ANL-22/54

Artificial Intelligence for Earth System Predictability (AI4ESP)

2021 Workshop Report

Environmental Science Division

ANL-22/54

Artificial Intelligence for Earth System Predictability (AI4ESP)

2021 Workshop Report

Prepared by
Nicki Hickmon[1], Charuleka Varadharajan[2], Forrest Hoffman[3], Haruko Wainwright[4], Scott Collis[1]
[1] Environmental Science Division, Argonne National Laboratory
[2] Energy Geosciences, Lawrence Berkeley National Laboratory
[3] Computational Earth Sciences, Oak Ridge National Laboratory
[4] Massachusetts Institute of Technology

Prepared for
Biological and Environmental Research, Advanced Scientific Computing Research
Office of Science, U.S. Department of Energy

September 2022

Acknowledgments

AI4ESP was an open collaborative effort and could not have been possible without the efforts of many from senior leadership to early career researchers and beyond. The core leadership team wishes to sincerely thank everyone who contributed materials, time and to any other aspect of the workshop. Specifically, we wish to acknowledge:

- Office of Science DOE BER and ASCR, particularly Gary Geernaert, Steven Lee, Justin Hnilo, Renu Joseph, Jeffrey Stehr, Jennifer Arrigo, Xujing Davis, Bob Vallario, Randall Laviolette, and Barbara Helland for guidance, funding, and leadership. We would also like to thank Harriet Kung, Deputy Director for Science Programs for motivating the AI4ESP workshop in her opening remarks.
- Olga Tweedy from the DOE Office of Science for continued management of the AI4ESP activity and ensuring that this report was released in a timely fashion.
- The AI4ESP laboratory leadership group Kerstin Kleese van Dam (BNL), Jim Ang (PNNL), Timo Bremer (LLNL), William Collins (LBNL), Ian Foster (ANL), Aric Hagberg (LANL), Andy Salinger (SNL) and David Womble (ORNL) for ensuring that AI4ESP was targeted and impactful in its creation and execution.
- The production, editorial and website team including Emily Zvolanek (ANL) for tireless reference support in AI4ESP report writing, Pam Richmond (ANL) for her leadership in website and visual design, Diana Swantek (LBNL), Nathan Armistead (ORNL), and Lee Turman (ANL) for graphics design, and to Andrea Manning (ANL) for an immense effort in formatting and editing the report.
- The dedicated efforts of the session chairs and co-leads, who worked to organize the 17 sessions around Earth science and computational topics, synthesize discussions, and write the report chapters.
- All plenary and session speakers who helped set the stage for the workshop discussions.
- White paper authors who contributed to so much of the original ideas behind the workshop and report.
- Workshop participants, who enthusiastically engaged in the session discussions and provided meaningful input on the path forward.

Our goal was to provide a forum and avenue for your great ideas to be heard by other scientists, agency managers and administration, and U.S. government decision makers. Your selfless participation and open communication of ideas invigorated hope in the scientific process and community to drive innovation and transformation for humankind. We feel confident to express, on behalf of all authors of this report, that we hope to have done justice to the information provided and discussed in the workshop and beyond. Thank you again for the hard work and sacrifices you made to participate.

This page intentionally left blank.

Table of Contents

Artificial Intelligence for Earth System Predictability: 2021 Workshop Report

Executive Summary

In October 2021, the U.S. Department of Energy (DOE) welcomed participants to the Artificial Intelligence for Earth System Predictability (AI4ESP) Workshop, hosted by the Office of Biological and Environmental Research (BER)—Advanced Scientific Computing Research (ASCR). The workshop is part of BER-ASCR's ambition to more radically and aggressively advance prediction capabilities in the climate, Earth, and environmental sciences through the use of modern data analytics and artificial intelligence (AI). Advances in these capabilities are needed to improve predictions of climate change and extreme events that provide actionable information for planning and building resilience to their impacts.

A distinguishing aspect of this workshop was the framing around BER's "Model-Experimentation" (ModEx) integrative research process, which involves integrating observations, experiments, and measurements performed in the field or laboratory, with model research that simulates these same processes. This iterative approach enables models to generate hypotheses that inform field and laboratory efforts to collect data, which are subsequently used to parameterize, drive, and test model predictions. Hence, this workshop was unique in seeking an immense breadth of AI applications to enhance Earth system models, observations, and theory, as well as the computational infrastructure and transdisciplinary collaborations that enable their seamless integration.

The scientific challenges framing these disciplines have become increasingly complex and beyond the reach of traditional approaches. Hence, BER and ASCR encouraged workshop attendees to be bold with out-of-the box thinking, even considering a paradigm shift in the approach to scientific discovery and enhanced predictability. Sessions were organized around nine Earth system predictability science topics and eight cross-cutting artificial intelligence/machine learning (AI/ML) topics (see A Community-driven Workshop below). All sessions included in-depth discussions of the following: (1) the grand challenges that must be tackled; (2) state-of-the-science; (3) opportunities to advance science using radical approaches; (4) research priorities; and (5) 2-, 5-, and 10-year goals to frame the community's engagement. This comprehensive report summarizes the major outcomes of the workshop with an overarching goal to define priorities that can yield the most impactful science.

A Community-driven Workshop

A total of 17 topics were addressed at the 2021 AI4ESP workshop sessions. These topics emerged from 156 white papers from the community in response to BER-ASCR's call for thought leadership on developing AI methods and applications in BER research areas. The workshop sessions emphasized quantifying and improving Earth system predictability, particularly related to the integrative water cycle and associated water cycle extremes.

Earth System Predictability Sessions	**Cross-Cut Sessions**
• Atmospheric Modeling • Land Modeling • Hydrology • Watershed Science • Ecohydrology • Aerosols and Clouds • Coastal Dynamics, Oceans, and Ice • Climate Variability and Extremes • Human Systems and Dynamics	• Data Acquisition to Distribution • Neural Networks • Surrogate Models and Emulators • Knowledge-Informed Machine Learning • Knowledge Discovery and Statistical Learning • Explainable/Interpretable/Trustworthy AI • Hybrid Modeling • AI Architectures and Co-design

ES.1 AI-enabled Earth System Science: Pressing Need for Paradigm Change

AI technologies have expanded exponentially over the past decade and are well positioned to accelerate predictions of Earth system processes. The current paradigm governing scientific discovery is that process understanding derived from measurements underpins our ability to create models that make predictions and long-term projections. However, the progress in physical understanding often does not transfer to actionable predictions due to the complexities associated with chaotic interacting components of Earth systems. Challenges include the wide range of relevant scales (microbe-to-global spatial scales, minutes-to-centuries temporal scales), the nonlinear interactions of multi-scale processes and human impacts, and the outsized impacts of extreme events on the environment.

Scientists and stakeholders are increasingly demanding predictions that have finer resolution, larger spatial domains, greater accuracy, and longer time horizons. Some examples include an urgent need for more accurate prediction of extreme events and their impacts, enhanced understanding of processes and strategies to make natural and human systems more resilient to climate change, and more complete characterization of uncertainty and biases in models and data to constrain scientific findings. While the increase in supercomputing capacity has allowed finer scales to be explicitly resolved in process-based models, achieving high-resolution simulations across large spatial domains is still beyond current capabilities for many models, especially when large ensemble runs and new process representations are needed to improve accuracies.
AI technologies—neural networks, classical machine learning models, optimized data acquisition and assimilation, computer vision, and unsupervised and reinforcement learning techniques—are tools that must be harnessed to accelerate progress in Earth system observations and models. The challenges to increase resolution and process fidelity and reduce the uncertainty of Earth system predictions require novel approaches that assimilate AI into traditional experiments, models, and data acquisition methodologies. This hybridized approach, while attractive, leads to new challenges, such as developing massive multi-scale datasets that require new AI techniques to understand cause and effect; overcoming huge computational costs where intelligent sensitivity analysis and uncertainty quantification must make progress; and requiring

new parameterizations of important nonlinear scales that bridge the natural and human components in complex systems such as those found in urban regions.

The AI4ESP vision aims to dramatically accelerate Earth system models, observation, and theory by taking advantage of rapid progress in machine learning methodologies in conjunction with advances in big data infrastructure, analytics, optimized hardware architectures, networking and edge computing technologies, and other computational tools that were unavailable even a decade ago (Figure ES-1). This vision requires co-design and co-investment in observational capabilities and platforms, models and software infrastructure, and computational hardware to develop AI approaches specifically aimed at the climate, Earth, and environmental system sciences. Notably, this goes beyond AI technologies that to date have benefitted from massive investment in the private sector as they are optimized for commercial applications. Although there are opportunities to apply commercial AI tools for Earth system predictions, new scientific AI methodologies that incorporate process understanding and respect physical laws are required to make Earth system models interpretable, trustworthy, and robust.

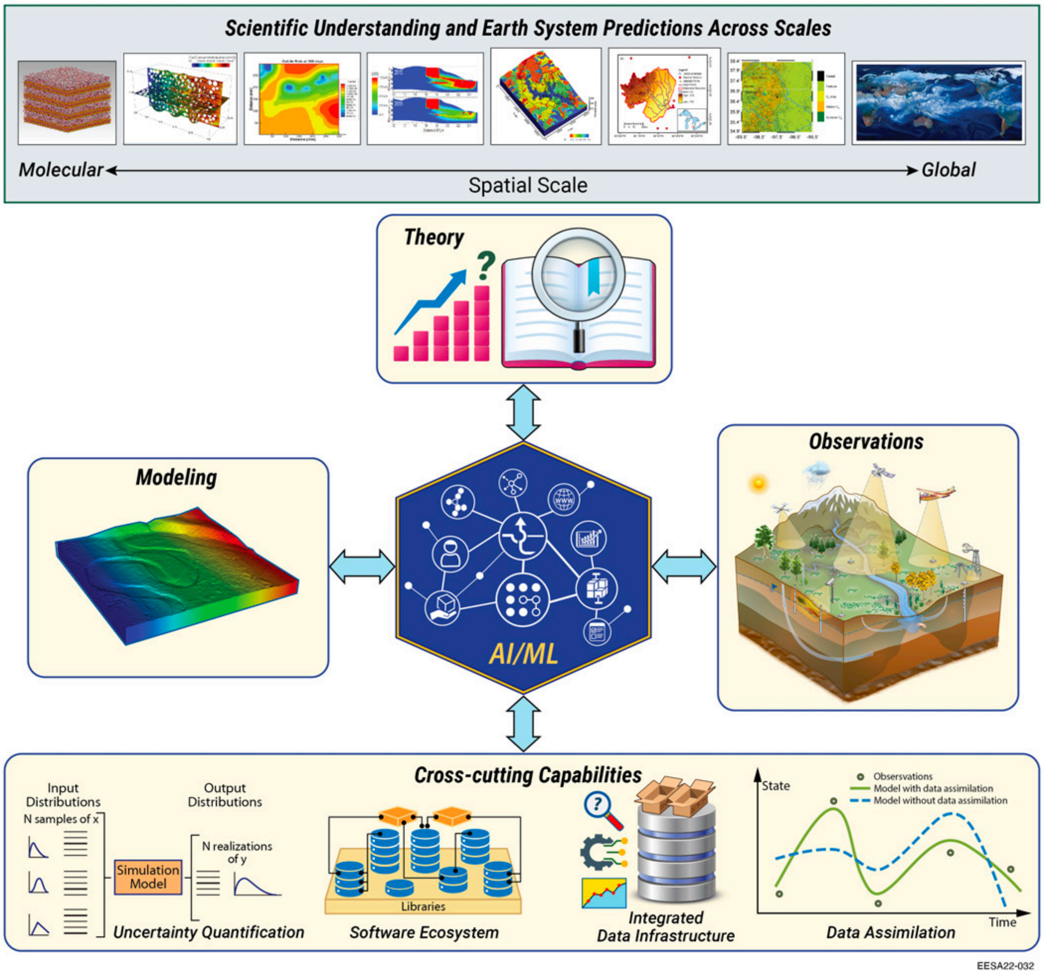


Figure ES-1. Conceptual diagram showing how Artificial Intelligence and Machine Learning (AI/ML) can enhance and be informed by the three pillars of Earth sciences—observations, modeling, and theory—as well as cross-cutting computational capabilities. The integrated outcomes from modeling, experimentation, and theoretical knowledge generation will enhance the scientific understanding and predictions of Earth systems processes across multiple spatial scales (Source: Lawrence Berkeley National Laboratory).

ES.2 Background of AI4ESP

AI4ESP began as a multi-laboratory collaboration within the DOE, which brought the national laboratories together to understand the key challenges and opportunities in AI/ML methods. The goal is to radically improve predictive capabilities by determining the most impactful AI/ML applications that span the continuum from observations to multi-scale modeling and analysis.

Planning for AI4ESP was initiated by scientists from three DOE national labs with leadership-class computing facilities—Argonne National Laboratory, Lawrence Berkeley National Laboratory, and Oak Ridge National Laboratory. However, the planning team rapidly expanded to engage scientists from five additional national laboratories: Brookhaven National Laboratory, Los Alamos National Laboratory, Lawrence Livermore National Laboratory, Pacific Northwest National Laboratory, and Sandia National Laboratories.

These scientists recognized the importance of broad participation and reached out to the Earth sciences, computing, and AI communities throughout the private and public international research enterprise through a white paper call. This outreach resulted in an encouraging response of 156 white papers from 640 unique authors across 112 unique institutions. Based on this response, AI4ESP leadership brought together a diverse group of approximately 100 people to design a workshop that would offer an open, collaborative environment to listen to and share ideas for understanding the opportunities for a paradigm shift in incorporating AI as a component of Earth system models and observations. This effort resulted in a virtual workshop, spanning five non-consecutive weeks in late October through early December of 2021 that engaged participants with diverse expertise across Earth and environmental sciences, computational sciences, and engineering (see Workshop Participation box).

Workshop Participation

The 156 white papers were from 640 unique authors across 112 unique institutions.
AI4ESP leadership brought together 100 researchers to design the virtual workshop, which:

- Spanned five non-consecutive weeks from October–December 2021.
- Engaged more than 740 participants from 178 institutions, of which 83% were domestic participants and 17% were international.

ES.3 Workshop Outcomes: Three Priorities

The AI4ESP workshop resulted in multiple overarching themes with recurring and common challenges, needs, and opportunities across many sessions. From these, three major categories of priorities emerged: (1) *Earth science*, (2) *computational science and methodology*, and (3) *programmatic and cultural changes to achieve a multidisciplinary and unified framework*. In particular, the workshop emphasized the need to incorporate AI into models, analytics, and data generation as a means to accelerate advancement, create new scientific opportunities, and revolutionize new approaches to predictive capabilities and capacity. The 2-, 5-, and 10-year priorities (Figure ES-2) identified across each of the workshop sessions could provide the basis for a roadmap to achieving the AI4ESP vision (Figure ES-3).

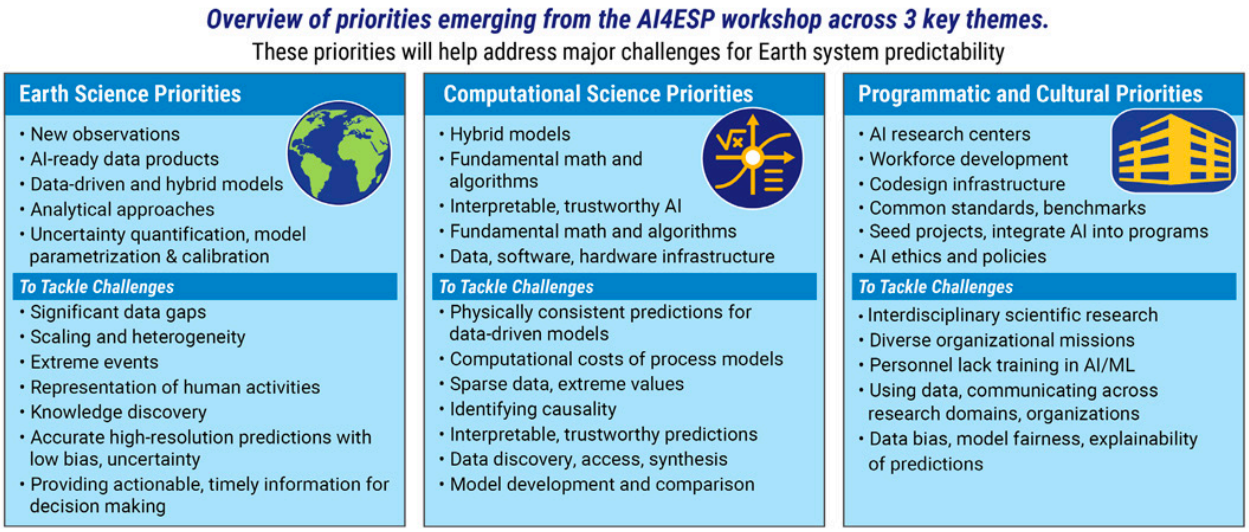


Figure ES-2. The AI4ESP workshop participants identified grand challenges and 2-, 5-, and 10-year research priorities to advance the use of AI/ML in Earth systems science. Several common themes emerged across the 17 sessions that fell broadly under the categories of Earth sciences, computational sciences, and programmatic and cultural changes to achieve a multidisciplinary framework (Source: Lawrence Berkeley National Laboratory).

ES.3.1 Earth Science Priorities

Some common challenges emerged from the various workshop discussions that involved Earth science topics. In general, the discussions all emphasized the central challenge of making dramatically improved predictions and observations across a wide range of spatial and temporal scales, and doing so with sufficient resolution and accuracy to accomplish this goal, which include (1) capturing heterogeneity in the relevant variables and processes, (2) overcoming the difficulty associated with observing and predicting extreme events, (3) managing and analyzing the immense volumes of data across a variety of ecosystems, and (4) launching a major effort to identify robust, interdisciplinary scientific approaches that integrate human activities.

The Earth system priorities focus on opportunities where AI can help address these challenges and reduce uncertainties. These included priorities for:

- Advancing scientific understanding and knowledge discovery
- Developing approaches for obtaining new measurements at desired scales and resolutions
- Prioritizing the collection, synthesis, and curation of data most valuable for advancing AI in different Earth science domains
- Creating AI-ready datasets, such as standardized benchmarks, and quality-checked and gap-filled data for model training, verification, and validation
- Incorporating AI/ML into Earth system models (e.g., surrogate models, emulators, and hybrid ML-/process-based models) to help address challenges related to scaling and process heterogeneity for accurate, high-resolution predictions with reduced bias and quantified uncertainties

- Improving predictive capabilities of extreme events and ecosystem disturbances including compounding events (e.g., coincident heat waves and droughts) and cascading impacts, given sparse data and lack of prior event analogs
- Improving representation of human-driven processes and interactions in models
- Quantifying the impact of all sources of error and uncertainties (e.g., arising from unknown model parameters, noisy data, missing processes, discretization errors in the solution of model equations, and approximations in reduced-order models or ML models)
- Developing a systematic framework, metrics, and workflows for model training, calibration and optimization, verification and validation, and intercomparison
- Employing AI/ML to provide the scientific foundations and identifying critical pieces of information to support decision-making at various scales
- Improving representation of human-driven processes and interactions in models
- Quantifying the impact of all sources of error and uncertainties (e.g., arising from unknown model parameters, noisy data, missing processes, discretization errors in the solution of model equations, and approximations in reduced-order models or ML models)
- Developing a systematic framework, metrics, and workflows for model training, calibration and optimization, verification and validation, and intercomparison
- Employing AI/ML to provide the scientific foundations and identifying critical pieces of information to support decision-making at various scales

ES.3.2 Computational Priorities

Common computational challenges that emerged across the sessions include the development of: (1) large, curated datasets for model training; (2) new mathematical approaches tailored for sparse data and extreme events; (3) novel approaches that address interpretability and potential physical inconsistencies of traditional ML model outcomes, driving the need for hybrid models; (4) innovative and consistent approaches to represent model uncertainties and trustworthiness; (5) software infrastructure for supporting hybrid model components across major Earth and environmental system science codes; and (6) efficient and interoperable frameworks and architectures that provide access to data and model resources across organizations.

The computational priorities identified future developments and advancements in AI and ML techniques, algorithms, mathematical frameworks, data management, tools and libraries, and hardware architectures, including for:

- Developing portable and efficient software infrastructure for systematically combining traditional process parameterizations with data-driven models for hybrid modeling and data assimilation, supporting plug-and-play parameterization swapping and online training
- Advancing fundamental math and algorithms for working with complex systems, sparse data, long system memory, and extreme values

- Developing methods to extract causal relationships and mechanisms and to offer robust interpretability and explainability of the ML model outcomes toward application-specific, explainable, and interpretable AI
- Developing AI-guided data acquisition frameworks that leverage adaptive observational capabilities such as edge computing and autonomous instrumentation, and that inform optimal data collection strategies
- Co-designing computational and storage infrastructure for automated ML model selection, design and training, integration of process and ML models, model intercomparison, and data assimilation
- Developing AI-assisted data discovery and synthesis, scientific data management archives, and tools that provide efficient access to and use of data across organizations and computational resources

ES.3.3 A Unified Framework to Incorporate Multidisciplinary Priorities and Cultural Change

Numerous programmatic and cultural needs were identified across all sessions, which include the need for (1) bridging multi-domain and multi-mission demands across different science and government communities; (2) having a trained workforce capable of interdisciplinary research across Earth and computational sciences; and (3) coordinating data generation, standards, synthesis, and model development efforts across different research groups. Development of a supporting framework to bridge these community-wide barriers would allow current and future activities to efficiently collaborate and accelerate development of AI research and technologies. In particular, workshop participants clearly signaled the need to create a radically different approach for future AI-enabled Earth system modeling and observational efforts that will enable and foster collaborations across disciplines and institutions. Notably, AI4ESP's focus on BER's ModEx approach, namely, using AI to enhance models, observations, and theory (Figure ES-1), makes the priorities identified by the AI4ESP community unique to the DOE scientific mission.

Workshop participants clearly signaled the need to create a radically different approach for future AI-enabled Earth system modeling and observational efforts that will enable and foster collaborations across disciplines and institutions.

Achieving the AI4ESP vision will require an unprecedented level of coordination across scientific disciplines and public, private, government, and scientific communities. Priorities identified to address these barriers include:

- Creating AI research centers tasked to coordinate and collaborate to more rapidly advance priorities across the various Earth science topics, where the centers would provide the supporting data and computational infrastructure, mathematical capabilities, and cross-disciplinary expertise to support community ambitions

- Co-designing frameworks or platforms to enable communities with different missions to efficiently share applicable results, techniques, data, and codes to decrease unnecessary duplication of effort and accelerate the application of AI
- Determining cross-disciplinary data-sharing standards, and creating shareable benchmark and training datasets that bridge organizations
- Supporting working group activities to investigate major and timely transdisciplinary research questions and quickly enhance or test developments such as workshops, challenges, and hackathons through a center or facility that is staffed to support commonly used data, models, and workflows
- Developing standards for trustworthy AI, including addressing data biases, ensuring fairness in models, and fostering the ethical and responsible development and use of AI
- Building public-private partnerships that enable use of commercial tools for research purposes and vice versa
- Focusing on new efforts to inspire and motivate the next-generation workforce, including training of multidisciplinary scientists, as well as outreach to a broader and more diverse set of academic and laboratory institutions
- Supporting early success stories to support training, inspiration, and strategic program design, such as through demonstration projects, infusion of AI into existing funded programs, and follow-up “implementation workshops” on key topics to chart roadmaps

ES.4 Beyond the AI4ESP Workshop

The success stories that are highlighted in this report and outcomes from the workshop deliberations clearly point to the potential for AI/ML to accelerate integrated, next-generation observations and models that incorporate complex natural and human processes at sufficient resolutions to support emerging science challenges, as well as to improve decision-making. There was broad consensus that AI can be transformational and help address long-standing grand challenges in Earth and environmental sciences, but that significant research and development in both AI and domain sciences are needed for this to happen.

Since completion of the workshop, participants have continued both high-level and specific topical discussions at conferences, such as the American Geophysical Union (AGU) 2021 Fall Meeting and the American Meteorological Society (AMS) 2022 Annual Meeting. Participants have also carried forward the information from the AI4ESP workshop to other community activities, including the SIAM AI4ESP workshop summary, National Academies Machine Learning and Artificial Intelligence to Advance Earth System Science workshop, and an upcoming special collection of the American Meteorological Society AI for Earth Systems (AIES) journal to promote information distribution from the workshop. Related workshops and meetings are expected in the future to capitalize on new collaborations and develop the underlying building blocks needed to develop a community-wide framework.

ES.5 Report Organization

The full report is designed to provide additional levels of detail in the following sections. The Workshop Summary provides references to past and ongoing activities, examples of AI applications involving Earth science, and both opportunities and research priorities identified across the 17 sessions. Individual chapter reports follow that dive into each of the Earth science domains and the AI/ML session topics. Finally, the appendices contain acronyms (Appendix A), the workshop agenda (Appendix B), call for white papers (Appendix C), lists of participants (Appendix D), and list of white papers (Appendix E).

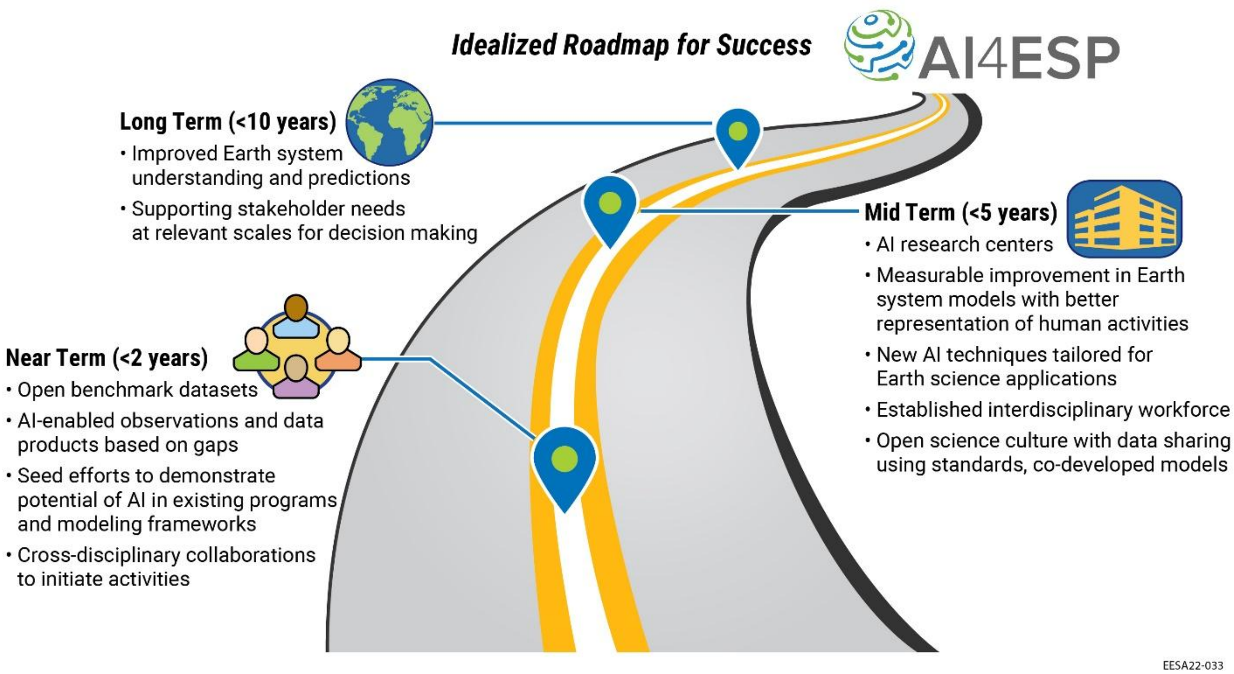


Figure ES-3. Roadmap to the execution of AI4ESP encompassing near- (2-year), mid- (5-year), and long-term (10-year) activities (Source: Lawrence Berkeley National Laboratory).

Workshop Summary

WS.1 Introduction

The Earth's natural resources are stressed due to climate and land use change, extreme events, and increased demand from growing populations. Advanced predictive capabilities for Earth systems are needed to provide actionable information to decision-makers for sustainably managing ecosystems and building resilience to disturbances. Recognizing the urgent need to improve the scientific understanding and state-of-the art predictions of the Earth system, the U.S. Department of Energy (DOE) prioritized a strategy to explore novel approaches that exploit developments in artificial intelligence (AI), edge computing, advanced wireless communications (5G and beyond), exascale-class computational architectures, and other emerging technologies. As part of this strategy, DOE launched the Artificial Intelligence for Earth System Predictability (AI4ESP, https://ai4esp.org/) effort in the fall of 2020, bringing together national laboratories, academia, and private sector participants. The Earth and Environmental Systems Science Division (EESSD) and Advanced Scientific Computing Research (ASCR) within DOE's Office of Science are jointly committed to exploring the paradigm shift necessary to bring together the expertise and capabilities for advancing the use of AI/machine learning (ML) for improving Earth system predictability.

Earth system prediction crosses an extremely heterogeneous set of scientific domains and spatiotemporal scales. This presents challenges to developing approaches for extracting scientific knowledge from data and making predictions that are relevant for a variety of stakeholders. Recent computational advancements in AI algorithms, big data analytics, and hardware present opportunities to dramatically improve Earth system understanding and prediction. However, realizing this potential requires co-design of computational algorithms and tools motivated by scientifically driven use cases and knowledge gaps. Hence, the AI4ESP workshop was themed around the DOE Biological and Environmental Research (BER) concept of "Model-Experimentation" ModEx (https://ess.science.energy.gov/modex/) that focuses on coupled development of models and observational capabilities in an iterative manner to test scientific hypotheses, generate knowledge, and improve predictions (Figure WS-1). Thus, the broad scope and scale of this workshop was unprecedented in comparison to prior efforts to identify AI opportunities in Earth sciences. Topics discussed spanned the use of AI for measurements, data generation, and modeling as well as the computational infrastructure and programmatic support needed to enable the adoption of AI across a wide variety of Earth science domains.

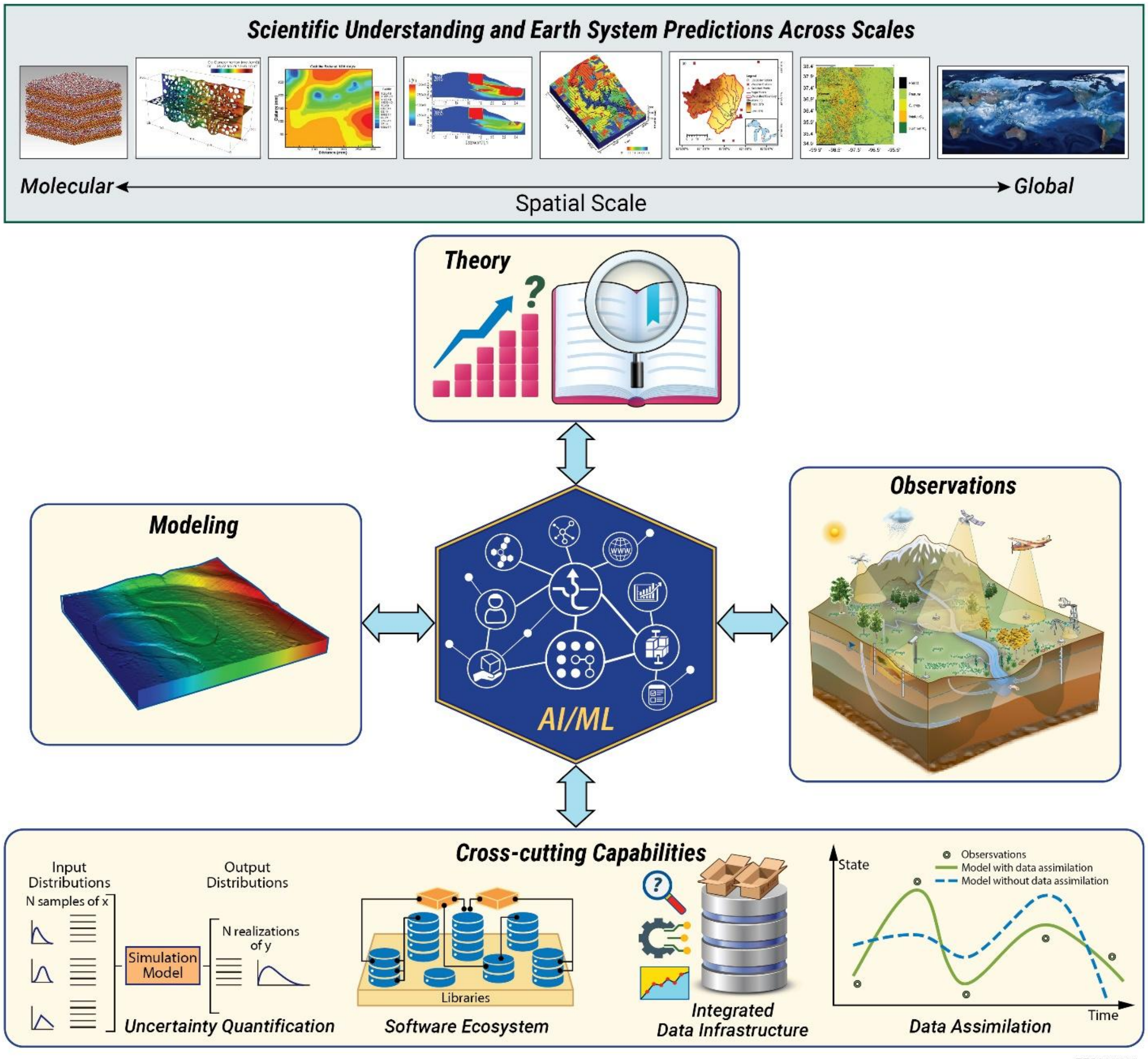


Figure WS-1. Conceptual diagram showing how Artificial Intelligence and Machine Learning (AI/ML) can enhance and be informed by the three pillars of Earth sciences—observations, modeling, and theory—as well as cross-cutting computational capabilities. The integrated outcomes from modeling, experimentation, and theoretical knowledge generation will enhance the scientific understanding and predictions of Earth systems processes across multiple spatial scales (Source: Lawrence Berkeley National Laboratory).

A multi-laboratory AI4ESP core team began with an open call for information in December 2020 relating to the potential of AI/ML to advance Earth system predictability, recognizing the need to bridge agencies (e.g., DOE, National Aeronautics and Space Administration [NASA], U.S. Department of Commerce/National Oceanic and Atmospheric Administration [DOC/NOAA]), institutions (e.g., university, public-private company), and scientific domains. The community response from 640 unique authors of 156 white papers from 112 unique institutions validated the understanding that the scientific community desires to collaborate in defining the vision and progressing forward. The team and DOE Program Managers engaged many of the authors by using the 156 papers to identify emerging opportunities and common themes and to organize a workshop that extended over a five-week, non-consecutive period starting on October 25, 2021.The AI4ESP workshop had 17 workshop sessions on topics related to Earth system predictability and AI/ML advances, which were titled by their primary focus and organized by a total of 100 session leads. Ultimately, the workshop engaged participants with diverse expertise across Earth and environmental sciences, computational sciences, and engineering (see Workshop Participation).

Workshop Participation

The 156 white papers were from 640 unique authors across 112 unique institutions. AI4ESP leadership brought together 100 researchers to design the virtual workshop, which:

- Spanned five non-consecutive weeks from October–December 2021.
- Engaged more than 740 participants from 178 institutions, of which 83% were domestic participants and 17% were international.

The AI4ESP workshop discussions were related to how AI could improve predictability for the different topics and what is needed to support the use of AI for Earth system predictability. Each week focused on different topics of Earth system predictability research in addition to other cross-cutting themes, where many topics involved co-design with new and/or emerging AI research concepts (Appendix B). The workshop was held in a very open style with presentations provided through the website, and authors were encouraged to share information and continuously develop teams and collaborations throughout the growing community.

Multiple overarching themes emerged throughout the workshop, which were recurring and common challenges, needs, and opportunities across many sessions. This summary categorizes these into three classes: *(1) Earth science, (2) computational science and methodology, and (3) programmatic and cultural activities to achieve a multidisciplinary and unified framework* (Figure WS-2).

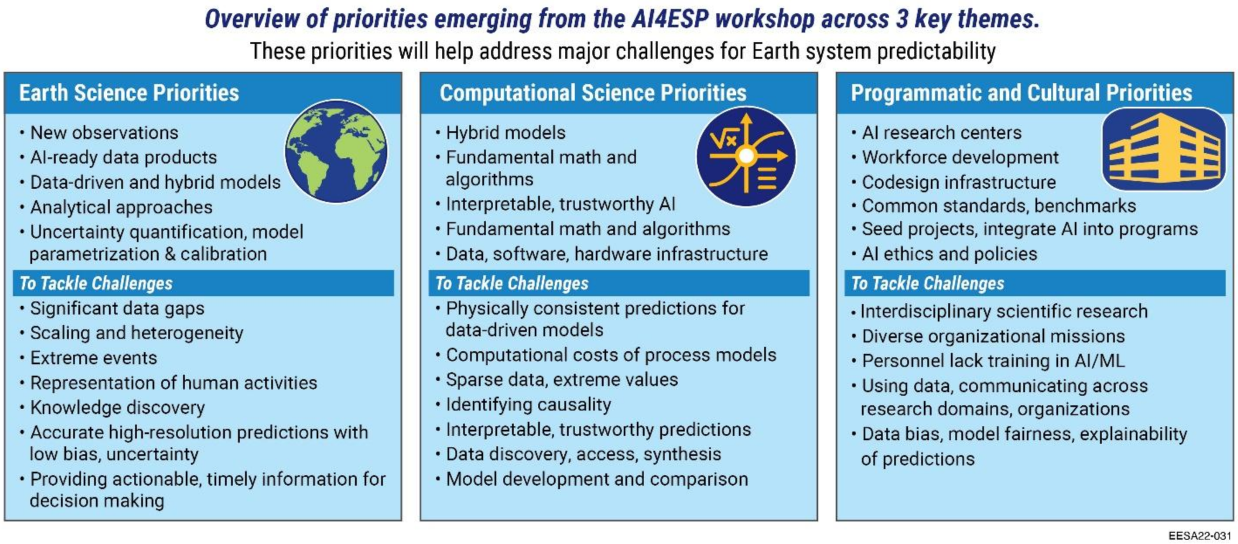


Figure WS-2. The AI4ESP workshop participants identified grand challenges and 2-, 5-, and 10-year research priorities to advance the use of AI/ML in Earth systems science. Several common themes emerged across the 17 sessions that fell broadly under the categories of Earth sciences, computational sciences, and programmatic and cultural changes to achieve a multidisciplinary framework (Source: Lawrence Berkeley National Laboratory).

WS.2 Earth Science Grand Challenges and Opportunities

WS.2.1Scaling and Heterogeneity

Scaling has appeared as a significant challenge and need in most of the Earth science sessions. Relevant scales range from microns (e.g., porous media, microbes, aerosols), to hundreds of meters (e.g., hillslope hydrobiogeochemistry, cloud physics), to hundreds of kilometers (e.g., ecosystems, regional climates and hydrology) to the global scale (e.g., global atmospheric and oceanic circulation). Many stakeholders need predictive capabilities at large scales; however, the presence of spatial and temporal heterogeneity in complex Earth system processes necessitates capturing fine-scale phenomena in models to attain accurate predictions. Ubiquitous challenges include the scales at which processes should be represented and aggregated in models, inconsistencies between model needs and the resolutions at which data are available, and model coupling to bridge scales and representation of subgrid processes. In many domains, models work well at small scales but often do not transfer easily to larger scales. Conducting high-resolution simulations at large scales while retaining fine-scale process representation with current models is computationally expensive if not impossible for many domains. Additionally, the approaches needed to obtain data at desired scales and resolutions must be identified and developed.

Incorporating ML into Earth system models (ESMs) can help address some of the challenges related to scaling and lead to development of regional-to-global scale models for predictions that work across a diversity of regions. This effort can be pursued in several ways. For example, statistical downscaling and upscaling have been an active research area in ML for many years in

Earth sciences. Deep learning algorithms are starting to enable hydrobiogeochemical predictions at continental scales (see Hydrology, chapter 3). Other strategies for scaling include replacing expensive process-based models or model components with faster-running ML-based surrogate models (see Ecohydrology, chapter 5) or using multifidelity approaches with high-fidelity simulations and fast-running, lower-complexity models. High-fidelity simulations could be performed for only selected regions of interest whereas ML could enable translation between high- and low-fidelity model outputs (see Watershed Science, chapter 4) or transfer of predictions to similar regions (see Hydrology, chapter 3). ML can also be used to develop efficient approaches for inverse modeling and uncertainty quantification (see Land Modeling, chapter 2). In addition, ML and deep learning methods for subgrid representation are emerging in several disciplines, particularly in atmospheric modeling and hydrology (see Atmospheric Modeling, chapter 1, and Hydrology, chapter 3). However, there are many gaps such that current subgrid parameterization does not sufficiently address uncertainty or does not efficiently incorporate multiscale or online data streams. From the ML sessions, there are many potential new methods suggested—such as hybrid physics/data-informed modeling, new classes of surrogate models, and two-way coupling among multiscale multiphysics models—which can be a paradigm shift to address scaling and subgrid representations in the Earth system models.

WS.2.2 Extremes, Disturbances, and Recovery

Weather and hydrological extremes such as hurricanes, floods, and droughts are projected to increase in a warming world and thus were a focus in many of the whitepapers solicited prior to the workshop. Fundamental and societally relevant questions related to extreme events include detection, attribution, and characterization; how their distributions are changing due to climate change; how they affect diverse natural and urban ecosystems or even trigger regime shifts that alter system behavior; how compound events interact to amplify consequences; and how we can identify adaptation measures that build system resilience. A primary challenge with modeling disturbances and their impacts is the lack of relevant data. Extremes, by definition, are rare events, and hence tend to have fewer observations. Moreover, the non-stationary conditions expected due to future extremes poses a substantial challenge to using data-driven approaches for modeling such disturbances without prior observations. Teasing apart the confounding effects of multiple co-occurring disturbances (e.g., due to heatwaves and drought) is also difficult with limited observations. Additionally, due to how they are trained, ML algorithms are biased toward predictions of mean values and need to be modified to explicitly simulate extremes and their impacts. Finally, there are no standardized definitions and data products available for most disturbances. For example, across the Earth system scales, there are many ways to define a drought or heatwave, often resulting in subjective choices for data selection and preprocessing, making it difficult to compare model results.

ML-based anomaly detection and feature identification methods have shown promise for identifying extreme events from multimodal data (e.g., time-series and remote sensing). Agnostic AI, unaware of manually applied labels of climate modes, along with relaxation of a priori criteria, could potentially lead to the discovery of new modes of variability, climate signals and precursors, and sources of predictability (see Climate Variability and Extremes, chapter 8). Approaches for modifying ML algorithms to better predict extremes and their impacts (e.g., modifying loss functions and using generative adversarial techniques) need to be developed. Predicting ecosystem disturbances and probable modes of recovery is critical to understanding vulnerability and resilience, but observational data are difficult to obtain and standardize (see Land Modeling, chapter 2, and Ecohydrology, chapter 5). AI-enabled technologies can be used to obtain automated observations of extreme events and post-disturbance system recovery (see Hydrology, chapter 3, and Data Acquisition to Distribution, chapter 10). ML can also be used to build standardized data products targeting extreme events, which are downscaled to appropriate resolutions and bias corrected (see section WS.2.6).

WS.2.3 Human Systems Integration

Human systems contribute to the uncertainty in climate change prediction and its impact on Earth systems (see Human Systems and Dynamics, chapter 9). In addition to the dedicated workshop session, human systems were of particular interest in the Land Modeling and Hydrology sessions (chapters 2 and 3, respectively), as terrestrial processes have been substantially altered by anthropogenic activities. Representation of human systems in Earth system models has been largely limited to the physical interface and is in its infancy in representing the science of decision-making—the action space. Workshop participants emphasized the importance of improving the spatiotemporal representation of human-induced processes, understanding multiscale interactions, developing generalizations of localized human systems, and identifying critical pieces of information to support decision-making at various scales (Figure WS-3). Current AI applications range from coupling integrated assessment models, like the Global Change Analysis Model (GCAM), with global Earth system models; integrating AI/ML with agent-based models; and developing emulators or less complex surrogate models for physical or human systems models to basic statistical and ML implementations (see Human Systems and Dynamics, chapter 9). One of the primary challenges with incorporating human systems into models is that relevant data are unavailable for various reasons, including the proprietary nature of these data, privacy concerns, and insufficient observations of human activities that affect the Earth system. Participants identified the need for new observations and data products relevant to human systems at fine scales.

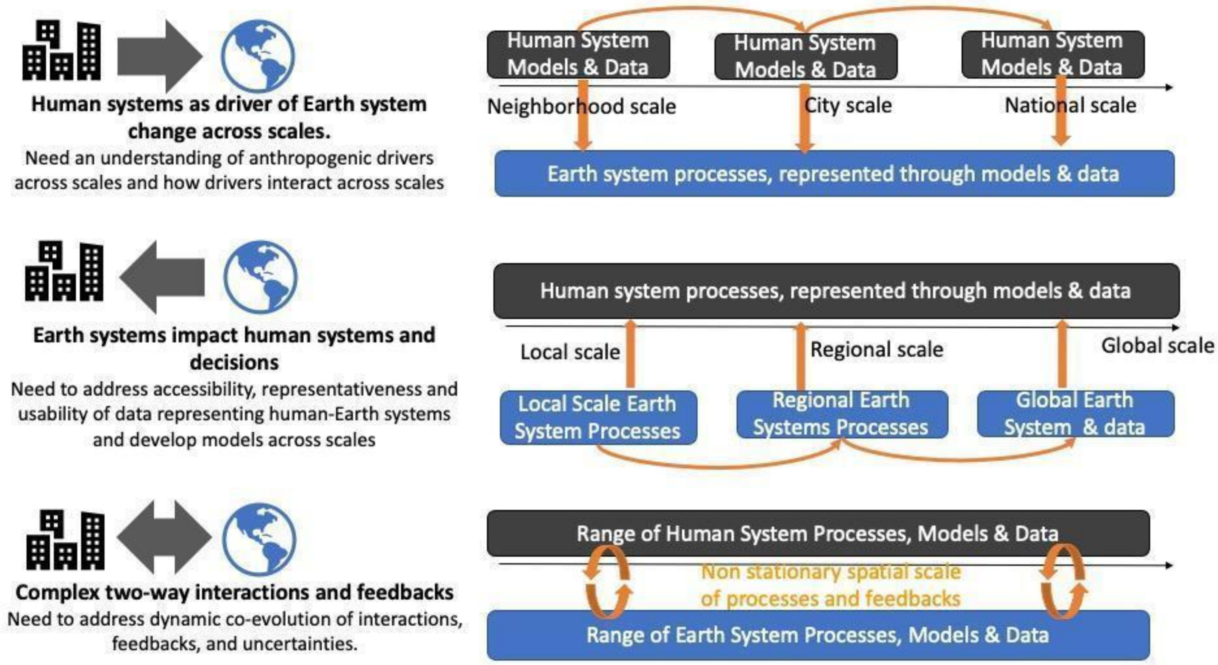


Figure WS-3. High-level synthesis of grand challenges in Earth-human system dynamics research (Sources: Pacific Northwest National Laboratory, Oak Ridge National Laboratory, Sandia National Laboratory; see Figure 9-2, chapter 9).

Despite the data limitations, using AI approaches is one of the key pathways to capturing anthropogenic activities in Earth system models because of the challenges in modeling human behavior in a purely mechanistic way. In this workshop, participants identified potential AI/ML targets such as developing data-driven models for representing human subsystems and using surrogate models to generate projections for different scenarios, reinforcement learning to drive decisions, and emulators to incorporate the outcomes of human actions into Earth system models. These would provide opportunities to advance and accelerate inclusion of complex human processes and dynamics in models; capture complex feedbacks among all components; and build ethically responsible decision-relevant models. The overarching goal is to use AI/ML technologies to develop fully coupled human-Earth system models and new platforms that integrate human systems science (see Human Systems and Dynamics, chapter 9).

WS.2.4 Model Improvements Combining Process Knowledge with ML

Hybrid AI/ML was another topic with multiple dedicated sessions (chapters 12, 13, 16) that also garnered significant discussion across Earth science sessions. Because of the availability of growing volumes of observational data and in situ data, the Earth system modeling community is increasingly adopting data-driven approaches for high resolution weather and climate simulations. While DOE supercomputers have traditionally been employed for process-based simulations across different science domains, the computational expense and challenges to mathematically representing key processes impose limits on model predictions and Earth system model projections. Combining ML-based process representations with traditional ordinary and partial differential equation representations offers new opportunities to improve model predictions, speed computational performance, and advance understanding of the complex Earth system.

Participants identified the need to develop a multifidelity framework of data-informed or custom-built model-informed surrogates using ML techniques such as Gaussian process regression, dynamic mode decomposition, random forest, and neural networks (NNs) as these techniques have shown promising results in addressing this problem. Emerging techniques, such as physics-informed NNs, have great potential to couple process-based simulation capabilities and state-of-the-art ML and improve our understanding of physical processes that are missing or poorly represented by current ESM parametrizations. In particular, two-way coupling (or online integration)—running physics simulations and AI/ML algorithms simultaneously and concurrently—is identified as a difficult but potential game changer. Furthermore, using explainable and interpretable NNs to replace parametrizations, could improve predictive understanding, accuracy, and uncertainty estimates of model outcomes. However, advances in hybrid modeling will remain challenging without making progress in domain-informed data processing methods and restructuring ESMs on modern architectures (e.g., most of the DOE codes are built on FORTRAN and thus not often supported by the modern AI/ML architecture).

WS.2.5 Uncertainty Quantification and Model Calibration

Uncertainty quantification (UQ) and parameter estimation (PE) have been active research areas for ML or general statistical methods for the last several decades. Although UQ and PE approaches have been established for relatively straightforward physical systems, there are still significant challenges to using them in complex multiphysics coupled systems. In many instances, computational resources are optimized for running one large simulation, or technique, rather than for determining UQ. At the same time, identifying a comprehensive UQ that propagates errors from heterogenous raw field data and experimental data to integrated data products and ultimately model simulations is quite challenging. Advancing the use of AI/ML in model calibration would also require addressing other challenges, including the lack of sufficient data and data quality, and multiple sources of uncertainty associated with data and model experimental setup (Atmospheric Modeling, chapter 1). Notably, UQ becomes particularly complex for hybrid models, which are sensitive to structural errors and parameter choices in process models, as well as affected by choices of input features, training data, and hyperparameters in ML models. Quantifying uncertainties caused by human activities and their impacts on the Earth system is an additional challenge.

To improve Earth system predictability, developing a systematic framework and workflow to advance UQ from observations to simulations is critical. During the workshop, various emerging ML methods were suggested to exploit modern computing including hybrid AI/ML approaches coupling physics-based simulations and data, as well as multifidelity approaches, which couple the results from model runs with different resolutions of complexity. For example, surrogate models (or emulators) were often raised as effective ways to speed up sensitivity analysis and allow for more efficient model parameter calibration. ML approaches were proposed for the auto-calibration of models, wherein multiple optimal parameter settings could be obtained instead of a single hand-tuned result that would enable parametric uncertainty to be included in model outcomes (Atmospheric Modeling, chapter 1). Such auto-calibration could leverage emerging AutoML frameworks that utilize advanced optimization algorithms for ML architecture design and hyperparameter choice. Other approaches include ML-based parametrization of ESMs, where dynamic parameters are learned from appropriate sets of statistical covariates, which have been used in hydrology and atmospheric modeling (Atmospheric Modeling, chapter 1; Land Modeling, chapter 2; Hydrology, chapter 3).

WS.2.6Knowledge Discovery and Hypothesis Generation

Several sessions identified opportunities for using data-driven analytical approaches and ML models to make new scientific discoveries from large datasets, as well as generate testable hypotheses. For example, AI/ML approaches can be used to identify patterns in data; classify landscapes or regimes with similar behavior; for detection and attribution of extreme events and compound disturbances; and discovery of unknown relationships or mechanisms hidden in big complex data. In addition, ML and hybrid models (including surrogates of process models) can be used to generate new hypotheses that can be tested with a combination of new observations and further model development in the classic ModEx approach.

To advance scientific understanding, it is not sufficient to improve prediction accuracy only but to determine "how" and "why" models make predictions. More approaches that investigate the underlying mechanics of how deep learning models make predictions are needed. Ultimately, the goal is to be able to extract the maximum amount of information from the data as possible. Toward this end, some of the more exciting approaches attempt to discover governing equations using data-driven methods; however, these require significant research to become applicable for predictions in complex systems (see Hydrology, chapter 3; Knowledge-Informed ML, chapter 13). Additionally, information theory and causal inference approaches that infer causality from data, in tandem with ML, can be used to determine system responses to different driving factors, extract causal relationships and mechanisms, and offer robust interpretability and explainability of ML model outcomes. It was noted by the workshop participants that applying large-scale causal inference within the Earth system communities remains in the development stage, and there is no single algorithm that detects complex causal inference robustly.

Other techniques such as natural language processing (NLP) and computer vision tools may help to extract valuable knowledge from published literature (Knowledge-Informed ML, chapter 13). It is difficult for researchers to master all available information across the Earth sciences and

AI/ML domains given the growth in volumes of publications and different ways of knowledge representations (e.g., text, tables, figures, datasets, codes). Meta-analysis and reviews across the literature in Earth sciences have proven to be useful in extracting generalizable insights across studies, and automation of such synthesis can not only accelerate scientific discovery but help generate hypotheses and identify future research directions.

Finally, several sessions pointed to the need for different types of integrated data products for model development, benchmarking, validation, and ultimately knowledge extraction (section WS.2.7).

WS.2.7Data: New Observations and Data Products

Although the interest in AI/ML has been spurred due to the rapid increase in Earth observations over the past two decades, the amount of data available is sparse and insufficient to capture the range, dimensions, complexity, and heterogeneity of several Earth system processes. The lack of data is identified as a significant challenge across many of the domains, and it affects both process-based and ML models. Observational data are required for model input, training, validation, benchmarking, parameter calibration, and process representation. Many data gaps were identified, both in terms of volumes of data collected, missing variables needed for accurate modeling, and regional coverage (see Human Systems and Dynamics, chapter 9; Hydrology, chapter 3; Aerosols and Clouds, chapter 6; and other chapters). For example, many currently available datasets are based on a single field campaign or a small set of process model simulations and do not scale beyond the measurement domain. Large-scale monitoring networks are often focused on a few variables of interest, and other co-located observations—needed to quantify processes and their interactions—are lacking. Sampling bias is a concern, and it results in underrepresented regions in current data products (e.g., mountainous areas are challenging to observe due to complex terrain). Data involving human systems and extreme events are extremely limited, and in some cases proprietary data may not be publicly available (e.g., see Human Systems and Dynamics, chapter 9; Climate Variability and Extremes, chapter 8). Sensor maintenance and data quality assurance and control (QA/QC) were identified as significant bottlenecks to deploying instrumentation at large scales. Another challenge is that data collected by a myriad of entities ranging from federal and state agencies, local governments, industrial partners, and academics are distributed across different databases in non-standard ways, making it difficult to discover, access, and integrate data.

Several sessions identified major data gaps and the need for more observations that have much greater spatiotemporal coverage and resolution than is possible with current infrastructure. A short-term need is for community efforts (e.g., workshops) to prioritize the collection, synthesis, and curation of data most valuable for advancing the use of AI in different Earth science domains. Some domains called out the need for data infrastructure to house stand-alone or federated datasets. There was universal consensus that new value-added data products were essential to make progress in each of the domains. These would include AI-ready datasets such as standardized benchmarks, quality-checked and gap-filled data for model training, products that synthesize and harmonize different types of observations to desired spatiotemporal resolutions, and ensembles of high-resolution simulation outputs that could be used to create hybrid models. Other needs include for data and metadata standards that enable interoperability between different systems and for tools that enable data synthesis and assimilation into models.

Advances in low-cost sensing, edge computing, 5G networks, and high-resolution remote sensing (e.g., through cubesats, drones) paired with ML-based classification and regression can be used to acquire and create new data products. ML-enabled technologies can expand our observational footprints; identify optimal sampling strategies; drive data collection when and where it is needed; quality-check data in near real time; and create imputed data products to make better use of coarse, sparse, and indirectly related information. For example, ML or hybrid model output could be used to determine optimum new sensor placement and drive autonomous "self-guiding" measurements. Natural language processing can enable advanced search and harmonization tools to create curated AI-ready datasets and to automate analysis through data mining.

WS.2.8 Applications for Informing Decision-Making

Enhancing model applications through AI/ML for decision-making was discussed in several of the domain science sessions. Example applications that were discussed included end-to-end tools that seamlessly integrate ML algorithms (e.g., neural networks) and process-based simulations to enable actionable predictive insights for stakeholders at relevant scales. AI/ML technologies are already changing the operational forecasting landscape and have the potential to improve state-of-the-art Earth system models to generate timely predictions at large spatial scales for use by practitioners and decision-makers in a manner that is currently not feasible.

Participants across sessions identified areas where improvements in AI/ML could assist with decision-making. These include but are not limited to: (1) identification, observation, and prediction of extreme events or other disturbances and their impacts to urban and natural systems; (2) interactions between human decision-making and changes in socio-ecological systems over time in response to environmental change and extreme events; (3) transferability of models to different local and regional settings to provide predictions at decision-relevant scales; (4) appropriate assessment of uncertainties in model predictions; and (5) the building of confidence and trust in AI/ML systems so that decision-makers, practitioners, and other end-users feel comfortable in employing these systems in practice. The importance of having trustworthy, explainable AI and understanding how ML and hybrid models make predictions was repeatedly emphasized across most of the sessions. Some noted the need to incorporate stakeholder engagement to determine different end-user needs into the development process for AI/ML systems. The Human Systems and Dynamics session (chapter 9) noted that

> [t]he communication of uncertainty extends from technical evaluations, accomplished in part by statistical and ML techniques, to dissemination of these analyses to policymakers and the interested public. Doing the latter in an iterative process might be needed to share questions, insight, data, and research outcomes, while following high ethical standards around data privacy.

WS.3 Computational Science and Methodological Challenges

In addition to applying existing state-of-the-art AI/ML techniques to science in the Earth system predictability space, AI4ESP challenges new research and developments into AI. This subsection summarizes the computational science and methodological challenges and opportunities facing AI4ESP.

WS.3.1Knowledge-informed ML and Hybrid Models

Participants in most sessions called for the development of new algorithms, tools, and mathematical frameworks that systematically combine traditional physics with data-driven models. These are needed for various reasons, including for generating physically consistent, interpretable predictions that are extensible under non-stationary conditions, enabling data-driven models to work with sparse datasets that are typical in Earth sciences, having data-driven representations of processes that are not adequately represented in current models (e.g., preferential flow, human activities), and improving the computational speed of process-based models. Furthermore, there is an opportunity to develop new methods for integration and performance analysis of ML models with complex domain constraints to avoid conservativeness or loss of expressivity of the ML model.

Hybrid models of various types were proposed during this workshop including embedding ML models in physics-models, using process model outputs to train and optimize ML algorithms (as emulators or surrogates), and fundamentally altering ML model architectures to incorporate physics.

Hybrid, ML, and reduced-order models have a strong potential to reduce the cost of running components of Earth system models (Surrogate Models and Emulators, chapter 12). Further advances in surrogates using ML techniques are needed to improve the predictive understanding, accuracy, and uncertainty estimates of model outcomes. To date, most surrogate models are surrogates of a single information source, usually an expensive computer model. If used for uncertainty quantification, they are designed to treat only a single source of uncertainty, usually model parametric uncertainty. The workshop participants envision future surrogates to be capable of making projections about the real Earth system by combining information from hierarchies of models and data across multiple modalities, as well as incorporating different sources of uncertainties.

In parallel with this development, it is critical that algorithms are developed that explicitly quantify the impact of all sources of error and uncertainties (e.g., arising from unknown model parameters, noisy data, missing physics, discretization errors in the solution of model equations, and approximations in reduced-order models or ML models). It has been acknowledged that hybrid models are often calibrated and tuned in a highly manual way, and there is no systematic approach for dealing with the out-of-distribution, tuning, and calibrating of the hybrid models to reduce waste of thousands of compute core-hours and the sheer time required for tuning the simulation (see Hybrid Modeling, chapter 16). There is a need for development of ML techniques that will allow uncertainty to propagate through the hybrid-models to better understand the final variability of the predictions and for methods that can inform about the sources of uncertainty in the model chain.

There is also a need to develop ML models that can quickly adapt to the change in the data. Developing models with data assimilation and continual learning capabilities will significantly improve rapid training without the situation of losing learning from the past and will enable fast adaptation to the new information from the newly collected data.

WS.3.2 Fundamental Math and Algorithmic Advances

Participants emphasized a growing need for additional advances in mathematics and computer science, such as for: (1) faster, scalable algorithms for mathematical optimization (including constrained optimization solvers); (2) improved numerical solvers for differential equations and numerical analysis (adaptivity; stability); (3) advanced model reduction for infinite dimensional systems; (4) improved stochastic gradient descent (SGD) and more research and development (R&D) involving scalable randomized algorithms; (5) the discovery of new physics/mathematics (e.g., governing equations) that govern the physical phenomenon, as well as (6) advances in software development, workflow design, and data management (see Knowledge-Informed Machine Learning, chapter 13).

One of the concerns from Earth system domains is that ML training is challenged by very small training set sizes (e.g., limited number of ensembles, short durations of simulation or observation) to provide many independent data points. Correspondingly, the AI4ESP community identified a need to advance ML approaches that work in the low data or data-sparse regime. In addition, the need for algorithmic advances were called out specifically for predictions of extreme events. Earth system models can be significantly improved by addressing data challenges related to non-Gaussian (i.e., multimodal or heavy-tailed) data distributions of extremes and lack of relevant observations (also discussed in section WS.2.6). Thus, future developments and advancements in ML techniques that do not rely on Gaussian assumptions have the potential to play an important role in improving our understanding and prediction ability of extremes. Furthermore, there is a need to improve existing and develop new ML algorithms that can inform data collection strategies (e.g., what additional measurements are needed, optimal sampling frequencies, and spatial distributions). For scalability, nonlinear dimension reduction methods are also needed (see Neural Networks, chapter 11). Preliminary work shows promise; however, advancements must be made to further develop these computational tools.

WS.3.3 Trustworthiness: Explainability, Interpretability, and Physics-based AI

Applied AI, which efforts in AI4ESP fall within, must be understandable, robust, and reproducible. For example, if a component of a microphysical scheme was replaced by an AI-driven surrogate, it would need to be explainable and improvable with new measurements. AI architectures, which have largely been developed primarily for classification and prediction on images and time series based on frequency components (convolutions), need to be more physically based (Figure WS-4). Current explainable AI techniques, like layer-wise relevance propagation (heatmaps, or attention maps), will have additional interpretability if the network layers have physical and not just spatial meaning. This would also mean that data-driven AI model development, the developing of surrogate models based purely on data observed from the past, would not only provide predictions but could drive knowledge discovery as scientists could probe why a certain prediction was made and the answer would not be just based on the propagation of a particular spatial structure through a long-short-term memory (LSTM) network. A need for new hybrid ML algorithms that will help to identify new and existing processes was also identified (see Hybrid Modeling, chapter 16).

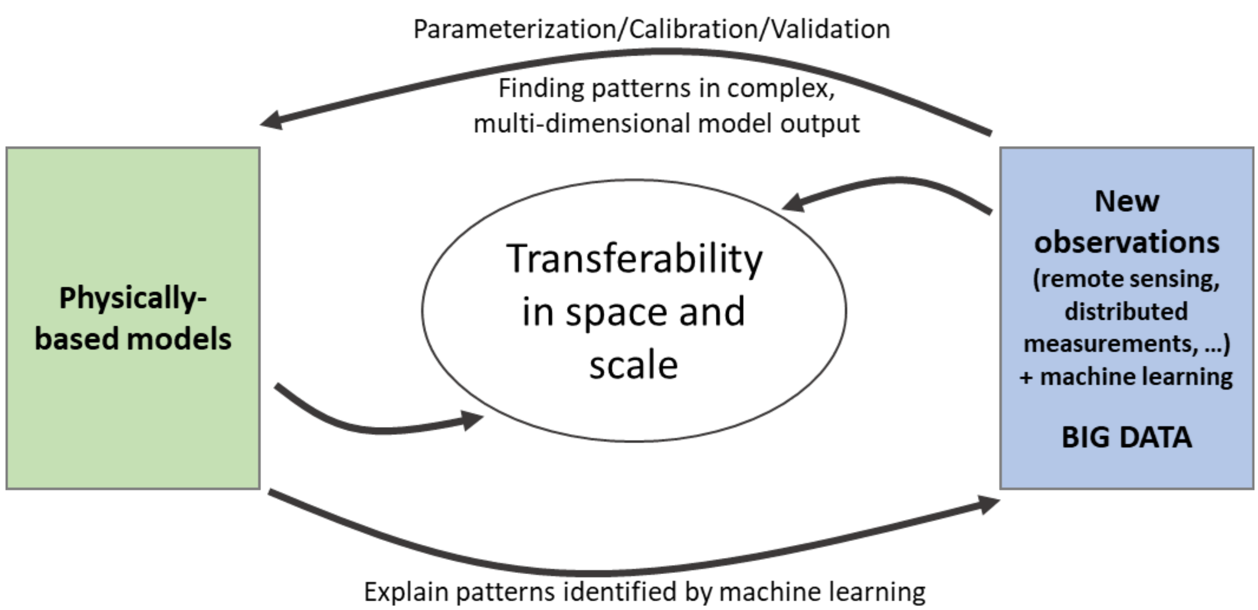


Figure WS-4. Combining physically based models with machine learning models enables identification of processes and patterns that can inform future model development and new observational campaigns. Such hybrid models provide transferability across space and time scales (Source: figure courtesy of Naomi Tague, UC Santa Barbara; see Figure 5-2, chapter 5).

The Earth system predictability (ESP) community often expresses concerns regarding the lack of robustness and reproducibility of AI/ML models, which can slow the progress of the field and significantly reduce researchers' efforts to accelerate new development and discovery. Currently, there are no community-wide guidelines, standard metrics, and computing infrastructure for ensuring and quantifying the robustness and reproducibility of AI/ML (see Explainable/Interpretable/Trustworthy AI, chapter 15). Establishing reproducibility standards for AI-based ESP research and developing community-wide guidelines and recommended practices will promote robust and reproducible AI/ML in Earth system predictability.

Numerous and sometimes conflicting definitions of explainable and interpretable AI make implementation of AI and ML methods even more difficult. Reaching a consensus on definitions of AI and ML is essential for making progress in this area, which involves formalizing AI explainable/interpretable methods in a rigorous framework that quantifies why a particular explanation is better than another.

WS.3.4 Data Acquisition, Edge Computing, and ModEx

As mentioned previously, Earth system observations are sparse and insufficient to capture the heterogeneity of relevant processes across scales. Our knowledge of Earth system processes often relies on instrumentation deployed or operated for monitoring or research purposes, sometimes in a sub-optimal manner due to logistical constraints, with insufficient data to guide sensor placement, and constrained by available funding and space or timing of events. Currently, the ModEx cycle takes so long to realize results between data collection, model refinement, identification of knowledge gaps, and redeployment that the state of the science has moved on during the process.

To accelerate advancement across ModEx, the workshop and this report identify the need for developing an AI-guided data acquisition framework, which will require exploration of optimal sampling strategies based on existing datasets and model projections; the leveraging of AI such as reinforcement learning to develop adaptive, agile data collection that can operate in autonomous modes; and development of emulators of complex processes to help define observational needs for ESMs (Figure WS-5). In the longer term, developing observing system simulation experiment (OSSE)-type strategies that bridge process models and AI approaches can help determine optimal sampling design for large, complex measurement networks and campaigns. Instruments and data collection methodologies need to be adapted for the science question being asked and the phenomena present. In the framework of AI, this means inference at the edge, across distributed sensor networks, and with linkages to training and modeling at high-performance computing (HPC with a continuum of computation (enabled by 5G and other technologies). This objective requires large scale heterogeneous cyberinfrastructure with an open community codes co-designed by Earth system scientists and computer science and AI researchers. Such cyberinfrastructure and specific efforts do not yet exist, and DOE offices (e.g., ASCR, BER) are well positioned to collaborate across agencies to bring it into existence.

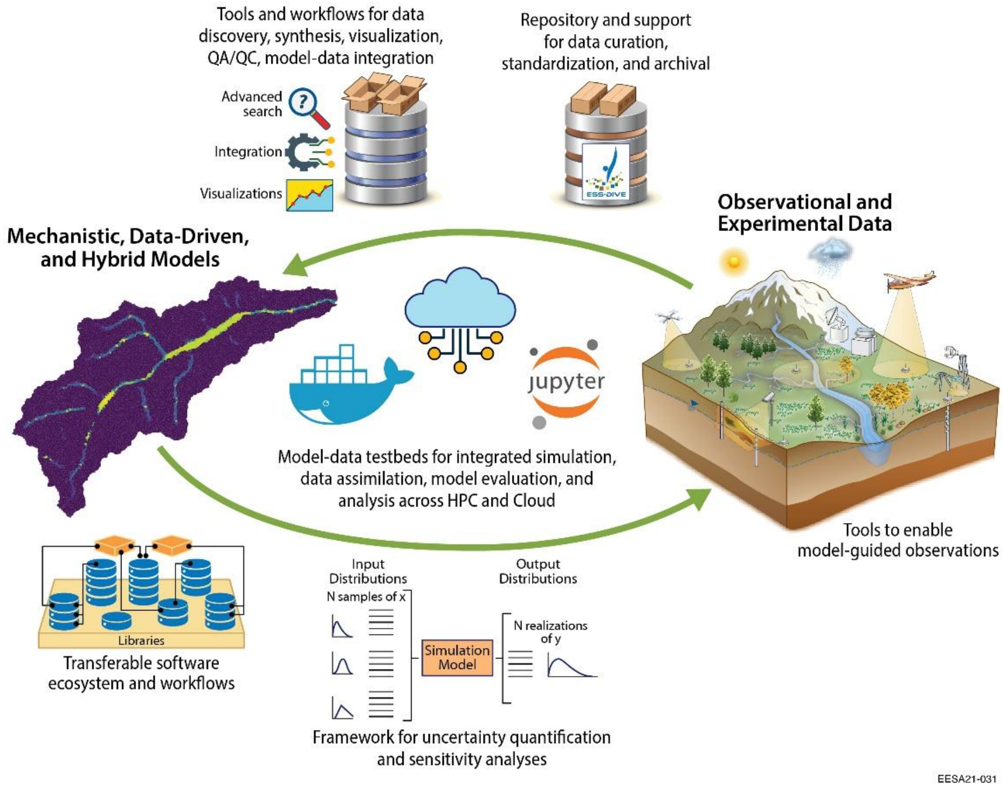


Figure WS-5. Data-model and model-data pipelines that include data management tools, software workflows to enable data discovery and integration, scaling and transformation of data for model needs, and testbeds to enable co-located big datasets that can be fed into models on HPC facilities. Examples of technologies that can be used in modeling testbeds include Jupyter notebooks (http://jupyter.org), Docker containers (http://docker.com), and tools that enable seamless execution of ML/hybrid models on HPC and cloud computing centers (Source: Lawrence Berkeley National Laboratory; see Hydrology, chapter 3).

WS.3.5 AI Architecture, Infrastructure, and Co-design

Co-design and implementation of an integrated data and computational infrastructure are essential for supporting AI/ML in Earth system science, which involves leveraging existing data centers, computational centers, and software infrastructure. Large-scale computing systems, such as DOE's Leadership Computing Facility (LCF) systems for high-performance computing and AI, can provide a foundation for advanced system concepts that range from centralized, large-scale modeling and training to edge computing inferencing and federated learning (see AI Architectures and Co-design, chapter 17). Capitalizing on current and future capabilities, including exascale and quantum computing, will require a significant investment both in foundational technology systems and co-design programs. This investment would enable Earth system scientists, mathematicians, AI/ML experts, computer scientists, and hardware engineers to collaborate on creating a radically different approach to future Earth system modeling efforts and AI-enabled integration with measurement and observation data. Among identified AI4ESP opportunities are development of (1) a new generation of proxy applications and benchmarks for both modeling and observation capabilities; (2) AI-assisted, physics-based models; and (3) AI/ML for adaptive data assimilation, observations, and edge sensor control. Furthermore, there is a need to co-design a collection of AI-ready datasets and improve the modularity of models, as well as for a framework for which to accelerate the progression of these. Long-term priorities of AI-at-scale opportunities for the envisioned concepts (i.e., HPC+Cloud+Edge) must be supported with near-term priorities for secure storage, authentication, provenance, and development of federated learning that integrates streaming analytics and AI at edge sensors with HPC modeling (see AI Architectures and Co-Design, chapter 17).

The research community currently has access to HPC capabilities at large computing centers, like DOE's Argonne Leadership Computing Facility (ALCF), National Energy Research Scientific Computing Center (NERSC) at Lawrence Berkeley National Laboratory, and Oak Ridge Leadership Computing Facility (OLCF). The next generation of computing architectures is likely to be custom-designed to better support deep learning capabilities that will be required to propel ever-larger ML models. In addition, the community has access to large collections of Earth system and environmental data at data centers like DOE's Atmospheric Radiation Measurement (ARM) Data Center (ADC), Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE), Earth System Grid Federation (ESGF), NASA's Distributed Active Archive Centers (DAACs), and many others. These data centers operate as stand-alone resources and usually require users to download data to their own computational resources or move data to the computational resources provided. This process of downloading, pre-processing, and integrating the data and then performing simulations and analysis is tedious and often unnecessary given recent technological developments. When developing and deploying AI/ML methods, researchers will experience the increased difficulty of this workflow because high-speed access to vastly larger data collections will be required for training ML models, potentially incorporating such training as part of the simulation itself. Thus, data-to-ML pipelines (model-observations integration) were suggested as a critical need. A domain-specific example of such a pipeline exists in the DOE's KBase project, which links to well-curated datasets and provides an application programming interface (API) to extract data within ML codes.

Advancements called for by the AI4ESP community require integrated computing capabilities (designed for both traditional and AI/ML workloads) and data infrastructure that eliminates, or greatly lessens the burden of, the challenges of finding, acquiring, downloading, and integrating data. Benchmark AI/ML data—including datasets created from model-data integrated, synthesized, and combined data products—should be created on an on-demand basis, based on scientific needs, and accessible from any large computing environment, no matter where those data reside or are archived. This standard could be accomplished through APIs and data transport services, like Globus (https://globus.org/), that hide the details of data movement and exploit high-bandwidth networks to deliver data as needed for simulation and analysis, combined with data synthesis and integration tools that are easy for scientists to use. A model-data integration center that could provide such integrated storage and computing resources for the growing Earth system science community would help accelerate the science identified in this report. This capability could be stand-alone or integrated within a broader AI research center (section WS.4); provide data hosting services, compute-near-the-data infrastructure, and "AI/ML as a service" capabilities; and sponsor training activities and multidisciplinary working groups focused on new or advanced research topics. Such a center could lower the barrier-to-entry for scientists to use advanced ML capabilities while enabling research with tools not otherwise easily accessible or usable.

WS.4 Programmatic Structure: Culture Change

The AI4ESP workshop was a demonstration of the scientific community reaching across agency, institution, public-private, and mission-specific boundaries to express the need for large-scale change. The specifics of how diversity, equality, and inclusion should improve are not the focus of this report; however, the community supports and is engaged in the advancement of these important factors across the scope of AI4ESP. The needed advancements described in this report require progressive steps to remove barriers to necessary collaborations. These advancements are beyond "grassroots effort" capabilities and require infrastructure investment and leadership direction.

Modernizing scientific culture efforts, typically isolated and stovepiped due to funding sources and heavy workload fatigue, are necessary to uniting the Earth system community to efficiently advance research and provide meaningfully improved results for stakeholders, including to:

- Support the existing workforce in staying current and leveraging hardware, software, AI, and domain science advancements
- Support the future workforce in learning and entering collaborative environments at the cutting edge
- Support diverse missions within a framework of accelerating testing and advancement for the community-based co-design, standardization, and open architectures

The need for a paradigm change is repeated from other perspectives, such as the National Academies ML and AI to Advance Earth System Science workshop report (National Academies of Sciences 2022) and is required to lead in Earth system predictability. Co-design, effective collaboration, infrastructure investment, and workforce development are all necessary to achieve this change.

WS.4.1 Support of Existing Workforce

The separate advancements in AI, hardware, software, models, simulators, and domain science processes make it increasingly difficult to stay current in the combined areas required to strategically advance Earth system predictability. Domain experts may have limited access to advanced ML techniques, and AI experts have limited knowledge of the science, both of which are necessary to solve known issues of Earth system predictions. In addition, tools (e.g., software, interfaces) that allow domain science expertise to learn from and utilize hybrid models and ML techniques quickly expand the validation and advancement of both techniques and critical stakeholder projections. Investment in this area allows for dedicated support, improving the access and sharing of both positive and negative results, learning/training, dedicated efforts of co-design, standardization enabling advanced data assimilation, and hybrid modeling.

WS.4.2Support for the Future Workforce

Open engagement with academia creates pathways to bringing a new workforce in at the cutting-edge to accelerate advancements. Building on the inclusion of academic researchers and educators in the platform supporting the existing workforce will allow for their projection of skills needed for future employment. This engagement can define and address the gap between traditional education and the training needed to prepare the future workforce. Training programs, workshops, webinar series, virtual and in-person hackathons for more rigorous training and engagement of graduate students, postdoctoral scholars, and early career scientists are urgently needed to keep domain scientists updated on ML research and necessary analytic tools that are relevant to Earth system prediction problems. Investment in these opportunities will accelerate the advancement of Earth system predictability through a prepared, inclusive future workforce.

WS.4.3 Support for Framework Development

Establishing partnerships between national laboratories, universities, industry, and the private sector was also identified as one of the priorities that will help facilitate open communication between AI and domain scientists. An interoperable framework is needed to enable the community to build specific AI methodologies for application in Earth system predictability. Benchmarking datasets must be curated, created, and managed to quickly advance with knowledge discovery. Cross-domain and architecture standards for observations, data assimilation, and hybrid modeling must be achieved for UQ inclusion and propagation through the ModEx environment, as well as to achieve collaborations resulting in explainable results and stakeholder information. This collaborative framework can serve as the backbone for the community's accelerated scientific advancement with mission-specific utilization forked off, eliminating the need for agency-level reorganization.

WS.5 Integrated Research and Infrastructure Centers

All of the above opportunities could be significantly accelerated through the development of an integrated AI4ESP research center that advances scientific understanding and predictions by connecting domain scientists with modelers, experimentalists, and computational and AI/data science experts. Through this center, the workforce can learn and communicate across the complex breadth of AI and ESP research, efficiently share results, catalyze standardization and co-design, and support the framework required for game-changing activities, all of which eliminate stovepiped barriers to advancement. This would drive effective multilateral communication and cross-culture networks and support collaborations across multidisciplinary research for domain scientists to learn advanced ML techniques, as well as for AI/ML experts to better understand the physics and principles associated with Earth system predictions. Such a center will provide computational and storage infrastructure, the necessary benchmark data, a wide variety of models at different scales, software tools for analysis and visualization, and model-data integration. It will also enable the ModEx approach by development of an easy-to-use data assimilation network, which would bring together new AI/ML capabilities with existing investments (e.g., ARM, DAAC, ESS-DIVE, Geo Data Portal, ILAMB, LTER, NGEE, OneStop). Coordination of data, modeling, and AI advances can also take place in these centers, accelerating knowledge-informed learning from AI. It can also lead to new scientific discoveries; improve modeling capabilities that leverage the combination of process models, hybrid approaches, or pure AI algorithms as appropriate; and produce explainable results. The center can also work to harmonize global datasets with consistent terminology and techniques to bridge gaps in spatial and temporal observations. To accomplish all of the above, the center would need to leverage additional resources (e.g., funding support, convening of workshops) toward collaborative projects among ML experts and domain scientists to tackle the critical Earth system prediction problems.

WS.6 Conclusions and Next Steps

To summarize, the DOE's workshop on AI for Earth system predictability covered an extremely wide range of scientific and computational domains and identified several approaches to move the field forward. To effectively leverage AI, significant cultural change and widespread collaboration are necessary. Many institutions across the world are moving towards interdisciplinary domain and computational research seeking to integrate machine learning into Earth sciences.

Achieving the AI4ESP vision will require an unprecedented level of coordination across scientific disciplines and public, private, government, and scientific communities. Priorities identified to address these barriers include (Figure WS-6):

- Creating AI research centers tasked to coordinate and collaborate to more rapidly advance priorities across the various Earth science topics, where the centers would

provide the supporting data and computational infrastructure, mathematical capabilities, and cross-disciplinary expertise to support community ambitions.
- Co-designing frameworks or platforms to enable communities with different missions to efficiently share applicable results, techniques, data, and codes to decrease unnecessary duplication of effort and accelerate the application of AI.
- Determining cross-disciplinary data-sharing standards and creating shareable benchmark and training datasets that bridge organizations.
- Supporting working group activities to investigate major and timely transdisciplinary research questions and quickly enhance or test developments such as workshops, challenges, and hackathons through a center or facility that is staffed to support commonly used data, models, and workflows.
- Developing standards for trustworthy AI, including addressing data biases, ensuring fairness in models, and fostering the ethical and responsible development and use of AI.
- Building public-private partnerships that enable use of commercial tools for research purposes and vice versa.
- Focusing on new efforts to inspire and motivate the next-generation workforce, including training of multidisciplinary scientists, as well as outreach to a broader and more diverse set of academic and laboratory institutions.
- Supporting early success stories to support training, inspiration, and strategic program design, such as through demonstration projects, infusion of AI into existing funded programs, and follow-up “implementation workshops” on key topics to chart roadmaps.

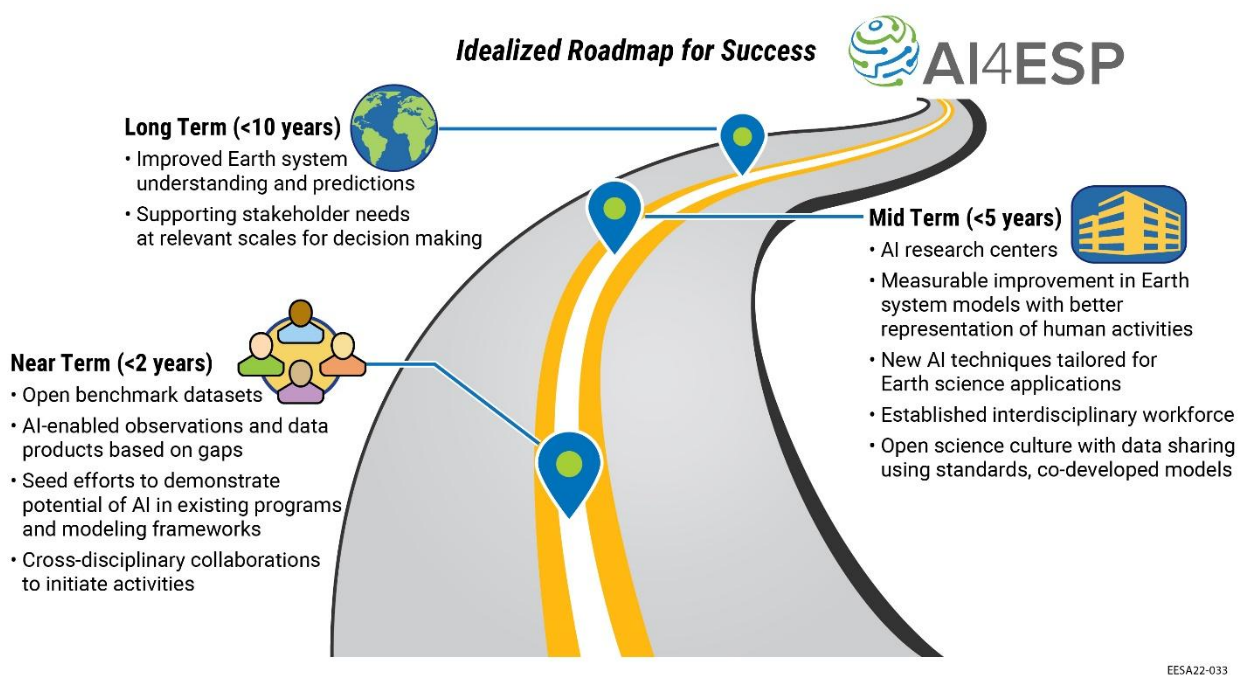


Figure WS-6. Roadmap to the execution of AI4ESP including near-, mid-, and long-term activities (Source: Lawrence Berkeley National Laboratory).

Since the workshop, AI4ESP participants have continued discussions and advancement of information through collaborations and workshops, such as the National Academies Machine Learning and Artificial Intelligence to Advance Earth System Science workshop and the open AMS AI4ESP: Challenges and Opportunities for Advancements special collection call for papers.

The AI4ESP workshop participants volunteered their service by participating and developing this information to provide the community's perspective. The collective intent is to support leaders and decision-makers of the grand challenges and present opportunities available to research priorities. Organizers and participants, actively engaged across the AI and ESP communities, are eagerly seeking opportunities to bring together the breadth of knowledge and expertise necessary to tackle the scientific hurdles to improve prediction through effective development and application of AI.

This report is intended to provide DOE with a high-level, yet forward-looking workshop report to inform its planning and investment agenda, while furthering the development of a community of researchers with common interests in AI and Earth system science with bold innovative thinking. More detailed information related to the sessions is provided in the 17 chapters of this report.

1 Atmospheric Modeling

Authors: Ruby Leung (PNNL), Andrew Bradley (SNL), Noah Brenowitz (AI2), Chris Bretherton (AI2/Univ. of Washington), Peter Caldwell (LLNL), Scott Collis (Argonne), Greg Elsaesser (Columbia Univ./NASA-Goddard), Chris Fletcher (Univ. of Waterloo), Ann Fridlind (NASA-Goddard), Yangang Liu (BNL), Amy McGovern (Univ. of Oklahoma), David Mechem (Univ. of Kansas), Andy Salinger (SNL), Istvan Szunyogh (Texas A&M), Paul Ullrich (UC-Davis), Marcus van Lier-Walqui (Columbia Univ./NASA-Goddard)

Representing a wide range of atmospheric processes that operate across spatiotemporal scales spanning many orders of magnitude, atmospheric models are a key component of climate and Earth system models (ESMs), which are fundamental tools for answering critical questions about the future climate. With typical grid spacings of 25–100 km, atmospheric models used in ESMs must divide processes into resolved and subgrid, with the latter represented by subgrid physics parameterizations. These parameterizations include many assumptions, such as fractional cloud cover, convective mass flux closures, and orographic drag associated with subgrid-scale processes. Despite decades of research, systematic errors in subgrid parameterizations and limited/subjective model calibration have contributed to substantial biases in atmospheric models. Such deficiencies in atmospheric models are a major source of the large uncertainties in multimodel projections of the future climate, particularly at regional scales that matter the most for addressing climate impacts and adaptation.

By more explicitly representing subgrid processes, cloud-resolving models (CRMs) with 1-km or finer resolution may avoid the use of some subgrid parameterizations or use simpler ones; however, they are and will remain too expensive for Earth system modeling at multidecadal to century timescales for the next decade. Nevertheless, CRMs may provide important global reference data for employing artificial intelligence and machine learning (AI/ML) to improve modeling of subgrid physics. Doing so could improve the fidelity and predictive skill of ESMs with computationally affordable grid spacings of 25 km and larger. More generally, AI/ML could help us develop and train atmospheric models to be optimally skillful from weather to climate timescales by improving modeling of subgrid physics and guiding/automating model calibration. By creating "differentiable" models with smooth relationships between inputs and outputs, ML will better enable sensitivity analysis, physical interpretation, parameter calibration, and data assimilation. Parallel to/in conjunction with ML-enabled model improvements, ML may also facilitate understanding of the atmosphere and climate systems (along with relevant applications), with the goal to provide more skillful predictions and credible projections, backed by physical explanations and interpretations, to support decision-making.

In sum, to advance atmospheric modeling, AI/ML offers the potential to substantially improve model skill by improving **modeling of subgrid physics** and **guiding/automating model calibration**. AI/ML also plays a critical role in **enhancing model applications** for decision-making. Each of these endeavors represents a grand challenge for atmospheric modeling that encompasses many conceptual and technical challenges. In what follows, we discuss these grand challenges, summarize the associated state-of-the-science, and highlight research opportunities and priorities to address the grand challenges. Crosscutting challenges related to uncertainty, use of observations, AI/ML methods, generalizability, infrastructure, and trustworthiness are briefly discussed in the context of each individual grand challenge.

1.1 Grand Challenges

1.1.1 Improving Modeling of Subgrid Physics for Atmospheric Modeling across Scales

The grand challenge of improving modeling of subgrid physics in atmospheric models centers on correcting biases in model simulations/predictions relative to observations/high-resolution models. For example, there is a gap in our ability to predict precipitation, including its regional and local statistics (i.e., precipitation intensity, duration, and frequency). Precipitation extremes, in particular, are poorly predicted and have outsized socio-economic impacts. Specific examples of challenges include using ML parameterizations to represent boundary layer processes, shallow and deep convection, cloud microphysics and aerosol-cloud interactions, all of which contribute to the diverse climate sensitivity and transient climate response in multimodel ensembles. Complementing the development of ML parameterizations, recent efforts to model the discrepancy or residual between low-resolution simulations and cloud-resolving simulations using AI/ML have shown some promise in correcting for biases in low-resolution atmospheric models (Bretherton et al. 2021) (Figure 1-1). As the correction acts as an ML column parameterization, this bias correction approach is analogous to the use of ML parameterizations for hybrid modeling that combines ML models of subgrid processes with conventional atmospheric models. Ideally, efforts directed at developing ML-derived parameterizations should account for: (1) uncertainty quantification (UQ); (2) knowledge transfer across scales; (3) ensuring that the domain of the training data is sufficient to span a broad range of relevant climate regimes for generalization; (4) creating and curating high-quality data; and (5) enabling physical understanding and interpretability of the ML-parameterization results, with the goal of creating efficient, generalizable, and robust hybrid models. Notably, many of the issues and approaches associated with AI/ML used in parameterizations are also common to data assimilation (DA), with parameter estimation being one example. Bringing ML and DA into a closer relationship could reap big rewards, particularly as applied to estimation of parameters in physically based atmospheric and Earth system models.

