physical understanding of the global Earth system, must be relevant for informing the decision-making process, and must be backed by expert judgment. As more and more data products have become available, multiple challenges have emerged around the production, evaluation, selection, and use of these data products. Consequently, to ensure crucial decisions leverage the best possible historical and future physical climate data, there is a pressing need to develop a coordinated national climate data strategy that is inclusive of all relevant communities of practice.

In response to this need, the "Understanding Decision-Relevant Regional Climate Data Projections" workshop was held in-person and virtually from November 14-16, 2023, in Berkeley, California. This workshop was coordinated by the U.S. Global Change Research Program (USGCRP) Interagency Group on Integrative Modeling (IGIM) and the Federal Adaptation and Resilience Group (FARG). Participation came from most major U.S. federal agencies and their partners who are involved in the production and dissemination of regional climate data products, including the U.S. Department of Energy (USDOE), the National Oceanic and Atmospheric Administration (NOAA), the National Aeronautics and Space Administration (NASA), the U.S. Environmental Protection Agency (EPA), Federal Emergency Management Agency (FEMA), U.S. Geological Survey (USGS), U.S. Bureau of Reclamation (USBR), and the Department of Defense's (DOD) Strategic Environmental Research and Development Program and Environmental Security Technology Certification Program (SERDP and ESTCP). The workshop brought together a wide range of researchers, data producers, end-users, and interagency representatives to understand the current state of the nation's decision-relevant regional climate projections and carry that understanding forward to enable the development of guidelines for the usage and evaluation of such projections. Numerous approaches for generating regional climate data were discussed, including statistical downscaling, dynamical downscaling, hybrid downscaling, regionally refined global modeling and artificial intelligence. This effort provided a forum for sharing knowledge, establishing common ground, and moving towards the development of a community of practice around decision-relevant data.

The workshop was organized into four sessions focused on 1) data production; 2) data use; 3) data evaluation; and 4) emerging topics. The session on data production featured 10 talks from a variety of data producers, representing multiple federal agencies and academic research groups, followed by breakout sessions that sought to frame the needs of a community of practice. The session on data use featured two panels, each with four panelists presenting brief talks on topics related to how they employ climate data and their perceptions of gaps among existing data products. The session on data evaluation again featured two panels, each with four panelists presenting brief talks related to ascertaining credibility of climate data. The final session on emerging topics featured 11 technically oriented talks on topics related to climate data, including bias correction, model weighting, ensembles, and performance across scales. For each of the first three sessions, there was an accompanying breakout discussion which featured a mix of participants who addressed key questions related to that session and the context of the broader workshop theme.

This executive summary provides a high-level synthesis of discussions at the workshop, focusing on the outstanding challenges identified during the workshop and also potential deliverables from a nascent community of practice to address these challenges.

Challenges for the Decision-Relevant Climate Data Community

Building a common vocabulary: In the course of the workshop, it became clear that a common vocabulary across related communities is needed. Terms such as "extreme event" can have different meanings depending on the needs of particular end-users and the impacts they are considering. The workshop itself was framed around "decision-relevant" or "actionable" data products, but the regional extent, spatial resolution, and temporal resolution for a product to be considered decision-relevant varies depending on the decision being made. "Uncertainty" and "confidence" also emerged as terms that are widely employed in the climate data space, but precise, quantitative definitions of these terms are rarely provided.

Filling data gaps: Despite rapid growth in the number of climate data products, conspicuous gaps remain. For instance, although some statistically downscaled products have global coverage, higher resolution coverage of areas outside the contiguous United States (OCONUS), including Alaska, Puerto Rico, and island territories, is still needed. Additionally, few high-temporal-resolution (hourly to sub-hourly) data products are available even in the contiguous United States (CONUS) (not to mention OCONUS), despite being needed for many applications (e.g., evaluating sufficiency of storm sewers and projections of renewable energy production). Many opportunities exist for addressing these gaps through new simulations or innovative downscaling methods.

Cataloging and characterizing decision-relevant climate data products: Dozens of regional climate data products have emerged in the past decade at local-to-global scales. They exhibit a variety of spatial and temporal resolutions and feature a variety of climate variables. However, in the absence of a central catalog of data products, end-users and researchers have largely relied on word of mouth and Internet searches to identify relevant data products. Consequently, other, equally relevant products have likely been underused or unused. A catalog of data products, their characteristics, relevant expert guidance, and evaluation metrics could benefit all members of the climate data community, and enable the identification of gaps and synergies among presently available products.

Provisioning common-format, decision-relevant climate data products: Related to the aforementioned challenge of cataloging these products, additional challenges exist in provisioning these data. Three bottlenecks generally stymie data producers interested in provisioning their data to a broader audience: firewalls at the data source, access restrictions and data provisioning support requirements. The sheer size of these data products creates provisioning challenges that are generally beyond the scope of the data producer's expertise and bandwidth. Data archiving and distribution portals, such as the Earth System Grid Federation and the National Center for Atmospheric Research Data Archive, have greatly accelerated science through the provisioning of relevant climate data sets. However, more archival systems (and/or the expansion of existing portals) are needed to support the variety of products currently being used across the community. Opportunities exist for leveraging cloud services and/or server-side compute to potentially address these needs.