1.1.2 *Transforming Climate Modeling through AI/ML-guided/automated Model Calibration*

Climate model tuning is among the most labor-intensive and challenging parts of the model development process, often taking years to complete and requiring many thousands of simulated years. Tuning is typically performed by hand using optimization criteria that are poorly defined and subjective (Hourdin et al. 2017; Schmidt et al. 2017). Moving to AI-based calibration could transform climate modeling by: (1) increasing objectivity and reproducibility of the model development process; (2) improving model skill relative to observational targets with well-defined uncertainties; (3) permitting rigorous quantification of parametric uncertainty in model predictions, such as climate sensitivity and extreme precipitation change; (4) making it possible to tune very-high-resolution models (e.g., by leveraging information from cheaper simulations); (5) exposing the actual impact of each model change by enabling full retuning after each major model change; and (6) identifying the observational deficiencies that contribute most to predictive uncertainty, for example, by comparing ensembles which either include or zero out uncertainty in particular parameters. Auto-calibration is a particularly attractive AI grand challenge because, unlike most climate applications, it does not involve extrapolation beyond the available data and reference observations.

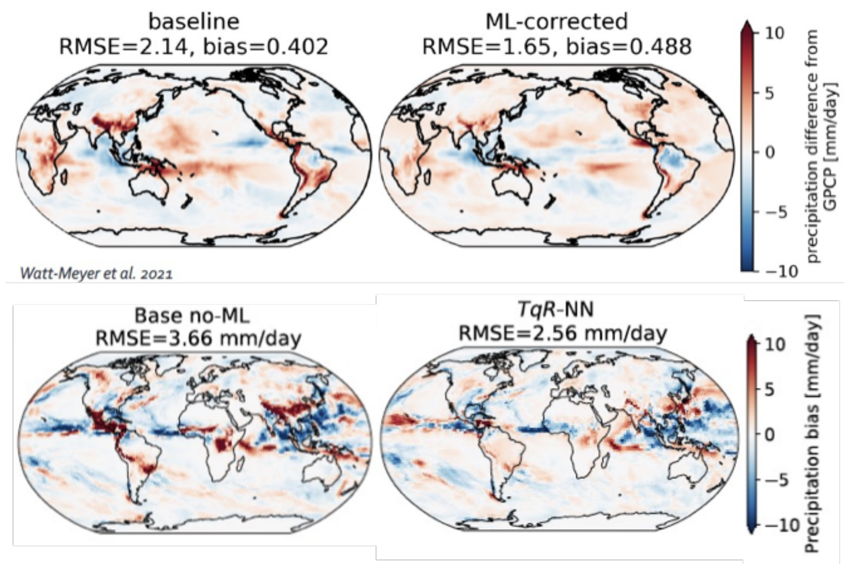


Figure 1-1. Using a corrective ML method in which the coarse model state is nudged to the reference state, with the nudging tendencies machine learned, shows promise. Upper panel: root-mean-square deviation (RMSE) of precipitation is reduced from 2.14 mm/day (left) to 1.65 mm/day (right) with ML-correction trained using observations (Watt-Meyer et al. 2021). Lower panel: Similar reduction in precipitation RMSE from 3.66 mm/day (left) to 2.56 mm/day (right) is also achieved with ML-correction trained using fine grid model output (Source: Bretherton et al. 2021; figure used with permission).

1.1.3 Enhancing Atmospheric Model Applications through AI/ML for Decision-Making

Either used as stand-alone models or in hybrid models, AI/ML has the potential to deliver salient information and data to practitioners and decision-makers. However, significant developments are needed to deliver on this promise, including in these areas:

(1) Identification and seamless prediction of extreme events or other outliers represent key areas where AI/ML systems hold significant promise. However, the capacity of AI/ML systems to produce out-of-sample prediction is poorly understood, particularly as these systems tend to emphasize the mean. There is a need to improve/adopt ML methodologies for systematically reducing climate drift and bias while accurately simulating the weather that leads to extreme events.

(2) New developments in AI/ML are needed to better represent complex relationships that emerge in the atmosphere and coupled systems including human impacts. Inevitably, decisions that are made based on climate data need to be contextualized within the scope of the coupled human-Earth system. Comprehensive AI/ML systems that incorporate impacts from multiple sectors may enable researchers to better identify optimal decisions in a given weather or climate context.

(3) Supervised, self-supervised, and unsupervised deep learning can be used in regime classification, model evaluation, and data-driven discovery, such as discovery of unknown relationships hidden in big complex data, in models, and observations.

(4) There is an outstanding need for building confidence and trust in AI/ML systems, so that decision-makers, practitioners, and other end-users feel comfortable in employing these systems in practice. Explainability and interpretability are essential components in the exchange between scientists and end-users to gain confidence that AI/ML systems are capturing relevant processes and interactions. How trust differs for different end-user needs should be incorporated into the development process for AI/ML systems.

1.2 State-of-the-Science

1.2.1 Modeling of Subgrid Physics

A number of state-of-the-art approaches for using AI/ML to improve modeling of subgrid physics are in current use. These include ML-only methods, hybrid ML/physics approaches, and emulators derived from perturbed physics ensembles. A growing number of efforts exist to replace or supplement column parameterizations with ML models that are either more accurate (being trained using higher-resolution simulations) or more computationally efficient (replacing complex physics code with a speedy surrogate) (e.g., Gentine et al. 2018; Brenowitz and Bretherton 2018, 2019; Brenowitz and Beucler et al. 2020; Brenowitz and Henn et al. 2020; McGibbon and Bretherton 2019; Wang et al. 2021; Bretherton et al. 2021; O'Gorman and Dwyer 2018). While some ML parameterizations are monolithic (e.g., Bretherton et al. 2021), there is an attempt to create separate ML models, for example, one for each conventional process. The latter is motivated either by an attempt to mimic conventional divisions among processes (e.g., convection and cloud microphysics are conventionally represented by different physics parameterizations) or for reasons of efficiency. For example, in Yuval and O'Gorman (2020), one ML column model predicts subgrid vertical advection, cloud microphysics, sedimentation, precipitation, and radiative heating, while a second predicts turbulent diffusivity and corrections to the surface fluxes. Wang et al. (2021) use four ML models for efficient training. Multiple efforts find that online use of an ML-based parameterization is not always robust (e.g., Brenowitz and Beucler et al. 2020; Brenowitz and Henn et al. 2020), and there is a related issue of preserving physical properties such as conservation in ML models. Closing the gap between offline and online ML model performance is a general challenge for hybrid modeling. Blending Bayesian and DA methods results in a natural fit for many ML approaches, but this approach is currently underutilized. Overall, developing ML parameterizations for hybrid modeling remains a significant challenge.

1.2.2 Model Calibration

Although automated calibration has been studied for years (e.g., Jackson et al. 2008; Lee et al. 2011; Zhang et al. 2015; Liu et al. 2021), most modeling centers still tune by hand, except for recent exceptions. Impediments to widespread adoption of automated calibration include: (1) ambiguity about what aspects of model skill to optimize (e.g., variables, mean vs. variance, or transient evolution); (2) difficult-to-use workflows (e.g., hard-coded parameter values, inflexibility in job submission process); (3) lack of clarity about the best techniques to use (e.g., emulation, optimization, Gaussian processes, convolutional neural networks [CNNs], etc.); (4) communication disconnect between those responsible for creating new models and those engaged in auto-calibration research; (5) lack of well-defined observational targets (including observational uncertainty bounds); and (6) the need for multidimensional cost functions for optimization. Auto-calibration typically occurs in three stages: (1) identify the most important uncertain parameters in the model and conduct a perturbed physics ensemble (PPE) to sample them; (2) create an emulator which predicts the model output variables included in the optimization as a function of the input parameter settings; and (3) apply an optimization algorithm to the emulator to find one or more sets of best parameter settings. Identifying the best approach to emulation and optimization are active research topics.

1.2.3 Model Applications

Many AI/ML technologies are now being investigated to tackle the challenges for enhancing model applications. Examples include these: (1) Many uses of AI/ML are being explored to improve forecasts (Figure 1-2), track/classify features in observations/simulations, evaluate models, discover new knowledge, and catch long/nonlinear dependence. (2) Digital twins are under development that aim to seamlessly integrate observations, modeling experiments, and model data to support decision-making (e.g., Bauer, Stevens, and Hazeleger 2021). This technology requires use of AI/ML to improve realism and efficiency of the digital representations of atmosphere and/or other Earth system processes and integration of high-performance computing, along with associated computational challenges to enable extreme-scale computing and real-time exploitation of observation data. (3) There are ongoing efforts to develop catalogs beyond data repositories to advance ML applications in Earth system science. For example, ML model catalogs (https://mlhub.earth/models) have been developed to provide a library of existing ML models that users can easily find and put into practice. (4) Adversarial models are being used to discover model limitations and identify unphysical behaviors of ML models. (5) Knowledge graphs have been used to better understand what the ML systems have learned. (6) Extreme ensembling and stacked learning are used to combine predictions from individual ML models to boost prediction skill. (7) To provide confidence in the knowledge and predictions advanced by AI/ML, explainable and interpretable AI is critically important as it helps to characterize model accuracy, fairness, transparency, and outcomes.

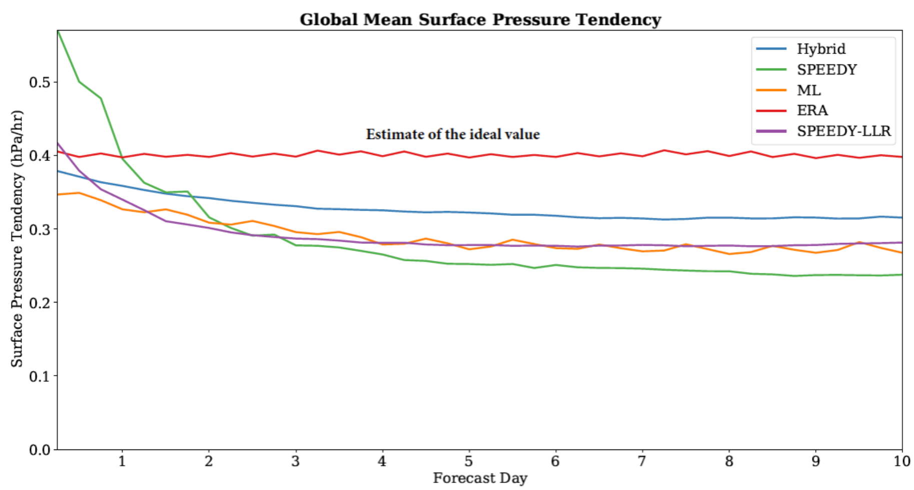


Figure 1-2. A hybrid model using the Combined Hybrid-Parallel Prediction (CHyPP) technique on the Simplified Parameterization, primitive Equation Dynamics (SPEEDY) atmospheric general circulation model is trained using ERA5 reanalysis data. One hundred 21-day forecasts have been verified and compared with the benchmarks (SPEEDY, ML only, and SPEEDY-LLR). The hybrid model state is well balanced throughout the forecasts and produces more realistic surface pressure tendencies than the benchmarks. The ML component of the hybrid model is based on parallel reservoir computing in local subdomains (Source: Arcomano et al. 2022; figure used with permission).

1.3 Experimental, Data, and Modeling Opportunities

1.3.1 Modeling Subgrid Physics

Current ML-derived parameterizations do not sufficiently address uncertainty. Characterizing uncertainty should be front and center in parameterization development, ideally with such uncertainty being characterized by type (structural, parametric, initial/boundary condition, etc.). Developing a better representation of uncertainty could employ a Bayesian approach and would have a number of advantages such as the ability to generate meaningful physics ensembles. An

ML-derived parameterization should transfer knowledge learned across scales. Ideally, an ML-derived parameterization should be scale-aware and include scale-dependent uncertainty estimates.

Progress is currently hampered by lack of sufficient data suitable for developing parameterizations, either in the form of observations or reference model output, for ML training. Training on datasets that insufficiently span the parameter space can result in extrapolation errors. These extrapolation errors become even more likely when using ML-derived parameterizations developed for the present climate and then extrapolated to future, high-CO_2 environments. Training should also include edge cases, such as high-latitude processes, that are particularly hard for ML to learn accurately because they do not occur everywhere and all the time. These are well-understood points, but too often training datasets do not sufficiently train a parameterization for all situations it may encounter. Additionally, a vast quantity of data at the weather scale is routinely collected, for example, by the ARM observatories or NWS radar network, but little of it is used to inform parameterizations via systematic inference. Instead, parameterizations are typically improved in global simulations using satellite climatologies, or via synthetically generated "reference model" data, which may have their own structural and parametric uncertainties and biases. ML techniques should be investigated that can make observational insights from the weather scale available to improve parameterizations in climate and weather models. Such techniques need to consider potential challenges presented by observation data such as sparse space/time distribution and coverage and how well budgets of mass, energy, and water are closed in observations from different sources.

ML-derived parameterizations should strive to enforce physical constraints such as water conservation, Clausius-Clapeyron relation, and causality that human designers deliberately build into parameterizations to improve generality. ML models developed without physical (conservation law) constraints should still provide physically consistent results. An ML-derived parameterization trained on model output that is physically inconsistent may induce unphysical behavior in an ESM based on conservation laws.

It is important to recognize that meaningful parameterizations can still be formulated even if the governing physical laws are not apparent or observable, especially for less well-understood processes and/or processes that are not governed only by physical processes. Hence, ML approaches may diverge significantly from physically based parameterizations, highlighting the need to interpret such differences for credible use of ML parameterizations. In the case of the more physical aspects of these systems, DA approaches with forward modeling may help constrain parameterizations using data that are available.

1.3.2 Model Calibration

The most exciting opportunity enabled by auto-calibration is the ability to derive multiple optimal parameter settings instead of the single "best" tuning provided by hand-tuning. With multiple tunings, parametric uncertainty could be easily included in any analysis or prediction made using a model. More ambitiously, a likelihood function covering all parameter choices could be established. Establishing the uncertainty in the predictions we make is essential for climate information to be useful to the public, so this capability should be prioritized.

Making training/testing datasets publicly available would accelerate auto-calibration research by avoiding the expensive and slow step of creating PPEs. Archives of observational datasets with well-defined uncertainty bounds would also be a major advance in model-data fusion for climate research. Making these collections of datasets publicly available would expand the group of researchers able to contribute to this effort. Having common datasets across research groups would also enable more rigorous cross comparison of the skill of proposed techniques, making it easier to identify the best approaches and simultaneously identify data gaps for future observing system design. Similarly, while choice of cost function will probably always need to be determined by guess-and-check approaches and will differ between modeling centers, having standardized cost functions for comparing methodologies would be useful. Creation of these public datasets and cost functions could be done very rapidly. Infrastructure to support dissemination of large PPEs would be needed.

Auto-calibration also opens the door for leveraging cheaper simulations in the tuning of very expensive model configurations. Cheaper configurations definitely contain information relevant to their more expensive cousins, but not all of their behavior is relevant. AI can easily separate the useful versus irrelevant information from these simulations (e.g., Anderson and Lucas 2018). Cheaper simulations could be lower-resolution global runs, but they could also be single-column runs, limited-area cloud-resolving simulations, or global simulations of very short duration.

Besides the aforementioned opportunities, advancing use of AI/ML in model calibration would also require addressing common challenges, including the lack of sufficient data and data quality and the multiple sources of uncertainty associated with data and model experimental setup (e.g., initial/boundary conditions, simulation length) used in auto-calibration.

1.3.3 Model Applications

To tackle the grand challenges described above, we have identified three outstanding opportunities that should be tackled. First, there is a need to entrain decision-makers, practitioners, and other end-users in the development process to learn where their needs for systems or datasets are presently unmet or where AI/ML can more rapidly or accurately deliver information. There are significant benefits to both scientists and stakeholders when a co-production model is pursued. For instance, researchers may be able to identify unique or previously unknown datasets, or may learn about the interactions, drivers, or thresholds that exist in decision networks. Similarly, end-users can learn about recent developments relevant to their needs and consequently can gain a better understanding of what is achievable with existing technology. The co-production model would also build trust in AI systems by allowing end-users to learn more about how these systems are designed and operate and to provide feedback to researchers on specific connections that need to be captured in these systems.

Second, to meet the data needs for ML, there is a clear and outstanding need for datasets with complete and consistent provenance, metadata, availability, and quality control. Observational uncertainty should be incorporated to avoid overconfidence. Tools should be available to both researchers and the general public to search and extract/slice data that are useful for AI studies. To limit the need for transferring large datasets, these capabilities should be available "in the cloud." Third, we need to build trust from the climate research community by showing that models that use AI/ML provide results that are consistent with the laws of physics. It should be demonstrated that AI/ML can be used, not only for prediction by a "black box," but also to answer questions of "why."

1.4 Research Priorities

1.4.1 Near-term Goals

1.4.1.1 Modeling Subgrid Physics

Near-term goals in this area include efforts to:

- Ensure that new ML-derived parameterizations are physically consistent (i.e., bound by conservation laws).
- Better understand the domain of the training dataset used for parameterization development.
- Discover the characteristics of ML-based parameterizations, which can lead to instabilities when incorporated in the atmospheric model, despite being good data fits.
- Augment cloud-resolving simulations with coarsening to account for topography and land surface types, careful selection of variables to support multiple ML model inputs/outputs (I/Os), and I/O modules to assemble training set members.
- Archive training data in standard formats and create a central repository of training data to be made available to all researchers. Training data should include not only cloud-resolving simulations and high-quality reanalysis data, but also integration of a wide range of observations such as data collected by DOE ARM for different climatic regimes, satellite data covering broad areas, routinely collected meteorological data and forecasts, and aircraft measurements collected by field campaigns/intensive observation periods (IOPs) or routine flights.

- Gain interdisciplinary knowledge from new teams that span atmospheric scientists to data scientists.
- Gather low-hanging fruit of training improvements to parameterizations-based high-resolution or observational data.
- Build infrastructure for datasets, software, testing, validation, and training workflows.
- Develop "super-high" resolution models such as direct numerical simulations (DNSs) to fill in critical gaps in physical understanding and measurements for generating synthetic data.

1.4.1.2 Model Calibration

Near-term goals in this area include efforts to:

- Make one or more perturbed physics ensembles for auto-calibration publicly available.
- Establish observational dataset characteristics needed for observational targets (e.g., minimum uncertainty characterization).
- Identify a suite of standard cost functions that researchers can implement to compare methodologies.
- Create or identify easy-to-use workflows for performing auto-calibration.
- Quantify uncertainty in observations used in model calibration, for example, based on discrepancies across multiple, independent datasets.

1.4.1.3 Model Applications

Near-term goals in this area include efforts to:

- Jointly develop standards for AI relevant datasets that enable their distribution and utility.
- Develop a searchable catalog of AI4ESP-relevant datasets that conforms to FAIR principles: findability, accessibility, interoperability, and reuse of digital assets.
- Develop a catalog of AI models that is searchable by problem or discipline.
- Develop a resource for tracking lessons learned, necessitated by the rapid developments being made in AI/ML.
- Support more widespread pursuit of open-source development practices, ensure comprehensive and complete documentation of models, and leverage modern development practices to ensure that high-quality and error-free codes are distributed to end-users.

1.4.2 5-year Goals

1.4.2.1 Modeling Subgrid Physics

Mid-term goals in this area include efforts to:

- Emphasize rudimentary uncertainty estimates in ML-derived parameterizations.
- Strive for physical understanding and interpretability.
- Use ML to discover physically salient features in observational data.

- Replace select parameterizations with ML-based parameterizations, particularly where the current parameterizations are themselves heuristic curve fits or where large efficiency gains can be made by ML-trained surrogates.
- Design a user-friendly ML system and workflow on DOE computing resources for improving or replacing the physical parameterizations of a 50-km DOE AGCM that runs end-to-end within a day.
- Experiment with different ML architectures and training methodologies to optimize the system, including ML methodologies for minimizing mean-state bias.
- Devise methods for the creation of ML-based parameterizations that automatically maintain the stability of the coupled atmospheric model at the same time steps as physics-based parameterizations.
- Test hybrid models of atmosphere coupled to other components for a fundamentally more accurate low-resolution, full-Earth simulator. Coupling hybrid models may not differ substantially from coupling conventional ones, since the coupling boundaries are defined by the resolved rather than subgrid models.

1.4.2.2 Model Calibration

Mid-term goals in this area include efforts to:

- Systematically test auto-calibration techniques in order to clarify best-practice approaches.
- Develop multifidelity methods for using cheaper simulations to tune higher-resolution configurations.
- Establish a community repository for observational datasets, with uncertainty characteristics adequately derived from more sophisticated methods, for auto-calibration.
- Develop collaborations with several modeling centers to use auto-calibration for their next model release, paying careful attention to the cost functions being used and the commonalities and differences, to determine if best practices for cost function design are being established.
- Construct multiple sets of optimal tunings for auto-tuned models and explore the resulting parametric uncertainty in model predictions.

1.4.2.3 Model Applications

Mid-term goals in this area include efforts to:

- Establish an operational and heavily utilized data and model service, recognizing a continuing need to invest in maintaining these systems, and ensuring that they are modernized and streamlined for user needs.
- Support greater availability, usability, and documentation of models that have demonstrable credibility for practitioner applications.
- Build/apply multiphysics and multiscale AI/ML to address the inherent multiphysics-multiscale atmospheric phenomena.

1.4.3 10-year Goals

1.4.3.1 Modeling Subgrid Physics

Long-term goals in this area include efforts to:

- Provide more sophisticated uncertainty estimates.
- Focus on addressing scale-aware, ML-derived parameterizations.
- Develop AI systems to automate the selection of parameterizations and resolutions to balance accuracy and computational costs. For some applications, full ML models may replace the current models.
- Evaluate robust use of hybrid atmospheric models in coupled Earth system models.
- Work with the applications community to adapt prototypes of hybrid models to their needs (e.g., by adding aerosols, chemistry, automatic ML downscaling, uncertainty estimation, observational simulators, etc.) and demonstrate to their satisfaction that the ML is improving the model based on metrics relevant to their objectives.
- Work with the IPCC ESM community to cross-fertilize ML best practices for atmospheric modeling.

1.4.3.2 Model Calibration

Long-term goals in this area include efforts to:

- Use auto-calibration at each step of model development to decide whether each new development effort truly improves the model.
- Develop a full Bayesian approach to auto-calibration calculating likelihood functions of potential tunings so the uncertainties in the tunings can be propagated through to uncertainties in predictions.
- Exploit the auto-calibration process across multiple climate models to design Earth-observing systems which will maximally reduce uncertainties in model predictions.

1.4.3.3 Model Applications

Long-term goals in this area include efforts to:

- Develop workflows that allow data from models and observations to be available in the cloud as they are produced.
- Provide on-demand services to run AI models on this data as the data are produced, providing targeted value products or analysis for associated researchers and communities.
- Provide AI models that are widely accepted, trusted, and used throughout the community.

1.5 References

Anderson, Gemma J., and Donald D. Lucas. 2018. "Machine Learning Predictions of a Multiresolution Climate Model Ensemble." *Geophysical Research Letters* 45 (9): 4273–80. https://doi.org/10.1029/2018GL077049.

Arcomano, Troy, Istvan Szunyogh, Alexander Wikner, Jaideep Pathak, Brian R. Hunt, and Edward Ott. 2022. “A Hybrid Approach to Atmospheric Modeling That Combines Machine Learning With a Physics-Based Numerical Model.” *Journal of Advances in Modeling Earth Systems* 14 (3): e2021MS002712. https://doi.org/10.1029/2021MS002712.

Bauer, Peter, Bjorn Stevens, and Wilco Hazeleger. 2021. “A Digital Twin of Earth for the Green Transition.” *Nature Climate Change* 11 (2): 80–83. https://doi.org/10.1038/s41558-021-00986-y.

Brenowitz, N. D., and C. S. Bretherton. 2018. “Prognostic Validation of a Neural Network Unified Physics Parameterization.” *Geophysical Research Letters* 45 (12): 6289–98. https://doi.org/https://doi.org/10.1029/2018GL078510.

Brenowitz, Noah D., Tom Beucler, Michael Pritchard, and Christopher S. Bretherton. 2020. “Interpreting and Stabilizing Machine-Learning Parametrizations of Convection.” *Journal of the Atmospheric Sciences* 77 (12): 4357–75. https://doi.org/10.1175/JAS-D-20-0082.1.

Brenowitz, Noah D., and Christopher S. Bretherton. 2019. “Spatially Extended Tests of a Neural Network Parametrization Trained by Coarse-Graining.” *Journal of Advances in Modeling Earth Systems* 11 (8): 2728–44. https://doi.org/https://doi.org/10.1029/2019MS001711.

Brenowitz, Noah D., Brian Henn, Jeremy McGibbon, Spencer K. Clark, Anna Kwa, W. Andre Perkins, Oliver Watt-Meyer, and Christopher S. Bretherton. 2020. “Machine Learning Climate Model Dynamics: Offline versus Online Performance.” https://arxiv.org/abs/2011.03081.

Bretherton, Christopher S., Brian Henn, Anna Kwa, Noah D. Brenowitz, Oliver Watt-Meyer, Jeremy McGibbon, W. Andre Perkins, Spencer K. Clark, and Lucas Harris. 2021. “Correcting Coarse-Grid Weather and Climate Models by Machine Learning from Global Storm-Resolving Simulations.” *Earth and Space Science Open Archive*, 39. https://doi.org/10.1002/essoar.10507879.1.

Gentine, P., M. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis. 2018. “Could Machine Learning Break the Convection Parameterization Deadlock?” *Geophysical Research Letters* 45 (11): 5742–51. https://doi.org/https://doi.org/10.1029/2018GL078202.

Hourdin, Frédéric, Daniel Williamson, Catherine Rio, Fleur Couvreux, Romain Roehrig, Najda Villefranque, Ionela Musat, Laurent Fairhead, F. Binta Diallo, and Victoria Volodina. 2021. “Process-Based Climate Model Development Harnessing Machine Learning: II. Model Calibration From Single Column to Global.” *Journal of Advances in Modeling Earth Systems* 13 (6): e2020MS002225. https://doi.org/10.1029/2020MS002225.

Jackson, Charles S., Mrinal K. Sen, Gabriel Huerta, Yi Deng, and Kenneth P. Bowman. 2008. “Error Reduction and Convergence in Climate Prediction.” *Journal of Climate* 21 (24): 6698–6709. https://doi.org/10.1175/2008JCLI2112.1.

Lee, H. J., Y. Liu, B. A. Coull, J. Schwartz, and P. Koutrakis. 2011. “A Novel Calibration Approach of MODIS AOD Data to Predict PM_2.5 Concentrations.” *Atmospheric Chemistry and Physics* 11 (15): 7991–8002. https://doi.org/10.5194/acp-11-7991-2011.

Liu, Wuji, Qianwen Ye, Chase Q. Wu, Yangang Liu, Xin Zhou, and Yunpeng Shan. 2021. "Machine Learning-Assisted Computational Steering of Large-Scale Scientific Simulations." In *2021 IEEE Intl Conf on Parallel Distributed Processing with Applications, Big Data Cloud Computing, Sustainable Computing Communications, Social Computing Networking (ISPA/BDCloud/SocialCom/SustainCom)*, 984–92. https://doi.org/10.1109/ISPA-BDCloud-SocialCom-SustainCom52081.2021.00138.

McGibbon, J., and C. S. Bretherton. 2019. "Single-Column Emulation of Reanalysis of the Northeast Pacific Marine Boundary Layer." *Geophysical Research Letters* 46 (16): 10053–60. https://doi.org/https://doi.org/10.1029/2019GL083646.

O'Gorman, Paul A., and John G. Dwyer. 2018. "Using Machine Learning to Parameterize Moist Convection: Potential for Modeling of Climate, Climate Change, and Extreme Events." *Journal of Advances in Modeling Earth Systems* 10 (10): 2548–63. https://doi.org/https://doi.org/10.1029/2018MS001351.

Schmidt, G. A., D. Bader, L. J. Donner, G. S. Elsaesser, J.-C. Golaz, C. Hannay, A. Molod, R. B. Neale, and S. Saha. 2017. "Practice and Philosophy of Climate Model Tuning across Six US Modeling Centers." *Geoscientific Model Development* 10 (9): 3207–23. https://doi.org/10.5194/gmd-10-3207-2017.

Wang, X., Y. Han, W. Xue, G. Yang, and G. J. Zhang. 2021. "Stable Climate Simulations Using a Realistic GCM with Neural Network Parameterizations for Atmospheric Moist Physics and Radiation Processes." *Geoscientific Model Development Discussions* 2021: 1–35. https://doi.org/10.5194/gmd-2021-299.

Watt-Meyer, Oliver, Noah D. Brenowitz, Spencer K. Clark, Brian Henn, Anna Kwa, Jeremy McGibbon, W. Andre Perkins, and Christopher S. Bretherton. 2021. "Correcting Weather and Climate Models by Machine Learning Nudged Historical Simulations." *Geophysical Research Letters* 48 (15): e2021GL092555. https://doi.org/https://doi.org/10.1029/2021GL092555.

Yuval, Janni, and Paul A. O'Gorman. 2020. "Stable Machine-Learning Parameterization of Subgrid Processes for Climate Modeling at a Range of Resolutions." *Nature Communications* 11 (1): 3295. https://doi.org/10.1038/s41467-020-17142-3.

Zhang, T., L. Li, Y. Lin, W. Xue, F. Xie, H. Xu, and X. Huang. 2015. "An Automatic and Effective Parameter Optimization Method for Model Tuning." *Geoscientific Model Development* 8 (11): 3579–91. https://doi.org/10.5194/gmd-8-3579-2015.

2 Land Modeling

Authors: Beth Drewniak (Argonne), Charlie Koven (LBNL), Dan Ricciuto (ORNL), William Riley (LBNL), Pamela Weisenhorn (Argonne), Xiaojuan Yang (ORNL), Qing Zhu (LBNL)

2.1 Grand Challenges

Many of the grand challenges in land modeling are related to processes that are not fully understood or lack quality measurements. These challenges are compounded since we cannot capture small-scale features and dynamics of the land surface with the coarse grid structure of Earth system models. There are many examples of patterns we do not understand in nature that affect the ability of models to predict behavior. Our discussion during the Land Modeling session focused on ways we can resolve small features across a wide range of scales, including processes in models that lack quantity and quality measurements, and how we might leverage AI/ML techniques to enhance our current modeling efforts. We highlight some of the grand challenges below that have historically been and continue to be difficult to overcome in land modeling.

2.1.1 Uncertainty Quantification

The difficulty quantifying uncertainty was highlighted in multiple sessions during the workshop and cannot be underestimated. Uncertainty quantification is desired across a variety of areas including uncertainty in Earth system models (parametric and structural), uncertainty in ML models, uncertainty in data (including data bias), and uncertainty in parametric bounds. Furthermore, understanding how uncertainty propagates is of concern. Addressing UQ in models requires a large number of model ensembles, which are typically limited by computational resources. We need methods to increase ensembles either through increased throughput or using surrogates and emulators that run faster than the traditional model. Improvements in retrievals of remote sensed data (e.g., plant traits) can also help with UQ by improving our understanding of parametric uncertainty and can help guide process representation in models. Finally, the role of urbanization on historical and future land use/cover and the impact from changes in land use, land cover, and management practices play a considerable role in uncertainty. Improving our predictions of these human behaviors and decision-making will be critical to improving our Earth system predictions.

2.1.2 Scaling

Land surface modeling covers a variety of different spatial and temporal scales that are difficult to bridge, for example, spanning from microns (e.g., microbes) to hundreds of kilometers (e.g., ecosystems) and processes that occur from seconds (e.g., gross primary productivity [GPP]) to hundreds of years (e.g., soil carbon turnover). In addition, data must be extrapolated from site measurements to the globe and account for changes in parameters across a variety of environments and ecosystems. Similarly, data collected at different temporal frequencies across experiments, remote sensing, and satellites must be integrated into a uniform framework. Identifying and generalizing trends across ecosystems that have different biogeochemistry responses and including fine-scale events such as hot spots and hot moments or land use and land cover change in these extrapolations are challenges. As such, models need subgrid scales, including subgrid parameterizations, to capture the full complexity within a grid. Although models have higher confidence simulating short-term processes and fluxes, emergent behavior leading to long-term dynamics is more difficult to capture. Finally, understanding slow processes, acclimation, and legacy is equally important. We need a change in culture that connects modelers with experimentalists. Understanding challenges, limitations, and the needs of each group would help design the next generation of science development.

2.1.3 Extreme Events and Disturbance

Observations of extreme events are limited. Systematic observation of extreme events and disturbance is difficult, because these events are by definition relatively rare. Thus, long records with high spatial resolution and large observing areas are needed to robustly train ML models. As a result, a key challenge is to build the datasets needed to characterize disturbances and their effects on terrestrial ecosystems. While some types of extremes, such as wildfire, are directly observable from remote sensing, others, such as drought-driven tree mortality or forest blowdowns, must be inferred by detecting temporal discontinuities in ecosystem variables. Because of the high resolution needed to detect such discontinuities from satellite observations, the data stream has historically been too enormous to process at global scales. Therefore, ML techniques for pattern recognition to categorize disturbance could be useful; see, for example, the detection of forest fires using a neural network trained on spatial data (Angayarkkani and Radhakrishnan 2010). As a result of the lack of training data, predicting extreme events and their effects on ecosystem structure and function remains a grand challenge. One solution is using space-for-time substitutions. Space-for-time substitution allows the calculation of a temporal trend from spatially variable data. This technique is useful when long-term observations are not available and relies on the assumption that spatial and temporal variability are equivalent. It has been widely used in ecology to study biodiversity (Blois et al. 2013), nutrient cycling (Frauendorf et al. 2020), and genetics (Wogan and Wang 2018), to name a few. A suitable application of space-for-time substitution is understanding recovery from disturbance, which requires long-term observational records that may not be available. Instead, data from multiple sites with varying time since disturbance can be used to recreate a recovery timeline. Alternatively, we need to identify proxies that can be used from nontraditional data sources (e.g., cell phones, WiFi camera networks). For example, cell phone signals can be used to gauge rainfall and improve weather forecast models (Messer, Zinevich, and Alpert 2006). WiFi networks can provide information that links people, objects, and places. This would allow ML to take good observations to inform data-poor extreme events or use process models in lieu of observations. Another approach is to use ML for adaptive sampling during a disturbance event to enhance current observations or determine where to optimally place sensors. Building a database for these types of observations is needed, but the challenge is still dealing with outlier predictions and extrapolating out-of-sample. Finally, we also need to understand impacts by relating drivers to consequences, for example, by combining multiple different datasets: on extreme event occurrence such as wildfire, on ecosystem structure such as lidar or canopy coverage, and on ecosystem function such as evapotranspiration or productivity, all at scales that resolve disturbance processes and for sufficiently long time periods to build statistical models.

2.1.4 Data Challenges

Limitations on available observational data represent a significant challenge for land models. Observational data are used and needed for model input, benchmarking, parameter calibration, and process representation. But given the distributed nature of measurements, questions remain about how much data is really available and what is unknown. Modelers need to map the unmeasured, which can be done by identifying and using proxies. There are challenges integrating historical data, particularly when the data are not digital. We also need to bridge gaps in spatial and temporal scales caused by data sparsity and scarcity. Observational data are collected across spatial scales ranging from the centimeter scale to hundreds of kilometers (e.g., satellite data) and across temporal scales ranging from a single measurement (snapshot) to continuous (e.g., flux towers). Adding to this challenge is the explosive growth in data collected across multiple disciplines that need to be stored and mined for information. Finding ways to integrate long-term observations, remote sensing retrievals, nutrient cycles under management schemes, policy impacts, and relationships between plant traits requires uniform ways to access data. This calls for the creation of a database with consistency in terminology and collection frequency that can be used for data mining. The database can help identify and deal with the lack of training data and guide connections to theory.

2.1.5 Capturing Heterogeneity

The coarse nature of the ESM grid prohibits capturing small-scale variability and heterogeneity. Therefore, models cannot predict the impact of land surface heterogeneity in land–atmosphere interactions. Increasing grid resolution can allow models to simulate more detail and new processes but comes at a computational expense. And the question still remains: is the average of the grid truly representative of the grid? One primary example where improvements in capturing heterogeneity would benefit is understanding the role of urbanization and decision-making on hydrological processes. Other examples are small-scale topographic variability (i.e., wetland hummocks and depressions) or permafrost's dynamic contributions to biogeochemical cycles. Heterogeneity is not limited to spatial features; models generally assume steady state when ecosystems are often in non-steady states. Knowledge of long-term trends and interannual variability are important and require improvements to temporal variability.

2.1.6 Building ML into ESMs

As we expand our experience with AI/ML, we need to find ways to integrate ML approaches into models. There is an opportunity to develop and substitute emulators and hybrid models for important or computationally expensive processes. This requires shifting our current approaches for model development to use a modular framework that is user friendly and utilizes computing language that can transfer across platforms. ESMs should be built to allow processes to be replaced or swapped for alternate representations. These new models should include transferable AI techniques that can be used across ESMs and/or across a range of climate/ecosystems. The methods should include explainable AI when biophysical rules and mechanistic understanding can be incorporated or pure ML when we lack theory. An example is growth and allocation of carbon where ML can provide an optimization approach using data from fluxes, tree rings, and satellite biomass. Another example could be leveraging rich aboveground datasets to inform belowground processes that are data limited. Some of the challenges are in understanding what these models might look like and what programming languages should be used and identifying which elements of the land model might be appropriate targets to benefit most from ML approaches.

2.2 State-of-the-Science

2.2.1 Model-Data Integration to Improve Predictive Skill

Quantifying uncertainty in land models and improving their predictive skill through model-data calibration require a large number of ensemble simulations, which are limited by computational resources. The number of required simulations rises exponentially as more uncertain parameters are considered, and land models often contain more than 100 such parameters. ML approaches are therefore critically important to improve the efficiency of these processes, especially as land models become more complex in terms of process representation and spatial resolution. Global sensitivity analysis (GSA) may be used to identify key parameters and processes for further investigation using a limited number of ensemble members at representative locations (e.g., Massoud 2019; Ricciuto, Sargsyan, and Thornton 2018). Model parameter calibration using methods like Markov Chain Monte Carlo (MCMC) is considerably more demanding because of the large number of simulations and serial iterations that quickly lead to infeasible integration times.

An increasingly popular use of ML in ESMs is the development of advanced emulation or surrogate modeling approaches to predict responses over parameter space and to enable constraining these predictions using observations. Surrogate models, which may take a number of different forms including polynomial functions or artificial neural networks (ANNs), may be fit using a relatively small number of model ensemble members, and then used to predict output quantities of interest (QoIs) at different points in the parameter space with a computational cost that is orders of magnitude smaller than the original model. Using surrogates therefore decreases the time required to generate the much larger number of ensembles needed for GSA and calibration. There are several examples where surrogate models have led to improved model predictions at individual sites or collections of sites (Sinha et al. 2021; Lu and Ricciuto 2019). Constructing surrogate models becomes considerably more cumbersome at regional to global scales because of the large number of QoIs contained in gridded outputs. However, a recent example using E3SM demonstrated the value of surrogate modeling to improve the wildfire model using a deep neural network approach (Zhu et al. 2022). A second example applied machine-learning surrogate modeling to reduce uncertainties in surface energy budget partitioning in the E3SM and CMIP6 models (Yuan, Zhu, Riley, et al. 2022). Land model outputs tend to have high spatial and temporal correlations, allowing application of AI-enabled dimension reduction approaches to reduce the number of required surrogate models to a more manageable number (Lu and Ricciuto 2019). Using these approaches, calibration can be performed with gridded observations (e.g., satellite measurements, ground based, fluxnet) guiding the process and can provide uncertainty quantification of the parameters; see an example for hydrological parameters in Ray et al. (2015).

In addition to quantifying model parameter uncertainty, it is also possible to use ML approaches to quantify model structural uncertainty using error-embedding approaches (Sargsyan, Huan, and Najm 2018), or by explicitly including multiple model forms in an uncertainty quantification framework (e.g., LeBauer et al. 2013; Walker et al. 2018). ML techniques such as ANNs have been used to correct biases in temperature, precipitation (Moghim and Bras 2017), and fluxes of land surface models (see Abramowitz et al. 2007). Causal network inference approaches have also been applied to assess and reduce uncertainty in global models, for example, to estimate precipitation dynamics over California throughout the 21st century (Li et al. 2022). Causality-guided machine learning approaches can also be used to improve estimates of CH_4 fluxes from site-level observations and to build extrapolatable models that can be integrated with land models (e.g., Yuan, Zhu, Fa, et al. 2022). Finally, interpretable machine learning approaches are being developed to extend current land-modeling approaches, for example, to improve estimates of wildfire and better interpret dependencies of fires on climate, fuel, and other forcings (Li et al. 2022). These methods have the potential to provide insight on the error sources and suggest improvements to the way processes are represented in models.

2.2.2 Addressing Observational Gaps

Large quantities of quality data are needed for model benchmarking, understanding relationships between variables, and developing representation of important processes in models. One example of AI applications in addressing data gaps is through the increasing development and use of low-cost sensors to expand our ability to collect data, where ML has been used to determine optimum new sensor placement and automate analysis through data mining to inform decision-making. The precision agriculture industry has a strong history of leveraging AI techniques to aid with decision-making, monitoring the health and resource needs of crops, and optimizing productivity. Some examples include using low-cost sensors with AI methods to evaluate land sustainability (Vincent et al. 2019), irrigation, and pesticides (Talaviya et al. 2020). Robotic sensors are used in the air and on the ground to gather real-time data on crop conditions in combination with satellite, weather, and soil data (Linaza et al. 2021). A second application of ML has been to bridge gaps across spatial and temporal scales. For example, FluxCom uses several ML methods (e.g., tree based, kernel methods, regression splines, and neural networks) to upscale eddy covariance for in situ to global-scale data. In addition, Radiant MLHub is a cloud-based open access resource for geospatial training data and ML techniques. Applications provided include agriculture, urban building footprints, wildfire, floods, and more. Finally, ML techniques are combined with physics to partition fluxes to extract data such as evapotranspiration (ET), sensible heat (SH), latent heat (LH), and GPP from other remotely sensed hydrological measurements (Pal and Sharma 2021). Combining an artificial neural network with physics led to the development of the first hybrid model to estimate ET globally (Zhao et al. 2019). The next step for land surface data is to expand data mining for discovery similar to the success in material sciences (see Bock et al. 2019 for review). Data mining can be carried out with Coupled Model Intercomparison Project (CMIP) model output, biogeochemistry, digital trace of humans for the urban footprint, population growth, and land use land cover change.

2.3 Experimental, Data, and Modeling Opportunities

2.3.1 Data

There are three opportunities where AI/ML can benefit data. First, AI techniques can aid with the unification and standardization of data sources collected across space, time, and disciplines. Second, AI methods can augment these datasets through gap filling and fusing multiple observations and data streams to build an improved dataset. This can include informing the deployment and optimization of sensor placement and measurements for both observational networks and experimental platforms. Finally, AI can be used to explore the data and identify correlations, causations, and relationships to inform process models.

Unification and Standardization: Observational data are critical to improving our understanding and model representation of land surface properties and dynamics, ultimately for better prediction of the Earth system. The rapid growth over the last few decades of remote sensing products, automatic ground sensor measurements, and observational and experimental facilities have led to the growing volumes and varieties of available data for land ecosystems. We need to understand what data we have and identify what data are still missing/needed. This can be achieved by standardizing, harmonizing, and integrating the diverse data and should include historical data from the published literature to generate a more complete and uniform digitized dataset that can be searched and indexed and that contains available geospatial and metadata. Classification and creation of labels should be an integral component of data processing and can be performed using AI/ML techniques. For example, ML can assist in processing labor-intensive data, such as digitizing soil surveys and smoothing soil classifications. These soil classifications were previously determined independently at regional scales with few data points extrapolated across large regions, which often resulted in sharp unnatural gradients of soil characteristics. This process can be accelerated through the use of a natural language to clarify and standardize confusing terminology used differently between scientific fields.

Data Augmentation Strategies: Data sources for land modeling are varied and include satellite and remote sensing data, automated sensors, and manual measurements on samples collected at specific points in space and time. Manual point measurements are often challenging and expensive to collect, resulting in considerable sparsity in the resultant datasets. This issue is most pronounced within the subsurface where few automated sensors can accommodate long-term field deployments. Opportunities to directly address the challenge of data sparsity within the land domain include the development of improved field-hardy sensors, use of nontraditional data collection approaches (e.g., citizen science), and improved methods and access to historical datasets. ML can be used to identify optimal deployment of sensor networks to fill in data gaps, inform decisions for experimental design, and assist in determining temporal frequency of data collection to aid in the detection of emergent properties. ML can be used to enhance data collection networks to be adaptive and respond to changes in environmental conditions (e.g., onset of drought, flood events, or hot spot/hot moment detection). Furthermore, ML can substitute space for time when data are sparse or lack sufficient time series. ML can use proxies as alternate data sources, extend time series, or classify events.

Data Exploration and Mining: ML is not limited to processing data. ML can be used for data assimilation and data–model integration, which requires quality data (not just quantity) and demonstrates the need to reduce observation bias. AI/ML should be used to improve land use/cover characterization and projection, which are large sources of uncertainty in models. Biogeophysical and socioeconomic data can be integrated with machine learning approaches to both generate historical reconstructions of and project future global land use/cover. Another opportunity should focus on training AI/ML models with remotely sensed products and site-level measurements. When combined with physical-based models, we can extrapolate site observations to the global scale to inform ESMs. ML trained on observations can be used to explore emergent behavior and outlier events, represent heterogeneity, and develop new theories. Long-term datasets collected at sites from ARM, National Ecological Observatory Network (NEON), Next-Generation Ecosystem Experiments (NGEE)-Arctic, NGEE-Tropics, Long-Term Ecological Research (LTER), and other projects can be used in generating these models. In addition, partnerships should be developed to share public/private data.

2.3.2 Representing Processes in Models

As AI/ML techniques advance, there are opportunities to improve the representation of processes in models that can reduce model bias and parameter uncertainty. Domain scientists, modelers, and computational experts should work together to identify which processes might benefit most from AI and how to replace them. Opportunities to communicate and learn via workshops, seminars, and cross-discipline funding opportunities can facilitate these connections. For example, deciding on whether to employ a hybrid or surrogate model to replace a certain process might require the following: an understanding of the drivers of the computational cost of the process, uncertainty related to parameter values and ranges required, knowledge and theory (or lack thereof) of the process, and whether there are observations to train the AI model and benchmark the model output. Priority should go to replacing processes that are most expensive, most uncertain, and least understood. Some early case studies could focus on integrating anthropogenic processes such as human decision-making related to land management, land use land cover change, crop and bioenergy productivity, and urban development. Another area of research we should explore is understanding memory and how processes across spatial and temporal scales can influence the legacy of an ecosystem. For example, understanding the extent and duration that GPP anomalies persist after a stress event (such as drought) can improve Earth system predictability. Methods focusing on explainable AI should also be prioritized, which can help grow confidence in ML-generated models and outputs. Synthetic data should be generated to test ML methods and models before they are integrated into larger ESMs. Of course, focusing on developing transferable AI methods will allow ESMs to share newly developed hybrid models, test them, and swap different representations of similar processes.

2.3.3 Advancements in AI

The above opportunities would not be possible without advancements in AI. This calls for the creation of a center to enable model-data integration, develop an easy-to-use data assimilation network, and bring together AI/ML, ModEx, ILAMB, ESSDIVE, ARM, NEON, NGEE, NOAA, USGS, LTER, and ESMs. We are still in the early stages of understanding the full possibilities of AI, and since applications for Earth system modeling can span a wide array of activities, a variety of methods and techniques employing AI/ML will be needed. Coordination of data, modeling, and AI advances will accelerate knowledge-informed learning from AI. It will also lead to discovery of when to value physics and process models, or when hybrid approaches or pure ML should be used.

2.4 Research Priorities

2.4.1 Data Preparation Strategy

AI/ML models can learn high-dimensional relationships between input variables (e.g., climate forcings) and output variables of interest (e.g., plant productivity, soil carbon flux). The high dimensionality (and often nonlinearity) of problems requires (1) a large amount of data samplings for training purposes, and (2) diverse data observed under different circumstances (normal, extreme, control, manipulated conditions) in order to achieve stable and smooth predictive relationships. One of the research challenges is a lack of datasets for those terrestrial ecosystem processes, including water, energy, carbon and nutrient cycles, and disturbances. Therefore, the research priority is to tackle the data scarcity in many innovative ways.

First, data synthesis research is needed to systematically survey available datasets that measured different aspects of the ecosystem processes ranging from in situ observations to large-scale remote sensing products, from high-frequency sampling (e.g., eddy covariance data) to long-term sparse samples (e.g., tree rings), and from more easily measurable near-surface processes to measurements of soil dynamics (e.g., microbial processes). And, more importantly, the data survey will give us a better understanding of which type of datasets are not available yet but critically important for training data-driven AI/ML land models, so that future data collection efforts could be more strategically targeted on those specific datasets.

Second, synthetic datasets can be extremely useful for pre-training AI/ML models and reduce the requirements for real measurements in order to obtain a stable predictive model. Synthetic datasets could be generated from modeling studies, such as model intercomparison projects or an individual land model. Particularly for disturbance events, the observations are sparse by nature. In that case, the AI/ML models pre-trained on synthetic datasets will become more reliant. For example, a transfer learning approach was adopted to develop the ML module of wildfire processes within E3SM (Zhu et al. 2022). The ML wildfire module was first pre-trained with E3SM process-based, model-generated outputs of burned area and then fine-tuned with the Global Fire Emissions Database (GFED) burned area product (Randerson et al. 2017) at global scale. The ML wildfire model was proven to be 90% more accurate than the process-based model and with two orders of magnitude reduction in model parameterization time. Studies that aim to generate, standardize, and archive synthetic datasets, as well as analyses and applications of the synthetic datasets to improve ML models, need to be prioritized for land modeling.

Third, data augmentation techniques are promising in terms of generating "new" datasets using real but sparse data samples. For example, spatial datasets (satellite images) could be augmented with simple mathematical operations (cropping, flipping, rotation, etc.), which create new versions of the spatial datasets and artificially increase the volume of a training dataset. Furthermore, super-resolution techniques could improve the quality of existing data and generate high-resolution representation of the observed processes.

2.4.2 Development of ML Models That Are Appropriate for Earth System Predictions

Earth system predictions have traditionally relied on process-based models (e.g., CMIP6) that simulate energy, water, carbon, and nutrient cycles among land, atmosphere, ocean, sea/land ice, and human components. However, the models' predictability is often limited by (1) incomplete knowledge of relevant processes, (2) highly uncertain model parameters, (3) coarse spatial resolutions, and (4) expensive computational cost for large simulation experiments. Recent advances in ML provide great potential for overcoming many aspects of these limitations. The intensive integration between ML and process-based modeling will be critical for the improvement of Earth system predictions. We have identified the following research areas as high priorities for the integration of ML and process-based modeling for better Earth system predictions.

Benchmarking and bias correction: First, ML models can automatically learn useful patterns and relationships from observational data, without having complete knowledge of the processes (Hengl et al. 2017). The identified patterns and derived correlative or causal (both linear or nonlinear) relationships could in turn serve as prior knowledge to further improve the theory and mathematical structures of process-based models. These patterns and data-derived relationships can also be used for model benchmarking, helping identify model deficiencies and bias and ultimately leading to future model development. Furthermore, having some prior knowledge (both quantitative or qualitative) will favor the ML model and help it derive more robust patterns or relationships (Kashinath et al. 2021).

Parameterization: Many of the parameters in land models are very uncertain, which could be due to scarcity of observational data, measurement errors, and simplification of process representations. ML can help improve the coarse-scale parameterizations in land models through optimizing parameters using observation data or detailed and high-resolution models. ML can also help make these parameters in ESMs more dynamic by allowing these parameters to be learned from appropriate sets of statistical covariates, rather than from fixed values for each plant function type (Reichstein et al. 2019). This ML-based parameterization has been used in hydrology and atmospheric modeling (Beck et al. 2016; Schirber et al. 2013; Gentine et al. 2018) and could help improve parametrizations in land surface models in ESMs.

Surrogate modeling: Land surface models are getting increasingly complicated in terms of processes and spatial and temporal resolutions. The high computational cost of running these models on the global scale has become a big technical challenge, hindering our capability for further model development, calibration, and uncertainty quantification. Emulation of the full model or surrogate modeling can greatly reduce the computation cost while maintaining the accuracy of model outputs. Surrogate modeling can help speed up sensitivity analysis and allow for more efficient model parameter calibration. Surrogate modeling therefore has gained traction in the land surface community in recent years (Sinha et al. 2021; Lu and Ricciuto 2019; Zhu et al. 2022). Further development of ML algorithms and more efficient surrogate modeling will be critical for future land model parameterization and uncertainty quantification.

Hybrid modeling: Some of the processes in land models are very computationally expensive. These processes can be replaced by a machine learning model if sufficient observations are available for model training. Hybrid modeling, combining physical modeling and ML, can be computationally efficient while maintaining the strengths of physical modeling. Some of the processes that are represented in a more empirical way can also be replaced with ML models if enough observational data exist.

2.4.3 Bridging ML Experts and Domain Scientists

Knowledge gaps exist among ML experts and Earth system domain scientists, and these have significantly hindered the development and applications of ML techniques to improve Earth system predictability. Therefore, we urgent need to bridge the gaps to: (1) educate domain scientists to learn the advantages and limitations of ML techniques; (2) facilitate ML experts to better understand the physics and principles associated with Earth system predictions; and (3) foster close collaborations between ML experts and Earth system domain scientists.

Nowadays, Earth system scientists not only need to be adept at the traditional process-based modeling tools but also ought to be trained in the advanced data science and ML techniques. Enabling domain scientists to perform ML-based research has tremendous benefits as domain scientists often have deep understanding of the major issues associated with poor Earth system predictability and where the largest prediction uncertainty is coming from. However, domain experts may have limited access to advanced ML techniques that can help break through the known issues of Earth system predictions. Oftentimes, domain scientists tend to apply simple off-the-shelf algorithms (e.g., random forest) that are far away from the state-of-the-art ML techniques; meanwhile, the ML research domain has been fast evolving, new ML tools emerge quickly, and existing ML tools also iterate and improve. In addition, there are often steep learning curves for domain scientists in terms of learning ML techniques, including using different programming languages (e.g., python versus Fortran) and different mathematical representations (e.g., physical processes versus data-based rules), as well as in the diversity of

ML modeling techniques. Training programs, workshops, and discussions are urgently needed to keep domain scientists updated on the-state-of-the-art ML research and necessary analytic tools that are relevant to Earth system prediction problems. Also, creating useful tools (e.g., software, interfaces) that lower the bar for domain scientists to learn and use ML models is highly desirable.

Collaboration among domain scientists, modelers, experimentalists, computational experts, and ML experts should be encouraged to maximize the benefits of AI/ML related to improving Earth system predictability. That should involve two-way communications and discussions that both help domain scientists to learn advanced ML techniques as well as facilitate ML experts to better understand the physics and principles associated with Earth system predictions. Creating such a collaborative and supportive research environment will need more resources (e.g., funding support, workshops) and allocation toward collaborative projects among ML experts and domain scientists to tackle the critical Earth system prediction problems.

2.5 Short-term (<5 years), 5-year, and 10-year Goals

2.5.1 Short-term Goals (<5 years)

2.5.1.1 Data Collection and Organization

Short-term goals in this area include efforts to:

- Start a data collection process to build a standardized data repository that connects disciplines and data sources that span the full breadth of space and time.
- Build the repository with consistent terminology and clear geospatial and metadata starting with currently available data and growing over time.
- Identify potential datasets (including nontraditional data sources) to build a comprehensive database for exploration. Examples include soil samples and surveys, microbial analysis and sequences, precision agriculture soil conditions and productivity, eddy covariance data (AmeriFlux, FLUXNET), and long-term experiments (LTER, ARM, NGEE, FACE, and NEON). Data should also include remote sensing data from satellites; ground observations from field measurements (long and short term); point and regional data; historical data from surveys; value-added products; proxy data such as human digital trace; and modeling output from ESMs, empirical models, or hybrid models.
- Include partnerships between private and public sectors for sharing data.
- Build trust that assures data privacy and ease of use while encouraging a wide variety of users to participate in uploading and using data.

2.5.1.2 Partnerships and Workforce Development

Short-term goals in this area include efforts to:

- Connect disciplines that normally work separately, including computational experts, domain experts, and experimentalists.

- Establish partnerships between entities, including national laboratories, universities, industry, and the private sector.
- Increase AI education through training and hiring practices to ensure that students, postdocs, and scientists experience a variety of AI methods and techniques.
- Establish partnerships between public and private industry and facilitate open communication between AI and domain scientists.

2.5.1.3 Establish Future Directions of AI

Regarding short-term goals in this area, we are in the early stages of understanding what is possible with AI/ML, so the focus should be on developing a roadmap for the future of AI. We can start to build a foundation of understanding what AI techniques work and which ones fail. On the modeling side, effort should focus on building models with more modularity to allow transferable AI for different processes. We should select a few ideal processes to use as test cases for AI, and identify which processes are better suited as process or mechanistic approaches. Fire processes make an ideal first case study, and many aspects of fire can be explored with AI including fire prediction, impacts to the hydrological cycle, and feedbacks to biogeochemistry.

2.5.2 Mid-term Goals (5 years)

2.5.2.1 Data Exploration

Regarding mid-term goals in this area, we need to establish methods to mine and explore the database efforts underway to help discover gaps in both observed data and knowledge of processes, scalability, and emergent properties. Initial data exploration will serve to find and bridge gaps in the spatial and temporal frequency of data collection, ensure that data classification is consistent across science domains, and address data scarcity. This process can lead to discoveries in experimental design, such as placement of new sensors, automation techniques for data collection, and/or changes in measurement frequency. Mining data can detect proxies to in turn detect extreme events, disturbance, or legacy or understand the role of human behavior and decision-making on the land use and land cover changes. Furthermore, we can determine the scale of observations (spatial and temporal) needed to capture large-scale or long-term processes, for example, what frequency of soil organic carbon (SOC) measurements are needed to predict carbon storage processes.

2.5.2.2 Data Preparation

Regarding mid-term goals in this area, new and emerging datasets should be integrated such that data are AI ready. Data should be identified for use including data for AI training, data for AI testing, data for model assimilation, and data for model benchmarking. We will need to generate synthetic data that can be used for testing ML models. This will be particularly important for extreme and rare events such as fire, floods, drought, and disturbance, and can also be used for land use land cover change prediction.

2.5.2.3 Establish Test Cases for AI Methods

During this stage, implementing a variety of AI techniques should be completed, which will increase confidence through experience for a range of land processes. The approaches should be used to reduce model and parameter uncertainty and identify gaps in knowledge (examples of early test cases). We also need to invest in new AI methodologies such as the capability of inverting prediction intervals to obtain confidence intervals. Conformal prediction and risk-controlling prediction sets now permit accurate UQ with virtually any black-box AI architecture; however, we cannot yet invert the resulting prediction intervals to obtain confidence intervals on relevant parameters. Solving this problem will transform data science with seismic impacts on the statistical and modeling communities.

2.5.2.4 AI Integration in Models

Furthermore, we need to begin to explore how to integrate AI into models and develop hybrid modeling approaches using transferable AI methods. These models have been developed for a variety of land processes including crop yields, evapotranspiration, soil moisture, momentum, and heat fluxes (Pal and Sharma 2021). However, scaling these ML models from site to global scales requires additional remote sensed data. Also, these methods should be expanded to other components of the land model, such as human decision and behavior, land use land cover change, allocation, and streamflow. Finally, we must investigate when domain knowledge should be integrated with AI and when pure AI is sufficient to represent processes.

2.5.3 Long-term Goals (10 years)

Overall, an infrastructure in the form of an institute should be established for long-term support. This structure will contribute to the development of a shared community effort to further the advancement of AI. An institute provides a means for communication and network building between science domains and AI experts. The institution will also work to harmonize global datasets with consistent terminology and techniques to bridge gaps in spatial and temporal observations. As experience grows, we will demonstrate successful AI/ML techniques including scaling and emulators, understand processes, and generate new hypotheses about ecosystems' behaviors and responses. Finally, these ML models will be integrated with confidence into a variety of ESMs with transferable AI.

2.6 References

Abramowitz, Gab, Andy Pitman, Hoshin Gupta, Eva Kowalczyk, and Yingping Wang. 2007. "Systematic Bias in Land Surface Models." *Journal of Hydrometeorology* 8 (5): 989–1001. https://doi.org/10.1175/JHM628.1.

Angayarkkani, K., and N. Radhakrishnan. 2010. "An Intelligent System For Effective Forest Fire Detection Using Spatial Data," February. https://doi.org/10.48550/arxiv.1002.2199.

Beck, Hylke E., Albert I. J. M. van Dijk, Ad de Roo, Diego G. Miralles, Tim R. McVicar, Jaap Schellekens, and L. Adrian Bruijnzeel. 2016. "Global-Scale Regionalization of Hydrologic Model Parameters." *Water Resources Research* 52 (5): 3599–3622. https://doi.org/https://doi.org/10.1002/2015WR018247.

Blois, Jessica L., John W. Williams, Matthew C. Fitzpatrick, Stephen T. Jackson, and Simon Ferrier. 2013. "Space Can Substitute for Time in Predicting Climate-Change Effects on Biodiversity." *Proceedings of the National Academy of Sciences* 110 (23): 9374–79. https://doi.org/10.1073/pnas.1220228110.

Bock, Frederic E., Roland C. Aydin, Christian J. Cyron, Norbert Huber, Surya R. Kalidindi, and Benjamin Klusemann. 2019. "A Review of the Application of Machine Learning and Data Mining Approaches in Continuum Materials Mechanics." *Frontiers in Materials* 6 (May): 110. https://doi.org/10.3389/fmats.2019.00110.

Frauendorf, Therese C., Richard A. MacKenzie, Ralph W. Tingley, Dana M. Infante, and Rana W. El-Sabaawi. 2020. "Using a Space-for-Time Substitution Approach to Predict the Effects of Climate Change on Nutrient Cycling in Tropical Island Stream Ecosystems." *Limnology and Oceanography* 65 (12): 3114–27. https://doi.org/10.1002/LNO.11577.

Gentine, P., M. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis. 2018. "Could Machine Learning Break the Convection Parameterization Deadlock?" *Geophysical Research Letters* 45 (11): 5742–51. https://doi.org/https://doi.org/10.1029/2018GL078202.

Hengl, Tomislav, Jorge Mendes de Jesus, Gerard B. M. Heuvelink, Maria Ruiperez Gonzalez, Milan Kilibarda, Aleksandar Blagotić, Wei Shangguan, et al. 2017. "SoilGrids250m: Global Gridded Soil Information Based on Machine Learning." Edited by Ben Bond-Lamberty. *PLOS ONE* 12 (2): e0169748. https://doi.org/10.1371/journal.pone.0169748.

Kashinath, K., M. Mustafa, A. Albert, J.-L. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, et al. 2021. "Physics-Informed Machine Learning: Case Studies for Weather and Climate Modelling." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379 (2194): 20200093. https://doi.org/10.1098/rsta.2020.0093.

LeBauer, David S., Dan Wang, Katherine T. Richter, Carl C. Davidson, and Michael C. Dietze. 2013. "Facilitating Feedbacks between Field Measurements and Ecosystem Models." *Ecological Monographs* 83 (2): 133–54. https://doi.org/https://doi.org/10.1890/12-0137.1.

Li, F., Q. Zhu, W. Riley, L. Zhao, L. Xu, K. Yuan, M. Chen, et al. 2022. "AttentionFire_v1.0: Interpretable Machine Learning Fire Model for Burned Area Predictions over Tropics." *Geoscientific Model Development Discussions* 2022: 1–28. https://doi.org/10.5194/gmd-2022-195.

Li, Fa, Qing Zhu, William J. Riley, Kunxiaojia Yuan, Huayi Wu, and Zhipeng Gui. 2022. "Wetter California Projected by CMIP6 Models With Observational Constraints Under a High GHG Emission Scenario." *Earth's Future* 10 (4): e2022EF002694. https://doi.org/https://doi.org/10.1029/2022EF002694.

Linaza, Maria Teresa, Jorge Posada, Jürgen Bund, Peter Eisert, Marco Quartulli, Jürgen Döllner, Alain Pagani, et al. 2021. "Data-Driven Artificial Intelligence Applications for Sustainable Precision Agriculture." *Agronomy* 11 (6): 1227. https://doi.org/10.3390/agronomy11061227.

Lu, D., and D. Ricciuto. 2019. "Efficient Surrogate Modeling Methods for Large-Scale Earth System Models on Machine-Learning Techniques." *Geoscientific Model Development* 12 (5): 1791–1807. https://doi.org/10.5194/gmd-12-1791-2019.

Massoud, Elias C. 2019. "Emulation of Environmental Models Using Polynomial Chaos Expansion." *Environmental Modelling & Software* 111 (January): 421–31. https://doi.org/10.1016/J.ENVSOFT.2018.10.008.

Messer, Hagit, Artem Zinevich, and Pinhas Alpert. 2006. "Environmental Monitoring by Wireless Communication Networks." *Science* 312 (5774): 713. https://doi.org/10.1126/science.1120034.

Moghim, Sanaz, and Rafael L. Bras. 2017. "Bias Correction of Climate Modeled Temperature and Precipitation Using Artificial Neural Networks." *Journal of Hydrometeorology* 18 (7): 1867–84. https://doi.org/10.1175/JHM-D-16-0247.1.

Pal, Sujan, and Prateek Sharma. 2021. "A Review of Machine Learning Applications in Land Surface Modeling." *Earth* 2 (1): 174–90. https://doi.org/10.3390/earth2010011.

Randerson, J. T., G. R. van der Werf, L. Giglio, G. J. Collatz, and P. S. Kasibhatla. 2017. "Global Fire Emissions Database, Version 4.1 (GFEDv4)." ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1293.

Ray, J., Z. Hou, M. Huang, K. Sargsyan, and L. Swiler. 2015. "Bayesian Calibration of the Community Land Model Using Surrogates." *SIAM/ASA Journal on Uncertainty Quantification* 3 (1): 199–233. https://doi.org/10.1137/140957998.

Reichstein, Markus, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and Prabhat. 2019. "Deep Learning and Process Understanding for Data-Driven Earth System Science." *Nature* 566 (7743): 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Ricciuto, Daniel, Khachik Sargsyan, and Peter Thornton. 2018. "The Impact of Parametric Uncertainties on Biogeochemistry in the E3SM Land Model." *Journal of Advances in Modeling Earth Systems* 10 (2): 297–319. https://doi.org/https://doi.org/10.1002/2017MS000962.

Sargsyan, Khachik, Xun Huan, and Habib N. Najm. 2018. "Embedded Model Error Representation for Bayesian Model Calibration," January. https://doi.org/10.48550/arxiv.1801.06768.

Schirber, Sebastian, Daniel Klocke, Robert Pincus, Johannes Quaas, and Jeffrey L. Anderson. 2013. "Parameter Estimation Using Data Assimilation in an Atmospheric General Circulation Model: From a Perfect toward the Real World." *Journal of Advances in Modeling Earth Systems* 5 (1): 58–70. https://doi.org/https://doi.org/10.1029/2012MS000167.

Sinha, E., K. Calvin, B. Bond-Lamberty, B. Drewniak, D. Ricciuto, K. Sargsyan, Y. Cheng, C. Bernacchi, and C. Moore. 2021. "Modeling Perennial Bioenergy Crops in the E3SM Land Model." *Geoscientific Model Development Discussions* 2021: 1–31. https://doi.org/10.5194/gmd-2021-244.

Talaviya, Tanha, Dhara Shah, Nivedita Patel, Hiteshri Yagnik, and Manan Shah. 2020. "Implementation of Artificial Intelligence in Agriculture for Optimisation of Irrigation and Application of Pesticides and Herbicides." *Artificial Intelligence in Agriculture* 4 (January): 58–73. https://doi.org/10.1016/J.AIIA.2020.04.002.

Vincent, Durai Raj, N. Deepa, Dhivya Elavarasan, Kathiravan Srinivasan, Sajjad Hussain Chauhdary, and Celestine Iwendi. 2019. "Sensors Driven AI-Based Agriculture Recommendation Model for Assessing Land Suitability." *Sensors* 19 (17): 3667. https://doi.org/10.3390/s19173667.

Walker, Anthony P., Ming Ye, Dan Lu, Martin G. de Kauwe, Lianhong Gu, Belinda E. Medlyn, Alistair Rogers, and Shawn P. Serbin. 2018. "The Multi-Assumption Architecture and Testbed (MAAT v1.0): R Code for Generating Ensembles with Dynamic Model Structure and Analysis of Epistemic Uncertainty from Multiple Sources." *Geoscientific Model Development* 11 (8): 3159–85. https://doi.org/10.5194/gmd-11-3159-2018.

Wogan, Guinevere O. U., and Ian J. Wang. 2018. "The Value of Space-for-Time Substitution for Studying Fine-Scale Microevolutionary Processes." *Ecography* 41 (9): 1456–68. https://doi.org/https://doi.org/10.1111/ecog.03235.

Yuan, Kunxiaojia, Qing Zhu, Li Fa, William Riley, Margaret Torn, Housen Chu, Gavin McNicol, et al. 2022. "Causality Guided Machine Learning Model on Wetland Ch4 Emissions Across Global Wetlands." *SSRN Electronic Journal*, January. https://doi.org/10.2139/ssrn.4034619.

Yuan, Kunxiaojia, Qing Zhu, William J. Riley, Fa Li, and Huayi Wu. 2022. "Understanding and Reducing the Uncertainties of Land Surface Energy Flux Partitioning within CMIP6 Land Models." *Agricultural and Forest Meteorology* 319 (May): 108920. https://doi.org/10.1016/J.AGRFORMET.2022.108920.

Zhao, Wen Li, Pierre Gentine, Markus Reichstein, Yao Zhang, Sha Zhou, Yeqiang Wen, Changjie Lin, Xi Li, and Guo Yu Qiu. 2019. "Physics-Constrained Machine Learning of Evapotranspiration." *Geophysical Research Letters* 46 (24): 14496–507. https://doi.org/https://doi.org/10.1029/2019GL085291.

Zhu, Qing, Fa Li, William J. Riley, Li Xu, Lei Zhao, Kunxiaojia Yuan, Huayi Wu, Jianya Gong, and James Randerson. 2022. "Building a Machine Learning Surrogate Model for Wildfire Activities within a Global Earth System Model." *Geoscientific Model Development* 15 (5): 1899–1911. https://doi.org/10.5194/gmd-15-1899-2022.

3 Hydrology

Authors: Charuleka Varadharajan (LBNL), Scott L. Painter (ORNL), Jitendra Kumar (ORNL), Chaopeng Shen (Penn State Univ.), Dan Lu (ORNL), David Moulton (LANL), Xingyuan Chen (PNNL), Mohammed Ombadi (LBNL), Dapeng Feng (Penn State Univ.), Soumendra Bhanja (ORNL), Helen Weierbach (LBNL), Wenping Tsai (Penn State Univ.), Jared Willard (Univ. of Minnesota/LBNL), Wei Zhi (Penn State Univ.), Alexander Sun (UT Austin)

3.1 Introduction

As an Earth science discipline, hydrology is concerned with "understanding the movement of water at all scales and environments and its interaction with climate and life on Earth" (National Research Council 2012). The DOE's interests in hydrology revolve around the Earth system and integrated water cycle considering topics such as coupling between the water and carbon cycles, decadal projections accounting for environmental change, extreme water cycle events, and combining models and observations to advance fundamental process understanding (U.S. Department of Energy Office of Science 2018). In understanding the potential role of ML and AI in hydrologic science, it is important to make the distinction between hydrology as an operations-focused engineering endeavor versus a fundamental yet application-inspired Earth science discipline. From an operations perspective, hydrology has traditionally focused on applications like flood resilience and protection of water supply for human and agricultural uses and energy production. While significant effort has been made on incorporating AI and ML (AI/ML) in an operational context, less progress has been made on its use to advance hydrologic science as a component of the broader Earth system, the focus of this workshop report.

3.2 Grand Challenges

Building on earlier synthesis reports (National Research Council 2012; Blöschl et al. 2019), we identify several persistent grand challenges in hydrology that may benefit from AI/ML. Several of these challenges also present exciting research opportunities and priorities for DOE, which are described in subsequent sections of this chapter. Figure 3-1 depicts the opportunities using ML as they relate to the grand challenges.

3.2.1 Exogenous Change

Regardless of whether the perspective is local, regional, or global, the hydrologic system is strongly influenced by exogenous variables such as climate forcing, land use and land cover, and water management practices, which are themselves subject to significant change. Understanding exogenous change and its effect on the hydrologic system is a central challenge in hydrology as a terrestrial science. Among the many fundamental questions related to exogenous change are the following: Where is the regional hydrologic cycle expected to intensify in a warming climate and what are the consequences of that intensification? Have abrupt transitions (tipping points) occurred or will they occur? What are the consequences of land-cover and land-use change on groundwater recharge and other water fluxes?

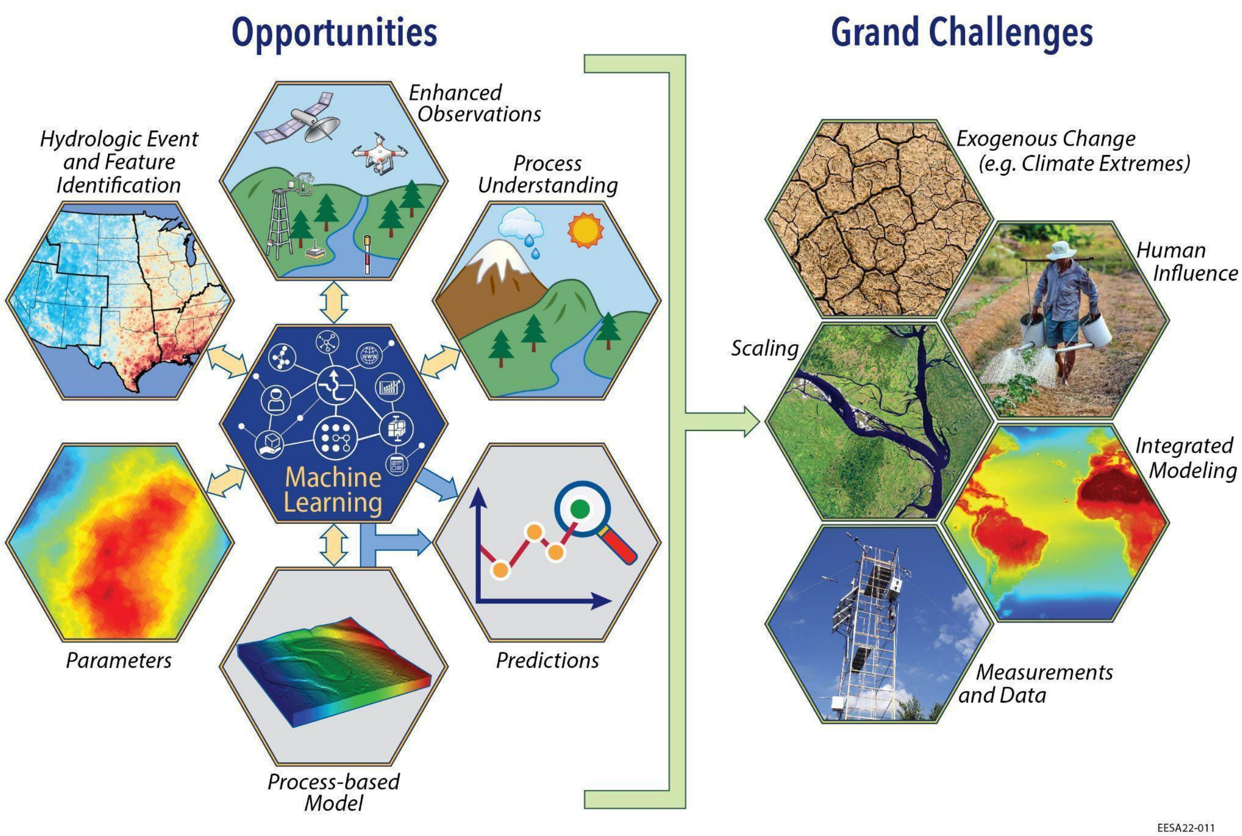


Figure 3-1. Machine learning (ML), statistics, and information theory have been used in various hydrological applications for decades, and recent advances have started to blend ML with process models. The use of ML in data collection, curation, and models can help address grand challenges in hydrology such as scaling, integrated modeling that includes the effects of human activities, and identifying the effects of extreme events on the hydrologic system (Source: Lawrence Berkeley National Laboratory. Figure adapted with permission from Xu and Liang 2021 © 2021 Wiley Periodicals LLC).

ML-based tools have significant potential as components in broader strategies by addressing such questions. As described further in section 3.3, feature extraction and detection techniques can be used to identify the exogenous change and hydrologic responses. Causal inference methods can be used to identify whether these external factors drive changes in the hydrologic system (Runge et al. 2019). Finally, regression analysis using data-driven or hybrid models can be used to understand and simulate the magnitude and duration of impacts resulting from the exogenous change on the hydrologic system. Current examples of the use of ML to address these challenges include change detection in land-cover/land-use patterns (Shi et al. 2020), early warning signs for tipping points (Bury et al. 2021), and attribution of climate impacts (Callaghan et al. 2021).

3.2.2 Spatial-temporal Variability and Scaling

Heterogeneity is ubiquitous in hydrologic systems across all scales, but typically not well characterized at the relevant scales of interest. Additionally, a long-standing fundamental challenge exist in understanding the relationship of heterogeneity in properties or hydrologic state variables and spatial variability in evaporation and surface and subsurface fluxes of water and waterborne material likes carbon, nutrients, sediments, and contaminants. Underlying this broad challenge are questions about the nature and distribution of flowpaths, the relationship between fine-scale processes and catchment-scale behavior, and the existence of catchment-scale theories. In addition to being of fundamental interest, these questions have practical implications about, for example, prediction in unmonitored basins (PUBS). Regional- to continental-scale predictions of river flow and water quality (e.g., temperature, salinity, nutrients, and contaminants) is one example of how spatial variability has hindered scaling. The treatment of flow and biogeochemistry in current models is limited to smaller scales, and bridging to larger regional scales is computationally challenging given the diversity of processes involved and complexity of the reactions to be represented at desired resolutions (Steefel 2019; Steefel et al. 2021). The hydrologic system is also subject to significant temporal variability, which can interact with spatial variability to determine emergent behavior at catchment scales. A particularly challenging aspect of temporal variability and its effect is hydroclimatic extreme events (floods, droughts, heat waves). Fundamental and societally relevant questions related to extreme events include detection, attribution, and characterization; how their distributions are changing due to global warming; how land-use and land-cover modulate their impacts; and how multiple extreme events interact (i.e., compounding events) to amplify consequences or trigger regime shifts that alter system behavior.

Machine learning offers significant potential to advance our understanding of transferability and spatial-temporal scaling in the terrestrial hydrologic system. For example, the long-standing challenge of runoff PUBS (Sivapalan 2003) has typically been approached using statistical methods for extrapolating observations from well-monitored catchments, including similarity-based, regression-based, and signature-based regionalization techniques (Guo et al. 2021). New ML approaches including deep learning models (e.g., long short-term memory [LSTM] and its variants) and transfer learning have the potential to address the PUBs challenge, not just for runoff but for several other variables with long-term datasets. Supervised and unsupervised classification methods have the capacity to uncover previously unknown patterns of catchment similarity, particularly when used with new datasets that include human activities. The inclusion of ML into Earth system models can address another hydrological grand challenge, that is, the development of continental-scale models for hydrobiogeochemical predictions that work across a diversity of catchment types, at decision-relevant spatial and temporal resolutions. AI/ML for detection and prediction of extreme events is challenged by lack of training data but is a promising research topic that can benefit from new ML techniques (chapter 8, Climate Variability and Extremes).

3.2.3 Integrated Modeling

Hydrologists have traditionally confronted the enormous process and spatial complexity of hydrologic systems by splitting into component subsystems (e.g., groundwater, surface water, ecosystems) that can be more easily understood and modeled. Such a reductionist approach is clearly necessary for scientific progress, but leaves significant scientific uncertainties associated with the resulting spatial and discipline interfaces. Developing holistic syntheses generally and integrated models specifically that span those interfaces has been a long-term challenge. Modern software and high-performance computing have resulted in powerful physics-based integrated models (Maxwell, Condon, and Kollet 2015; Painter et al. 2016) that couple surface water, shallow groundwater, and land surface processes. However, using those integrated models at societally relevant scales while still honoring process understanding developed at much finer scales remains a difficult challenge. Calibration and uncertainty quantification of these models is computationally expensive, potentially requiring hundreds of thousands of forward model simulations.

ML has considerable potential to help with the challenge of integrated modeling by improving process models, for example, by providing accurate surrogate models for component subsystems, data-driven representations for subsystems with immature or uncertain process representations, or more efficient approaches for inverse modeling and uncertainty quantification. In particular, ML-based surrogate modeling approaches can speed up model run times and reduce computational costs (Willard et al. 2020). Conversely, process models can improve ML models by incorporating finer-scale process knowledge into data-driven models implemented at coarser scales, providing training data particularly in situations with sparse data, enabling more robust predictions under non-stationary conditions, and improving the explainability of black box models. The dichotomy of *"process-first"* and *"AI-first"* approaches for integrating mechanistic and ML codes was a persistent theme in the workshop, and opportunities for both paths are described below.

3.2.4 Measurements and Data

Hydrology is a data-limited science. Long-term monitoring networks, integrated field observatories, and satellites that make use of advances in remote sensing, unmanned aerial vehicles (UAVs), wireless networked sensors, hydrogeophysical techniques, and other technologies are producing large amounts of data. However, most of those data are either sparse, or at the wrong scale to be most useful, or at best indirectly related to quantities of interest, hydrologic state variables, and fluxes. Water fluxes and constituents in particular are difficult to measure over large areas and difficult to measure in the subsurface at any scale. Sensor technologies to collect high-resolution data are limited to a few variables, and their cost as well as the lack of automated quality assurance methods limits the extent to which these can be deployed (Kruse 2018). The challenges of data sparsity are exacerbated by the fact that many observations are currently spread across multiple databases that lack interoperable standards. A bottleneck for applying ML to real-world datasets is being able to discover and integrate data located across numerous databases, which are stored in different formats and units.

AI/ML has the potential to enable autonomous technologies that collect measurements over much greater spatial areas than is currently possible with higher spatial and temporal resolution. Advances in edge computing, 5G networks, and ML-based classification and regression can be used to optimally design sensor networks, drive data collection when and where it is needed, quality-check data in near real-time, and create imputed data products. Natural language processing (NLP) can enable advanced search and harmonization to create curated, AI-ready datasets. The long-standing approach to address the problem of unobserved variables is to combine proxy data – observations that are indirectly related to quantities of interest – with sparse, high-quality data. ML is well suited to that problem of data fusion, particularly for multimodal data (Gao et al. 2020), and offers the potential to make better use of coarse, sparse, and indirectly related information.

3.2.5 Interfaces with Human Systems

In the age of the Anthropocene, human actions are responsible for major changes in the water cycle driving regime shifts across hydrologic systems (Abbott et al. 2019; Gleeson et al. 2020). Traditionally, hydrologists have represented the effect of human activities through scenario-based analyses that represent human activities as external forcings to the hydrologic system. This approach is problematic for long-term projections because it neglects bi-directional interactions between the human and hydrologic systems. Developing a predictive understanding of the "human-water system in a holistic sense" (Blair and Buytaert 2016), accounting for bi-directional interactions is the subject of socio-hydrology (Sivapalan, Savenije, and Blöschl 2012; di Baldassarre et al. 2013; Elshafei et al. 2014) and an important challenge for hydrologic science.

Data-driven approaches are central to this emerging field (Mount et al. 2016), for example, by using surrogate approaches to generate projections for different scenarios, employing reinforcement learning to drive decisions, and using data from observed variables to implicitly incorporate the outcomes of human actions into hydrological models. ML applications in this space are also described in chapter 9, Human Systems and Dynamics.

3.3 State-of-the-Science

There has been a long history of using ML in hydrologic sciences. Several recent reviews comprehensively capture the state-of-the-art scientific ML, deep learning, and knowledge-guided

ML that have been used in numerous hydrologic studies (Shen 2018; Sun and Scanlon 2019; Shen and Lawson 2021; Xu and Liang 2021; Varadharajan et al. 2022). These studies have used ML for a variety of hydrologic applications ranging from precipitation estimation and forecasting (e.g., estimation of rainfall from radars, creation of high-resolution precipitation products), rainfall-runoff modeling (e.g., streamflow predictions), quantification of ecohydrological fluxes (e.g., evapotranspiration), estimation of subsurface storage and flow (e.g., groundwater, soil moisture), water quality predictions (e.g., of stream temperature and chemical properties), determining impacts of hydrological events on human systems (e.g., urban flooding), and studies of complex interrelated hydrologic systems (e.g., surface-groundwater interactions and terrestrial aquatic interfaces). These studies have used a variety of ML methods including (1) classical models such as support vector machines, ensemble-tree approaches (random forests, XGBoost), artificial neural networks (ANNs), and hierarchical clustering; (2) deep learning models such as convolutional neural networks (CNNs), several variants of long- and short-term memory networks (LSTMs), multilayer perceptrons (MLPs), and generative adversarial networks (GANs); and (3) physics-constrained ML that couple process models (e.g., for stream flow and temperature) with machine learning in various ways. A brief summary of some hydrological ML applications is presented in Table 3-1. This is not meant to be a comprehensive list and covers only a sampling of topics to illustrate recent advances in ML for hydrology. Additional studies are described in the context of research opportunities in section 3.4.

Table 3-1. Synthesis of recent advances in AI for hydrology.

Topic	Example AI Application(s) and Related Work(s)
Regional to continental modeling	• Streamflow and temperature forecasts using LSTM and its variants (Kratzert et al. 2018; Feng, Fang, and Shen 2020; Rahmani et al. 2020), graph neural networks (Zhao et al. 2020; Jia et al. 2021; Sun and Tang 2020)
PUBS	• Streamflow and temperature predictions using LSTM and its variants (Kratzert et al. 2019; Rahmani et al. 2021; Li et al. 2022), ensemble classical ML approaches (Weierbach et al. 2022)
Data Acquisition	• ML-assisted UAV for soil moisture (Araya et al. 2021) • CNNs for river stage estimates from camera imagery (vanden Boomen, Yu, and Liao 2021)

Table 3-1. (Cont.)

Topic	Example AI Application(s) and Related Work(s)
Variable/ Parameter Estimation	● CNNs, conditional GANS for precipitation estimation from remote sensing (Hayatbini et al. 2019; Sadeghi et al. 2019), Resnet CNN for precipitation estimation from radar measurements (Chen and Chandrasekar 2021) ● Physics-informed neural networks for subsurface parameter estimation (Tartakovsky et al. 2020) ● Global 30-min evapotranspiration estimates using random forest (Bodesheim et al. 2018) ● Estimation of subsurface properties from streamflow using DNNs and integrated hydrology models (Cromwell et al. 2021)
Downscaling and Imputation	● Attention-based CNNs for precipitation downscaling (Sun and Tang 2020), random forest for imputation of missing precipitation data (Mital et al. 2020) ● LSTM for Soil Moisture Active Passive (SMAP) soil moisture estimates (Fang et al. 2017)
Hybrid/Surrogate Modeling	● Short-term forecasts of water temperature pre-training hybrid-LSTM with a process model (Read et al. 2019; Jia et al. 2021; Zwart et al. 2021), model outputs as inputs to LSTM (Konapala et al. 2020) ● Differentiable parameter learning coupling a neural-network-based parameterization scheme to a process-based model (Tsai et al. 2020) ● Streamflow predictions using hybrid unsupervised ML non-negative matrix factorization with k-means cluster (Fleming, Vesselinov, and Goodbody 2021) ● Neural network model symmetry (Daw et al. 2020)
Knowledge Discovery/Feature Detection	● Surface-groundwater fluxes using CNN and other methods (Moghaddam et al. 2022) ● River network classification using CNN with different optimizers (Donadio et al. 2021) ● Landscape attributes affecting streamflow using LSTM with watershed attributes (Kratzert et al. 2019) ● Numerous works using random forest/XGBoost feature importance to identify primary factors influencing the variable of interest
Hydrological Extremes Detection and Impacts	● Flood detection using ANN, support vector machines (SVMs), MLP, random forest, fuzzy inference systems, wavelet NN, and ensembles (Mosavi, Ozturk, and Chau 2018); conditional density networks (Cannon 2012); LSTM (Frame et al. 2021); and modified training datasets or loss functions (Fleming et al. 2015; Xie et al. 2021) ● Harmful algal bloom detection (HABNet) with CNN+ (LSTM/RF/SVR) and remote sensing inputs (P. R. Hill et al. 2020)
Information theory and causal analysis	● Information entropy for identifying hydrologic similarity (Loritz et al. 2018) ● Ecohydrologic responses to precipitation using temporal information partitioning networks (Goodwell et al. 2018) ● Causal inference PC algorithm for estimating drivers of evapotranspiration (Ombadi et al. 2020)
Uncertainty Quantification	● Bayesian LSTM for UQ (Dan Lu, Konapala, et al. 2021) ● Permutation feature importance on groundwater-level forecasts for MLP with surrogate model-based hyperparameter optimization (Sahu et al. 2020)

Table 3-1. (Cont.)

Topic	Example AI Application(s) and Related Work(s)
Data Products	● Meteorological products: Daymet (Thornton et al. 2021), PRISM (PRISM Climate Group 2019), NLDAS (Xia et al. 2012), reanalysis (https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5), PERSIANN (https://chrsdata.eng.uci.edu/) ● Rivers and watershed products: CAMELS (Addor et al. 2017), GAGES-II (Falcone et al. 2010), StreamCat (R. A. Hill et al. 2016) ● Evapotranspiration: FLUXNET (Pastorello et al. 2020)
Hydrology Workflows and Tools	● Data Integration: CUAHSI HIS (Horsburgh et al. 2015), BASIN-3D data broker (Varadharajan et al. 2022) ● Model-data integration: Watershed workflow for ATS model (https://github.com/ecoon/watershed-workflow/)

3.4 Experimental, Data, and Modeling Opportunities

We provide selected examples of opportunities for transformative scientific advances in different fields of hydrology using AI/ML that were identified through this workshop. This is not intended to be an exhaustive list. Instead, these examples illustrate the types of opportunities identified for a few major themes: experiments and data collection; data curation, fusion and imputation; data-driven and hybrid modeling; knowledge discovery and transferability; and model-data co-development.

3.4.1 Experiments and Data Collection

Data-driven models and particularly deep learning models need large amounts of high-resolution data. New and continued data acquisition of important variables is essential to advance the use of AI/ML for hydrology. There are numerous opportunities to use ML to improve data collection by either direct or proxy measurements. These include approaches to collect new measurements of variables of interest (e.g., nutrients) across heterogeneous landscapes at much greater scales and resolutions through the use of automated ML-assisted technologies such as next-generation sensor networks, camera and video imagery processed using mature computer vision methods, autonomous UAV (Song et al. 2017; Araya et al. 2021), mobile aquatic drones (Matos and Postolache 2016), and robotics (e.g., Figure 3-2). Classification methods that combine different data layers can be used to determine optimal measurement strategies and sampling network design (Wainwright et al. 2022). Other ML approaches (e.g., active or reinforcement learning) combined with edge computing could be used to guide autonomous instrumentation in near real-time to collect optimal observations that capture processes of interest, such as during an anomalous event or across the spatial gradients encountered at interfaces or ecological control points (i.e., hot spots). Surrogate ML models serving as "soft sensors or electronic noses" can also be used to make predictions of variables that are difficult to measure directly using proxy measurements of easily observable variables (Paepae, Bokoro, and Kyamakya 2021). Finally, ML approaches can be used to automate quality assurance and quality control (QA/QC) of data streams in near real time moving beyond current semi-automated statistical and rule-based methods. Studies using statistical, ML-based anomaly and change-point detection algorithms for hydrological measurements are mostly bespoke demonstration applications for small-scale datasets, and there is potential to scale existing methods to larger datasets from sensor networks or explore the use of newer algorithms such as hierarchical temporal memory networks (Wu, Zeng, and Yan 2018; Cho et al. 2020). Quality-checking frameworks that incorporate outlier detection and imputation algorithms are being applied for other fields such as electric grid monitoring and have potential for use in Earth sciences as well (Stewart et al. 2017).

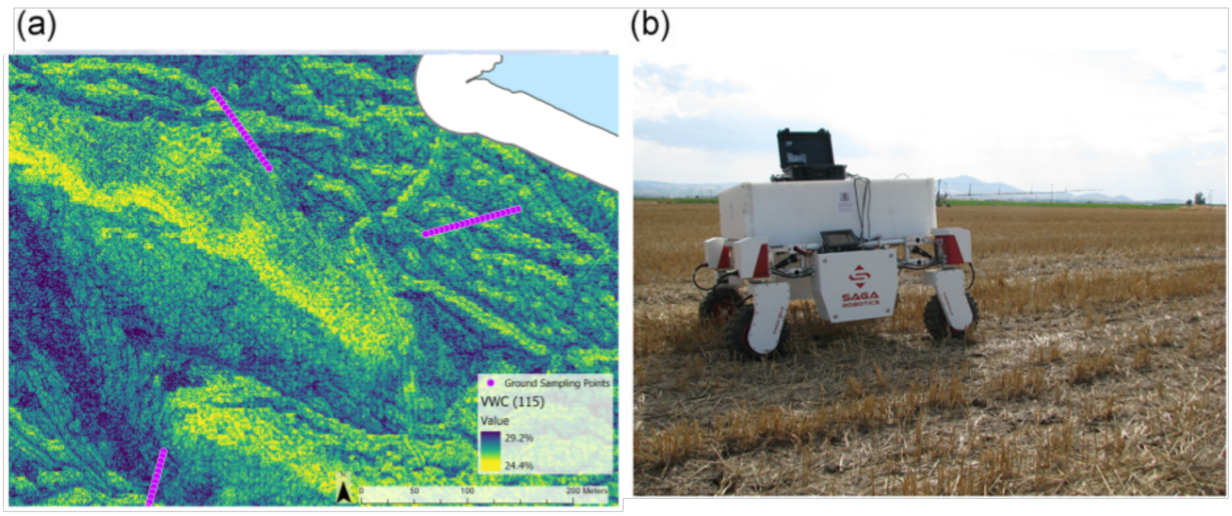


Figure 3-2. New ML-assisted technologies can generate high-resolution data at greater spatiotemporal scales. Examples shown include high spatial-resolution soil moisture measurements collected using (a) ML-driven unoccupied aircraft (Araya et al. 2021), and (b) a robot equipped with a cosmic-ray sensor (Source: Reproduced from Araya et al. 2021 and Pulido Fentanes et al. 2020 under Creative Commons CC BY 4.0).

3.4.2 Curation, Synthesis, and Imputation to Improve Data Products

Research in AI/ML was revolutionized by the introduction of large benchmark databases such as ImageNet (Deng et al. 2009) and MNIST (Deng 2012). Classic examples of curated data products that have accelerated ML in hydrology include CAMELS (Addor et al. 2017) and FLUXNET (Pastorello et al. 2020), and several new products are emerging such as ResOpsUs (Steyeart et al. 2022), TWSA (Adusumilli et al. 2019). Efforts to make data publicly available in open repositories with sufficient metadata to enable reuse are important and relevant for ML. Additionally, benchmark hydrological and related datasets – that is, curated, high-quality, ready-to-use data products – for ML model inputs and evaluation need to be developed and maintained to include the latest available data (Crystal-Ornelas et al. 2021). Federated databases or data discovery and synthesis tools (e.g., Varadharajan et al. 2022) can make it easier to identify relevant databases, to enable more uniform access and intercomparisons of data across providers, and to extract information from multiple data sources that have traditionally been analyzed separately. Infrastructure to integrate datasets across data sources located in heterogeneous environments (cloud, high-performance file systems, databases etc.) and transformed into a uniform format would accelerate data model integration. For example, natural language processing methods could be applied to discover datasets and enable semantic harmonization. In addition, ML holds great potential to improve model inputs or evaluation datasets. For example, several gridded climate forcing datasets (Table 3-1) have proven to be valuable for spinning up ML and physical model simulations. However, these have different variables, spatial resolutions, and uncertainties, making it difficult to merge or compare across products. Furthermore, the meteorological products are not sufficiently accurate for high-altitude regions with complex terrain (Lundquist et al. 2019; Feldman et al. 2021). Machine learning methods can be used for imputation of sparse observations or for downscaling gridded products to create new higher-resolution datasets (Sun and Tang 2020; Risser, Rhoades, and Mahesh 2021; Mittal et al. 2020). Finally, ML can be used for data synthesis to extract the maximum amount of information from conventional or nontraditional data sources (e.g., cell phone towers; Overeem, Leijnse, and Uijlenhoet 2013). Information retrieval from remote sensing datasets is an example where ML has added significant value (Table 3-1). Additionally, deep learning applied to data fusion of multimodal (e.g., geospatial, remote sensing, and time-series) datasets (Gao et al. 2020)—such as high spatial-resolution satellite imaging available from cubesats such as Planet Labs (www.planet.com) with high-temporal time-series from long-term, ground-based sensor networks—can generate products with unprecedented spatiotemporal resolution.

3.4.3 Data-driven Models for Prediction

Machine learning has great potential to improve understanding and representation of hydrologic subsystems that lack mature and reliable process descriptions. Examples where data-driven models can improve or even provide an alternative to process-based representation include preferential flow in unsaturated soils, evapotranspiration, and human influences on the water cycle. Many of the ML methods that have been used in operations-focused applications could be applied directly to this task. For example, recent implementations of data-driven models demonstrate their potential for predictions of hydrological variables at basin to continental scales including in dammed and unmonitored basins (Table 3-1). Attention-based transformers (Vaswani et al. 2017) have superseded LSTM architectures for NLP but have not been broadly applied for time-series modeling (Pouchard et al. 2021). While such models are useful for operational purposes, their use in Earth system modeling is yet unexplored and requires additional considerations. In particular, it is important that the AI/ML methods be interpretable, able to accommodate constraints derived from fundamental process understanding, and robust to non-stationarity in exogenous variables. Preferential flow is a representative example of how ML could be used to develop data-driven representations that could be used alone or as components in integrated hydrological models (Sprenger et al. 2021). The preferential flow of water in localized pathways that bypasses significant fractions of the soil porous medium is a widely observed phenomenon in soils that significantly affects response to precipitation events. Preferential flow can be the result of macropores (cracks, biopores) or unstable "fingering" in coarse soil. Despite its known importance, capturing the effect of macropore preferential flow in catchment-scale models has proved to be challenging. In a recent review, Jarvis, Koestel, and Larsbo (2016) note that empirical process understanding developed from high temporal and spatial monitoring at catchment scales has advanced significantly and is now outpacing the ability to represent the process in models. These new datasets and sources of data are ripe for ML applications. Indeed, some initial steps have been made in that direction. For example, Koestel and Jorda (2014) applied random forest regression to a meta-dataset to identify soil properties and environmental conditions that trigger preferential flow. It is envisioned to go beyond that analysis of controlling variables to build ML-based, data-driven representations of preferential flow that can be used as components in catchment-scale integrated hydrological models. Success would address one of the most persistent challenges in catchment hydrology – *how to tractably represent the effects of preferential flow.*

3.4.4 Integration with Process-based Models

Several opportunities exist to more deeply integrate machine learning into hydrological models such as the use of surrogate models for accelerating simulations, subgrid parameterizations for use in larger models, use of hybrid ML-process models to improve robustness and accuracy, and integrated model hierarchies (Collins 2021; Painter, Coon, and Lu 2021; Steefel et al. 2021) (Figure 3-3). Advances in software design and high-performance computing have renewed interest in models that are "as physically based as possible" (Bierkens 2015), following the Freeze and Harlan (1969) blueprint for integrated models. That class of fully resolved physics-oriented models, which includes ATS (Painter et al. 2016; Coon et al. 2019; Coon et al. 2020) and PARFLOW (Maxwell, Condon, and Kollet 2015), combine 3D representations of variably saturated subsurface flow, 2D representations of overland flow, and land surface processes. Such *virtual watersheds* are "widely considered to be the gold standard in hydrologic modeling" (Fleming and Gupta 2020). Applying this class of models at basin or continental scales remains a significant computational challenge, and applications at river basin to continental scales are relatively rare. Similar computational demands will be encountered as the global land surface modeling community increases spatial resolution, which will require better representation of hydrologic processes like lateral subsurface flow (Bierkens 2015). Machine learning has significant potential to significantly accelerate those computationally demanding models.

An appealing strategy for accelerating physics-oriented watershed-to-basin models is to not repeat calculations for subdomains with similar climate forcings and physiographic properties. This strategy based on hydrologic similarity is well suited for ML (see next section), which in the short term could be implemented by first building surrogate models from simulations on smaller domains, then combining them in larger-scale simulations. Longer term, it is envisioned that this strategy could be implemented dynamically, using AI/ML to replace subdomains of a fully resolved simulation with surrogate models on-the-fly, switching back to fully resolved simulations if needed (Painter, Coon, and Lu 2021). That approach would require significant research and development at the nexus of ML, algorithm design, process-oriented modeling, and software design. The resulting capability would be truly transformative, making it possible to routinely apply fully resolved, physics-oriented models at river basin scales and enabling, for the first time, calibration, sensitivity, and scenario analyses and UQ for virtual watersheds.

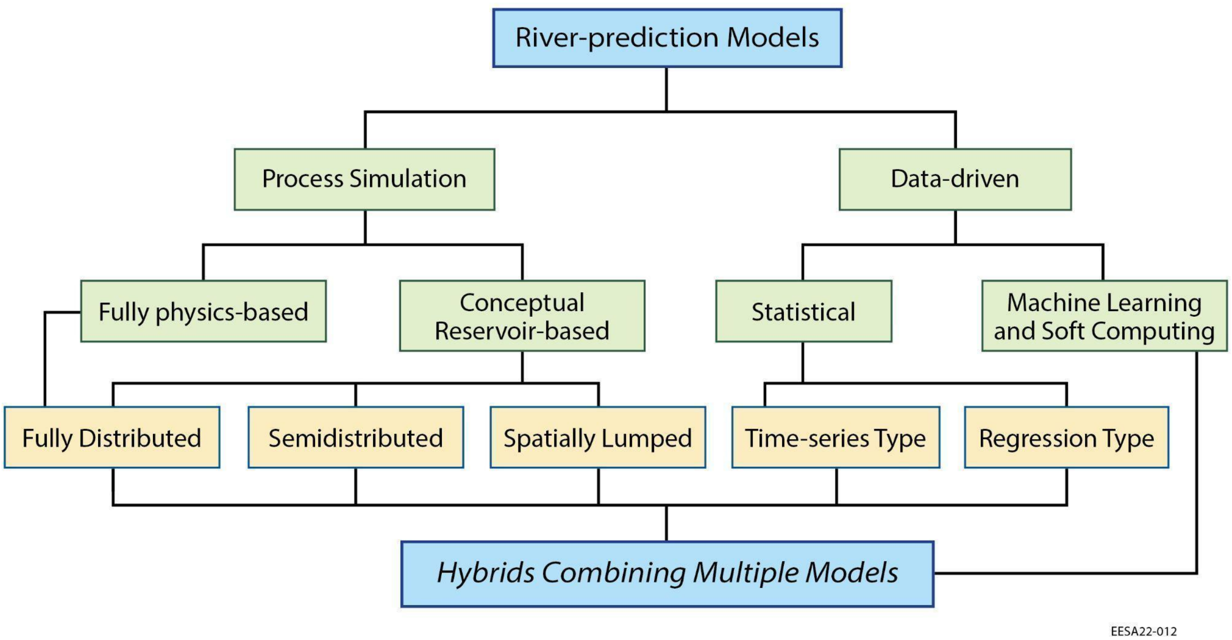


Figure 3-3. Traditionally river-prediction models have either been mechanistic, process-based codes or data-driven models that use ML or other statistical algorithms. There is significant potential to combine these approaches to create a new class of physically informed machine learning models that incorporate process information and physical consistencies for more accurate and scalable hydrological predictions (Source: Reproduced from Fleming and Gupta 2020, "The physics of river prediction," *Physics Today*, 73(7), pp. 46–52. doi:10.1063/PT.3.4523, with the permission of AIP Publishing).

Building a fast-to-evaluate surrogate model needs training data from the high-fidelity hydrologic model simulations; however, generating such simulation data at large scales is computationally demanding. ML methods can also address this problem, for example, using dimension reduction and Bayesian optimization techniques to simplify and optimize the neural network structure (Lu and Ricciuto 2019) and Bayesian neural networks (Lu, Ricciuto, et al. 2021) to build accurate surrogates with a small amount of training data. Another strategy is a direct and fundamental coupling between ML and process-based modeling in a differentiable manner such as in the differentiable parameter learning framework proposed by Tsai et al. (2020). This framework can be extended to learn better process descriptions and improve hydrologic model performance with the addition of data until it eventually reaches the performance of a purely data-driven model. It is important to use ML as a fixed, restricted set of model architectures, but to also leverage it as a collection of useful, fundamental techniques and philosophies. Other directions to explore include the use of physics-informed neural networks (Raissi, Perdikaris, and Karniadakis 2019) for forward modeling, inference of parameters, or constitutive relationships under a limited-data regime (Shen and Lawson 2021).

3.4.5 Data-driven Models for Knowledge Discovery

A topic that is ripe for ML applications is their use to derive insights from data to inform process understanding, generate hypotheses, and inform model improvements. For example, the concept of similarity mentioned above has been long studied in hydrology. A variety of AI/ML methods are available (see, for example, Table 4 of Sun and Scanlon 2019) to classify the landscape into subdomains that are expected to have similar hydrologic responses (semantic segmentation) and to build surrogate models for subdomains that are not modeled explicitly. (Meta) transfer learning methods are being explored to translate models built for well-observed locations to data-sparse or unmonitored sites, exploiting the power of the meta-learning model to determine similarities between sites (Willard et al. 2020). Although a suite of methods has been used for identification of hydrologic events including extremes and compound disturbances (Table 3-1), there is considerable room for improvement to use ML in this realm, particularly given considerations of non-stationary conditions in a future climate (chapter 8). ML can also be used to identify ecosystem control points (Bernhardt et al. 2017) and extreme spatial gradients that occur along transitions (e.g., terrestrial-aquatic interfaces, urban-natural boundaries) that may have an outsized impact on hydrologic functioning. Methods to treat rivers as connected networks, such as graph neural networks, and other network analysis methods should be considered. Additionally, methods adapted for sparse datasets such as few-shot or one-shot learning (Wang et al. 2020) need to be explored further given that hydrological studies are almost always data limited. There is a long history of using information theoretic approaches and causal inference in hydrology to extract information on relationships between variables (Kumar and Gupta 2020), and using such approaches in tandem with ML can be applied to determine system responses to different driving factors. Explainable ML tools like Shapley additive explanations (SHAP; Štrumbelj and Kononenko 2014), local interpretable model-agnostic explanation (LIME; Ribeiro, Singh, and Guestrin 2016), and interpretable LSTM models have significant potential to provide insight into the predictions generated by ML models (Lu, Ricciuto, and Liu 2022). Notably, more approaches that investigate how deep learning models make accurate predictions of hydrological processes (e.g., Lees et al. 2021) are needed, and can also provide insights into how these models learn and generalize behavior across diverse catchments. Ultimately, the goal is to be able to extract the maximum amount of information from the data as is possible (Nearing et al. 2021). Toward this end, some of the more exciting approaches attempt to discover governing equations using data-driven methods (e.g., Champion et al. 2019), although these require significant research to become applicable for predictions in complex hydrological systems.

3.4.6 Model-data Co-design

Most long-term monitoring networks have been designed and optimized to serve different stakeholder needs, which can result in a disconnect between observational data being acquired and data needed for modeling. Arguably the best way to improve model predictions is to acquire relevant data when and where needed for reducing model uncertainties in an iterative model-data design strategy. An example of this approach is Observing System Simulation Experiments (OSSEs) that aim to identify how many and what types of observations are needed to reach a desired model performance metric without actual observations being available (Masutani et al. 2010; Zeng et al. 2020). However, OSSEs are limited in their design and could be improved to expand beyond using process-based models to also include ML models, which may value a different set of observations. Other ML techniques in conjunction with OSSEs such as reinforcement learning methods can inform data collection (Hardin et al. 2021; Varadharajan et al. 2021), and such approaches should be considered in designing future observations. Edge computing for near-real time sampling decisions and connectivity with high performance computing centers with 5G or other high-speed networking can accelerate the transfer of information between field instrumentation and models. Fast training and adaptive learning methods that allow continuous integration or assimilation of data into ML models (Rao et al. 2021) are needed for successful model-data co-design strategies. This requires not only QA/QC but also automated approaches to gap-filling, data harmonization across diverse sensor streams, and new ways to represent complex data (e.g., probabilistic graph models) in ML models. Software pipelines that enable seamless assimilation of data from sensors into models and integration of model output into observational strategies are needed (Cholia, Varadharajan, and Pastorello 2021). Finally, uncertainty quantification (UQ) arising from sampling error, measurement noise, differences in spatiotemporal scales between data and models, and various types of structural and operational model errors must be taken into consideration. Frameworks for determining total uncertainties in ML or hybrid-ML models based on techniques such as Bayesian approaches or ensembles are needed (e.g., Psaros et al. 2022). The UQ can also be used to guide additional data collection in field (or lab measurements) and to advance model development.

3.5 Research Priorities

The workshop participants identified several priority activities that could accelerate the integration of AI/ML into hydrologic science for data, modeling, and capability development.

3.5.1 Data: Enhanced Observations and Improved Products

Data on hydrological processes and their drivers across a diverse range of ecosystems are the foundation for both ML-related research and process-based modeling. Participants in this workshop identified gaps in several data types, including: (1) precipitation and snowpack estimates in high-altitude and high-latitude regions where complex topography and partitioning between rain and snow pose significant challenges; (2) subsurface datasets such as groundwater levels, soil moisture and other vadose zone measurements, infiltration rates, soil and rock properties, etc., which are difficult to obtain and upscale; (3) water quality observations, a majority of which are collected using manually intensive sampling and lab characterization; (4) datasets at interfaces with steep gradients (e.g., terrestrial-aquatic boundaries and urban-natural divides) wherein fine spatial resolution is required to capture highly heterogeneous processes; (5) human activities such as water use, groundwater withdrawals, etc., which may not be publicly available; and (6) observations during extreme events, which are rare by definition and also tend to result in poor measurements or data gaps due to sensor failure. Data gaps were also identified with respect to the volume of data collectively generated, given that current hydrology ML models utilize large-sample, long-term datasets available from networks such as Fluxnet (Pastorello et al. 2020), USGS stream gage and water quality measurements, the Natural Resources Conservation Service's SNOTEL (snow telemetry), widely deployed weather stations, or remote sensing.

Here we identify three key research priorities related to (1) observations to address data gaps, (2) ML-assisted measurements, and (3) model-ready data products.

Addressing Data Gaps across a Wide Range of Ecosystems: Several of the identified data gaps are pertinent to DOE interests and could be prime targets for new data collection efforts. To date, data obtained from intensive field investigations at testbed sites have been heavily used to parametrize and validate process models. However, existing datasets and more broadly this approach to data collection may not be sufficient to capture processes across diverse catchment types or used directly for ML-based predictions at larger spatial scales. Future measurement strategies need to balance the depth of data obtained from measurements of smaller-scale process investigations with the breadth of data obtained through large-sample measurements in regional- and national-scale efforts (Gupta et al. 2014). Long-term measurements and curated data products such as those from the DOE's Ameriflux network and ARM atmospheric observatories are considered of high value for ML research. Other priorities include conducting long-term, integrated measurements to address some of the data gaps identified, and developing strategies to use data collected from smaller-scale intensive field investigations in large-scale ML models.

ML-Assisted Automated Measurement Technologies: Priority research areas include developing new approaches to increase the spatiotemporal coverage and resolution of hydrological observations, including the use of autonomous, ML-guided observational technologies (e.g., reinforcement learning), edge computing, and advances in 5G and wireless networking. Another priority research area that expands on the current DOE ModEx (model-experiment strategy) is the iterative co-design and integration of models and observations. This would result in a "self-guiding" capability to assimilate multiscale streaming of diverse data into models, and in turn, use of the models to guide field observations in near-real time for optimal data collection and resource utilization.

Model-Ready Data Products: Priority topics include: expanding standardized benchmark datasets for methods testing and development, such as curated datasets with key hydrologic variables like runoff, soil moisture, and groundwater levels; dedicated and automated QA/QC and gap-filling capabilities or products that handle the complex problems encountered in diverse measurements across a range of ecosystems; and data imputation and synthesis to generate high-resolution, gridded, model-ready products for a range of hydrologic and driver variables that are harmonized in variable nomenclature as well as spatiotemporal resolutions and scales. In particular, gridded data products like soil properties, soil thicknesses, and depth to bedrock are based on statistical relationships between the quantity of interest and readily observable co-varying properties, and can be improved using ML. Similarly, gridded evapotranspiration data products, derived from satellite observations and sparse ground-based observations, are needed for model evaluation. In addition, standardized climate forcing from downscaled ESM projections—while correcting for bias and underrepresented extreme events—is envisioned for use in integrated hydroterrestrial models to assess local impacts.

3.5.2 Integrated Hydroterrestrial Modeling

There were diverging views among the workshop participants on the relative importance of data-driven and process-based methods, with some arguing for a predominant role for data science and others arguing for combining traditional process-based models with ML. That divergence of opinion is likely due to differences in the classes of applications under consideration. Although there have been a number of recent papers showing superior performance of ML models over process models for hydrologic forecasting, short-term prediction is less likely to benefit from incorporating detailed process understanding than, say, efforts to understand consequences of climate change over decadal scales. Indeed, the data to support a purely data-driven approach to understanding decadal-scale change in a non-stationary climate does not exist in general. Although a space-for-time approach can sometimes be used to support a purely data-driven approach, that strategy breaks down for the many locations that are expected to experience future conditions with no current-day analog. Moreover, many hydrological processes cannot be observed well; direct observation of large-scale water flux is largely limited to stream discharges, and subsurface fluxes of water are difficult to observe at any scale. Those data limitations and the need to account for non-stationarity and potential threshold (tipping point) behaviors mean that process-based models will continue to be important for hydrologic science even as the role of ML increases. A key priority then is to have approaches that allow data-driven and process-oriented approaches to be combined. A variety of such approaches have already been explored, but it is important to test those and other emerging ML algorithms in real-world conditions and not just theoretical or proof-of-concept cases. Overall, the consensus from this workshop was that the coupling of mechanistic and ML codes into an integrated Earth system modeling framework is necessary to advance DOE priorities and is a potent target for future research regardless of whether the path chosen is *"process-first"* or *"AI-first."* For this, it will be desirable to transcend the dichotomy between the two approaches and investigate how data-driven models and process models can complement and leverage each other for hydrological predictions at decision-relevant scales.

Here we identify five research priorities focused on (1) accelerating high-resolution regional-to-continental scale process-informed models that are applicable across pristine and human-impacted watersheds, (2) improving subgrid parameterizations of hydrologic processes in large-scale models, (3) representing hydrological changes caused by extreme events and non-stationary conditions under future climate scenarios, (4) using AI/ML for knowledge discovery, and (5) improving model interpretability and uncertainty quantification.

Accelerating High-resolution Regional-to-Continental Scale Process-informed Models: Priority activities for accelerating process-informed hydrological models include the following: use ML methods to improve the accuracy of surrogate models with high-dimensional inputs (e.g., meteorological forcings or inflows from upstream catchments) and to reduce the number of forward runs required to train such surrogates; pursue tighter integration of surrogate models into large-scale simulations to enable on-the-fly construction of surrogate models and on-demand switching between process-resolving and ML models; advance AI/ML methods for partitioning of large regions into subregions with similar hydrological response; and advance use and development of data-driven model architectures and foundational ML methods for large-scale predictions with sparse data. A particular focus should be understanding and predictions of hydrological (sub)systems that are heavily empirical and lack reliable process descriptions (e.g., human-impacted catchments, subsurface with preferential flow pathways).

Subgrid Parameterizations: Important activities for improving parameterizations of hydrological processes in large-scale hydrology models include the following: ML-based representations of runoff and inundation fraction in permafrost landscapes for use in ESMs, ML-based models for lateral subsurface flow and surface flow within ESM grid cells, and representation of subgrid depression storage in regional-scale hydrology models. These subgrid parameterizations could be developed as surrogates for expensive process-based models or directly from data.

Extreme Events and Non-stationarity: Priority activities for representing extreme events include the following: improve AI/ML-based approaches to post-process ESM projections to better represent extreme events (tail distribution correction and downscaling); using ML to improve detection and attribution of extreme events; improving the ability to learn from small sample sizes; and exploring methods to extrapolate model projections that account for non-stationarity under future climate scenarios. A particular challenge is to consider the cascading impacts of compound events (e.g., alternating flood and drought conditions). Other priorities include modeling the impacts of extreme events on natural and urban systems.

AI/ML for Knowledge Discovery: ML can help improve understanding of the processes governing feedback between natural systems and human actions, which will contribute to improvements in modeling capabilities. Besides identification of subregions of hydrological similarity identified previously, other important topics include using ML and other data-driven methods (e.g., causal inference and information theory) to identify relationships between driving variables and hydrologic responses, considering river networks and order in determining hydrologic functioning, and exploring the importance of interfaces with spatial gradients or ecosystem control points on hydrologic behavior.

Building Interpretable Models with Uncertainty Quantification: The priority areas described above use ML-based surrogate models or ML-based component models. Quantifying uncertainty in those models and constructing them to be interpretable are common priority topics. Uncertainty quantification is important to help identify when conditions (climatic or land use) are encountered that would render the model inaccurate, and what data we should collect to improve model accuracy. Additionally, a priority is to improve the interpretability and explainability of ML models, and particularly understand the mechanisms by which classical ML/deep learning models learn to make accurate predictions.

3.5.3 Capability Development

There are several barriers to realizing the full potential of AI/ML in the hydrologic sciences, and actions to resolve these can be identified. Software tools that allow process-based models and ML to be combined exist only at a coarse-grained level. Modular tools that allow for such integration at a finer granularity would greatly facilitate numerical experimentation and allow for more rapid progress in hybridizing process-based and ML approaches. High-quality training and input data are clearly an issue. Combining ML with indirectly related information (proxy data) is a promising approach. However, much of the existing data is siloed in isolated data systems, making it difficult to combine them to extract the maximum amount of information. Data-broker tools to facilitate the discovery, extraction, and fusion of siloed data would be valuable, as would the development of standardized benchmark datasets. Cyberinfrastructure advancements are needed that bring together the community's data, model, and analytical capabilities (Figure 3-4). Other important needs noted by several participants were workforce development efforts to cross-train staff in hydrologic science and in computational AI/ML methods, and the need for future opportunities to spur collaborative engagement between domain and computational scientists.

Here we identify four priority research areas to develop capabilities for (1) modeling, (2) model-data integration, (3) data management and synthesis, and (4) workforce development.

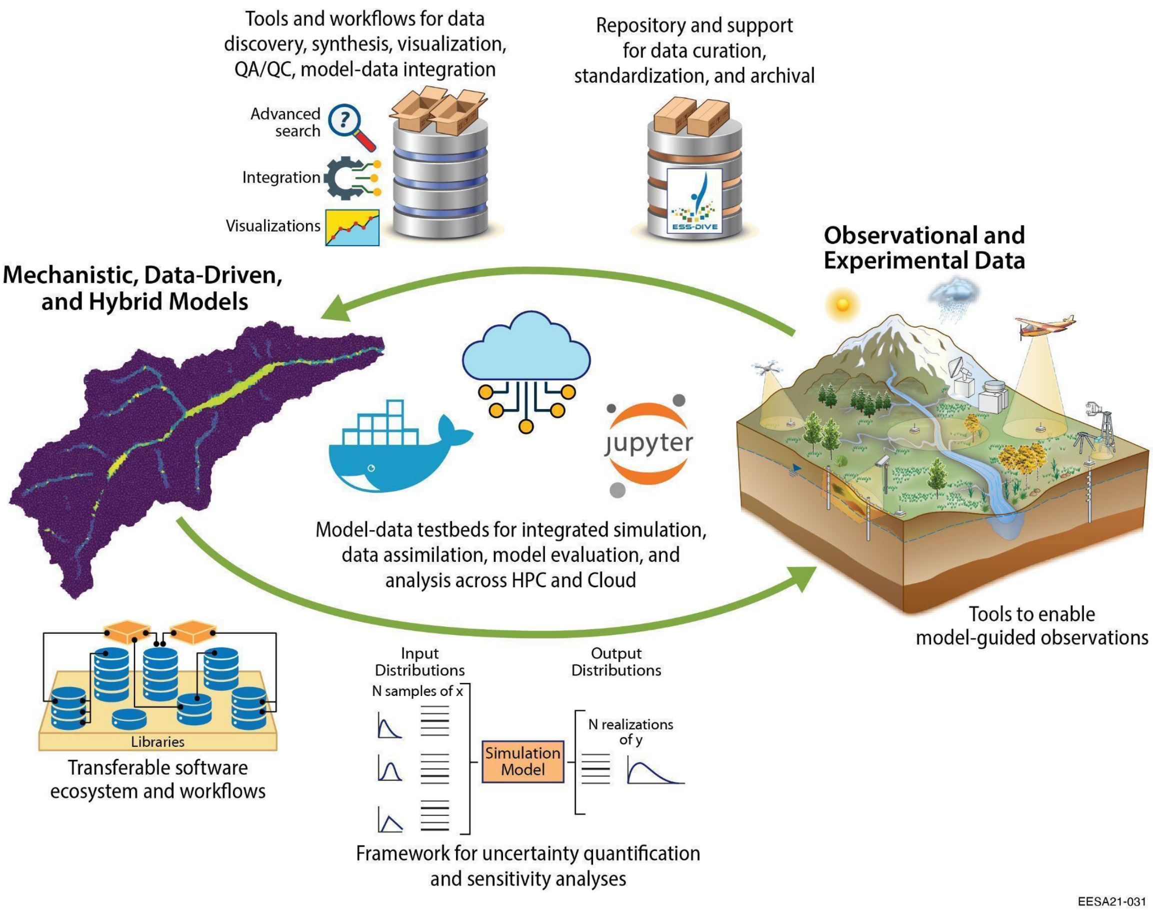


Figure 3-4. Data-model and model-data pipelines that include data management tools, software workflows to enable data discovery and integration, scaling and transformation of data for model needs, and testbeds to enable co-located big datasets that can be fed into models on HPC facilities. Examples of technologies that can be used in modeling testbeds include Jupyter notebooks (http://jupyter.org), Docker containers (http://docker.com), and tools that enable seamless execution of ML/hybrid models on HPC and cloud computing centers (Source: Lawrence Berkeley National Laboratory).

Capabilities for Modeling: Priority areas include developing capabilities to speed up use of ML models such as AutoML frameworks (Hutter, Kotthoff, and Vanschoren 2019) with automated architecture selection and hyperparameter optimization; creating modeling testbeds and tools to enable model ensembles and intercomparisons; and software frameworks that make it easier to explore combinations of ML and process-based models. While combinations of ML and process-based models have been demonstrated at a systems level, software frameworks are needed to allow finer-grained integration where both ML- and process-based models are components in a larger system. Availability of mid-range computing and sufficient graphical processing unit (GPU)-based computational resources is critical to advance the widespread adoption of ML in hydrologic sciences.

Software Workflows and Data-Model Pipelines: Priority directions include developing data-to-model pipelines that enable integration of observational data with simulation codes, which would dramatically improve the efficiency of modeling workflows; expanding the use of modeling testbeds to include associated datasets with co-located and covarying observations; and developing tools to harness model output to inform measurement collection in a feedback loop.

Data Capabilities: A priority direction is to develop infrastructure and software tools to make it easier to extract and combine data from multiple, independent data systems with the DOE and across other data sources (including federal agencies). This infrastructure could take the form of federated, cloud-based infrastructure or extensions of brokering tools. Separately, efforts to create and adopt data standards and interoperability between DOE data systems (e.g., ESS-DIVE, ARM, ESGF) would make it easier to curate and synthesize diverse data funded by different programs. Other priorities include developing tools for data discovery (e.g., ML-enabled metadata extraction and search capabilities), subsetting, and visualization that lower the burden on scientists to obtain the data needed for ML research.

Workforce Development: Priority directions include enabling domain scientists to obtain a deep understanding of ML approaches through trainings and attending computational workshops (e.g., NeurIPs), creating opportunities to engage cross-functional teams of computational and domain scientists, and supporting the building of an international hydrological-ML community that can promote the use of ML in hydrological sciences.

3.6 Short-Term (<5 years), 5-year, and 10-year Goals

3.6.1 Short-term (<5 years) Goals

A near-term goal is organizing workshops or other community efforts to help prioritize collection, synthesis, and curation of data that would be most valuable for advancing the use of AI in hydrology. Expanding standardized benchmark datasets for ML testing and development is another. In particular, approaches to obtain data at desired scales and resolutions need to be identified and initiated. For integrated hydroterrestrial models, software frameworks that make it easier to combine process-based and ML-based models is a near-term goal, as are data-driven methods for identifying hydrologically similar catchments. Developing ML-based surrogate models for a variety of applications is achievable in the next few years. In the short-term, a critical examination of when, where, and how data-driven models add value for modeling and hypothesis generation would guide future research directions.

3.6.2 5-yearGoals

A 5-year goal is to use ML to improve key data products for model inputs and evaluation (e.g., soil properties, evapotranspiration, climate forcing) and develop software frameworks that enable assimilation and synthesis of data in models. Efforts to close identified data gaps and collect appropriate data for ML models need to be underway in 5 years to attain the long-term goals. Identifying promising approaches for a deep integration of process-based and ML (e.g., surrogate models trained on high-resolution process-based hydrology models, physics-informed neural networks) to achieve regional- to continental-scale predictive capabilities is an important 5-year goal. Another goal is to develop frameworks to improve explainability and uncertainty quantification for the hybrid models. Development of a workforce with deep knowledge across hydrology and AI/ML in 5 years is essential to making future progress.

3.6.3 10-year Goals

In 10 years, ML should be deeply woven into hydrology, and not considered two distinct fields as they are now. This likely will parallel the development of HPC to exascale modeling development currently being carried out by domain scientists. An order-of-magnitude increase in observations, particularly in under-represented regions and hydrologic systems, may be necessary to achieve the 10-year goals. Developing technologies that seamlessly assimilate the latest available data into models and use models to guide observations in a near real-time feedback loop is a 10-year goal. The use of ML-based models to improve representation of lateral surface and subsurface flow, biogeochemical processes, human impacts, and extreme events in ESMs is another 10-year goal. An important long-term goal is to enable basin-scale, decadal-timeframe, integrated hydrology model projections with sufficient spatial and temporal resolution to provide actionable information to stakeholders.

3.7 References

Abbott, Benjamin W., Kevin Bishop, Jay P. Zarnetske, David M. Hannah, Rebecca J. Frei, Camille Minaudo, F. Stuart Chapin III, et al. 2019. "A Water Cycle for the Anthropocene." *Hydrological Processes* 33 (23): 3046–52. https://doi.org/https://doi.org/10.1002/hyp.13544.

Addor, Nans, Andrew J. Newman, Naoki Mizukami, and Martyn P. Clark. 2017. "The CAMELS Data Set: Catchment Attributes and Meteorology for Large-Sample Studies." *Hydrology and Earth System Sciences* 21 (10): 5293–5313. https://doi.org/10.5194/hess-21-5293-2017.

Adusumilli, S., A. A. Borsa, M. A. Fish, H. K. McMillan, and F. Silverii. 2019. "A Decade of Water Storage Changes Across the Contiguous United States From GPS and Satellite

Gravity." *Geophysical Research Letters* 46 (22): 13006–15. https://doi.org/10.1029/2019GL085370.

Araya, S. N., A. Fryjoff-Hung, A. Anderson, J. H. Viers, and T. A. Ghezzehei. 2021. "Advances in Soil Moisture Retrieval from Multispectral Remote Sensing Using Unoccupied Aircraft Systems and Machine Learning Techniques." *Hydrology and Earth System Sciences* 25 (5): 2739–58. https://doi.org/10.5194/hess-25-2739-2021.

di Baldassarre, G., A. Viglione, G. Carr, L. Kuil, J. L. Salinas, and G. Blöschl. 2013. "Socio-Hydrology: Conceptualising Human-Flood Interactions." *Hydrology and Earth System Sciences* 17 (8): 3295–3303. https://doi.org/10.5194/hess-17-3295-2013.

Bernhardt, Emily S., Joanna R. Blaszczak, Cari D. Ficken, Megan L. Fork, Kendra E. Kaiser, and Erin C. Seybold. 2017. "Control Points in Ecosystems: Moving Beyond the Hot Spot Hot Moment Concept." *Ecosystems* 20 (4): 665–82. https://doi.org/10.1007/s10021-016-0103-y.

Bierkens, Marc F. P. 2015. "Global Hydrology 2015: State, Trends, and Directions." *Water Resources Research* 51 (7): 4923–47. https://doi.org/10.1002/2015WR017173.

Blair, P., and W. Buytaert. 2016. "Socio-Hydrological Modelling: A Review Asking ``why,~what~and~how?''." *Hydrology and Earth System Sciences* 20 (1): 443–78. https://doi.org/10.5194/hess-20-443-2016.

Blöschl, Günter, Marc F. P. Bierkens, Antonio Chambel, Christophe Cudennec, Georgia Destouni, Aldo Fiori, James W. Kirchner, et al. 2019. "Twenty-Three Unsolved Problems in Hydrology (UPH) – a Community Perspective." *Hydrological Sciences Journal* 64 (10): 1141–58. https://doi.org/10.1080/02626667.2019.1620507.

Bodesheim, Paul, Martin Jung, Fabian Gans, Miguel D. Mahecha, and Markus Reichstein. 2018. "Upscaled Diurnal Cycles of Land–Atmosphere Fluxes: A New Global Half-Hourly Data Product." *Earth System Science Data* 10 (3): 1327–65. https://doi.org/10.5194/essd-10-1327-2018.

Boomen, Ryan L vanden, Zeyun Yu, and Qian Liao. 2021. "Application of Deep Learning for Imaging-Based Stream Gaging." *Water Resources Research* 57 (11): e2021WR029980. https://doi.org/https://doi.org/10.1029/2021WR029980.

Bury, Thomas M., R. I. Sujith, Induja Pavithran, Marten Scheffer, Timothy M. Lenton, Madhur Anand, and Chris T. Bauch. 2021. "Deep Learning for Early Warning Signals of Tipping Points." *Proceedings of the National Academy of Sciences* 118 (39): e2106140118. https://doi.org/10.1073/pnas.2106140118.

Callaghan, Max, Carl-Friedrich Schleussner, Shruti Nath, Quentin Lejeune, Thomas R. Knutson, Markus Reichstein, Gerrit Hansen, et al. 2021. "Machine-Learning-Based Evidence and Attribution Mapping of 100,000 Climate Impact Studies." *Nature Climate Change* 11 (11): 966–72. https://doi.org/10.1038/s41558-021-01168-6.

Cannon, Alex J. 2012. "Neural Networks for Probabilistic Environmental Prediction: Conditional Density Estimation Network Creation and Evaluation (CaDENCE) in R." *Comput. Geosci.* 41 (April): 126–35. https://doi.org/10.1016/j.cageo.2011.08.023.

Champion, Kathleen, Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. 2019. "Data-Driven Discovery of Coordinates and Governing Equations." *Proceedings of the National Academy of Sciences of the United States of America* 116 (45): 22445–51. https://doi.org/10.1073/PNAS.1906995116/SUPPL_FILE/PNAS.1906995116.SAPP.PDF.

Chen, H., and V. Chandrasekar. 2021. "Deep Learning for Surface Precipitation Estimation Using Multidimensional Polarimetric Radar Measurements." In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 359–62. https://doi.org/10.1109/IGARSS47720.2021.9554847.

Cho, Brian, Teresa Dayrit, Yuan Gao, Zhe Wang, Tianzhen Hong, Alex Sim, and Kesheng Wu. 2020. "Effective Missing Value Imputation Methods for Building Monitoring Data." In *2020 IEEE International Conference on Big Data (Big Data)*, 2866–75. https://doi.org/10.1109/BigData50022.2020.9378230.

Cholia, Shreyas, Charuleka Varadharajan, and Gilberto Pastorello. 2021. "Integrating Models with Real-Time Field Data for Extreme Events: From Field Sensors to Models and Back with AI in the Loop." United States. https://doi.org/10.2172/1769727.

Collins, William D. 2021. "Model Hierarchy for Mountainous Hydrological Observatories (MH2O)," April. https://doi.org/10.2172/1769748.

Coon, Ethan, Daniil Svyatsky, Ahmad Jan, Eugene Kikinzon, Markus Berndt, Adam Atchley, Dylan Harp, et al. 2019. "Advanced Terrestrial Simulator." https://doi.org/10.11578/dc.20190911.1.

Coon, Ethan T., J. David Moulton, Evgeny Kikinzon, Markus Berndt, Gianmarco Manzini, Rao Garimella, Konstantin Lipnikov, and Scott L. Painter. 2020. "Coupling Surface Flow and Subsurface Flow in Complex Soil Structures Using Mimetic Finite Differences." *Advances in Water Resources* 144 (October): 103701. https://doi.org/10.1016/J.ADVWATRES.2020.103701.

Cromwell, Erol, Pin Shuai, Peishi Jiang, Ethan T. Coon, Scott L. Painter, J. David Moulton, Youzuo Lin, and Xingyuan Chen. 2021. "Estimating Watershed Subsurface Permeability From Stream Discharge Data Using Deep Neural Networks." *Frontiers in Earth Science* 9 (February): 3. https://doi.org/10.3389/feart.2021.613011.

Crystal-Ornelas, Robert, Charuleka Varadharajan, Danielle Christianson, Joan Damerow, Helen Weierbach, Emily Robles, Lavanya Ramakrishnan, Boris Faybishenko, and Gilberto Pastorello. 2021. "A Library of AI-Assisted FAIR Water Cycle and Related Disturbance Datasets to Enable Model Training, Parameterization and Validation." United States. https://doi.org/10.2172/1769646.

Daw, Arka, R. Quinn Thomas, Cayelan C. Carey, Jordan S. Read, Alison P. Appling, and Anuj Karpatne. 2020. "Physics-Guided Architecture (PGA) of Neural Networks for Quantifying Uncertainty in Lake Temperature Modeling." In *Proceedings of the 2020 SIAM International Conference on Data Mining (SDM)*, 532–40. Proceedings. Society for Industrial and Applied Mathematics. https://doi.org/doi:10.1137/1.9781611976236.60.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. "ImageNet: A Large-Scale Hierarchical Image Database." In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–55. https://doi.org/10.1109/CVPR.2009.5206848.

Deng, Li. 2012. "The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]." *IEEE Signal Processing Magazine* 29 (6): 141–42. https://doi.org/10.1109/MSP.2012.2211477.

Donadio, Carlo, Massimo Brescia, Alessia Riccardo, Giuseppe Angora, Michele Delli Veneri, and Giuseppe Riccio. 2021. "A Novel Approach to the Classification of Terrestrial Drainage Networks Based on Deep Learning and Preliminary Results on Solar System Bodies." *Scientific Reports* 11 (1): 5875. https://doi.org/10.1038/s41598-021-85254-x.

Elshafei, Y., M. Sivapalan, M. Tonts, and M. R. Hipsey. 2014. "A Prototype Framework for Models of Socio-Hydrology: Identification of Key Feedback Loops and Parameterisation Approach." *Hydrology and Earth System Sciences* 18 (6): 2141–66. https://doi.org/10.5194/hess-18-2141-2014.

Falcone, James A., Daren M. Carlisle, David M. Wolock, and Michael R. Meador. 2010. "GAGES: A Stream Gage Database for Evaluating Natural and Altered Flow Conditions in the Conterminous United States." *Ecology* 91 (2): 621–621. https://doi.org/10.1890/09-0889.1.

Fang, Kuai, Chaopeng Shen, Daniel Kifer, and Xiao Yang. 2017. "Prolongation of SMAP to Spatiotemporally Seamless Coverage of Continental U.S. Using a Deep Learning Neural Network." *Geophysical Research Letters* 44 (21): 11, 11–30, 39. https://doi.org/https://doi.org/10.1002/2017GL075619.

Feldman, Daniel, V. Chandrasekar, P. Dennedy-Frank, Dipankar Dwivedi, Mayur Mudigonda, Karhik Mukkavilli, Utkarsh Mital, and Erica Siirila-Woodburn. 2021. "Reliable Modeling and Prediction of Precipitation & Radiation for Mountainous Hydrology," April. https://doi.org/10.2172/1769771.

Feng, Dapeng, Kuai Fang, and Chaopeng Shen. 2020. "Enhancing Streamflow Forecast and Extracting Insights Using Long-Short Term Memory Networks With Data Integration at Continental Scales." *Water Resources Research* 56 (9): e2019WR026793. https://doi.org/https://doi.org/10.1029/2019WR026793.

Fleming, Sean W., Dominique R. Bourdin, Dave Campbell, Roland B. Stull, and Tobi Gardner. 2015. "Development and Operational Testing of a Super-Ensemble Artificial Intelligence Flood-Forecast Model for a Pacific Northwest River." *JAWRA Journal of the American Water Resources Association* 51 (2): 502–12. https://doi.org/https://doi.org/10.1111/jawr.12259.

Fleming, Sean W., and Hoshin v. Gupta. 2020. "The Physics of River Prediction." *Physics Today* 73 (7): 46–52. https://doi.org/10.1063/PT.3.4523.

Fleming, Sean W., Velimir v. Vesselinov, and Angus G. Goodbody. 2021. "Augmenting Geophysical Interpretation of Data-Driven Operational Water Supply Forecast Modeling for a Western US River Using a Hybrid Machine Learning Approach." *Journal of Hydrology* 597 (June): 126327. https://doi.org/10.1016/J.JHYDROL.2021.126327.

Frame, J., F. Kratzert, D. Klotz, M. Gauch, G. Shelev, O. Gilon, L. M. Qualls, H. v Gupta, and G. S. Nearing. 2021. "Deep Learning Rainfall-Runoff Predictions of Extreme Events." *Hydrology and Earth System Sciences Discussions* 2021: 1–20. https://doi.org/10.5194/hess-2021-423.

Freeze, R. Allan, and R. L. Harlan. 1969. "Blueprint for a Physically-Based, Digitally-Simulated Hydrologic Response Model." *Journal of Hydrology* 9 (3): 237–58. https://doi.org/10.1016/0022-1694(69)90020-1.

Gao, Jing, Peng Li, Zhikui Chen, and Jianing Zhang. 2020. "A Survey on Deep Learning for Multimodal Data Fusion." *Neural Computation* 32 (5): 829–64. https://doi.org/10.1162/neco_a_01273.

Gleeson, Tom, Lan Wang-Erlandsson, Miina Porkka, Samuel C. Zipper, Fernando Jaramillo, Dieter Gerten, Ingo Fetzer, et al. 2020. "Illuminating Water Cycle Modifications and Earth System Resilience in the Anthropocene." *Water Resources Research* 56 (4): e2019WR024957. https://doi.org/https://doi.org/10.1029/2019WR024957.

Goodwell, Allison E., Praveen Kumar, Aaron W. Fellows, and Gerald N. Flerchinger. 2018. "Dynamic Process Connectivity Explains Ecohydrologic Responses to Rainfall Pulses and Drought." *Proceedings of the National Academy of Sciences* 115 (37): E8604–13. https://doi.org/10.1073/pnas.1800236115.

Guo, Yuhan, Yongqiang Zhang, Lu Zhang, and Zhonggen Wang. 2021. "Regionalization of Hydrological Modeling for Predicting Streamflow in Ungauged Catchments: A Comprehensive Review." *WIREs Water* 8 (1): e1487. https://doi.org/https://doi.org/10.1002/wat2.1487.

Gupta, H. v., C. Perrin, G. Blöschl, A. Montanari, R. Kumar, M. Clark, and V. Andréassian. 2014. "Large-Sample Hydrology: A Need to Balance Depth with Breadth." *Hydrology and Earth System Sciences* 18 (2): 463–77. https://doi.org/10.5194/hess-18-463-2014.

Hardin, Joseph C., Jr., William I. Gustafson, Adam Varble, Jim Marquis, Jennifer Comstock, Andrew Geiss, Sam Silva, et al. 2021. "The Usage of Observing System Simulation Experiments and Reinforcement Learning to Optimize Experimental Design and Operation," April. https://doi.org/10.2172/1769782.

Hayatbini, Negin, Bailey Kong, Kuo-lin Hsu, Phu Nguyen, Soroosh Sorooshian, Graeme Stephens, Charless Fowlkes, Ramakrishna Nemani, and Sangram Ganguly. 2019. "Conditional Generative Adversarial Networks (CGANs) for Near Real-Time Precipitation Estimation from Multispectral GOES-16 Satellite Imageries—PERSIANN-CGAN." *Remote Sensing* 11 (19): 2193. https://doi.org/10.3390/rs11192193.

Hill, Paul R., Anurag Kumar, Marouane Temimi, and David R. Bull. 2020. "HABNet: Machine Learning, Remote Sensing-Based Detection of Harmful Algal Blooms." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13: 3229–39. https://doi.org/10.1109/JSTARS.2020.3001445.

Hill, Ryan A., Marc H. Weber, Scott G. Leibowitz, Anthony R. Olsen, and Darren J. Thornbrugh. 2016. "The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States." *JAWRA Journal of the American Water Resources Association* 52 (1): 120–28. https://doi.org/https://doi.org/10.1111/1752-1688.12372.

Horsburgh, Jeffery S., Stephanie L. Reeder, Amber Spackman Jones, and Jacob Meline. 2015. "Open Source Software for Visualization and Quality Control of Continuous Hydrologic and Water Quality Sensor Data." *Environmental Modelling & Software* 70 (August): 32–44. https://doi.org/10.1016/J.ENVSOFT.2015.04.002.

Hutter, Frank, Lars Kotthoff, and Joaquin Vanschoren, eds. 2019. *Automated Machine Learning: Methods, Systems, Challenges*. 1st ed. Springer Cham. https://doi.org/https://doi.org/10.1007/978-3-030-05318-5.

Jarvis, Nicholas, John Koestel, and Mats Larsbo. 2016. "Understanding Preferential Flow in the Vadose Zone: Recent Advances and Future Prospects." *Vadose Zone Journal* 15 (12): 1–11. https://doi.org/10.2136/VZJ2016.09.0075.

Jia, Xiaowei, Jacob Zwart, Jeffrey Sadler, Alison Appling, Samantha Oliver, Steven Markstrom, Jared Willard, et al. 2021. "Physics-Guided Recurrent Graph Model for Predicting Flow and Temperature in River Networks." *Proceedings*, 612–20. https://doi.org/10.1137/1.9781611976700.69.

Koestel, John, and Helena Jorda. 2014. "What Determines the Strength of Preferential Transport in Undisturbed Soil under Steady-State Flow?" *Geoderma* 217–218 (April): 144–60. https://doi.org/10.1016/J.GEODERMA.2013.11.009.

Konapala, Goutam, Shih-Chieh Kao, Scott L. Painter, and Dan Lu. 2020. "Machine Learning Assisted Hybrid Models Can Improve Streamflow Simulation in Diverse Catchments across the Conterminous US." *Environmental Research Letters* 15 (10): 104022. https://doi.org/10.1088/1748-9326/aba927.

Kratzert, Frederik, Daniel Klotz, Claire Brenner, Karsten Schulz, and Mathew Herrnegger. 2018. "Rainfall–Runoff Modelling Using Long Short-Term Memory (LSTM) Networks." *Hydrology and Earth System Sciences* 22 (11): 6005–22. https://doi.org/10.5194/hess-22-6005-2018.

Kratzert, Frederik, Daniel Klotz, Mathew Herrnegger, Alden K. Sampson, Sepp Hochreiter, and Grey S. Nearing. 2019. "Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning." *Water Resources Research* 55 (12): 11344–54. https://doi.org/https://doi.org/10.1029/2019WR026065.

Kruse, Peter. 2018. "Review on Water Quality Sensors." *Journal of Physics D: Applied Physics* 51 (20): 203002. https://doi.org/10.1088/1361-6463/aabb93.

Kumar, Praveen, and Hoshin v. Gupta. 2020. "Debates—Does Information Theory Provide a New Paradigm for Earth Science?" *Water Resources Research* 56 (2): e2019WR026398. https://doi.org/10.1029/2019WR026398.

Lees, T., S. Reece, F. Kratzert, D. Klotz, M. Gauch, J. de Bruijn, R. Kumar Sahu, P. Greve, L. Slater, and S. Dadson. 2021. "Hydrological Concept Formation inside Long Short-Term Memory (LSTM) Networks." *Hydrology and Earth System Sciences Discussions* 2021: 1–37. https://doi.org/10.5194/hess-2021-566.

Li, Xiang, Ankush Khandelwal, Xiaowei Jia, Kelly Cutler, Rahul Ghosh, Arvind Renganathan, Shaoming Xu, et al. 2022. "Regionalization in a Global Hydrologic Deep Learning Model: From Physical Descriptors to Random Vectors." https://doi.org/10.1002/ESSOAR.10510083.1.

Loritz, Ralf, Hoshin Gupta, Conrad Jackisch, Martijn Westhoff, Axel Kleidon, Uwe Ehret, and Erwin Zehe. 2018. "On the Dynamic Nature of Hydrological Similarity." *Hydrology and Earth System Sciences* 22 (7): 3663–84. https://doi.org/10.5194/hess-22-3663-2018.

Lu, D., and D. Ricciuto. 2019. "Efficient Surrogate Modeling Methods for Large-Scale Earth System Models on Machine-Learning Techniques." *Geoscientific Model Development* 12 (5): 1791–1807. https://doi.org/10.5194/gmd-12-1791-2019.

Lu, Dan, Goutam Konapala, Scott L. Painter, Shih Chieh Kao, and Sudershan Gangrade. 2021. "Streamflow Simulation in Data-Scarce Basins Using Bayesian and Physics-Informed Machine Learning Models." *Journal of Hydrometeorology* 22 (6): 1421–38. https://doi.org/10.1175/JHM-D-20-0082.1.

Lu, Dan, Daniel Ricciuto, Marcia Branstetter, Eric Pierce, Seung-Swan Lim, and David Womble. 2021. "A Bayesian Neural Network Ensemble Approach for Improving Large-Scale Streamflow Predictability," March. https://doi.org/10.2172/1769641.

Lu, Dan, Daniel Ricciuto, and Siyan Liu. 2022. "An Interpretable Machine Learning Model for Advancing Terrestrial Ecosystem Predictions." In *The International Conference on Learning Representations (ICLR): AI for Earth Sciences*, 1–10. https://www.ornl.gov/publication/interpretable-machine-learning-model-advancing-terrestrial-ecosystem-predictions.

Lundquist, Jessica, Mimi Hughes, Ethan Gutmann, and Sarah Kapnick. 2019. "Our Skill in Modeling Mountain Rain and Snow Is Bypassing the Skill of Our Observational Networks." *Bulletin of the American Meteorological Society* 100 (12): 2473–90. https://doi.org/10.1175/BAMS-D-19-0001.1.

Masutani, Michiko, Thomas W. Schlatter, Ronald M. Errico, Ad Stoffelen, Erik Andersson, William Lahoz, John S. Woollen, G. David Emmitt, Lars-Peter Riishøjgaard, and Stephen J. Lord. 2010. "Observing System Simulation Experiments." In *Data Assimilation: Making Sense of Observations*, edited by William Lahoz, Boris Khattatov, and Richard Menard, 647–79. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74703-1_24.

Matos, João, and Octavian Postolache. 2016. "IoT Enabled Aquatic Drone for Environmental Monitoring." In *2016 International Conference and Exposition on Electrical and Power Engineering (EPE)*, 598–603. https://doi.org/10.1109/ICEPE.2016.7781410.

Maxwell, R. M., L. E. Condon, and S. J. Kollet. 2015. "A High-Resolution Simulation of Groundwater and Surface Water over Most of the Continental US with the Integrated Hydrologic Model ParFlow V3." *Geoscientific Model Development* 8 (3): 923–37. https://doi.org/10.5194/gmd-8-923-2015.

Mital, Utkarsh, Dipankar Dwivedi, James B. Brown, Boris Faybishenko, Scott L. Painter, and Carl I. Steefel. 2020. "Sequential Imputation of Missing Spatio-Temporal Precipitation Data Using Random Forests." *Frontiers in Water* 2 (August). https://doi.org/10.3389/frwa.2020.00020.

Moghaddam, Mohammad A., Ty P. A. Ferre, Xingyuan Chen, Kewei Chen, and Mohammad Reza Ehsani. 2022. "Application of Machine Learning Methods in Inferring Surface Water Groundwater Exchanges Using High Temporal Resolution Temperature Measurements," January. https://doi.org/10.48550/arxiv.2201.00726.

Mosavi, Amir, Pinar Ozturk, and Kwok-wing Chau. 2018. "Flood Prediction Using Machine Learning Models: Literature Review." *Water* 10 (11): 1536. https://doi.org/10.3390/w10111536.

Mount, N. J., H. R. Maier, E. Toth, A. Elshorbagy, D. Solomatine, F.-J. Chang, and R. J. Abrahart. 2016. "Data-Driven Modelling Approaches for Socio-Hydrology: Opportunities and Challenges within the Panta Rhei Science Plan." *Hydrological Sciences Journal* 61 (7): 1192–1208. https://doi.org/10.1080/02626667.2016.1159683.

National Research Council. 2012. *Challenges and Opportunities in the Hydrologic Sciences*. Washington, DC: The National Academies Press. https://doi.org/10.17226/13293.

Nearing, Grey S., Frederik Kratzert, Alden Keefe Sampson, Craig S. Pelissier, Daniel Klotz, Jonathan M. Frame, Cristina Prieto, and Hoshin v. Gupta. 2021. "What Role Does Hydrological Science Play in the Age of Machine Learning?" *Water Resources Research* 57 (3): e2020WR028091. https://doi.org/10.1029/2020WR028091.

Ombadi, Mohammed, Phu Nguyen, Soroosh Sorooshian, and Kuo-lin Hsu. 2020. "Evaluation of Methods for Causal Discovery in Hydrometeorological Systems." *Water Resources Research* 56 (7): e2020WR027251. https://doi.org/https://doi.org/10.1029/2020WR027251.

Overeem, Aart, Hidde Leijnse, and Remko Uijlenhoet. 2013. "Country-Wide Rainfall Maps from Cellular Communication Networks." *Proceedings of the National Academy of Sciences of the United States of America* 110 (8): 2741–45. https://doi.org/10.1073/PNAS.1217961110.

Paepae, Thulane, Pitshou N. Bokoro, and Kyandoghere Kyamakya. 2021. "From Fully Physical to Virtual Sensing for Water Quality Assessment: A Comprehensive Review of the Relevant State-of-the-Art." *Sensors* 21 (21): 6971. https://doi.org/10.3390/s21216971.

Painter, Scott L., Ethan Coon, and Dan Lu. 2021. "AI-Directed Adaptive Multifidelity Modeling of Water Availability and Quality at River Basin Scales," April. https://doi.org/10.2172/1769669.

Painter, Scott L., Ethan T Coon, Adam L Atchley, Markus Berndt, Rao Garimella, J. David Moulton, Daniil Svyatskiy, and Cathy J. Wilson. 2016. "Integrated Surface/Subsurface Permafrost Thermal Hydrology: Model Formulation and Proof-of-Concept Simulations."

Water Resources Research 52 (8): 6062–77. https://doi.org/https://doi.org/10.1002/2015WR018427.

Pastorello, Gilberto, Carlo Trotta, Eleonora Canfora, Housen Chu, Danielle Christianson, You-Wei Cheah, Cristina Poindexter, et al. 2020. "The FLUXNET2015 Dataset and the ONEFlux Processing Pipeline for Eddy Covariance Data." *Scientific Data* 7 (1): 225. https://doi.org/10.1038/s41597-020-0534-3.

Pouchard, Line, Carlos Soto, Marcia Branstetter, and Giri Prakash. 2021. "FAIR Data Infrastructure and Tools for AI-Assisted Streamflow Prediction," April. https://doi.org/10.2172/1769710.

PRISM Climate Group. 2019. "Parameter-Elevation Regression on Independent Slopes Model (PRISM) Climate Data." *Oregon State University*. https://prism.oregonstate.edu.

Psaros, Apostolos F., Xuhui Meng, Zongren Zou, Ling Guo, and George Em Karniadakis. 2022. "Uncertainty Quantification in Scientific Machine Learning: Methods, Metrics, and Comparisons," January. https://doi.org/10.48550/arxiv.2201.07766.

Pulido Fentanes, Jaime, Amir Badiee, Tom Duckett, Jonathan Evans, Simon Pearson, and Grzegorz Cielniak. 2020. "Kriging-Based Robotic Exploration for Soil Moisture Mapping Using a Cosmic-Ray Sensor." *Journal of Field Robotics* 37 (1): 122–36. https://doi.org/10.1002/ROB.21914.

Rahmani, Farshid, Kathryn Lawson, Wenyu Ouyang, Alison Appling, Samantha Oliver, and Chaopeng Shen. 2020. "Exploring the Exceptional Performance of a Deep Learning Stream Temperature Model and the Value of Streamflow Data." *Environmental Research Letters*. https://doi.org/10.1088/1748-9326/abd501.

Rahmani, Farshid, Chaopeng Shen, Samantha Oliver, Kathryn Lawson, and Alison Appling. 2021. "Deep Learning Approaches for Improving Prediction of Daily Stream Temperature in Data-Scarce, Unmonitored, and Dammed Basins." *Hydrological Processes* 35 (11): e14400. https://doi.org/https://doi.org/10.1002/hyp.14400.

Raissi, M., P. Perdikaris, and G. E. Karniadakis. 2019. "Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations." *Journal of Computational Physics* 378: 686–707. https://doi.org/https://doi.org/10.1016/j.jcp.2018.10.045.

Rao, Vishwas, Sandeep Madireddy, Carlo Graziani, Pengfei Xue, and Romit Maulik. 2021. "Probabilistic Machine Learning and Data Assimilation," April. https://doi.org/10.2172/1769766.

Read, Jordan S., Xiaowei Jia, Jared Willard, Alison P. Appling, Jacob A. Zwart, Samantha K. Oliver, Anuj Karpatne, et al. 2019. "Process-Guided Deep Learning Predictions of Lake Water Temperature." *Water Resources Research* 55 (11): 9173–90. https://doi.org/https://doi.org/10.1029/2019WR024922.

Ribeiro, Marco, Sameer Singh, and Carlos Guestrin. 2016. *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*. https://doi.org/10.18653/v1/N16-3020.

Risser, Mark, Alan Rhoades, and Ankur Mahesh. 2021. "Machine Learning to Generate Gridded Extreme Precipitation Data Sets for Global Land Areas with Limited in Situ Measurements," April. https://doi.org/10.2172/1769784.

Runge, Jakob, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, et al. 2019. "Inferring Causation from Time Series in Earth System Sciences." *Nature Communications* 10 (1): 2553. https://doi.org/10.1038/s41467-019-10105-3.

Sadeghi, Mojtaba, Ata Akbari Asanjan, Mohammad Faridzad, Phu Nguyen, Kuolin Hsu, Soroosh Sorooshian, and Dan Braithwaite. 2019. "PERSIANN-CNN: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks–Convolutional Neural Networks." *Journal of Hydrometeorology* 20 (12): 2273–89. https://doi.org/10.1175/JHM-D-19-0110.1.

Sahu, Reetik Kumar, Juliane Müller, Jangho Park, Charuleka Varadharajan, Bhavna Arora, Boris Faybishenko, and Deborah Agarwal. 2020. "Impact of Input Feature Selection on Groundwater Level Prediction From a Multi-Layer Perceptron Neural Network." *Frontiers in Water* 2. https://doi.org/https://doi.org/10.3389/frwa.2020.573034.

Shen, Chaopeng. 2018. "A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists." *Water Resources Research* 54 (11): 8558–93. https://doi.org/10.1029/2018WR022643.

Shen, Chaopeng, and Kathryn Lawson. 2021. "Applications of Deep Learning in Hydrology." In *Deep Learning for the Earth Sciences*, 283–97. John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.1002/9781119646181.ch19.

Shi, Wenzhong, Min Zhang, Rui Zhang, Shanxiong Chen, and Zhao Zhan. 2020. "Change Detection Based on Artificial Intelligence: State-of-the-Art and Challenges." *Remote Sensing* 12 (10). https://doi.org/10.3390/rs12101688.

Sivapalan, Murugesu. 2003. "Prediction in Ungauged Basins: A Grand Challenge for Theoretical Hydrology." *Hydrological Processes* 17 (15): 3163–70. https://doi.org/https://doi.org/10.1002/hyp.5155.

Sivapalan, Murugesu, Hubert H. G. Savenije, and Günter Blöschl. 2012. "Socio-Hydrology: A New Science of People and Water." *Hydrological Processes* 26 (8): 1270–76. https://doi.org/https://doi.org/10.1002/hyp.8426.

Song, Keunyea, Austin Brewer, Shawyan Ahmadian, Ajay Shankar, Carrick Detweiler, and Amy J. Burgin. 2017. "Using Unmanned Aerial Vehicles to Sample Aquatic Ecosystems." *Limnology and Oceanography: Methods* 15 (12): 1021–30. https://doi.org/https://doi.org/10.1002/lom3.10222.

Sprenger, Matthias, Sheila Saia, P. Dennedy-Frank, and Erica Woodburn. 2021. "Preferential Flow in Subsurface Hydrology: From a Century of Denial to a Decade of Addressing It via ML?," April. https://doi.org/10.2172/1769765.

Steefel, Carl, Dipankar Dwivedi, Guillen Sole-Mari, Zexuan Xu, Ilhan Ozgen, Allan Leal, and Utkarsh Mital. 2021. "On Demand Machine Learning for Multi-Fidelity Biogeochemistry in

River Basins Impacted by Climate Extremes." United States. https://doi.org/10.2172/1769757.
Steefel, Carl I. 2019. "Reactive Transport at the Crossroads." *Reviews in Mineralogy and Geochemistry* 85 (1): 1–26. https://doi.org/10.2138/rmg.2019.85.1.
Stewart, Emma M., Philip Top, Michael Chertkov, Deepjyoti Deka, Scott Backhaus, Andrey Lokhov, Ciaran Roberts, et al. 2017. "Integrated Multi-Scale Data Analytics and Machine Learning for the Distribution Grid." In *2017 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 423–29. https://doi.org/10.1109/SmartGridComm.2017.8340693.
Steyaert, Jennie C., Laura E. Condon, Sean W. D. Turner, and Nathalie Voisin. 2022. "ResOpsUS, a Dataset of Historical Reservoir Operations in the Contiguous United States." *Scientific Data 2022 9:1* 9 (1): 1–8. https://doi.org/10.1038/s41597-022-01134-7.
Štrumbelj, Erik, and Igor Kononenko. 2014. "Explaining Prediction Models and Individual Predictions with Feature Contributions." *Knowledge and Information Systems* 41 (3): 647–65. https://doi.org/10.1007/s10115-013-0679-x.
Sun, Alexander Y., and Bridget R. Scanlon. 2019. "How Can Big Data and Machine Learning Benefit Environment and Water Management: A Survey of Methods, Applications, and Future Directions." *Environmental Research Letters* 14 (7): 073001. https://doi.org/10.1088/1748-9326/ab1b7d.
Sun, Alexander Y., and Guoqiang Tang. 2020. "Downscaling Satellite and Reanalysis Precipitation Products Using Attention-Based Deep Convolutional Neural Nets." *Frontiers in Water* 2 (November): 56. https://doi.org/10.3389/frwa.2020.536743.
Tartakovsky, A. M., C. Ortiz Marrero, Paris Perdikaris, G. D. Tartakovsky, and D. Barajas-Solano. 2020. "Physics-Informed Deep Neural Networks for Learning Parameters and Constitutive Relationships in Subsurface Flow Problems." *Water Resources Research* 56 (5): e2019WR026731. https://doi.org/https://doi.org/10.1029/2019WR026731.
Thornton, Peter E., Rupesh Shrestha, Michele Thornton, Shih-Chieh Kao, Yaxing Wei, and Bruce E. Wilson. 2021. "Gridded Daily Weather Data for North America with Comprehensive Uncertainty Quantification." *Scientific Data* 8 (1): 190. https://doi.org/10.1038/s41597-021-00973-0.
Tsai, Wen-Ping, Kuai Fang, Xinye Ji, Kathryn Lawson, and Chaopeng Shen. 2020. "Revealing Causal Controls of Storage-Streamflow Relationships With a Data-Centric Bayesian Framework Combining Machine Learning and Process-Based Modeling." *Frontiers in Water* 2 (November): 40. https://doi.org/10.3389/frwa.2020.583000.
U.S. Department of Energy Office of Science. 2018. "Climate and Environmental Sciences Division Strategic Plan 2018–2023." https://science.osti.gov/-/media/ber/pdf/workshop-reports/2018_CESD_Strategic_Plan.pdf?la=en&hash=0823BBAAF5C13D42AC1F193257BEC66300A7E068.
Varadharajan, Charuleka, Valerie C. Hendrix, Danielle S. Christianson, Madison Burrus, Catherine Wong, Susan S. Hubbard, and Deborah A. Agarwal. 2022. "BASIN-3D: A

Brokering Framework to Integrate Diverse Environmental Data." *Computers & Geosciences* 159 (February): 105024. https://doi.org/10.1016/J.CAGEO.2021.105024.

Varadharajan, Charuleka, Yuxin Wu, Andrew Wiedlea, Kolby Jardine, Haruko Wainwright, Robert Crystal-Ornelas, Joan Damerow, et al. 2021. "Observational Capabilities to Capture Water Cycle Event Dynamics and Impacts in the Age of AI." United States. https://doi.org/10.2172/1769755.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Ł. Lukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need." In *Advances in Neural Information Processing Systems*, edited by I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Wainwright, Haruko M., Sebastian Uhlemann, Maya Franklin, Nicola Falco, Nicholas J. Bouskill, Michelle E. Newcomer, Baptiste Dafflon, et al. 2022. "Watershed Zonation through Hillslope Clustering for Tractably Quantifying Above- and below-Ground Watershed Heterogeneity and Functions." *Hydrology and Earth System Sciences* 26 (2): 429–44. https://doi.org/10.5194/hess-26-429-2022.

Wang, Yaqing, Quanming Yao, James T. Kwok, and Lionel M. Ni. 2020. "Generalizing from a Few Examples." *ACM Computing Surveys (CSUR)* 53 (3). https://doi.org/10.1145/3386252.

Weierbach, Helen, Aranildo R. Lima, Jared D. Willard, Valerie C. Hendrix, Danielle S. Christianson, Michaelle Lubich, and Charuleka Varadharajan. 2022. "Stream Temperature Predictions for River Basin Management in the Pacific Northwest and Mid-Atlantic Regions Using Machine Learning." *Water* 14 (7): 1032. https://doi.org/10.3390/w14071032.

Willard, Jared D., Jordan S. Read, Alison P. Appling, Samantha K. Oliver, Xiaowei Jia, and Vipin Kumar. 2020. "Predicting Water Temperature Dynamics of Unmonitored Lakes with Meta Transfer Learning," November. https://doi.org/10.1029/2021WR029579.

Willard, Jared, Xiaowei Jia, Shaoming Xu, Michael Steinbach, and Vipin Kumar. 2020. "Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems," March. https://doi.org/10.48550/arxiv.2003.04919.

Wu, Jia, Weiru Zeng, and Fei Yan. 2018. "Hierarchical Temporal Memory Method for Time-Series-Based Anomaly Detection." *Neurocomputing* 273 (January): 535–46. https://doi.org/10.1016/J.NEUCOM.2017.08.026.

Xia, Youlong, Kenneth Mitchell, Michael Ek, Justin Sheffield, Brian Cosgrove, Eric Wood, Lifeng Luo, et al. 2012. "Continental-Scale Water and Energy Flux Analysis and Validation for the North American Land Data Assimilation System Project Phase 2 (NLDAS-2): 1. Intercomparison and Application of Model Products." *Journal of Geophysical Research: Atmospheres* 117 (D3): 3109. https://doi.org/10.1029/2011JD016048.

Xie, Kang, Pan Liu, Jianyun Zhang, Dongyang Han, Guoqing Wang, and Chaopeng Shen. 2021. "Physics-Guided Deep Learning for Rainfall-Runoff Modeling by Considering Extreme

Events and Monotonic Relationships." *Journal of Hydrology* 603 (December): 127043. https://doi.org/10.1016/J.JHYDROL.2021.127043.

Xu, Tianfang, and Feng Liang. 2021. "Machine Learning for Hydrologic Sciences: An Introductory Overview." *WIREs Water* 8 (5): e1533. https://doi.org/https://doi.org/10.1002/wat2.1533.

Zeng, Xubin, Robert Atlas, Ronald J. Birk, Frederick H. Carr, Matthew J. Carrier, Lidia Cucurull, William H. Hooke, et al. 2020. "Use of Observing System Simulation Experiments in the United States." *Bulletin of the American Meteorological Society* 101 (8): E1427–38. https://doi.org/10.1175/BAMS-D-19-0155.1.

Zhao, Qun, Yuelong Zhu, Kai Shu, Dingsheng Wan, Yufeng Yu, Xudong Zhou, and Huan Liu. 2020. "Joint Spatial and Temporal Modeling for Hydrological Prediction." *IEEE Access* 8: 78492–503. https://doi.org/10.1109/ACCESS.2020.2990181.

Zwart, Jacob, Samantha Kay Oliver, David Watkins, Jeffrey Sadler, Alison Appling, Hayley Corson-Dosch, Xiaowei Jia, Vipin Kumar, and Jordan Read. 2021. "Near-Term Forecasts of Stream Temperature Using Process-Guided Deep Learning and Data Assimilation." August 13, 2021. https://doi.org/https://doi.org/10.31223/X55K7G.

4 Watershed Science

Authors: Mavrik Zavarin (LLNL), Carl Steefel (LBNL), David Moulton (LANL), Xingyuan Chen (PNNL), Scott Painter (ORNL), Dipankar Dwivedi (LBNL), Li Li (Penn State)

4.1 Introduction

Watershed science is the interdisciplinary study of the natural processes and human activities that affect freshwater resources. It necessitates the integration of hydrology with geochemistry, plant physiology, microbial community dynamics, human system dynamics, and other processes to predict emergent behavior of a complex system. Watersheds are a basic unit of Earth's land surface system and link human activities (consumption, agriculture, energy production, transportation, and recreation) to ecosystem health and resilience and climate-impacted extreme events (e.g., fires, precipitation, extreme weather). The management of freshwater resources and prediction of future watershed function are increasingly important and complex challenges in the United States and worldwide.

4.2 Grand Challenges

Watershed science is, by nature, an interdisciplinary research area. Conducting a full treatment of water flow and biogeochemical reactive transport at watershed to river basin scales is not currently possible, and yet there is a critical need to provide improved estimates of fluxes of nutrients and contaminants for both scientific and water management purposes at these scales (Figure 4-1). Our capacity to predict watershed system behavior is challenged by the evolving nature of extreme events (fire, precipitation, etc.) that are outside the bounds of historical conditions, by system and process complexity, and by the computational demands. Simulations of watershed processes are additionally challenged by both sparsity (insufficient spatial density) and scarcity (insufficient and infrequent observations) of data. This is particularly the case in the subsurface domain. It is apparent that growth in computing power will not be sufficient to account for the scale of complexity and the spatial and temporal modeling fidelity that are needed. AI is poised to bridge the gap between process complexity, model fidelity, and computing power to allow for the effective prediction of watershed function within the context of large Earth system modeling. We envision that AI will play a critical role in (1) co-design of data collection and modeling, (2) process scaling, (3) hybrid-ML approaches, (4) trait-based and surrogate ecosystem ecology modeling, (5) extreme event modeling, and (6) incorporation of human and engineered systems into watershed models. However, significant challenges to the application of AI in watershed systems exist.

4.2.1 Co-designing Data Collection and Assimilation for Soil and Watershed System Applications (Sensors, Complex Data, and Sparse Data)

The iteration between model and experiments (ModEx) is essential to improve the predictability of watershed models under both baseline and perturbed conditions (Chen et al. 2021). The increase in model complexity and data volume has led to a substantial increase in computational cost and an exponential increase in data dimensionality that have both hampered ModEx. However, observational data for extreme events are scarce due to their rare occurrence and the practical challenges in collecting data under those conditions. Targeted data collection assisted by modeling can increase the information content of data for model improvement. We need a systematic way to integrate data and modeling with varying complexity and evaluate the models' performance against each other and against observational data to identify gaps. We are facing several challenges to boost the adoption of AI/ML methods (Maskey et al. 2020): (1) incorporating physics in ML models; (2) improving the interpretability of ML models; (3) enabling reliable extrapolations beyond the training conditions; (4) quantifying and propagating uncertainty in model results; and (5) developing publicly available benchmark training datasets that can be used to aid and test new ML methods.

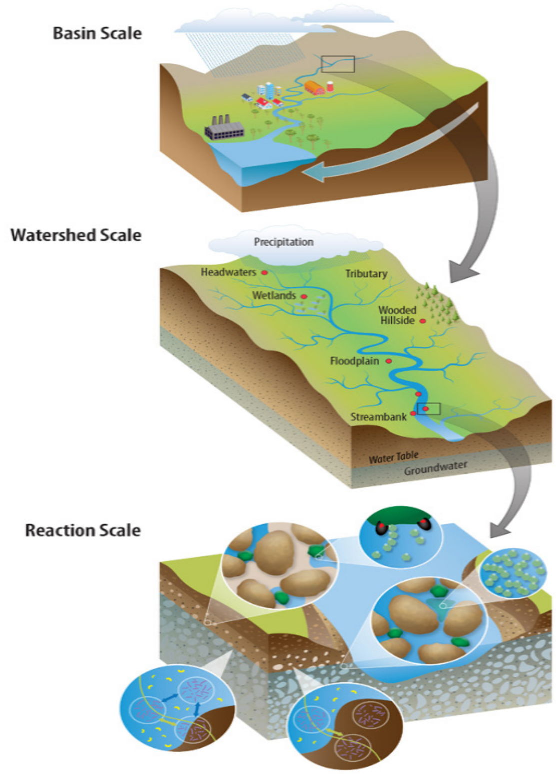


Figure 4-1. Watershed processes span a range of spatial scales while also integrating across these scales (Source: U.S. Department of Energy 2019).

4.2.2 Strategies of Scaling (Multifidelity ML, Surrogate Models)

Watershed modelers have traditionally relied on lumped-parameter or semi-distributed spatial representations to make models tractable at watershed or river-basin scales. When properly calibrated, that approach has been valuable for quantifying watershed-scale hydrology and biogeochemical processes. However, calibrated, semi-distributed models have a large empirical component and significantly underestimate the diversity of flowpaths and the potential for storm-driven changes in those flowpaths, which introduces large uncertainties when used to evaluate the impacts of changing climate or land use. Focusing on denitrification as an example, Helton et al. (2010) point out the limitations of watershed modeling approaches and conclude those limitations "restrict our ability to simulate biogeochemical dynamics among diverse river networks."

Modern software and high-performance computing have resulted in powerful physics-based integrated models (e.g., ATS) that couple surface water, shallow groundwater, and land surface processes. Those hydrology-based models have recently been extended to form integrated surface/subsurface reactive transport models (Molins et al. 2019; Painter 2021). This emerging class of models has great potential for increasing our understanding of watershed response to changes in external forcings. However, applications at basin scales have been thwarted by the large computational demand. ML can potentially accelerate those models and make them more tractable at basin scales by, for example, providing accurate surrogate models for component subsystems or more efficient approaches for inverse modeling and uncertainty quantification.

4.2.3 Hybrid AI Models (e.g., On-demand ML) for Biogeochemical Reactive Transport Simulation and Scaling

Conducting a full treatment of water flow and biogeochemical reactive transport at watershed to river basin scales is not currently possible, and yet there is a critical need to provide improved estimates of fluxes of nutrients and contaminants for both scientific and water management purposes at these scales. The use of hybrid, multifidelity predictive models that take advantage of ML techniques offers an attractive option to overcome the obstacles associated with computational expense, especially insofar as it is possible to maintain process fidelity for heterogeneously distributed biogeochemical processes interacting with the hydrological cycle—a situation that is expected to be particularly pronounced during climate extremes. Explainable/interpretable AI is key to its integration with physics-based watershed and Earth system predictive models (i.e., application of AI from the realm of interpolation to extrapolation). We must create new hybrid ML algorithms that identify processes (both known and unknown) and associations to discover emergent properties that help predict system behavior.

4.2.4 Watershed Microbial Dynamics and Ecology

Integration of microbial ecology at watershed scales: Integrating mechanistic microbiological processes in biogeochemical models is not scalable at this moment. While trait-based approaches are being developed (e.g., MicroTrait, BioCrunch), it is as yet unclear how these approaches can be scaled effectively. Surrogate models, which can be aided by AI, will likely need to be developed. Surrogate models can be used in large-scale models (watershed, basin, Earth) to achieve computational tractability. These models can, for example, be developed from mechanistic metabolic genome-informed biogeochemical reaction network models (e.g., trait-based models).

Data access and transferability: The data needed for mechanistic/trait-based models (direct and indirect data) are becoming increasingly available across diverse watershed settings (e.g., those from WHONDRS and the GROW database). However, methodologies to harness knowledge from these data are challenging because of issues related to transferability (space for time, etc.), extrapolation vs. interpolation, and inference, among others. AI can facilitate the application of community datasets to address the sparse data issue. This will require development of transferability models that can predict microbial potential function and expressed function through space and time, across watershed "components" distributed across global watersheds, and distributed across both "normal" conditions and extreme events. These kinds of predictions can be used to feed the surrogate models for large-scale application.

Optimization of sparse sample collection: Microbial community composition and activity data are intrinsically expensive and are naturally sparse at the watershed scale. Microbial functional data are accessed through relatively expensive and low-throughput approaches that require physical sample collection and intensive and costly laboratory analysis (i.e., metagenomics, metatranscriptomics, metaproteomics). AI can be implemented to guide optimal sample collection (in space and time) and optimal sequencing/analysis depth of each sample, to more efficiently generate data with maximum information content. For example, AI approaches may help improve the translation of "cheap" microbial surveys to infer the more expensive functional information (e.g., metabolic potential).

4.2.5 AI Application to Extreme and Future Effects and Events (Fires, Precipitation, Flooding, and Compounding)

Watersheds are becoming more susceptible to climatic and anthropogenic disturbances, as well as extreme events such as droughts, floods, extreme weather, wildfires, and land-use changes. Predicting extreme events and their future compounding effects through transformative science solutions is critical to understanding watershed function and managing freshwater resources. The increasing frequency and intensity of extreme precipitation events are anticipated to cause numerous problems beyond the storm and flooding, for example, failures to provide clean drinking water and excessive nutrient export from agriculture. Although AI is poised to dramatically impact watershed science, several challenges must be addressed in the application of AI to extreme and future effects. Data sparsity and obtaining relevant data, particularly with respect to extreme events, have remained the largest of these challenges.

4.2.6 Impacts and Feedback of Human/Engineered Systems on Watersheds (e.g., Reservoir Management, Urban, Agricultural)

Human impacts and engineering in watersheds (e.g., agriculture, dams, flood controls, land management) have significantly altered these systems to support services for human use, whether directly for drinking water and recreation, or indirectly for agriculture and power generation. However, climate change is challenging the assumptions made in the design of these managed systems, from low-to-no snow futures in the west to increasing frequency and intensity of extreme precipitation events. The impact of these changes will go well beyond the flooding observed already and will include failures to provide clean drinking water, contaminant releases from once-stable legacy waste sites, and excessive nutrient export from agriculture. Unfortunately, data covering this range of human-impacted watersheds are sparse and fragmented because water management-related data (irrigation, reservoirs, drinking water, storm drains) are often not publicly available and are controlled by different local or regional agencies. Addressing this data shortage, along with developing a typology of human-impacted watersheds and their function, are grand challenges that underpin new AI approaches to building understanding, making predictions over a range of scales, and exploring adaptation of engineered watersheds.

4.3 State-of-the-Science

4.3.1 Co-designing Data Collection and Assimilation for Soil and Watershed System Applications (Sensors, Complex Data, and Sparse Data)

Neural networks have a long history of applications in watershed science, ranging from streamflow forecast to groundwater and water quality modeling. Increasing the depths of the network, that is, the number of hidden layers, could improve its ability to represent more complex system behaviors (Shen 2018; Raghu et al. 2017), especially for mapping highly nonlinear relationships between the model inputs and outputs. More recently, machine learning methods are used to replace computationally expensive forward models in uncertainty quantification and model parameter estimation via surrogates and other reduced-order representation. For example, Mo et al. (2019) successfully employed a deep autoregressive neural network-based surrogate approach to estimate the heterogeneous aquifer permeability as well as groundwater contaminant sources with high accuracy and computational efficiency. Canchumuni, Emerick, and Pacheco (2019) found that a convolutional variational autoencoder outperformed the standard ensemble data assimilation methods in reconstructing the spatial distribution of geologic facies. Cromwell et al. (2021) successfully used deep learning to map the nonlinear relationship between permeability and stream discharge, presenting new opportunities for improving the subsurface characterization of large-scale watersheds. It also paves the way to help develop more generalized watershed model calibration strategies for complex systems that involve multiple parameters and multiple types of observation data.

To facilitate co-design strategies, national and international efforts are emerging to build standardized databases and associated field sites (e.g., ESS-DIVE, NEON, LTER, CZNet), in parallel to advancements in sensing and monitoring technologies. High-resolution satellite and remote-sensing data products with global coverage are becoming increasingly available, ranging from surface properties (e.g., land use/land cover types, elevation), to states (e.g., leaf area index, biomass, surface soil moisture), and to energy, water, and carbon fluxes (e.g., evapotranspiration and gross primary production).