Avoiding redundancy and leveraging limited computational resources: Production of climate data products, particularly high-resolution products generated from process-based models, generally requires extensive computational resources and substantial human investments of time and effort to both run the models and archive the data. Facilitating better communication among data producers could identify needs that are addressable through coordinated

simulations and make better use of existing computational resources. For example, better lines of communication could make an air quality modeler aware of community needs for wind power projections, and subsequently lead them to include high-frequency hub-height wind speeds as a model output. Additionally, the aforementioned identification of gaps among existing products could allow the community to identify high-priority simulations that have the broadest potential value.

Developing expert analysis and insights for data users: The choice of climate data products employed for decision-making is often based on existing research networks or data availability. In general, there is little guidance available to end-users on whether these products and their associated parent climate models meet a minimum standard of quality for their purposes. Community-developed and supported templates for metadata, which could include criteria for data documentation and licensing, along with requirements for publication of metric scores from an established and community-support evaluation protocol, and guidelines on best practices and/or pitfalls for parent climate model and data product averaging and weighting, would be helpful for informing decision-makers and building confidence in those data products. This information would further support data selection for widely-used, government led community activities, such as the National Climate Assessment and National Nature Assessment. An increasing focus of the climate data community on "scientific co-production" has also highlighted the increasing need for researchers and end-users to work together to address relevant knowledge gaps, and suggests that efforts should be made to identify questions about climate data products of greatest importance for decision-makers.

Continuously improving observational (training) data products: High-quality observational data sets underpin any climate data product. Observational data products are constructed through various means, generally from meteorological station, airborne, or satellite observations or a combination thereof. The need for continuous improvement arises from the sheer number of unconstrained choices made to develop a product in terms of gridding and/or managing data outages, changes in measurement technology, instrument relocation, and other requirements to produce (with or without homogenization) long-term, high-temporal-and-spatial-resolution fields. In addition to improving these observational products, uncertainties around these products need to be quantified since they can translate to corresponding uncertainties in future impact projections.

Nurturing a cohesive regional climate data product community to address these challenges: The November 2023 Workshop was not the first workshop to address regional climate data issues. Many of the topics discussed echoed themes of previous workshops, but none of them resulted in a sustained, coordinated set of research activities to address long-standing, and more importantly, growing challenges with regional climate data and their connections to decision-making. The lack of a cohesive community to address the challenges discussed in the November 2023 Workshop was glaringly apparent. Workshop participants concluded that a follow-up workshop in 12-18 months would allow us to ascertain progress and plan for the future.

Research Needs for the Decision-Relevant Climate Data Community

The workshop concluded that substantial near-term progress could be made in addressing the eight challenges above, and laid out several potential deliverables that could also support longer-term improvements.

A community of practice: Conversations at the workshop highlighted the pressing need to ensure that lines of communication remain open between data producers, evaluators, and end-users. A community of practice, involving regular meetings and other means of facilitating communications between affiliated parties, would allow the climate data community to evolve to meet the ever-changing needs of this space. Beyond improving communication, we need to set forth a governance structure, a scope of activities, and incentives. These are critical for ensuring cohesion for the nascent community so that it can achieve and regularly measure its progress on the challenges identified in the workshop.

A common format for decision-relevant climate data: Early efforts by the Coupled Model Intercomparison Project (CMIP) led to the creation of a metadata standard and set of common variable names that would ensure interoperability of model data. Researchers have benefited greatly from this foresight, as most analysis tools and workflows can now be rapidly applied to model outputs, whether they be from Europe, Asia, or North America. Publicly-available tools such as the Climate Model Output Rewriter (CMOR) allow operational modeling centers to convert their native model outputs to data that conforms to a community standard. However, these practices have not been widely adopted by the regional climate modeling community, leading to workflows typically tailored to a particular data product. A common framework, decided upon by the climate data community, that specifies file format, metadata requirements and variable naming conventions would accelerate the usefulness of decision relevant climate data.

A common framework for climate data product evaluation: Quantification of the performance of climate data products is an important step in ascertaining confidence in the data for decision support. With no commonly accepted standards for climate data evaluation, it is difficult to compare climate data products and understand issues that may support or preclude their use. Consequently, there is a substantial and outstanding need for a community-developed framework for decision-relevant climate data product evaluation that leverages observation datasets and physical principles. Such a framework would identify and prioritize metrics, diagnostics, and other criteria relevant to the credibility of the data product. Providing accompanying expert guidance would assist in explaining observed differences between data and observations, and support the development of new strategies for climate data generation. This framework must also accommodate and navigate the differences inherent in the different types of downscaling and bias correction approaches.

Climate data cyberinfrastructure: Cyberinfrastructure to support the climate data community could include a maintained catalog of climate data products, disk space, and bandwidth to support archiving and provisioning of climate data and a computing platform for server-side analysis of climate data. Coordination among agencies could avoid redundant investments, ensure greater sharing of data, and allow users to avoid difficulties associated with accessing data through multiple platforms

Cover description:

A hypothetical tropical cyclone makes landfall near Philadelphia, as simulated using the Energy Exascale Earth System Model (E3SM) in its Simple Cloud Resolving E3SM Atmosphere Model (SCREAM) configuration on a regionally refined grid. Simulation by Colin Zarzycki, Pennsylvania State University. Visualization by Paul Ullrich, Lawrence Livermore National Laboratory and University of California Davis.

Recommended Citation: U.S. DOE. 2024. *Understanding Decision-Relevant Regional Data Products Workshop Report*, DOE/SC-2021, Department of Energy's Biological and Environmental Research Program, U.S. Department of Energy, Office of Science. **https://doi.org/10.2172/2474992**

U.S. Department of ENERGY | Office of Science

Biological and Environmental Research Program