Iterative ModEx loops are being implemented. For example, Chen et al. (2021) describe an iterative model-experiment process in the PNNL River Corridor Science Focus Area (SFA). In this example, a series of field- and lab-based studies demonstrated that thermodynamic properties of organic matter exert significant control on biogeochemical processes in river corridors. This experimental finding motivated development of a new theoretical modeling framework published by Song et al. (2020) that defined a new model parameter that quantifies the impacts of thermodynamics on biogeochemical reactions. Additional field sampling was then conducted through the WHONDRS consortium to identify spatial and temporal patterns across diverse systems, and the resulting data were used to simulate intersystem biogeochemical process variability using the KBase modeling platform (e.g., Napieralski and Roden 2020; Klasek et al. 2021). The ModEx cycle is ongoing in this example and extends into the AI/ML space, as ML analyses of the WHONDRS data/metadata are currently being used to plan the next WHONDRS sampling campaign, the outcomes of which will in turn be used to refine watershed model parameters.

Better ways to collect data are needed. Presently, modeling has not been used to inform data collection on a regular basis. Techno-economic analysis can be conducted before deploying sensors to identify the optimum location of sensors and value of information collected. More and better low-cost sensors are required for AI-informed data collection and targeted field design. Sensors that are flexible (movable and controllable in space and time) and linked to models in near-real time are needed, as well as the ability to decide sampling sites/times using AI/ML and ingest data into models using AI/ML in real time. Combining data from different scales (both space and time) is necessary to connect from the point scale to the continental scale and enable causal inference to identify cross-scale linkages.

4.3.2 Strategies of Scaling (Multifidelity ML, Surrogate Models)

Opportunities to incorporate ML into process-based watershed or water quality models have been noted (e.g., Fu et al. 2020), but those opportunities are yet to be realized. ML has been used to address scaling challenges in watershed hydrology (see Hydrology, chapter 3) and those approaches are generally applicable to watershed reactive transport modeling. Approaches include using ML as an alternative to traditional inverse modeling (Cromwell et al. 2021), for surrogate modeling in support of uncertainty quantification (see Surrogate Models and Emulators, chapter 9), or to construct data-driven representations of component subsystems. While multifidelity models (e.g., Meng and Karniadakis 2020; Peherstorfer, Willcox, and Gunzburger 2018) that use ML to relate results from fast-running but lower-fidelity models with more expensive, higher-fidelity models could be promising to accelerate watershed modeling (e.g., Fu et al. 2020), but only initial steps have been taken in that direction in the watershed modeling context (e.g., Mao et al. 2021).

4.3.3 Hybrid AI Models (e.g., On-demand ML) for Biogeochemical Reactive Transport Simulation and Scaling

Much of our understanding of ecosystem function, including the integrated water and biogeochemical cycles, stems from high-fidelity, physics-based models. Although these high-fidelity models can provide a detailed simulation of how the future climate extremes will impact biogeochemistry in river basins, they are computationally expensive and extremely time-consuming, making them unsuitable for the multitudinous runs needed to evaluate the complex interactions of processes ranging from the bedrock to the canopy (Leal et al. 2020). The use of surrogates or emulators within the on-demand machine learning, multifidelity framework can significantly limit the prohibitive computational costs of high-fidelity simulations while capturing the dynamics of underlying processes. The overarching benefit of surrogates is their ability to reduce complexity by learning the state variables' dynamics directly from the observational data or full output of a custom-built model (i.e., synthetic high-resolution data). As a whole, these data-informed or custom-built, model-informed surrogates can easily be developed on the fly using ML techniques, such as Gaussian process regression, dynamic mode decomposition, random forest, and neural networks (e.g., Lu and Tartakovsky 2021), and be seamlessly integrated with the larger multifidelity framework. The computational costs associated with modeling chemical processes in reactive transport simulations can substantially compromise performance, such as when rich chemical descriptions of fluids and rocks are considered. In this case, the on-demand ML strategy presented in Leal et al. (2020) and Kyas et al. (2022) can speed up the sheer number of chemical equilibrium and kinetics calculations by one to three orders of magnitude and provide significant overall speed up.

4.3.4 Watershed Microbial Dynamics and Ecology

Integrating microbial community function, generally through a trait-based approach at the watershed scale, is still nascent. Notable codes that are attempting to implement trait-based microbial community function into reactive transport models include microtrait and biocrunch. However, these models are likely to remain untenable at the watershed or larger scales. Surrogate model development will be needed to scale processes.

As data scarcity and sparsity with regard to sequencing and omics data will remain an issue into the future, indirect (and more cost-effective) measurements (e.g., satellite and sensor data) will likely need to be harnessed to bridge the gap between model calibration requirements and data availability. Indirect measurement of subsurface microbial function is a promising development (e.g., Kim et al. 2021). Causal inference, the need to establish that relationships are asymmetric in time, sets a certain requirement of data frequency—this interacts with the variance of a system process/component.

4.3.5 AI Application to Extreme and Future Effects and Events (Fires, Precipitation, Flooding, and Compounding)

ML methods are versatile and can handle a range of watershed-related research questions, including extreme events to some extent. For example, several studies have utilized machine learning for post-fire applications (Saxe, Hogue, and Hay 2018; Schmidt and Eloy 2020) and have estimated several critical watershed characteristics, including soil burn severity following wildfires in California (e.g., Wilder et al. 2021). However, deep learning models for extreme events have been underutilized because of the unavailability of adequate extreme events data needed to train such models. A new paradigm is needed to overcome the dearth of data and obtain relevant data utilizing deep learning methods.

4.3.6 Impacts and Feedbacks of Human/Engineered Systems on Watersheds (e.g., Reservoir Management, Urban, Agricultural)

Exploration of AI/ML techniques in water management and many aspects of human-impacted watersheds is widespread and growing, with more than 7,000 papers published in this area in 2019 (Allen-Dumas et al. 2021). The majority of this work has focused on data-driven methods, with each study focused on a single facet of an overarching challenge. For example, Sapitang et al. (2020) examined ML algorithms for reservoir water level forecasting for managing hydropower generation. Kim and Han (2020) evaluated the urban flood prediction skill of a deep neural network with data augmentation. In the first of these studies, water level and dam operations data were obtained directly from the Kenyir Operation Unit, while in the second study, simulation output was used to overcome the lack of urban runoff data. These papers highlight the challenges we face in data access and sparsity. Furthermore, although this broad but isolated approach is a natural starting point for the exploration and integration of these new methods and workflows, it does not begin to address the integrated challenge of water security in human and engineered systems (Allen-Dumas et al. 2021). Additionally, it does not address the challenge of a changing climate where historical data may not reflect future conditions. As this field matures, we anticipate not only the integration of more comprehensive datasets, but also the integration of process-based models through physics-informed or hybrid ML techniques to help mitigate the sparsity of the data, as well as to support predictions under future climate scenarios.

Ultimately, this integration will help support a more holistic approach to guide management and adaptation of an increasingly comprehensive view of the watershed and its services, spanning, for example, reservoir management, agricultural best practices, and urban drinking water, along with overall drought and flood risk.

4.4 Experimental, Data, and Modeling Opportunities

4.4.1 Co-designing Data Collection and Assimilation for Soil and Watershed System Applications (Sensors, Complex Data, and Sparse Data)

There is great opportunity (for domain researchers and knowledge engineers and data scientist) in exploring how to codevelop ontologies and semantic tools to both annotate data on repositories and inform AI/ML (via traditional knowledge graphs or otherwise). AI techniques can be used to automate data curation (e.g., through apps or devices that can automatically collect required metadata or help with sensor maintenance) and optimize data collection. AI can be used to integrate, unify, and harmonize a hierarchy of observational datasets that range from high to low frequency (e.g., hourly, daily, one-time measurement, etc.), covering various timescales (e.g., once, seasonal, annual, decadal) across different regions. For example, using AI to map multiple irregular sensor outputs onto regular grids can significantly lower the barrier for researchers to use such information in improving watershed models through model parameterization, calibration, or benchmarking. In addition, because a lot of historic data exist in non-digital forms, there is potential opportunity in using natural language processing (NLP) methods to extract such data from historic records and making them accessible to watershed system science. Such efforts are nearly impossible to be undertaken by researchers without machine learning.

AI can inform sensors for better data collection. Self-correcting sensors or intelligent sampling are possible on the edge when AI algorithms are deployed on the edge devices to identify problematic data points or find the most valuable data to improve models. Such adaptive control will also enable intelligent data collection triggered by certain events (e.g., extreme event or disturbances) to address data sparsity under conditions, leveraging recent advances in AI@Edge, low footprint sensors, and data acquisition (DAQ) systems.

The greatest opportunity is in building a community computational platform to allow the sharing of ML-assisted ModEx pipelines, with easy access to pre-trained ML models (e.g., similar to Model Zoo, https://modelzoo.co/), standardized application-ready datasets, interoperable process-based models, and supercomputing and/or cloud computing resources.

4.4.2 Strategies of Scaling (Multifidelity ML, Surrogate Models)

Several opportunities to accelerate large-scale hydrobiogeochemical models can be identified. Replacing expensive process-based models with faster-running ML-based surrogate models is a well-established strategy for model calibration that can be adapted to scale process-based models to large scales (Fu et al. 2020). One approach is to not repeat calculations for subdomains with similar climate forcings, biogeochemical inputs, and physiographic properties (see, e.g., Hydrology, chapter 3). In the short term, this strategy could be implemented by first building surrogate models from simulations on smaller domains and then combining them in larger-scale simulations. Longer term, it is envisioned that this strategy could be implemented dynamically, using AI/ML to replace the subdomains of a fully resolved simulation with surrogate models on-the-fly, switching back to fully resolved simulations if needed (e.g., Steefel et al. 2021; Painter, Coon, and Lu 2021). A closely related approach would be to combine fast-running but more approximate models with computationally demanding full-physics simulations. In this multifidelity approach, high-fidelity models would be performed only for selected catchments within the basin of interest while fast-running intermediate-complexity models would be used for all catchments. ML would then be used to develop mappings that relate high-fidelity model output to model inputs and intermediate-fidelity output. Once those mappings are learned, the intermediate-fidelity model would then be used in projections and model calibrations.

4.4.3 Hybrid AI Models (e.g., On-demand ML) for Biogeochemical Reactive Transport Simulation and Scaling

Considering biogeochemical processes in reactive transport simulations is computationally expensive. By using an on-demand machine learning (ODML) algorithm (Leal et al. 2020), however, the computing costs for biogeochemistry and transport calculations can be reduced by orders of magnitude. The ODML model will start with zero knowledge at the beginning of the simulation. It will then gradually learn key biogeochemical calculations during the reactive transport simulation. These key calculations are then used as often as possible to predict similar calculations. The predictions are much faster because they do not require iterative algorithms, just a fast matrix-vector multiplication. In addition to ODML, other ML approaches can provide an uncertainty quantification (UQ)-based approach relying on selective comparison with observational data and high-resolution physics-based simulations to automatically choose the fidelity of a biogeochemical approach (e.g., high-resolution RTM versus 1D flowtube versus surrogate) to balance the demands of computational efficiency and process fidelity (Figure 4-2).

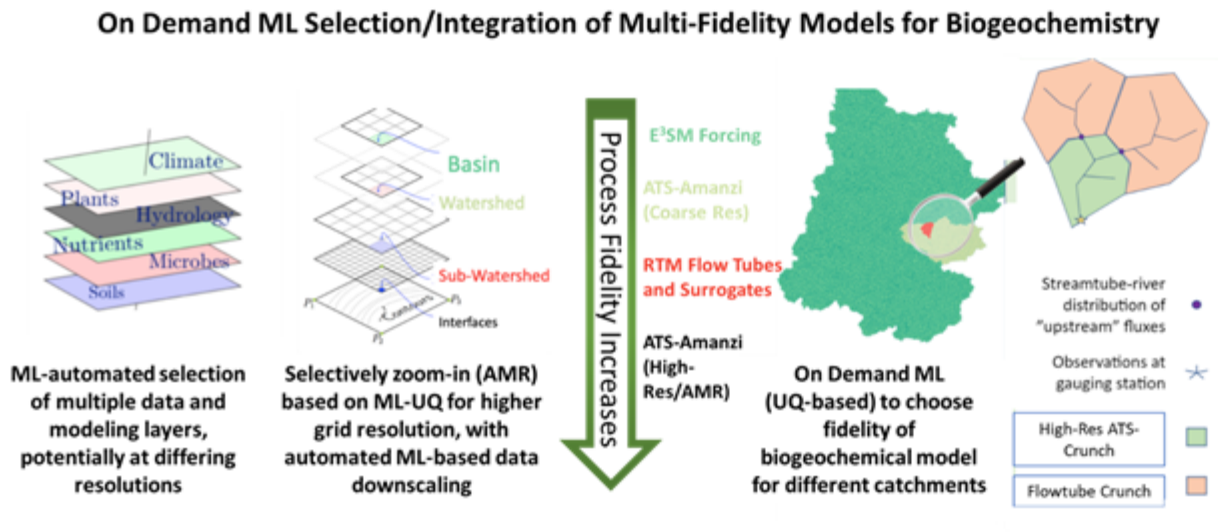


Figure 4-2. Watershed microbial dynamics and ecology (Source: Steefel et al., 2021).

Distributed data collection efforts are growing exponentially and show great promise to develop direct and indirect data for knowledge discovery, particularly with regard to integrating ecology at the watershed scale (e.g., WHONDRS, GROW, and EXCHANGE). Leveraging KBase efforts to link microbial 'omics with detailed chemistry (e.g., FTICR-MS, LC-MS, NMR) will open new avenues to explore microbial data transferability and surrogate model development. ICON principles are critical: as described in a recent BER workshop, they are to intentionally design research efforts as Integrated, Coordinated, Open, and Networked in order to build the interoperable data foundation needed to feed AI modeling efforts that are designed to learn new "physics" and predict microbial function across diverse watershed settings (Goldman 2021) (Figure 4-1).

A new generation of libraries focused on trait-based microbial function (Sokol et al. 2022) could provide a pathway for scaling and surrogate model development (Figure 4-2). A Unified Biogeochemical Reaction/Trait Database/Library conforming to community standards would provide the necessary data harmonization to harness AI for process discovery, surrogate model development, benchmarking, and robust uncertainty quantification (Sokol et al. 2022). Community standards would provide opportunities to harmonize complex field data with simplified and controlled laboratory microbial consortia experiments. Detailed laboratory microbiology experiments yield deep knowledge of genomic potential and rich data on expressed function and associated contextual chemistry and physical properties. Integrating laboratory functional data with field observations will help in the development of hybrid and surrogate modeling approaches at the watershed scale.

Due to the inherently high costs of sampling and processing of samples to obtain genomic and multi-omic data necessary to capture subsurface microbial ecology and its impacts on watershed function, development of methodologies to optimize costly data collection activities and data curation (adaptive sampling) is essential. AI can also contribute to the generation of direct and indirect (e.g., soil characterization) data through the application of natural language processing of decades of historical data and application of deep learning to causal inference and elucidating biogeochemical function and traits.

4.4.4 AI Application to Extreme and Future Effects and Events (Fires, Precipitation, Flooding, and Compounding)

To handle the dearth of data for extreme events, relying on physics-based hydrological models can be valuable. Especially because data will never be abundant for extreme events, the idea is to use physics-based models iteratively to create scenarios and generate data. However, physics-based models are computationally expensive. There is also a need to develop an efficient workflow that can run these models tractably. AI/ML can provide a viable option to enhance seamless integration of physics-based models with field observations to generate reliable data for deep learning models. Moreover, AI/ML methods can develop surrogates and emulators installed in the field for ModEx-guided data collection. Indeed, ModEx-guided data collection can be very efficient for collecting relevant extreme event data by informing how, when, and where to collect data. However, extreme events such as fires are not what we can plan, so rapid data collection and deployment capabilities are key. Integrated field laboratories, automated samplers, and on-site AI-driven decision systems for data collection are needed to collect relevant extreme event data.

4.4.5 Impacts and Feedback of Human/Engineered Systems on Watersheds (e.g., Reservoir Management, Urban, Agricultural)

The fragmentation in data access has naturally led to fragmentation in the community, with the scope of each application of data-driven AI/ML techniques constrained by the limited data that a researcher can use. To accelerate a more complete and holistic approach to watershed science, there is a need for a community of practice that can develop and promote best practices in a wide range of AI/ML techniques, help coordinate the creation of an open data portal that can act as a central location for researchers to access common datasets and share AI/ML workflows and results. Significant effort is needed here to collect and coordinate existing open datasets, potentially anonymize datasets that are unavailable due to privacy or proprietary concerns, create synthetic datasets, and explore and curate historical and ancestral best management practices. In addition, to address the challenge of watershed resilience under the impact of future climate and continued human development, there are modeling opportunities at all scales and across scales. For example, process-based models (e.g., ATS, ParFlow) can help fill some of the data gaps, particularly as we look at future climate scenarios for which we have no real data. In this case, some model development is needed to add engineered features and water management into these models, and better support workflows that integrate operational data. This development would help support advancing our current understanding of watershed function as well. Modeling across scales is needed as the domain of interest moves from local to regional or even to national. Here, AI/ML can help inform the transition in models from process-based (fully resolved) to semi-distributed and statistical and can help leverage new approaches to building surrogates or directly combining models of varying fidelity (see breakout on Strategies of Scaling) to enable more robust and transferable predictions of system function.

4.5 Research Priorities

4.5.1 Co-designing Data Collection and Assimilation for Soil and Watershed System Applications (Sensors, Complex Data, and Sparse Data)

Generating public benchmark training datasets (similar to ImageNet, http://www.image-net.org/) is the key to advancing applications of ML in Earth science domains (Maskey et al. 2020; Dramsch 2020). There is a unique opportunity to enhance the use of the new generation of remote sensing (RS) products that capture components of the water cycle (precipitation, snow, soil moisture, evapotranspiration, groundwater, and runoff), as well as coupled carbon and nutrient cycle components, with increasing spatial and temporal resolutions. Training data may also be generated from process-based models. Leveraging open-source resources from federal agencies is necessary for the success of such an extensive and expensive effort. For example, NASA's Earth Sciences Data Systems (ESDS) has generated high-quality training datasets that are open and easily accessible. NOAA, USGS, and other federal agencies have been maintaining extensive observation networks and are developing a large number of integrated Earth system models. Standardized data management practices would significantly increase data usability.

We need computational infrastructure to address long-standing challenges of complexity and heterogeneity in watershed models that would otherwise be overwhelmed by the tremendous complexity in managing software, hardware, workflows, and computational cost. Addressing these challenges requires developing and maintaining open-source scientific software and ML frameworks for deploying Earth science ML and process-based models. To achieve this goal, existing frameworks can be expanded or integrated through collaborative efforts for efficiency. DOE's Systems Biology Knowledgebase (KBase, https://www.kbase.us/) is a good example of such computational infrastructure, which is designed to meet the grand challenge of predicting and designing biological functions. In addition to facilitating data access/sharing and building reusable bioinformatic pipelines, KBase uses a Narrative (an interactive digital notebook) to capture workflows for various scientific discoveries, which can be shared with other researchers to enhance scientific reproducibility and adaptability to answer other questions. The use of a Jupyter Notebook-based narrative interface to encode workflows makes the computational framework much more accessible to the broader community. Another example is Pangeo (https://pangeo.io/), which is an open-source architecture that provides interconnected software packages and deployments of the software in cloud and high-performance computing environments for ocean, atmosphere, land, and climate science.

4.5.2 Strategies of Scaling (Multifidelity ML, Surrogate Models)

Priority activities for enabling scaling of process-based models to basin scales cluster into two general groups: development of the necessary software tools to implement various strategies, and real-world example applications to provide the knowledge base necessary for routine applications. Specific priorities include:

- Flexible and composable modeling systems that make it possible to easily switch out ML and process-based models.
- Modeling tools that accommodate watershed-based domain decompositions, thus allowing for switching between data-driven and process-based simulations of individual watersheds.
- Fully resolved process-based modeling results at sufficient scale to provide benchmark data for testing various efficient scaling approaches.

- Software tools to manage the complex workflows around multifidelity models and surrogate model construction, testing, and integration into hybrid models.
- Research into AI/ML methods to detect when a given subdomain can be represented by a surrogate and when it needs to be represented explicitly.
- Real-world tests of ML-based multifidelity models.
- Community model comparison projects to test and compare different approaches for using ML to accelerate large models.

4.5.3 Hybrid AI Models (e.g., On-demand ML) for Biogeochemical Reactive Transport Simulation and Scaling

The need for dramatically improved prediction of river basin–scale biogeochemical function is clear, but the computational challenges are daunting, But ML can play an important role in at least five ways: (1) ML can facilitate the inclusion of diverse big data in physics-based models for water and biogeochemistry through downscaling and upscaling approaches (Mital et al. 2020); (2) ML can achieve improved predictability by enabling calibration and validation of models for given river basins and watersheds (Cromwell et al. 2021); (3) ML can enable the development of surrogate and reduced order/dimension models that capture watershed and river basin function with reduced computational expense; (4) on the fly ML can be used for automation of uncertainty quantification to choose dynamically the level of fidelity and computational expense that is adequate for a given river basin-scale simulation; and (5) on demand ML can be used to gradually reduce the number of full predictive calculations that are needed to describe the watershed to river basin-scale biogeochemical function, essentially replacing full physics-based simulation with continuously improving surrogate models (Leal et al. 2020).

4.5.4 Watershed Microbial Dynamics and Ecology

Bridging genomic and metagenomics information into models at watersheds and Earth system scales is a key priority. Trait-based approaches appear to be the most promising mechanism to achieve this objective. However, the complexity and data limitations, particularly for the subsurface community diversity, spatial distribution, and function, will continue to be a challenge. Surrogate approaches must be explored to bridge scales, and knowledge-informed AI is poised to help scale genomic information to make genomically informed watershed function modeling tractable.

Key research that knowledge-informed AI may help to resolve include questions such as how evolution of microbial community composition and function will constrain or accelerate response to perturbation in Earth systems. In addition, can AI identify the rules/patterns/rates of ecosystem response by mining data from ever-increasing data collections across the Earth system? AI may help inform alternatives to trait-based approaches that identify the most sensible/meaningful units of biodiversity at watershed and Earth system scales, for example, of genetic groups versus species/populations/strains/clades or taxonomic versus phylogenetic versus genetic data as input variables into models.

Much of the data needed to apply AI to watershed and Earth systems already exists. However, ICON/FAIR principles in data collection, mining, curating, and access need significant investment. Linking controlled laboratory microbial function information with field observations can yield more cause-and-effect data instead of correlation data. Natural language processing may yield new forms of indirect information that can accelerate surrogate model development. Field studies associated with perturbed environments (e.g., chemical pollutants, fire impacts, erosion, extreme weather) can inform trait-based and surrogate models that connect microbial responses to future climate extremes.

4.5.5 AI Application to Extreme and Future Effects and Events (Fires, Precipitation, Flooding, and Compounding)

Although ML faces data scarcity, we are far from fully utilizing all the observational data for two reasons: our limited data-ingesting abilities and data availability with unknown sources. Several small research groups collect valuable data across the nation and are willing to share that information. However, we do not have any mechanism to maximize data availability through aggregating all the information through these groups. Nevertheless, doing so will require

QA/QC. We argue that there is an urgent need to develop ways to structure complex amalgamations of data through various sources and enhance data-ingesting abilities.

Finally, no universal strategy exists regarding what machine learning method and variables should be used to address extreme events and examine future compounding effects. For example, there are several deep learning approaches, and there is no guidance on using method A vs. B. Further, we choose to rely on a few variables, assuming them to be important rather than considering the full range of data variables. There is an urgent need to develop domain-aware AI and benchmarking AI methods with mechanistic models (or otherwise) to build confidence. However, benchmarks are only as useful as the metrics themselves, so appropriate metrics are also required.

4.5.6 Impacts and Feedback of Human/Engineered Systems on Watersheds (e.g., Reservoir Management, Urban, Agricultural)

Establishing a data portal is a critical research priority and offers a multifaceted challenge. First, the AI/ML workflows and analysis that researchers need to advance are integrating data and models across scientific and engineering disciplines; and hence, development of a metadata standard (and possibly an ontology) is necessary for the portal to function effectively. Next, the portal needs a straightforward API that enables data discovery, remote data operations (e.g., subsetting), and download. Remote data operations depend on the functionality that is supported by the data repository that holds the data, such as ESGF or ESS-DIVE. Next, there is significant research and development surrounding data collection itself. For example, we can survey the community and identify datasets that have been published through successful collaborations with local governments or industry, such as the Argonne National Laboratory (Argonne) collaboration with AT&T on infrastructure locations for analysis of flood risk. We can reach out to other agencies (e.g., U.S. Department of Agriculture) and explore the potential of anonymizing their data, or creating synthetic collections inspired by their data, so that it could be made accessible to the broader community. We can explore the collection and analysis of more remote sensing data, particularly in relation to agricultural practices, but also for water management in general. Finally, we can consider natural language processing to explore historical (including ancestral) as well as current land management practices, and develop ontologies needed to express this knowledge in useful ways for AI.

The data portal should not be limited to traditional observational data. It is imperative that it also provides a catalog of models, documenting the scales and processes that a model supports. In addition, it should support model data (both input and output data), as this data may be integrated in AI/ML workflows for training or evaluation purposes. An additional research priority is development of process-based models that include the specific and unique features of human-impacted watersheds and that can address questions of water quality, not just water quantity.

This includes advances in process-based models to include water management structures such as reservoirs and diversions; agricultural fields, such as tile drains and ditches; and urban features such as riparian zones and storm drains. In addition, the efficacy of the various process representations at various scales should be documented and discoverable through the data portal. To help develop this modeling component of the data portal and advance these models, a number of use cases could be defined that examine current and future scenarios in real or synthetic watersheds.

4.6 Short-term (<5 years), 5-year, and 10-year Goals

At the outset, embedding ML capabilities in watershed science teams requires strong collaboration with AI expertise and is critical to AI application in watershed system science. This may require the formation of Applied AI Research Centers that embed domain scientists with AI/data science expertise to expedite collaboration across an ever-increasing multidisciplinary research area.

In the short term, an effort should be initiated to make available the necessary software infrastructure for integrating ML-surrogate models and process-based models, highly resolved process-based simulations at sufficient scale for benchmarking different approaches, proof-of-concept simulations of scaling approaches with static spatial structure (e.g., ML-based and process-based subcatchments are pre-selected), and a community workshop to define model comparison studies.

Trait-based, alternative, and surrogate approaches that capture genomic and metagenomic information in watershed and Earth system models must be developed. These approaches will require significant investment in ICON/FAIR principles in data collection to enable knowledge-information AI to play a major role in knowledge discovery in complex watershed systems that will continue to be plagued by sparse/scarce data and require data transferability and indirect data to calibrate and scale ecological processes to the watershed and Earth system.

Generating benchmark datasets, co-design of flexibly deployable sensor systems, development of AI-assisted physics-based models, and AI/ML for data assimilation and sensor control are needed. Using AI analysis of existing models is also a good near-term step—models provide rich data on the structure and dynamics of process interactions, as there are non-intuitive interactions that emerge that are excellent hypotheses for us to test empirically. We either falsify those and therefore show that our model structures or parameterizarion needs work, or we accept them and validate our conceptual models/theories.

In the medium term (~5 years), developing a Subsurface Biogeochemistry (BGC) network patterned along the format of Ameriflux should be a priority. The network development can be initiated through the existing field observatories to demonstrate value. The data collection, curation, storage, and accessibility should be patterned, ab initio, to the application of knowledge-informed AI with data science expertise integrated into the initiative.

Using AI to identify patterns of interaction across scales to identify what governs the distributions of traits in ecosystems (plants, microbes, animals, and physical/chemical traits) is also achievable in the near to mid term, for example, by combining multiple remotely sensed data layers and complementary ground-based surveys. We need to do this with theory in mind—that way a community advances together even if we work in disparate systems. The time is right for coupling AI with these exploding data sources to build integrated, soil-aware, land-atmospheric domain datasets that can both improve watershed and Earth system models, and at the same time inform development/deployment of targeted observation networks to reduce uncertainties and increase predictive understanding.

In the longer term (~10 years), our goals should include extending the scaling approach to become dynamic, using AI/ML to replace subdomains of a fully resolved simulation with surrogate models on-the-fly, switching back to fully resolved simulations as needed. In addition, results from a community model comparison would provide the necessary knowledge base to support applications.

Integration of genomic, metagenomic, and ecological datasets across space and time to identify interactions between different scales through biological and environmental trait analysis should be developed. Fully developed hybrid or surrogate models of microbial dynamics and ecology should be developed whereby the coarse-scale grid represents climate factors or fluxes (gasses/particulates) and the fine-grid solver for biological processes.

Building a software infrastructure that allows seamless iterations between the model and experiments, making it easier to test different combinations of ML and process-based models, is critical. We can work with the ESS cyberinfrastructure working groups to collect the design requirements of the computational infrastructure from the broad community for maximum impact. Once built, it will provide transferable scientific tools to understand watershed systems by iteratively learning from both process-based models, observational data, and data-driven approaches, paving our way toward a hybrid modeling approach that couples physical process models with the versatility of data-driven ML to improve the predictability of watershed models and Earth system models.

4.7 References

Allen-Dumas, Melissa R., Haowen Xu, Kuldeep R. Kurte, and Deeksha Rastogi. 2021. "Toward Urban Water Security: Broadening the Use of Machine Learning Methods for Mitigating Urban Water Hazards." *Frontiers in Water* 2. https://doi.org/10.3389/frwa.2020.562304.

Canchumuni, Smith W. A., Alexandre A. Emerick, and Marco Aurélio C. Pacheco. 2019. "Towards a Robust Parameterization for Conditioning Facies Models Using Deep Variational Autoencoders and Ensemble Smoother." *Computers and Geosciences* 128 (January): 87–102. https://doi.org/10.1016/j.cageo.2019.04.006.

Chen, Xingyuan, Raymond Mark Lee, Dipankar Dwivedi, Kyongho Son, Yilin Fang, Xuesong Zhang, Emily Graham, et al. 2021. "Integrating Field Observations and Process-Based Modeling to Predict Watershed Water Quality under Environmental Perturbations." *Journal of Hydrology* 602 (November): 125762. https://doi.org/10.1016/J.JHYDROL.2020.125762.

Cromwell, Erol, Pin Shuai, Peishi Jiang, Ethan T Coon, Scott L Painter, J David Moulton, Youzuo Lin, and Xingyuan Chen. 2021. "Estimating Watershed Subsurface Permeability From Stream Discharge Data Using Deep Neural Networks." *Frontiers in Earth Science* 9: 3. https://doi.org/10.3389/feart.2021.613011.

Dramsch, Jesper Sören. 2020. "70 Years of Machine Learning in Geoscience in Review: A Preprint." https://arxiv.org/abs/2006.13311.

Fu, B., J. S. Horsburgh, A. J. Jakeman, C. Gualtieri, T. Arnold, L. Marshall, T. R. Green, et al. 2020. "Modeling Water Quality in Watersheds: From Here to the Next Generation." *Water Resources Research* 56 (11): e2020WR027721. https://doi.org/https://doi.org/10.1029/2020WR027721.

Goldman, A. E.; S. R. Emani, L. C. Pérez-Angel, J. A. Rodríguez-Ramos, J. C. Stegen, 2021. "Integrated, Coordinated, Open, and Networked (ICON) Science to Advance the Geosciences: Introduction and Synthesis of a Special Collection of Commentary Articles." *Earth and Space Science Open Archive*. https://www.essoar.org/doi/10.1002/essoar.10508554.1.

Helton, Ashley M., Geoffrey C. Poole, Judy L. Meyer, Wilfred M. Wollheim, Bruce J. Peterson, Patrick J. Mulholland, Emily S. Bernhardt, et al. 2011. "Thinking Outside the Channel: Modeling Nitrogen Cycling in Networked River Ecosystems." *Frontiers in Ecology and the Environment* 9 (4): 229–38. https://doi.org/https://doi.org/10.1890/080211.

Kim, Hyun il, and Kun Yeun Han. 2020. "Urban Flood Prediction Using Deep Neural Network with Data Augmentation." *Water* 12 (3): 899. https://doi.org/10.3390/w12030899.

Kim, Junhyeong, Allen H. Goldstein, Romy Chakraborty, Kolby Jardine, Robert Weber, Patrick O. Sorensen, Shi Wang, Boris Faybishenko, Pawel K. Misztal, and Eoin L. Brodie. 2021. "Measurement of Volatile Compounds for Real-Time Analysis of Soil Microbial Metabolic Response to Simulated Snowmelt." *Frontiers in Microbiology* 12 (June). https://doi.org/10.3389/fmicb.2021.679671.

Klasek, Scott A., Marcus T. Brock, Hilary G. Morrison, Cynthia Weinig, and Loïs Maignien. 2021. "Soil Microsite Outweighs Cultivar Genotype Contribution to Brassica Rhizobacterial

Community Structure.” *Frontiers in Microbiology* 12. https://doi.org/10.3389/fmicb.2021.645784.

Kyas, Svetlana, Diego Volpatto, Martin O. Saar, and Allan M. M. Leal. 2022. “Accelerated Reactive Transport Simulations in Heterogeneous Porous Media Using Reaktoro and Firedrake.” *Computational Geosciences* 26 (2): 295–327. https://doi.org/10.1007/s10596-021-10126-2.

Leal, Allan M. M., Svetlana Kyas, Dmitrii A. Kulik, and Martin O. Saar. 2020. “Accelerating Reactive Transport Modeling: On-Demand Machine Learning Algorithm for Chemical Equilibrium Calculations.” *Transport in Porous Media* 133 (2): 161–204. https://doi.org/10.1007/s11242-020-01412-1.

Lu, Hannah, and Daniel M. Tartakovsky. 2021. “Dynamic Mode Decomposition for Construction of Reduced-Order Models of Hyperbolic Problems with Shocks.” *Journal of Machine Learning for Modeling and Computing* 2 (1): 1–29. https://doi.org/10.1615/JMachLearnModelComput.2021036132.

Mao, Yongjing, Daniel L. Harris, Zunyi Xie, and Stuart Phinn. 2021. “Efficient Measurement of Large-Scale Decadal Shoreline Change with Increased Accuracy in Tide-Dominated Coastal Environments with Google Earth Engine.” *ISPRS Journal of Photogrammetry and Remote Sensing* 181 (November): 385–99. https://doi.org/10.1016/J.ISPRSJPRS.2021.09.021.

Maskey, Manil, Hamed Alemohammad, Kevin Murphy, and Rahul Ramachandran. 2020. “Advancing AI for Earth Science: A Data Systems Perspective.” *Eos* 101 (November). https://doi.org/10.1029/2020EO151245.

Meng, Xuhui, and George Em Karniadakis. 2020. “A Composite Neural Network That Learns from Multi-Fidelity Data: Application to Function Approximation and Inverse PDE Problems.” *Journal of Computational Physics* 401 (January): 109020. https://doi.org/10.1016/J.JCP.2019.109020.

Mital, Utkarsh, Dipankar Dwivedi, James B. Brown, Boris Faybishenko, Scott L. Painter, and Carl I. Steefel. 2020. “Sequential Imputation of Missing Spatio-Temporal Precipitation Data Using Random Forests.” *Frontiers in Water* 2 (August). https://doi.org/10.3389/frwa.2020.00020.

Mo, Shaoxing, Nicholas Zabaras, Xiaoqing Shi, and Jichun Wu. 2019. “Deep Autoregressive Neural Networks for High-Dimensional Inverse Problems in Groundwater Contaminant Source Identification.” *Water Resources Research* 55 (5): 3856–81. https://doi.org/10.1029/2018WR024638.

Molins, Sergi, David Trebotich, Bhavna Arora, Carl I. Steefel, and Hang Deng. 2019. “Multi-Scale Model of Reactive Transport in Fractured Media: Diffusion Limitations on Rates.” *Transport in Porous Media* 128 (2): 701–21. https://doi.org/10.1007/S11242-019-01266-2/FIGURES/10.

Napieralski, Stephanie A., and Eric E. Roden. 2020. "The Weathering Microbiome of an Outcropping Granodiorite." *Frontiers in Microbiology* 11. https://doi.org/10.3389/fmicb.2020.601907.

Painter, Scott L. 2021. "On the Representation of Hyporheic Exchange in Models for Reactive Transport in Stream and River Corridors." *Frontiers in Water* 2. https://doi.org/10.3389/frwa.2020.595538.

Painter, Scott L., Ethan Coon, and Dan Lu. 2021. "AI-Directed Adaptive Multifidelity Modeling of Water Availability and Quality at River Basin Scales," April. https://doi.org/10.2172/1769669.

Peherstorfer, Benjamin, Karen Willcox, and Max Gunzburger. 2018. "Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization." *SIAM Review* 60 (3): 550–91. https://doi.org/10.1137/16M1082469.

Raghu, Maithra, Ben Poole, Jon Kleinberg, Surya Ganguli, and Jascha Sohl-Dickstein. 2017. "On the Expressive Power of Deep Neural Networks." In *Proceedings of the 34th International Conference on Machine Learning*, edited by Doina Precup and Yee Whye Teh, 70:2847–54. Proceedings of Machine Learning Research. PMLR. https://proceedings.mlr.press/v70/raghu17a.html.

Sapitang, Michelle, Wanie M. Ridwan, Khairul Faizal Kushiar, Ali Najah Ahmed, and Ahmed El-Shafie. 2020. "Machine Learning Application in Reservoir Water Level Forecasting for Sustainable Hydropower Generation Strategy." *Sustainability* 12 (15): 6121. https://doi.org/10.3390/su12156121.

Saxe, Samuel, Terri S. Hogue, and Lauren Hay. 2018. "Characterization and Evaluation of Controls on Post-Fire Streamflow Response across Western US Watersheds." *Hydrology and Earth System Sciences* 22 (2): 1221–37. https://doi.org/10.5194/hess-22-1221-2018.

Schmidt, Isabel Belloni, and Ludivine Eloy. 2020. "Fire Regime in the Brazilian Savanna: Recent Changes, Policy and Management." *Flora* 268 (July): 151613. https://doi.org/10.1016/J.FLORA.2020.151613.

Shen, Chaopeng. 2018. "A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists." *Water Resources Research* 54 (11): 8558–93. https://doi.org/10.1029/2018WR022643.

Sokol, Noah W., Eric Slessarev, Gianna L. Marschmann, Alexa Nicolas, Steven J. Blazewicz, Eoin L. Brodie, Mary K. Firestone, et al. 2022. "Life and Death in the Soil Microbiome: How Ecological Processes Influence Biogeochemistry." *Nature Reviews Microbiology 2022 20:7* 20 (7): 415–30. https://doi.org/10.1038/s41579-022-00695-z.

Song, Hyun-Seob, James C Stegen, Emily B Graham, Joon-Yong Lee, Vanessa A. Garayburu-Caruso, William C. Nelson, Xingyuan Chen, J. David Moulton, and Timothy D. Scheibe. 2020. "Representing Organic Matter Thermodynamics in Biogeochemical Reactions via Substrate-Explicit Modeling." *Frontiers in Microbiology* 11. https://doi.org/10.3389/fmicb.2020.531756.

Steefel, Carl, Dipankar Dwivedi, Guillen Sole-Mari, Zexuan Xu, Ilhan Ozgen, Allan Leal, and Utkarsh Mital. 2021. "On Demand Machine Learning for Multi-Fidelity Biogeochemistry in River Basins Impacted by Climate Extremes." United States. https://doi.org/10.2172/1769757.

U.S. DOE (U.S. Department of Energy). 2019. "Open Watershed Science by Design: Leveraging Distributed Research Networks to Understand Watershed Systems Workshop Report." DOE/SC-0200, Environmental System Science program, Office of Science. https://ess.science.energy.gov/open-watershed-workshop/.

Wilder, Benjamin T., Catherine S. Jarnevich, Elizabeth Baldwin, Joseph S. Black, Kim A. Franklin, Perry Grissom, Katherine A. Hovanes, et al. 2021. "Grassification and Fast-Evolving Fire Connectivity and Risk in the Sonoran Desert, United States." *Frontiers in Ecology and Evolution* 9. https://doi.org/10.3389/fevo.2021.655561.

5 Ecohydrology

Speakers: Forrest M. Hoffman (ORNL), Zheng Shi (Univ. of Oklahoma), Jinyun Tang (LBNL), Mallory Barnes (Indiana Univ.), Octavia Crompton (Duke Univ.), James Dennedy-Frank (LBNL), Sagar Gautam (SNL), Alexandra Konings (Stanford Univ.), Paul Levine (NASA JPL), Natasha MacBean (Indiana Univ.), Jiafu Mao (ORNL), Elias Massoud (UC-Berkeley), Richard Tran Mills (Argonne), Umakant Mishra (SNL), Sarah Scott (SNL), Matthias Sprenger (LBNL), Naomi Tague (UC-Santa Barbara), Erica Siirila-Woodburn (LBNL), Chonggang Xu (LANL), and Claire Zarakas (Univ. of Washington)

Acknowledgement: Session leaders & workshop organizers would like to acknowledge the significant contribution to report of Yaoping Wang.

5.1 Grand Challenges

Ecohydrology research sits at the intersection of ecosystem ecology and water cycle science, and it incorporates land surface processes and atmospheric and watershed science in addressing ecophysiological responses and feedbacks to the hydrological cycle. Vegetation and benthic organisms, soils, and surface and subsurface hydrology are key components of the Earth system encompassed in ecohydrology. Ecohydrology research focuses on plant transpiration and water use, plant productivity, ecophysiology, plant-soil interactions, and biogeochemistry of terrestrial ecosystems. These processes span scales from stomates and microbes to canopies, watersheds, continents, and the entire globe (Figure 5-1). Understanding interactions among important mechanisms across these scales is challenging. Constraining ecohydrological models is limited by mechanistic knowledge gaps and by observations that are available at only a few of the scales of interest.

Artificial intelligence (AI) and machine learning (ML) approaches will likely provide new avenues for extracting mechanistic understanding from the diversity of data available at different scales. Here we identify fundamental grand challenges in ecohydrology research that are likely to be transformatively addressed by AI/ML approaches.

Grand Challenge #1: Develop multiscale representations of land processes that incorporate heterogeneous patterns of water stores and fluxes, vegetation patterns and processes, physiological function, heterogeneous soil properties and processes, and biogeochemical cycling to understand and predict responses to climate change and climate extremes.

Current models inadequately capture the necessary land processes for simulating ecohydrology at the plant scale. Biological data for root network density and depth, root trait variability, and root responses to varying stresses are sparse, and data on soil properties and processes are insufficient for constraining models. While above-ground processes are better understood, there are still significant gaps in the data needed to resolve species differences in ecophysiological processes, such as carbon allocation, and the degree to which species-specific traits are plastic and adapt to local conditions. Moreover, traditional methods for integrating available data have been inadequate for developing insights into plant-soil interactions and how those interactions respond to and feed back into hydrology at the watershed and larger scales.

Grand Challenge #2: Develop models of climate extremes, pulse and press stresses, ecophysiological responses, and ecosystem structure and function to understand and predict ecosystem disturbances and recovery.

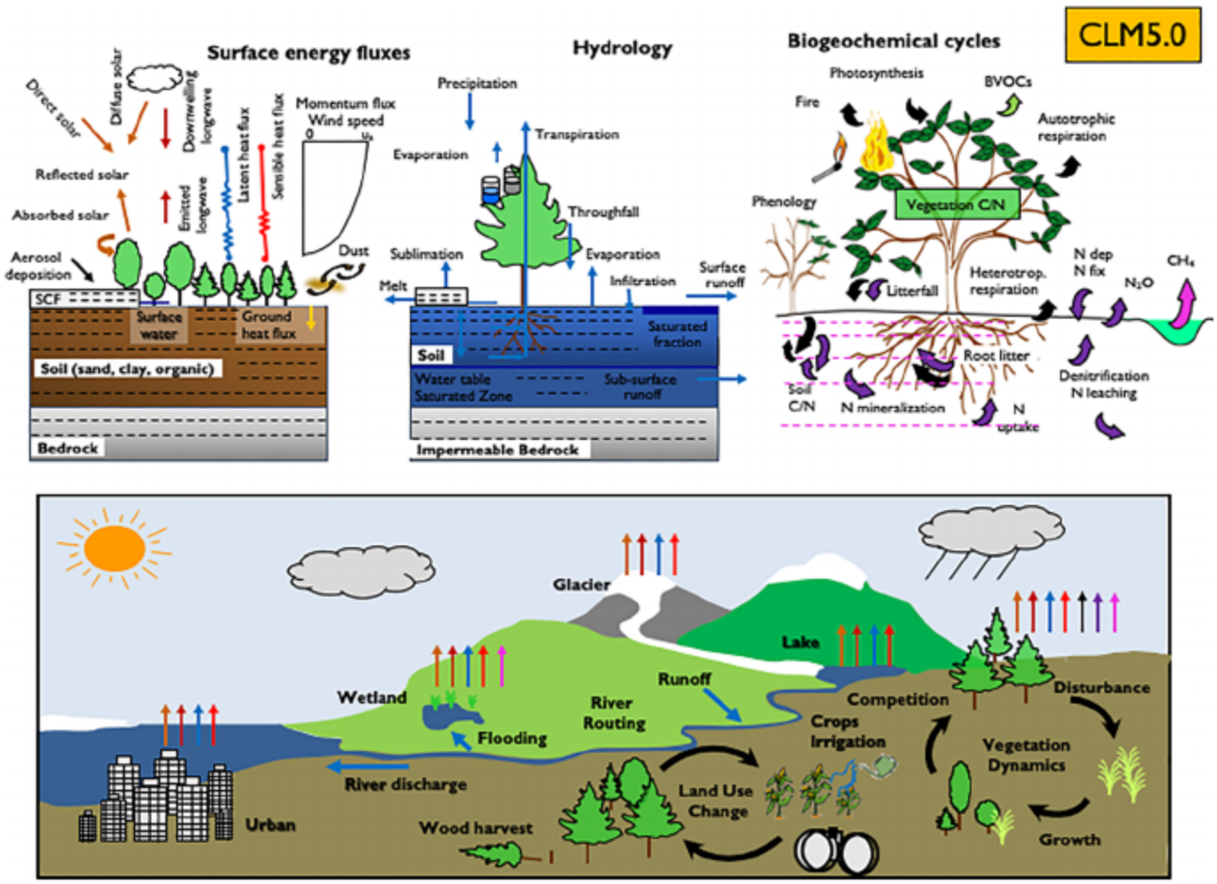


Figure 5-1. Schematic representation of primary processes and functionality in the Community Land Model version 5 (CLM5) shown as an example of integrated ecohydrological processes required for modeling across scales. SCF = snow cover fraction; BVOC = biogenic volatile organic compounds; and C/N = carbon and nitrogen. For biogeochemical cycles, black arrow denotes carbon flux, and the purple arrow denotes nitrogen flux. Note that not all soil levels are shown. Not all processes are depicted. Optional features that are not active in default configurations are italicized (Source: Reproduced from Figure 1 in Lawrence et al. 2019 under Creative Commons [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Land surface models adequately simulate mean state behavior of vegetation, soils, and interactions with the atmosphere, but they often fail to capture responses to climate extremes either because of missing processes or sensitivity to changes in temperature and precipitation that are too weak or too strong. Ecosystem disturbances and recovery patterns are especially challenging because traditional big-leaf models do not incorporate vegetation's structural elements required to mechanistically account for changes in structure and function induced by climate or meteorological extremes, like windthrow, frost, drought, and insect or pathogen outbreaks. Biases in atmosphere models and forcing data, as well as scaling issues, can also strongly affect land surface model responses.

Grand Challenge #3: Apply machine learning to assimilate and calibrate emerging datasets, constrain model complexity, develop functional model benchmarks, and quantify the magnitude and sources of model and data uncertainty.

A wide diversity of satellite and airborne remote sensing and in situ measurements are available to support ecohydrology research; however, the data are not well integrated and often do not include observations of quantities needed by models. In an effort to improve model performance, scientists tend to increase the complexity of models to capture processes for which sufficient measurements are not available and parameters are highly uncertain. AI/ML approaches are already being used to improve data through multisensor data fusion and quantitative methods for extrapolation and accounting for heterogeneity. Similar approaches are showing promise for calibrating model parameters and quantifying model structural uncertainty. AI/ML approaches are needed to improve data, develop multivariate model benchmarks of functional performance, and constrain the ever-increasing complexity of models.

5.2 State-of-the-Science

Ecohydrology involves coupling of soils, plants, and the atmosphere, requiring computationally intensive iterative solutions, which are difficult to integrate with limited observations. Machine learning approaches are already being used to (1) interpolate, extrapolate, and integrate data and models, accounting for nonlinear relationships among variables, to constrain and improve models; (2) build data-driven model components or parameterizations of processes from measurements and observational data products; and (3) develop emulators and surrogate models of complex, nonlinear process representations for parameter optimization and model tuning.

Bi-linear interpolation, kriging, cluster analysis, random forests, model tree ensembles, convolutional neural networks, and other machine learning methods have been applied to spatially sparse measurements to understand their representativeness (e.g., Hoffman et al. 2013; Kumar et al. 2016), to design optimal sampling networks (e.g., Keller et al. 2008; Hoffman et al. 2013), to analyze multidimensional model output (e.g., Burke et al. 2021), and to intelligently upscale and extrapolate environmental fluxes and characteristics over larger spatial domains (e.g., Langford et al. 2019; Jung et al. 2020; Konduri et al. 2020) using inferred relationships with environmental gradients, ecosystem dynamics, and remote sensing radiances. Recently, Mishra et al. (2020) compared four different machine learning approaches with traditional regression kriging to estimate surface soil organic carbon (SOC) stocks for the northern circumpolar permafrost region. Their results showed that the ensemble median of the results of these machine learning techniques exhibited the highest prediction accuracy.

Widening adoption of deep neural networks and the growth of meteorological and climate data have fueled interest in adopting machine learning technologies for use in weather and climate models (Dueben and Bauer 2018). Leveraging the successes in rainfall prediction (Miao et al. 2015; Tao et al. 2016), soil moisture retrievals (Santi et al. 2016; Kolassa et al. 2017), and surface turbulent flux retrievals (Alemohammad et al. 2017; Jung et al. 2020), researchers are training deep neural networks as model parameterizations, initially for convection and subgrid-scale processes (Rasp et al. 2018; Gentine et al. 2018; Brenowitz and Bretherton 2018, 2019; Brenowitz et al. 2020), which are poorly captured by current models or are computationally prohibitive for decadal or longer timescale simulations.

Massoud (2019) used polynomial chaos expansion (PCE) emulators and sparse grid sampling for models of increasing complexity, including a hydrology, ecohydrologic, and vegetation dynamics model. The study showed that emulators performed better for lower-complexity models, as opposed to those models that were more complex. However, the results demonstrated that dimensionality reduction improved the emulation of even the most complex model in the study.

Burke et al. (2021) employed random forests to identify the relative importance of biophysical and climatic parameters in predicting effects of fuel treatment in forests. These researchers found that interactions between biophysical settings, climate, and fuel treatments are complex and have nonlinear effects on forests, water, and fire behavior. The importance of individual drivers emerged from their analysis. They further indicated that random forest models could be used to test additional scenarios without needing to run the complex model.

5.3 Experimental, Data, and Modeling Opportunities

Advancing Earth system predictions with AI/ML methods requires large quantities of data regarding relevant processes across multiple spatial and temporal scales. Data requirements for training ML algorithms typically exceed the data needs for traditional process model development, verification, and validation. Therefore, additional data may be required from new laboratory and field measurements, manipulative experiments, airborne and satellite remote sensing, multisensor fusion and data synthesis, and modeling studies. Collecting, aggregating, sharing/distributing, and archiving these larger quantities of data and newly derived data products require a systematic and organized approach to data management. Creating, finding, accessing, analyzing, visualizing, and utilizing these data to train ML algorithms necessitate an integrated storage and computational infrastructure available across projects, institutions, and individual investigators.

5.3.1 New Data Products

Ecohydrology research suffers from a lack of sufficient data across spatial scales from microbial and leaf scales to watershed and continental scales. In particular, because of its high spatial heterogeneity and difficulty in sampling, much more belowground data are needed to reduce characterization uncertainties and understand relationships between soil organic carbon and environmental factors (e.g., soil moisture, soil texture) that influence its formation and turnover and to understand root density, distribution and how roots change with environmental conditions. Similarly, species-specific plant data are needed globally to improve the representation of vegetation communities in models and better characterize and simulate responses to environmental change. A wide variety of measurement techniques is required across scales, and a hierarchical modeling approach is needed to simulate processes across spatial scales (Figure 5-2). AI/ML can be useful in acquiring such data through optimization of sampling or monitoring networks (e.g., Keller et al. 2008; Hargrove et al. 2003), autonomous control of measurement or sampling devices under changing conditions and extreme events, intelligent gap-filling and extrapolation of point measurements (e.g., Mishra et al. 2020; Jung et al. 2020), and fusion of data from multiple sensors and in situ data from different agencies and measurement campaigns (e.g., Langford et al. 2017). New data products should be constructed in a manner that makes them easily accessible as a collection in standard, well-documented formats to facilitate ease of use and testing with a wide range of AI/ML approaches. One prominent example of such benchmark datasets, called ImageNet (Russakovsky et al. 2015), consists of a collection of images with associated labels (nouns) that can be used by the research community to train and test any number of object detection algorithms. Building collections of labeled Earth science data and offering them to the community would facilitate rapid testing of existing AI/ML methods and faster development of new AI/ML methods aimed specifically at addressing the needs of ecohydrology and related Earth science research.

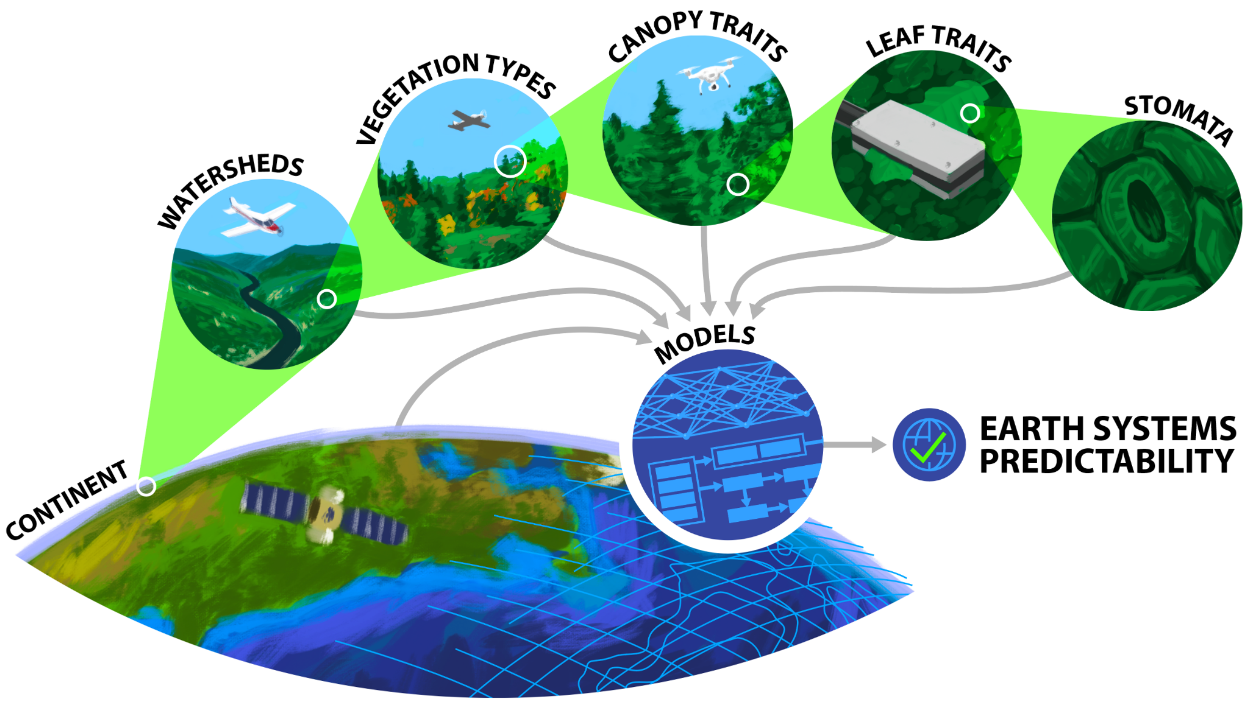


Figure 5-2. A wide variety of measurement techniques is required across spatial scales from stomata to biomes to improve representation of ecohydrological processes with machine learning approaches, and a hierarchy of process-based and machine learning-based models is needed to simulate important processes across those scales and improve Earth system predictability (Source: Figure courtesy of Nathan Armistead, Oak Ridge National Laboratory).

5.3.2 Hybrid Models

Improving and developing new model parameterizations of ecohydrology processes in models is inherent in the Grand Challenges presented above. However, where sufficient data are available, the opportunity exists to train deep neural networks as model parameterizations. Such efforts have begun, initially for convection and subgrid-scale processes (Rasp et al. 2018; Gentine et al. 2018; Brenowitz and Bretherton 2018, 2019; Brenowitz et al. 2020), which are poorly captured by current models or are computationally prohibitive for decadal or longer timescale simulations. Adding such capabilities in land surface models for simulating ecohydrological processes could greatly advance the utility and performance of these models. Envisioned is a framework that employs such methods for data-driven, hybrid process-/ML-based Earth system models (Schneider et al. 2017). As can be seen from this early work, lack of adequate data, numerical instabilities in coupling, and “out-of-sample” problems must be overcome, but the outlook for these approaches is promising. Employing similar approaches for adding machine learning capabilities in land surface models for simulating ecohydrological processes could greatly advance the utility and performance of these models.

By explaining patterns identified by machine learning, physically based models can be improved, providing transferability across space and time scales (Figure 5-3). Alternatively, ML can be used to reduce the complexity of multidimensional output from physically based models. The framework envisioned employs machine learning methods for data-driven process representation alongside traditional differential equation-based representation of ecohydrology processes, resulting in a hybrid process-based/ML-based Earth system model (Schneider et al. 2017). To facilitate that vision, existing models must be made more modular so that individual process-based or ML-based parameterizations with a model may be swapped in and out as desired.

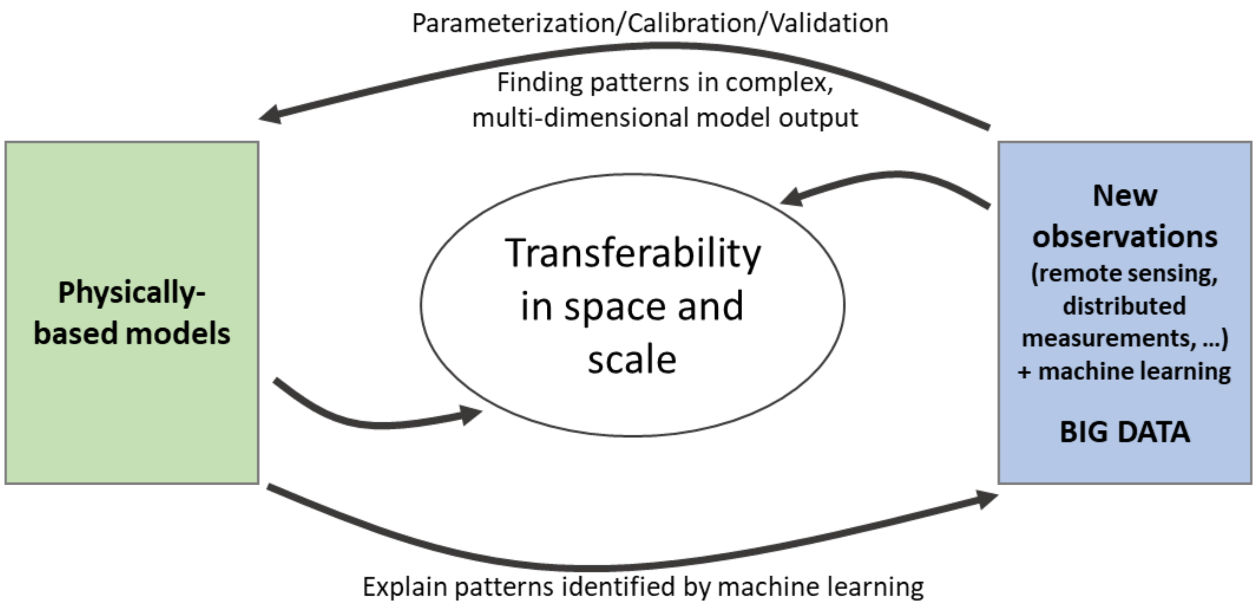


Figure 5-3. Combining physically based models with machine learning models enables identification of processes and patterns that can inform future model development and new observational campaigns. Such hybrid models provide transferability across space and time scales (Source: Figure courtesy of Naomi Tague, UC Santa Barbara).

5.3.3 Computing and Data Infrastructure

The research community currently has access to high-performance computing capabilities at large computing centers within DOE, like NERSC at Lawrence Berkeley National Laboratory, the Oak Ridge Leadership Computing Facility, and the Argonne Leadership Computing Facility. The community has access to large collections of data at data centers like DOE's Atmospheric Radiation Measurement (ARM) Data Center (ADC), Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE), Earth System Grid Federation (ESGF), NASA's Distributed Active Archive Centers (DAACs), and others. These data centers operate as stand-alone resources and require data users to download data to their own computational resources. This process of downloading data, pre-processing and integrating the data, and then performing simulations and analysis is tedious and unnecessary given recent technological developments. When developing and deploying AI/ML methods, the difficulty of this workflow will increase since high-speed access to vastly larger data collections will be required for training ML models, potentially doing such training as part of simulation itself.

The opportunity before the community is to build integrated computing and data infrastructure that eliminates the challenges of finding, acquiring, and downloading data. Benchmark AI/ML data should be accessible from any large computing environment, no matter where those data reside or are archived. This could be accomplished through application programming interfaces and data transport services, like Globus (https://globus.org/), that hide the details of data movement and exploit high-bandwidth networks to deliver data as needed for simulation and analysis. Funding agencies might coordinate in the creation of a model-data integration center that could provide such integrated storage and computing resources for the growing Earth system science community. The center could provide data hosting services, offer compute-near-the-data computational infrastructure and "AI/ML as a service" capabilities, and sponsor training activities and multidisciplinary working groups focused on advancing new or advanced research topics that may have some element of risk. Such a center could lower the bar of entry for laboratory and university scientists while enabling research with tools not otherwise easily accessible or usable.

5.4 Research Priorities

Priorities for near-term research in ecohydrology should aim to prepare the research community to address the Grand Challenges enumerated above. This includes improving characterization of soil and vegetation properties, improving representation of water stores and fluxes, developing models of extremes and ecosystem disturbance and recovery, and developing new assimilation and analysis capabilities to help constrain models and quantify sources of both model and data uncertainty. The research community is at a stage where progress can be made in creating benchmark "AI-ready" datasets and developing initial machine learning parameterizations and process emulators. Initial research and development activities should engage a broader, more multidisciplinary community of researchers, particularly in mathematics and computer science. Transitioning the community to significant use of AI/ML approaches in ecohydrology and climate science will require enhanced efforts to train the next generation of researchers to use new tools and methods. National scientific workforce development activities should consider how best to deliver the additional knowledge and training to early career scientists.

5.4.1 Benchmark Datasets

While Earth system and environmental data centers distribute and archive a wide variety of data collections from in situ measurements, monitoring networks, and airborne and satellite remote sensing platforms, they do not typically lead activities to synthesize data products across those collections for specific research purposes. Instead, funded or volunteer working groups are often formed to synthesize data to address specific science questions or hypotheses. Such working groups may be catalyzed by existing projects (e.g., RUBISCO Soil Carbon Dynamics Working Group, RUBISCO-AmeriFlux Working Group), data collection activities or databases (e.g., various AmeriFlux and Fluxnet working groups, International Soil Carbon Network [ISCN], International Soil Radiocarbon Database [ISRaD], TRY Plant Trait Database, International Soil Moisture Network), or data synthesis centers (e.g., National Center for Ecological Analysis and Synthesis [NCEAS], National Institute for Mathematical and Biological Synthesis [NIMBioS], Powell Center for Analysis and Synthesis, Aspen Global Change Institute). They may be sponsored by the National Science Foundation, U.S. Geological Survey, U.S. Department of Agriculture, and other agencies and nongovernmental organizations. However, because these working group activities are often narrowly focused, they may not produce synthesized data products that are of general use, well documented, easily distributed, archived, and maintained over time. A more systematic approach with a broader vision for reusability and maintainability is required to generate benchmark datasets for training, testing, and benchmarking AI/ML models.

Producing and maintaining large collections of understandable and reusable data, like that from ImageNet (Russakovsky et al. 2015), will be of great utility to the ecohydrology research community and will facilitate wider engagement of the mathematics and computer science communities already involved in developing and applying AI/ML methods. Some of these datasets will be similar to climate reanalysis data products (e.g., ERA5) or synthesized data used for model evaluation by software like the International Land Model Benchmarking (ILAMB) package (Collier et al. 2018). However, they must be highly multivariate for ML methods to uncover relationships, integrated in a consistent manner for direct use without translation or conversion, and available across multiple spatial and temporal scales and must contain long time series of a large number of samples, points, or grid cells. Such datasets should draw upon many independent data sources, such as data fused from multiple remote sensing platforms, and be calibrated with in situ measurements and continental-scale monitoring networks. To be of greatest utility, these data must be maintained and distributed by existing or new data centers, and integrated computing and storage infrastructure should be developed to facilitate data discovery and eliminate barriers to data movement and download.

5.4.2 Hybrid Modeling

Given the availability of growing volumes of observational data and in situ measurements, the Earth system modeling community is beginning to adopt data-driven approaches for high-resolution weather and climate simulations (Schneider et al. 2017). An ML framework could be used to integrate the wealth of leaf-level fluorescence and gas exchange measurements (e.g., Leafweb), AmeriFlux and FLUXNET ecosystem fluxes, and Free Air Carbon Dioxide Enrichment (FACE) and Spruce and Peatland Responses Under Changing Environments

(SPRUCE) data to develop a unified treatment of stomatal responses, assimilation, and acclimation to changes in hydrology and soil moisture. ML-based models of stomatal conductance and plant hydraulics should be employed to produce a hybrid process-based/ML-based land model with the aim of reducing the uncertainty of soil moisture and carbon assimilation (Figure 5-4). Such hybrid ecohydrology models could also inform watershed models to deliver dynamic ecological process representations often absent in such models. In addition, ML models can be created to improve the characterization of soil organic carbon and soil bulk properties to further reduce soil moisture uncertainties. ML methods should be explored to scale leaf-level and ecosystem processes to the watershed scale for season-to-interannual predictions, through a hierarchy of ML and process-based models, and further to regional and continental scales for interannual-to-decadal predictions. For research questions involving disturbance and recovery, new mechanistic modeling approaches to disturbance and disturbance recovery (e.g., Hanan et al. 2021) are advancing understanding, and these models would benefit from detailed information about change to vegetation structure to both support model parameterization and evaluation. Modeling disturbance is an area, due to its complexity, that would particularly benefit from hybrid approaches.

Since plant and soil processes respond to climate change, FACE and SPRUCE data should be used to develop climate-adaptive ML models for the processes described above. This approach could enable significant steps forward in developing and integrating new and alternative parameterizations within Earth system models, like DOE's Energy Exascale Earth System Model (E3SM), to produce a hybrid process-based/ML modeling framework (Reichstein et al. 2019). The requirement for reducing uncertainties in ecohydrological processes dictates prioritizing process representations of land–atmosphere interactions (energy, water, and carbon) that are (1) highly uncertain but for which observational data are available and (2) computationally expensive. Measurements of leaf-level responses to environmental variations can be related to measurements made at the canopy scale to reduce uncertainties in canopy integration schemes. ML methods can be applied to scale up plant responses—informed by ecosystem- and watershed-scale measurements, upscaled soil properties, and remote sensing data—to bound water budgets for watersheds and quantify risks of flooding and drought, particularly under water cycle extremes. While the primary motivation is to improve mechanistic understanding of these processes across scales, by connecting a chain of hierarchical ML-empowered models to weather forecasting systems, the results may be useful for informing probabilistic risk analysis to quantify risks for urban areas and other built infrastructure and to better quantify drought impacts on streamflow for energy and water utilities.

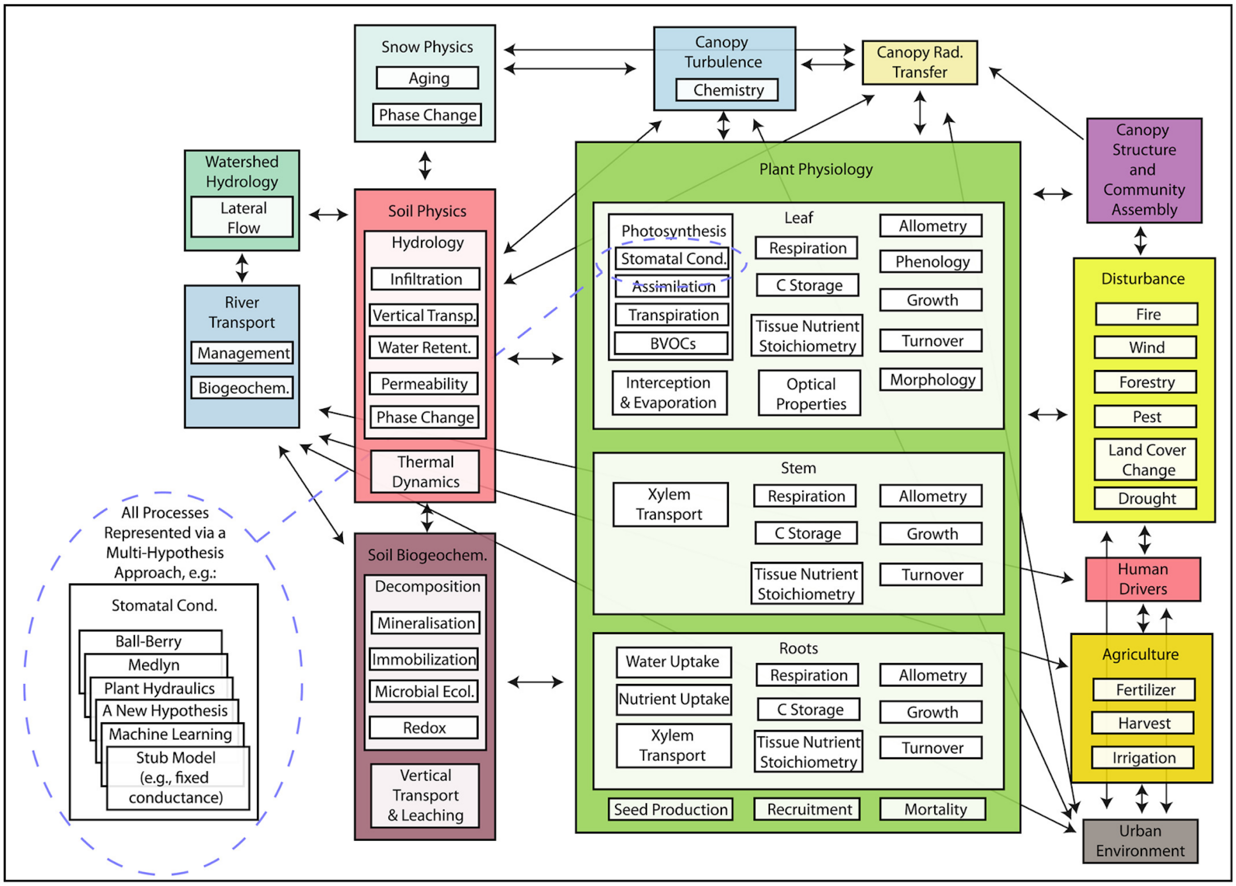


Figure 5-4. A process schematic of a full-complexity land surface model. Processes, and sets of processes, are represented as boxes in the diagram, with information connections represented as arrows. All processes—though here shown only for stomatal conductance—are intended to allow alternative specification, including possibly multiple hypothetical process realizations, empirical or machine learning-derived formulations, and/or simplified stub or null representations to allow for holding a given process constant while other processes vary (Source: Figure adapted from Fisher and Koven 2020 under Creative Commons CC BY 4.0).

5.4.3 Multidisciplinary Engagement

New research in ecohydrology employing AI/ML approaches will benefit from strong collaboration with scientists in mathematics and computer science, who routinely apply such methods in other disciplines and who are actively developing new methods specific to research needs in other domains. Strengthening such collaborations will require frequent interaction between domain experts and computer scientists, mathematical generalization of specific process representations in models, and well-documented benchmark datasets. For ecohydrology, engaging with mathematicians and computer scientists will enable leveraging of research and development activities already underway, and it will foster long-lasting collaborations that will benefit both sets of communities. For long-lasting changes fostering intense cross-disciplinary collaborations, mathematics and computer science should become more prominent in Earth system science education at both the undergraduate and graduate levels.

5.4.4 Integrated Data and Computational Research Infrastructure

Adding AI/ML approaches for data acquisition, processing, assimilation, modeling, and analysis will require improved infrastructure for large datasets and computational capacity. The growth opportunity is to build integrated computing and data infrastructure that eliminates the challenges of finding, acquiring, and downloading data. Benchmark AI/ML data should be accessible from all large computing environments, no matter where those data reside or are archived. This could be accomplished through application programming interfaces and data transport services, like Globus (https://globus.org/), that hide the details of data movement and exploit high-bandwidth networks to deliver data as needed for simulation and analysis. Funding agencies might coordinate in the creation of a model-data integration center that could provide such integrated storage and computing resources for the growing Earth system science community. The center could provide data hosting services, offer compute-near-the-data computational infrastructure and "AI/ML as a service" capabilities, and sponsor training activities and multidisciplinary working groups focused on advancing new or advanced research topics that may have some element of risk. Such a center could lower the bar of entry for laboratory and university scientists, fostering multidisciplinary engagement, while enabling research with tools not otherwise easily accessible or usable.

5.4.5 Training and Workforce Development

In order to advance research with AI/ML approaches, current and next-generation researchers need training on the wide variety of ML methods, data management, large-scale analytics techniques, and use of integrated computational and data resources. This could be accomplished through fellowships that support national laboratory internships for promising graduate students, training courses for postdoctoral and early career scientists (akin to open access online classes for hydrology at https://www.hydrolearn.org/), and seminars and hackathons for existing staff (similar to hydrology seminars provided by the CUAHSI Community at https://www.cuahsi.org/community). These activities could begin with webinars that highlight existing research in national laboratories and universities and virtual hackathons that demonstrate analysis techniques, useful software packages, and strategies for applying emerging datasets. These education and training activities should be an integrated part of training the next-generation workforce of diverse research scientists to meet the needs of the nation.

5.5 Short-term (<5 years), 5-year, and 10-year Goals

Addressing the research priorities identified above will lead to completion of goals in the short term, in the mid term, and in the longer term. Incremental progress through these goals is expected to reduce model uncertainties and improve predictions, leading to actionable science outcomes. The following short-term, mid-term, and long-term goals provide a roadmap for ecohydrology experiments, data, and models.

5.5.1 Short-Term (<5 years) Goals

Shorter-term goals include efforts to:

- Develop a collection of "AI-ready" benchmark datasets for leaf-level measurements of fluxes of energy, water, and carbon; canopy-level observations of evapotranspiration and productivity; and continental-scale estimates of carbon and water cycle time series from in situ measurements and airborne and satellite remote sensing.
- Synthesize existing data in a network-of-networks approach to provide "AI-ready" datasets on subsurface characterization (e.g., high-frequency soil moisture dynamics, soil water tracer data) across large environmental gradients to study the soil-plant feedbacks.
- Improve the modularity of current models so that individual parameterizations can be isolated and swapped with ML-based versions of parameterizations.
- Develop an initial set of ML-based parameterizations for photosynthesis and soil processes that can be integrated as components into hybrid models.
- Establish collaborative opportunities across Earth system science, mathematics, and computer science directed at developing and applying novel and domain-specific ML methods to improve the accuracy of ecohydrology process representations in Earth system models.
- Design and begin implementation of an integrated data and computational infrastructure to support AI/ML in Earth system science. This could leverage existing data centers, computational centers, and software infrastructure, and potentially be transitioned to its own center or facility for broader engagement of the research community.
- Initiate a webinar series for educating and training cross-disciplinary researchers across career stages about the use of AI/ML methods and tools.
- Conduct virtual and in-person hackathons for more rigorous training of graduate students, postdoctoral scholars, and early career scientists.

5.5.2 Mid-Term (5 years) Goals

Mid-term goals include efforts to:

- Develop an initial modeling framework for swapping or interchanging process-based and ML-based parameterizations within Earth system models.
- Foster cross-disciplinary research and training by sponsoring trans-disciplinary working groups that include observational scientists, modelers, data scientists, mathematicians, and computer scientists to take advantage of the benchmark data, ML model frameworks, and integrated computational and storage resources to address specific science questions in ecohydrology.
- Develop accurate and efficient science-guided ML systems or models to predict effects of different ecohydrological disturbances and post-disturbance responses and feedbacks.
- Employ ML to generate new synthetic data for training ML algorithms, for example, photographing each root core collected and developing an ML algorithm to help understand and fast-track improvements in data observations of this kind.

- Develop ML algorithms that can infer what additional measurements are needed and what optimal sampling frequencies and spatial distributions will lead to improvements in ecohydrology models.

5.5.3 Long-Term (10 years) Goals

Long-term goals include efforts to:

- Deploy a fully functioning modeling framework for easily configuring and monitoring ML-based parameterizations alongside process-based parameterizations within Earth system models, supporting online training and in situ analysis and visualization.
- Deploy a fully functioning, explainable ML framework that can identify where to collect data (space/time gaps), what processes need to be improved (physics/chemistry/biology gaps), and how to better manage and analyze data for ecohydrological applications.
- Deploy a fully functioning ecohydrological modeling subsystem for Earth system models that is tested and calibrated for accurate predictions across relevant space and time scales, and which includes ecosystem disturbance and recovery process representations.
- Establish a multiagency AI center to provide computational and storage infrastructure, necessary benchmark data, a wide variety of models at different scales, software tools for analysis and visualization, and staff to support a collection of working groups proposed to address key science questions in ecohydrological predictability.

5.6 References

Alemohammad, S. H., B. Fang, A. G. Konings, F. Aires, J. K. Green, J. Kolassa, D. Miralles, C. Prigent, and P. Gentine (2017), Water, Energy, and Carbon with Artificial Neural Networks (WECANN): A statistically based estimate of global surface turbulent fluxes and gross primary productivity using solar-induced fluorescence, *Biogeosci.*, 14(18):4101–4124, doi:10.5194/bg-14-4101-2017.

Brenowitz, N. D., and C. S. Bretherton (2018), Prognostic validation of a neural network unified physics parameterization, *Geophys. Res. Lett.*, 45(12):6289–6298, doi:10.1029/2018GL078510.

Brenowitz, N. D., and C. S. Bretherton (2019), Spatially extended tests of a neural network parameterization trained by coarse-graining, *J. Adv. Model. Earth Sy.*, 11(8):2728–2744, doi:10.1029/2019MS001711.

Brenowitz, N. D., T. Beucler, M. Pritchard, and C. S. Bretherton (2020), Interpreting and stabilizing machine-learning parameterizations of convection, *J. Atmos. Sci.*, 77(12):4357–4375, doi:10.1175/JAS-D-20-0082.1.

Burke, W. D., C. Tague, M. C. Kennedy, and M. A. Moritz (2021), Understanding How Fuel Treatments Interact With Climate and Biophysical Setting to Affect Fire, Water, and Forest Health: A Process-Based Modeling Approach, *Front. For. Glob. Change*, 3, 143, doi:10.3389/ffgc.2020.591162.

Collier, N., F. M. Hoffman, D. M. Lawrence, G. Keppel-Aleks, C. D. Koven, W. J. Riley, M. Mu, and J. T. Randerson (2018), The International Land Model Benchmarking (ILAMB) System: Design, Theory, and Implementation, *J. Adv. Model. Earth Syst.*, 10(11):2731–2754, doi:10.1029/2018MS001354.

Dueben, P. D., and P. Bauer (2018), Challenges and design choices for global weather and climate models based on machine learning, *Geosci. Model Dev.*, 11(10):3999–4009, doi:10.5194/gmd-11-3999-2018.

Fisher, R. A., and C. D. Koven (2020), Perspectives on the Future of Land Surface Models and the Challenges of Representing Complex Terrestrial Systems, *J. Adv. Model. Earth Sy.*, 12(4), e2018MS001453, doi:10.1029/2018MS001453.

Hanan, E. J., J. Ren, C. L. Tague, C. A. Kolden, J. T. Abatzoglou, R. R. Bart, M. C. Kennedy, M. Liu, and J. C. Adam (2021), How Climate Change and Fire Exclusion Drive Wildfire Regimes at Actionable Scales, *Environ. Res. Lett.*, 16(2), 024051, doi:10.1088/1748-9326/abd78e.

Hargrove, W. W., F. M. Hoffman, and B. E. Law (2003), New analysis reveals representativeness of the AmeriFlux network, *Eos Trans. AGU*, 84(48):529, 535, doi:10.1029/2003EO480001.

Hoffman, F. M., J. Kumar, R. T. Mills, and W. W. Hargrove (2013), Representativeness-based sampling network design for the State of Alaska, *Landscape Ecol.*, 28(8):1567–1586, doi:10.1007/s10980-013-9902-0.

Jung, M., C. Schwalm, M. Migliavacca, S. Walther, G. Camps-Valls, S. Koirala, P. Anthoni, S. Besnard, P. Bodesheim, N. Carvalhais, F. Chevallier, F. Gans, D. S. Goll, V. Haverd, P. Köhler, K. Ichii, A. K. Jain, J. Liu, D. Lombardozzi, J. E. M. S. Nabel, J. A. Nelson, M. O'Sullivan, M. Pallandt, D. Papale, W. Peters, J. Pongratz, C. Rödenbeck, S. Sitch, G. Tramontana, A. Walker, U. Weber, and M. Reichstein (2020), Scaling carbon fluxes from eddy covariance sites to globe: Synthesis and evaluation of the FLUXCOM approach, *Biogeosci.*, 17(5):1343–1365, doi:10.5194/bg-17-1343-2020.

Keller, M., D. S. Schimel, W. W. Hargrove, and F. M. Hoffman (2008), A continental strategy for the National Ecological Observatory Network, *Front. Ecol. Environ.*, 6(5):282–284, doi:10.1890/1540-9295(2008)6[282:ACSFTN]2.0.CO;2.

Kolassa, J. R. H. Reichle, and C. S. Draper (2017), Merging active and passive microwave observations in soil moisture data assimilation, *Remote Sens. Environ.*, 191:117–130, doi:10.1016/j.rse.2017.01.015.

Konduri, V. S., J. Kumar, W. W. Hargrove, F. M. Hoffman, and A. R. Ganguly (2020), Mapping crops within the growing season across the United States, *Remote Sens. Environ.*, 251, 112048, doi:10.1016/j.rse.2020.112048.

Kumar, J., F. M. Hoffman, W. W. Hargrove, and N. Collier (2016), Understanding the representativeness of FLUXNET for upscaling carbon flux from eddy covariance measurements, *Earth Syst. Sci. Data Discuss.*, 1–25, doi:10.5194/essd-2016-36.

Langford, Z. L., J. Kumar, and F. M. Hoffman (2017), Convolutional Neural Network Approach for Mapping Arctic Vegetation using Multi-Sensor Remote Sensing Fusion, *Proceedings of the 2017 IEEE International Conference on Data Mining Workshops (ICDMW 2017)*, Institute of Electrical and Electronics Engineers (IEEE), Conference Publishing Services (CPS), doi:10.1109/ICDMW.2017.48.

Langford, Z. L., J. Kumar, F. M. Hoffman, A. L. Breen, and C. M. Iversen (2019), Arctic vegetation mapping using unsupervised training datasets and convolutional neural networks, *Remote Sens.*, 11(1):69, doi:10.3390/rs11010069.

Lawrence, D. M., R. A. Fisher, C. D. Koven, K. W. Oleson, S. C. Swenson, G. Bonan, N. Collier, B. Ghimire, L. van Kampenhout, D. Kennedy, E. Kluzek, P. J. Lawrence, F. Li, H. Li, D. Lombardozzi, W. J. Riley, W. J. Sacks, M. Shi, M. Vertenstein, W. R. Wieder, C. Xu, A. A. Ali, A. M. Badger, G. Bisht, M. van den Broeke, M. A. Brunke, S. P. Burns, J. Buzan, M. Clark, A. Craig, K. Dahlin, B. Drewniak, J. B. Fisher, M. Flanner, A. M. Fox, P. Gentine, F. M. Hoffman, G. Keppel-Aleks, R. Knox, S. Kumar, J. Lenaerts, L. R. Leung, W. H. Lipscomb, Y. Lu, A. Pandey, J. D. Pelletier, J. Perket, J. T. Randerson, D. M. Ricciuto, B. M. Sanderson, A. Slater, Z. M. Subin, J. Tang, R. Q. Thomas, M. Val Martin, and X. Zeng (2019), The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty, *J. Adv. Model. Earth Sy.*, 11(12):4245–4287, doi:10.1029/2018MS001583.

Massoud, E. C. (2019), Emulation of Environmental Models Using Polynomial Chaos Expansion, *Environ. Model. Softw.*, 111:421–431, doi:10.1016/j.envsoft.2018.10.008.

Miao, C., H. Ashouri, K.-L. Hsu, S. Sorooshian, and Q. Duan (2015), Evaluation of the PERSIANN-CDR daily rainfall estimates in capturing the behavior of extreme precipitation events over China, *J. Hydrometeor.*, 16(3):1387–1396, doi:10.1175/JHM-D-14-0174.1.

Mishra, U., S. Gautam, W. J. Riley, and F. M. Hoffman (2020), Ensemble Machine Learning Approach Improves Predicted Spatial Variation of Surface Soil Organic Carbon Stocks in Data-Limited Northern Circumpolar Region, *Front. Big Data*, 3:40, doi:10.3389/fdata.2020.528441.

Rasp, S., M. S. Pritchard, and P. Gentine (2018), Deep learning to represent subgrid processes in climate models, *Proc. Nat. Acad. Sci.*, 115(39):9684–9689, doi:10.1073/pnas.1810286115.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat (2019), Deep learning and process understanding for data-driven Earth system science, *Nature*, 566(7743): 195–204, doi:10.1038/s41586-019-0912-1.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei (2015), ImageNet Large Scale Visual Recognition Challenge, *Int. J. Comput. Vis.*, 115(3):211–252, doi:10.1007/s11263-015-0816-y.

Santi, E., S. Paloscia, S. Pettinato, and G. Fontanelli (2016), Application of artificial neural networks for the soil moisture retrieval from active and passive microwave spaceborne sensors, *Int. J. Appl. Earth Obs. Geoinf.*, 48:61–73, doi:10.1016/j.jag.2015.08.002.

Schneider, T., S. Lan, A. Stuart, and J. Teixeira (2017), Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations, *Geophys. Res. Lett.*, 44:12,396–12,417, doi:10.1002/2017GL076101.

Tao, Y., X. Gao, K. Hsu, S. Sorooshian, and A. Ihler (2016), A deep neural network modeling framework to reduce bias in satellite precipitation products, *J. Hydrometeor.*, 17(3):931–945, doi:10.1175/JHM-D-15-0075.1.

6 Aerosols and Clouds

Authors: Po-Lun Ma (PNNL), Paloma Borque (PNNL), Pavlos Kollias (Stony Brook Univ./BNL), Salil Mahajan (ORNL), Sam Silva (PNNL), Adam Varble (PNNL), Yunyan Zhang (LLNL)

6.1 Grand Challenges

Aerosol and cloud processes (Figures 6-1, 6-2) are extremely complicated and not well understood. Biases and uncertainties in predicting aerosol and cloud properties and processes are the primary sources of uncertainty in long-term projections of global temperature and precipitation. Reducing these uncertainties to provide more robust predictions of temperature and precipitation, both globally and regionally, is a foremost grand challenge in climate science, and AI/ML can be a powerful tool that provides a viable path forward. Three grand challenges for using AI/ML to improve the understanding and predictability of aerosols and clouds have been identified.

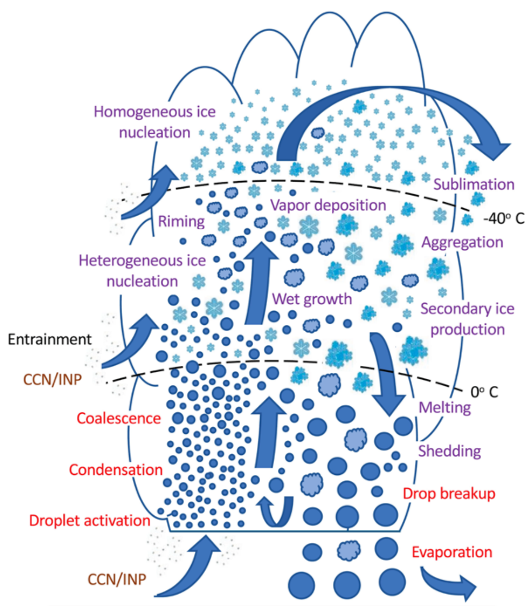


Figure 6-1. Schematic illustration of cloud microphysical processes within a typical cumulonimbus cloud. Specific microphysical processes are listed in red (involving only liquid drops) and purple (involving ice particles only or both liquid and ice). Cloud droplet activation occurs on aerosol particles serving as cloud condensation nuclei (CCN) in supersaturation conditions; cloud droplets then grow by condensation. Further growth by collision-coalescence produces raindrops. Above the 0°C level, there is heterogeneous ice nucleation on aerosols serving as ice nucleating particles (INPs). Ice particles grow by vapor deposition and riming (i.e., accretion and freezing of supercooled drops). If riming is especially heavy, not all of the collected liquid water freezes onto the ice particles and some is shed, representing wet growth. Above approximately the −40°C level, homogeneous ice nucleation can generate additional ice particles. Sublimation of ice particles detrained from the cloud occurs in subsaturated conditions. Ice crystals can grow by aggregation when they collide and stick together. Secondary ice production, not associated with heterogeneous or homogeneous ice nucleation, can generate more ice particles. Below the 0°C level, ice particle melting generates raindrops, and shedding of meltwater occurs for some ice particles. Raindrop collision-coalescence produces larger drops, while raindrop breakup produces smaller ones. Below cloud base, evaporation of falling raindrops occurs in subsaturated air (Source: Reproduced from Morrison et al. 2020 under Creative Commons CC BY 4.0).

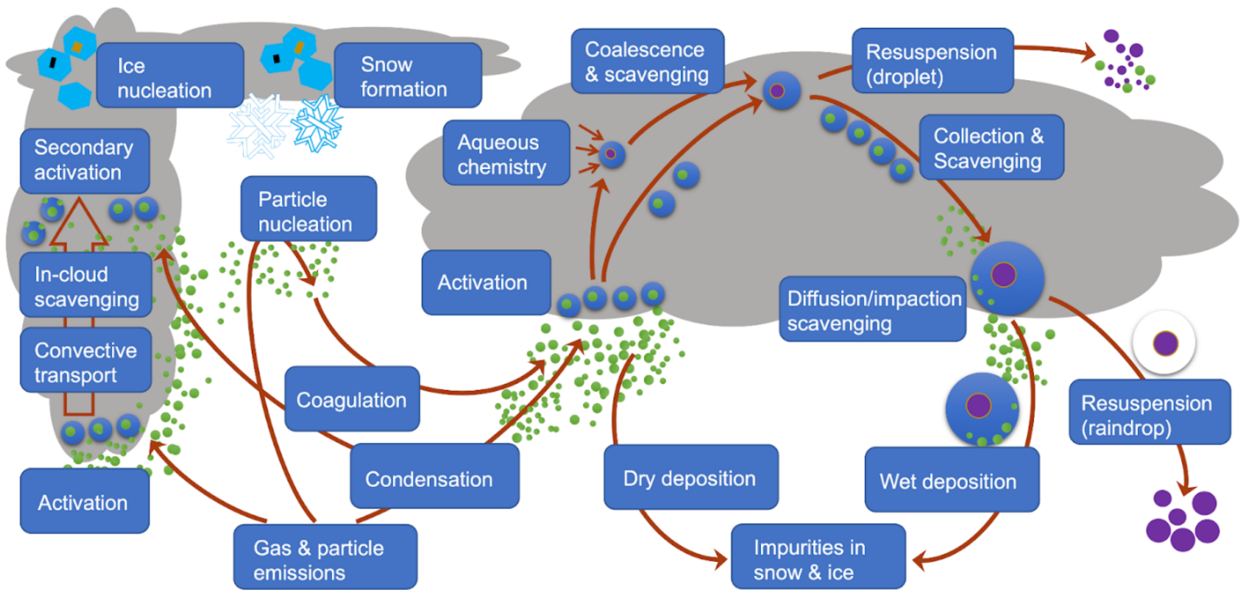


Figure 6-2. Schematic diagram of aerosol-related processes, denoted by blue boxes in clear air, stratiform cloud (shaded area on the right), convective cloud (shaded area on the left), or ice cloud (shaded area on the upper left) (Source: Figure reproduced and adapted from Wang et al. 2020 under Creative Commons CC BY 4.0).

6.1.1 Insufficient Data

Many important aerosol and cloud properties and processes remain poorly characterized. These include but are not limited to new particle formation; aerosol growth and activation; cloud processing and wet scavenging of aerosols; cloud droplet, rain, and ice nucleation and growth; secondary ice formation; convective circulations; and turbulent cloud entrainment/detrainment, cloud transitions, and organizations. Parameterizing these processes in climate models requires a combination of theory, process models, and measurements of aerosol and cloud properties such as size distributions across the range of atmospheric conditions covering Earth that influence and interact with them. However, measurements and high-resolution process models are limited, both in sampling and uncertainty, which presents a hurdle for producing sufficient extensive and high-fidelity datasets for AI/ML purposes. Aerosol and cloud studies that are based on a single field campaign or a small set of process model simulations do not represent the worldwide spectrum of aerosol, cloud, and meteorological regimes (Kogan and Ovchinnikov 2020; Possner et al. 2020). Specific cloud regimes and transitions between them are often critical to climate prediction including the stratocumulus to trade cumulus transition, high-latitude mixed-phase clouds, and deep convection; but these are not sampled equally, while extremes are sampled even less. Many collocated observations are required for isolating and quantifying aerosol-cloud-precipitation interactions due to confounding factors and feedback that operate across a range of scales. Moreover, uncertainties and deficiencies in process models (such as numerical diffusion in the bin microphysics approach) can also lead to biases. Both simulation and observational datasets are currently insufficient in number and often have uncharacterized uncertainties.

6.1.2 Model Calibration and Uncertainty Quantification

Studies have shown that the predictions of aerosols, clouds, effective radiative forcing (ERF) associated with anthropogenic aerosols (ERFaer), and cloud feedback are sensitive to model parameter settings. However, best practices for this resource-intensive procedure have not been established. Deriving signals from the nonlinear system, such as decoupling aerosol signals from meteorological co-variability, disentangling large-scale controls, and aerosol-cloud interactions (ACI), etc., is challenging. The effectiveness of existing emergent constraints is unclear (Schneider et al. 2017; Schlund et al. 2020), so continuous efforts on the development of process-oriented constraints as calibration targets are desirable. Furthermore, as Earth system models become more complex, the dimensionality of the parameter space also grows significantly. The scientific community has not established a prioritization of metrics for calibration targets.

6.1.3 Extreme-scale Separation

Processes affecting aerosols, clouds, and their interactions with the Earth system span a vast range of scales from 10^{-9} to 10^{6} m. Finding the right technology and data to bridge these scales is a major challenge. A hierarchy of models exists to resolve different portions of this vast scale spectrum from DNS to LES, CPM, LAM, and RRMs with limited domain sizes before reaching ESMs with global coverage. However, the workflows and best practices needed for bridging these models and scales are unclear. Similarly, observations vary in resolved scale and spatiotemporal coverage, but the ideal methods for overcoming scale mismatches to connect a range of different datasets are also unclear. While promising process-oriented test cases at select scales exist including applications of ML, the challenge remains of how to implement those results into global ESMs in a way that does not degrade other aspects of ESM predictions.

6.2 State of the Science

There have been many successful applications of AI/ML in understanding and improving the predictability of aerosols and clouds. Below we list five research areas.

6.2.1 Surrogates and Emulators

Plenty of studies have demonstrated success in applying AI/ML approaches to develop emulators for aerosol or cloud processes, including chemistry (Kelp et al. 2020; Keller and Evans 2019), aerosol activation (Silva et al. 2021), aerosol mixing (Zheng et al. 2021), and the warm rain process (Chiu et al. 2021; Gettelman et al. 2021). These emulators are developed to improve the accuracy, computational performance, or both. DNNs have been demonstrated to be more skillful than traditional, physically based parameterizations (Figure 6-3). Additional work has been done exploring the role of maintaining physical consistency when applying machine learning methods (Sturm and Wexler 2020; Beucler et al. 2021).

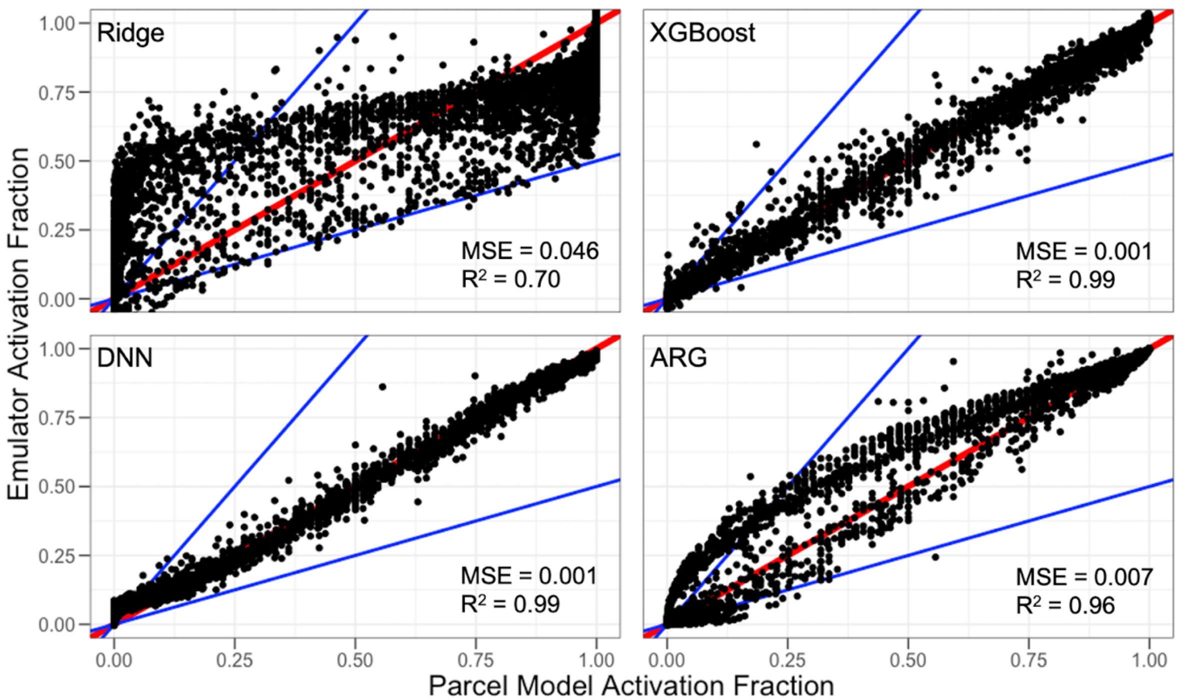


Figure 6-3. Scatterplot comparisons of the three physically naïve machine learning emulators using ridge regression, XGBOOST, and deep neural network (DNN) and the Abdul-Razzak and Ghan (2000) scheme (ARG) predicted activation fraction with the detailed parcel model. The 1:1 line is in red, and the blue lines represent a factor of 2 difference. Performance statistics are given in each panel
(Source: Reproduced from Silva et al. 2021 under Creative Commons CC BY 4.0).

6.2.2 Bridging Spatial Scales

Model hierarchies consisting of numerical models of varying resolution have been used by the research and operational communities for providing predictions across scales. Modern AI/ML techniques for coarse graining (Bretherton et al. 2021) and downscaling (Sachindra et al. 2018) show promise for bridging such scales.

6.2.3 Feature/Signature Detection and Causal Inference

Multiple studies have highlighted the importance and usefulness of interpretable AI (IAI), explainable AI (XAI), feature/signature detection, and causal inference techniques in climate and weather science (Barnes et al. 2020; McGovern et al. 2019; Toms, Barnes, and Ebert-Uphoff 2020; Tao Zhang et al. 2021). These methods can be used to identify indicator patterns of forced changes and emergent properties of the real and simulated climate system. These indicator patterns and emergent properties provide a path toward knowledge discovery, understanding what the AI/ML learned, and revealing missing mechanisms.

6.2.4 Model Optimization and Uncertainty Quantification

AI/ML approaches have shown promise in correcting model biases with respect to observations or a high-fidelity model simulation, optimizing model fidelity (Watt-Meyer et al. 2021; Kennedy and O'Hagan 2001; Couvreux et al. 2021; Hourdin et al. 2021; Zhang et al. 2015; Zhang et al. 2018; Xu et al. 2018; Cleary et al. 2021). Emulating a complex model's parameter sensitivities following human-constructed trial simulations have been used to aid model calibration and uncertainty quantification.

6.3 Experimental, Data, and Modeling Opportunities

6.3.1 Developing Representative Datasets for AI/ML Applications

This requires more extensive, accurate aerosol and cloud microphysics datasets consisting of in situ measurements, surface and satellite-based remote sensing retrievals, and high-fidelity process model simulations. Uncertainties associated with those datasets need to be quantified and cataloged. Synthesizing different datasets to overcome data gaps is essential for building representative datasets that can sufficiently apply to a range of climate scenarios. In addition, data collected for natural experiments such as volcanic eruptions and reduced socioeconomic activities due to COVID can provide insights into impacts of low-frequency or extreme events. Targeted observations are helpful for closing critical data gaps by identifying sensitive regions, collecting high-resolution/high-fidelity data, and designing the campaigns. Another good application of AI/ML is to improve retrievals because they are quasi-linear operators that can be inverted and used for data assimilation.

Process models can be used to generate high-fidelity data. For example, explicit particle-resolved models (Riemer et al. 2009, 2010) can be used to generate data for aerosol microphysics and chemistry. Bin models (Khain et al. 2015; Tzivion [Tzitzvashvili], Feingold, and Levin 1987; 1989; Feingold, Tzivion [Tzitzvashvili], and Leviv 1988) and the new Lagrangian-based super droplets approach (Grabowski et al. 2019) are useful for providing detailed cloud microphysics information. Models using the 4-stream approximation (Liou, Fu, and Ackerman 1988) can be used to generate data for estimating the 3D radiative effects of aerosols and clouds. Optics models (such as Mie code) can provide information about aerosol and cloud radiative properties. Large eddy simulations (LES) may provide detailed turbulence structure and transport of mass, momentum, and energy in cloud and sub-cloud layers. Model simulation ensembles (with perturbed physics, emissions and forcings, or other model configurations) provide useful information about the projection uncertainties and sensitivities to the perturbations. These datasets can be generated from high-resolution LES or ESMs and be used for emulator development and for improving the understanding of predictability. Unsupervised variational autoencoder (VAE) compression can be applied to mine the datasets for feature detection through the transformation of high-dimensional data into a quasi-Gaussian "latent-space."

6.3.2 Developing Emulators for Aerosol and Cloud Parameterizations in ESMs

Many existing aerosol and cloud parameterizations are based on empirical fits of observational or simulation data. Faster and more accurate emulators can be developed to replace or augment these parameterizations. Furthermore, the right level of complexity required in ESMs has not been systematically assessed. The ad hoc determination regarding, for example, the number of moments or the number of distinct separable sub-regimes used for parameterizing cloud microphysics leads to uncharacterized uncertainties. The appropriate architectures for different emulation purposes remain unclear. Unsupervised ML dimensionality reduction techniques such as VAEs applied to high-resolution data can be used to enhance our view of how many and which sub-regimes merit separate treatment in heuristic parameterization, ensuring the generalizability of the parameterizations. In the meantime, equation discovery and interpretable AI are important for emulator development to ensure physical interpretability. A verification and validation framework for emulators needs to be established. For processes without appropriate benchmark data, a Bayesian framework can provide a proper constraint. In addition to developing emulators for individual aerosol and cloud processes, developing emulators for systems can help us better understand and predict various feedback and adjustments by focusing on multivariate co-variabilities rather than effects of a single variable.

6.3.3 Model Calibration

The traditional manual calibration methods are resource-intensive and subjective, and their use makes it difficult to achieve local or global optimality in complex ESMs. While some techniques are established, such as emulating a complex model's parameter sensitivities following human-constructed trial simulations, end-to-end optimization that involves AI/ML in the model tuning process is not here yet. Yet deploying AI/ML-assisted calibration techniques online could significantly reduce the computational cost or enhance the quality of model calibration. Spatiotemporal Bayesian inference and multiagent reinforcement learning can efficiently link various timescales of the Earth system. These approaches also provide valuable information on the model uncertainty and limitations throughout simultaneous surrogate modeling. Instead of running expensive ESM simulations, surrogate models using AI/ML regression methods can be utilized to describe the relationship between the uncertain parameters and the output variables of an ESM to aid model calibration. Combining the data-driven tuning and domain knowledge from traditional manual tuning can help optimization algorithms converge effectively and analyze the mechanism of different local optimums.

6.3.4 Developing New Software Infrastructure to Seamlessly Incorporate AI/ML Approaches in the Modeling Framework

The software architecture for state-of-the-art ESMs is largely based on FORTRAN. However, AI/ML tool-chains are typically developed in a Python ecosystem. Efficient use of Python-based AI/ML trained networks in FORTRAN-based ESMs remains a challenge. Modelers have developed different approaches to address this gap. Libraries connecting the two software architectures are only a short-term solution. The software architecture needs to be redesigned to achieve computational performance portability for ESMs implemented with AI/ML methods to enable online training, calibration, and bias correction and to facilitate testing of process (emulator) splitting and coupling.

6.4 Research Priorities

The session participants suggested that the socioeconomic benefits and scientific uncertainties should be two important factors for determining priority science questions. Research resource allocation should be proportional to the scientific and socioeconomic impact. In addition, a set of use cases should be developed, such as (1) understanding and predicting aerosol emissions associated with wildfires (location, intensity, emission height, etc.); (2) exploring the influence of decarbonization on aerosol microphysics and chemistry; (3) characterizing the highly nonlinear chemical interactions; and (4) building faster, more accurate, and physically constrained emulators. Datasets for these use cases should be made accessible to the scientific community. The session participants also identified three priority research areas.

6.4.1 Data Compilation and Harmonization

Collecting and harmonizing data from different sources are critical. This includes data from ARM, satellites, detailed process-level and high-resolution-high-fidelity models, and other detailed data such as cloud chamber measurements. Data properties including the spatial and temporal scales they represent, as well as their uncertainty, strength, and weakness, need to be cataloged.

6.4.2 Improving the Representation of Aerosols and Clouds in ESMs

Many process representations can be replaced by emulators based on large training datasets that contain all variables. With increasing emulator development efforts, a framework for identifying good architecture of the AI model for different data needs to be established. A procedure for systematically evaluating the performance, trustworthiness, and generalizability of the emulators is critical. In addition to verification and validation techniques in the data science community, representative spatiotemporal information and process-level metrics as emergent constraints developed by the climate science community should also be used. For highly uncertain processes that do not have sufficient benchmark data, Bayesian representation of uncertainties can be helpful. Surrogates for the system, rather than individual processes, should also be developed to efficiently and systematically assess system response to perturbations.

6.4.3 Improving Understanding Using XAI and IAI

New XAI and IAI techniques have transformed the AI/ML applications from black box to identifying intrinsic properties and relationships of the data. These techniques are helpful in finding variables that are driving a system and can be used for feature detection and signature identification. These techniques can yield insights to aerosol and cloud scientists so they can provide physical interpretation. New insights can further drive new knowledge discovery and next-generation model development.

6.5 Short-term (<5 years), 5-year, and 10-year Goals

The short-term goals are to establish a set of good and bad practices. At this stage, the community should carry out a large number of exploratory projects to accumulate experience and establish best practices. These efforts will include:

1. Exploration of different architecture for building emulators for well-defined aerosol and cloud processes.
2. Application of XAI or IAI techniques for feature/signature detection to reduce ERFaci uncertainty in observations and models and to better isolate and quantify contributing factors.
3. Generation and collection of a large amount of observational and simulation data from high-resolution/high-fidelity models.

The 5-year goals are to complete the groundwork for using AI/ML to transform the prediction and understanding of aerosols and clouds. These efforts will include:

1. Data curation, harmonization, and uncertainty characterization.
2. Development of a framework for evaluating AI/ML techniques against observations, in addition to evaluating models against observations.
3. Next-generation model development driven by new knowledge discovered by AI/ML feature detection techniques.
4. Application of AI/ML for automated model calibration.
5. Targeted observations through AI/ML-assisted OSSEs.
6. Establishment of the standard for the AI/ML models to be more interpretable, transparent, and trustworthy.

The 10-year goals are to significantly improve the understanding and predictability of aerosols, clouds, and their roles in the Earth system through the integration of AI/ML techniques in the research and operational communities. These efforts will include:

1. Development of digital twins for science applications of significant impact. This includes identifying new mechanisms with AI/ML and exploring correlations in the system to find minimal subspaces or manifolds that could represent the full system efficiently.
2. Better understanding and model treatment of multiscale interactions.
3. Better characterization, and reduction, of model uncertainty including parametric, structural, and representative uncertainties.
4. Better representation of the system in data-limited regimes.
5. A movement toward machine reasoning where the AI/ML model trained from one dataset can be applied for another science application.

6.6 References

Barnes, Elizabeth A., Benjamin Toms, James W. Hurrell, Imme Ebert-Uphoff, Chuck Anderson, and David Anderson. 2020. "Indicator Patterns of Forced Change Learned by an Artificial Neural Network." *Journal of Advances in Modeling Earth Systems* 12 (9): e2020MS002195. https://doi.org/https://doi.org/10.1029/2020MS002195.

Beucler, Tom, Michael Pritchard, Stephan Rasp, Jordan Ott, Pierre Baldi, and Pierre Gentine. 2021. "Enforcing Analytic Constraints in Neural Networks Emulating Physical Systems." *Physical Review Letters* 126 (9): 98302. https://doi.org/10.1103/PhysRevLett.126.098302.

Bretherton, Christopher S., Brian Henn, Anna Kwa, Noah D. Brenowitz, Oliver Watt-Meyer, Jeremy McGibbon, W. Andre Perkins, Spencer K. Clark, and Lucas Harris. 2021. "Correcting Coarse-Grid Weather and Climate Models by Machine Learning from Global Storm-Resolving Simulations." *Earth and Space Science Open Archive*, 39. https://doi.org/10.1002/essoar.10507879.1.

Chiu, J. Christine, C. Kevin Yang, Peter Jan van Leeuwen, Graham Feingold, Robert Wood, Yann Blanchard, Fan Mei, and Jian Wang. 2021. "Observational Constraints on Warm Cloud Microphysical Processes Using Machine Learning and Optimization Techniques."

Geophysical Research Letters 48 (2): e2020GL091236. https://doi.org/10.1029/2020GL091236.
Cleary, Emmet, Alfredo Garbuno-Inigo, Shiwei Lan, Tapio Schneider, and Andrew M. Stuart. 2021. “Calibrate, Emulate, Sample.” *Journal of Computational Physics* 424 (January): 109716. https://doi.org/10.1016/J.JCP.2020.109716.
Couvreux, Fleur, Frédéric Hourdin, Daniel Williamson, Romain Roehrig, Victoria Volodina, Najda Villefranque, Catherine Rio, et al. 2021. “Process-Based Climate Model Development Harnessing Machine Learning: I. A Calibration Tool for Parameterization Improvement.” *Journal of Advances in Modeling Earth Systems* 13 (3): e2020MS002217. https://doi.org/10.1029/2020MS002217.
Feingold, Graham, Shalvn Tzivion (Tzitzvashvili), and Zev Leviv. 1988. “Evolution of Raindrop Spectra. Part I: Solution to the Stochastic Collection/Breakup Equation Using the Method of Moments.” *Journal of Atmospheric Sciences* 45 (22): 3387–99. https://doi.org/10.1175/1520-0469(1988)045<3387:EORSPI>2.0.CO;2.
Gettelman, A., D. J. Gagne, C. C. Chen, M. W. Christensen, Z. J. Lebo, H. Morrison, and G. Gantos. 2021. “Machine Learning the Warm Rain Process.” *Journal of Advances in Modeling Earth Systems* 13 (2): e2020MS002268. https://doi.org/10.1029/2020MS002268.
Grabowski, Wojciech W., Hugh Morrison, Shin-Ichiro Shima, Gustavo C. Abade, Piotr Dziekan, and Hanna Pawlowska. 2019. “Modeling of Cloud Microphysics: Can We Do Better?” *Bulletin of the American Meteorological Society* 100 (4): 655–72. https://doi.org/10.1175/BAMS-D-18-0005.1.
Hourdin, Frédéric, Daniel Williamson, Catherine Rio, Fleur Couvreux, Romain Roehrig, Najda Villefranque, Ionela Musat, Laurent Fairhead, F. Binta Diallo, and Victoria Volodina. 2021. “Process-Based Climate Model Development Harnessing Machine Learning: II. Model Calibration From Single Column to Global.” *Journal of Advances in Modeling Earth Systems* 13 (6): e2020MS002225. https://doi.org/10.1029/2020MS002225.
Keller, C. A., and M. J. Evans. 2019. “Application of Random Forest Regression to the Calculation of Gas-Phase Chemistry within the GEOS-Chem Chemistry Model V10.” *Geoscientific Model Development* 12 (3): 1209–25. https://doi.org/10.5194/gmd-12-1209-2019.
Kelp, Makoto M., Daniel J. Jacob, J. Nathan Kutz, Julian D. Marshall, and Christopher W. Tessum. 2020. “Toward Stable, General Machine-Learned Models of the Atmospheric Chemical System.” *Journal of Geophysical Research: Atmospheres* 125 (23): e2020JD032759. https://doi.org/https://doi.org/10.1029/2020JD032759.
Kennedy, Marc C., and Anthony O’Hagan. 2001. “Bayesian Calibration of Computer Models.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (3): 425–64. https://doi.org/10.1111/1467-9868.00294.
Khain, A. P., K. D. Beheng, A. Heymsfield, A. Korolev, S. O. Krichak, Z. Levin, M. Pinsky, et al. 2015. “Representation of Microphysical Processes in Cloud-Resolving Models: Spectral

(Bin) Microphysics versus Bulk Parameterization." *Reviews of Geophysics* 53 (2): 247–322. https://doi.org/10.1002/2014RG000468.

Kogan, Yefim, and Mikhail Ovchinnikov. 2020. "Formulation of Autoconversion and Drop Spectra Shape in Shallow Cumulus Clouds." *Journal of the Atmospheric Sciences* 77 (2): 711–22. https://doi.org/10.1175/JAS-D-19-0134.1.

Liou, Kuo-Nan, Qiang Fu, and Thomas P. Ackerman. 1988. "A Simple Formulation of the Delta-Four-Stream Approximation for Radiative Transfer Parameterizations." *Journal of Atmospheric Sciences* 45 (13): 1940–48. https://doi.org/10.1175/1520-0469(1988)045<1940:ASFOTD>2.0.CO;2.

McGovern, Amy, Ryan Lagerquist, David John Gagne, G. Eli Jergensen, Kimberly L. Elmore, Cameron R. Homeyer, and Travis Smith. 2019. "Making the Black Box More Transparent: Understanding the Physical Implications of Machine Learning." *Bulletin of the American Meteorological Society* 100 (11): 2175–99. https://doi.org/10.1175/BAMS-D-18-0195.1.

Morrison, H., et al. 2020. "Confronting the Challenge of Modeling Cloud and Precipitation Microphysics." *Journal of Advances in Modeling Earth Sys*tems 12 (8). https://doi.org/10.1029/2019MS001689.

Possner, Anna, Ryan Eastman, Frida Bender, and Franziska Glassmeier. 2020. "Deconvolution of Boundary Layer Depth and Aerosol Constraints on Cloud Water Path in Subtropical Stratocumulus Decks." *Atmospheric Chemistry and Physics* 20 (6): 3609–21. https://doi.org/10.5194/acp-20-3609-2020.

Riemer, N., M. West, R. A. Zaveri, and R. C. Easter. 2009. "Simulating the Evolution of Soot Mixing State with a Particle-Resolved Aerosol Model." *Journal of Geophysical Research: Atmospheres* 114 (D9): 9202. https://doi.org/10.1029/2008JD011073.

Riemer, Nicole, Matthew West, Rahul Zaveri, and Richard Easter. 2010. "Estimating Black Carbon Aging Time-Scales with a Particle-Resolved Aerosol Model." *Journal of Aerosol Science* 41 (1): 143–58. https://doi.org/10.1016/J.JAEROSCI.2009.08.009.

Sachindra, D. A., K. Ahmed, Md Mamunur Rashid, S. Shahid, and B. J. C. Perera. 2018. "Statistical Downscaling of Precipitation Using Machine Learning Techniques." *Atmospheric Research* 212 (November): 240–58. https://doi.org/10.1016/J.ATMOSRES.2018.05.022.

Schlund, Manuel, Axel Lauer, Pierre Gentine, Steven C. Sherwood, and Veronika Eyring. 2020. "Emergent Constraints on Equilibrium Climate Sensitivity in CMIP5: Do They Hold for CMIP6?" *Earth System Dynamics* 11 (4): 1233–58. https://doi.org/10.5194/esd-11-1233-2020.

Schneider, Tapio, Shiwei Lan, Andrew Stuart, and João Teixeira. 2017. "Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations." *Geophysical Research Letters* 44 (24): 12,396-12,417. https://doi.org/10.1002/2017GL076101.

Silva, Sam J., Po-Lun Ma, Joseph C. Hardin, and Daniel Rothenberg. 2021. "Physically Regularized Machine Learning Emulators of Aerosol Activation." *Geoscientific Model Development* 14 (5): 3067–77. https://doi.org/10.5194/gmd-14-3067-2021.

Sturm, Patrick Obin, and Anthony S. Wexler. 2020. "A Mass- and Energy-Conserving Framework for Using Machine Learning to Speed Computations: A Photochemistry Example." *Geoscientific Model Development* 13 (9): 4435–42. https://doi.org/10.5194/gmd-13-4435-2020.

Toms, Benjamin A., Elizabeth A. Barnes, and Imme Ebert-Uphoff. 2020. "Physically Interpretable Neural Networks for the Geosciences: Applications to Earth System Variability." *Journal of Advances in Modeling Earth Systems* 12 (9): e2019MS002002. https://doi.org/https://doi.org/10.1029/2019MS002002.

Tzivion (Tzitzvashvili), Shalva, Graham Feingold, and Zev Levin. 1987. "An Efficient Numerical Solution to the Stochastic Collection Equation." *Journal of Atmospheric Sciences* 44 (21): 3139–49. https://doi.org/10.1175/1520-0469(1987)044<3139:AENSTT>2.0.CO;2.

———. 1989. "The Evolution of Raindrop Spectra. Part II: Collisional Collection/Breakup and Evaporation in a Rainshaft." *Journal of Atmospheric Sciences* 46 (21): 3312–28. https://doi.org/10.1175/1520-0469(1989)046<3312:TEORSP>2.0.CO;2.

Wang, H., R. C. Easter, R. Zhang, P.-L. Ma, B. Singh, and K. Zhang, 2020. Aerosols in the E3SM Version 1: New developments and their impacts on radiative forcing. *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001851. https://doi.org/10.1029/2019MS001851.

Watt-Meyer, Oliver, Noah D. Brenowitz, Spencer K. Clark, Brian Henn, Anna Kwa, Jeremy McGibbon, W. Andre Perkins, and Christopher S. Bretherton. 2021. "Correcting Weather and Climate Models by Machine Learning Nudged Historical Simulations." *Geophysical Research Letters* 48 (15): e2021GL092555. https://doi.org/https://doi.org/10.1029/2021GL092555.

Xu, Haoyu, Tao Zhang, Yiqi Luo, Xin Huang, and Wei Xue. 2018. "Parameter Calibration in Global Soil Carbon Models Using Surrogate-Based Optimization." *Geoscientific Model Development* 11 (7): 3027–44. https://doi.org/10.5194/gmd-11-3027-2018.

Zhang, T., L. Li, Y. Lin, W. Xue, F. Xie, H. Xu, and X. Huang. 2015. "An Automatic and Effective Parameter Optimization Method for Model Tuning." *Geoscientific Model Development* 8 (11): 3579–91. https://doi.org/10.5194/gmd-8-3579-2015.

Zhang, Tao, Wuyin Lin, Andrew M. Vogelmann, Minghua Zhang, Shaocheng Xie, Yi Qin, and Jean-Christophe Golaz. 2021. "Improving Convection Trigger Functions in Deep Convective Parameterization Schemes Using Machine Learning." *Journal of Advances in Modeling Earth Systems* 13 (5): e2020MS002365. https://doi.org/https://doi.org/10.1029/2020MS002365.

Zhang, Tao, Minghua Zhang, Wuyin Lin, Yanluan Lin, Wei Xue, Haiyang Yu, Juanxiong He, et al. 2018. "Automatic Tuning of the Community Atmospheric Model (CAM5) by Using

Short-Term Hindcasts with an Improved Downhill Simplex Optimization Method." *Geoscientific Model Development* 11 (12): 5189–5201. https://doi.org/10.5194/gmd-11-5189-2018.

Zheng, Zhonghua, Jeffrey H. Curtis, Yu Yao, Jessica T. Gasparik, Valentine G. Anantharaj, Lei Zhao, Matthew West, and Nicole Riemer. 2021. "Estimating Submicron Aerosol Mixing State at the Global Scale With Machine Learning and Earth System Modeling." *Earth and Space Science* 8 (2): e2020EA001500. https://doi.org/https://doi.org/10.1029/2020EA001500.

7 Coastal Dynamics, Oceans, and Ice

Authors: William J. Pringle (Argonne), Matthew Hoffman (LANL), Steven Brus (Argonne), Paul J. Durack (LLNL), Carolyn B. Begeman (LANL), Veronica Nieves (Univ. of Valencia)

7.1 Grand Challenges

To tackle this question, we asked, "What are the Grand Challenges in Coastal Dynamics, Oceans, and Ice prediction that could be uniquely and transformatively addressed by AI/ML approaches?"

We have identified four Grand Challenges across our diverse session, as follows.

7.1.1 Grand Challenge 1 (GC1)

GC1 is to accurately represent multiple spatiotemporal scale processes across the ice-land-ocean system, which includes efforts to:

- Capture coastal, ocean, and cryosphere processes that span a wide range of interacting scales in space and time. Examples include eddy turbulence in the ocean affecting large-scale transport and distribution of sea ice floe size affecting large-scale rheology.
- Develop parameterizations that are scale-aware/valid for all model resolutions, including the use of stochastic subgrid-scale models.
- Resolve model physics appropriate for the question being addressed.
- Improve model/data ensemble use through ML (e.g., by making ensembles more efficient to capture the range of uncertainty in a coupled system).

7.1.2 Grand Challenge 2 (GC2)

GC2 is to accurately represent complex coupled ice-land-ocean systems and the nonlinear interactions between individual components in both observational systems and models, which includes efforts to:

- Address many critical coastal, ocean, and cryosphere processes that occur at interfaces, and these boundary effects have a large impact on the overall behavior of the system. Examples include freshwater fluxes affecting the stratification of the ocean, and glacier hydrologic processes on the surface (e.g., crevasse propagation) and bed (e.g., basal friction).
- Obtain a better understanding of sensitivities and uncertainty/bias propagation in a coupled system.
- Build connections between fully resolved process models.

- Describe interfacial fluxes, which are not well modeled at the current grid scales.
- Improve coupled model tuning.

7.1.3 Grand Challenge 3 (GC3)

GC3 is to address incompleteness in observed data and theory, which includes efforts to:

- Overcome the lack of data for the vast ocean and remote polar regions, which particularly suffer from challenges in data collection and data sparsity.
- Overcome incomplete/unrepresented physics.
- Use AI/ML to develop insights where data do not exist (short term).
- Use AI/ML to generate new platforms to fill holes where no data exist (long term).

7.1.4 Grand Challenge 4 (GC4)

GC4 is to improve the prediction of extremes, the identification of tipping points, and the influence of human actions, which includes efforts to:

- Incorporate marine and cryosphere processes, which often have strong nonlinearities and regime changes related to phase change and other threshold processes. Predicting extremes is exacerbated by the Grand Challenges listed above (multiscale, controls at interfaces, data sparsity).
- Recognize that traditional physics approaches may struggle; continuum simulations may be poor at extrema.
- Navigate the reality that the effects of human scenarios and future decisions are difficult to incorporate into predictions.

In addition, we developed a work cloud summarizing the phrases/words related to Grand Challenges expressed by workshop participants, shown in Figure 7-1.

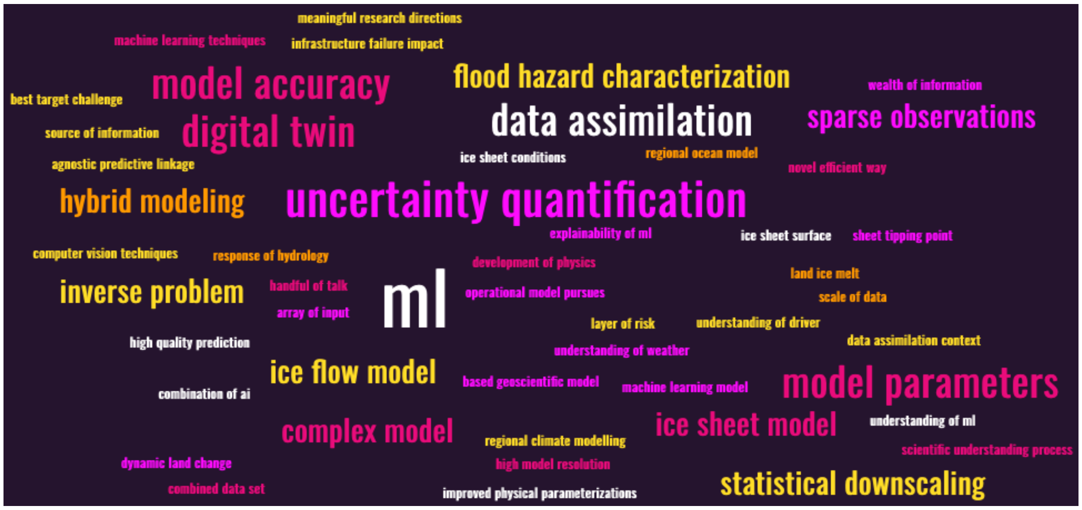


Figure 7-1. Word cloud indicating Grand Challenges that come to mind for session participants with regard to coastal dynamics, oceans, and ice.

7.2 State-of-the-Science

Next, given the Grand Challenges we identified, we asked, "What is the state-of-the-science for development and application of AI/ML approaches for these Grand Challenges?"

As shown in Figure 7-2, ML techniques that have been most commonly employed by workshop participants are neural networks, model emulators/surrogates, and data synthesis.

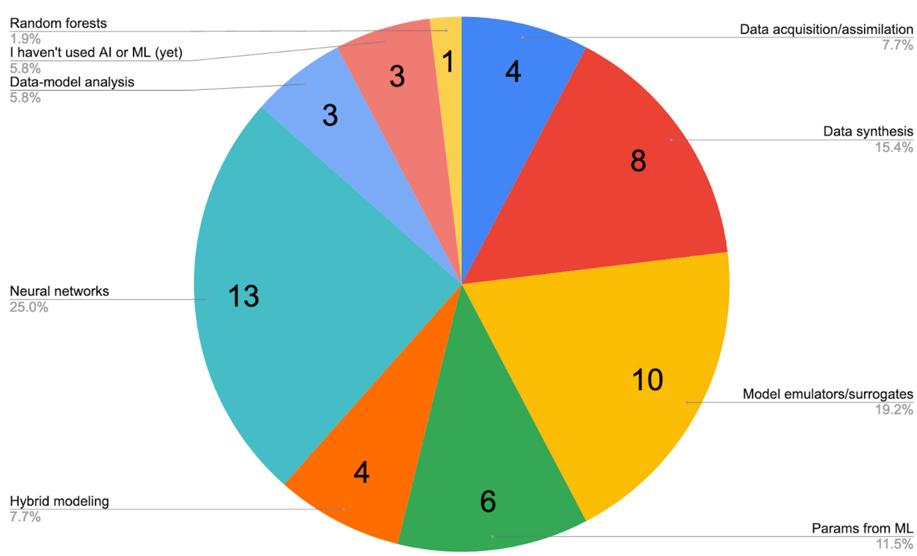


Figure 7-2. Summary of AI/ML techniques that the workshop participants have previously used.

7.2.1 Coastal Dynamics

Coastal dynamics encompasses multiple interacting systems and scales across sharp horizontal gradients in the coastal zone, including aspects such as ocean and lake dynamics, flooding, sediment transport and geomorphology, biogeochemistry (BGC), and human interactions. This complexity is hard to model and predict, and strongly motivates past and emerging ML applications.

Extreme coastal water level and flooding prediction are the most common areas of ML development in coastal dynamics. In particular, storm surge surrogate models have had wide attention using methodologies such as Gaussian process regression (Kriging), polynomial chaos, CNN, ANN, and SVR (Kajbaf and Bensi 2020; Jia et al. 2016; Kim et al. 2015; Kyprioti et al. 2021a; Lee et al. 2021; Plumlee et al. 2021; Sochala et al. 2020; Zhang et al. 2018). In these studies, tropical cyclones (TCs), which are responsible for the strongest surge events, are modeled through parametric models. This allows researchers to characterize TCs through a small set of parameters and thus form a map from these inputs to the modeled surge. Emerging areas for ML-based coastal flooding include: incorporating sea-level rise (Kyprioti et al. 2021b), coupling to and including rainfall and hydrological processes for compound flooding prediction (Bass and Bedient 2018; Li, Kiaghadi, and Dawson 2021), and using more advanced ML techniques such as deep learning (Tiggeloven et al. 2021) and Fourier Neural Operators (Jiang, et al. 2021). An additional component to sea levels and coastal flooding originates from wind waves and associated wave setup, runup, and overtopping. Wind wave models, especially so-called third-generation models, are however very computationally expensive as they are solved in five dimensions. One study drastically reduced this cost using MPL ANNs and SVM classification methods for predicting significant wave heights and characteristic periods (James, Zhang, and O'Donncha 2018).

In the area of coastal sediment transport, suspended sediment concentrations and fluxes have been predicted using ANN and Boosted Regression Trees (BRTs) (see Goldstein, Coco, and Plant 2019 for a review). Settling velocities have been estimated using Random Forest (RF) techniques (Cao et al. 2020). In morphodynamics, sandbars tend to be studied using ANN, while Bayesian Networks (BNs) are often applied to shoreline and dune erosion (Goldstein, Coco, and Plant 2019). Probabilistic ML is important for the highly uncertain nature of coastal morphodynamics, an example of which was using Gaussian processes for prediction of wave runup for dune erosion estimates (Beuzen, Goldstein, and Splinter 2019).

BGC models encompass a whole range of chemical kinetics from biological to gas phase. These include stiff reaction systems and often biological reactions that are either too slow or very fast and poorly understood. Developing surrogate models for this entire reaction chain that are an order of magnitude faster will allow us to include these multi-timescale kinetics in larger models (regional and global) and explore the feedbacks/nonlinearities fully. The typical approach to deal with stiffness and a wide range of timescales in reaction kinetic networks is to use different time integrators for various parts of the timescales of the problem. With AI/ML we can do something similar using neural ordinary differential equation (ODEs). An emerging topic is data analysis and surrogate modeling of coupled water, sediment, and nutrient fluxes, as well as dissolved organic carbon (DOC) dynamics using satellite data in tidally influenced wetland–estuarine systems (Cao and Tzortzio 2021).

7.2.2 Ocean

The global ocean covers 71% of Earth's surface, and yet it is one of the least understood and least mapped domains in the climate system. The open ocean directly interacts with the sub-domains of coastal, cryosphere, and the atmosphere. Observational ocean measurements are spatially and temporally sparse and mostly limited to the surface, and few continuous measurements span more than several decades. The timescales relevant to accurate measurements span seconds to millennia, and spatial scales of importance range from micro with relevance of ocean turbulence, to synoptic with relevance to fronts and wave propagation. The complexity of ocean observation and modeling strongly motivates developers of past and emerging ML applications to improve their understanding and predictability.

For the open ocean, AI/ML has been used for observational analysis in addition to modeling applications. For observational applications, examples include:

- Data fusing using satellite and in situ data products for biological and physical insights (e.g., Castellani 2006; Chapman and Charantonis 2017; Denvil-Sommer et al. 2019; Duncan et al. 2019; Kavanaugh et al. 2016; Martinez et al. 2020; ben Mustapha et al. 2014) or combining data from disparate satellite platforms (e.g., Guimbard et al. 2012).
- Quantifying deep ocean currents (Manucharyan, Siegelman, and Klein 2021) and heat fluxes (George, Manucharyan, and Thompson 2021) from satellite altimetry.
- Deciphering three-dimensional North Atlantic ocean circulation, and ocean dynamical regions from an ocean state estimate (Sonnewald, Wunsch, and Heimbach 2019) in addition to numerous ocean model simulations (Sonnewald and Lguensat 2021).
- Estimating global ocean heat content from tidal satellite observations (Irrgang, Saynisch, and Thomas 2019), predicting Indian Ocean Dipole events (Ratnam, Dijkstra, and Behera 2020), and identifying the ENSO state from sea surface temperature (SST) inputs, along with predicting near surface temperature on land from SSTs (Toms, Barnes, and Ebert-Uphoff 2020).

In addition, leveraging unsupervised ML methods has facilitated new objective approaches that complement conventional classification techniques, including: developing physical insights as a function of vertical ocean temperature/stratification in the North Atlantic (Maze et al. 2017), Southern Ocean (Jones et al. 2019), mid-latitude and equatorial Pacific (Houghton and Wilson

2020), subpolar North Atlantic (Desbruyères, Chafik, and Maze 2021), and Southern Ocean mixing off the Kerguelen Plateau (Rosso et al. 2020) from Argo profile data. This has also resulted in developing an understanding of the pathways and variability of modified Circumpolar Deep Water in the Amundsen Sea (Boehme and Rosso 2021), along with biogeochemical insights, such as mapping subsurface oxygen (Giglio, Lyubchich, and Mazloff 2018), CO_2 fluxes (Bushinsky et al. 2019), and silicate and phosphate (Park et al. 2021) in the Southern Ocean, along with estimates of global atmosphere-ocean CO_2 fluxes (Landschützer et al. 2014; Watson et al. 2020).

The application of AI/ML to ocean modeling is less developed than its observational counterpart. Some preliminary applications include data-driven meso/submesoscale eddy parameterizations to better anchor parameterizations from data, rather than idealized theories (Zanna and Bolton 2020; Bolton and Zanna 2019), parameterizing unresolved mesoscale ocean dynamics using deep learning (Guillaumin and Zanna 2021), reconstructing an index of the Atlantic Meridional Overturning Circulation (AMOC; DelSole and Nedza 2021), and ascertaining the predictive sensitivity of September sea ice across ten representative climatological quantities in a coupled climate model (Nichol et al. 2021).

There are numerous opportunities for AI/ML to tackle ocean science challenges. Like the cryosphere, the poor temporal and spatial sampling of the global ocean lends itself to being a rich opportunity for ML-assisted data fusion and exploration (GC3). The role of climate extremes (GC4) and their role as drivers of climate impacts are another area where AI/ML could aid discovery, deriving insights from crude-resolution ocean model data that poorly resemble the Earth system. In addition, we can leverage ML to quantitatively guide the development and updating of parameterizations directly from observed data, rather than the conventional approach, using idealized theories.

Figure 7-3 shows how ML has started to converge with research developments in oceanography (top), while the lower box shows the ML process with oceanography components assigned to each ML step.

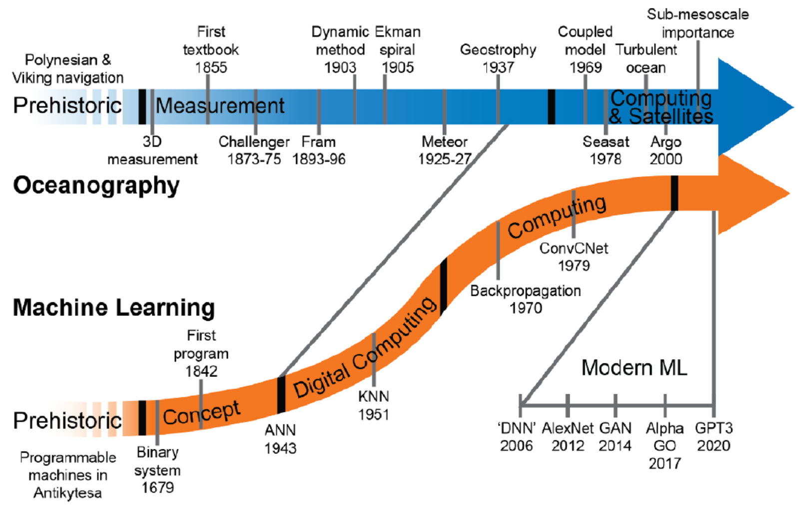


Figure 1. Timeline sketch of oceanography (blue) and ML (orange). The timelines of oceanography and ML are moving towards each other, and interactions between the fields where ML tool as are incorporated into oceanography has the potential to accelerate discovery in the future. Distinct 'events' marked in grey. Each field has gone through stages (black), with progress that can be attributed to the available tools. With the advent of computing, the fields were moving closer together in the sense that ML methods generally are more directly applicable. Modern ML is seeing an very fast increase in innovation, with much potential for adoption by oceanographers. See table A1 for acronyms.

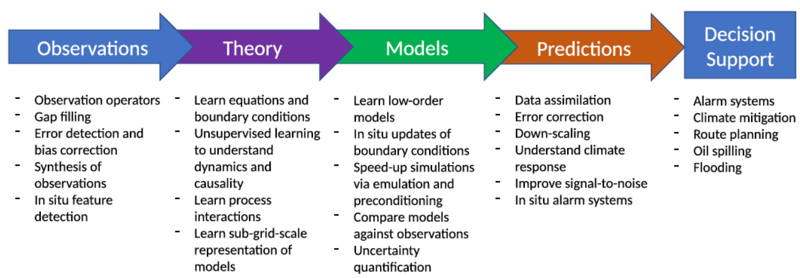


Figure 2. Machine learning within the components of oceanography. A diagram capturing the general flow of knowledge, highlighting the components covered in this review. Separating the categories (arrows) is artificial, with ubiquitous feed-backs between most components, but serves as an illustration.

Figure 7-3. Convergence of oceanography research with the emergence of ML (top); and ML applications within components of oceanography (bottom) (Source: Reproduced from Sonnewald et al. 2021b under Creative Commons CC BY 4.0).

7.2.3 Sea-Level Prediction

Sea-level changes occur for many reasons and on a range of timescales, from fast changes, such as storm surges or tides, to slow changes linked to persistent natural processes like decadal ocean temperature variations. The influence of these factors also varies around the globe, where other local effects (e.g., subsidence or variations in land height) can also affect sea-level rising, making it harder to estimate regional sea-level changes at specific locations. Developing more accurate predictions will bring huge benefits for the world population living within 100 kilometers of the coast – which represent 40% of all territories (Stocker et al. 2014). We also need a diverse set of tools with different informational needs to help planners and decision makers identify key regions vulnerable to sea-level changes. Among all available tools and techniques, the application of ML to this topic has been largely underexplored, even though it seems to be particularly suitable for capturing nonlinear complex relationships. In this context, nearshore and short-timescale processes (e.g., extreme coastal sea levels) have been the primary target of ML efforts (Sztobryn 2003; Bajo and Umgiesser 2010; French et al. 2017), as discussed in section 7.3. More recently, the impact of offshore (natural) processes, such as internal climate variability, was examined through observational ocean temperature data (Nieves, Marcos, and Willis 2017) and ML techniques (Nieves, Radin, and Camps-Valls 2021) to predict regional coastal sea levels on timescales from 1 to 3 years. Input for short-term predictions may be obtained as well through statistical ML downscaling of climate models (Sithara, Pramada, and Thampi 2021). On longer timescales, ML can also be used to fill data gaps or identify outliers in the dataset to aid in the evaluation of past sea-level variability (Hieronymus, Hieronymus, and Hieronymus 2019; Radin and Nieves 2021).

7.2.4 Ice

In the domains of sea ice and land ice, AI/ML has been used both in interpreting observations and improving models. Within observational applications, automated feature classification from satellite imagery is a common application. For sea ice, this has been applied to sea ice ridges, surface melt ponds, and open ocean leads, for example, using convolutional neural networks (Reinisch et al. 2021). Land ice applications also include morphological and hydrologic features, including crevasses and fracture and surface ponds, lakes, and streams (e.g., Lai et al. 2020). To date, these applications have generally been performed using single satellite sensors, and analysis of multiple coincident sensors or satellite sensors with other types of airborne or in situ measurements remains a challenging, but promising, direction. Variations in spatial and temporal scales (GC1) and data sparsity associated both with in situ data as well as satellite repeat intervals (GC3) are also difficult, although AI/ML approaches have made some progress in these areas where traditional remote sensing studies have struggled (e.g., Braakmann-Folgmann and Donlon 2019).

In modeling applications, AI/ML has been applied most to learning parameterizations for unresolved processes and for emulation for uncertainty quantification. Sea ice parameterization applications have included wave-ice interactions (Horvat and Roach 2022). Land ice applications have been most focused on representing fracturing and calving, but are also starting to consider surface mass balance processes in snow and ice and subglacial hydrology. Uncertainty quantification applications have included ranking parameter space sensitivity for stand-alone sea-ice models and building sea-level emulators from large-scale ice sheet models (Urrego-Blanco et al. 2016).

Ice sheet contributions to regional sea level constitute a highly nonlocal phenomenon, through glacial isostatic adjustment and changes to the gravitational field with impacts on the sea-level fingerprint and tides. ML has been used primarily to statistically emulate ice sheet mass loss scenarios to constrain uncertainty, as mentioned above. However, the differential regional sea-level rise associated with ice loss from Greenland and Antarctica remains a challenge to ML approaches to sea-level predictability on decadal timescales.

7.3 Experimental, Data, and Modeling Opportunities

We next asked, “What experimental, data, and modeling opportunities exist for advancing development and use of AI/ML approaches for these Grand Challenges?”

In the various breakout groups, there was an emphasis on advancing the ModEx approach of using data to improve models, and using models to inform data gathering and experiments, in a repeating cycle. For instance, models can be continuously updated with real-time data paired and continuously interrogated and tested. Different ways AI/ML approaches could be Integrated with the ModEx approach include by:

- Comparing models to data in order to understand underlying processes and improve model parameterizations and therefore their predictive capabilities (using better physics or AI/ML).
- Assuming that models are skillful/perfect, using AI/ML techniques to improve our prior knowledge of environmental variables (e.g., data assimilation).
- Using models to improve observational data by removing effects of other unwanted environmental variables and similarly removing contamination from observations that are not relevant to models.
- Using ML to link observations with process modeling.

7.3.1 Coastal Dynamics

The main limitations on coastal data are twofold: the lack of availability of consistent and continuing time series of coastal measurements and a lack of well-distributed observations. Having more co-located, multi sensor measurements to be combined with satellite imagery, NWP, and reanalysis products would be most helpful (e.g., by using 4D CNN ML methods) and could include, for example, nearshore wave measurements and current profilers, or reflectors in marshes to enable InSAR analyses of these sensitive areas. In the case of extreme flooding events, it is important to have detailed damage and recovery information to test model predictions and develop ML approaches that go beyond the empirical damage curve approach.

In the BGC community, measurements to figure out the dynamic and heterogeneous areas are required, including landscape and spatial structures. For example, as tides ebb and flow, they create a dynamic environment for microbes to respond along with the plants, and measurements targeting this area and processes are needed. Similarly, observations on the arctic permafrost are lacking at all relevant scales, and there is very limited sampling. The state of the permafrost is that of extreme heterogeneity and in scales beyond the reach of current physics-based models, motivating the need for more data to develop alternative AI/ML modeling approaches and to reduce the uncertainty and accuracy of the models. ML approaches for optimization of data placement sensors, to better tap into remote sensing imagery and to fill gaps in data (e.g., on the permafrost) could be transformative to improving data coverage and quality.

Modeling opportunities include wave-current interaction and the nonlinear interaction of sea level, tide, surge, and wave-setup interaction for coastal flooding, especially across large regions as opposed to point-based output. Also, incorporating inland hydrological processes with ocean processes (coupling watershed ML models with coastal surge models) is a big opportunity to improve both climate modeling and extreme flood prediction under a future climate of wetter storms. These nonlinear coupled processes are challenging to model using physics-based approaches across all scales, and ML approaches that excel in mapping nonlinear high-dimensional outputs will be useful. ML could be applied for testing where certain processes matter (e.g., where wave setup is an important contribution to total water levels that could cause coastal flooding). Furthermore, for emulators involving human systems and couplings, there may also be lessons from mechanistic surrogates to bring in dynamical knowledge. For example, response models are used for SLR, where (assuming linearity) you can predict the response to any forcing, without knowing in advance what those forcings might look like.

Furthermore, in exposure modeling, usage of AI/ML approaches could be in generating synthetic data to develop and represent built and environmental infrastructure. Using graph-based algorithms and ML models, there are opportunities to create power outage modeling and combine these with hazard and impact modeling to capture the actual risk. This helps decision makers to have necessary information. Impact modeling can be developed with the available data. Besides this, usage of linear models such as logistic regression can be used, along with new datasets from customers to then apply transfer learning techniques for use of models across different regions of interest.

Developing improved parameterizations for physics-based models that are based on idealized lab experiments are necessary. Such idealized parameterizations are ubiquitous in BGC models and have not been deeply scrutinized. The same argument can be made for turbulence closure and boundary layer stress models in hydrodynamic models that are derived from idealized fluid dynamics experiments. The use of more versatile and scrutinizable ML/hybrid ML methods in model parameterizations could provide opportunities for large modeling advances and help to ensure that these parameterizations preserve multiscale features across the coastal zone.

7.3.2 Ocean and Sea Level

There are numerous opportunities for AI/ML to make progress on ocean science challenges. Like the cryosphere, the poor temporal and spatial sampling of the global ocean lends this to be a rich opportunity for ML-assisted data fusion and exploration (GC3). In particular, using unsupervised ML techniques may enable new insights to be extracted from existing single or composite datasets, elucidating new relationships and new opportunities for predictability (Sonnewald, Wunsch, and Heimbach 2019). There may also be opportunities to leverage ML to extract information from sparse and "messy" unmanaged observational data such as those from individual biologically focused research campaigns. The role of climate extremes and in particular marine heatwaves for the ocean (GC4) and their role as drivers of climate impacts are another area where AI/ML could aid discovery, deriving insights from crude-resolution ocean model data that poorly resemble the Earth system. In addition, leveraging ML to quantitatively guide the development and updating of parameterizations – generated directly from observed data – rather than the conventional approach using idealized theories is a rich opportunity (e.g., Zanna and Bolton 2020). There is also an opportunity to leverage ML to replicate current PDE-based models, validating and reproducing existing capabilities, while potentially leveraging efficiency gains provided by emerging hardware that does not scale well with legacy Fortran libraries.

7.3.3 Ice

7.3.3.1 Sea Ice

Perhaps the greatest possibilities for AI/ML in sea ice prediction exist in simulating small-scale morphological features and in accelerating codes, thereby allowing larger and more physically realistic ensembles simulating the sea ice state in Earth system models. A great volume of synthetic aperture radar (SAR) data is expected to become available in the coming decade to support this development, which will be useful for characterizing the upper surface of sea ice to aid the development of models. However, data sparsity will continue to be a challenge for many aspects of sea ice, especially for characterizing the submerged ice surface, which will continue to be limited by submarine track and mooring measurements. This also poses an opportunity for machine learning. Within the available model toolset of the coming decade, new classes of basin-scale Lagrangian sea ice models are emerging for synoptic to centennial prediction (e.g., Turner, Peterson, and Bolintineanu 2022), offering the potential for machine learning to develop improved sea-ice element contact models based on observations and high-resolution floe-resolving simulations. It is also possible that nonlocal physics-informed Neural Networks (e.g., Pang et al. 2020) may better capture aspects of sea ice rheology than traditional PDE-based continuum models. As part of improvements in sea ice dynamics, thermodynamics, and morphology in sea ice models, feature shifts between near-coincident imagery sources (e.g., SAR, MSI, etc.) would expand the availability of data for ML problems as well as the potential for quantitative comparison between imaging modalities. With increases in model resolution and large increases in observational data, AI approaches could have a significant impact by discovering patterns in sea ice evolution, revealing connections between Earth system processes associated with the atmosphere-ice-ocean boundary layer. There is also potential for development of surrogates or emulators that could be used to more efficiently to create large ensembles to investigate uncertainty in predictions. Targeted AI/ML applications to individual problems in sea ice simulation may offer the most scientifically sound path to explainable outcomes.

7.3.3.2 Land Ice

Experimental, data, and modeling opportunities in land ice tend to be less mature than in other areas, but there has been tremendous improvement in recent years. Observational data for glaciers and ice sheets in particular suffer from data sparsity in space and time exacerbated by the slow response time of many glacier processes (decades to millennia). Due to the relative lack of data, neural networks hold promise, but non-physically informed neural networks will be more difficult to evaluate. A proliferation of satellite and airborne data products for ice velocity, thickness, thinning rate, and surface conditions has made ice-sheet-wide products for each of these quantities available at increasing temporal frequency. The relative uniformity and frequency of remotely sensed data products makes them the best target for AI/ML applications. In situ observations are significantly more time-consuming and resource-intensive to obtain and suffer from more problems with data availability and standards. However, these methods represent the only observations of englacial and subglacial conditions, which are critical for ice sheet evolution. There are few experimental approaches being used in glaciology, but there has been a slight resurgence of these methods, and they offer unique opportunities for AI/ML applications. Laboratory experiments of basal and ice rheology using ring-shear and other mechanical devices provide direct measures of poorly understood processes that are often crudely parameterized in large-scale models and theory. Glaciological models range from process-scale, to full-continent ice-sheet models, to Earth system models. Linkages between these levels of complexity are relatively few in glaciology, and AI/ML could accelerate establishing connections between these efforts. Transitioning to software frameworks that support both physical models and AI/ML would help accelerate these connections.

7.4 Research Priorities

Given these active research focus areas, we asked: "What research must be conducted as next steps for addressing the Grand Challenges, and what are the research priorities?"

We identified four overarching research priorities for addressing the Grand Challenges, as follows.

7.4.1 Improving Data Standardization and Consistency and Exploring How AI/ML Can Be Used to Address Sparse Data Coverage and Merge Disparate Datasets

Workshop participants across all breakout groups often brought up the challenges of the lack or poor data standards, data inconsistencies among datasets, and low machine-readability of the data (GC3). As a result, most of the work when using programs like TensorFlow becomes interfacing with data input/output. Furthermore, as Internet of Things (IoT) devices become more ubiquitous, manual cleansing of data will not be feasible. Participants suggested focusing on the need for better data standards (community-agreed data formats, tools, metadata, test databases, set of metrics, etc.) and developing templates for people to use to contribute data. Producers of data should ensure that the data are in a readily usable form for modeling groups and available through open-source repositories that can be easily cited by data users.

Workshop participants also brought up data scarcity and sparse spatiotemporal data issues, which are true across the coastal, ocean, and cryosphere space (GC3). An AI/ML approach can shed new light on this problem in several ways, including hole filling and representativeness error in data-model comparisons (point observation vs. model grid cell). However, the relatively short temporal coverage of some observational records can limit the training performance of ML models on these datasets. Current strategies for training the models include using proxy/substitute data to recreate the missing data (e.g., Radin and Nieves 2021). The forecasting ability of the ML models may be assessed where longer records are available. Further targeted ML-based research is needed to improve prediction of data gaps when there is a limited number of reliable databases.

Other specific examples efforts to:

- Develop and synthesize datasets from satellite and surface remote sensing as an urgent precursor for building NN models and validating them.
- Quantify the required and desired observational data volumes and types for predicting detailed features of sea ice including leads, ridges, and floes.
- Develop data assimilation approaches that can be used to improve the predictions with ground information and AI-based remote sensing approaches that can be used to generate new calibration datasets.
- Interrogate existing satellite data, which have not been well-used by the coastal community, and merge with in situ observations.
- Create AI applications for use in data analysis and model validation: process-based studies.
- Convert data that we have on human systems, which are often aspatial, into spatial and grid-based form for use in ESMs.
- In the coastal space, merge and fill in data for groundwater and BGC observations, which are much further inland than other coastal observations (GC2).

7.4.2 Utilizing Transferable AI/Transfer Learning, and Using AI/ML for Investigating Parameter Sensitivity and Scale-dependence in Models

Many AI/ML coastal process and sea ice models will be trained on data from specific well-instrumented regions and well-measured physical regimes. Their broader applicability will need to be addressed with transfer learning approaches in order to allow these models to extend from data-rich to data-poor regions (GC3). Transfer learning techniques are also important to apply ML methods to multiple situations such as various coastal regions with different attributes and physical processes, or multiple sea ice regimes. Mitigating site-specific bias is also related to environmental justice issues, since training datasets tend to be disproportionately distributed across different socioeconomic regions. It will be important to identify regions where statistical ML methods are sufficient and where they are not. Another related priority is determining how to ensure accuracy under extreme conditions and improve out-of-sample prediction with AI/ML approaches (GC4). Hybrid modeling is an important focus for reducing the uncertainty of AI/ML predictions for extremes.

AI/ML techniques that can help determine scale-dependence and parameter sensitivity are of particular interest to applications across the ocean, ice, and coastal spaces. These approaches are needed to better describe cascades from large to small and from small to large scales (GC1). A particular pressing application would be to the Arctic permafrost landscape in which its structural makeup is changing across all scales and leading to associated changes to heat and BGC fluxes. It is also important to better understand the scale dependence and sensitivity of certain parameters, especially in the context of fully coupled simulations (GC2). This includes couplings with human dimensions at the coast, and more focus needs to be placed on how we can make predictions on human impacts from physical coastal models (GC4).

7.4.3 Increasing Trust in ML through Advances in Interpretability, Interoperability, and Explainability

One pervasive priority for advancing the use of AI/ML in ocean, ice, and coastal science is to increase the transparency of AI/ML models through explainable and interpretable approaches. Not only is this important for building trust in AI-based models within the scientific and stakeholder communities, but it will also provide a powerful framework for better understanding the complex systems being modeled. Interpretability will better demonstrate the impact AI has on Earth system science, and it will improve how the results of AI models can be communicated and adopted into actionable policy. UQ will form a large part of this in being able to quantify uncertainty from ML models and data.

As real-time data are made more AI-ready, AI models will need to become more interoperable with various data sources and uses. Fully leveraging an increase in AI-ready data will require a capability to continuously update models, which necessitates that models be continuously evaluated to ensure that they remain physically constrained. This will involve effectively interfacing them with domain experts. Furthermore, ML requires more direct interfacing with HPC climate models to extend the state of the art – an effort that needs to consider improved accuracy, improved precision, and importantly enhanced connectivity to impact industry and society and demonstrate how ML improves our world. ESMs should be upgraded to use more modern programming tools (e.g., CliMA), which will better allow for direct interfacing with AI/ML tools.

Although improved explainability is critical to the application and advancement of AI models in the long-term, it should not limit the efforts in pushing the boundaries of what is possible with current techniques. It is important that we fully explore the potential accuracy gains made possible by AI-based approaches even if they are lacking in interpretability, since in any case assumptions and statistical relationships (e.g., for subgrid-scale processes) that may not be fully understood are widely employed in existing mechanistic models.

7.4.4 Community Building / Finding a Common Language

Important factors in community building include the following:

- At Geophysical Fluid Dynamics Laboratory (GFDL), a lot of the momentum comes from the younger scientists in terms of engaging other communities and building bridges.
- Trying to leverage what we already have (e.g., AGU/DOE, etc.), do we pursue workshops? What are the tangible steps to make this happen? While it might be more clear to build bridges within an institution, other efforts are more fractured.
- While there are lots of high-level comments/discussions, very specific skills are needed, and there are lots of details to sort out. A ground-up approach might be needed.
- People who often fuel these efforts are those not doing the publishing of papers, and efforts can fall through the cracks. We need a better career pathway for those who are really dependent on publishing.
- Lots of work is needed to bridge the gap between work in climate research/physical understanding and what we can do with the ML to work together on specific problems, approaching them from different perspectives and seeing what we can merge and what new understandings can emerge.
- How do we discover the usefulness of ML, and who do we work with on it? How do we set time aside to build bridges across communities (data, code, infrastructure, regridding, etc.)?
- We need a solid way to recognize the work critical to a project but that is not easy to recognize through papers. Some DOI recognition is possible now, and there is also the journal of open-source software (at https://github.com/openjournals/joss).

7.5 Short-term (<5 years), 5-year, and 10-year Goals

Finally, we tackled the question of, "What are the near-term (< 5 years), 5-year, and 10-year goals for developing and applying AI/ML approaches for Coastal Dynamics, Oceans, and Ice Prediction?"

7.5.1 Short-term (< 5 years) Goals

Short-term goals include efforts to:

1. Develop various AI-ready standardized datasets and distribute them to the community to work on/collaborate on through open-source repositories.
    - Example 1: Deltares CLASH database for ANN development prediction of wave overtopping.
    - Example 2: obs4MIPs project to coalesce data. This provides a standard format (netcdf-4), and standard metadata that facilitates indexing/discovery. This also allows trivial conversion to cloud formats (if required, e.g., Zarr).
2. For coastal AI/ML research, develop an end-to-end flood modeling pipeline and risk assessment solutions using exposure, hazard, and vulnerability (including human evacuation) models for a smaller region within the next five years before expanding this to greater spatial coverage as computationally feasible. Develop wave-current interaction AI/ML-based parameterizations or emulators and employ AI/ML analysis techniques for coupled coastal-hydrology applications (e.g., using causal networks).
3. For ocean AI/ML research, develop ML-based parameterizations or emulators for unresolved physics.
4. For ice AI/ML research, develop efficient methods for representing sea ice morphology and associated physical interactions efficiently, including floe, ridge, and lead generation and evolution. Begin to explore a framework for AI/ML within Earth system models that caters to grand challenges for sea ice. For land ice, a short-term goal is the demonstration of AI/ML, either from observations or full complexity models, for reducing the cost and complexity of key processes that are currently expensive (glacier ice flow) or difficult to model (subglacial hydrology, iceberg calving).
5. For combined efforts, investigate connections between atmosphere-ocean-ice processes using causal networks. Explore scale-dependent parametric space aided by AI/ML. Find fast and accurate parameterizations for ice-shelf basal melting (e.g., Rosier, Bull, and Gudmundsson 2022).

7.5.2 Medium-term (5-year) Goals

Medium-term goals include efforts to:

1. Establish a large-scale AI/ML working group to build bridges across the communities that can help modelers, observationalists, and computer scientists to connect on shared problems and projects.
    - Example: Regular virtual hybrid seminars, with the goal to create a list of research projects in the big areas we are involved in. Participants pick projects that they are

most interested in and fracture into subgroups to work on the different projects, while keeping each other updated and cited since the work is so closely related. The seminar is to be envisioned as an "ongoing conference" so that the community stays up to date with the diverse ideas and approaches fresh to all of us.

2. Establish "official" domain-focused AI/ML code repositories (Github) to which the entire community can collaborate on/contribute.
    - Currently, generalized ML software development has high community engagement. We need a higher level of community engagement around coastal/ocean/ice applications for AI. It requires interagency cooperation.
    - As part of this, the numerical systems/prediction systems (e.g., E3SM) and data, we have a need to be prepared for easy use with AI/ML methods.
3. Have ocean-/ice-/coastal-focused AI/ML tools and data available to every oceanographer (and ice/coastal scientist) so that it becomes a core part of every scientist's toolkit, building from demonstration AI/ML models developed as part of short-term goals.
4. Develop AI for "smart" instruments (i.e., "learning" drifters, argo profilers, automated weather stations, etc.), self-adjusting parameterization, and data assimilation in models.

7.5.3 Long-term (10-year) Goals

Long-term goals include efforts to:

1. Establish a new workforce stream that is native in AI/ML and uses it as par for the course in the areas of ocean, ice, or coastal areas (also see recommendations in Fleming et al. 2021).
    - Example 1: In Germany, the question of how to bridge marine science and ML/data science is being tackled by launching a doctoral training program: https://www.mardata.de/program.
    - DOE could be in a unique position to fund Scientific Discovery through Advanced Computing (SciDac)-style collaborations or graduate and postdoctoral fellowships that build bridges between climate and AI/ML expertise.
2. Establish trust and reliability as a core part of AI/ML through better interpretability and explainability, allowing users to interact with the models. The speed of AI/ML models opens opportunities for education and outreach by letting students and members of the public run efficient models that convey Earth system concepts.
3. Develop comprehensive AI/ML-capable modular components within ESMs (e.g., sea ice, ice sheet, surface wave, ocean mixing, coastal flooding).
4. Develop scale-aware coupling of PDE-based models with AI/ML models to capture important physical processes more efficiently within Earth system models.

7.6 References

Bajo, Marco, and Georg Umgiesser. 2010. "Storm Surge Forecast through a Combination of Dynamic and Neural Network Models." *Ocean Modelling* 33 (1–2): 1–9. https://doi.org/10.1016/J.OCEMOD.2009.12.007.

Bass, Benjamin, and Philip Bedient. 2018. "Surrogate Modeling of Joint Flood Risk across Coastal Watersheds." *Journal of Hydrology* 558: 159–73. https://doi.org/https://doi.org/10.1016/j.jhydrol.2018.01.014.

Beuzen, Tomas, Evan B. Goldstein, and Kristen D. Splinter. 2019. "Ensemble Models from Machine Learning: An Example of Wave Runup and Coastal Dune Erosion." *Natural Hazards and Earth System Sciences* 19 (10): 2295–2309. https://doi.org/10.5194/nhess-19-2295-2019.

Boehme, Lars, and Isabella Rosso. 2021. "Classifying Oceanographic Structures in the Amundsen Sea, Antarctica." *Geophysical Research Letters* 48 (5): e2020GL089412. https://doi.org/https://doi.org/10.1029/2020GL089412.

Bolton, Thomas, and Laure Zanna. 2019. "Applications of Deep Learning to Ocean Data Inference and Subgrid Parameterization." *Journal of Advances in Modeling Earth Systems* 11 (1): 376–99. https://doi.org/https://doi.org/10.1029/2018MS001472.

Braakmann-Folgmann, Anne, and Craig Donlon. 2019. "Estimating Snow Depth on Arctic Sea Ice Using Satellite Microwave Radiometry and a Neural Network." *The Cryosphere* 13 (9): 2421–38. https://doi.org/10.5194/tc-13-2421-2019.

Bushinsky, Seth M., Peter Landschützer, Christian Rödenbeck, Alison R. Gray, David Baker, Matthew R. Mazloff, Laure Resplandy, Kenneth S. Johnson, and Jorge L. Sarmiento. 2019. "Reassessing Southern Ocean Air-Sea CO_2 Flux Estimates With the Addition of Biogeochemical Float Observations." *Global Biogeochemical Cycles* 33 (11): 1370–88. https://doi.org/https://doi.org/10.1029/2019GB006176.

Cao, Fang, and Maria Tzortziou. 2021. "Capturing Dissolved Organic Carbon Dynamics with Landsat-8 and Sentinel-2 in Tidally Influenced Wetland–Estuarine Systems." *Science of The Total Environment* 777 (July): 145910. https://doi.org/10.1016/J.SCITOTENV.2021.145910.

Cao, Zhendong, Phillip J. Wofram, Joel Rowland, Yu Zhang, and Donatella Pasqualini. 2020. "Estimating Sediment Settling Velocities from a Theoretically Guided Data-Driven Approach." *Journal of Hydraulic Engineering* 146 (10): 04020067. https://doi.org/10.1061/(ASCE)HY.1943-7900.0001798.

Castellani, M. 2006. "Identification of Eddies from Sea Surface Temperature Maps with Neural Networks." *International Journal of Remote Sensing* 27 (8): 1601–18. https://doi.org/10.1080/01431160500462170.

Chapman, Christopher, and Anastase Alexandre Charantonis. 2017. "Reconstruction of Subsurface Velocities From Satellite Observations Using Iterative Self-Organizing Maps." *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS* 14 (5). https://doi.org/10.1109/LGRS.2017.2665603.

DelSole, Timothy, and Douglas Nedza. 2021. "Reconstructing the Atlantic Overturning Circulation Using Linear Machine Learning Techniques." *Atmosphere-Ocean*, August 1–13. https://doi.org/10.1080/07055900.2021.1947181.

Denvil-Sommer, Anna, Marion Gehlen, Mathieu Vrac, and Carlos Mejia. 2019. "LSCE-FFNN-v1: A Two-Step Neural Network Model for the Reconstruction of Surface Ocean PCO_2 over the Global Ocean." *Geoscientific Model Development* 12 (5): 2091–2105. https://doi.org/10.5194/gmd-12-2091-2019.

Desbruyères, Damien, Léon Chafik, and Guillaume Maze. 2021. "A Shift in the Ocean Circulation Has Warmed the Subpolar North Atlantic Ocean since 2016." *Communications Earth & Environment* 2 (48). https://doi.org/10.1038/s43247-021-00120-y.

Duncan, David Ian, Patrick Eriksson, Simon Pfreundschuh, Christian Klepp, and Daniel C. Jones. 2019. "On the Distinctiveness of Observed Oceanic Raindrop Distributions." *Atmospheric Chemistry and Physics* 19 (10): 6969–84. https://doi.org/10.5194/acp-19-6969-2019.

Fleming, Sean W., James R. Watson, Ashley Ellenson, Alex J. Cannon, and Velimir C. Vesselinov. 2021. "Machine Learning in Earth and Environmental Science Requires Education and Research Policy Reforms." *Nature Geoscience* 14 (12): 878–80. https://doi.org/10.1038/s41561-021-00865-3.

French, Jon, Robert Mawdsley, Taku Fujiyama, and Kamal Achuthan. 2017. "Combining Machine Learning with Computational Hydrodynamics for Prediction of Tidal Surge Inundation at Estuarine Ports." *Procedia IUTAM* 25 (January): 28–35. https://doi.org/10.1016/J.PIUTAM.2017.09.005.

George, Tom M., Georgy E. Manucharyan, and Andrew F. Thompson. 2021. "Deep Learning to Infer Eddy Heat Fluxes from Sea Surface Height Patterns of Mesoscale Turbulence." *Nature Communications* 12 (800). https://doi.org/10.1038/s41467-020-20779-9.

Giglio, D., V. Lyubchich, and M. R. Mazloff. 2018. "Estimating Oxygen in the Southern Ocean Using Argo Temperature and Salinity." *Journal of Geophysical Research: Oceans* 123 (6): 4280–97. https://doi.org/https://doi.org/10.1029/2017JC013404.

Goldstein, Evan B., Giovanni Coco, and Nathaniel G. Plant. 2019. "A Review of Machine Learning Applications to Coastal Sediment Transport and Morphodynamics." *Earth-Science Reviews* 194: 97–108. https://doi.org/https://doi.org/10.1016/j.earscirev.2019.04.022.

Guillaumin, Arthur P, and Laure Zanna. 2021. "Stochastic-Deep Learning Parameterization of Ocean Momentum Forcing." *Journal of Advances in Modeling Earth Systems* 13 (9): e2021MS002534. https://doi.org/https://doi.org/10.1029/2021MS002534.

Guimbard, Sébastien, Jérôme Gourrion, Marcos Portabella, Antonio Turiel, Carolina Gabarró, and Jordi Font. 2012. "SMOS Semi-Empirical Ocean Forward Model Adjustment." *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING* 50 (5). https://doi.org/10.1109/TGRS.2012.2188410.

Hieronymus, Magnus, Jenny Hieronymus, and Fredrik Hieronymus. 2019. "On the Application of Machine Learning Techniques to Regression Problems in Sea Level Studies." *Journal of Atmospheric and Oceanic Technology* 36 (9): 1889–1902. https://doi.org/10.1175/JTECH-D-19-0033.1.

Horvat, Christopher, and Lettie A. Roach. 2022. "WIFF1.0: A Hybrid Machine-Learning-Based Parameterization of Wave-Induced Sea Ice Floe Fracture." *Geoscientific Model Development* 15 (2): 803–14. https://doi.org/10.5194/gmd-15-803-2022.

Houghton, Isabel A., and James D. Wilson. 2020. "El Niño Detection Via Unsupervised Clustering of Argo Temperature Profiles." *Journal of Geophysical Research: Oceans* 125 (9): e2019JC015947. https://doi.org/https://doi.org/10.1029/2019JC015947.

Irrgang, Christopher, Jan Saynisch, and Maik Thomas. 2019. "Estimating Global Ocean Heat Content from Tidal Magnetic Satellite Observations." *Scientific Reports* 9 (1): 7893. https://doi.org/10.1038/s41598-019-44397-8.

James, Scott C., Yushan Zhang, and Fearghal O'Donncha. 2018. "A Machine Learning Framework to Forecast Wave Conditions." *Coastal Engineering* 137: 1–10. https://doi.org/https://doi.org/10.1016/j.coastaleng.2018.03.004.

Jia, Gaofeng, Alexandros A. Taflanidis, Norberto C. Nadal-Caraballo, Jeffrey A. Melby, Andrew B. Kennedy, and Jane M. Smith. 2016. "Surrogate Modeling for Peak or Time-Dependent Storm Surge Prediction over an Extended Coastal Region Using an Existing Database of Synthetic Storms." *Natural Hazards* 81 (2): 909–38. https://doi.org/10.1007/s11069-015-2111-1.

Jiang, Peishi, Nis Meinert, Helga Jordão, Constantin Weisser, Simon Holgate, Alexander Lavin, Björn Lütjens, et al. 2021. "Digital Twin Earth – Coasts: Developing a Fast and Physics-Informed Surrogate Model for Coastal Floods via Neural Operators," October. https://doi.org/10.48550/arxiv.2110.07100.

Jones, Daniel C., Harry J. Holt, Andrew J. S. Meijers, and Emily Shuckburgh. 2019. "Unsupervised Clustering of Southern Ocean Argo Float Temperature Profiles." *Journal of Geophysical Research: Oceans* 124 (1): 390–402. https://doi.org/https://doi.org/10.1029/2018JC014629.

Kajbaf, Azin al, and Michelle Bensi. 2020. "Application of Surrogate Models in Estimation of Storm Surge: A Comparative Assessment." *Applied Soft Computing* 91: 106184. https://doi.org/https://doi.org/10.1016/j.asoc.2020.106184.

Kavanaugh, M. T., M. J. Oliver, F. P. Chavez, R. M. Letelier, F. E. Muller-Karger, and S. C. Doney. 2016. "Seascapes as a New Vernacular for Pelagic Ocean Monitoring, Management and Conservation." *ICES Journal of Marine Science* 73: 1839–50. https://doi.org/10.1093/icesjms/fsw086.

Kim, Seung-Woo, Jeffrey A. Melby, Norberto C. Nadal-Caraballo, and Jay Ratcliff. 2015. "A Time-Dependent Surrogate Model for Storm Surge Prediction Based on an Artificial Neural Network Using High-Fidelity Synthetic Hurricane Modeling." *Natural Hazards* 76 (1): 565–85. https://doi.org/10.1007/s11069-014-1508-6.

Kyprioti, Aikaterini P., Alexandros A. Taflanidis, Norberto C. Nadal-Caraballo, and Madison Campbell. 2021a. "Storm Hazard Analysis over Extended Geospatial Grids Utilizing Surrogate Models." *Coastal Engineering* 168: 103855. https://doi.org/https://doi.org/10.1016/j.coastaleng.2021.103855.

Kyprioti, Aikaterini P., Alexandros A. Taflanidis, Norberto C. Nadal-Caraballo, and Madison O. Campbell. 2021b. "Incorporation of Sea Level Rise in Storm Surge Surrogate Modeling." *Natural Hazards* 105 (1): 531–63. https://doi.org/10.1007/s11069-020-04322-z.

Lai, Ching-Yao, Jonathan Kingslake, Martin G. Wearing, Po-Hsuan Cameron Chen, Pierre Gentine, Harold Li, Julian J. Spergel, and J. Melchior van Wessem. 2020. "Vulnerability of Antarctica's Ice Shelves to Meltwater-Driven Fracture." *Nature* 584 (7822): 574–78. https://doi.org/10.1038/s41586-020-2627-8.

Landschützer, P., N. Gruber, D. C. E. Bakker, and U. Schuster. 2014. "Recent Variability of the Global Ocean Carbon Sink." *Global Biogeochemical Cycles* 28 (9): 927–49. https://doi.org/https://doi.org/10.1002/2014GB004853.

Lee, Jun-Whan, Jennifer L. Irish, Michelle T. Bensi, and Douglas C. Marcy. 2021. "Rapid Prediction of Peak Storm Surge from Tropical Cyclone Track Time Series Using Machine Learning." *Coastal Engineering* 170: 104024. https://doi.org/https://doi.org/10.1016/j.coastaleng.2021.104024.

Li, Wei, Amin Kiaghadi, and Clint Dawson. 2021. "High Temporal Resolution Rainfall–Runoff Modeling Using Long-Short-Term-Memory (LSTM) Networks." *Neural Computing and Applications* 33 (4): 1261–78. https://doi.org/10.1007/s00521-020-05010-6.

Manucharyan, Georgy E., Lia Siegelman, and Patrice Klein. 2021. "A Deep Learning Approach to Spatiotemporal Sea Surface Height Interpolation and Estimation of Deep Currents in Geostrophic Ocean Turbulence." *Journal of Advances in Modeling Earth Systems* 13 (1): e2019MS001965. https://doi.org/https://doi.org/10.1029/2019MS001965.

Martinez, Elodie, Thomas Gorgues, Matthieu Lengaigne, Clement Fontana, Raphaëlle Sauzède, Christophe Menkes, Julia Uitz, Emanuele di Lorenzo, and Ronan Fablet. 2020. "Reconstructing Global Chlorophyll-a Variations Using a Non-Linear Statistical Approach." *Frontiers in Marine Science* 7. https://www.frontiersin.org/article/10.3389/fmars.2020.00464.

Maze, Guillaume, Herlé Mercier, Ronan Fablet, Pierre Tandeo, Manuel Lopez Radcenco, Philippe Lenca, Charlène Feucher, and Clément le Goff. 2017. "Coherent Heat Patterns Revealed by Unsupervised Classification of Argo Temperature Profiles in the North Atlantic Ocean." *Progress in Oceanography* 151 (February): 275–92. https://doi.org/10.1016/J.POCEAN.2016.12.008.

Mustapha, Zied ben, Séverine Alvain, Cédric Jamet, Hubert Loisel, and David Dessailly. 2014. "Automatic Classification of Water-Leaving Radiance Anomalies from Global SeaWiFS Imagery: Application to the Detection of Phytoplankton Groups in Open Ocean Waters." *Remote Sensing of Environment* 146 (April): 97–112. https://doi.org/10.1016/J.RSE.2013.08.046.

Nichol, J. Jake, Matthew G. Peterson, Kara J. Peterson, G. Matthew Fricke, and Melanie E. Moses. 2021. "Machine Learning Feature Analysis Illuminates Disparity between E3SM Climate Models and Observed Climate Change." *Journal of Computational and Applied Mathematics* 395 (October): 113451. https://doi.org/10.1016/J.CAM.2021.113451.

Nieves, Veronica, Marta Marcos, and Josh K. Willis. 2017. "Upper-Ocean Contribution to Short-Term Regional Coastal Sea Level Variability along the United States." *Journal of Climate* 30 (11): 4037–45. https://doi.org/10.1175/JCLI-D-16-0896.1.

Nieves, Veronica, Cristina Radin, and Gustau Camps-Valls. 2021. "Predicting Regional Coastal Sea Level Changes with Machine Learning." *Scientific Reports* 11 (1): 7650. https://doi.org/10.1038/s41598-021-87460-z.

Pang, G., M. D'Elia, M. Parks, and G. E. Karniadakis. 2020. "NPINNs: Nonlocal Physics-Informed Neural Networks for a Parametrized Nonlocal Universal Laplacian Operator. Algorithms and Applications." *Journal of Computational Physics* 422 (December): 109760. https://doi.org/10.1016/J.JCP.2020.109760.

Park, Ellen, Jae Deok Kim, Nadege Aoki, Yumeng Melody Cao, Yamin Arefeen, Matthew Beveridge, David Nicholson, and Iddo Drori. 2021. "Predicting Critical Biogeochemistry of the Southern Ocean for Climate Monitoring," October. https://doi.org/10.48550/arxiv.2111.00126.

Plumlee, Matthew, Taylor G. Asher, Won Chang, and Matthew v Bilskie. 2021. "High-Fidelity Hurricane Surge Forecasting Using Emulation and Sequential Experiments." *The Annals of Applied Statistics* 15 (1): 460 – 480. https://doi.org/10.1214/20-AOAS1398.

Radin, Cristina, and Veronica Nieves. 2021. "Machine-Learning Based Reconstructions of Past Regional Sea Level Variability From Proxy Data." *Geophysical Research Letters* 48 (23): e2021GL095382. https://doi.org/https://doi.org/10.1029/2021GL095382.

Ratnam, J. v, H. A. Dijkstra, and Swadhin K. Behera. 2020. "A Machine Learning Based Prediction System for the Indian Ocean Dipole." *Scientific Reports* 10 (1): 284. https://doi.org/10.1038/s41598-019-57162-8.

Reinisch, Elena C., Christopher X. Ren, Andrew Roberts, Cathy Wilson, Patrick B. Eriksson, and Amanda Ziemann. 2021. "Remote Sensing of Polar Ice: Combining Synthetic Aperture Radar and Machine Learning for Operational Navigability." In *Algorithms for Synthetic Aperture Radar Imagery XXVIII*, edited by Edmund Zelnio and Frederick D Garber, 11728:20 – 33. SPIE. https://doi.org/10.1117/12.2587518.

Rosier, S. H. R., C. Y. S. Bull, and G. H. Gudmundsson. 2022. "Predicting Ocean-Induced Ice-Shelf Melt Rates Using a Machine Learning Image Segmentation Approach." *The Cryosphere Discussions* 2022: 1–26. https://doi.org/10.5194/tc-2021-396.

Rosso, Isabella, Matthew R. Mazloff, Lynne D. Talley, Sarah G. Purkey, Natalie M. Freeman, and Guillaume Maze. 2020. "Water Mass and Biogeochemical Variability in the Kerguelen Sector of the Southern Ocean: A Machine Learning Approach for a Mixing Hot Spot." *Journal of Geophysical Research: Oceans* 125 (3): e2019JC015877. https://doi.org/https://doi.org/10.1029/2019JC015877.

Sithara, S., S. K. Pramada, and Santosh G. Thampi. 2021. "Statistical Downscaling of Sea Level by Support Vector Machine and Regression Tree Approaches." In *Climate Change Impacts on Water Resources: Hydraulics, Water Resources and Coastal Engineering*, edited by Ramakar Jha, Vijay P. Singh, Vivekanand Singh, L. B. Roy, and Roshni Thendiyath, 183–

92. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-64202-0_17.

Sochala, Pierre, Chen Chen, Clint Dawson, and Mohamed Iskandarani. 2020. "A Polynomial Chaos Framework for Probabilistic Predictions of Storm Surge Events." *Computational Geosciences* 24 (1): 109–28. https://doi.org/10.1007/s10596-019-09898-5.

Sonnewald, Maike, and Redouane Lguensat. 2021a. "Revealing the Impact of Global Heating on North Atlantic Circulation Using Transparent Machine Learning." *Journal of Advances in Modeling Earth Systems* 13 (8): e2021MS002496. https://doi.org/https://doi.org/10.1029/2021MS002496.

Sonnewald, Maike, Redouane Lguensat, Daniel C. Jones, Peter D. Dueben, Julien Brajard, and V. Balaji. 2021b. "Bridging Observations, Theory and Numerical Simulation of the Ocean Using Machine Learning." *Environmental Research Letters* 16 (7): 073008. https://doi.org/10.1088/1748-9326/ac0eb0.

Sonnewald, Maike, Carl Wunsch, and Patrick Heimbach. 2019. "Unsupervised Learning Reveals Geography of Global Ocean Dynamical Regions." *Earth and Space Science* 6 (5): 784–94. https://doi.org/10.1029/2018EA000519.

Stocker, Thomas, Dahe Qin, Gian-Kasper Plattner, Melinda Tignor, Simon Allen, Judith Boschung, Alexander Nauels, Yu Xia, Vincent Bex, and Pauline Midgley, eds. 2014. *Climate Change 2013: The Physical Science Basis*. New York: Cambridge University Press. https://www.cambridge.org/us/academic/subjects/earth-and-environmental-science/climatology-and-climate-change/climate-change-2013-physical-science-basis-working-group-i-contribution-fifth-assessment-report-intergovernmental-panel-climate-change?format=PB&isbn=9781107661820#bookPeople.

Sztobryn, Marzenna. 2003. "Forecast of Storm Surge by Means of Artificial Neural Network." *Journal of Sea Research* 49 (4): 317–22. https://doi.org/10.1016/S1385-1101(03)00024-8.

Tiggeloven, Timothy, Anaïs Couasnon, Chiem van Straaten, Sanne Muis, and Philip J. Ward. 2021. "Exploring Deep Learning Capabilities for Surge Predictions in Coastal Areas." *Scientific Reports* 11 (1): 17224. https://doi.org/10.1038/s41598-021-96674-0.

Toms, Benjamin A., Elizabeth A. Barnes, and Imme Ebert-Uphoff. 2020. "Physically Interpretable Neural Networks for the Geosciences: Applications to Earth System Variability." *Journal of Advances in Modeling Earth Systems* 12 (9): e2019MS002002. https://doi.org/https://doi.org/10.1029/2019MS002002.

Turner, Adrian K., Kara J. Peterson, and Dan Bolintineanu. 2022. "Geometric Remapping of Particle Distributions in the Discrete Element Model for Sea Ice (DEMSI v0.0)." *Geoscientific Model Development* 15 (5): 1953–70. https://doi.org/10.5194/gmd-15-1953-2022.

Urrego-Blanco, Jorge R., Nathan M. Urban, Elizabeth C. Hunke, Adrian K. Turner, and Nicole Jeffery. 2016. "Uncertainty Quantification and Global Sensitivity Analysis of the Los Alamos Sea Ice Model." *Journal of Geophysical Research: Oceans* 121 (4): 2709–32. https://doi.org/https://doi.org/10.1002/2015JC011558.

Watson, Andrew J., Ute Schuster, Jamie D. Shutler, Thomas Holding, Ian G. C. Ashton, Peter Landschützer, David K. Woolf, and Lonneke Goddijn-Murphy. 2020. "Revised Estimates of Ocean-Atmosphere CO_2 Flux Are Consistent with Ocean Carbon Inventory." *Nature Communications* 11 (1): 4422. https://doi.org/10.1038/s41467-020-18203-3.

Zanna, Laure, and Thomas Bolton. 2020. "Data-Driven Equation Discovery of Ocean Mesoscale Closures." *Geophysical Research Letters* 47 (17): e2020GL088376. https://doi.org/https://doi.org/10.1029/2020GL088376.

Zhang, Jize, Alexandros A. Taflanidis, Norberto C. Nadal-Caraballo, Jeffrey A. Melby, and Fatimata Diop. 2018. "Advances in Surrogate Modeling for Storm Surge Prediction: Storm Selection and Addressing Characteristics Related to Climate Change." *Natural Hazards* 94 (3): 1225–53. https://doi.org/10.1007/s11069-018-3470-1.

8 Climate Variability and Extremes

Authors: Maria J. Molina (NCAR), Travis A. O'Brien (Indiana University/LBNL), Gemma Anderson (LLNL), Moetasim Ashfaq (ORNL), Katrina Eleanor Bennett (LANL), William Collins (LBNL/UC Berkeley), Scott Collis (Argonne), Katherine Dagon (NCAR), Stephen Klein (LLNL), Juan M. Restrepo (ORNL/University of Tennessee), and Paul A. Ullrich (UC-Davis)

8.1 Grand Challenges

Grand challenges in relation to the application of artificial intelligence (AI) and Earth system models (ESMs) for climate variability and extremes are described in the five subsections that follow, which constituted breakout group discussions facilitated by the chapter authors and attended by the active session participants. Climate variability comprises phenomena across spatiotemporal scales, encompassing teleconnections between large-scale modes of variability and regional to local-scale climate and weather, which can have natural and anthropogenic sources. Extremes include hazards such as heatwaves, floods, droughts, tropical cyclones, and severe thunderstorms.

8.1.1 Climate Variability, Signal Identification, and Sources of Predictability

There are several climate variability phenomena with strong teleconnections, such as the El Niño-Southern Oscillation (ENSO) and Madden Julian Oscillation (MJO), and these should be further explored. In this regard, leveraging new AI tools that can capture nonlinearities and quantify causal relationships in very large climate simulations and datasets may help parametrize and account for across-scale dynamics that are considered important but are as yet not fully understood. It remains unclear whether climate variability and its associated uncertainty are sufficiently captured in our ESMs due to spurious trends in observational data, data gaps, and differences among observational products and ESMs. Agnostic AI, unaware of our own labels of climate modes, along with relaxation of a priori criteria, could potentially lead to the discovery of new modes of variability, climate signals, and sources of predictability. Machine learning (ML) and AI specific to climate applications should be further developed, refined, and leveraged to potentially provide new ways of thinking about variability (Figure 8-1).

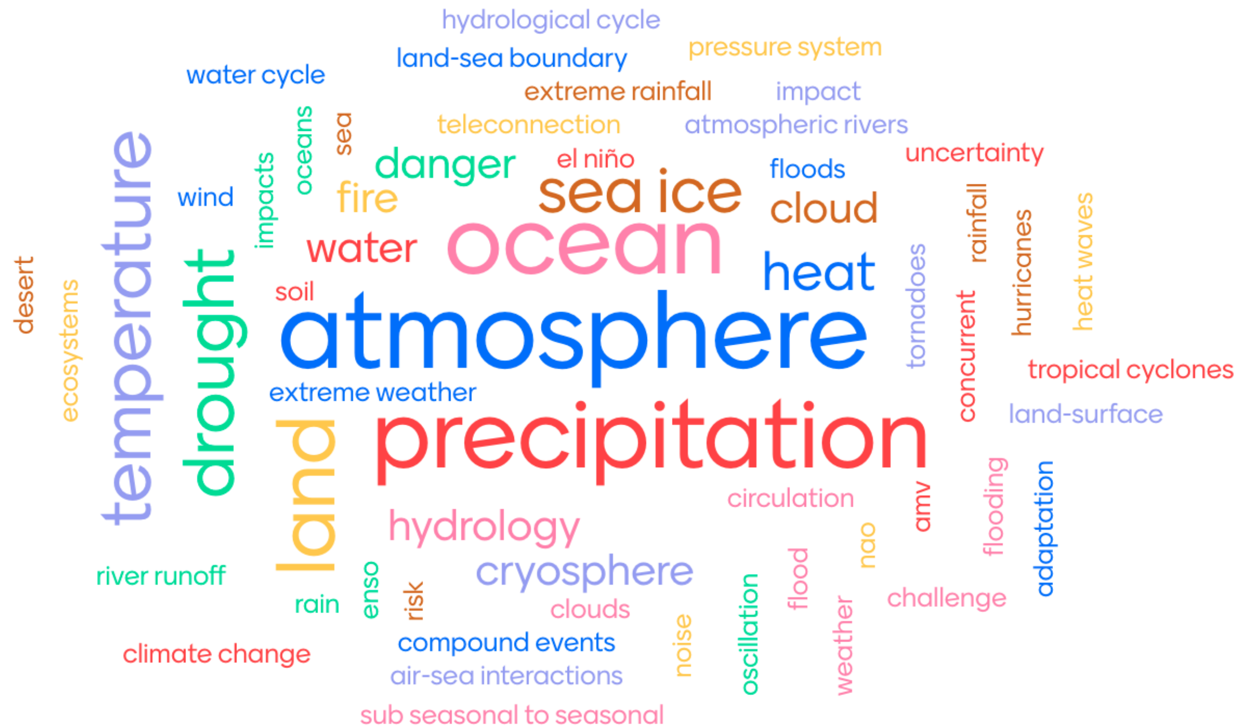


Figure 8-1. Word cloud indicating Earth system components that come to mind for session participants with regard to climate variability and extremes.

8.1.2 Feature Identification and Characterization

ML for feature identification (e.g., tropical cyclones, fronts, and atmospheric rivers) is a rich area which has experienced progress in recent years, but many challenges remain. For example, current ML methods can identify features and make inferences, even if the data lie outside of the training domain, rather than abstaining from making predictions when there is very low confidence. Non-stationarity in Earth system processes also presents issues, as changing characteristics may lead to incorrect identification of extremes through trained ML models, although work focused on conditional stationarity has been recently published (see Table 8-1). The genesis, dissipation, and transition of features of interest (e.g., tropical cyclones to extratropical cyclones) are also not well handled by ML methods. Detection of localized extremes (such as hailstones, tornadoes, and ice storms) that pose significant societal danger are a gap in current feature detection work. It is also unclear to what extent ML models trained in a specific geographic region or season may be transferable to another. Additionally, the unavailability of spatiotemporally resolved Earth system variables, which are needed for feature identification, forces the use of analogues whose appropriateness requires thorough investigation.

Table 8-1. Synthesis of recent advances in AI for climate variability and extremes.

Topic	AI Application(s) and Related Work(s)
Bias Correction	● Bias correction of the MJO using LSTM (Kim et al. 2021). ● CNN-based precipitation observation quality control (Sha et al. 2021). ● Bias correction of climate projections using Regularized Adversarial Domain Adaptation (Pan et al. 2021).
Causal Analysis	● PCMCI: PC-based multivariate causal discovery for time series datasets (Runge et al. 2019; Tigramite github repository). ● Causal inference for quantification of teleconnection pathways (Kretschmer et al. 2021).
Climate and Extremes Prediction	● ML for prediction of extremes, associated hazards, and real-time decision-making (McGovern et al. 2017). ● Forecasting extreme precipitation with random forests (Herman and Schumacher 2018). ● Multiyear ENSO prediction using DL (Ham, Kim, and Luo 2019). ● Analog forecasting of extreme-causing weather patterns using DL (Chattopadhyay, Nabizadeh, and Hassanzadeh 2020). ● DL for near-term tornado prediction (Lagerquist et al. 2020). ● Rainfall-runoff prediction following extremes (Frame et al. 2021). ● Identification of subseasonal forecasts of opportunity using explainable neural networks (Mayer and Barnes 2021). ● Deep generative models for short-term skillful prediction of precipitation and associated radar imagery (Ravuri et al. 2021).
Conditional Stationarity	● Local causal states and discrete coherent structures (Rupe and Crutchfield 2020). ● Discovering Causal Structure with Reproducing-Kernel Hilbert Space ε-Machines (Brodu and Crutchfield 2021).
Downscaling	● DL for statistical downscaling (Baño-Medina, Manzanas, and Gutiérrez 2020). ● DL for downscaling precipitation over complex topography (Sha et al. 2020).
Emulation	● ML for predicting output of high-resolution climate models (Anderson and Lucas 2018). ● ML for emulation and parameter estimation (Dagon et al. 2020). ● ML for clouds and associated processes in general circulation models (Gettelman et al. 2021).
Extremes and Related-Feature Detection	● DL for detection of extremes (Liu et al. 2016). ● Self-organizing maps for climate extremes (Gibson et al. 2017). ● DL for simultaneous tracking of weather phenomena (Mudigonda et al. 2017). ● DL-Front: Detection of fronts using a CNN (Biard and Kunkel 2019). ● DL for front detection (Lagerquist, McGovern, and Gagne 2019).

Table 8-1. (Cont.)

Topic	AI Application(s) and Related Work(s)
	• Uncertainty quantification for detection of extremes (Collins et al. 2020). • Statistical ML for detecting atmospheric rivers (O'Brien et al. 2020). • Unsupervised ML for improved understanding of future drought conditions in the Colorado River basin (Talsma, Bennett, and Vesselinov 2021) • ClimateNet: CNN/ResNet architecture for extremes detection using hand-labeled targets (Prabhat et al. 2021).
Hybrid Modeling	• Hybrid modeling: Combining physics and data-driven approaches with ML (Reichstein et al. 2019). • Application of physics-guided ML for rainfall-runoff and extremes (Xie et al. 2021).
Robustness to Non-stationarity	• Physics constraints and normalization (Beucler et al. 2020). • Physics-informed ML for weather and climate (Kashinath et al. 2021). • Assessment of DL-classification robustness to nonstationarity for severe thunderstorms (Molina, Gagne, and Prein 2021). • Climate-Invariant ML (Beucler et al. 2021).
Signal Separation and Anomaly Detection	• Viewing forced climate signals through an AI lens (Barnes et al. 2019). • Physics-based unsupervised discovery of spatiotemporal coherent structures (Rupe et al. 2019). • Using DL for near-term hail prediction and explainable AI for extraction of prediction signals (Gagne et al. 2019). • Explainable AI methods (McGovern et al. 2019 and Toms, Barnes, and Ebert-Uphoff 2020). • Anomaly detection for physics analysis (Nachman 2020). • Analysis of physical causes of climate change on Midwest extreme precipitation using ML (Davenport and Diffenbaugh 2021).
Synthetic Data	• MJO-index time series reconstruction using one-dimensional convolutional neural networks (Dasgupta et al. 2020).

8.1.3 Extreme Weather Predictors and Precursors

Extremes by definition are rare events and, therefore, the separation of signal from noise in extremes statistics remains demanding due to limited observational records. The limited observational record also presents challenges to understanding co-variability and compounding extremes, in addition to the nonstationarity of extremes and the diversity of teleconnections between natural modes of climate variability and extremes. Currently, many statistical and explainable AI methods do not consider causality, which may lead to spurious relationships, including over-confidence in the teleconnectivity of natural modes of climate variability to extremes; causal predictors and precursors for extremes need to be better quantified. Characterization and definitions for extremes (e.g., extreme precipitation and floods) also need to be improved given different inherent properties based on spatial and temporal scales and the implications on short-term forecasting and climate projections. Furthermore, challenges exist in predicting the statistics of extremes skillfully in a changing climate, where new pathways for water and heat may emerge. Despite existing identified features, precursors to extreme events are still poorly understood and require further identification and characterization.

8.1.4 Observation-Model Integration

Verification and validation of climate model simulations is challenging, particularly during longer timescales (e.g., decadal) and for rare events due in part to limited data, which is a barrier for major development of ESMs. A risk of over-parameterization when integrating observations into models also exists, and thus there is a need for methods that indicate when observation and model agreement are suboptimal. Uncertainty quantification at the intersection of observations and model simulations is important to assess confidence in predictive capability but is lacking and a challenge to compute. Data are also limited over critical climate regions that contain potential tipping points, such as the Arctic and Antarctic. There is also a temporal lag in the incorporation of observations into models; earlier incorporation of data as they are collected is needed. Additionally, connections between modelers and observationalists are difficult to maintain, which can stymie progress on tool development and application. Observations are often inhomogeneous and disparate over time, which presents a challenge when gap filling observations and models. Moreover, a heavy emphasis has been placed on the development of data-driven methods in AI; more development of physics-informed AI methods is needed in observation-model integration.

8.1.5 Downscaling and Bias Correction

The use of AI/ML in the downscaling and bias correction of climate models at finer scales is currently very limited. There is some success in the development of trained deep learning (DL) models to represent components of regional models, such as the Weather Research and Forecasting model, and in the downscaling of climate models outputs, such as temperature. However, most of these efforts are currently limited to snapshot modeling or downscaling. There are challenges in the sampling to capture tails in small-scale physics, and in the application of ML models without breaking physical laws. Fine-scale benchmarking datasets are also very limited or lacking, which could be used for standardized evaluations. There are also challenges in the selection and availability of reasonable priors for neural networks, especially for uncertainty quantification. Future advancements would require ML models that are specifically designed and trained for the domain science, downscaling in both space and time, and blending of traditional methods with ML to fully exploit the strengths of multiple data analytics approaches. Sufficiently long and high-resolution data streams would be key to exploit ML as a tool to emulate physical models representing small-scale physics.

8.2 State-of-the-Science

Motivation driving the use of ML for climate variability and extremes includes improving predictive skill across timescales and gaining new process understanding, given the ability of AI to learn nonlinear, complex relationships in large datasets and across multiple variables. Improved climatology and trend assessment of climate variability and extreme events, in addition to creating improved representation of extremes, also motivate the use of AI within these science areas. ML also provides scalability and resolves issues associated with rule-based and heuristic methods, which are sensitive to data distributions, in the detection of extremes. The use of AI is also motivated by the need to perform reduced-order modeling, overcome data I/O challenges with exascale, and improve assimilation of observations into modeling systems.

Literature that documents applications of ML for climate variability and extremes (Table 8-1) include extreme event identification (e.g., atmospheric rivers, tropical cyclones, and fronts), improvement of subseasonal-to-seasonal and decadal prediction, explainable AI for signal separation from noise (in both space and time), data extrapolation and interpolation (e.g., gap filling and generation of pseudo-observational data), representation of subgrid processes (e.g., parameterizations of subgrid processes), and bias correction as a model post-processing step.

Figure 8-2 shows attendees' primary usage of AI/ML.

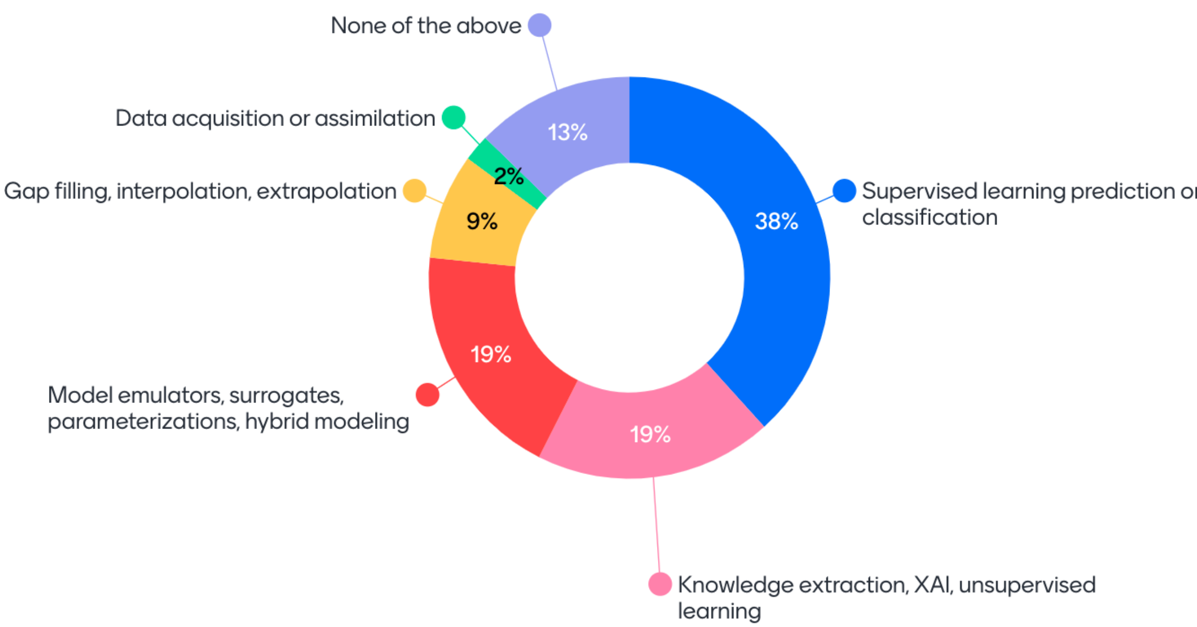


Figure 8-2. Attendee responses indicating their primary usage of AI and ML.

While not specific to the application of AI, compound extremes are a critically emerging field. Predictive understanding of compound and cascading extremes is an emerging area where robust frameworks for defining such extremes and evaluating their impacts are still in their infancy stage. Complexity in compound extremes is evident from the fact that these can arise due to interaction between multiple extremes (e.g., concurrently occurring temperature, precipitation, and wind extremes) or due to the confluence of non-extreme states of Earth system variables (e.g., persistent increases in mean temperatures) across space and time. Cascading events are even more challenging to identify as those often are separated in space and time. ML can guide the development of robust analytical frameworks for efficient detection and attribution of such interdependent and interacting states in the land-atmosphere-ocean continuum.

8.3 Experimental, Data, and Modeling Opportunities

Using existing ML tools for mechanisms and precursors discovery. Existing tools, such as artificial neural networks and linear inverse modeling, present opportunities to discover new signals in the Earth system and should be further applied to interrogate existing data. Causal inference and causal discovery methods, along with the application of trained ML models across varying numerical modeling systems, should be further incorporated into AI workflows to help with the quantification of physical drivers and the assessment of robust sources of predictability, while also considering their spatiotemporal nonstationarity.

ML and AI also present opportunities for filling observational gaps and/or combining observational datasets to leverage their complementary strengths (e.g., transfer learning). The use of ML for the creation of pseudo-observational data is another opportunity, which can include extending observational products into time periods when remote sensing and certain observational technologies were unavailable.

ML model development and benchmarking for climate variability and extremes. Massive catalogs from feature trackers, such as the atmospheric river ARTMIP catalog, already exist in consistent format and with uniform standards, and could serve as an example of how to use such data catalogs to leverage and train ML models for extremes. Already trained ML models for feature detection, such as ClimateNet, can also be applied to other numerical model output and observations to gain a better understanding of extremes and assess robustness across products, and new ML models could be created to detect other features associated with extremes (e.g., monsoon depressions and extratropical cyclones). Numerous open-source software packages are well-documented and available in Python for development, training, and deployment of ML (e.g., scikit-learn) and DL (e.g., tensorflow, pytorch) models. Available software, in combination with various large data products, present an opportunity to aid ML model development and benchmarking for climate variability and extremes.

Development and open sourcing of datasets that could be used for ML model benchmarking, including ultra-high-resolution simulations for sufficient representation of extremes, are an opportunity to catalyze robust advancements in downscaling science and the application of ML thereof. Numerous reanalyses, multimodel archives, and cloud-resolving regional climate datasets also exist (e.g., ERA5, E3SM large ensemble) that can be used for unsupervised or supervised ML approaches. These datasets could also be leveraged by educators to train physical science and computer science students.

Data set compression tools and online learning techniques. Data set compression tools present an opportunity to store large amounts of data, enabling subsequent training of data-hungry ML models. ML methods could also potentially help assess how much compression is possible with minimal data loss, and in the case of a reduced set of numerical model output variables, help determine how necessary variables can be reconstructed. Online learning techniques, such as during a numerical model simulation or during ML model training, also present opportunities for observation-model integration.

8.4 Research Priorities

8.4.1 Quantification of Teleconnection Pathways

The teleconnections of ENSO and other climate modes have complex and varying pathways (e.g., inter-basin interactions, interactions with other modes), and thus a clear understanding of these pathways is important, including identification of model-specific signals and those that apply in observations. In regard to methods for climate variability and extremes, further development of AI methods incorporating causality, uncertainty, and physics are of priority, including, but not limited to, methods such as physics-informed neural nets (PINNs), Bayesian multifidelity PINNs, operator regressions, transfer learning for prediction of extremes, and active learning for dealing with lack of data for extremes and longer timescales. ML models that focus on probabilistic forecasts and those that can handle non-Gaussian distributions should also be of priority. The creation of an ML model hierarchy, generation of large ML-based ensembles, modification of ML model architectures, and development of domain-specific loss functions are also priorities that may aid with advancing the science of climate variability and extremes.

8.4.2 Understanding Drivers and Extending Observations of Extremes

The application of ML for detection of extremes and associated features has been successful, and research priorities should lie in gaining a better understanding of the characteristics of extreme events and their inception (e.g., convection initiation or cyclogenesis) both within specific ESMs and in observations. The use of ML methods for developing symbolic relationships (robust to nonstationarity) between quantities (e.g., sea surface temperatures and precipitation from tropical cyclones) were also identified as a priority. In order to extend observational products of extremes and modes of variability specific to longer timescales, a research priority includes using ML to gap fill, statistically downscale, and extend observations back in time (e.g., using GANs). ML should also be leveraged to identify regions that need better sampling (i.e., more observations). Augmentation of current data assimilation methods was also identified as a research priority (e.g., KF, 4DVAR). Anomaly detection using ML is another area that has experienced much progress in recent years, but distinguishing extreme events from outliers that are "bad" or have incorrect data remains difficult and is of priority.

8.4.3 Improving Fine-scale Processes and Model Development

Research priorities include the continued development of emulators to replace fine-scale physics and parameterizations in regional and global models. Better understanding the connections between climate models and weather extremes is another research priority, given that climate models cannot resolve or accurately represent processes associated with extremes. Embedding of ML into numerical models as they are running is another research priority in order to extract information and statistics between time steps. The inclusion of ML in the workflow of model development to diagnose model errors, rather than using ML as a post diagnosis, was identified as a research priority, along with application of Bayesian calibration tools to better understand where numerical model deficiencies are. More robust methods for comparing observations and models as dynamical systems were identified as a community need, along with new ways of capturing and representing unresolved and poorly understood physics and interactions. Research priorities also include the development of physical science synthetic datasets for benchmarking of ML methods, in addition to the creation/emulation of factual and counterfactual scenarios that incorporate uncertainty for climate change attribution.

8.4.4 Metrics and Robustness Assessment

Metrics are scalar measures of skill that are extensively used in model development to compare performance, and the use of AI for their development could lead to metrics that can better capture process complexities and nonlinearities. Various existing metrics do not describe extremes well (e.g., precipitation), such as generalized extreme value theory, partly due to distributions with a heavy tail, and thus AI could be potentially used to create distributions or more useful metrics. Development of benchmarking metrics that measure more than just ML model performance in prediction tasks, such as scalability across hardware and software systems, in addition to trustworthiness of the respective ML method, were identified as research priorities. On the topic of robustness, further development and benchmarking of explainable AI methods, assessment of the transferability of trained ML models from one ESM to another, and whether knowledge extracted is consistent across models and data products and in a nonstationary system (e.g., new water pathways due to changing climate) were all identified as research priorities. Assessing and ensuring the robustness of trained ML models to adversarial data are also of research priority due to security and societal implications.

8.5 Short-term (<5 years), 5-year, and 10-year Goals

To achieve short-, medium-, and long-term goals, numerous considerations were discussed. Collaboration and co-development by computer scientists, domain experts, and software engineers were deemed important to make visionary changes and push scientific boundaries. Training and development of the future and current workforce was also seen as imperative, particularly when considering the experienced workforce with an already clear understanding of Earth science data, and strong partnerships between laboratories and universities to allow for cross pollination of ideas and training of students. High-risk/high-reward research was also identified as critical, with needed support by funding agencies and the broader scientific community for creativity and risk taking in research to create transformative change. Focus on stakeholder and end-user engagement was also emphasized for ML and predictive analytics.

Short-term (< 5 years) goals related to climate variability and extremes include the continued development of ML-based prediction models for modes of climate variability and the application of explainable AI methods to these prediction models to identify where new observations should be collected to enhance predictability. In a related vein, there should be acquisition of more and new observational data for extreme events, such as the use of drifters for tropical cyclones, along with consistent and standardized baselines (e.g., how, when, and where to measure them). Other short-term goals include: (1) continued simulation and exploration of extreme weather phenomena across different ESMs for training ML, (2) development of well-documented geoscience datasets for benchmarking ML applications for domain-specific problems, (3) better quantification of the linkages between climate variability and change to extremes using causal methods, (4) continued development and benchmarking of explainable AI methods for improved understanding of ML model decisions and to build community trust, (5) development of ML-based methods to run and analyze ultra-high resolution simulations and continued assessment of needed resolutions for extremes, (6) development of ML-based analytical frameworks for detecting the environmental stressors causing compound and cascading extremes, and (7) exploration of ML for bounded but long-tailed distributions of extremes.

Medium-term (5-year) goals for climate variability and extremes include the creation of extensive catalogs of features (e.g., atmospheric rivers, tropical cyclones) detected using ML, along with clear documentation using FAIR principles of the features, the preceding training processes, and the trained models themselves. Another medium-term goal is leveraging the extensive availability of ESMs and their ensembles, along with data-driven ML methods, to extract sources of predictability and assess their robustness across modeling systems and available observations. Related to this, the development of ML-prediction models for modes of variability and the capability of ML for handling large volumes of data should allow the community to identify new sources of predictability and to overcome traditional predictability limits. Once overcome, these ML-prediction models can be interrogated to physically understand the new sources of predictability and to identify (via transfer learning) the prediction-relevant biases of ESMs that need fixing. Medium-term goals using ML also include: (1) detection of features and phenomena to regions where focus has been limited in past years (e.g., Arctic, Antarctica, and the southern ocean); (2) modified loss functions for detection of a range of extremes robust to a changing climate; (3) application and development of models that can appropriately handle long-term memory sequences, effects of compound events, and varying causal pathways and teleconnections; (4) efficient detection and attribution of compound and cascading extremes; (5) development of scaling theories for extreme-causing events; (6) production of an easy-to-use toolbox for extreme value distributions; and (7) development of a model hierarchy for use by studies focusing on climate variability and extremes.

Long-term (10-year) goals in relation to climate variability and extremes include the use of AI for discovery of new modes of variability and better quantification of teleconnections to extremes, including an understanding of teleconnection variations for different flavors of modes of climate variability. Long-term goals also include the use of ML for the following: (1) robust identification of reasons for prediction failures, (2) enhanced understanding of sources and limits of predictability across timescales and across ESM components, (3) data assimilation, (4) automated identification of water cycle extremes within observations and ESMs, (5) co-evolved climate model ensembles with a focus on initialized prediction, (6) extraction of coherent structures from model output with robust strategies that can separate long-term forcing signals from short-term variability, (7) uncovering of processes involved in the genesis and initiation of extreme weather phenomena, (8) assessment of predictability of and extending predictions of compound and cascading extremes, and (9) use of transfer operators (or similar) providing an unsupervised learning alternative for the detection of extremes, robust to a changing climate. Other long-term goals include improvement of AI model emulation for uncertainty estimation of climate variability and extremes, the use of AI to enable ultra-high-resolution simulations of extreme phenomena (such as tornadoes, lightning, etc.), and the use of transfer operators with climate models to map history of observations to predictions.

Table 8-2 summarizes long-term goals for Climate Variability and Extremes.

Table 8-2. Overarching long-term and potentially transformative advances in the identified challenge areas.

Grand Challenges	Long-term Vision (10-years)
Climate variability, signal identification, and sources of predictability	Dramatically enhance the predictability of modes of variability via an ML-enabled fusion of observations and models incorporating the discovery of new sources of predictability, the identification of where additional observations are needed, and the discernment of the optimum ways to improve the predictive capabilities of ESMs.
Feature identification and characterization	Feature identification and characterization is seamlessly automated using suites of publicly available toolsets and event catalogs for validation of and within ESMs.
Extreme weather predictors and precursors	The processes involved in the genesis and evolution of extreme weather are revealed within a nonstationary system, compound and cascading extremes are better understood and predictions thereof extended, and ultra-high-resolution simulations of extremes are facilitated.
Observation- model integration	Observation-model integration evolves to simultaneously evaluate and improve both observations and models, leverage AI/ML to target observational and modeling priorities, and encourage effective communication by building connections between research communities.
Downscaling and bias correction	Seamless integration of AI/ML methods with conventional downscaling and bias correction approaches.

8.6 References

Anderson, Gemma J., and Donald D. Lucas. 2018. "Machine Learning Predictions of a Multiresolution Climate Model Ensemble." *Geophysical Research Letters* 45 (9): 4273–80. https://doi.org/10.1029/2018GL077049.

Baño-Medina, Jorge, Rodrigo Manzanas, and José Manuel Gutiérrez. 2020. "Configuration and Intercomparison of Deep Learning Neural Models for Statistical Downscaling." *Geoscientific Model Development* 13 (4): 2109–24.https://doi.org/10.5194/gmd-13-2109-2020.

Barnes, Elizabeth A., James W. Hurrell, Imme Ebert-Uphoff, Chuck Anderson, and David Anderson. 2019. "Viewing Forced Climate Patterns Through an AI Lens." *Geophysical Research Letters* 46 (22): 13389–98. https://doi.org/10.1029/2019GL084944.

Beucler, Tom, Michael Pritchard, Pierre Gentine, and Stephan Rasp. 2020. "Towards Physically-Consistent, Data-Driven Models of Convection," February. https://doi.org/10.48550/arxiv.2002.08525.

Beucler, Tom, Michael Pritchard, Janni Yuval, Ankitesh Gupta, Liran Peng, Stephan Rasp, Fiaz Ahmed, et al. 2021. "Climate-Invariant Machine Learning," December. https://doi.org/10.48550/arxiv.2112.08440.

Biard, James C., and Kenneth E. Kunkel. 2019. "Automated Detection of Weather Fronts Using a Deep Learning Neural Network." *Advances in Statistical Climatology, Meteorology and Oceanography* 5 (2): 147–60. https://doi.org/10.5194/ascmo-5-147-2019.
Brodu, Nicolas, and James P. Crutchfield. 2020. "Discovering Causal Structure with Reproducing-Kernel Hilbert Space ϵ-Machines," November. https://doi.org/10.48550/arxiv.2011.14821.
Chattopadhyay, Ashesh, Ebrahim Nabizadeh, and Pedram Hassanzadeh. 2020. "Analog Forecasting of Extreme-Causing Weather Patterns Using Deep Learning." *Journal of Advances in Modeling Earth Systems* 12 (2): e2019MS001958. https://doi.org/https://doi.org/10.1029/2019MS001958.
Collins, William, Travis O'Brien, M. R. Prabhat, and Karthik Kashinath. 2020. "Machine Learning for Detection of Climate Extremes: New Approaches to Uncertainty Quantification." In *EGU General Assembly 2020, Online, 4–8 May 2020, EGU2020-9820*. https://doi.org/https://doi.org/10.5194/egusphere-egu2020-9820.
Dagon, K., B. M. Sanderson, R. A. Fisher, and D. M. Lawrence. 2020. "A Machine Learning Approach to Emulation and Biophysical Parameter Estimation with the Community Land Model, Version 5." *Advances in Statistical Climatology, Meteorology and Oceanography* 6 (2): 223–44. https://doi.org/10.5194/ascmo-6-223-2020.
Dasgupta, Panini, Abirlal Metya, C. v. Naidu, Manmeet Singh, and M. K. Roxy. 2020. "Exploring the Long-Term Changes in the Madden Julian Oscillation Using Machine Learning." *Scientific Reports 2020 10:1* 10 (1): 1–13. https://doi.org/10.1038/s41598-020-75508-5.
Davenport, Frances v., and Noah S. Diffenbaugh. 2021. "Using Machine Learning to Analyze Physical Causes of Climate Change: A Case Study of U.S. Midwest Extreme Precipitation." *Geophysical Research Letters* 48 (15): e2021GL093787. https://doi.org/10.1029/2021GL093787.
Frame, J., F. Kratzert, D. Klotz, M. Gauch, G. Shelev, O. Gilon, L. M. Qualls, H. v Gupta, and G. S. Nearing. 2021. "Deep Learning Rainfall-Runoff Predictions of Extreme Events." *Hydrology and Earth System Sciences Discussions* 2021: 1–20. https://doi.org/10.5194/hess-2021-423.
Gagne, David John, Sue Ellen Haupt, Douglas W. Nychka, and Gregory Thompson. 2019. "Interpretable Deep Learning for Spatial Analysis of Severe Hailstorms." *Monthly Weather Review* 147 (8): 2827–45. https://doi.org/10.1175/MWR-D-18-0316.1.
Gettelman, A., D. J. Gagne, C. C. Chen, M. W. Christensen, Z. J. Lebo, H. Morrison, and G. Gantos. 2021. "Machine Learning the Warm Rain Process." *Journal of Advances in Modeling Earth Systems* 13 (2): e2020MS002268.https://doi.org/10.1029/2020MS002268.
Gibson, Peter B., Sarah E. Perkins-Kirkpatrick, Petteri Uotila, Acacia S. Pepler, and Lisa v. Alexander. 2017. "On the Use of Self-Organizing Maps for Studying Climate Extremes." *Journal of Geophysical Research: Atmospheres* 122 (7): 3891–3903. https://doi.org/10.1002/2016JD026256.

Ham, Yoo-Geun, Jeong-Hwan Kim, and Jing-Jia Luo. 2019. "Deep Learning for Multi-Year ENSO Forecasts." *Nature* 573 (7775): 568–72. https://doi.org/10.1038/s41586-019-1559-7.

Herman, Gregory R., and Russ S. Schumacher. 2018. "Money Doesn't Grow on Trees, but Forecasts Do: Forecasting Extreme Precipitation with Random Forests." *Monthly Weather Review* 146 (5): 1571–1600. https://doi.org/10.1175/MWR-D-17-0250.1.

Kashinath, K., M. Mustafa, A. Albert, J.-L. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, et al. 2021. "Physics-Informed Machine Learning: Case Studies for Weather and Climate Modelling." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379 (2194): 20200093. https://doi.org/10.1098/rsta.2020.0093.

Kim, H., Y. G. Ham, Y. S. Joo, and S. W. Son. 2021. "Deep Learning for Bias Correction of MJO Prediction." *Nature Communications* 12 (1): 3087. https://doi.org/10.1038/s41467-021-23406-3.

Kretschmer, Marlene, Samantha v. Adams, Alberto Arribas, Rachel Prudden, Niall Robinson, Elena Saggioro, and Theodore G. Shepherd. 2021. "Quantifying Causal Pathways of Teleconnections." *Bulletin of the American Meteorological Society* 1 (aop): 1–34. https://doi.org/10.1175/BAMS-D-20-0117.1.

Lagerquist, Ryan, A. M. Y. McGovern, and David John Gagne. 2019. "Deep Learning for Spatially Explicit Prediction of Synoptic-Scale Fronts." *Weather and Forecasting* 34 (4): 1137–60. https://doi.org/10.1175/WAF-D-18-0183.1.

Lagerquist, Ryan, Amy McGovern, Cameron R. Homeyer, David John Gagne, and Travis Smith. 2020. "Deep Learning on Three-Dimensional Multiscale Data for Next-Hour Tornado Prediction." *Monthly Weather Review* 148 (7): 2837–61. https://doi.org/10.1175/MWR-D-19-0372.1.

Liu, Yunjie, Evan Racah, Prabhat, Joaquin Correa, Amir Khosrowshahi, David Lavers, Kenneth Kunkel, Michael Wehner, and William Collins. 2016. "Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets," May. https://doi.org/10.48550/arxiv.1605.01156.

Mayer, Kirsten J., and Elizabeth A. Barnes. 2021. "Subseasonal Forecasts of Opportunity Identified by an Explainable Neural Network." *Geophysical Research Letters* 48 (10): e2020GL092092. https://doi.org/https://doi.org/10.1029/2020GL092092.

McGovern, Amy, Kimberly L. Elmore, David John Gagne, Sue Ellen Haupt, Christopher D. Karstens, Ryan Lagerquist, Travis Smith, and John K. Williams. 2017. "Using Artificial Intelligence to Improve Real-Time Decision-Making for High-Impact Weather." *Bulletin of the American Meteorological Society* 98 (10): 2073–90. https://doi.org/10.1175/BAMS-D-16-0123.1.

McGovern, Amy, Ryan Lagerquist, David John Gagne, G. Eli Jergensen, Kimberly L. Elmore, Cameron R. Homeyer, and Travis Smith. 2019. "Making the Black Box More Transparent: Understanding the Physical Implications of Machine Learning." *Bulletin of the American Meteorological Society* 100 (11): 2175–99. https://doi.org/10.1175/BAMS-D-18-0195.1.

Molina, Maria J., David John Gagne, and Andreas F. Prein. 2021. “A Benchmark to Test Generalization Capabilities of Deep Learning Methods to Classify Severe Convective Storms in a Changing Climate.” *Earth and Space Science* 8 (9): e2020EA001490. https://doi.org/10.1029/2020EA001490.

Mudigonda, Mayur, Sookyung Kim, Ankur Mahesh, Samira Ebrahimi Kahou, Karthik Kashinath, Dean N. Williams, Vincent Michalski, Travis O’Brien, and Mr. Prabhat. 2017. “Segmenting and Tracking Extreme Climate Events Using Neural Networks.” In *31st Conference on Neural Information Processing Systems*. https://dl4physicalsciences.github.io/files/nips_dlps_2017_20.pdf.

Nachman, Benjamin. 2020. “Anomaly Detection for Physics Analysis and Less than Supervised Learning,” October. https://doi.org/10.48550/arxiv.2010.14554.

O’Brien, Travis A., Mark D. Risser, Burlen Loring, Abdelrahman A. Elbashandy, Harinarayan Krishnan, Jeffrey Johnson, Christina M. Patricola, et al. 2020. “Detection of Atmospheric Rivers with Inline Uncertainty Quantification: TECA-BARD v1.0.1.” *Geoscientific Model Development* 13 (12): 6131–48. https://doi.org/10.5194/gmd-13-6131-2020.

Pan, Baoxiang, Gemma J. Anderson, André Goncalves, Donald D. Lucas, Céline J. W. Bonfils, Jiwoo Lee, Yang Tian, and Hsi-Yen Ma. 2021. “Learning to Correct Climate Projection Biases.” *Journal of Advances in Modeling Earth Systems* 13 (10): e2021MS002509. https://doi.org/10.1029/2021MS002509.

Prabhat, K. Kashinath, M. Mudigonda, S. Kim, L. Kapp-Schwoerer, A. Graubner, E. Karaismailoglu, et al. 2021. “ClimateNet: An Expert-Labeled Open Dataset and Deep Learning Architecture for Enabling High-Precision Analyses of Extreme Weather.” *Geoscientific Model Development* 14 (1): 107–24. https://doi.org/10.5194/gmd-14-107-2021.

Ravuri, Suman, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, et al. 2021. “Skillful Precipitation Nowcasting Using Deep Generative Models of Radar.” *Nature* 597 (7878): 672–77. https://doi.org/10.1038/s41586-021-03854-z.

Reichstein, Markus, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and Prabhat. 2019. “Deep Learning and Process Understanding for Data-Driven Earth System Science.” *Nature* 566 (7743): 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Runge, Jakob, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. 2019. “Detecting and Quantifying Causal Associations in Large Nonlinear Time Series Datasets.” *Science Advances* 5 (11): 4996–5023. https://doi.org/10.1126/SCIADV.AAU4996/SUPPL_FILE/AAU4996_SM.PDF.

Rupe, Adam, and James P. Crutchfield. 2018. “Local Causal States and Discrete Coherent Structures.” *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28 (7): 075312. https://doi.org/10.1063/1.5021130.

Rupe, Adam, Nalini Kumar, Vladislav Epifanov, Karthik Kashinath, Oleksandr Pavlyk, Frank Schlimbach, Mostofa Patwary, et al. 2019. "DisCo: Physics-Based Unsupervised Discovery of Coherent Structures in Spatiotemporal Systems," September. https://doi.org/10.48550/arxiv.1909.11822.

Sha, Yingkai, David John Gagne, Gregory West, and Roland Stull. 2020. "Deep-Learning-Based Gridded Downscaling of Surface Meteorological Variables in Complex Terrain. Part II: Daily Precipitation." *Journal of Applied Meteorology and Climatology* 59 (12): 2075–92. https://doi.org/10.1175/JAMC-D-20-0058.1.

Talsma, Carl James, Katrina E. Bennett, and Velimir v Vesselinov. 2021. "Characterizing Drought Behavior Using Unsupervised Machine Learning for Improved Understanding of Future Drought in the Colorado River Basin," October. https://doi.org/10.1002/ESSOAR.10508427.1.

Toms, Benjamin A., Elizabeth A. Barnes, and Imme Ebert-Uphoff. 2020. "Physically Interpretable Neural Networks for the Geosciences: Applications to Earth System Variability." *Journal of Advances in Modeling Earth Systems* 12 (9): e2019MS002002. https://doi.org/10.1029/2019MS002002.

Xie, Kang, Pan Liu, Jianyun Zhang, Dongyang Han, Guoqing Wang, and Chaopeng Shen. 2021. "Physics-Guided Deep Learning for Rainfall-Runoff Modeling by Considering Extreme Events and Monotonic Relationships." *Journal of Hydrology* 603 (December): 127043. https://doi.org/10.1016/j.jhydrol.2021.127043.

9 Human Systems and Dynamics

Authors: Christa Brelsford (ORNL), Melissa Allen-Dumas (ORNL), Thushara Gunda (SNL), Donatella Pasqualini (LANL), Abigail Snyder (PNNL), Nathalie Voisin (PNNL)

9.1 Introduction

Human processes are the single largest driver of uncertainty in the future Earth system, encompassing everything from global emissions pathways to the farmers' decisions that impact future algal blooms. However, most Earth systems research is not designed to estimate human-scale consequences (Coen 2021). Research at human-centric scales is critical for performing, understanding, and executing actionable climate research because we know that the consequences of climate hazards spread far beyond direct geographic impact via economic and infrastructural connections (Figure 9-1; Shughrue, Werner, and Seto 2020). In particular, Earth systems research needs to enable cross-system and cross-sector analysis of complex climate risks including representations of human systems, building on more complex ways of characterizing risk determinants and their interactions (Simpson et al. 2021). On the following pages, we describe the grand challenges facing human system dynamics in Earth systems predictability and the opportunity space enabled by artificial intelligence (AI) and machine learning (ML) to tackle these science questions.

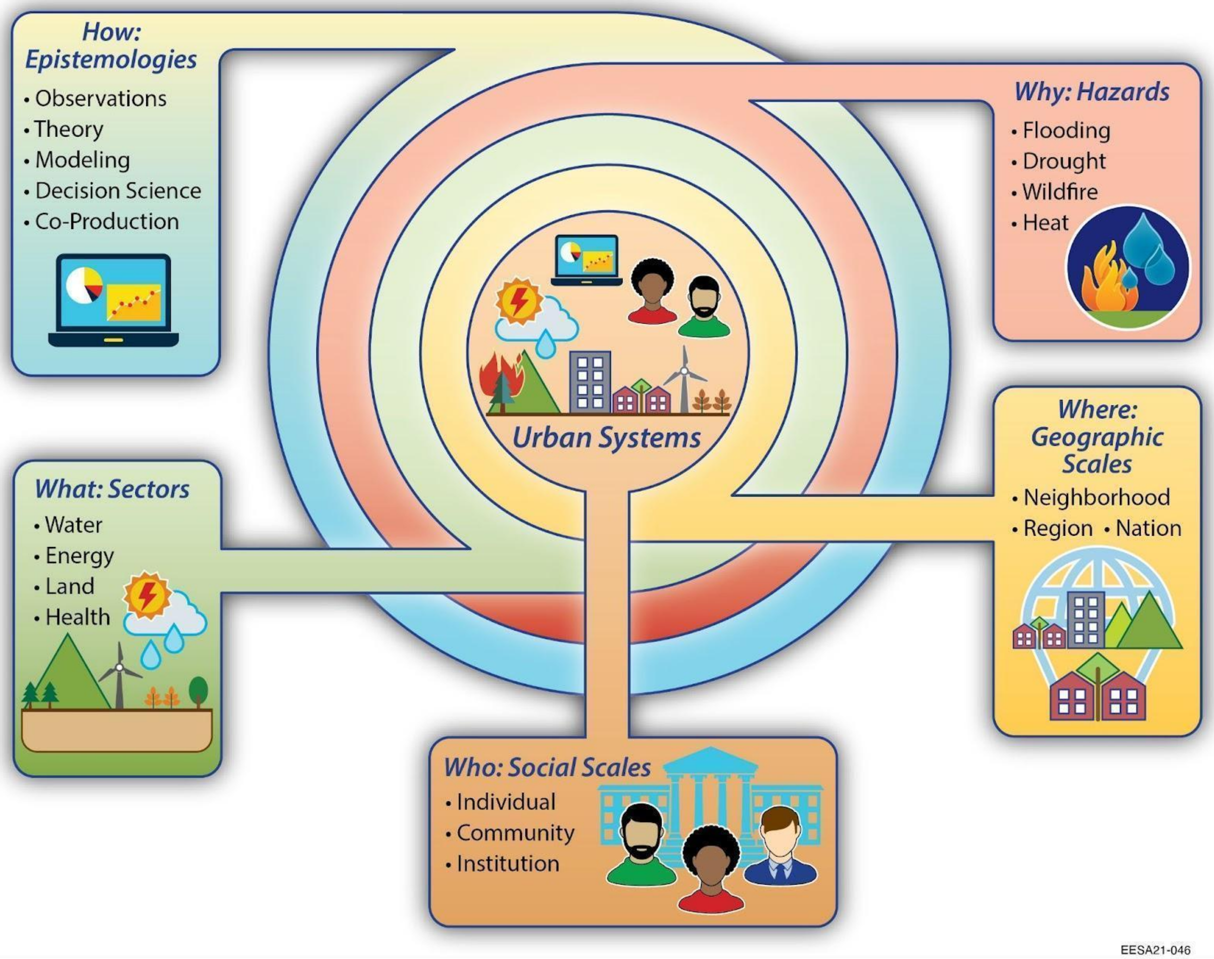


Figure 9-1. Complex Interactions within human systems. As with geophysical processes, important process interactions and heterogeneities in human contexts differ across geographic scales and sectors. These interactions are sensitive to various hazards as well as social mechanisms for organization and response. The understanding of these complex interactions can manifest in multiple ways, including through different statistical and artificial intelligence/machine learning techniques (Source: Reproduced from Brelsford and Jones 2021 with permission from creators Jones, Brelsford, and Swantek).

9.2 Grand Challenges

The grand challenges facing human system dynamics in Earth system predictability can be grouped into three overarching categories: (1) prediction of primary drivers of Earth system change; (2) the impact of Earth system change on human processes and decisions; and finally, (3) complex, coupled two-way interactions and feedbacks between the human processes and Earth systems. The primary human-driven cause of Earth system change over the next century is anthropogenic greenhouse gas (GHG) emissions, while influences of Earth system change on human processes occur at finer scales. Complex feedbacks between the two encompass a range of processes, spanning natural resources (land and water) use and industrial processes that result in changes in net emissions (Figure 9-2).

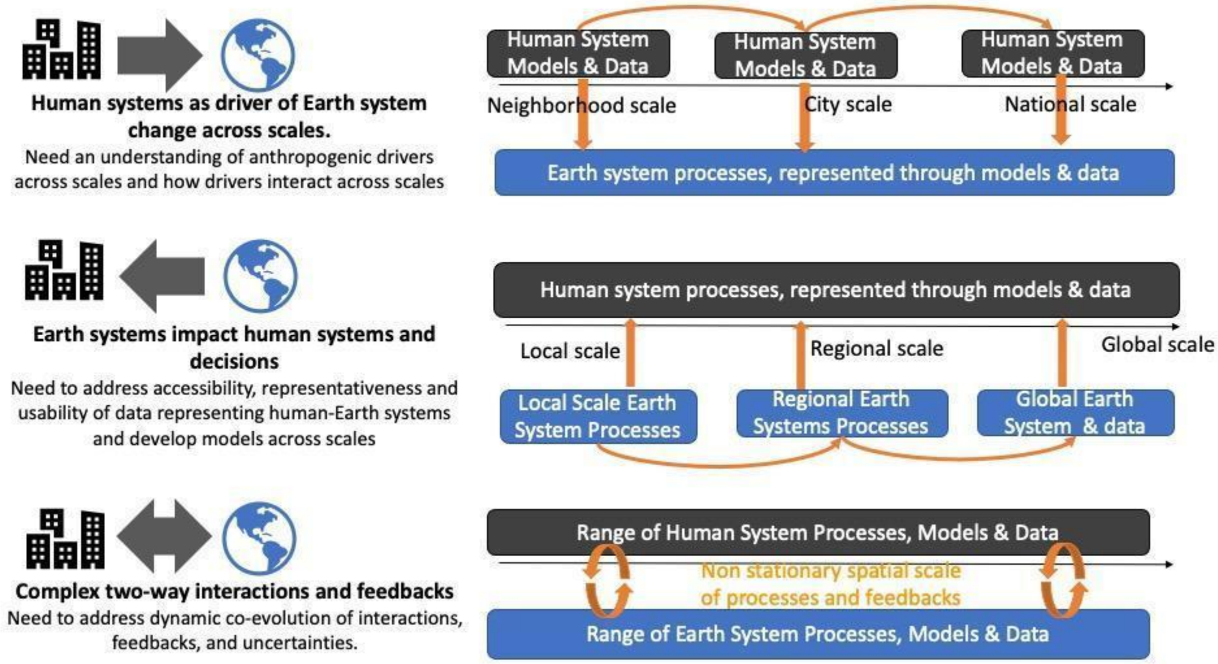
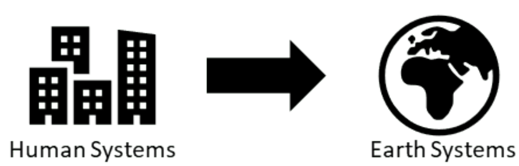


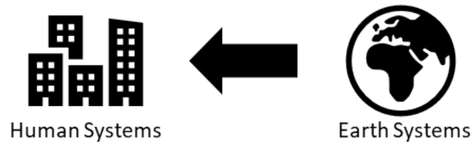
Figure 9-2. High-level synthesis of grand challenges in Earth-human system dynamics research (Sources: Pacific Northwest National Laboratory, Oak Ridge National Laboratory, and Sandia National Laboratory).

9.2.1 Primary Drivers of Earth System Change



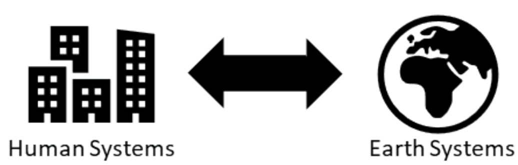
Human GHG emissions are the primary anthropogenic driver of Earth system change. However, there are fundamental limits to the extent to which aggregate GHG emissions from human activities are predictable since human behavior—both individually and collectively—responds to both current conditions and our expectations about the future. To improve our capacity for Earth system prediction, we need to assess how far into the future, across which social scales of aggregation, and over what action spaces our predictions of the evolution of human systems can be improved to identify stable patterns at the Earth system level. While we note that substantial progress has been made in developing a "predictive" capability at smaller scales – for example, extensions of urban scaling theory into human mobility theory (Alessandretti, Aslak, and Lehmann 2020; Bettencourt 2013; Pappalardo et al. 2015) – the development of a predictive capability at the Earth system level is far more challenging, especially as institutions start to explore larger-scale geoengineering options (Bull et al. 2021).

9.2.2 Earth System Impacts to Humans Processes and Decisions



Decision makers need information about how future Earth system changes (particularly hazards whose risk and distributions are changing with the climate) influence and interact with human processes. In order to improve our understanding of Earth system processes at decision-relevant scales, we need to build models that are sufficiently sensitive (e.g., building simulations that demonstrate impacts from heating/cooling degree days) and can traverse across scales (i.e., capture large-scale variations but also enable explorations of smaller-scale influences). However, there is not sufficient information at regional and local scales about how to understand and interpret changes in climate risks and hazards, nor an understanding of how global-scale climate influences global-scale human processes (e.g., Mach et al. 2019). In general, the accessibility, representativeness, and usability of data for supporting integrated human and physical systems understanding (especially for informing evolving priorities, e.g., justice and equity) are underdeveloped and face substantial computational challenges. Furthermore, models coupling human and Earth system processes need to cover a wide range of spatial and temporal scales, as well as calibrating high-dimensional parameter spaces. Thus, the intensive computational demands involved in coupling human systems and physical systems models is a notable challenge. Model interpretability is particularly critical in this area, as the information that decision makers actually see and use can be crucial to understanding the dynamics of the problems they are attempting to address.

9.2.3 Complex, Two-Way Interactions and Feedbacks



Finally, to address the grand challenge of evaluating complex feedbacks between human and Earth system processes, we must determine the appropriate scales and interactions that are needed to produce usable conclusions about these feedbacks and also to adequately capture the co-evolutionary aspect of humans as an active participant in Earth systems. Toward these goals, the first obstacle to overcome may be the problem of limited compatibility between physical models and human models. To improve predictive skill in human systems and dynamics, we need to address spatial nonstationarity and feedbacks between human decision-making and environmental impacts on socioecological systems. This includes (1) interactions between and decision-making dynamics and change in socioecological systems over time as a function of environmental change and extreme events, (2) transferability of models to different local and regional settings to capture spatial heterogeneity in human behavior/social processes and how they interoperate with Earth systems, and (3) accounting for uncertainty in feedbacks between human and Earth system responses. The latter is a critical challenge for all parts of human-Earth system interactions but is particularly significant when doing climate research at decision-relevant scales. The communication of uncertainty extends from technical evaluations, accomplished in part by statistical and machine learning techniques, to dissemination of these analyses to policymakers and the interested public. Doing the latter in an iterative process might be needed to share questions, insight, data, and research outcomes, while following high ethical standards around data privacy (Zipper et al. 2019).

9.3 State-of-the-Science

There are several disciplines that focus on interactions between human and physical systems:

- Engineers and urban scholars have focused on low- and zero-emission cities (Chen et al. 2021; Ramaswami et al. 2016, 2021; Seto et al. 2021).
- Socio-hydrology looks at interactions between human and hydrological processes (Brelsford et al. 2020; Mazzoleni et al. 2021; Müller and Levy 2019).
- Socio-ecological-systems research highlights the dynamics of interactions between ecological processes and human processes (Anderies 2015; Olsson and Jerneck 2018; Olsson and Ness 2019).
- The science of cities looks for universal patterns in urban systems (Alessandretti, Aslak, and Lehmann 2020; Bettencourt and Zünd 2020; Verbavatz and Barthelemy 2020).

Generally, inclusion of human system dynamics in Earth system predictability research spans the triple intersection of datasets, representations, and computational approaches for understanding these complex processes.

Recent advances in human-centric fields have been enabled by newly available digital trace data and advances in our ability to use machine learning to infer characteristics of human activities from satellites designed primarily to measure the non-human world (Golder and Macy 2014). Satellite and remotely sensed information have been processed using machine learning tools to create stand-alone datasets, models, and inputs to models that characterize aggregate human activity in a globally scalable manner (Allen-Dumas et al. 2021) These approaches have been used to gain insights into urban systems, mobility, and water resource activities across scales—from urban to global systems (Table 9-1).

Representations of Earth system effects on human sectors have spanned small energy systems (Alemazkoor et al. 2020; Mosavi et al. 2019) to large wildfire, agriculture, and adaptation policies (Biesbroek, Badloe, and Athanasiadis 2020; Hamrani, Akbarzadeh, and Madramootoo 2020; Wang et al. 2021). Methods used vary from coupling the Global Change Analysis Model (GCAM) with global climate models, integrating AI/ML with agent-based models, developing emulators or less complex surrogate models for physical or human systems models, and performing basic statistical and ML implementations (Table 9-1). Computer simulations have established practices for credibility generation (e.g., verification, validation, uncertainty quantification, sensitivity analyses, counterfactuals). Although some of these practices can be leveraged for AI/ML, credibility processes across the full pipeline are not yet well-established for scientific machine learning (Rushdi and Acquesta 2021). From accounting for data collection-related errors to generalizability assessments beyond initial training sets, much still needs to be done to improve the usability of hybrid simulations, particularly as they relate to human activities.

Table 9-1. Synthesis of recent advances in AI for human system dynamics.

Topic	AI Application(s) and Related Work(s)
Human systems data extension via inference models	• Digital trace data and machine learning to gain insights into urban prediction and long-term urban policy (Alessandretti, Aslak, and Lehmann 2020). • Analysis of high-frequency, big urban data to inform long-term urban policy (Kandt and Batty 2021). • Mechanistic explanation of emerging spatial structure of cities predicts human flows within and through cities through an individual mobility model based on exploration of local neighborhoods and preferential return (Schläpfer et al. 2021). • A scalable computational approach based on the topological properties of digital maps identifies local infrastructural deficits and proposes context-appropriate minimal solutions (Soman et al. 2020). • Addresses the linkages between human systems and Earth systems by coupling GCAM with global climate models through market equilibrium models to represent intricate linkages between energy, water, land, climate, and economic systems (Calvin et al. 2019). • Deterministic classification of human contributions to determine drivers of water scarcity across climate change scenarios (Graham et al. 2020). • Machine learning to explore climate change mitigation in infrastructure and urban resource sectors (Milojevic-Dupont and Creutzig 2021) and building energy performance forecasting (Fathi et al. 2020). • Machine learning discovers and extracts Earth features from satellite images such as land use and land cover, flood inundation extent, and water reservoir storage (Hodgson, Davis, and Kotelenska 2010). • GeoAI infers from remotely sensed imagery and Google Streetview images human activities such as mobility or land use (Bhaduri et al. 2021). • Deep learning and big data analytics fuse global radar and multispectral satellite data to create global/urban/local climate zones classifications covering cities with populations greater than 300,000 (Zhu et al. 2022).

Table 9-1. (Cont.)

Topic	AI Application(s) and Related Work(s)
Human process representation using AIML/hybrid models	● Connecting infrastructure models with physics-based hurricane models to evaluate power outage risk under climate change (Alemazkoor et al. 2020). ● Adaptive neuro-fuzzy inference systems for superior energy demand forecasting (Mosavi et al. 2019). ● Leveraging text analysis to train artificial neural networks to support early detection of climate change adaptation practices (Biesbroek, Badloe, and Athanasiadis 2020). ● Long-short term memory-based deep learning models and random forests to predict GHG emissions from soils (Hamrani, Akbarzadeh, and Madramootoo 2020). ● Ensemble learning methods and game theory to identify drivers of wildfires (Wang et al. 2021). ● Modeling typology to bridge human systems research communities, facilitate the synthesis of scientific advances, and chart new research directions including AIML models and coupling strategies (Yoon et al. 2022). ● Summary statistics to frame new ML applications for domain insights (Zaidi et al. 2018). ● A reinforcement learning approach for irrigation-related decision-making (Chen et al. 2021).
AI/ML as surrogates for complex, computationally intensive physical system models	● Artificial feed-forward neural networks to emulate Community Land Model 5 outputs (Dagon et al. 2020). ● Reducing model complexity through probabilistic calibration of Earth system models (Nicholls et al. 2021). ● Multiyear ENSO prediction using DL (Ham, Kim, and Luo 2019). ● Machine learning prediction of monthly fire emissions over the contiguous United States and comparison to process-based models (Wang et al. 2021a). ● Identification of drivers of wildfire in the contiguous United States via ensemble learning models and game theory Wang et al. 2021b). ● Hybrid machine learning and process-based crop modeling to improve crop yield predictions in the U.S. corn belt (Shahhosseini et al. 2021).

9.4 Experimental, Data, and Modeling Opportunities

> "Just as the invention of the telescope revolutionized the study of the heavens, so too by rendering the unmeasurable measurable, the technological revolution in mobile, Web, and Internet communications has the potential to revolutionize our understanding of ourselves and how we interact … . [T]hree hundred years after Alexander Pope argued that the proper study of mankind should lie not in the heavens but in ourselves, we have finally found our telescope. Let the revolution begin." —(Watts 2012)

A major gap in Earth system predictability is the lack of coupling between Earth and human systems models, including human-induced dynamics, human response patterns, human decision processes, and the dynamic interactions across human and Earth systems. However, artificial intelligence and machine learning coupled with new sets of digital trace data provide a transformational opportunity for our understanding of human-Earth system interactions—that is, digital trace data can serve as the telescope and machine learning as a tool with which we can see patterns in these generally large, inconsistent, and low-quality data (Golder and Macy 2014). Specifically, machine learning can improve the coupling between human-Earth systems in the following ways: (1) "fill in" sparse data about human systems; (2) enable the synchronous bidirectional coupling of human and Earth system models, with uncertainty; and (3) emulate climate data and model outputs to efficiently generate spatially resolved climate information for any future scenario and support scenario discovery.

We know that environmental, social, and built system processes interact to produce outcomes of relevance to resilience, equity, sustainable use of resources, and the global Earth system. However, generating actionable and useful insights for decision-making requires effective integration of data and modeling, including incorporation of relevant data into models, delivery of modeling results, and inclusion of human responses into the model. Because capturing these details is primarily a data collection process, data analytical techniques may be more useful than mechanistic or process-based models for understanding these relationships. Machine learning and artificial intelligence methods may provide the appropriate platforms for understanding multiple types and sources of data and explaining relationships that are not yet well understood.

9.4.1 Develop Data about Human Systems

AI/ML could also be leveraged to generate data for coupled or integrated human system-Earth system models that may be unavailable or challenging/expensive to acquire. Some human activity data that might be useful (e.g., agriculture and infrastructure characteristics) are proprietary, and data at the individual scale must also respect the individual's right to privacy. Despite these challenges, developing sustainable and FAIR (i.e., findable, accessible, interoperable, and reproducible) data streams to represent anthropogenic processes and coupled human-natural systems is one of the most actionable opportunities to increase the pace of scientific discovery in Human Systems and Dynamics. For example, established work in natural language processing (NLP) and other data mining techniques provides an opportunity to both extract more out of existing data sources and to develop novel data streams to address the need for data on land use change, resource consumption, mobility, behavior, infrastructure demand, infrastructure characteristics, and hazard mitigation strategies (e.g., through web scraping and digitization of historical records). In some cases, data that are currently mostly privately owned or proprietary come with commercial, privacy, and ethics concerns. Using application programming interfaces (APIs) to access anonymized versions of data still held and maintained by the owners in mutually beneficial partnerships with the data owners is another opportunity for enabling access to the diverse data needed for human systems research. Finally, collaborations with citizen-level data entities (e.g., https://www.flowminder.org/) could be pursued to open up a new realm of possible studies that can connect human activity to environmental stimulus and response. AI/ML can also be used to provide estimates of observational data at the high-resolution scales that are becoming increasingly important for Earth system research, especially as issues such as equity are becoming a focal research perspective. This is particularly the case for the socioeconomic data necessary for adequate characterization/representation of human systems. For example, AI/ML techniques to downscale block-level U.S. Census data to even finer resolutions and develop artificial agent populations (e.g., Graetz, Ummel, and Aldana Cohen 2021; Tuccillo 2021; Tuccillo and Spielman 2022) could be incredibly useful for enabling high-resolution agent-based modeling (ABM) efforts in the Earth system space. With the large amount of newly available data from remote sensing imagery, in situ observations, and unconventional sector, social, and economic data complemented by citizen science observations, AI/ML is providing an opportunity to rethink coupled human-Earth system science in the context of big data.

9.4.2 Coupling of Human and Earth System Models

We can develop integrated frameworks for these systems by leveraging data from both human and natural systems and integrating these systems using machine learning and artificial intelligence. One type of integration could be a hybrid model encompassing machine learning, data flow, and ABM. These integrated methods could also leverage insights from non-ML implementations of interactions between human and Earth systems at varying scales (Arneth, Brown, and Rounsevell 2014; Nazemi and Wheater 2015; Thornton et al. 2017; Calvin et al. 2019) by using data-driven causal inference learning processes such as causal complex network analysis to identify feedback among the processes. We can also harness machine learning to relate patterns in data across scales and to facilitate reduction of model bias (Molajou, Pouladi, and Afshar 2021; Sert, Bar-Yam, and Morales 2020). Additional opportunities within this space include the synchronous, bidirectional integration of existing human system models with existing Earth system or specific domains (Calvin et al. 2019) as well as exploitation of transfer learning methods. Such methods will both improve the individual existing models and aid direct integration (Simpson et al. 2021) by allowing for the use of diverse data sources and model results across heterogeneous scales. To facilitate these tasks, method development and testing should incorporate synthetic datasets and benchmarking standards. Relevant and hard test problems that share key challenges should be developed among different research communities.

Machine learning can help facilitate this compatibility by representing human activity as distributions of inputs to physical model processes and by providing analysis of physical model output. In addition to employing machine learning in these coupled ways, machine learning may help us determine which human activities (e.g., infrastructure construction, land use change, fertilizer use) or results of human activities (e.g., changes in land, water, and atmospheric processes and net emissions) most impact Earth system predictability and should be captured in a coupled context. With further AI research, this process can enable improvements in, for example, rapid detections in changes in demand, demand-side management during stresses and shocks, and automatic remote infrastructure (re)configuration such as virtual power plants (Gilrein et al. 2021; Helmrich et al. 2021).

9.4.3 Integration of Social Sciences in Representation of Human-Earth System Interactions, including Uncertainties

We can build on methodologies developed and lessons learned from the social sciences by incorporating social insights and associated datasets into algorithms and modeling approaches (e.g., is there a "digital twin" equivalent of cultural preferences toward tilling?). Such nuances will become especially critical for addressing barriers to adoption of adaptation and mitigation activities. We need careful consideration of the generalizability of algorithms (i.e., is a model trained on region A able to generate valuable/robust insights for region B?). This requires explicit attention to the assumptions about human dynamics that are being encoded into the algorithm. Finally, quantifying the uncertainty of large coupled systems is a challenge that will require large-scale computing for multiple-scenario simulations and new modeling methodology. Human system data used to inform models typically have unknown and highly variable uncertainty, but inferring a meaningful signal from these data is necessary to improve our predictive skill. Such use of AI/ML directly addresses the Grand Challenge of integrating complex, coupled human system-physical system models for Earth system predictability.

9.5 Research Priorities

Human systems are both the locus of impact for Earth system dynamics and the locus of decision-making for climate hazard mitigation and adaptation. As such, the core research priorities in this space center around expanding the scope of Earth system research to directly include assessment of climate risks on human systems, tuning research questions so that they are relevant for informing decision makers' choices, and addressing current gaps (Table 9-2).

Table 9-2. Gaps identified by white papers.

Topic	Gaps Identified	Whitepaper ID
Using AI for addressing data gaps	Address data sparsity and data fusion of disparate, multisource data.	AI4ESP1114
	Develop the more refined datasets (e.g., more refined LULC details in urban systems) that are needed.	AI4ESP1016
	Integrate LU with LC portions of datasets.	AI4ESP1137
	Develop ML algorithms to downscale/upscale data.	AI4ESP1093
	Decompose global environmental change data into locally resolved processes using probabilistic components within RNN (LSTM) to capture uncertainty.	AI4ESP1001

Table 9-2. (Cont.)

Topic	Gaps Identified	Whitepaper ID
Integrating knowledge into AI implementations	Address lack of foundational (i.e., theoretical, systemic) understanding of different approaches to knowledge-guided AI.	AI4ESP1138
	Fundamentally advance uncertainty quantification by integrating knowledge- and physics-inspired models.	AI4ESP1019
	Use Driver-Pressure-State-Impact-Response framework for supporting decision-making activities.	AI4ESP1024
	Address "data floods" by using heuristics methods to incorporate model/theory-based information into remote sensing model parameterization.	AI4ESP1040
	Extend beyond simple economic representations of agents to understand behavior.	AI4ESP1137
Use AI to drive understanding of causality	Use DNN/RNN to find latent features at multiple scales of analysis (e.g., in disease outbreaks).	AI4ESP1106
	Use ML techniques for predicting model bias and uncertainty as a way to explore data-driven causal inferences within interactions/feedbacks.	AI4ESP1093
	Dynamically trace pathways from south to impact of dominant drivers of observed climate changes.	AI4ESP1020
	Extract information across scales using Bayesian network models.	AI4ESP1029
Increase computational efficiency	Use NN for fast solvers (e.g., for urban hydrodynamics).	AI4ESP1016
	Pursue hierarchical integration of high-fidelity physics-based models with ML-derived surrogates.	AI4ESP1101
	Develop hybrid models to replace simulation of empirical processes.	AI4ESP1093
	Use ML techniques in combination with probabilistic methods and hierarchy of models to increase efficiency of multiple runs.	AI4ESP1020
Evaluation of impacts	Use AI/ML to extend lead times for extreme weather event-related, engineered infrastructure impacts (e.g., power outages) to longer-term planning.	AI4ESP1041 AI4ESP1068
	Use explainable AI to identify human influence.	AI4ESP1029

Priority: Using AI for addressing human systems data gaps by creating better datasets, especially by extending and validating the sparse data available. Top priority datasets include those related to (1) social vulnerability, (2) physical vulnerability, (3) exposure to hazards, (4) social and economic impacts, (5) responses to change, and (6) assessment of existing response. Much of the climate research community is wrestling with the computational and logistical demands of the data-intensive research ecosystem that ML requires. Human-centric data bring additional challenges: data about human systems must both respect the dignity and privacy of individuals represented in the data and maintain the security of information about built infrastructure data. Much of the data about humans is owned by large, private corporations, which brings additional challenges for data access and reuse. Facilitating data workflows which address these challenges is a critical step for integrating human systems into Earth systems research.

In particular, one focused objective under these priorities could be to obtain a meaningful and semantically enriched representation of the high dimensional, multimodal geospatial data in a lower dimensional vector representation such that similar objects in high dimension tend to be closer in the embedded space. Such embedded representation is useful for various downstream tasks such as urban change detections or prediction of demographic-related information. These geospatial representations of anthropogenic processes are also then in a form familiar to Earth scientists. Another focused objective could be expanding the use of machine learning to make inferences about the state of human systems based on textual, remote sensing, and digital trace datasets. For example, convolutional neural networks can be used to infer characteristics of the built environment from remotely sensed imagery, and AI-driven data fusion can be used to integrate diverse datasets.

Priority: Integrating human systems knowledge into AI4ESP implementations. This can improve Earth system predictability by improving both the representations of human drivers of uncertainty in the various Earth system sectors, and the predictability of human systems as a fundamental component/sector of the Earth system in their own right. For example, machine learning can improve our understanding of the human-Earth system coupling through climate emulation, which enables us to generate spatially resolved climate information for any future scenario, capture different drivers/processes for scenario discovery, and analyze large ensembles of model results. Which human sectors are the determining factors for Earth system processes, and which are the biggest Earth system influences on human system dynamics and resilience? Emulators provide options for fast, cheap, and flexible model coupling and exploration of feedbacks between human and Earth systems. Unsupervised methods can help identify what information is critical and decision-relevant and identify patterns across scales. AI-driven, agent-based models can be a bridge between data and simulation models, and Long Short-Term Memory networks based on Recurrent Neural Networks (LSTM based on RNN) can represent regional Earth and human system dynamic processes. Further, this integration is an intrinsically interdisciplinary research area, and it requires that the dissemination of research across disciplines becomes a research priority in parallel with the actual domain and integration research. To aid dissemination and integration of results as a research priority, support for and expected adherence to FAIR standards in research methods and data will be crucial. Additionally, the use of existing ML-driven information extraction techniques to create shared catalogs of data and model results in a searchable location can aid this dissemination. Successfully prioritizing dissemination of results will ultimately improve these multidisciplinary models for Earth system predictability, expanding on current efforts (Peng et al. 2021) to map different dimensions of human systems and interactions according to their usefulness, implementability in code, and importance to different stakeholders.

Priority: Exploring new, fast-evolving human-Earth systems dynamics and validation. Multiple questions arise regarding the credibility of AI implementations as they expand into complex, human system dynamics. Namely, questions of verification (i.e., did the model accurately execute), validation (i.e., does the model accurately represent the real world), and uncertainty quantification will need to be addressed to increase the confidence of implementations. Initial efforts to reconcile definitions of important terms that are used differently across disciplines will be critical to successful implementations of credibility assessments. Traditional AI implementations rely on ground truth data for credibility assessments, especially for validation (i.e., training/testing sets). However, such ground truth data will be limited as AI implementations expand into data assimilation, causality evaluation, and knowledge integration methods for human systems. Therefore, new research techniques will be required to build both confidence of AI implementations to support subsequent analyses and decision-making as well as model generalizability to support advancement of science in human system dynamics. Such techniques could range from comparisons to theory-based estimates, evaluation of encoded assumptions, and systematic explorations (from parameter-level to scenario-level) of uncertainty within dynamically coupled system models.

9.6 Short-, Mid-, and Long-term Goals: <5 years, 5-year, and 10-year Goals

We have established that the representation of human systems in Earth systems models has so far been largely limited to the physical interface and is in its infancy in representing the science of decision-making—the action space. The representation of human systems and behaviors is mature, yet we identified gaps across individual and aggregated social scales of decision-making, and we also established that the science of human decision-making across systems is still in its infancy. Activities to advance the current state of the art span science objectives, methodological process, and development of resources across time (from <5 years to 10 years+).

9.6.1 Short-Term Goals

Short-term priorities span across: (1) development of science objectives, (2) methodological progress in developing benchmarks for AI/ML as well as new modeling techniques, and (3) development of compatible platforms to support data sharing and model executions. Given the infancy of the integration of human and Earth systems, independently from the AI/ML opportunities, science efforts should focus on improving: the spatiotemporal representation of human-induced processes, our understanding of multiscale interactions, and developing generalization of localized human systems for accelerating dissemination across scales by identifying critical pieces of information to support decision-making at various scales.

Methodological activities in support of these science priorities require flexible approaches to support advancements in both process-based understanding and incremental understanding that enable substitution of models for exploration of the unknowns. In the case of AI/ML, the latter requires development of benchmarks (e.g., AI-driven ABMs as a "bridge" between data and simulation models) as well as new modeling conceptions (e.g., hybrid/surrogate models and exploration of parameter space). Finally, with regard to resources, development is needed of compatible or hybrid platforms that can handle the different required execution timeframes and modeling time horizons—as well as partnerships that can generate mutually beneficial and privacy-informed approaches to enable use of (often proprietary) fine-scale data most relevant to human systems to advance critical research in this space.

9.6.2 Mid-Term Goals

While the short term is about scientific and methodological advances for human systems predictability, we envision that the mid-term priorities will be making progress with workflows toward more systematic and reproducible and generalizable integration of human and Earth systems. Specific priorities for the mid-term include: (1) development of an institutional environment for conducting AI4ESP science, along with the (2) implementation of the science (hardware/software and workflows under FAIR principles), and (3) education. Specifically, the institutional environment for research approaches will need to continue evolving toward complementary paradigms (process-based and AI/ML) to enable more rapid discovery of human systems interactions to keep pace with the fast co-evolution of complex processes of human and Earth systems. Additionally, we also need workflows that move us toward better standards and protocols to allow different models to intercommunicate. Data-/model-sharing protocols will need to become seamless as supported by new hardware and software, dramatically advancing scientific discovery. Finally, with regard to teaming and education, we need to invest in pipeline development to generate expertise through closer collaborations between domain scientists (and especially bringing in social science expertise) and methodological scientists. Potential ideas for the latter include establishing dedicated research centers, possibly in partnership with universities that have established method expertise, and in developing more dedicated internship opportunities to enhance relationships with advisors at universities as a long-term investment.

9.6.3 Long-Term Goals

The short- and mid-term goals will support a long-term goal focused on development of AI/ML-supported, fully coupled human-Earth systems and platforms with multiple complementary ways to carry out science related to human systems including their interactions with Earth systems. Specifically, the long-term goal involves: (1) workflows, hardware, and software that support hybrid process-based and AI4ESP-based science for both incremental and fundamental discoveries and fast-evolving fundamental co-evolution discoveries; (2) sustainable data streams (generalizable, FAIR, supporting process-based and AI4ESP types of science) to support the coupling of natural and human systems; and (3) established human system science at the interface of Earth-human systems. Effective execution of these activities will likely require an iterative process as well as a pipeline of expertise and closer collaborations between domain scientists (especially social scientists) and methodological scientists. This is an ambitious agenda, yet it is key to meeting the nation's needs to achieve intended objectives given that societal transitions are rapidly occurring in these tightly connected systems.

9.7 References

Alemazkoor, Negin, Benjamin Rachunok, Daniel R. Chavas, Andrea Staid, Arghavan Louhghalam, Roshanak Nateghi, and Mazdak Tootkaboni. 2020. "Hurricane-Induced Power Outage Risk under Climate Change Is Primarily Driven by the Uncertainty in Projections of Future Hurricane Frequency." *Scientific Reports* 10 (1): 15270. https://doi.org/10.1038/s41598-020-72207-z.

Alessandretti, Laura, Ulf Aslak, and Sune Lehmann. 2020. "The Scales of Human Mobility." *Nature* 587 (7834): 402–7. https://doi.org/10.1038/s41586-020-2909-1.

Allen-Dumas, Melissa R., Haowen Xu, Kuldeep R. Kurte, and Deeksha Rastogi. 2021. "Toward Urban Water Security: Broadening the Use of Machine Learning Methods for Mitigating Urban Water Hazards." *Frontiers in Water* 2. https://doi.org/10.3389/frwa.2020.562304.

Anderies, John M. 2015. "Managing Variance: Key Policy Challenges for the Anthropocene." *Proceedings of the National Academy of Sciences* 112 (47): 14402–3. https://doi.org/10.1073/pnas.1519071112.

Arneth, A., C. Brown, and M. D. A. Rounsevell. 2014. "Global Models of Human Decision-Making for Land-Based Mitigation and Adaptation Assessment." *Nature Climate Change* 4 (7): 550–57. https://doi.org/10.1038/nclimate2250.

Bettencourt, Luís M. A. 2013. "The Origins of Scaling in Cities." *Science* 340 (6139): 1438–41. https://doi.org/10.1126/science.1235823.

Bettencourt, Luís M. A., and Daniel Zünd. 2020. "Demography and the Emergence of Universal Patterns in Urban Systems." *Nature Communications* 11 (1): 4584. https://doi.org/10.1038/s41467-020-18205-1.

Bhaduri, Budhendra, Ryan McManamay, Olufemi Omitaomu, Jibo Sanyal, and Amy Rose. 2021. "Urban Energy Systems: Research at Oak Ridge National Laboratory." In *Urban Informatics*, edited by Wenzhong Shi, Michael F. Goodchild, Michael Batty, Mei-Po Kwan, and Anshu Zhang, 281–308. Singapore: Springer Singapore. https://doi.org/10.1007/978-981-15-8983-6_18.

Biesbroek, Robbert, Shashi Badloe, and Ioannis N. Athanasiadis. 2020. "Machine Learning for Research on Climate Change Adaptation Policy Integration: An Exploratory UK Case Study." *Regional Environmental Change* 20 (3): 85. https://doi.org/10.1007/s10113-020-01677-8.

Brelsford, Christa, Marion Dumas, Edella Schlager, Brian J. Dermody, Michael Aiuvalasit, Melissa R. Allen-Dumas, Janice Beecher, et al. 2020. “Developing a Sustainability Science Approach for Water Systems.” *Ecology and Society* 25 (2). https://doi.org/10.5751/ES-11515-250223.

Brelsford, Christa, and Andrew Jones. 2021. “Multi-Sectoral Urban Interactions. Fundamental Science Needs to Inform Pathways to More Resilient Communities in a Changing Climate [Workshop Report].” https://multisectordynamics.org/working-groups/urban/workshop-report-multi-sectoral-urban-interactions/.

Bull, Diana, Laura Swiler, Kara Peterson, Irina Tezaur, Lyndsay Shand, Warren Davis, Benjamin Wagman, Matt Peterson, and Erika Roesler. 2021. “Climate Intervention Assessment and Attribution.” United States. https://doi.org/10.2172/1769687.

Calvin, K., P. Patel, L. Clarke, G. Asrar, B. Bond-Lamberty, R. Y. Cui, A. di Vittorio, et al. 2019. “GCAM v5.1: Representing the Linkages between Energy, Water, Land, Climate, and Economic Systems.” *Geoscientific Model Development* 12 (2): 677–98. https://doi.org/10.5194/gmd-12-677-2019.

Calvin, Katherine, Ben Bond-Lamberty, Andrew Jones, Xiaoying Shi, Alan di Vittorio, and Peter Thornton. 2019. “Characteristics of Human-Climate Feedbacks Differ at Different Radiative Forcing Levels.” *Global and Planetary Change* 180: 126–35. https://doi.org/https://doi.org/10.1016/j.gloplacha.2019.06.003.

Chen, Mengting, Yuanlai Cui, Xiaonan Wang, Hengwang Xie, Fangping Liu, Tongyuan Luo, Shizong Zheng, and Yufeng Luo. 2021. “A Reinforcement Learning Approach to Irrigation Decision-Making for Rice Using Weather Forecasts.” *Agricultural Water Management* 250: 106838. https://doi.org/https://doi.org/10.1016/j.agwat.2021.106838.

Chen, Shaoqing, Kai Fang, Shobhakar Dhakal, Ali Kharrazi, Kangkang Tong, and Anu Ramaswami. 2021. “Reshaping Urban Infrastructure for a Carbon-Neutral and Sustainable Future.” *Resources Conservation and Recycling* 174 (November): 105765. https://doi.org/10.1016/j.resconrec.2021.105765.

Coen, Deborah R. 2021. “A Brief History of Usable Climate Science.” *Climatic Change* 167 (3): 51. https://doi.org/10.1007/s10584-021-03181-2.

Dagon, K., B. M. Sanderson, R. A. Fisher, and D. M. Lawrence. 2020. “A Machine Learning Approach to Emulation and Biophysical Parameter Estimation with the Community Land Model, Version 5.” *Advances in Statistical Climatology, Meteorology and Oceanography* 6 (2): 223–44. https://doi.org/10.5194/ascmo-6-223-2020.

Fathi, Soheil, Ravi Srinivasan, Andriel Fenner, and Sahand Fathi. 2020. “Machine Learning Applications in Urban Building Energy Performance Forecasting: A Systematic Review.” *Renewable and Sustainable Energy Reviews* 133: 110287. https://doi.org/https://doi.org/10.1016/j.rser.2020.110287.

Gilrein, Erica J., Thomaz M. Carvalhaes, Samuel A. Markolf, Mikhail v Chester, Braden R. Allenby, and Margaret Garcia. 2021. “Concepts and Practices for Transforming

Infrastructure from Rigid to Adaptable." *Sustainable and Resilient Infrastructure* 6 (3–4): 213–34. https://doi.org/10.1080/23789689.2019.1599608.
Golder, Scott A., and Michael W. Macy. 2014. "Digital Footprints: Opportunities and Challenges for Online Social Research." *Annual Review of Sociology* 40 (1): 129–52. https://doi.org/10.1146/annurev-soc-071913-043145.
Graetz, Nick, Kevin Ummel, and Daniel Aldana Cohen. 2021. "Small-Area Analyses Using Public American Community Survey Data: A Tree-Based Spatial Microsimulation Technique." *Sociological Methodology* 52 (1): 53–74. https://doi.org/10.1177/00811750211057572.
Graham, Neal T., Mohamad I. Hejazi, Min Chen, Evan G. R. Davies, James A. Edmonds, Son H. Kim, Sean W. D. Turner, et al. 2020. "Humans Drive Future Water Scarcity Changes across All Shared Socioeconomic Pathways." *Environmental Research Letters* 15 (1): 14007. https://doi.org/10.1088/1748-9326/ab639b.
Ham, Yoo-Geun, Jeong-Hwan Kim, and Jing-Jia Luo. 2019. "Deep Learning for Multi-Year ENSO Forecasts." *Nature* 573 (7775): 568–72. https://doi.org/10.1038/s41586-019-1559-7.
Hamrani, Abderrachid, Abdolhamid Akbarzadeh, and Chandra A. Madramootoo. 2020. "Machine Learning for Predicting Greenhouse Gas Emissions from Agricultural Soils." *Science of The Total Environment* 741: 140338. https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.140338.
Helmrich, Alysha, Samuel Markolf, Rui Li, Thomaz Carvalhaes, Yeowon Kim, Emily Bondank, Mukunth Natarajan, Nasir Ahmad, and Mikhail Chester. 2021. "Centralization and Decentralization for Resilient Infrastructure and Complexity." *Environmental Research: Infrastructure and Sustainability* 1 (2): 21001. https://doi.org/10.1088/2634-4505/ac0a4f.
Hodgson, Michael, Bruce Davis, and Jitka Kotelenska. 2010. "Remote Sensing and GIS Data/Information in the Emergency Response/Recovery Phase." In book: *Geospatial Techniques in Urban Hazard and Disaster Analysis*, 327–54. https://doi.org/10.1007/978-90-481-2238-7_16.
Kandt, Jens, and Michael Batty. 2021. "Smart Cities, Big Data and Urban Policy: Towards Urban Analytics for the Long Run." *Cities* 109: 102992. https://doi.org/https://doi.org/10.1016/j.cities.2020.102992.
Mach, Katharine J., Caroline M. Kraan, W. Neil Adger, Halvard Buhaug, Marshall Burke, James D. Fearon, Christopher B. Field, et al. 2019. "Climate as a Risk Factor for Armed Conflict." *Nature* 571 (7764): 193–97. https://doi.org/10.1038/s41586-019-1300-6.
Mazzoleni, Maurizio, Vincent O Odongo, Elena Mondino, and Giuliano di Baldassarre. 2021. "Water Management, Hydrological Extremes, and Society: Modeling Interactions and Phenomena." *Ecology and Society* 26 (4). https://doi.org/10.5751/ES-12643-260404.
Milojevic-Dupont, Nikola, and Felix Creutzig. 2021. "Machine Learning for Geographically Differentiated Climate Change Mitigation in Urban Areas." *Sustainable Cities and Society* 64: 102526. https://doi.org/https://doi.org/10.1016/j.scs.2020.102526.

Molajou, Amir, Parsa Pouladi, and Abbas Afshar. 2021. “Incorporating Social System into Water-Food-Energy Nexus.” *Water Resources Management* 35 (13): 4561–80. https://doi.org/10.1007/s11269-021-02967-4.

Mosavi, Amir, Mohsen Salimi, Sina Faizollahzadeh Ardabili, Timon Rabczuk, Shahaboddin Shamshirband, and Annamaria R Varkonyi-Koczy. 2019. “State of the Art of Machine Learning Models in Energy Systems, a Systematic Review.” *Energies* 12 (7). https://doi.org/10.3390/en12071301.

Müller, Marc F., and Morgan C. Levy. 2019. “Complementary Vantage Points: Integrating Hydrology and Economics for Sociohydrologic Knowledge Generation.” *Water Resources Research* 55 (4): 2549–71. https://doi.org/https://doi.org/10.1029/2019WR024786.

Nazemi, A., and H. S. Wheater. 2015. “On Inclusion of Water Resource Management in Earth System Models – Part 1: Problem Definition and Representation of Water Demand.” *Hydrology and Earth System Sciences* 19 (1): 33–61. https://doi.org/10.5194/hess-19-33-2015.

Nicholls, Z., M. Meinshausen, J. Lewis, M. Rojas Corradi, K. Dorheim, T. Gasser, R. Gieseke, et al. 2021. “Reduced Complexity Model Intercomparison Project Phase 2: Synthesizing Earth System Knowledge for Probabilistic Climate Projections.” *Earth’s Future* 9 (6): e2020EF001900. https://doi.org/https://doi.org/10.1029/2020EF001900.

Olsson, Lennart, and Anne Jerneck. 2018. “Social Fields and Natural Systems: Integrating Knowledge about Society and Nature.” *Ecology and Society* 23 (3). https://doi.org/10.5751/ES-10333-230326.

Olsson, Lennart, and Barry Ness. 2019. “Better Balancing the Social and Natural Dimensions in Sustainability Research.” *Ecology and Society* 24 (4). https://doi.org/10.5751/ES-11224-240407.

Pappalardo, Luca, Filippo Simini, Salvatore Rinzivillo, Dino Pedreschi, Fosca Giannotti, and Albert-László Barabási. 2015. “Returners and Explorers Dichotomy in Human Mobility.” *Nature Communications* 6 (1): 8166. https://doi.org/10.1038/ncomms9166.

Peng, Wei, Gokul Iyer, Valentina Bosetti, Vaibhav Chaturvedi, James Edmonds, Allen Fawcett, Stéphane Hallegatte, David Victor, Detlef Vuuren, and John Weyant. 2021. “Climate Policy Models Need to Get Real about People—Here’s How.” *Nature* 594 (June): 174–76. https://doi.org/10.1038/d41586-021-01500-2.

Ramaswami, Anu, Armistead G. Russell, Patricia J. Culligan, Karnamadakala Rahul Sharma, and Emani Kumar. 2016. “Meta-Principles for Developing Smart, Sustainable, and Healthy Cities.” *Science* 352 (6288): 940–43. https://doi.org/10.1126/science.aaf7160.

Ramaswami, Anu, Kangkang Tong, Josep G. Canadell, Robert B. Jackson, Eleanor (Kellie) Stokes, Shobhakar Dhakal, Mario Finch, et al. 2021. “Carbon Analytics for Net-Zero Emissions Sustainable Cities.” *Nature Sustainability* 4 (6): 460–63. https://doi.org/10.1038/s41893-021-00715-5.

Rushdi, Ahmad, and Erin Acquesta. 2021. “Credibility in Scientific Machine Learning: Data Verification and Model Qualification.” June 2021.

https://cfwebprod.sandia.gov/cfdocs/CompResearch/templates/insert/newsitem.cfm?news=2148.
Schläpfer, Markus, Lei Dong, Kevin O'Keeffe, Paolo Santi, Michael Szell, Hadrien Salat, Samuel Anklesaria, Mohammad Vazifeh, Carlo Ratti, and Geoffrey B. West. 2021. "The Universal Visitation Law of Human Mobility." *Nature* 593 (7860): 522–27. https://doi.org/10.1038/s41586-021-03480-9.
Sert, Egemen, Yaneer Bar-Yam, and Alfredo J Morales. 2020. "Segregation Dynamics with Reinforcement Learning and Agent Based Modeling." *Scientific Reports* 10 (1): 11771. https://doi.org/10.1038/s41598-020-68447-8.
Seto, Karen C., Galina Churkina, Angel Hsu, Meredith Keller, Peter W. G. Newman, Bo Qin, and Anu Ramaswami. 2021. "From Low- to Net-Zero Carbon Cities: The Next Global Agenda." *Annual Review of Environment and Resources* 46 (1): 377–415. https://doi.org/10.1146/annurev-environ-050120-113117.
Shahhosseini, Mohsen, Guiping Hu, Isaiah Huber, and Sotirios v Archontoulis. 2021. "Coupling Machine Learning and Crop Modeling Improves Crop Yield Prediction in the US Corn Belt." *Scientific Reports* 11 (1): 1606. https://doi.org/10.1038/s41598-020-80820-1.
Shughrue, Chris, B. T. Werner, and Karen C. Seto. 2020. "Global Spread of Local Cyclone Damages through Urban Trade Networks." *Nature Sustainability* 3 (8): 606–13. https://doi.org/10.1038/s41893-020-0523-8.
Simpson, Nicholas P., Katharine J. Mach, Andrew Constable, Jeremy Hess, Ryan Hogarth, Mark Howden, Judy Lawrence, et al. 2021. "A Framework for Complex Climate Change Risk Assessment." *One Earth* 4 (4): 489–501. https://doi.org/https://doi.org/10.1016/j.oneear.2021.03.005.
Soman, Satej, Anni Beukes, Cooper Nederhood, Nicholas Marchio, and Luís M. A. Bettencourt. 2020. "Worldwide Detection of Informal Settlements via Topological Analysis of Crowdsourced Digital Maps." *ISPRS International Journal of Geo-Information* 9 (11). https://doi.org/10.3390/ijgi9110685.
Thornton, Peter E., Katherine Calvin, Andrew D. Jones, Alan v di Vittorio, Ben Bond-Lamberty, Louise Chini, Xiaoying Shi, et al. 2017. "Biospheric Feedback Effects in a Synchronously Coupled Model of Human and Earth Systems." *Nature Climate Change* 7 (7): 496–500. https://doi.org/10.1038/nclimate3310.
Tuccillo, Joe. 2021. "An Individual-Centered Approach for Geodemographic Classification." In: United States. https://www.osti.gov/biblio/1823347.
Tuccillo, Joseph v., and Seth E. Spielman. 2022. "A Method for Measuring Coupled Individual and Social Vulnerability to Environmental Hazards." *Annals of the American Association of Geographers*, January, 1–24. https://doi.org/10.1080/24694452.2021.1989283.
Verbavatz, Vincent, and Marc Barthelemy. 2020. "The Growth Equation of Cities." *Nature* 587 (7834): 397–401. https://doi.org/10.1038/s41586-020-2900-x.
Wang, S. S.-C., Y. Qian, L. R. Leung, and Y. Zhang. 2021a. "Interpretable Machine Learning Prediction of Fire Emissions and Comparison with FireMIP Process-Based Models."

Atmospheric Chemistry and Physics Discussions 2021: 1–33. https://doi.org/10.5194/acp-2021-634.

Wang, Sally S.-C., Yun Qian, L. Ruby Leung, and Yang Zhang. 2021b. "Identifying Key Drivers of Wildfires in the Contiguous US Using Machine Learning and Game Theory Interpretation." *Earth's Future* 9 (6): e2020EF001910. https://doi.org/https://doi.org/10.1029/2020EF001910.

Watts, Duncan J. 2012. *Everything Is Obvious: How Common Sense Fails Us*. Currency. https://www.penguinrandomhouse.com/books/187477/everything-is-obvious-by-duncan-j-watts/.

Yoon, Jim, Patricia Romero-Lankao, Yi-Chen Ethan Yang, Christian Joachim Anton Klassert, Nathan M. Urban, Kendra Elena Kaiser, Klaus Keller, et al. 2022. "A Typology for Characterizing Human Action in MultiSector Dynamics Models." *Earth and Space Science Open Archive*, 35. https://doi.org/10.1002/essoar.10510028.1.

Zaidi, Syed Mohammed Arshad, Varun Chandola, Melissa R. Allen, Jibonananda Sanyal, Robert N. Stewart, Budhendra L. Bhaduri, and Ryan A. McManamay. 2018. "Machine Learning for Energy-Water Nexus: Challenges and Opportunities." *Big Earth Data* 2 (3): 228–67. https://doi.org/10.1080/20964471.2018.1526057.

Zhu, Xiao Xiang, Chunping Qiu, Jingliang Hu, Yilei Shi, Yuanyuan Wang, Michael Schmitt, and Hannes Taubenböck. 2022. "The Urban Morphology on Our Planet – Global Perspectives from Space." *Remote Sensing of Environment* 269: 112794. https://doi.org/https://doi.org/10.1016/j.rse.2021.112794.

Zipper, Samuel C., Kaitlin Stack Whitney, Jillian M. Deines, Kevin M. Befus, Udit Bhatia, Sam J. Albers, Janice Beecher, et al. 2019. "Balancing Open Science and Data Privacy in the Water Sciences" 55. https://doi.org/https://doi.org/10.1029/2019WR025080.

10 Data Acquisition to Distribution

Authors: Giri Prakash (ORNL), Shawn Serbin (BNL), Prasanna Balaprakash (Argonne), Baptiste Dafflon (LBL), Scott Giangrande (BNL), Chongai Kuang (BNL), Line Pouchard (BNL), Cory Stuart (ORNL), Pamela Weisenhorn (Argonne)

Acknowledgement: Session leaders & workshop organizers would like to acknowledge the significant contribution to report of Kyle Pressel.

10.1 Introduction

The Earth sciences have entered an era of "big data" with a high diversity of data types, resolutions (space and time), and formats across the geosciences, Earth System Science (ESS), and atmospheric communities (Dietze, Lebauer, and Kooper 2013). This large volume and diversity of data stem from a range of efforts that cross multiple agencies (e.g., NASA, NOAA, National Science Foundation [NSF], and DOE), measurement platforms (e.g., in-situ observations, automated sensors, measurement networks, remote sensing), application types (e.g., water cycle, atmosphere, land-ocean-cloud interaction), and collection objectives (e.g., monitoring, scaling, reporting and verification, model development and testing).

Typically, the collection and distribution of observational data have remained largely siloed around a domain-specific thematic area (e.g., atmospheric, land-surface, ocean). Legacy data systems, insufficient metadata standards, and ontologies—as well as a clear lack of communication and cross-agency collaboration on the development of accessible, curated, harmonized, and distribution standards across research domains—have created new and significant challenges for efforts that require large, diverse, well-curated datasets (e.g., Serbin et al. 2021), including AI/ML methods (e.g., Devarakonda et al. 2021; Chantry et al. 2021; Varadharajan et al. 2021). To foster the increased use of AI and ML methods in the Earth sciences, a new comprehensive and fully interoperable cross-domain data infrastructure is needed. As described in Figure 10-1, this includes the new development or extension of current metadata and ontology standards, new methods for optimizing the acquisition of data, automated data characterization and quality assessment, and new tools to enable cross-domain data discovery and distribution. At the same time, emerging automated and distributed data streams (e.g., 5G, Internet of Things [IoT], e.g., Beckman et al. 2020; Kollias et al. 2021; Varadharajan et al. 2021), novel collection platforms (e.g., unoccupied aerial systems [UASs]; Yang et al. 2021), together with edge computing capability and AI/ML-guided observations (Balaprakash et al. 2021) will require modern data and file standards that can be flexible enough to capture key spatial information (e.g., projection, resolution), allow for real-time updating of information (e.g., streaming data storage and distribution), and account for any data reduction (e.g., extracting a specific signal from a larger data stream) while maintaining provenance to original data sources. File formats also need to allow for capturing and storing QA/QC and uncertainty information from collection to distribution, including as many sources of uncertainty as possible. Some existing file format and metadata standards are already designed for modernized data systems to support AI/ML efforts (e.g., netCDF and CF conventions), but a much wider adoption of flexible file formats with robust ontologies is needed. Given these urgent needs, a critical requirement for building out comprehensive AI/ML capabilities will be a data infrastructure that provides the necessary tools to enable this acquisition to distribution of analysis-ready data (Figure 10-1). With the advent of modern AI/ML methods (e.g., Huntingford et al. 2019) together with new advancements in the curation and documenting of datasets that allow for federated data discovery (e.g., Wang et al. 2021; L. Pouchard et al. 2021; Wilkinson et al. 2016), new opportunities exist for increasing the utility and accessibility of critical observations across space, time, and agencies.

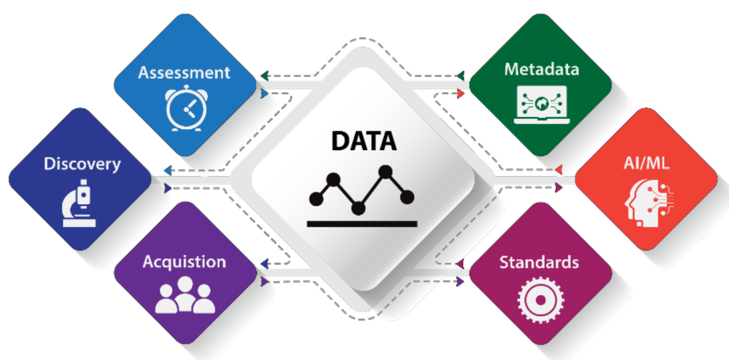


Figure 10-1. Key Components for Enabling Cross-domain Interoperability (Source: Oak Ridge National Laboratory).

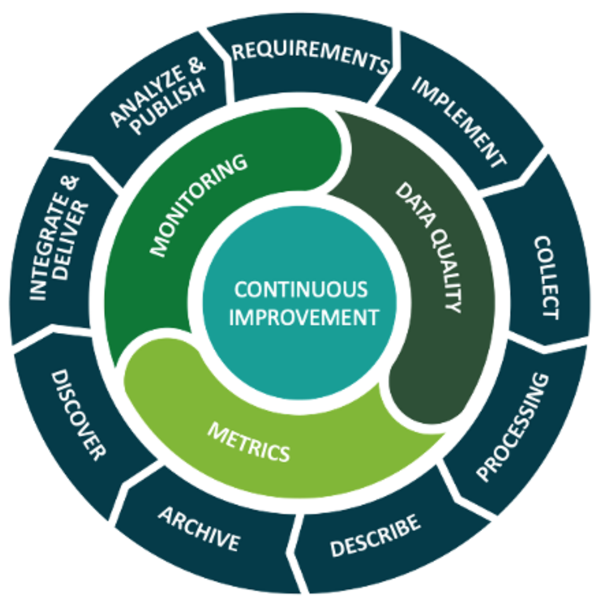


Figure 10-2. End-to-end Data Lifecycle Components (Source: Oak Ridge National Laboratory).

Developing a comprehensive AI/ML-enabled ModEx approach together with a data and modeling cyberinfrastructure is critical to efficiently integrating observations, linking models, and assimilating new data with climate simulations and disciplines to transform traditional measurement campaigns or synthesize multiple multiscaled efforts (e.g., Fer et al. 2021; Lu et al. 2021; Mueller et al. 2021; Serbin et al. 2021). Observational data networks use complex data lifecycle pipelines to manage the end-to-end data (Figure 10-2). These networks are expected to utilize increasingly more and diverse AI/ML methods to improve their data components. For example, applying AI to the source instrument will enable real-time data quality assessment, targeted data collection, and data dimensionality reductions. Similarly, there are opportunities to improve data processing, metadata preparation, and data recommendations based on user research needs. In addition, there are opportunities to provide new data services for AI/ML use cases. These can include, for example, edge computing to enable data analysis at the source, adaptive data collection, and new ontology capabilities to improve cross-domain data discovery (e.g., Pouchard et al. 2013; Beckman et al. 2020; Balaprakash et al. 2021).

10.2 Grand Challenges

In our session, a number of key challenges were identified that have slowed the collection and dissemination of the Earth and atmospheric science data required for improving model fidelity (Table 10-1). For example, the vast majority of surface measurement and observation systems that are deployed are biased because they are located in less remote, or low heterogeneity regions (e.g., Schimel et al. 2015); more distributed observations are needed in model and scaling efforts to represent much larger domains. This creates sampling biases or issues related to the overall "representativeness" of the observations, an additional data uncertainty that may not be easily quantified and integrated into data uncertainty characterization or modeling studies. Measurement networks are also often deployed in a more ad-hoc fashion, which typically creates a suboptimal sampling design to capture the processes of interest (e.g., Kaminski and Rayner 2017). This biased sampling can lead to challenges and errors when developing statistical or mechanistic methods meant to upscale these measurements and extrapolate over space and time. These issues then translate through data systems but are not adequately captured in metadata, QA/QC or uncertainty accounting. Instead, more representative sampling in remote regions or those locations with high spatial and temporal gradients are needed to reduce these sampling biases.

Beyond data representation issues, additional significant challenges remain (Table 10-1). These include data volume and diversity challenges, requiring development of new methods to manage the volume and mix of data types across spatiotemporal scales. Addressing this challenge will also require new investments in hardware and software to manage the increasing flow of complex and novel datasets. Similarly, improved data documentation and automated data valuation will allow for improved data synthesis efforts and allow for developing better AI/ML training datasets. However, this also requires new efforts for cross-agency metadata and file standardization with consistent ontologies, together with improved lifecycle management, that will enable more efficient, federated data discovery and dissemination, provenance management, and versioning. Table 10-1 summarizes the main challenges within the end-to-end data services thematic areas identified during the workshop.

Table 10-1. There are a number of critical challenges and requirements that need to be addressed to improve the data acquisition-to-distribution pipeline of Earth and atmospheric science data. Here we provide a detailed list of primary grand challenges that need to be addressed to foster large-scale AI/ML methods.

Topic	Primary Grand Challenges
Data Acquisition	Representative adaptive and on-demand sampling in environments with high spatial and temporal gradients, sub-optimal observing systems.
Data Transfer and monitoring	Remotely deployed instruments, data reduction, both network and power requirements to operate and transfer data, cyber security.
Data Processing	Protocols and data standards to enable real-time data processing to feed to AI use cases. Coordination across distributed networks.
Data Access and Distribution	Discovery, near real time data access, availability FAIR assessment for AI-ready data, data volume and diversity.
Protocols	Standards used for sensors, UAV, IoT, unrepresentative data, transfer, adaptive data sampling.
Ontology	Lack of standards and knowledgebase for cross -domain semantic ontology.
Metadata and data standards	Inconsistent metadata requirements create significant challenges for federated search and discovery; inconsistent file and data formatting and standards require translation and harmonization when compiled with data from multiple sources.
Data mining and integration	Integration of datasets from multiple acquisition platforms and/or observational networks with ability to generate and assess data products.
QA/QC	Automated data quality assessment, with instant feedback to data submitter/provider/user.

10.2.1 AI/ML-informed Data Collection and Edge Computing

The AI/ML-informed data collection and edge computing "grand challenges" covered a wide range of data acquisition, assimilation, and emerging capability efforts enabled by machine learning, AI, and advanced methods (Table 10-1). The primary challenges discussed involved experimental/network design and its optimization, ability to perform online and continual learning at the edge, and hardware-related efforts involving AI, for example, edge computing (Balaprakash et al. 2021). This emphasis was attributed to one overarching challenge common to the water cycle and its extremes (i.e., droughts, flooding, severe convective storms)—the current limitations of observational capabilities to constrain model treatments of such events that operate over extended spatiotemporal scales (Cholia, Varadharajan, and Pastorello 2021). Specifically, it was highlighted that current water cycle studies are often focused on episodic events (e.g., Varadharajan et al. 2021); however, detailed field campaigns for ideal process-level insights often rely on instrumentation co-deployed and/or operated in an ad hoc, suboptimal manner and often based on limited experience. These strategies are currently informed by expert guidance, with concerns that these strategies may not reflect objective methods but are concepts perpetuated by anecdotal evidence. Discussions specifically highlighted the demand to target several forms of coordinated water cycle processes or key quantities, for example, designing optimal retrieval strategies for key quantities of interest by using ML efforts bolstered by high-resolution model outputs (e.g., Kollias et al. 2021). Other challenges discussed included concerns that most water cycle observations are less representative in regions of high temporal and spatial gradients—a problem especially for tracking episodic, isolated extremes in precipitation and related fields. Similar discussions extended to the challenges of global representativeness when considering limited sampling in higher-latitude and/or other remote tropical or oceanic environments. Separately, many water cycle process studies require distributed and reliable atmospheric state or related quantity retrievals over relatively finer scales and must be available in a timely fashion for model assimilation, which may also be prohibitive to deploy, operate, or coordinate with existing capabilities (e.g., Cholia, Varadharajan, and Pastorello 2021). Finally, several aforementioned challenges are exacerbated by cross-cutting DAQ themes (Table 10-1), necessitating improved data conditioning, data quality control, accessibility to datasets, and high-resolution modeling outputs for AI/ML training and testing.

10.2.2 DAQ for Developing Training and Test Datasets

An overall challenge in advancing Earth system predictability is to improve the acquisition of high-quality datasets and the inference of data products that help develop, parameterize, or validate Earth system models (ESMs) (e.g., Cholia, Varadharajan, and Pastorello 2021; Devarakonda et al. 2021; Ghate et al. 2021; Serbin et al. 2021; Wu et al. 2021). The development of automated frameworks for assessing, processing, and coupling various datasets is key for overcoming the above-mentioned challenge, as well as for taking full advantage of the quickly increasing resolution, coverage, and diversity of ground-based, airborne, and satellite observations capturing atmospheric, surface, and subsurface processes. Particular challenges identified during our session included automated data quality assessment and instant feedback to the data submitter/provider/user, advanced data analysis for error quantification and data discovery by considering historical dataset and physics-based estimations, and assessment of where and when additional measurements are needed.

Further, discussions highlighted the requirements for computational frameworks that can help to advance the development of data products for model training or testing. Particular challenges include merging datasets with various coverages/resolutions to infer spatially and/or temporally resolved products, and interrogating datasets to estimate data product accuracy (Crystal-Ornelas et al. 2021). To that end, FAIR frameworks for data (Wilkinson et al. 2016) to ensure findability, accessibility, interoperability, and reusability are critical. An additional challenge is to leverage both field data and physically based models to improve the analysis of multidimensional relationships needed for generating reliable data products. Overcoming these challenges is also key to advancing the evaluation and identification of trustworthy datasets, guiding and optimizing data acquisition or selection based on physics-based models and prior knowledge, and developing emulators of complex processes. It is noted that the above challenges are all linked to a range of technical challenges associated with software and hardware.

10.2.3 Data Lifecycle, Discovery and Ontology, Standards and Protocols

Arguably one of the of the biggest challenges and potentially most limiting step to the full-scale rollout of an integrated DOE AI/ML framework for improvement of Earth system predictability is the management, processing, and dissemination of properly prepared, documented, and standardized analysis-ready datasets that have adequate uncertainty information and that adhere to FAIR data principles and broader data interoperability (Figure 10-1). To be most impactful, the data system needed to support transformation research in AI/ML will also need to effectively utilize external datasets from partner agencies (e.g., NASA, NOAA, USGS) such that DOE and non-DOE datasets can be efficiently harmonized and incorporated into statistical and mechanistic modeling workflows.

The primary challenges (Table 10-1) to the development of the datasystem needed to support these AI/ML needs that were identified during our session included managing volume, velocity (rate), and veracity of data; inconsistent standards and ontologies; inadequate data lifecycle methods to capture data and model artifacts; data quality; and full data error propagation during collection to dissemination. A key challenge, but also an important opportunity, for increasing the diversity of datasets for AI/ML efforts is the improvement of cross-agency data coordination. This would foster enhanced multiscale datasets by integrating data streams across different scientific domains. To achieve this, challenges associated with ontologies and standards also need to be addressed to remove issues related to ad-hoc or “re-inventing the wheel” efforts to generalized workflows that work across agencies and data types. In addition, improved data dissemination, including automated AI/ML processing, synthesis, and QA/QC, as well as community cloud storage and compute “near” the data system, are needed to overcome data volume challenges but also minimize the requirements of moving data packages and facilitating the effort of analyzing “in place.” This requires a fundamental shift in the way that data for the new lifecycle paradigm are implemented, as well as how ontology, standards, and protocols are developed for effective operations. In addition, issues related to provenance and versioning of data will need to be addressed. Finally, another challenge and opportunity exists in developing ways for error propagation, much of which could leverage AI/ML methods during the data management pipeline and/or during data use.

10.3 State-of-the-Science and Current Challenges

10.3.1 AI/ML-informed Data Collection and Edge Computing

Optimizing the deployment of observational resources is a critical but presently unrealized priority in Earth science and atmospheric research. This is because the reduction of uncertainties in climate predictions has been hindered by a lack of targeted observations that provide the spatially and temporally representative information and surface-atmosphere coupling required to inform atmospheric motions, capture the impacts of environmental heterogeneity, and track the temporal evolution of key properties and processes (e.g., Reddington et al. 2017). Many large-scale measurement campaigns intending to target these processes are expensive, time-consuming endeavors requiring years of careful planning before an observational facility or site locations are identified. While it is critical to maximize the scientific value of the data, these observational campaigns largely rely on heuristic planning processes grounded in domain scientists' intuition. It is also exceedingly rare for process models to inform the siting and measurement strategies of observational networks during their design phase, although a small number of examples do exist (e.g., Lahoz and Schneider 2014; Metzger et al. 2021). Because of this, it is highly likely that resources are misallocated, and critical insights are not identified during the lifetime of the campaign. This challenge will only be exacerbated with more complex measurement campaigns and model needs that are not easy to identify in the project-planning stage. Achieving the goal of intelligent data collection is currently limited by a number of factors, including: (1) lack of dense observation networks to capture the full spatiotemporal spectrum and heterogeneity in the drivers of extreme events; (2) lack of synergistic use of instruments to maximize the information available from observations; (3) physical limitations of modern sensor technology, and the failure to utilize the full amounts of data available from current sensors; (4) lack of low-cost, miniaturized, and easily deployable instrumentation; (5) slower data pipelines that require significant human intervention in the acquisition to distribution workflow that limits the ability to use AI-guided measurement optimization approaches; and (6) lack of AI-enabled, model-driven experiments that perform targeted data collection based on different terrain and locations. Most campaigns are not optimally designed to inform processes that are critical for improving model predictions across space and time, may not efficiently utilize measurement resources for informing models, and are rarely set up for rapid model-data assimilation to iteratively inform new measurement and model priorities during the lifetime of the campaign. However, these issues will need to be addressed to enable widespread AI/ML methods in the climate sciences.

10.3.2 DAQ for Developing Training and Test Datasets

Developing training and test data products for physics- or statistics-based models is key to improving Earth system processes' predictability and discovery (Chantry et al. 2021). Many data products are provided by large sensor networks (e.g., AmeriFlux, IMPROVE, LTER, NEON) and/or remote sensing observations (e.g., aerosol optical depth, cloud fraction, surface reflectance, land-surface temperature) that aim at capturing specific ecosystem properties with a predefined spatial and temporal resolution. These products are the results of carefully designed and applied strategies for data acquisition, QA/QC, processing, management, and dissemination (Pastorello et al. 2020). At the same time, a significant fraction of available datasets are produced from smaller, short-term, ad hoc, targeted or PI-driven measurement campaigns (e.g., Dafflon et al. 2022). Data from these campaigns are then archived in specific agency data systems and later used in other synthesis or modeling efforts; however, this process of archiving and use may take several years. Despite possible delays in dissemination, the value of these products for improving the understanding of ecosystem and convective processes and aerosol properties; quantifying energy, water, and carbon cycle fluxes and their trends; and developing and validating AI/ML or physically based models have been demonstrated in numerous studies (e.g., Jung et al. 2019; Ojha et al. 2021; Shiklomanov et al. 2021). ML techniques have also demonstrated their value, including for automated data QA/QC and processing, estimating data at locations or temporal periods outside the observation window, and for the evaluation of short- and long-term behaviors (e.g., Mylavarapu, Thomas, and Viswanathan 2019; Okafor, Alghorani, and Delaney 2020; Sanhudo, Rodrigues, and Filho 2021). Yet, the resolution, coverage, and diversity of the existing products; the level of automation to generate/update them; and the potential use of physical-based models to guide their development constitute areas where improvements are particularly needed.

10.3.3 Data Lifecycle, Discovery and Ontology, Standards and Protocols

Data repositories and collection sources use various capabilities and standards for producing FAIR-ready datasets that the data analytics platform can readily utilize. In addition, a combination of DOE leadership computing and commercial cloud-compute and storage capabilities are available to projects for data storage and computation needs. Large observatories and data-intensive projects also use many tools and processes to monitor the data flow, perform data quality analysis, and create value-added products (Prakash et al. 2021).

The data management communities are currently discussing methods to extend the rubric score used by FAIR to evaluate datasets ready for AI analysis (e.g., sessions during 2021 SciDataCon and ESIP 2022 January meeting). In addition, there are many opportunities to improve the data provenance, globally persistent identifiers, and citation standards for cross-agency federated search and discovery (Devarakonda et al. 2021). There is a considerable gap in common standards for managing and generating data from the IoT, sensors, and smart devices, which can produce highly diverse data and metadata records (e.g., Dafflon et al. 2022). Additionally, there are emerging technologies to gather feedback between collection and distribution facilitated by AI/ML, targeted model-data infrastructures, or both, enabling uncertainty quantification using complex models and diverse measurements (e.g., Dietze et al. 2014) and self-supervised and semi-supervised learning methodologies. However, new difficulties are created by the fact that AI models and data or data integration methods are tightly integrated; the AI/ML models drive the data collection, which in turn changes the models used for data collection. Currently, there is no data standard or protocol that takes into account the iterative nature of data-model dependency.

10.4 Experimental, Data, and Modeling Opportunities

Developing a modern data system (Figures 10-1 and 10-2) capable of supporting complex, multiscale or "big data" AI/ML approaches would transform our capacity to collect, process, and use Earth and atmospheric measurement data to improve our capacity to predict changes to the Earth system. We identified a number of future opportunities (Table 10-2) within three main thematic areas where AI/ML can be used to enhance current measurement capabilities but also provide or improve novel, new capabilities (e.g., edge computing). In Table 10-2, we describe these opportunities in more detail.

A major opportunity is to develop an integrated framework (including cyber-infrastructure, AI/ML tools, physically based models) for advanced model data experiments (sometimes referred as ModEx). The use of advanced data analytics (including AI/ML) can address a wide range of challenges associated with quantifying data error during their entire lifecycle, merging datasets from various platforms and resolutions into advanced data products, using data products for science discovery and parameterization or validation of physic-based models, and improving data acquisition (property, sensor, resolution, etc.) based on model data experimentation. Importantly, combining AI/ML with physics-based models can revolutionize the model-data experiment through identifying where datasets and process representation are trustworthy and developing emulators of complex processes (e.g., Fer et al. 2018) for their inclusion in complex ESMs (e.g., E3SM) to reduce the computational burdens associated with parameterizing and calibrating ESMs.

Table 10-2. Core focal research priority areas and the associated main short-term and long-term goals to address

	Focal area	<5-year Goal	>5-year Goals	10+-year Goals

	Data Collection	Explore optimal experimental designs and strategies using upcoming field campaigns	Support increased coordination between data collection and analysis methods	AI for ModEx: from data collection to AI and back again AI-guided data collection strategies based on real-time observation needs

Table 10-2. (Cont.)

	Focal area	<5-year Goal	>5-year Goals	10+-year Goals
	Data Processing	Support the development of new protocols to enable real-time data processing Foster intra- and inter-agency coordination on data collection, discovery, and distribution	AI-guided symbiotic framework for integrating models and data to enhance training Federated data search, discovery and distribution w/ dataset uncertainties Cross-agency collaborations on development of data standards, protocols and ontologies	Instantaneous data QC and feedback to data submitters Quantify and control the propagation of error in training datasets to AI/ML and AI/ML in physical-based models AI/ML to identify trustworthy data Cross-agency AI/ML-guided data reduction and synthesis capabilities across domains and spatiotemporal scales Integrate AI/ML into lifecycle and QA/QC to identify gaps more quickly
	Data Distribution		Natural language processing (NLP) technologies for ontology New standards and protocols for beyond 5G network and edge computing	
	Data QA/QC	Create working groups with domain, field scientists and computational scientists focused on defining needs for AI/ML data QA/QC and UQ		AI/ML driven QA/QC
	Data Standards		New standards and protocols for beyond 5G network and edge computing Standards for IOT, sensors, unrepresentative	Provenance, globally persistent identifiers and citation standards for cross-agency federated search and discovery

		data/metadata Provenance, globally persistent identifiers and citation standards for cross-agency federated search and discovery	

Table 10-2. (Cont.)

Focal area	<5-year Goal	>5-year Goals	10+-year Goals
Computational Needs	Computing platforms for “big data” hosting together with computing resources	Coordination with ASCR to build out necessary dedicated computing continuum cyber-infrastructure, better coordinate of DOE flagship compute	Accessible and affordable community cloud compute and storage Compute and data co-existing on GPU/central processing unit (CPU)/HPCs enabling rapid exploration of new methods
AI/ML Testbeds	Leverage existing campaign data to test AI/ML scalability	Leverage upcoming campaigns (e.g., AMF3 SE US) explore optimal experimental designs and strategies	Experiment with novel measurement and processing technologies Edge Computing: Beyond 5G networks, neuromorphic computing, and quantum sensors
AI-guided data acquisition framework	Explore optimal sampling based on existing datasets	Leverage AI/ML to develop adaptive/agile data collection Development of emulators of complex processes to help define obs. needs for ESMs (e.g., E3SM)	Connect physical and AI approaches with optimal sampling design to optimize return on investment (ROI) on measurements campaigns

DOE campaigns collect a wide range of datasets, including high-frequency land-surface, cloud, and aerosol measurements, often resulting in significant data volumes (e.g., Mather and Voyles 2013; Miller et al. 2016; Uin et al. 2019; Kollias et al. 2021; Hardin et al. 2021). End-to-end ModEx simulations, leveraging AI tools including ML model surrogates, and reinforcement-learning can be used to develop optimal sampling for upcoming campaigns, including AMF3 SEUS, and to address key challenges associated with operations design (Hardin et al. 2021; Kollias et al. 2021; Serbin et al. 2021). Similarly, these methods can iteratively adapt sampling strategies or define optimal measurement frequencies over time (seasonally or annually) by proposing new collection or enhanced deployment strategies that target specific improvements in model performance, together with hybrid or emulator model UQ (e.g., Cholia, Varadharajan, and Pastorello 2021; Hardin et al. 2021; Lu et al. 2021; Mueller et al. 2021). Similarly, AI/ML edge-computing for rapid data QA/QC, pattern recognition, and/or AI-assisted dimensionality reduction could help inform adaptive atmospheric measurement strategies, reduce data volume by rapidly targeting data collection and/or separating observations from background conditions (e.g., Balaprakash et al. 2021; Dafflon et al. 2022; Li et al. 2021). This illustrates an added pathway wherein model needs are used to iteratively inform measurement requirements (e.g., Figure 10-3). DOE campaigns like SEUS are also ideal testbeds to advance AI-assisted 5G “smart sensor” networks targeting key variables (e.g., micromet, PM2.5/10) and distributed methods for integrating data across multiple facilities.

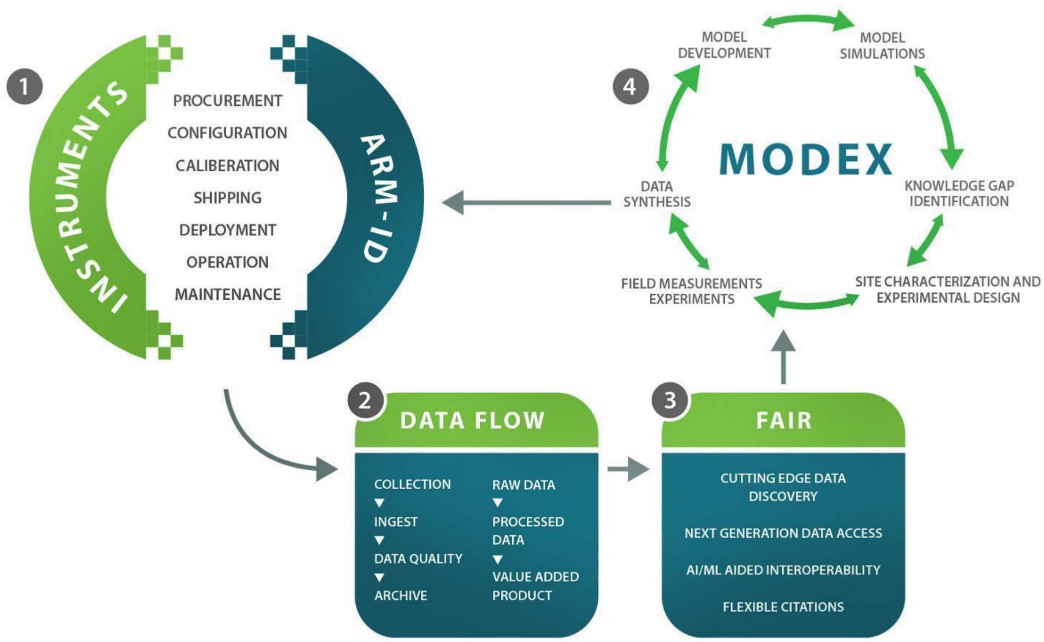


Figure 10-3. Observational and data sources connecting with models and experiments using FAIR and direct feedback mechanisms (Source: Reproduced from Prakash et al. 2021).

Sensor instruments with edge AI/ML capabilities deployed in the field can help validate climate models emulated and simulated in DOE supercomputing facilities (e.g., Dafflon et al. 2022). Quantum sensors that leverage atoms and photons as measurement probes offer high-resolution measurements that were not previously feasible. Utilizing AI/ML models running at the edge can provide high-level information, for example, detection of objects of interest and forecasting short-term events (Balaprakash et al. 2021). This enables any appropriate automated changes in the sensing strategy in near-real time at the edge. For example, local weather changes (such as cloud cover) can be sensed by local edge devices and be used to assimilate the model for accurate short-term forecasts.

Looking beyond 5G, future mobile networks can be a potential game changer for AI/ML-enabled data collection. The 6G and beyond networks will enable a higher degree of programmability

(Saad, Bennis, and Chen 2020), which can deliver a wide range of new use cases, including (1) enhanced mobile broadband (eMBB), to support high-bandwidth, data-driven use cases such as ecohydrology and tropospheric water vapor and temperature data collection; (2) ultra-reliable low-latency communications (URLLC) to support mission-critical communications, for example, remote control of autonomous aerial vehicles such as UASs for targeted data collection; and (3) massive machine-type communications, to support dense deployments of sensor devices that enable the capture of the full spatiotemporal spectrum and heterogeneity in the drivers of extreme events. These new services will leverage the beyond-5G network's revolutionary design of its software-defined core and transport networks and the radio access network that can support advanced wireless communication. With its 100-×100-km footprint, highly heterogeneous land surface, and world-class instrumentation, ARM's Southern Great Plains (SGP) site is ripe for the implementation of a 5G-enabled observational network.

A comprehensive AI-guided, end-to-end ModEx framework requires the ability to synthesize and integrate observations to inform our predictive understanding of ecosystems and climate, address uncertainties and advise observation requirements, or test competing hypotheses (e.g., Fer et al. 2021; Lu et al. 2021; Serbin et al. 2021). Computationally efficient AI/ML tools and processes are needed to manage information flows and provide tractable approaches for model UQ and assimilation to inform predictions, feeding back to data needs (e.g., Cholia, Varadharajan, and Pastorello 2021; Mueller et al. 2021). A key future opportunity to foster much tighter relationships between model simulations and forecasting and data collection is the development of a FAIR model-data integration cyberinfrastructure (Prakash et al. 2021). Presently, there are various examples representing different parts of this much-needed ModEx community infrastructure (e.g., Fer et al. 2021); however, scaling up these efforts based on the current individual PI or smaller team paradigm is not feasible to support the larger institutional needs associated with Earth system predictability through data-informed modeling. Instead, there needs to be a more coordinated effort to develop a flexible cyberinfrastructure (Figure 10-1) that can facilitate formal, modular model-data integration that tracks data and simulation provenance, as well as changes associated with model development efforts (e.g., model versioning tied to specific simulations/ensembles); rapidly integrate datasets from new measurement campaigns; account for data and model uncertainties; provide more rapid feedbacks to data collection based on model needs; and identify and link the most impactful datasets to inform models at the appropriate scales (e.g., Lu et al. 2021; Prakash et al. 2021; Serbin et al. 2021). This infrastructure would leverage AI/ML as both part of the data-to-model pipeline, but also as a means for developing statistical predictors of different phenomenon or model emulators (e.g., Hardin et al. 2021; Mueller et al. 2021; Xu et al. 2021).

Earth and atmospheric observations data are well-suited for the application of a modern AI/ML cyberinfrastructure and data analytics tools to efficiently curate, harmonize, and synthesize diverse, often noisy or of varying quality datasets into analysis-ready, scale-aware observations suitable for multiscale modeling. Working with data communities, developing an extension to FAIR principles (Wilkinson et al. 2016) to evaluate AI-readiness, and upgrading data repositories could provide easy access to datasets sourced from multiple archives to serve a wide-range of AI scientific applications (e.g., Prakash et al. 2021). Given the wealth of data, accessible, harmonized, trustworthy, and QA/QC datasets (with quantitative uncertainties) are essential for future broad-scale applications of AI/ML and ModEx. Development of AI/ML and ModEx testbeds around specific measurement and modeling campaigns (e.g., AMF3 S.E. U.S.) could be used to evaluate modern data processing and analytics tools, including globally persistent unique identifiers to facilitate provenance tracking across diverse data archives, federated data searches, and to maintain standards compliance. In this modern data acquisition to distribution framework, data curation informed by AI can be used to identify and/or gap-fill missing or erroneous data through clustering across variables and scales. AI's ability to process large data volumes enables the discovery of functional relationships between variables (e.g., temperature and evapotranspiration), which are useful for predictive modeling and benchmarking (Fer et al. 2021). Integration of ontology into metadata will support discoverability in persistent data pipelines and workflows and will ensure interoperability across data and analysis (e.g., Python, R) platforms. Supporting end-to-end ModEx, an AI data testbed would facilitate hyperparameter sweeps together with the tracking of simulations through metadata acquisition to compare ML/process model results and automate model retraining or updated simulations, given new data.

Advanced real-time monitoring and access to distributed and remote sensors will be key for building modular edge computing for real-time data quality analysis and data processing at the source. The real-time analysis would enable detection and flagging of problematic data by cross-analyzing datasets before their use in a dynamic selection process (Prakash et al. 2021). Technologies need to be developed to enable the next-generation monitoring and access to distributed sensors that lack power and network. Newer and modular network solutions such as satellite-based, high-throughput networks need to be deployed, along with onsite computational capabilities to perform the data processing at the source.

10.5 Research Priorities: Short-term (<5 years), 5-year, and Longer-term (10-year+) Goals

In recent years, there has been a substantial increase in the amount and diversity of data available for improving the predictability of the Earth system, and as such, a commensurate level of effort in new methods to efficiently collect, process, quality control, synthesize, and distribute these data is required to fully realize the potential for increased ESM fidelity. During our session, a number of focal areas were identified as critical for shorter-term (<5 years) and long-term (5+ years) investment to both increase the utility and ROI of data collection and measurements campaigns, as well as to dynamically target those variables and observations that are most needed for reducing the uncertainty in ESM projections (Table 10-2). These priority areas fall within the main areas of data collection, processing, and distribution; computational needs; AI/ML testbeds; AI/ML synthesis and data reduction approaches across scales; and the development of an AI-guided data acquisition framework.

10.6 References

Balaprakash, Prasanna, Scott Collis, Yongho Kim, Pete Beckmann, Maria Cadeddu, Miquel Gonzalez-Meler, Ryan Sullivan, Sandeep Madireddy, and Rao Kotamarthi. 2021. "AI-Enabled MODEX and Edge-Computing over 5G for Improving the Predictability of Water Cycle Extremes." United States. https://doi.org/10.2172/1769672.

Beckman, Pete, Charlie Catlett, Moinuddin Ahmed, Mohammed Alawad, Linquan Bai, Prasanna Balaprakash, Kevin Barker, et al. 2020. "5G Enabled Energy Innovation: Advanced Wireless Networks for Science (Workshop Report)." United States. https://doi.org/10.2172/1606538.

Chantry, Matthew, Hannah Christensen, Peter Dueben, and Tim Palmer. 2021. "Opportunities and Challenges for Machine Learning in Weather and Climate Modelling: Hard, Medium and Soft AI." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379 (2194): 20200083. https://doi.org/10.1098/rsta.2020.0083.

Cholia, Shreyas, Charuleka Varadharajan, and Gilberto Pastorello. 2021. "Integrating Models with Real-Time Field Data for Extreme Events: From Field Sensors to Models and Back with AI in the Loop." United States. https://doi.org/10.2172/1769727.

Crystal-Ornelas, Robert, Charuleka Varadharajan, Danielle Christianson, Joan Damerow, Helen Weierbach, Emily Robles, Lavanya Ramakrishnan, Boris Faybishenko, and Gilberto Pastorello. 2021. "A Library of AI-Assisted FAIR Water Cycle and Related Disturbance Datasets to Enable Model Training, Parameterization and Validation." United States. https://doi.org/10.2172/1769646.

Dafflon, Baptiste, Stijn Wielandt, John Lamb, Patrick McClure, Ian Shirley, Sebastian Uhlemann, Chen Wang, et al. 2022. "A Distributed Temperature Profiling System for Vertically and Laterally Dense Acquisition of Soil and Snow Temperature." *The Cryosphere* 16 (2): 719–36. https://doi.org/10.5194/tc-16-719-2022.

Devarakonda, Ranjeet, Jitendra Kumar, Dalton Lunga, Jong Choi, and Giri Prakash. 2021. "AI-Driven Data Discovery to Improve Earth System Predictability." United States. https://doi.org/10.2172/1769671.

Dietze, Michael C., David S. Lebauer, and Rob Kooper. 2013. "On Improving the Communication between Models and Data." *Plant Cell and Environment* 36: 1575–85. https://doi.org/10.1111/pce.12043.

Dietze, Michael C., Shawn P. Serbin, Carl Davidson, Ankur R. Desai, Xiaohui Feng, Ryan Kelly, Rob Kooper, et al. 2014. "A Quantitative Assessment of a Terrestrial Biosphere

Model's Data Needs across North American Biomes." *Journal of Geophysical Research: Biogeosciences* 119 (3): 286–300. https://doi.org/https://doi.org/10.1002/2013JG002392.

Fer, Istem, Anthony K. Gardella, Alexey N. Shiklomanov, Eleanor E. Campbell, Elizabeth M. Cowdery, Martin G. de Kauwe, Ankur Desai, et al. 2021. "Beyond Ecosystem Modeling: A Roadmap to Community Cyberinfrastructure for Ecological Data-Model Integration." *Global Change Biology* 27 (1): 13–26. https://doi.org/https://doi.org/10.1111/gcb.15409.

Fer, Istem, Ryan Kelly, Paul R. Moorcroft, Andrew D. Richardson, Elizabeth M. Cowdery, and Michael C. Dietze. 2018. "Linking Big Models to Big Data: Efficient Ecosystem Model Calibration through Bayesian Model Emulation." *Biogeosciences* 15 (19): 5801–30. https://doi.org/10.5194/bg-15-5801-2018.

Ghate, Virendra P., Maria P. Cadeddu, Amanda Lenzi, and Zhengchun Liu. 2021. "Geophysical Retrievals in an Artificial Intelligence (AI) Framework for Illuminating Processes Controlling Water Cycle." United States. https://doi.org/10.2172/1769714.

Hardin, Joseph C., Jr., William I. Gustafson, Adam Varble, Jim Marquis, Jennifer Comstock, Andrew Geiss, Sam Silva, et al. 2021. "The Usage of Observing System Simulation Experiments and Reinforcement Learning to Optimize Experimental Design and Operation," April. https://doi.org/10.2172/1769782.

Huntingford, Chris, Elizabeth S. Jeffers, Michael B. Bonsall, Hannah M. Christensen, Thomas Lees, and Hui Yang. 2019. "Machine Learning and Artificial Intelligence to Aid Climate Change Research and Preparedness." *Environmental Research Letters* 14 (12): 124007. https://doi.org/10.1088/1748-9326/ab4e55.

Jung, Martin, Sujan Koirala, Ulrich Weber, Kazuhito Ichii, Fabian Gans, Gustau Camps-Valls, Dario Papale, Christopher Schwalm, Gianluca Tramontana, and Markus Reichstein. 2019. "The FLUXCOM Ensemble of Global Land-Atmosphere Energy Fluxes." *Scientific Data* 6 (1): 74. https://doi.org/10.1038/s41597-019-0076-8.

Kaminski, Thomas, and Peter Julian Rayner. 2017. "Reviews and Syntheses: Guiding the Evolution of the Observing System for the Carbon Cycle through Quantitative Network Design." *Biogeosciences* 14 (20): 4755–66. https://doi.org/10.5194/bg-14-4755-2017.

Kollias, Pavlos, Edward Luke, Katia Lamer, Fan Yang, Mariko Oue, Nathan Urban, and Andrew Voelmann. 2021. "Multisensor Agile Adaptive Sampling of Convective Storms Driven by Real-Time Analytics." United States. https://doi.org/10.2172/1769753.

Lahoz, William A., and Philipp Schneider. 2014. "Data Assimilation: Making Sense of Earth Observation." *Frontiers in Environmental Science* 2 (MAY): 16. https://doi.org/10.3389/FENVS.2014.00016/BIBTEX.

Li, Xinya, Chris Vernon, Min Chen, Heng Wang, and Zhangshuan Hou. 2021. "A Self-Evolution Data Fusion Platform for Large-Scale Water Models," April. https://doi.org/10.2172/1769652.

Lu, Dan, Eric Pierce, Shih-Chieh Kao, Guannan Zhang, Juan Restrepo, David Womble, Li Li, and Daniella Rempe. 2021. "An AI-Enabled MODEX Framework for Improving

Predictability of Subsurface Water Storage across Local and Continental Scales." United States. https://doi.org/10.2172/1769675.

Mather, James H., and Jimmy W. Voyles. 2013. "The Arm Climate Research Facility: A Review of Structure and Capabilities." *Bulletin of the American Meteorological Society* 94 (3): 377–92. https://doi.org/10.1175/BAMS-D-11-00218.1.

Metzger, Stefan, David Durden, Sreenath Paleri, Matthias Sühring, Brian J. Butterworth, Christopher Florian, Matthias Mauder, et al. 2021. "Novel Approach to Observing System Simulation Experiments Improves Information Gain of Surface–Atmosphere Field Measurements." *Atmospheric Measurement Techniques* 14 (11): 6929–54. https://doi.org/10.5194/amt-14-6929-2021.

Miller, M. A., K. Nitschke, T. P. Ackerman, W. R. Ferrell, N. Hickmon, and M. Ivey. 2016. "The ARM Mobile Facilities." *Meteorological Monographs* 57 (1): 9.1-9.15. https://doi.org/10.1175/AMSMONOGRAPHS-D-15-0051.1.

Mueller, Juliane, Charuleka Varadharajan, Erica Siirila-Woodburn, and Charles Koven. 2021. "Machine Learning for Adaptive Model Refinement to Bridge Scales." United States. https://doi.org/10.2172/1769741.

Mylavarapu, G., J. P. Thomas, and K. A. Viswanathan. 2019. "An Automated Big Data Accuracy Assessment Tool." In *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA)*, 193–97. https://doi.org/10.1109/ICBDA.2019.8713218.

Ojha, Narendra, Imran Girach, Kiran Sharma, Amit Sharma, Narendra Singh, and Sachin S. Gunthe. 2021. "Exploring the Potential of Machine Learning for Simulations of Urban Ozone Variability." *Scientific Reports* 11 (1): 22513. https://doi.org/10.1038/s41598-021-01824-z.

Okafor, Nwamaka U., Yahia Alghorani, and Declan T. Delaney. 2020. "Improving Data Quality of Low-Cost IoT Sensors in Environmental Monitoring Networks Using Data Fusion and Machine Learning Approach." *ICT Express* 6 (3): 220–28. https://doi.org/10.1016/J.ICTE.2020.06.004.

Pastorello, Gilberto, Carlo Trotta, Eleonora Canfora, Housen Chu, Danielle Christianson, You-Wei Cheah, Cristina Poindexter, et al. 2020. "The FLUXNET2015 Dataset and the ONEFlux Processing Pipeline for Eddy Covariance Data." *Scientific Data* 7 (1): 225. https://doi.org/10.1038/s41597-020-0534-3.

Pouchard, Line C., Marcia L. Branstetter, Robert B. Cook, Ranjeet Devarakonda, Jim Green, Giri Palanisamy, Paul Alexander, and Natalya F. Noy. 2013. "A Linked Science Investigation: Enhancing Climate Change Data Discovery with Semantic Technologies." *Earth Science Informatics* 6 (3): 175–85. https://doi.org/10.1007/s12145-013-0118-2.

Pouchard, Line, Carlos Soto, Marcia Branstetter, and Giri Prakash. 2021. "FAIR Data Infrastructure and Tools for AI-Assisted Streamflow Prediction," April. https://doi.org/10.2172/1769710.

Prakash, Giri, Nicki Hickmon, Adam Theisen, Debjani Singh, Cory Stuart, Ranjeet Devarakonda, and Jitendra Kumar. 2021. "AI-Based Upgrades to Observational Data Centers to Facilitate Data Interoperability," April. https://doi.org/10.2172/1769667.

Reddington, C. L., K. S. Carslaw, P. Stier, N. Schutgens, H. Coe, D. Liu, J. Allan, et al. 2017. "The Global Aerosol Synthesis and Science Project (GASSP): Measurements and Modeling to Reduce Uncertainty." *Bulletin of the American Meteorological Society* 98 (9): 1857–77. https://doi.org/10.1175/BAMS-D-15-00317.1.

Saad, W., M. Bennis, and M. Chen. 2020. "A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems." *IEEE Network* 34 (3): 134–42. https://doi.org/10.1109/MNET.001.1900287.

Sanhudo, Luís, João Rodrigues, and Ênio Vasconcelos Filho. 2021. "Multivariate Time Series Clustering and Forecasting for Building Energy Analysis: Application to Weather Data Quality Control." *Journal of Building Engineering* 35 (March): 101996. https://doi.org/10.1016/J.JOBE.2020.101996.

Schimel, David, Ryan Pavlick, Joshua B. Fisher, Gregory P. Asner, Sassan Saatchi, Philip Townsend, Charles Miller, Christian Frankenberg, Kathy Hibbard, and Peter Cox. 2015. "Observing Terrestrial Ecosystems and the Carbon Cycle from Space." *Global Change Biology* 21 (5): 1762–76. https://doi.org/https://doi.org/10.1111/gcb.12822.

Serbin, Shawn P., Scott E. Giangrande, Chongai Kuang, Nathan Urban, and Line Pouchard. 2021. "AI to Automate ModEx for Optimal Predictive Improvement and Scientific Discovery." United States. https://doi.org/10.2172/1769662.

Shiklomanov, Alexey N., Michael C. Dietze, Istem Fer, Toni Viskari, and Shawn P. Serbin. 2021. "Cutting out the Middleman: Calibrating and Validating a Dynamic Vegetation Model (ED2-PROSPECT5) Using Remotely Sensed Surface Reflectance." *Geoscientific Model Development* 14 (5): 2603–33. https://doi.org/10.5194/gmd-14-2603-2021.

Uin, Janek, Allison C. Aiken, Manvendra K. Dubey, Chongai Kuang, Mikhail Pekour, Cynthia Salwen, Arthur J. Sedlacek, et al. 2019. "Atmospheric Radiation Measurement (ARM) Aerosol Observing Systems (AOS) for Surface-Based In Situ Atmospheric Aerosol and Trace Gas Measurements." *Journal of Atmospheric and Oceanic Technology* 36 (12): 2429–47. https://doi.org/10.1175/JTECH-D-19-0077.1.

Varadharajan, Charuleka, Yuxin Wu, Andrew Wiedlea, Kolby Jardine, Haruko Wainwright, Robert Crystal-Ornelas, Joan Damerow, et al. 2021. "Observational Capabilities to Capture Water Cycle Event Dynamics and Impacts in the Age of AI." United States. https://doi.org/10.2172/1769755.

Wang, Chengshan, Robert M Hazen, Qiuming Cheng, Michael H Stephenson, Chenghu Zhou, Peter Fox, Shu-zhong Shen, et al. 2021. "The Deep-Time Digital Earth Program: Data-Driven Discovery in Geosciences." *National Science Review* 8 (9): nwab027. https://doi.org/10.1093/nsr/nwab027.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for

Scientific Data Management and Stewardship." *Scientific Data* 3 (1): 160018. https://doi.org/10.1038/sdata.2016.18.
Wu, John, Deb Agarwall, Boris Faybishenko, Marta Gonzalez, Junmin Gu, Tianzhen Hong, Ling Jin, Alina Lazar, Alex Sim, and C. Anna Spurlock. 2021. "Quality Data Essential for Modeling Water Cycles Effectively." United States. https://doi.org/10.2172/1769769.
Xu, Haowen, Melissa R. Allen-Dumas, Anne Berres, Kuldeep R. Kurte, Yan Liu, Guannan Zhang, and Jibonananda Sanyal. 2021. "A HPC Theory-Guided Machine Learning Cyberinfrastructure for Communicating Hydrometeorological Data Across Scales," March. https://doi.org/10.2172/1769644.
Yang, Dedi, Bailey D. Morrison, Wouter Hantson, Amy L. Breen, Andrew McMahon, Qianyu Li, Verity G. Salmon, Daniel J. Hayes, and Shawn P. Serbin. 2021. "Landscape-Scale Characterization of Arctic Tundra Vegetation Composition, Structure, and Function with a Multi-Sensor Unoccupied Aerial System." *Environmental Research Letters* 16 (8): 085005. https://doi.org/10.1088/1748-9326/ac1291.

11 Neural Networks

Authors: Auroop R. Ganguly (Northeastern Univ.), Richard Archibald (ORNL), Craig Bakker (PNNL), Kate Duffy (NASA-ERI), Romit Maulik (ANL), Juliane Mueller (LBNL), Khachik Sargsyan (SNL), Puja Das (Northeastern Univ.), Jack Watson (Northeastern Univ.)

Acknowledgement: Session leaders & workshop organizers would like to acknowledge the significant contribution to report of Nathan Hodas.

11.1 Grand Challenges

Given the recent advancements and success of neural information processing (especially deep learning), three major requirements of advances in NNs from an Earth system predictability (ESP) perspective were identified. These NN requirements are true for AI4SP (Artificial Intelligence for ESP) in general and for NN4ESP in particular, and they include:

(1) The development of credible and explainable NNs for enabling novel scientific discovery and advances in predictive understanding through causal discoveries (rather than correlations only).

(2) Development of end-to-end analysis tools that seamlessly integrate NNs and ESP simulations to enable actionable predictive insights for stakeholders and decision and policymakers for relevant variables at scale.

(3) Integration of physics and domain understanding in NNs, and combining them with data, analysis tools, and Earth system models (ESMs) process models to address (1) and (2) above by enabling uncertainty quantification and attribution, by developing theoretical guarantees of model predictive performance, and finally by enabling NN learning in the face of the availability of big and small data.

The development of NNs and analysis tools with the above requirements has the potential to advance modeling and understanding of ESP processes in major ways, including:

(a) Understanding the conditions for weather and hydrological extremes such as heat waves, heavy precipitation events, and the knock-on effects of these extremes is critical. These extremes exhibit non-Gaussian (i.e., multimodal or heavy-tailed) data distributions and are currently not adequately handled by current-generation ESMs, leading to their inability to accurately predict extreme events. State-of-the-art generative models have not yet been applied to this task.

(b) With a changing climate and an increase in the frequency of extreme events, there is an urgent need to enable the prediction of when regional climate-related tipping points will be reached and exceeded. Tipping points are a function of trends and patterns of changes at regional scales, and natural variability (deterministic such as sensitivity to initial conditions or even chaos, or stochastic such as low-frequency variability or even *1/f* noise). Extreme events can have short- and long-term impacts on a huge variety of sectors, including public health, emergency management, infrastructure and ecosystem resilience, water and energy security, agriculture and food security, biodiversity and conservation, and similar.

(c) There is a need to fill gaps in our understanding of physics processes that are currently included in ESMs in the form of parametrizations, thus limiting the reliability of the predictions made even with the most advanced simulation models. Explainable and interpretable NNs have the potential to help us advance our physics understanding and thus mechanistic modeling. The development of hybrid models (e.g., using NNs to replace parametrizations) improves predictive accuracy and provides uncertainty estimates, as well as other data-driven models for predictive understanding of missing physics, and is a key grand challenge (see Figure 11-1).

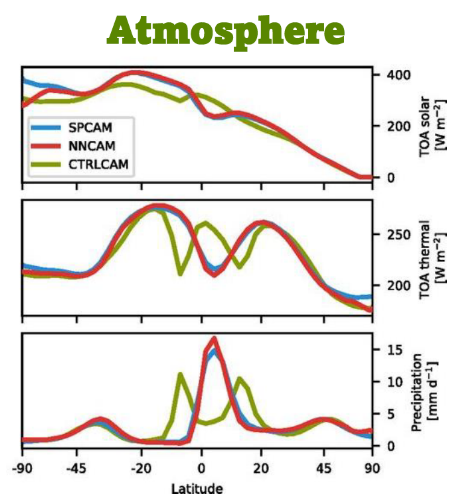


Figure 11-1. Community Atmospheric Model (CAM/NCAR): NN (NNCAM) replaces state-of-the-art CAM (CTRLCAM) and CRM (SPCAM) to demonstrate proof of principle in cloud parameterizations (Source: Figure from Rasp, Pritchard, and Gentine 2018 under Creative Commons CC BY-NC-ND 4.0).

(d) Non-stationary processes such as changes expected in the Earth's radiative budgets in the future must be reflected in parameterizations and data-driven functional mappings in order for NNs to make accurate predictions. A difficulty here is the potentially vast spatial and temporal scales that must be bridged when making predictions. Domain knowledge could be exploited here and incorporated in the NNs; for example, it is more likely that the climate of the U.S. Northeast begins to resemble the current climate in southern parts of the U.S. eastern seaboard within the next several decades, rather than the western United States. Similarly, the average climate of a region in the future may become warmer, but it may resemble outlier observations we have already observed occasionally due to natural variability. This kind of constrained nonstationarity should be considered when designing parameterizations and NN (including deep learning) approaches, as well as when designing systematic evaluation strategies by reflecting time- and space-dependent deviations.

(e) Quantifying the uncertainty (model, parameter, data uncertainty) is essential for enabling trustworthy predictions with ESP simulations. The computational expense associated with UQ on simulations quickly becomes computationally intractable. Here, NNs, in particular differentiable models, can serve as emulators of compute-intensive modules in ESP simulations, thus allowing for conducting efficient and effective UQ and parameter estimation.

11.2 State-of-the-Science

In order to fully exploit the abilities of upcoming NN and AI capabilities, advances have been made on various fronts, even though significant challenges remain. With the increasing popularity of NNs, they have been more frequently used in the Earth sciences for prediction (Bonavita et al. 2021), pattern recognition (Camps-Valls et al. 2021), classification (Srivastava, Nemani, and Steinhaeuser 2017), and uncertainty quantification (Vandal et al. 2018). NNs, in particular DL models, have found broad applicability for predicting groundwater levels (Srivastava, Nemani, and Steinhaeuser 2017; Müller et al. 2021), water quality (Solanki, Agrawal, and Khare 2015), etc. In the literature, DL models have been shown to be excellent replacements of parameterizations (Chattopadhyay, Subel, and Hassanzadeh 2020). DL has been used to represent subgrid processes in climate models (Rasp, Pritchard, and Gentine 2018); deep-learned process parameterizations have been shown to provide better representations of turbulent heat fluxes in hydrologic models (Bennett and Nijssen 2021); stochastic deep learning parameterizations have been used in ocean momentum forcing (Guillaumin and Zanna 2021) (see Figure 11-2). DL models have the advantage that, once trained, they can make predictions significantly faster than parametrizations (see, e.g., Brajard et al. 2021). Moreover, when directly trained on observation data, DL models can achieve higher prediction accuracy than parametrizations (Yuval and O'Gorman 2020), and they can enable the development of new parametrizations from long observation records (Yimin Liu, Sun, and Durlofsky 2019). Most applications of NNs in the Earth sciences are, however, still experimental, with practitioners often trying out different types of NN models. In order to select the best NN for a given application, objective model performance metrics for comparing different models are needed. These objectives must align with the science goals, that is, the widely used Euclidean norm may not be suitable for all applications. A recent paper (Belkin et al. 2019) attempted to reconcile current versus classical practices of deep learning with the bias-variance trade-off. Recent research in developing predictive insights with Earth system models, including for ocean-atmosphere-land coupling, have attempted to use explainable deep learning (Yumin Liu et al. 2022). Often, only a handful of models are tried and compared due to the computational cost required for training, and thus suboptimal solutions are often used. Moreover, most DL deployments have not been performed on operational resolutions, for example, in the Weatherbench project (Rasp et al. 2020).

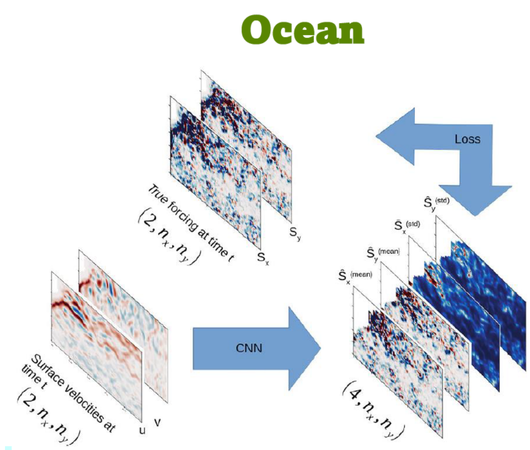


Figure 11-2. Ocean Model (CM2.6/GFDL): Stochastic parameterization with convolutional neural network (CNN) showed that subgrid momentum forcing can be generated with macroscale surface velocities (Source: Reproduced from Guillaumin and Zanna 2021 under Creative Commons CC BT 4.0).

The need for R&D has also been appreciated in terms of data acquisition and data availability. For example, the ARM (Atmospheric Radiation Measurement) facility is just beginning to intensify its activities, including improving their algorithms for filling in missing data, classification, and anomaly detection. NASA synchronizes satellite data to enable downstream emulation, including Bayesian deep learning, GANs, and VAEs. However, most of the data processing does not take place on-satellite, but rather at HPC centers, making an adaptive data collection approach hard.

NNs were employed to downscale climate variables circumventing the need for reanalysis (Yumin Liu, Ganguly, and Dy 2020; Vandal et al. 2018). Another major use case of NNs in climate modeling is the construction of computationally inexpensive surrogates, or emulators, for studies that otherwise require prohibitively many model simulations, such as global sensitivity analysis, model calibration, or uncertainty quantification. A supervised ML is carried out using perturbed parameter ensembles as training sets to build highly parameterized NNs as approximations of input-output relationships in climate models or parts thereof. For example, Scher (2018) used deep learning to emulate the complete physics and dynamics of a simple general circulation model. Another example of NN surrogate construction is in Lu and Ricciuto (2019), where the authors developed NN surrogates for principal components of E3SM land model outputs. Surrogate modeling has also been applied in a classification context in Prabhat et al. (2021). The authors trained an NN to emulate hand-drawn atmospheric river labels. A major challenge as well as an opportunity in NN surrogate modeling is ensuring that physical principles are respected, such as, for example, positivity, conservation laws, and output QoI relations. In this regard, the state-of-the-art scientific ML methods generally belong to two classes: (1) soft constraints that penalize NN training for violation of physical constraints, and (2) hard constraints that enforce the constraints *exactly* by the choice of NN architecture (Karpatne et al. 2017). A popular framework for soft constraints is the physics-informed NNs (PINNs), pioneered by Raissi, Perdikaris, and Karniadakis (2019) where boundary conditions of partial differential equations form the penalty during the NN learning of the state evolution. Hard constraints, on the other hand, are less straightforward to implement and remain problem specific. Beucler et al. (2019) have proposed NN architectural changes to enforce energy conservation and further developed the work to incorporate general analytical constraints (Beucler et al. 2021).

Another promising avenue of NN application to climate is discovering or learning governing equations, such as learning ocean mesoscale closures in Zanna and Bolton (2020). Furthermore, two NN flavors are particularly well suited for ESP application due to the spatiotemporal nature of Earth system model output quantities of interest. These are recurrent NNs (RNNs) that are built to handle temporal predictions and forecasts (Vandal et al. 2017; Shen, Liu, and Wang 2021; Xu et al. 2020; Lees et al. 2021; Yu et al. 2021) and convolutional NNs (CNNs) that efficiently target spatial relationships and patterns (Ise and Oba 2019; Chattopadhyay, Hassanzadeh, and Pasha 2020; Weyn, Durran, and Caruana 2020; Steininger et al. 2020; Baño-Medina, Manzanas, and Gutiérrez 2021). Finally, generative modeling approaches, such as GANs (Besombes et al. 2021; Klemmer et al. 2021; Wang, Tang, and Gentine 2021) and VAEs (Tibau Alberdi et al. 2018; Scher 2018; Mooers et al. 2020; Zadrozny et al. 2021) have demonstrated early promise in the Earth sciences (e.g., Ravuri et al. 2021). The VAEs, in particular, show significant potential via their ability to compactly represent and reproduce the climate state via careful construction of a nonlinear latent space.

11.3 Experimental, Data, and Modeling Opportunities

There are multiple opportunities to improve and accelerate the state of the art of exploiting NNs for advancing ESP. First, in collaboration with agencies such as DOE ARM or NASA, investments must be made for conducting field experiments in data-poor regions, such as the arctic. Local conditions and geography may pose accessibility issues that could potentially be overcome with autonomous data collection technology, such as drones. As shown in Figure 1 from Reichstein et al. (2019) data from a variety of sources must be integrated. If multiple sensors collect the same type of data, then differences in the provided data products (le Coz and van de Giesen 2020) could be an indicator of data uncertainties that can then be exploited in parameter estimation and UQ tasks for ESP simulation models. For data-poor applications, there is also an opportunity to create synthetic data (e.g., by resampling NNs such as GANs and VAEs) or to develop transfer learning approaches that allow us to generate data at an undersampled area by using NN models that were trained on similar data-rich areas.

When the goal is to build and use NNs as emulators, relevant data can be generated using ESMs. In particular, when extreme events are of interest, there could be opportunities to preferentially sample data from the tail ends of the distributions. Data should be collected and generated with the FAIR principles (**F**indability, **A**ccessibility, **I**nteroperability, **R**euse) in mind, which will also enable the development of AI benchmarks for ESP applications. While there are several common benchmarks for ML in computer vision (Venkata et al. 2009; L. Deng 2012; J. Deng et al. 2009), it is evident that scientific ML as well as UQ generally lack such datasets. In particular, ESMs could use established datasets with unified formatting, both from the modeling side and from observational campaigns (Collier et al. 2018), to help hone scientific ML algorithms for performance and predictive skill improvements.

Research opportunities also exist on the modeling side. This includes the improvement of weather and regional climate simulations, the development of model simulation "testbeds" of varying complexity, neural processing of multiscale temporal and spatial processes with UQ, and the enabling of interpretability (see Figure 11-3), NNs for causal inference engine testing, NNs as computationally tractable emulators of expensive simulations, and NNs in connection with active learning strategies to tell us which data to collect in order to reduce model predictive uncertainty and improve our trust in model predictions.

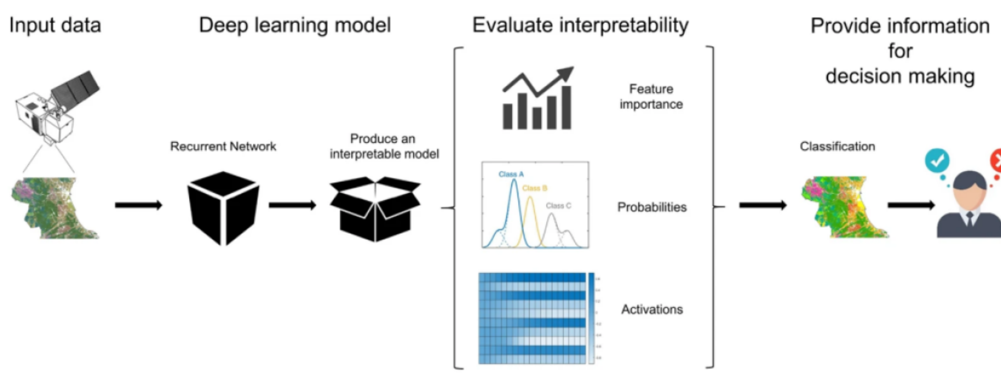


Figure 11-3. DL models for enabling interpretability of data and supporting decision-making (Source: Reproduced from Campos-Taberner et al. 2020 under Creative Commons CC BY 4.0).

11.4 Research Priorities

There are a multitude of complementary angles and approaches that can be taken to address NN4ESP challenges.

The science barriers they must address are the "constrained nonstationarity" issues outlined above, the heterogeneity and complex nonlinearities of the processes involved in Earth system modeling, and the missing understanding of certain ESPs. Advancements in NNs and their usage and in physics modeling must address the issues with extreme values (extreme events) and distributions that are non-Gaussian. The lack of observational data of extreme events makes this difficult, but the advance of learning models that do not make assumptions about underlying Gaussianity, such as normalizing flows, have the potential to play an important role to further our understanding and prediction ability of extremes. A second urgent need and research priority should be in accurately predicting regional changes, tipping points, and knock-on effects, which can span a large variety of spatial and temporal scales. This will become increasingly important for making lasting infrastructure investments in the face of a changing climate.

Research priorities should also focus on the development of theoretical guarantees for NN model performance. This is a necessary step for building trust in their predictions and for motivating their use as part of decision and policymaking frameworks (rather than using NNs as magic wands). Together with performance guarantees, we need uncertainty quantification of the NN predictions. Uncertainty can here arise due to multiple factors, including from the data and the model choice, and ideally different sources of uncertainty can be controlled independently of each other. It is also important to identify the limits of prediction ability of NNs, ESM simulations, parameterizations, and hybrid models, in particular with regard to extreme events that have not yet been observed: when and which model should be used and trusted? Furthermore, developments are needed to endow NN models with expert knowledge and physics information such as conservation laws and constraints on physics process interactions. The currently widely used approach of adding penalties to the NN model's loss function (soft constraints) cannot capture these constraints well, and there is no guarantee that the final trained NN model satisfies the constraints.

Another research priority should lie in the development of tighter collaborations between domain scientists, mathematicians, and computer scientists to enable (semi-)automated workflows for efficiently using and tuning NN models. In several scientific publications, the authors do not report on how their DL model architectures were chosen (Rahmani et al. 2021). Although several tools for hyperparameter optimization of DL models exist (Yang and Shami 2020), they are often not used by practitioners, perhaps because of the learning curve associated with using the tools or not knowing they exist. Further research is needed in the development of tools that allow for the extraction of causal relationships and mechanisms, and that offer robust interpretability and explainability of the NN model outcomes. Sustainable software development practices are needed as much as community data standards for "AI-ready" data that are created with FAIR principles in mind. Model and data repositories are needed to allow the reproducibility of scientific analyses. Currently, if not given exactly the same NN model and data, results cannot be reproduced, leading to questions about whether the reported results are robust and thus support any of the conclusions drawn. Scientific benchmarks of varying complexity must be developed together with objective comparison metrics in order to provide a comparison of NN model and hybrid model performance. Simulation models and learning algorithms must be developed such that they can exploit future HPC facilities. Taken together with the FAIR principles, containers (e.g., docker) are sometimes employed. However, depending on the size of the NN model and the datasets, this approach for achieving reproducible results may not be sustainable either.

Last, another research priority lies in the data challenges. Some applications are data-rich and others are data-poor. NN models are known to require vast amounts of data in order not to suffer from overfitting, but not all applications satisfy this necessity. Guidance is required in order to decide when to use which type of learning model (Gaussian process models, NN models). Transfer learning has shown promise when transitioning from a big data to a small data scenario. Questions arise regarding when a model trained for a data-rich region has no longer any predictive ability when transferred to a data-poor region. There may also be benefits in iteratively augmenting datasets with new measurements. NN models may help us to motivate additional data collection by providing a value for new data, which could be measured in reduced model uncertainty, for example.

11.5 Short-term (<5 years), 5-year, and 10-year Goals

In the short term (less than 5 years), research efforts should focus on the development of data repositories, simulation testbeds, and effective benchmarks. Active learning strategies as often used in optimization and uncertainty quantification should be enhanced and applied to ESP models in data collection and calibration tasks. Progress can and must be made in the adoption of NN architecture optimization and the development of objective comparison metrics. Extreme events and non-Gaussian data distributions must be given more consideration, and developments of UQ methods for NNs that do not rely on Gaussian assumptions are needed. Attention must also be paid to regional changes, which can span large timescales. NNs that are able to ingest these long time series and provide predictions of future regional changes and predict the nearing of tipping points are essential to provide decision makers with means for effective management strategies.

The research focus in the medium term should be on the development of interpretable and robust DL models. Causal inference and transfer learning methods must be improved in order to enable a broader applicability of the models. UQ of NN model predictions are lacking and must be developed. When using NN models to replace compute-intensive parts of mechanistic simulation models, we need the ability to propagate uncertainty through the hybrid-models to better understand the final variability of the predictions. Uncertainty attribution is equally important, and methods must be developed that can shed light on the sources of uncertainty in the model chain, which will thus allow us to quantify and reduce it. For scalability, nonlinear dimension reduction methods are needed.

In the long term, many of the research topics outlined above should be further developed and refined. In addition, theoretical guarantees and generalization bounds on NN/ML for ESP and theoretical guarantees for hybrid physics-NN combinations are needed to improve trust in DL models.

11.6 References

Baño-Medina, Jorge, Rodrigo Manzanas, and José Manuel Gutiérrez. 2021. "On the Suitability of Deep Convolutional Neural Networks for Continental-Wide Downscaling of Climate Change Projections." *Climate Dynamics* 57 (11): 2941–51. https://doi.org/10.1007/s00382-021-05847-0.

Belkin, Mikhail, Daniel Hsu, Siyuan Ma, and Soumik Mandal. 2019. "Reconciling Modern Machine-Learning Practice and the Classical Bias-Variance Trade-Off." *Proceedings of the National Academy of Sciences* 116 (32): 15849–54. https://doi.org/10.1073/pnas.1903070116.

Bennett, Andrew, and Bart Nijssen. 2021. "Deep Learned Process Parameterizations Provide Better Representations of Turbulent Heat Fluxes in Hydrologic Models." *Water Resources Research* 57 (5): e2020WR029328. https://doi.org/https://doi.org/10.1029/2020WR029328.

Besombes, C, O Pannekoucke, C Lapeyre, B Sanderson, and O Thual. 2021. "Producing Realistic Climate Data with Generative Adversarial Networks." *Nonlinear Processes in Geophysics* 28 (3): 347–70. https://doi.org/10.5194/npg-28-347-2021.

Beucler, Tom, Michael Pritchard, Stephan Rasp, Jordan Ott, Pierre Baldi, and Pierre Gentine. 2021. "Enforcing Analytic Constraints in Neural Networks Emulating Physical Systems." *Physical Review Letters* 126 (9): 98302. https://doi.org/10.1103/PhysRevLett.126.098302.

Beucler, Tom, Stephan Rasp, Michael Pritchard, and Pierre Gentine. 2019. "Achieving Conservation of Energy in Neural Network Emulators for Climate Modeling," June. https://doi.org/10.48550/arxiv.1906.06622.

Bonavita, Massimo, Rossella Arcucci, Alberto Carrassi, Peter Dueben, Alan J Geer, Bertrand le Saux, Nicolas Longépé, Pierre-Philippe Mathieu, and Laure Raynaud. 2021. "Machine Learning for Earth System Observation and Prediction." *Bulletin of the American Meteorological Society* 102 (4): E710–16. https://doi.org/10.1175/BAMS-D-20-0307.1.

Brajard, Julien, Alberto Carrassi, Marc Bocquet, and Laurent Bertino. 2021. "Combining Data Assimilation and Machine Learning to Infer Unresolved Scale Parametrization." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379 (2194): 20200086. https://doi.org/10.1098/rsta.2020.0086.

Campos-Taberner, Manuel, Francisco Javier García-Haro, Beatriz Martínez, Emma Izquierdo-Verdiguier, Clement Atzberger, Gustau Camps-Valls, and María Amparo Gilabert. 2020.

"Understanding Deep Learning in Land Use Classification Based on Sentinel-2 Time Series." *Scientific Reports* 10 (1): 17188. https://doi.org/10.1038/s41598-020-74215-5.

Camps-Valls, Gustau, Devis Tuia, Xiao Xiang Zhu, and Markus Reichstein, eds. 2021. *Deep Learning for the Earth Sciences*. Wiley. https://doi.org/10.1002/9781119646181.

Chattopadhyay, Ashesh, Pedram Hassanzadeh, and Saba Pasha. 2020. "Predicting Clustered Weather Patterns: A Test Case for Applications of Convolutional Neural Networks to Spatio-Temporal Climate Data." *Scientific Reports* 10 (1): 1317. https://doi.org/10.1038/s41598-020-57897-9.

Chattopadhyay, Ashesh, Adam Subel, and Pedram Hassanzadeh. 2020. "Data-Driven Super-Parameterization Using Deep Learning: Experimentation With Multiscale Lorenz 96 Systems and Transfer Learning." *Journal of Advances in Modeling Earth Systems* 12 (11): e2020MS002084. https://doi.org/https://doi.org/10.1029/2020MS002084.

Collier, Nathan, Forrest M Hoffman, David M Lawrence, Gretchen Keppel-Aleks, Charles D Koven, William J Riley, Mingquan Mu, and James T Randerson. 2018. "The International Land Model Benchmarking (ILAMB) System: Design, Theory, and Implementation." *Journal of Advances in Modeling Earth Systems* 10 (11): 2731–54. https://doi.org/https://doi.org/10.1029/2018MS001354.

Coz, Camille le, and Nick van de Giesen. 2020. "Comparison of Rainfall Products over Sub-Saharan Africa." *Journal of Hydrometeorology* 21 (4): 553–96. https://doi.org/10.1175/JHM-D-18-0256.1.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. "ImageNet: A Large-Scale Hierarchical Image Database." In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–55. https://doi.org/10.1109/CVPR.2009.5206848.

Deng, Li. 2012. "The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]." *IEEE Signal Processing Magazine* 29 (6): 141–42. https://doi.org/10.1109/MSP.2012.2211477.

Guillaumin, Arthur P, and Laure Zanna. 2021. "Stochastic-Deep Learning Parameterization of Ocean Momentum Forcing." *Journal of Advances in Modeling Earth Systems* 13 (9): e2021MS002534. https://doi.org/https://doi.org/10.1029/2021MS002534.

Ise, Takeshi, and Yurika Oba. 2019. "Forecasting Climatic Trends Using Neural Networks: An Experimental Study Using Global Historical Data." *Frontiers in Robotics and AI* 6. https://www.frontiersin.org/article/10.3389/frobt.2019.00032.

Karpatne, Anuj, Gowtham Atluri, James H Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, and Vipin Kumar. 2017. "Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data." *IEEE Transactions on Knowledge and Data Engineering* 29 (10): 2318–31. https://doi.org/10.1109/TKDE.2017.2720168.

Klemmer, Konstantin, Sudipan Saha, Matthias Kahl, Tianlin Xu, and Xiao Xiang Zhu. 2021. "Generative Modeling of Spatio-Temporal Weather Patterns with Extreme Event Conditioning," April. https://doi.org/10.48550/arxiv.2104.12469.

Lees, T, S Reece, F Kratzert, D Klotz, M Gauch, J de Bruijn, R Kumar Sahu, P Greve, L Slater, and S Dadson. 2021. "Hydrological Concept Formation inside Long Short-Term Memory (LSTM) Networks." *Hydrology and Earth System Sciences Discussions* 2021: 1–37. https://doi.org/10.5194/hess-2021-566.

Liu, Yimin, Wenyue Sun, and Louis Durlofsky. 2019. "A Deep-Learning-Based Geological Parameterization for History Matching Complex Models." *Mathematical Geosciences* 51 (March). https://doi.org/10.1007/s11004-019-09794-9.

Liu, Yumin, Kate Duffy, Jennifer G. Dy, and Auroop R. Ganguly. 2022. "Explainable Deep Learning for Insights in El Nino and River Flows," January. https://doi.org/10.48550/arxiv.2201.02596.

Liu, Yumin, Auroop R Ganguly, and Jennifer Dy. 2020. "Climate Downscaling Using YNet: A Deep Convolutional Network with Skip Connections and Fusion." In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 3145–53. KDD '20. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3394486.3403366.

Lu, D, and D Ricciuto. 2019. "Efficient Surrogate Modeling Methods for Large-Scale Earth System Models on Machine-Learning Techniques." *Geoscientific Model Development* 12 (5): 1791–1807. https://doi.org/10.5194/gmd-12-1791-2019.

Mooers, Griffin, Jens Tuyls, Stephan Mandt, Michael Pritchard, and Tom Beucler. 2020. "Generative Modeling for Atmospheric Convection," July. https://doi.org/10.48550/arxiv.2007.01444.

Müller, Juliane, Jangho Park, Reetik Sahu, Charuleka Varadharajan, Bhavna Arora, Boris Faybishenko, and Deborah Agarwal. 2021. "Surrogate Optimization of Deep Neural Networks for Groundwater Predictions." *Journal of Global Optimization* 81 (1): 203–31. https://doi.org/10.1007/s10898-020-00912-0.

Prabhat, K Kashinath, M Mudigonda, S Kim, L Kapp-Schwoerer, A Graubner, E Karaismailoglu, et al. 2021. "ClimateNet: An Expert-Labeled Open Dataset and Deep Learning Architecture for Enabling High-Precision Analyses of Extreme Weather." *Geoscientific Model Development* 14 (1): 107–24. https://doi.org/10.5194/gmd-14-107-2021.

Rahmani, Farshid, Chaopeng Shen, Samantha Oliver, Kathryn Lawson, and Alison Appling. 2021. "Deep Learning Approaches for Improving Prediction of Daily Stream Temperature in Data-Scarce, Unmonitored, and Dammed Basins." *Hydrological Processes* 35 (11): e14400. https://doi.org/https://doi.org/10.1002/hyp.14400.

Raissi, M, P Perdikaris, and G E Karniadakis. 2019. "Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations." *Journal of Computational Physics* 378: 686–707. https://doi.org/https://doi.org/10.1016/j.jcp.2018.10.045.

Rasp, Stephan, Peter D Dueben, Sebastian Scher, Jonathan A Weyn, Soukayna Mouatadid, and Nils Thuerey. 2020. "WeatherBench: A Benchmark Data Set for Data-Driven Weather Forecasting." *Journal of Advances in Modeling Earth Systems* 12 (11): e2020MS002203. https://doi.org/https://doi.org/10.1029/2020MS002203.

Rasp, Stephan, Michael S. Pritchard, and Pierre Gentine. 2018. "Deep Learning to Represent Subgrid Processes in Climate Models." *Proceedings of the National Academy of Sciences of the United States of America* 115 (39): 9684–89. https://doi.org/10.1073/PNAS.1810286115/SUPPL_FILE/PNAS.1810286115.SAPP.PDF.

Ravuri, Suman, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, et al. 2021. "Skilful Precipitation Nowcasting Using Deep Generative Models of Radar." *Nature* 597 (7878): 672–77. https://doi.org/10.1038/s41586-021-03854-z.

Reichstein, Markus, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and Prabhat. 2019. "Deep Learning and Process Understanding for Data-Driven Earth System Science." *Nature* 566 (7743): 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Scher, S. 2018. "Toward Data-Driven Weather and Climate Forecasting: Approximating a Simple General Circulation Model With Deep Learning." *Geophysical Research Letters* 45 (22): 12, 612–16, 622. https://doi.org/https://doi.org/10.1029/2018GL080704.

Shen, Jiangrong, Jian K Liu, and Yueming Wang. 2021. "Dynamic Spatiotemporal Pattern Recognition With Recurrent Spiking Neural Network." *Neural Computation* 33 (11): 2971–95. https://doi.org/10.1162/neco_a_01432.

Solanki, Archana, Himanshu Agrawal, and Kanchan Khare. 2015. "Predictive Analysis of Water Quality Parameters Using Deep Learning." *International Journal of Computer Applications* 125 (September): 29–34. https://doi.org/10.5120/ijca2015905874.

Srivastava, Ashok N., Ramakrishna Nemani, and Karsten Steinhaeuser, eds. 2017. *Large-Scale Machine Learning in the Earth Sciences*. Boca Raton : Taylor & Francis, 2017. | Series: Chapman & Hall/CRC data mining & knowledge discovery series ; 42 | "A CRC title, part of the Taylor & Francis imprint, a member of the Taylor & Francis Group, the academic division of T&F Informa plc.": Chapman and Hall/CRC. https://doi.org/10.4324/9781315371740.

Steininger, Michael, Daniel Abel, Katrin Ziegler, Anna Krause, Heiko Paeth, and Andreas Hotho. 2020. "Deep Learning for Climate Model Output Statistics," December. https://doi.org/10.48550/arxiv.2012.10394.

Tibau Alberdi, Xavier-Andoni, Christian Requena-Mesa, Christian Reimers, Joachim Denzler, Veronika Eyring, Markus Reichstein, and Jakob Runge. 2018. "SupernoVAE : VAE Based Kernel PCA for Analysis of Spatio-Temporal Earth Data." In: https://www.researchgate.net/publication/330984865_SupernoVAE_VAE_based_kernel_PCA_for_analysis_of_spatio-temporal_Earth_data.

Vandal, Thomas, Evan Kodra, Jennifer Dy, Sangram Ganguly, Ramakrishna Nemani, and Auroop R Ganguly. 2018. "Quantifying Uncertainty in Discrete-Continuous and Skewed Data with Bayesian Deep Learning." In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2377–86. KDD '18. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3219819.3219996.

Vandal, Thomas, Evan Kodra, Sangram Ganguly, Andrew Michaelis, Ramakrishna Nemani, and Auroop R Ganguly. 2017. "DeepSD: Generating High Resolution Climate Change Projections through Single Image Super-Resolution." In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1663–72. KDD '17. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3097983.3098004.

Venkata, Sravanthi Kota, Ikkjin Ahn, Donghwan Jeon, Anshuman Gupta, Christopher Louie, Saturnino Garcia, Serge Belongie, and Michael Bedford Taylor. 2009. "SD-VBS: The San Diego Vision Benchmark Suite." In *2009 IEEE International Symposium on Workload Characterization (IISWC)*, 55–64. https://doi.org/10.1109/IISWC.2009.5306794.

Wang, Cunguang, Guoqiang Tang, and Pierre Gentine. 2021. "PrecipGAN: Merging Microwave and Infrared Data for Satellite Precipitation Estimation Using Generative Adversarial Network." *Geophysical Research Letters* 48 (5): e2020GL092032. https://doi.org/https://doi.org/10.1029/2020GL092032.

Weyn, Jonathan A, Dale R Durran, and Rich Caruana. 2020. "Improving Data-Driven Global Weather Prediction Using Deep Convolutional Neural Networks on a Cubed Sphere." *Journal of Advances in Modeling Earth Systems* 12 (9): e2020MS002109. https://doi.org/https://doi.org/10.1029/2020MS002109.

Xu, Wei, Yanan Jiang, Xiaoli Zhang, Yi Li, Run Zhang, and Guangtao Fu. 2020. "Using Long Short-Term Memory Networks for River Flow Prediction." *Hydrology Research* 51 (6): 1358–76. https://doi.org/10.2166/nh.2020.026.

Yang, Li, and Abdallah Shami. 2020. "On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice." *Neurocomputing* 415: 295–316. https://doi.org/https://doi.org/10.1016/j.neucom.2020.07.061.

Yu, M, F Xu, W Hu, J Sun, and G Cervone. 2021. "Using Long Short-Term Memory (LSTM) and Internet of Things (IoT) for Localized Surface Temperature Forecasting in an Urban Environment." *IEEE Access* 9: 137406–18. https://doi.org/10.1109/ACCESS.2021.3116809.

Yuval, Janni, and Paul A O'Gorman. 2020. "Stable Machine-Learning Parameterization of Subgrid Processes for Climate Modeling at a Range of Resolutions." *Nature Communications* 11 (1): 3295. https://doi.org/10.1038/s41467-020-17142-3.

Zadrozny, Bianca, Campbell D. Watson, Daniela Szwarcman, Daniel Civitarese, Dario Oliveira, Eduardo Rodrigues, and Jorge Guevara. 2021. "A Modular Framework for Extreme Weather Generation," February. https://arxiv.org/abs/2102.04534.

Zanna, Laure, and Thomas Bolton. 2020. "Data-Driven Equation Discovery of Ocean Mesoscale Closures." *Geophysical Research Letters* 47 (17): e2020GL088376. https://doi.org/https://doi.org/10.1029/2020GL088376.

12 Surrogate Models and Emulators

Authors: Kenny Chowdhary (SNL), Nathan Urban (BNL), Dan Lu (ORNL), Earl Lawrence (LANL), Vanessa Lopez-Marrero (BNL), Panagiotis Stinis (PNNL), Draguna Vrabie (PNNL), Shinjae Yoo (BNL), Mira Berdahl (Univ. of Washington), Chad Sockwell (SNL), Simon Topp (USGS), Tarun Verma (Princeton Univ.)

A cross-cutting theme in artificial intelligence and machine learning for Earth system predictability is the concept of *emulators* or *surrogate models*. These terms, used in different communities, refer to replacing a complex representation of a physical system (e.g., a computationally expensive computer model) with a simpler and computationally efficient representation. This representation is often, but need not always be, a data-driven statistical or machine learning model. *Reduced order modeling* is a related concept that may refer to mechanistic or dynamical surrogates, and *digital twins* may refer to data-driven or physics-based simulators of real-world observations or instruments.

Emulators have a variety of applications:

- Fast surrogates for uncertainty quantification (enabling large ensemble studies)
- Emulation of Earth system processes for scientific understanding and model improvement
- Generative models of synthetic data for impacts and predictability studies
- Reduced models for integrated multisector dynamics modeling
- Model-data fusion, multimodel/multimodal synthesis, downscaling, and bias correction
- Surrogates for observing systems or instrument simulators
- Surrogates for scientific workflows or computational pipelines

In what follows, we outline the three main challenges with surrogate modeling for Earth system predictability (ESP), followed by a discussion of the current state of the science, future data and research opportunities, and a timeline of long- and short-term goals.

12.1 Grand Challenges

12.1.1 Data and Computational Challenges

While individual samples of simulation or observation data (e.g., Earth system model [ESM] state vectors) may contain millions of degrees of freedom, the total number of samples or training set sizes (number of statistically independent snapshots, e.g., days, months, or simulation runs) are oftentimes orders of magnitude smaller (e.g., dozens or even less for very high-resolution model runs). This makes emulating data at multiple scales in space and time very difficult. Furthermore, data can be highly dynamic (e.g., turbulent flows) and not easily reducible (e.g., projection onto some fixed basis). Therefore, it is essential that (1) our training methods can perform well in both data-sparse and data-rich environments; (2) are scalable with high-dimensional quantities of interest, large training sets, and multimodal data; and (3) we employ more computational resources to generate larger training set sizes.

12.1.2 Uncertainty Quantification

Can we build surrogates for scenarios that are not (well) represented in training data: different model structures (e.g., parameterization schemes), different boundary conditions, or rare or no-analog events, etc.? Many surrogate techniques that take a purely data-driven approach are physics-agnostic, which can lead to predictions that are not realistic. How do we hybridize physics and ML machinery to improve model predictability, leveraging existing methods, tools, and software? Furthermore, can we characterize or bound the uncertainties in such an extrapolation, especially in cases where accuracy is poor? We need surrogates that can account for multiple sources and types of uncertainty (parametric, boundary condition, initial condition, model structural error/multimodel).

12.1.3 Actionable, Trustworthy Surrogates

As a natural follow-up to the previous discussion, building actionable and trustworthy surrogates that are both meaningful and interpretable is of great importance for future advancements in Earth system predictability. This challenge, perhaps, represents the culmination of the previous two, where generalizability is crucial. We need surrogates that can adapt, for example, correcting biases and scale mismatches, imputing missing data, combining hierarchies of models and data, fusing multimodel/multimodal data, adding prediction error bars, etc. While there is no one-size-fits-all surrogate method, we need approaches that are reproducible and generalizable to more than a single application. To accomplish this, we need more automated training tools and meta-learning approaches for surrogate construction that can adapt to, for example, data-rich or data-sparse environments. Furthermore, in a very sparse data setting, it may even be necessary to combine expert decision-making with ML and/or physics information for the greatest impact.

12.2 State of the Science

In this section, we discuss past and current approaches in surrogate modeling for ESP. This is an active area of research, and while many advancements have been made, there are common pitfalls and roadblocks, such as resource limitations, lack of reproducibility and generalization, inadequate quantification of uncertainty, and interpretability issues. Nonetheless, progress is ongoing, and it is gaining speed given the advent of scalable software tools and the sheer momentum of the machine learning world.

12.2.1 Reduced Order Modeling (ROM)

ROMs represent a class of surrogates that use dimension reduction techniques in combination with classical machine learning approaches to reduce the burden of training high-dimensional or multi-objective targets. They do this, primarily, by learning low-dimensional structure from covariance information of samples or snapshots of the solution space. There are generally two types of ROM approaches: (1) an intrusive approach which requires a reformulation of the simulation code in terms of the reduced order ROM basis, similar to a spectral or FEM approach, and (2) a purely data-driven ROM approach (DDROMs) where the surrogate construction is performed in a separate offline stage. The latter is often more practical, especially with complex simulation code such as those in global climate models. But like any data-driven approach, its accuracy is heavily dependent on the training data sample size, which requires repeated simulations of the underlying physical system. Lu and Riccuito (2019) have shown these approaches to be effective, especially in the limited data realm, applied to a simplified version of E3SM. Salter et al. (2019) demonstrated the scalability and accuracy of these approaches to the Canadian atmosphere model, CanAM4. Chowdhary et al. (2021) have successfully applied these surrogate techniques to high-fidelity exascale simulations in aerodynamics, which have similar computational resource burdens to large-scale climate models.

12.2.2 Stochastic Emulators

Stochastic emulators represent a different class of surrogates that provide a probabilistic, rather than a deterministic, prediction model. That is, the predictions are given in terms of the statistical properties of the system (e.g., means and variances). Perhaps the most widely known and commonly used stochastic emulators are Gaussian process (GP) models. They are flexible in that they can emulate linear trends, periodic information, wave-like behavior, etc., all by choosing an appropriate covariance kernel. While these approaches have been around for decades and are well understood, they traditionally scale poorly in the presence of large amounts of training data. Recent advancements, however, have broken through this barrier with the advent of scalable GP training methods like GPyTorch (Gardner et al. 2018). Alternatively, Bayesian neural networks (BNNs) and ensemble deep learning approaches represent a more modern class of stochastic emulators that scale very well with large amounts of data but are not as well understood. Ensemble deep learning methods provide a conceptually similar but alternate probabilistic approach by generating a suite of ensembles for the quantity of interest, similar to a random forest type approach. While these approaches scale well with large amounts of data, training these probabilistic neural network models can be cost-prohibitive for very large networks. Methods like variational inference techniques, which constrain the probabilistic form of the target to standard normal distributions, can greatly reduce this cost, but at the expense of explainability and expressivity. Many of these methods are aimed at quantifying epistemic uncertainty (such as emulator and model errors). Generative models, a different class of stochastic emulator, seek to quantify aleatoric uncertainty (such as internal/chaotic natural variability) and sample random realizations of this uncertainty. These may include statistical models (Vesely et al. 2019; Link et al. 2019; Mészáros et al. 2021; Verdin et al. 2019), dynamical reduced models (Foster, Comeau, and Urban 2020), variational autoencoders (Tibau et al. 2021), and normalizing flows (Groenke, Madaus, and Monteleoni 2020). Dunbar et al. (2021), Berdahl et al. (2021), and Beusch, Gudmundsson, and Seneviratne (2020) have all utilized a Gaussian process emulator approach for the calibration of an idealized global climate model (GCM) and for the CISM ice sheet model, respectively (see Figure 12-1), Cleary et al. (2021) also proposed a calibrate-emulate-sample approach using GPs, and Watson-Parris et al. (2021) have released open-source software for Earth system emulation, which is built on top of GPyTorch. Yang, Zhang, and Karniadakis (2018) and Warner et al. (2020) have explored the intersection of physics-informed networks and general adversarial networks (PIGANs) for stochastic models in stochastic PDEs and solid mechanics, respectively. In summary, stochastic emulators are able to provide a measure of uncertainty and trustworthiness that their deterministic counterparts cannot, but at the expense of increased training time and model complexity.

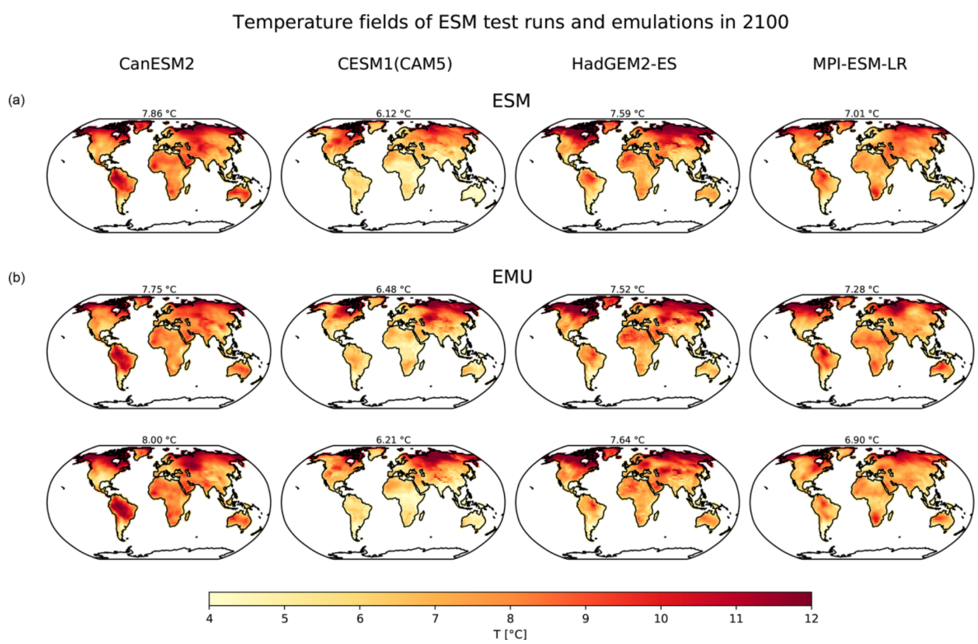


Figure 12-1. Gaussian process emulation for the global temperature field for four different Earth system models. (a) shows a simulation test run for each ESM (indexed by columns) and (b) shows two realizations of the emulated field (Source: Reproduced from Beusch, Gudmundsson, and Seneviratne 2020 under Creative Commons CC BY 4.0).

12.2.3 Physics Informed Machine Learning

Physics-informed machine learning, in general, is a type of ML technique that utilizes prior knowledge of the physics of a particular system to help train surrogates. In the broadest sense, using data generated from a physics simulation to train a surrogate or emulator is a basic example of a physics-informed ML approach. A more advanced and increasingly popular type of approach is to add physics constraints directly into the training procedure (e.g., the loss function). This can lead to significantly better extrapolative model performance (e.g., predicting events well outside the scope of the training data). The caveat is that adding physics constraints into the training process is not easy, especially when the constraints are in the form of differential equations. However, with the advent of automatic differentiation-enabled software, e.g., Tensorflow (Abadi et al. 2016), PyTorch (Pazske et al. 2019), and JAX (Bradbury et al. 2018), these types of constraints, which at one point were nearly impossible to implement, are now fairly easy to encode. This has led to significant advancements in physics-informed neural networks (PINNS) (Karniadakis et al. 2021). Most of the work in this field has been concentrated on idealized CFD and climate simulations and has not been applied to more complex or real-world GCMs (Beucler et al. 2019; Raissi, Perdikaris, and Karniadakis 2019), in particular, because of the difficulty and code-complexity of the respective systems. Kashinath et al. (2021) has a comprehensive survey of physics-informed machine learning methods where the key takeaways are the need to properly quantify uncertainty, improve interpretability, and encourage reproducibility. Oftentimes these approaches are too narrow in their focus and need to be built for broader applications and more general use.

12.2.4 Multifidelity / Multiresolution Modeling

Multifidelity approaches are becoming increasingly necessary when the training data sizes are small, oftentimes due to the heavy computational burden of state-of-the-art climate models such as E3SM (E3SM Project 2019). In these cases, efficient sampling approaches and adaptive surrogates can be used, but when data are extremely sparse, even these methods can fail. In extremely data-sparse settings, using data from both low-fidelity and high-fidelity simulations, where the former is in much greater abundance, is a better option. Fletcher, McNally, and Virgin (2021) and Anderson and Lucas (2018) have shown that multifidelity training is feasible and can lead to improvements in accuracy for CAM4 and CAM5 examples. More recent approaches by Liu, Pareschi, and Zhu (2021) and Xu and Narayan (2021) seek to find efficient and mathematically rigorous ways of learning a surrogate model with bi- or multifidelity training sources, but these approaches have not yet been tested on global climate models.

12.2.5 Neural Networks and Deep Learning

Neural networks continue to play an important role in the advancement of surrogate models for applications in climate due to their flexibility to fit many types of quantities of interest, from single scalar targets to multi-objective spatially varying fields (Fletcher, McNally, and Virgin (2021; see Figure 12-2), the ubiquity and availability of automatic differentiation-enabled software tools like tensorflow (Abadi et al. 2015), and the ability to incorporate complicated physics constraints (Karniadakis et al. 2021). These approaches with respect to climate modeling are still in their infancy, and trustworthiness and interpretability can be lacking. Rasp, Pritchard, and Gentine (2018) have shown the feasibility of applying deep learning surrogates on a smaller scale to replace subgrid physics processes, but they also note difficulties when extrapolating far beyond the training data—a common problem with data-driven machine learning methods. For a more complete discussion, see Neural Networks (chapter 11).

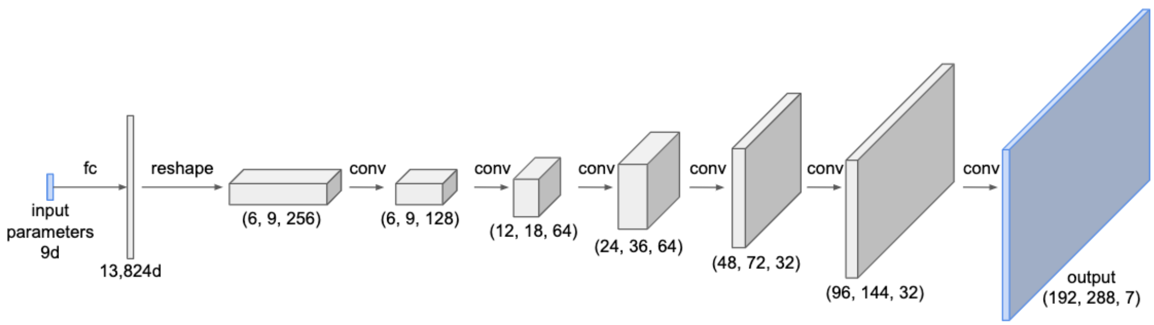


Figure 12-2. Convolutional neural network surrogate model for multitarget (spatially varying fields), multiresolution, and multimodal (7 different climatological fields) prediction for CESM (Source: Reproduced from Fletcher, McNally, and Virgin 2021 under Creative Commons CC BY 4.0).

12.2.6 Automated ML and Metalearning

Model tuning, hyper-parameter tuning, or meta-learning is an often-overlooked challenge in training a robust and trustworthy surrogate model. Without proper model tuning, for example, exploration of different layer widths and depths for neural network models, how do we know we are selecting the best model architecture for the task at hand? The most practical and clearest approach is through trial and error using model selection criteria such as cross-validation. This is oftentimes neglected, because training is much more expensive, i.e., requiring multiple fits with different models. Swischuk et al. (2019), Chowdhary et al. (2021), and Penwarden et al. (2021) are among few authors addressing this problem, and even fewer provide automated ML tools and software to implement methodologies such as *tesuract* (Chowdhary 2022). Furthermore, hyper-parameter tuning can sometimes reveal surprising results—taking existing approaches like multivariate polynomial regression and showing that, in some cases, it can be competitive with or even better than state-of-the-art approaches like neural networks (Chowdhary 2022).

Automated ML and meta-learning may also result in more robust and generalizable models, making it a critical component of model training.

12.3 Experimental, Data, and Modeling Opportunities

There are three major opportunities that can lead to significant technological advances in the use of surrogate models for AI4ESP. First, ***a new paradigm must be adopted by the climate modeling world, in which data generation is tightly integrated with model development***. Similar to how resources are allocated for regression and unit testing of code bases, resources should be strictly allocated to running perturbed parameter simulations for the purposes of enriching surrogate training data. Too often, running uncertainty quantification or machine learning studies have been low on the priority list, but this must change. The greatest advances in deep learning, for example, have occurred when scalable models meet huge amounts of training data, as with AlexNet (Krizhevsky, Sutskever, and Hinton 2017); DALL-E (Ramesh et al. 2021); and CLIP (Radford et al. 2021). In lieu of more training datasets, however, we should look to enrich existing datasets with different modes of solutions. For example, generating spatiotemporal data instead of time-averaged datasets would allow researchers to utilize different types of surrogates, such as ROM-based or operator neural networks, which could lead to significant improvements in understanding the dynamics of the problem (Kovachki et al. 2021; Lu et al. 2021) or characterization of non-parametric (e.g., structural) uncertainties (DeGennaro et al. 2019).

But how do we determine what data to generate and which modes to collect? This brings us to the second major data and experimental opportunity, which is ***the creation of a comprehensive database of benchmark test problems for surrogate modeling in ESP***. Upon consultation with ML and climate experts, and under the guidance of the FAIR (Findability, Accessibility, Interoperability, Reuse) principals, we should develop a suite of test problems for assessing the accuracy and effectiveness of different surrogate approaches. These test problems should test the range of surrogate models outlined in the previous State of the Science section (12.2). A unified test bed of experiments will also prevent overlapping discovery and progress in surrogate methodologies. Furthermore, this can mitigate the need to recreate methods from scratch, and instead, build upon existing approaches. Oftentimes surrogate approaches are developed and validated independently of one another, and it is difficult to make fair comparisons between different approaches.

The third major modeling opportunity involves ***the implementation of adjoint-based or automatic/approximate derivative calculations or differentiable programming for future Earth system models***, if at all possible. This could significantly improve surrogate construction methods by enabling adaptive sampling, adjoint-based error estimation (Jakeman and Wildey 2014), parameterization learning (Yang, Aziz Bhouri, and Perdikaris 2020; Melland, Albright, and Urban 2021), and utilization of ROM-based projection methods via existing software tools like Pressio (Rizzi et al. 2020), which could potentially speed up existing GCMs by an order of magnitude or more.

12.4 Research Priorities

In this section, we outline four main research priorities for the advancement of next-generation surrogate modeling for ESP.

12.4.1 Scalable (HPC-exploiting) Surrogates for Very Small or Large Training Sets

One of the grand challenges in Earth system predictability is that surrogate training may involve very small training set sizes, where ensembles are limited or even impossible, or simulation or observation durations are too short to provide many independent data points. Correspondingly, ML approaches are needed that work in the *low data* or *data sparse* regime (small training sets), requiring generalization from "few-shot learning" techniques. The needed advances may not come purely from computer science, but also from using physics knowledge as a regularization or constraint (knowledge-informed ML), or otherwise exploiting aspects of the problem structure in ML modeling and training. Surrogates with quantified uncertainties may become particularly important in the low-data regime. Data imputation or generative synthetic data approaches may be needed to augment sparse training sets, including scenarios where data are proprietary (situations where federated learning and differential privacy may also come into play). On the other end of the spectrum, the more traditional *big data* or *data-rich* regime also exists in Earth science, where datasets can be so large they cannot even be stored to disk, potentially requiring online/streaming/in situ training.

12.4.2 Domain-aware Surrogate Models

New types of surrogate models may be needed to tackle grand challenges related to Earth system predictability, customized to be either more reliable or more efficient on particular applications in this domain. This includes the whole scope of physics- or knowledge-informed AI, as well as hybrid physics-ML modeling; ML models that respect physical symmetries, conservation laws, locality, causality, and so on may be not only more accurate but also more efficient to train, by exploiting a known problem structure. "Surrogate models" in this context need not be limited to ML surrogates but could also be mechanistic or dynamical reduced-order models that might be informed or improved by ML. Completely different types of surrogates might be needed to model coupled natural-human systems, including the possibility of modeling decision-making endogenously to the model by treating humans as anticipatory, goal-seeking agents.

12.4.3 Information-fusing, Uncertainty-aware Surrogates

Most surrogates to date are surrogates of a single information source, usually an expensive computer model, and (if used for uncertainty quantification) only designed to treat a single source of uncertainty, usually model parametric uncertainty. Surrogates will increasingly be asked to make projections about the real Earth system, combining information from hierarchies of models and data across multiple modalities, rather than simply replicating the behavior of one imperfect representation of the Earth system. These may extend to "digital twins" of entire scientific workflows or observational campaigns in order to prioritize the active collection of new observational or simulation data. Surrogates may be inherently probabilistic in nature, incorporating epistemic uncertainties (such as model error) and stochastic generative models of aleatoric uncertainties (such as internal or chaotic natural variability), and they may be designed to compose and propagate high-dimensional probability distributions across system components.

12.4.4 Software Ecosystems and Workflow for Surrogate Development and Deployment

To enable more extensive use of surrogate models, it will be necessary to improve workflow and productivity. Automated (e.g., self-tuning) machine learning and meta-learning or reinforcement learning approaches may reduce both developer time, particularly among non-expert domain scientists, as well as computational time (e.g., by devising intelligent or adaptive training protocols). Better user guidelines and meta-heuristics may be used to decide what type of surrogate would be appropriate for a given problem. Generative models or other approaches may be used as synthetic data augmentation or missing data imputation for training. High-level programming environments for surrogate model development, inspired by probabilistic programming languages and potentially interacting with compiler technology, may allow users to spend more time on problem formulation (specifying data relationships, distributional assumptions, physics constraints, etc.) while allowing software to generate efficient implementations. Software frameworks may facilitate closer integration of ML surrogates with physical simulation codes, including online coupled training and prediction (see Figure 12-3 for an example workflow for climate model calibration using GP emulators). New advances in interpretability and reproducibility will be needed to facilitate user trust in surrogate model approximations, working closely with uncertainty quantification methods to self-critique their trustworthiness and domain of applicability.

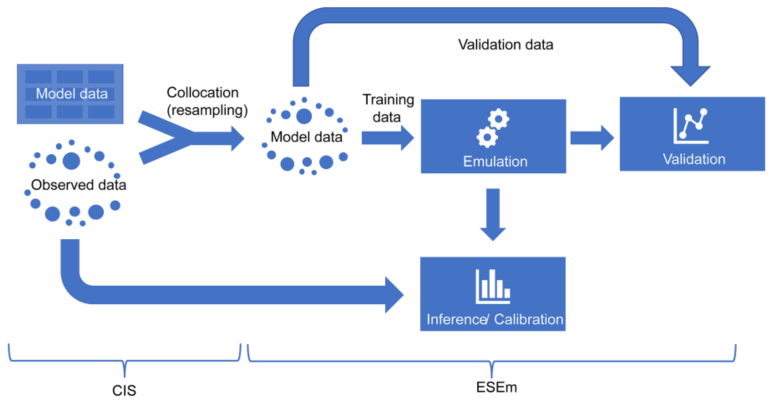


Figure 12-3. Watson-Parris et al. (2021) outline a general-purpose workflow (ESEm) for surrogate model calibration that can integrate into existing GCMs. Model data comes from simulations, after which a surrogate or emulator is used to infer model parameters for calibration (Source: Reproduced from Watson-Parris et al. 2021 under Creative Commons CC BY 4.0).

12.5 Short-term (<5 years), 5-year, and 10-year Goals

Immediate and short-term goals (< 5 years) should focus on reproducibility (publishable and testable code, open-source datasets) of current surrogate methodologies. This would be followed by the creation of interdisciplinary working groups composed of experts in both AI and climate science to better understand specific data needs in order to create a set of benchmark validation problems and datasets for surrogate model development and testing. Furthermore, these working groups would also help develop a taxonomy of scalable, domain-aware, and multimodal surrogate models under different types of model and structural uncertainties to act as a roadmap for long-term goals.

An intermediate/5-year milestone should be the development of a comprehensive set of benchmark problems and datasets for validation surrogate methods in ESP. This should be pursued in parallel with the development and testing of rigorous surrogate methodologies that address the research priorities described in section 12.4, on a variety of Earth system models (e.g., regional, global, and/or fully coupled). Software design workflows and automated tuning algorithms should also be leveraged. Equally important, tighter integration between ML data generation and climate model development should be made, as outlined in Experimental, Data, and Modeling Opportunities, section 12.3. And pursuit of new numerical features like adjoint-based derivatives and hybrid ML/physics approaches should be well underway.

In the long term, somewhere near the 10-year mark, a comprehensive testbed and a rigorous exploration of surrogate approaches should be complete, for example, with a clearer understanding of the advantages, limitations, trustworthiness, and mathematical properties. In particular, surrogate capabilities should include multiple modalities, multiresolution/ multifidelity, uncertainty quantification of different structural forms, both data-sparse and data-rich environments, and quantifiable out-of-distribution or rare event predictions, and meta-learning approaches would automatically select the right approach for the task at hand.

12.6 References

Abadi, Martín, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, et al. 2016. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," March. https://doi.org/10.48550/arxiv.1603.04467.

Anderson, Gemma J., and Donald D. Lucas. 2018. "Machine Learning Predictions of a Multiresolution Climate Model Ensemble." *Geophysical Research Letters* 45 (9): 4273–80. https://doi.org/10.1029/2018GL077049.

Berdahl, Mira, Gunter Leguy, William H. Lipscomb, and Nathan M. Urban. 2021. "Statistical Emulation of a Perturbed Basal Melt Ensemble of an Ice Sheet Model to Better Quantify Antarctic Sea Level Rise Uncertainties." *The Cryosphere* 15 (6): 2683–99. https://doi.org/10.5194/tc-15-2683-2021.

Beucler, Tom, Stephan Rasp, Michael Pritchard, and Pierre Gentine. 2019. "Achieving Conservation of Energy in Neural Network Emulators for Climate Modeling," June. https://doi.org/10.48550/arxiv.1906.06622.

Beusch, Lea, Lukas Gudmundsson, and Sonia I. Seneviratne. 2020. "Emulating Earth System Model Temperatures with MESMER: From Global Mean Temperature Trajectories to Grid-Point-Level Realizations on Land." *Earth System Dynamics* 11 (1): 139–59. https://doi.org/10.5194/esd-11-139-2020.

Bradbury, James, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, et al. 2018. "JAX: Composable Transformations of Python+NumPy Programs." http://github.com/google/jax.

Chowdhary, Kenny. 2022. "Tesuract: Tensor Surrogate Automation and Computation." Accessed June 27, 2022. https://github.com/kennychowdhary/tesuract.

Chowdhary, Kenny, Chi Hoang, Kookjin Lee, and Jaideep Ray. 2021. "Calibrating Hypersonic Turbulence Flow Models with the HIFiRE-1 Experiment Using Data-Driven Machine-Learned Models," December. https://doi.org/10.48550/arxiv.2112.09681.

Cleary, Emmet, Alfredo Garbuno-Inigo, Shiwei Lan, Tapio Schneider, and Andrew M. Stuart. 2021. "Calibrate, Emulate, Sample." *Journal of Computational Physics* 424 (January): 109716. https://doi.org/10.1016/J.JCP.2020.109716.

DeGennaro, Anthony M., Nathan M. Urban, Balasubramanya T. Nadiga, and Terry Haut. 2019. "Model Structural Inference Using Local Dynamic Operators." *International Journal for Uncertainty Quantification* 9 (1): 59–83. https://doi.org/10.1615/INT.J.UNCERTAINTYQUANTIFICATION.2019025828.

Dunbar, Oliver R. A., Alfredo Garbuno-Inigo, Tapio Schneider, and Andrew M. Stuart. 2021. "Calibration and Uncertainty Quantification of Convective Parameters in an Idealized GCM." *Journal of Advances in Modeling Earth Systems* 13 (9): e2020MS002454. https://doi.org/https://doi.org/10.1029/2020MS002454.

E3SM Project, DOE. 2019. "Energy Exascale Earth System Model v1.3." https://doi.org/10.11578/E3SM/dc.20210924.5.

Fletcher, Christopher G., William McNally, and John G. Virgin. 2021. "Toward Efficient Calibration of Higher-Resolution Earth System Models." https://doi.org/10.1002/ESSOAR.10508056.1.

Foster, Dallas, Darin Comeau, and Nathan M. Urban. 2020. "A Bayesian Approach to Regional Decadal Predictability: Sparse Parameter Estimation in High-Dimensional Linear Inverse Models of High-Latitude Sea Surface Temperature Variability." *Journal of Climate* 33 (14): 6065–81. https://doi.org/10.1175/JCLI-D-19-0769.1.

Gardner, Jacob R., Geoff Pleiss, David Bindel, Kilian Q. Weinberger, and Andrew Gordon Wilson. 2018. "GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration," September. https://doi.org/10.48550/arxiv.1809.11165.

Groenke, Brian, Luke Madaus, and Claire Monteleoni. 2020. "ClimAlign: Unsupervised Statistical Downscaling of Climate Variables via Normalizing Flows." In *Proceedings of the 10th International Conference on Climate Informatics*, 60–66. CI2020. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3429309.3429318.

Jakeman, John D., and Timothy Wildey. 2014. "Enhancing Adaptive Sparse Grid Approximations and Improving Refinement Strategies Using Adjoint-Based a Posteriori Error Estimates," July. https://doi.org/10.1016/j.jcp.2014.09.014.

Karniadakis, George Em, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. 2021. "Physics-Informed Machine Learning." *Nature Reviews Physics 2021 3:6* 3 (6): 422–40. https://doi.org/10.1038/s42254-021-00314-5.

Kashinath, K., M. Mustafa, A. Albert, J.-L. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, et al. 2021. "Physics-Informed Machine Learning: Case Studies for Weather and Climate Modelling." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379 (2194): 20200093. https://doi.org/10.1098/rsta.2020.0093.

Kovachki, Nikola, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. 2021. "Neural Operator: Learning Maps Between Function Spaces," August. https://doi.org/10.48550/arxiv.2108.08481.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2017. "ImageNet Classification with Deep Convolutional Neural Networks." *Communications of the ACM* 60 (6): 84–90. https://doi.org/10.1145/3065386.

Link, Robert, Abigail Snyder, Cary Lynch, Corinne Hartin, Ben Kravitz, and Ben Bond-Lamberty. 2019. "Fldgen v1.0: An Emulator with Internal Variability and Space–Time Correlation for Earth System Models." *Geoscientific Model Development* 12 (4): 1477–89. https://doi.org/10.5194/gmd-12-1477-2019.

Liu, Liu, Lorenzo Pareschi, and Xueyu Zhu. 2021. "A Bi-Fidelity Stochastic Collocation Method for Transport Equations with Diffusive Scaling and Multi-Dimensional Random Inputs," July. https://doi.org/10.1016/j.jcp.2022.111252.

Lu, D., and D. Ricciuto. 2019. "Efficient Surrogate Modeling Methods for Large-Scale Earth System Models on Machine-Learning Techniques." *Geoscientific Model Development* 12 (5): 1791–1807. https://doi.org/10.5194/gmd-12-1791-2019.

Lu, Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. 2021. "Learning Nonlinear Operators via DeepONet Based on the Universal Approximation Theorem of Operators." *Nature Machine Intelligence 2021 3:3* 3 (3): 218–29. https://doi.org/10.1038/s42256-021-00302-5.

Melland, Pake, Jason Albright, and Nathan M. Urban. 2021. "Differentiable Programming for Online Training of a Neural Artificial Viscosity Function within a Staggered Grid Lagrangian Hydrodynamics Scheme." *Machine Learning: Science and Technology* 2 (2): 025015. https://doi.org/10.1088/2632-2153/abd644.

Mészáros, Lőrinc, Frank van der Meulen, Geurt Jongbloed, and Ghada el Serafy. 2021. "A Bayesian Stochastic Generator to Complement Existing Climate Change Scenarios: Supporting Uncertainty Quantification in Marine and Coastal Ecosystems." *Stochastic Environmental Research and Risk Assessment* 35 (3): 719–36. https://doi.org/10.1007/S00477-020-01935-5/FIGURES/17.

Paszke, Adam, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, et al. 2019. "PyTorch: An Imperative Style, High-Performance Deep Learning Library," December. https://doi.org/10.48550/arxiv.1912.01703.

Penwarden, Michael, Shandian Zhe, Akil Narayan, and Robert M. Kirby. 2021. "Physics-Informed Neural Networks (PINNs) for Parameterized PDEs: A Metalearning Approach," October. https://doi.org/10.48550/arxiv.2110.13361.

Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, et al. 2021. "Learning Transferable Visual Models From Natural Language Supervision," February. https://doi.org/10.48550/arxiv.2103.00020.

Raissi, M., P. Perdikaris, and G. E. Karniadakis. 2019. "Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations." *Journal of Computational Physics* 378: 686–707. https://doi.org/https://doi.org/10.1016/j.jcp.2018.10.045.

Ramesh, Aditya, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. "Zero-Shot Text-to-Image Generation," February. https://doi.org/10.48550/arxiv.2102.12092.

Rasp, Stephan, Michael S. Pritchard, and Pierre Gentine. 2018. "Deep Learning to Represent Subgrid Processes in Climate Models." *Proceedings of the National Academy of Sciences of the United States of America* 115 (39): 9684–89. https://doi.org/10.1073/PNAS.1810286115/SUPPL_FILE/PNAS.1810286115.SAPP.PDF.

Rizzi, Francesco, Patrick J. Blonigan, Eric J. Parish, and Kevin T. Carlberg. 2020. "Pressio: Enabling Projection-Based Model Reduction for Large-Scale Nonlinear Dynamical Systems," March. https://doi.org/10.48550/arxiv.2003.07798.

Salter, James M., Daniel B. Williamson, John Scinocca, and Viatcheslav Kharin. 2019. "Uncertainty Quantification for Computer Models With Spatial Output Using Calibration-Optimal Bases." *Journal of the American Statistical Association* 114 (528): 1800–1814. https://doi.org/10.1080/01621459.2018.1514306.

Swischuk, Renee, Laura Mainini, Benjamin Peherstorfer, and Karen Willcox. 2019. "Projection-Based Model Reduction: Formulations for Physics-Based Machine Learning." *Computers & Fluids* 179 (January): 704–17. https://doi.org/10.1016/J.COMPFLUID.2018.07.021.

Tibau, Xavier-Andoni, Christian Reimers, Christian Requena-Mesa, and Jakob Runge. 2021. "Spatio-Temporal Autoencoders in Weather and Climate Research." *Deep Learning for the Earth Sciences*, September, 186–203. https://doi.org/10.1002/9781119646181.CH13.

Verdin, Andrew, Balaji Rajagopalan, William Kleiber, Guillermo Podestá, and Federico Bert. 2019. "BayGEN: A Bayesian Space-Time Stochastic Weather Generator." *Water Resources Research* 55 (4): 2900–2915. https://doi.org/https://doi.org/10.1029/2017WR022473.

Vesely, Fosco M., Livia Paleari, Ermes Movedi, Gianni Bellocchi, and Roberto Confalonieri. 2019. "Quantifying Uncertainty Due to Stochastic Weather Generators in Climate Change Impact Studies." *Scientific Reports 2019 9:1* 9 (1): 1–8. https://doi.org/10.1038/s41598-019-45745-4.

Warner, James E., Julian Cuevas, Geoffrey F. Bomarito, Patrick E. Leser, and William P. Leser. 2020. "Inverse Estimation of Elastic Modulus Using Physics-Informed Generative Adversarial Networks," May. https://doi.org/10.48550/arxiv.2006.05791.

Watson-Parris, Duncan, Andrew Williams, Lucia Deaconu, and Philip Stier. 2021. "Model Calibration Using ESEm v1.1.0-an Open, Scalable Earth System Emulator." *Geoscientific Model Development* 14 (12): 7659–72. https://doi.org/10.5194/GMD-14-7659-2021.

Xu, Yiming, and Akil Narayan. 2021. "Budget-Limited Distribution Learning in Multifidelity Problems," May. https://doi.org/10.48550/arxiv.2105.04599.

Yang, Liu, Dongkun Zhang, and George Em Karniadakis. 2018. "Physics-Informed Generative Adversarial Networks for Stochastic Differential Equations," November. https://doi.org/10.48550/arxiv.1811.02033.

Yang, Yibo, Mohamed Aziz Bhouri, and Paris Perdikaris. 2020. "Bayesian Differential Programming for Robust Systems Identification under Uncertainty." *Proceedings of the Royal Society A* 476 (2243). https://doi.org/10.1098/RSPA.2020.0290.

13 Knowledge-Informed Machine Learning

Authors: Francis J. Alexander (BNL), Mahantesh Halappanavar (PNNL), Karthik Kashinath (LBNL/NVIDIA), Shinjae Yoo (BNL), Arvind Thanam Mohan (LANL), Jayaraman Thiagarajan, Sutanay Choudhury (PNNL), Vipin Kumar (Univ. of Minnesota), Ján Drgoňa (PNNL), Charity Plata (BNL), Yansaka S. Horawalavithana (PNNL)

Acknowledgement: Session leaders & workshop organizers would like to acknowledge the significant contribution to report of Ramakrishna Tipireddy.

13.1 Introduction

Artificial intelligence (AI) has the potential to revolutionize climate science by providing tools to bridge information from raw scientific datasets (experiments/observations/simulations) to machine-readable representations of scientific knowledge. Maximizing the positive impact of AI on climate will require going beyond a purely data-driven approach and integrating information and knowledge of various disparate forms of data along the way to deep human understanding of processes. This **knowledge-informed machine learning** approach, or KIML, confers several benefits to "traditional" data-driven learning algorithms. For example, domain knowledge can be used effectively to enhance data-sparse situations. These situations occur in complex systems, such as climate. Domain knowledge (i.e., physics constraints in terms of symmetries and equations) can act as a regularization for the loss function in training, leading to improved generalization and better performance guarantees (stability, constraints satisfaction, optimality conditions). Such knowledge also ensures that the learned model is consistent and respects the laws of physics. In addition, domain knowledge improves the interpretability of the learned model through parsimonious and human-readable representations. Carefully chosen prior information—gleaned from parameter initialization, warm-starting, or domain-informed update strategies—also can significantly improve the convergence of model training.

Domain knowledge can be added in these various forms during multiple stages of the learning process. Augmenting data with this "extra" information is important because for complex systems, data alone may not be enough to construct good machine learning (ML) algorithms (predictors, classifiers, etc.). Integration of the aforementioned multimodal knowledge types can occur via gray box digital twins, physics-informed ML, or graphical models (among others). Options for integrating all such knowledge into ML models include loss functions, pretraining, architecture design, learning with constrained optimization solvers, ensemble/federated learning, meta-learning, transfer learning, and inductive/learning bias. Unsurprisingly, there is considerable overlap with other topics covered by this AI4ESP report, spanning knowledge discovery, knowledge representation, explainability, uncertainty quantification (UQ), surrogate models, bias, ethics (human-in-the-loop), and interfaces/standards.

The following considers three types of *domain* knowledge. Notably, there are multiple variations where these domain knowledge representations can be developed (Figure13-1), including:

- *Scientific knowledge* that focuses on incorporating information about a domain captured in the form of various mathematical relationships (partial differential equations [PDEs], integro-differential equations, etc.) and constraints.
- *World knowledge* that seeks to integrate information from a broader set of external data sources beyond traditional scientific datasets, including spatial invariances, logic rules, qualitative information, causal dependencies, and knowledge graphs.
- *Expert knowledge* that arises from combining information across different modeling paradigms to simulate and predict natural phenomena with higher fidelity. This includes probabilistic relations and human feedback.

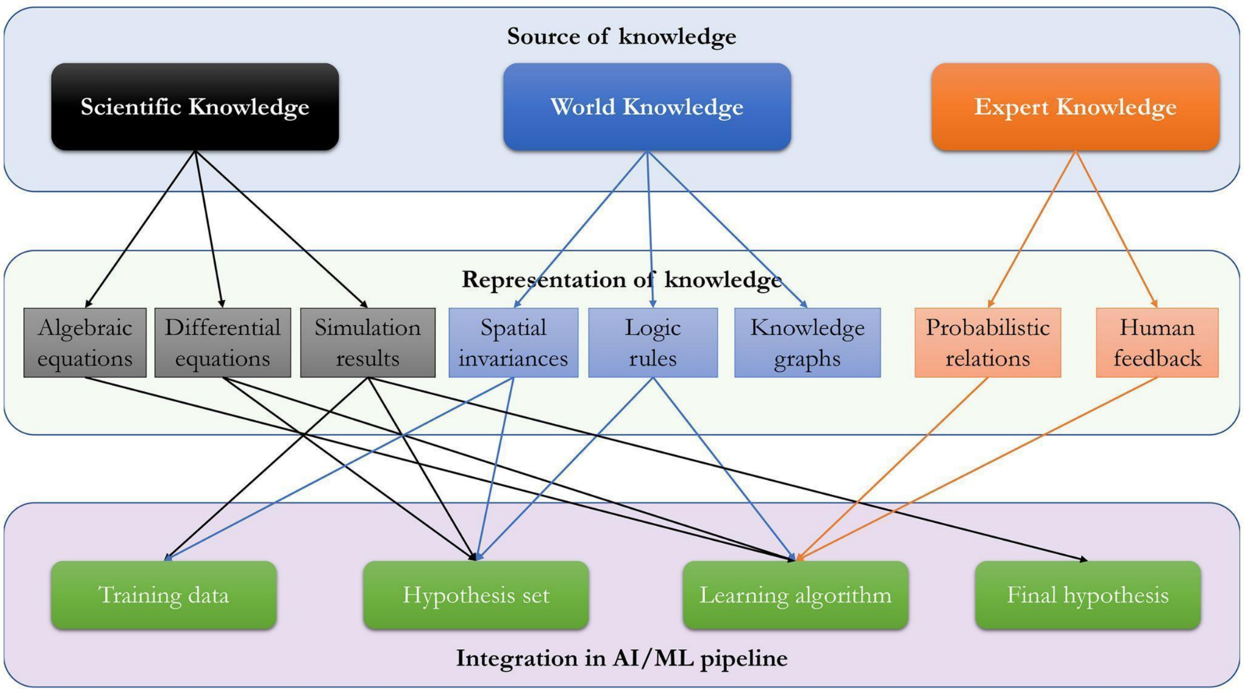


Figure 13-1. Taxonomy of informed ML (Source: Pacific Northwest National Laboratory. Adapted from von Rueden et al. 2021).

13.2 Grand Challenges

Contributors in this session identified the following Grand Challenges.

13.2.1 Extracting Actionable Climate Insights from Vast Open-source Data, including Climate Literature

Due to the inherent volume of literature growth, it is difficult for researchers to master all available climate-related information. The logical relationships or causation in climate communities are represented in literature text, tables, and figures. Named entities and their relationship findings could help to infer logical relationships. Sometimes, they are also represented in figures or tables. Natural language processing (NLP) and computer vision tools may help to extract valuable knowledge.

13.2.2 Creating Causal Inference Algorithms on Climate/Earth Datasets, Which Are Often Sparse

Causal inference algorithms in other scientific disciplines (e.g., medicine or economics) are not designed to scale or to detect the challenges in climate/Earth system science (ESS) data. Brute force application of such algorithms would either produce spurious causation or be unable to scale the problems. Although there is preliminary work in these areas, opportunities remain to further develop causal inference algorithms, including, for example, robust, nonlinear causation and UQ.

13.2.3 Utilizing Transfer and Self-supervised Learning as Tools to Store and Incorporate Parsimonious Representations of Climate Data

The inherent size of spatial resolution and potentially temporal context size makes it difficult to have a compressed representation of climate data. As the compressed representation model is trained on a domain, the ability to quickly adapt to a different domain without significant retraining would be critical for sharing the knowledge.

13.2.4 Guaranteeing Faster Convergence of KIML Methods

Integration of domain knowledge, while beneficial, potentially can make ML models complex and difficult to implement. Therefore, providing theoretical guarantees on optimality and constraints satisfaction and analytical bounds on convergence becomes important. While knowledge inclusion often can lead to lower requirements for data, theoretical estimations and guarantees on the amount of data are important Grand Challenge components.

13.2.5 Incorporating Uncertainties and Errors in Domain Knowledge

Epistemic and aleatory uncertainties from the inclusion of domain knowledge need to be carefully considered and propagated through the ML models. Uncertainty estimates and bounds on downstream inference tasks encompass an important Grand Challenge. Uncertainties are also generated when knowledge is exploited for data generation purposes, which consequently leads to uncertainties in learned models.

13.2.6 Choice of Models, ML Frameworks, and Algorithms

Although there is a large body of literature regarding combining knowledge and ML methods, such as physics-informed neural networks (PINNs) or DeepONet, there is no guidance or "rule of thumb" for choosing the amount of knowledge (e.g., mathematical models), ML architectures, or data needed to efficiently train and infer quantities of interest (QoI). The main challenge is to analyze these techniques mathematically and test on a set of benchmark problems and datasets.

13.2.7 Developing Standardized Benchmark Problems, Datasets, and Metrics

The success of ML methods for many applications, such as image classification, stems from the availability of large benchmark datasets and standardized metrics to assess new algorithms on the test data. Currently, development of a similar set of data, benchmark problems, and standardized metrics that can potentially help accelerate research in this direction does not keep up with the fast-evolving KIML field.

13.2.8 Using KIML to Discover New Physics/Mathematics/Theory

Currently, KIML methods are being developed to combine knowledge, while ML methods infer the QoI in a physical phenomenon. Given a set of observations/data and partial knowledge, is it possible to discover new physics/mathematics (e.g., governing equations) that govern the physical phenomenon? The development of new algorithms, tools, and mathematical frameworks that systematically combine traditional physics with data-driven models will be needed to tackle this challenge.

13.2.9 Augmenting Unknown Physics and Incomplete Data

In disciplines such as biological and behavioral sciences, observational data are incomplete, and the underlying physical phenomena are not completely understood. A major challenge is to integrate knowledge and ML algorithms in such a way that the ML algorithm can explore design spaces and identify correlations, while knowledge (mathematical equations; domain expertise) can accurately predict system dynamics and identify causality. There are several challenges that KIML can address, including solving ill-posed problems, identifying missing information, creating surrogate models, automating/eliminating space-time discretization, supplementing training data, quantifying uncertainty, exploring massive datasets, elucidating unknown mechanisms, preventing overfitting, minimizing data bias, and increasing rigor and reproducibility.

13.2.10 Determining Sufficiency and Optimality of Domain Constraints in KIML Methods

Another key challenge that arises is the necessity and sufficiency of knowledge, that is, *how much knowledge is sufficient*? Imposing equality and inequality constraints on the ML model, for example, in the form of algebraic equations, can provide feasibility and consistency checks of the modeled phenomena. On the other hand, inclusion of domain knowledge in the form of hard constraints potentially can lead to conservativeness or loss of expressivity of the ML model. New methods will need to be developed for integration and performance analysis of ML models with complex domain constraints.

13.3 State of KIML Science

While there has been recent work to integrate knowledge of the three categories described in the Introduction—namely, scientific, world, and expert knowledge—into modern ML methods, even more must be done to achieve the full potential for climate science. The reviews by Willard et al. (2020), Kashinath et al. (2021), and von Rueden et al. (2021) contain up-to-date and comprehensive sets of references.

To ensure that KIML achieves maximum impact, additional advances in mathematics and computer science are needed, such as faster, scalable algorithms for mathematical optimization (including constrained optimization solvers), improved numerical solvers for differential equations, and numerical analysis (adaptivity; stability).

As most climate systems of interest are based on dynamical systems theory, advances in model reduction for infinite dimensional systems also will be needed. Given the goal-driven nature of many climate KIML problems, there must be added research and development (R&D) in mixed-precision computation, especially for training large neural network models. In addition, improvements to stochastic gradient descent (SGD) and more R&D involving scalable randomized algorithms likely will be required, along with advances in software development, workflow design, and data management.

The following sections summarize the current state of the science for each of the preceding Grand Challenges.

13.3.1 Extracting Actionable Climate Insights from Vast Open-source Data, including Climate Literature

Extracting information and actionable insights from vast amounts of multimodal data in the form of text, images, videos, equations, and scientific simulations broadly falls in the general class called *foundation models* (Bommasani et al. 2021). These models enable advanced functionalities, such as in-context learning through large-scale pretraining with the popular Transformer architecture (Vaswani et al. 2017; Devlin et al. 2018). While these models are trained with zero supervision on unlabeled data (via self-supervised learning), they can be easily adopted to solve various downstream tasks via fine-tuning. There are few attempts on developing multimodal foundation models to jointly learn across language and vision data, such as FLAVA (Singh et al. 2021) and Perceiver (Jaegle et al. 2021). For example, all relevant information can be fused from a domain (e.g., medical images, patient databases, and clinical text in healthcare) into the model pretraining. Then, this model can be adopted to solve tasks that span multiple modalities, such as electronic health records and medical images for clinical outcome prediction.

There are potential avenues in the development of climate-related multimodal foundation models for intelligent transportation systems (e.g., mobile sensor data or public transit times), smart energy consumption systems (e.g., gas emissions or electricity consumption), and disaster management tools (e.g., aerial imagery or social media) (Rolnick et al. 2022). On the other hand, scholarly publications provide abundant data to augment the learning process in multimodal foundation models.

Given the untenable computational and energy costs of foundation models (Strubell, Ganesh, and McCallum 2019), considerable work remains to be explored to adapt the technology for specific climate domain needs.

13.3.2 Creating Causal Inference Algorithms on Climate/Earth Datasets, Which Are Often Sparse

Climate data are essentially spatiotemporal data. Thus, spatiotemporal causal analysis algorithms, such as a regression-based approach, such as Granger Causality (Yao, Yoo, and Yu 2015), information-theoretic approach (e.g., Transfer Entropy, Lobier et al. 2014), dynamic Bayesian approach (Ghahramani 1998), PCMCI (Runge, Nowack, et al. 2019), etc., are widely adapted. There are several limitations on these causal inference algorithms to apply to complex and nonlinear dynamics systems (i.e., non-stationary causal structure, nonlinear causal link, various direct/indirect lagged causal links, etc.). To overcome these limitations, some have used an autoencoder (Ramachandra 2019; Varando, Fernández-Torres, and Camps-Valls 2021), or low-rank causal structure learning has been incorporated directly to the deep learning algorithms (Huang, Xu, and Yoo 2019). However, still there is no single algorithm that detects complex causal inference robustly. Causal inference requires careful interpretation, as well as understanding the limitations of each approach. Applying large-scale causal inference within the climate community remains in the development stage. In the interim, the climate community has adopted the PCMCI approach and has evaluated its effectiveness on Arctic climate (Nichol et al. 2021) or El Niño Southern Oscillation (ENSO) (Zhang et al. 2020), while Varando, Fernández-Torres, and Camps-Valls (2021) use Granger causal inference on autoencoder space. Some have evaluated the difference of causal structures between the observation and simulation, such as the Energy Exascale Earth System Model (E3SM), which could be a good use case for improving the E3SM simulation model.

13.3.3 Utilizing Transfer and Self-supervised Learning Methods as Tools to Store and Incorporate Parsimonious Representations of Climate Data

Current state of knowledge shows that the inherent size of spatial resolution and temporal context make it imperative to have compressed representations of climate data (Kadow, Hall, and Ulbrich 2020). The compressed representation models need the ability to adapt quickly to new domains with different spatial or temporal context without significant retraining, which is also critical for sharing the knowledge (Notarangelo et al. 2021). Transfer learning methods have shown promising results for filling observational gaps in climate model data (Hu, Zhang, and Zhou 2016; Notarangelo et al. 2021). Recently developed self-supervised learning (SSL) methods (Devlin et al. 2018) have shown the ability to capture underlying structure in data by masking part of the data and using the remainder for predicting the masked sample. Furthermore, the multiphase training strategy in SSL allows for learning a generalized data representation during the pre-training phase, which later can be fine-tuned for task-specific outcomes. Adapting such SSL architectures and multiphase strategies for transfer learning across different spatial and temporal contexts has the potential to vastly expand and enable more efficient, effective use of typically sparse data and potentially compressed models in climate science.

13.3.4 Guaranteeing Faster Convergence of KIML Methods

Integration of domain knowledge into ML models has been shown to provide faster solutions for PDEs (Raissi, Perdikaris, and Karniadakis 2019), learn surrogate models for PDEs (Li et al. 2020), support learning with hard constraints for optimization problems (Donti, Rolnick, and Kolter 2021), and solve constrained optimal control problems (Drgona, Tuor, and Vrabie 2020). While beneficial, the infusion of domain knowledge can make ML models more complex and difficult to implement, and providing theoretical guarantees on optimality and analytical bounds on convergence becomes important (Li and Orabona 2018). While knowledge inclusion often can lead to lower requirements for data, theoretical estimations and guarantees on the amount of data will be an important Grand Challenge.

13.3.5 Choice of Models, ML Frameworks, and Algorithms

Although there is a large body of literature regarding combining knowledge and ML methods (Paullada et al. 2021), such as physics-informed neural networks (PINNs) (Karniadakis et al. 2021) or DeepONet, there is no guidance or "rule of thumb" for choosing the amount of knowledge (e.g., mathematical models), ML architectures, or data needed to efficiently train and infer QoI. The deep learning community has broadly converged to some guiding principles for ML architectures (Zbontar et al. 2021; Sanchez-Gonzalez et al. 2020; de Brouwer et al. 2019), while developing systematic approaches for engineering training datasets has become an important focus area in the ML community (Paullada et al. 2021). However, it is important to analyze these mature ML frameworks mathematically and test on a set of datasets and computational tasks that represent key challenges in climate science. While efforts such as Cachay et al.'s (2021) are representative of such ambitions, these benchmarks need to be expanded significantly and swiftly to cover the range of climate science applications.

13.3.6 Developing Standardized Benchmark Problems, Datasets, and Metrics

To advance research in KIML, standardized, high-quality, and large-scale datasets that form a comprehensive suite of real-world benchmarks are needed. These benchmarks should facilitate scalable, robust, and reproducible KIML research across different application domains and different types of tasks using a diverse set of datasets. The performance needs to be evaluated with meaningful metrics to measure the research progress in a consistent and reproducible way.

Domain experts have identified that mitigation (reducing emissions) and adaptation (preparing for unavoidable consequences) techniques developed for addressing climate change are well suited for ML research (Rolnick et al. 2022; Rasp 2021). The problems can be prioritized in high-impact application domains (e.g., electricity systems; Perera, Aung, and Woon 2014; Ramchurn et al. 2012), transportation (Davis et al. 2018), and urban infrastructure (Kreider et al. 1995) with a standard benchmark suite across both mitigation and adaptation techniques. They can be further grouped according to the interest of different practitioners (e.g., local and national governments, corporate leaders, entrepreneurs).

There are many potential problems in the respective application domains. For example, forecasting supply and demand in electricity systems could inform real-time electricity scheduling and longer-term system planning (Rolnick et al. 2022). In the transportation sector, KIML can be useful in modeling demand and planning new infrastructure, freight routing and consolidation, and for electric vehicles (e.g., charge scheduling, congestion management, and improving vehicle-to-grid algorithms). Another potential avenue would be to improve and accelerate climate models (e.g., clouds and aerosols, ice sheets, and sea-level rise) via KIML research.

13.3.7 *Using KIML to Discover New Physics/Mathematics/Theory*

Currently, KIML methods are being developed to combine knowledge, while ML methods infer the QoI in a physical phenomenon (Brunton, Proctor, and Kutz 2015). Given small datasets and partial domain knowledge, some authors have shown it is possible to discover new governing equations for various physical systems (Brunton, Proctor, and Kutz 2015; Xu, Zhang, and Wang 2021). Others have shared that learning coordinate transformations can map complex nonlinear problems to their simpler linear representations in higher dimensions. This method, known as the Koopman operator approach (Budišić, Mohr, and Mezić 2012), is of particular interest in control theory (Korda and Mezić 2018). Combining these two powerful principles, coordinate transformation and equation discovery, provides a promising tool for automated scientific discovery (Champion et al. 2019). Despite encouraging preliminary results, more applied research must be conducted in this direction, especially dealing with the development of software tools that provide domain scientists with user-friendly and scalable solutions for a range of domain-specific tasks. The development of new algorithms, tools, and mathematical frameworks that systematically combine traditional physics with data-driven models will be needed to tackle this challenge.

13.4 Experimental, Data, and Modeling Opportunities

There are three areas where the opportunities exist for the broader community to both standardize and demonstrate the process of expert-driven combinations of multiple modeling paradigms: (1) combining scientific knowledge and data-driven knowledge (Rasp et al. 2020), (2) propagating uncertainty across multiple events, and (3) modeling across different time-spatial scales.

The ability to extrapolate beyond observed data is a natural expectation that often rises from end users. However, providing performance guarantees is always a challenge for these capabilities. Integrating scientific knowledge into data-driven models via expert supervision can provide an intermediate path to realize these goals. Using this method, the scientific knowledge captured in the form of equations or property constraints ensures that an ML model's prediction respects the essential rules of the underlying domain. Such complementary integration is ideal when predictive models purely based on scientific considerations can be further improved by exploiting additional data sources or settings where data-driven predictive models cannot provide high-performance guarantees when input variables are drawn from sparser regions of the domain of all input variables. Overall, the community identified two important objectives for this focus area:

1. Generation of synthetic data for training models that use the knowledge of physics to constrain model architectures and make the model learn its inductive biases.
2. Integration of different knowledge representations, such as knowledge graphs and Bayesian networks, to yield a new flexible class of "Uncertainty-aware Knowledge Graphs" that treat uncertainties as a first-order feature.

Integrating predictive models from different spatial and temporal scales is both an important need for the climate science community and a driver for science-inspired AI research. Also, results generated by models may not match observational data for a number of well-understood technical factors (Stephens et al. 2010). A good approach would be to follow paradigms, such as DOE's ModEx (U.S. Department of Energy Environmental System Science Program 2022), to identify where new data need to be collected through experiments or simulations, or models need to be improved for higher fidelity. Such identification is likely to be semi-automated, including a combination of automated data analysis with human supervision. However, opening up the process to human supervision requires model development to be aware of the cognitive biases of those driving the process. This amplifies the challenges already described for the extrapolation objective (summarized as follows):

- Development of methodologies to unify computational models, where each targets different temporal or spatial scales.
- Such unification needs to be cognizant of its potential technical challenges, such as the propagation of uncertainties across varied input domains, as well as the human cognitive biases of different communities developing individual models, including confirmation biases, prioritization for causation (Runge, Bathiany, et al. 2019), etc.

In terms of KIML, several opportunities for the AI community were identified during the workshop that involve advances in differentiable programming:

- ***Algorithmic.*** Development of more computationally efficient automatic differentiation (AD) tools and methods for higher-order derivatives. New solvers for KIML models inspired by classical constrained optimization theory and algorithms.
- ***Theoretical.*** Progress in combining learning theory with mathematical optimization, dynamical systems, optimal control, and physical/chemical theories.
- ***Software development.*** Development of new tools for user-friendly and scalable solutions of KIML models, including but not limited to neural differential equations, differentiable constrained optimization, constrained graph neural networks, and large-scale PINNs coupled with domain simulation models (Figure 13-2).
- ***Using NLP and computer vision*** for meta-analysis of the literature (e.g., Intergovernmental Panel on Climate Change [IPCC] reports).
- ***Computational graphs of models*** are especially useful for interpretability.
- ***Transfer learning*** can be interrogated to obtain insights not only on model interpretability but the data itself.
- ***Knowledge graph and logic rule-constrained differentiable programming*** can be used in online learning for PDE emulators as surrogates.

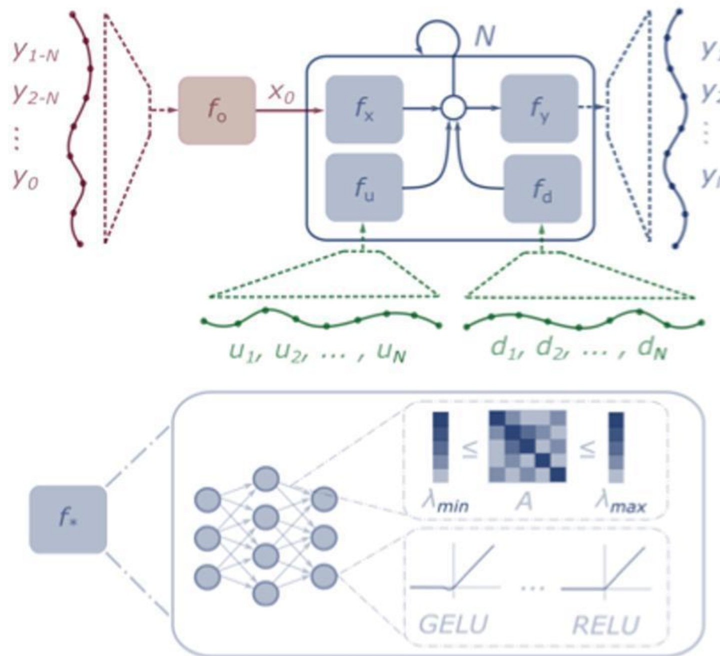


Figure 13-2. Generic structure of physics-inspired recurrent neural dynamics model architecture (Source: Reproduced and adapted from Drgoňa et al. 2021 under Creative Commons CC BY 4.0).

13.5 Research Priorities

Overall, the AI/climate research community recognizes two sets of goals aimed at tangible impacts in a five- and 10-year timeframe. The following research priorities will enable and accelerate the adoption of scientific knowledge in AI/ML methods:

13.5.1 Five-year Goals

Mid-term goals include:

- *Development of scalable, computationally efficient techniques for KIML*. Fundamental improvements in KIML algorithms for inverse, surrogate, reduced-order models, equation discovery, and parameter calibration will need to be prioritized for scaling and wider adoption of KIML techniques.
- *Development of inter/cross-disciplinary teams with expertise in multiple domains working together closely*. Tools and practices, such as agile programming, need to be adopted for effective integration of KIML techniques.
- *Tools for effective collaboration across different communities*. Tools for enabling collaboration and sharing of knowledge must be prioritized (e.g., ontologies for KIML).
- *Articulation of laws and constraints that ML models should be compliant to for a given domain.*

- *Development of KIML tools for easy accessibility for experts and non-experts*. Tools will include software libraries, trained and untrained models, standard datasets, and clearly written documentation of best practices.
- *Establishment of new cross-domain conferences, journals, workshops, and special issues focused on KIML and its maturation.*
- *Articulation of failure modes, gaps, and explainability of a model's behavior.*
- *A scalable, nonlinear causal interaction inference with UQ for climate data is a Grand Challenge, but it will accelerate scientific discovery and build additional knowledge graphs.* Current causal methods make assumptions on "no latent confounders," yet this can be improved with KIML, i.e., blend scientific knowledge with causal inference methods.
- *Fast, accurate surrogate models that blend AI and scientific knowledge for several PDEs of climate/Earth science relevance.*
- *Balanced, physics-based mechanistic models and data-driven ML models.*
- *Development of mathematical frameworks to integrate multiple spatial and temporal scales and account for heterogeneous phenomena within the same model.*

13.5.2 10-year Goals

Long-term goals include the following:

- AI must be able to do a "quick pass" to represent copious information/academic literature in a succinct manner. Climate literature has math and equations that are much more difficult to parse in NLP than pure text. New techniques are needed to understand these data. NLP and computer vision also can be used to digitize old journals and scientific records to feed these data to modern models.
  - Multidisciplinary teams, spanning theoretical ML/computer science to climate modelers, are necessary to frame a common "language" and terminology and make such collaborations easier.
  - Any absolute limitations—not only with ML methods, but also what can be accomplished with available climate/Earth data—must be identified. This is especially important for causal inference, where algorithms typically require large datasets to make an assessment.
  - For dimensionality reduction, can multi model data be reduced to a few key parameters and sensitivity by mapping everything to low-dimensional space? Transfer learning does this in an implicit manner, but it requires interpretation.
- Knowledge graph and logic rule-constrained differentiable programming have bottlenecks that can prevent deployment to scale, such as computational cost of a forward pass in a climate model, differentiability, and smoothness constraints. These need to be addressed.
  - Building models from climate data that generalize across multiple spatial-temporal scales with awareness of factors such as cognitive and inductive biases and uncertainty propagation.
  - AI/ML methods for connecting the climate modeling community with the observational community will be standardized beyond today's exploratory research efforts.

13.6 References

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, et al. 2021. "On the Opportunities and Risks of Foundation Models," August. https://doi.org/10.48550/arxiv.2108.07258.

Brouwer, Edward de, Jaak Simm, Adam Arany, and Yves Moreau. 2019. "GRU-ODE-Bayes: Continuous Modeling of Sporadically-Observed Time Series," May. https://doi.org/10.48550/arxiv.1905.12374.

Brunton, Steven L., Joshua L. Proctor, and J. Nathan Kutz. 2015. "Discovering Governing Equations from Data: Sparse Identification of Nonlinear Dynamical Systems," September. https://doi.org/10.1073/pnas.1517384113.

Budišić, Marko, Ryan Mohr, and Igor Mezić. 2012. "Applied Koopmanism." *Chaos* 22 (4): 047510. https://doi.org/10.1063/1.4772195.

Cachay, Salva Rühling, Venkatesh Ramesh, Jason N. S. Cole, Howard Barker, and David Rolnick. 2021. "ClimART: A Benchmark Dataset for Emulating Atmospheric Radiative Transfer in Weather and Climate Models," November. https://doi.org/10.48550/arxiv.2111.14671.

Champion, Kathleen, Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. 2019. "Data-Driven Discovery of Coordinates and Governing Equations." *Proceedings of the National Academy of Sciences of the United States of America* 116 (45): 22445–51. https://doi.org/10.1073/PNAS.1906995116/SUPPL_FILE/PNAS.1906995116.SAPP.PDF.

Davis, Steven J., Nathan S. Lewis, Matthew Shaner, Sonia Aggarwal, Doug Arent, Inês L. Azevedo, Sally M. Benson, et al. 2018. "Net-Zero Emissions Energy Systems." *Science* 360 (6396): eaas9793. https://doi.org/10.1126/science.aas9793.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," October. https://doi.org/10.48550/arxiv.1810.04805.

Donti, Priya L., David Rolnick, and J. Zico Kolter. 2021. "DC3: A Learning Method for Optimization with Hard Constraints." In *International Conference on Learning Representations*. https://openreview.net/forum?id=V1ZHVxJ6dSS.

Drgoňa, Ján, Aaron R. Tuor, Vikas Chandan, and Draguna L. Vrabie. 2021. "Physics-Constrained Deep Learning of Multi-Zone Building Thermal Dynamics." *Energy and Buildings* 243 (July): 110992. https://doi.org/10.1016/J.ENBUILD.2021.110992.

Drgona, Jan, Aaron Tuor, and Draguna Vrabie. 2020. "Learning Constrained Adaptive Differentiable Predictive Control Policies With Guarantees," April. https://doi.org/10.48550/arxiv.2004.11184.

Ghahramani, Zoubin. 1998. "Learning Dynamic Bayesian Networks." In *Adaptive Processing of Sequences and Data Structures: International Summer School on Neural Networks "E.R. Caianiello" Vietri Sul Mare, Salerno, Italy September 6–13, 1997 Tutorial Lectures*, edited by C Lee Giles and Marco Gori, 168–97. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/BFb0053999.

Hu, Qinghua, Rujia Zhang, and Yucan Zhou. 2016. "Transfer Learning for Short-Term Wind Speed Prediction with Deep Neural Networks." *Renewable Energy* 85 (January): 83–95. https://doi.org/10.1016/J.RENENE.2015.06.034.

Huang, Hao, Chenxiao Xu, and Shinjae Yoo. 2019. "Bi-Directional Causal Graph Learning through Weight-Sharing and Low-Rank Neural Network." In *2019 IEEE International Conference on Data Mining (ICDM)*, 319–28. https://doi.org/10.1109/ICDM.2019.00042.

Jaegle, Andrew, Sebastian Borgeaud, Jean-Baptiste Alayrac, Carl Doersch, Catalin Ionescu, David Ding, Skanda Koppula, et al. 2021. "Perceiver IO: A General Architecture for Structured Inputs & Outputs," July. https://doi.org/10.48550/arxiv.2107.14795.

Kadow, Christopher, David Matthew Hall, and Uwe Ulbrich. 2020. "Artificial Intelligence Reconstructs Missing Climate Information." *Nature Geoscience 2020 13:6* 13 (6): 408–13. https://doi.org/10.1038/s41561-020-0582-5.

Karniadakis, George Em, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. 2021. "Physics-Informed Machine Learning." *Nature Reviews Physics 2021 3:6* 3 (6): 422–40. https://doi.org/10.1038/s42254-021-00314-5.

Kashinath, K., M. Mustafa, A. Albert, J.-L. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, et al. 2021. "Physics-Informed Machine Learning: Case Studies for Weather and Climate Modelling." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379 (2194): 20200093. https://doi.org/10.1098/rsta.2020.0093.

Korda, Milan, and Igor Mezić. 2018. "Linear Predictors for Nonlinear Dynamical Systems: Koopman Operator Meets Model Predictive Control." *Automatica* 93 (July): 149–60. https://doi.org/10.1016/J.AUTOMATICA.2018.03.046.

Kreider, J. F., D. E. Claridge, P. Curtiss, R. Dodier, J. S. Haberl, and M. Krarti. 1995. "Building Energy Use Prediction and System Identification Using Recurrent Neural Networks." *Journal of Solar Energy Engineering* 117 (3): 161–66. https://doi.org/10.1115/1.2847757.

Li, Xiaoyu, and Francesco Orabona. 2018. "On the Convergence of Stochastic Gradient Descent with Adaptive Stepsizes," May. https://doi.org/10.48550/arxiv.1805.08114.

Li, Zongyi, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. 2020. "Fourier Neural Operator for Parametric Partial Differential Equations," October. https://doi.org/10.48550/arxiv.2010.08895.

Lobier, Muriel, Felix Siebenhühner, Satu Palva, and J. Matias Palva. 2014. "Phase Transfer Entropy: A Novel Phase-Based Measure for Directed Connectivity in Networks Coupled by Oscillatory Interactions." *NeuroImage* 85 (January): 853–72. https://doi.org/10.1016/J.NEUROIMAGE.2013.08.056.

Nichol, Jeffrey J., Matthew Peterson, George M. Fricke, and Kara Peterson. 2021. "Learning Why: Data-Driven Causal Evaluations of Climate Models." In *ICML 2021 Workshop on Tackling Climate Change with Machine Learning*. https://www.climatechange.ai/papers/icml2021/80.

Notarangelo, Nicla, Kohin Hirano, Raffaele Albano, and Aurelia Sole. 2021. "Transfer Learning with Convolutional Neural Networks for Rainfall Detection in Single Images." *Water* 13 (5): 588. https://doi.org/10.3390/w13050588.

Paullada, Amandalynne, Inioluwa Deborah Raji, Emily M. Bender, Emily Denton, and Alex Hanna. 2021. "Data and Its (Dis)Contents: A Survey of Dataset Development and Use in Machine Learning Research." *Patterns* 2 (11): 100336. https://doi.org/10.1016/J.PATTER.2021.100336.

Perera, Kasun S., Zeyar Aung, and Wei Lee Woon. 2014. "Machine Learning Techniques for Supporting Renewable Energy Generation and Integration: A Survey." In *Data Analytics for Renewable Energy Integration*, edited by Wei Lee Woon, Zeyar Aung, and Stuart Madnick, 81–96. Cham: Springer International Publishing. https://doi.org/https://doi.org/10.1007/978-3-319-13290-7_7.

Raissi, M., P. Perdikaris, and G. E. Karniadakis. 2019. "Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations." *Journal of Computational Physics* 378: 686–707. https://doi.org/https://doi.org/10.1016/j.jcp.2018.10.045.

Ramachandra, Vikas. 2019. "Causal Inference for Climate Change Events from Satellite Image Time Series Using Computer Vision and Deep Learning," October. https://doi.org/10.48550/arxiv.1910.11492.

Ramchurn, Sarvapali D., Perukrishnen Vytelingum, Alex Rogers, and Nicholas R. Jennings. 2012. "Putting the 'smarts' into the Smart Grid." *Communications of the ACM* 55 (4): 86–97. https://doi.org/10.1145/2133806.2133825.

Rasp, Stephan. 2021. "Weather and Climate Datasets for AI Research - Pangeo ML Datasets." 2021. http://mldata.pangeo.io/index.html.

Rasp, Stephan, Peter D. Dueben, Sebastian Scher, Jonathan A. Weyn, Soukayna Mouatadid, and Nils Thuerey. 2020. "WeatherBench: A Benchmark Data Set for Data-Driven Weather Forecasting." *Journal of Advances in Modeling Earth Systems* 12 (11): e2020MS002203. https://doi.org/https://doi.org/10.1029/2020MS002203.

Rolnick, David, Priya L. Donti, Lynn H. Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, et al. 2022. "Tackling Climate Change with Machine Learning." *ACM Computing Surveys (CSUR)* 55 (2): 96. https://doi.org/10.1145/3485128.

Rueden, Laura von, Sebastian Mayer, Katharina Beckh, Bogdan Georgiev, Sven Giesselbach, Raoul Heese, Birgit Kirsch, et al. 2021. "Informed Machine Learning - A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems." *IEEE Transactions on Knowledge and Data Engineering*, 1. https://doi.org/10.1109/TKDE.2021.3079836.

Runge, Jakob, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, et al. 2019. "Inferring Causation from Time Series in Earth System Sciences." *Nature Communications* 10 (1): 2553. https://doi.org/10.1038/s41467-019-10105-3.

Runge, Jakob, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. 2019. "Detecting and Quantifying Causal Associations in Large Nonlinear Time Series Datasets." *Science Advances* 5 (11): 4996–5023. https://doi.org/10.1126/SCIADV.AAU4996/SUPPL_FILE/AAU4996_SM.PDF.

Sanchez-Gonzalez, Alvaro, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter W. Battaglia. 2020. "Learning to Simulate Complex Physics with Graph Networks," February. https://doi.org/10.48550/arxiv.2002.09405.

Singh, Amanpreet, Ronghang Hu, Vedanuj Goswami, Guillaume Couairon, Wojciech Galuba, Marcus Rohrbach, and Douwe Kiela. 2021. "FLAVA: A Foundational Language And Vision Alignment Model," December. https://doi.org/10.48550/arxiv.2112.04482.

Stephens, Graeme L., Tristan L'Ecuyer, Richard Forbes, Andrew Gettelmen, Jean Christophe Golaz, Alejandro Bodas-Salcedo, Kentaroh Suzuki, Philip Gabriel, and John Haynes. 2010. "Dreary State of Precipitation in Global Models." *Journal of Geophysical Research: Atmospheres* 115 (D24): 24211. https://doi.org/10.1029/2010JD014532.

Strubell, Emma, Ananya Ganesh, and Andrew McCallum. 2019. "Energy and Policy Considerations for Deep Learning in NLP," June. https://doi.org/10.48550/arxiv.1906.02243.

U.S. Department of Energy Environmental System Science Program. 2022. "ModEx Approach." 2022. https://ess.science.energy.gov/modex/.

Varando, Gherardo, Miguel-Ángel Fernández-Torres, and Gustau Camps-Valls. 2021. "Learning Granger Causal Feature Representations." In *ICML 2021 Workshop on Tackling Climate Change with Machine Learning*. https://www.climatechange.ai/papers/icml2021/34.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Ł. Lukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need." In *Advances in Neural Information Processing Systems*, edited by I Guyon, U von Luxburg, S Bengio, H Wallach, R Fergus, S Vishwanathan, and R Garnett. Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Willard, Jared, Xiaowei Jia, Shaoming Xu, Michael Steinbach, and Vipin Kumar. 2020. "Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems," March. https://doi.org/10.48550/arxiv.2003.04919.

Xu, Hao, Dongxiao Zhang, and Nanzhe Wang. 2021. "Deep-Learning Based Discovery of Partial Differential Equations in Integral Form from Sparse and Noisy Data." *Journal of Computational Physics* 445 (November): 110592. https://doi.org/10.1016/J.JCP.2021.110592.

Yao, Shun, Shinjae Yoo, and Dantong Yu. 2015. "Prior Knowledge Driven Granger Causality Analysis on Gene Regulatory Network Discovery." *BMC Bioinformatics* 16 (1): 1–18. https://doi.org/10.1186/S12859-015-0710-1/FIGURES/13.

Zbontar, Jure, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. 2021. “Barlow Twins: Self-Supervised Learning via Redundancy Reduction,” March. https://doi.org/10.48550/arxiv.2103.03230.

Zhang, T., W. Lin, Z. Wu, Y. Liu, and A.~M. Vogelmann. 2020. “Causality Linking ENSO with North American Rainfall: Inference by Data-Driven Causal Discovery Methods and Physical Evaluation.” In *AGU Fall Meeting Abstracts*, 2020:A043-0003. https://ui.adsabs.harvard.edu/abs/2020AGUFMA043.0003Z/abstract.

14 Knowledge Discovery and Statistical Learning

Authors: Xingyuan Chen (PNNL), Richard Tran Mills (Argonne), Juan M. Restrepo (ORNL), Michael Wehner (LBNL), Julie Bessac (Argonne), Shashank Konduri (NASA-Goddard), Youzuo Lin (LANL), Emil Constantinescu (Argonne), Forrest M. Hoffman (ORNL), Muralikrishnan Gopalakrishnan Meena (ORNL), Chandrika Kamath (LLNL), Juliane Mueller (LBNL), Zachary Langford (ORNL), John Jakeman (SNL), Elias Massoud (UC-Berkeley)

14.1 Introduction

Statistical and classical machine learning methods originating in various fields (e.g., geostatistics, data mining, statistical learning, pattern recognition, etc.) have a long history in the Earth sciences for regression, supervised and unsupervised classification, model reduction, density estimation, and other applications. Although researchers have had great success recently using deep neural networks (DNNs) for supervised and reinforcement learning, classical methods remain relevant because they can be less computationally expensive, more interpretable, and better suited to small training datasets than DNNs. Furthermore, as available datasets continue to grow in size and richness, unsupervised methods that enable automatic knowledge discovery in these large datasets are increasingly important. The paradigm of Bayesian statistics provides a unifying framework for knowledge discovery and statistical learning in the context of Earth system prediction. It enables critical work in multisensor data fusion, uncertainty quantification, data assimilation, and scale-bridging, and it has taken on special relevance in the DNN era: it is central to quantifying epistemic uncertainty in large DNNs while, in turn, DNNs aid statistical inference by enabling fast approximation of posterior distributions.

14.2 Grand Challenges

The challenges we identified for knowledge discovery and statistical learning in Earth system prediction can be categorized into (1) issues of data acquisition, assimilation, fusion, sparsity, and reduction; (2) inherent challenges in representing multiscale processes and model-data integration (see Penny et al. 2019); (3) quantification of the uncertainty propagated from various sources to the end prediction through increasingly more complex modeling structure, as well as in machine learning assisted components; and (4) unsupervised learning for automated knowledge discovery–including estimation of strongly non-Gaussian distributions and discovery of causal relationships and governing equations–from large datasets.

14.2.1 Data Acquisition, Assimilation, Fusion, Sparsity, and Reduction

There are needs to develop new representations, augmentation, and standardization of data from various sources at various scales and resolutions in order to improve the extraction of knowledge and their use in standard methodologies in AI/ML by providing benchmark datasets. Many fields in the Earth science discipline, such as soil sciences, generate many sparse and isolated datasets that would benefit from being merged and enriched from one another in order to enable large-scale, standardized analysis. Instrument and simulation data are typically "skinny" and lack information and context resulting from the acquisition. In particular, the metadata are generally not informative enough to allow for a straight harmonization across multiple sources. Provenance tracking and tightly related uncertainty or confidence intervals are likely one of the top challenges to overcome, especially with datasets associated with publications.

The data harmonization process is performed manually and is labor intensive. It is also error prone and subject to investigator biases. Physics-informed neural networks (PINNs) and other ML strategies like hierarchical temporal memory have been used to infer data nonstationarity and guard against out-of-sample errors; however, the lack of standards for data collection is a significant challenge. A secondary aspect is related to the extrapolation of certain datasets (e.g., resulting from data assimilation and assuming climate stationarity), which pose significant challenges for harmonization of various simulation and instrument data sources and products. For nonstationary dynamic problems, a major challenge is in finding practical data assimilation methods that can handle the nonlinear/non-Gaussian estimation problem (see Majda and Harlim 2012; Särkkä 2013; van Leeuwen, Cheng, and Reich 2015). Reanalysis data have been created in that regard and successfully used in atmospheric applications; however, the concept is less developed in other fields.

Interpolation and extrapolation routines with limited data availability for highly nonlinear and multiscale problems (e.g., turbulent flows in the atmosphere and ocean) are a common challenge, including choice of variable being used in models (e.g., appropriate variable needed in flux-based models), geographical locations, limited time, and bottlenecks and biases from other geophysical processes (e.g., cloud cover hindering satellite data). Tying sampling techniques with geometrical attributes to take advantage of spatial auto-correlations can be helpful. The advancement of sparse linear algebraic techniques from dynamical systems theory can be used for optimizing sparse locations. Limited data also pose a challenge for reduced-order modeling. Choosing the appropriate dimensionality can be challenging in this scenario.

Lack of quality data can also arise when modeling processes which can only be measured indirectly, like using radiance for capturing ecology physics. Frameworks for effectively capturing such processes involve clearly identifying the processes, which ones are missed, and the scales at which their effects are relevant. When integrating instrument data with data from other sources like simulations to create reanalysis data, the provenance is typically lost; moreover, there is no confidence interval associated with the final product, which hampers the use of data product in projections, model development, and other analyses.

14.2.2 Multiscale Modeling and Model-data Integration

Subgrid-scale variability (unresolved processes at scales finer than the discretization resolution) modeling and parameterization is a significant challenge in representing multiscale processes. Subgrid-scale issues are naturally under-determined and require adequate treatment such as regularization and a stochastic treatment when scales are not separated. Subgrid-scale parameterizations often introduce new types of errors and uncertainty that are hard to characterize and quantify.

As high-fidelity and high-resolution simulations are computationally prohibitive, there is a need to develop surrogate models for fine-scale processes to alleviate these computational costs and in particular in multiscale applications. Associated challenges arise from accessing fine-scale or high-fidelity training data and in the embedding of fine-scale surrogate models within other components (this is linked to some stability issues discussed in the Hybrid Modeling session, chapter 16). Last but not least, there are workflow challenges with high-volume data streams or high-dimensional models.

14.2.3 Uncertainty Quantification and Propagation

Some grand challenges in statistical learning (Bayesian methods, statistical inference, uncertainty quantification, etc.) for AI applications in Earth system models (ESMs) consist of dealing with uncertainty in deep learning methods and data products for ESMs. For example, satellite-based, radar-based, and model-derived precipitation data products often do not agree (Kim et al. 2020), and methods are needed that can incorporate this data uncertainty into ESMs. Deep learning techniques are inherently over-parameterized, making them great interpolators; however, this creates large parameter spaces that are challenging for their associated uncertainty quantification, and this limits their generalizability. We must accurately quantify uncertainty in coupled ESM predictions that are computationally expensive to generate and are subject to a large number of uncertainties.

14.2.4 Unsupervised Learning

Grand challenges for unsupervised learning center around moving beyond assumptions of linearity, statistical stationarity, and Gaussianity, as well as automatic discovery or improvement of model equations or causal relationships of underlying processes. Effective techniques for working with very large datasets – many of which combine data with different sources, modalities, and spatiotemporal distribution – are also an important challenge.

Explicit data reduction techniques can be very helpful in scenarios with insufficient high-quality data. The variability of the ML/DL models in the reduced space is a great challenge when using reduced representation of the data (or system). One solution is to implicitly reduce the dimension of the data rather than using explicit methods. This involves coupling the data reduction technique with the ML model. The low-dimensional features in a system are learned in the "middle" layers of an ML model, and thus learning needs to be done in a reduced or latent space where variability would be limited. Nonetheless, the challenge for such implicit data reduction methods is the requirement of high-quality data. Moreover, the structure of the data considering the geometrical attributes is also important. The issue is with those geometrical attributes that cannot be transferable using a change of basis, which lie on a union of subspaces. Possible solutions to effectively capture the structure of data can be sparse representation using compressed sensing, use of a manifold representation rather than various subspaces, and tractable optimization techniques with better relaxation techniques. The latter would add an additional challenge of the computational complexity trade-off.

14.3 State-of-the-Science

Unsupervised methods have a long history of application in the Earth sciences, with linear methods such as principal components analysis (PCA) or empirical orthogonal functions (EOF) (Lorenz 1956) commonly applied for dimensionality reduction and trend identification, and various clustering methods used for applications such as quantitative identification of ecoregions (Hoffman et al., 2013; see an example in Figure 14-1 from Langford et al. [2019] and watershed zonation in Wainwright et al. [2022]). In some cases, these methods have been successfully scaled to the solution of "big data" problems on large parallel computing resources (Kumar et al. 2011; Mills et al. 2013; Sreepathi et al. 2017; Mills et al. 2018), although most existing tools are suited for use only on single workstations.

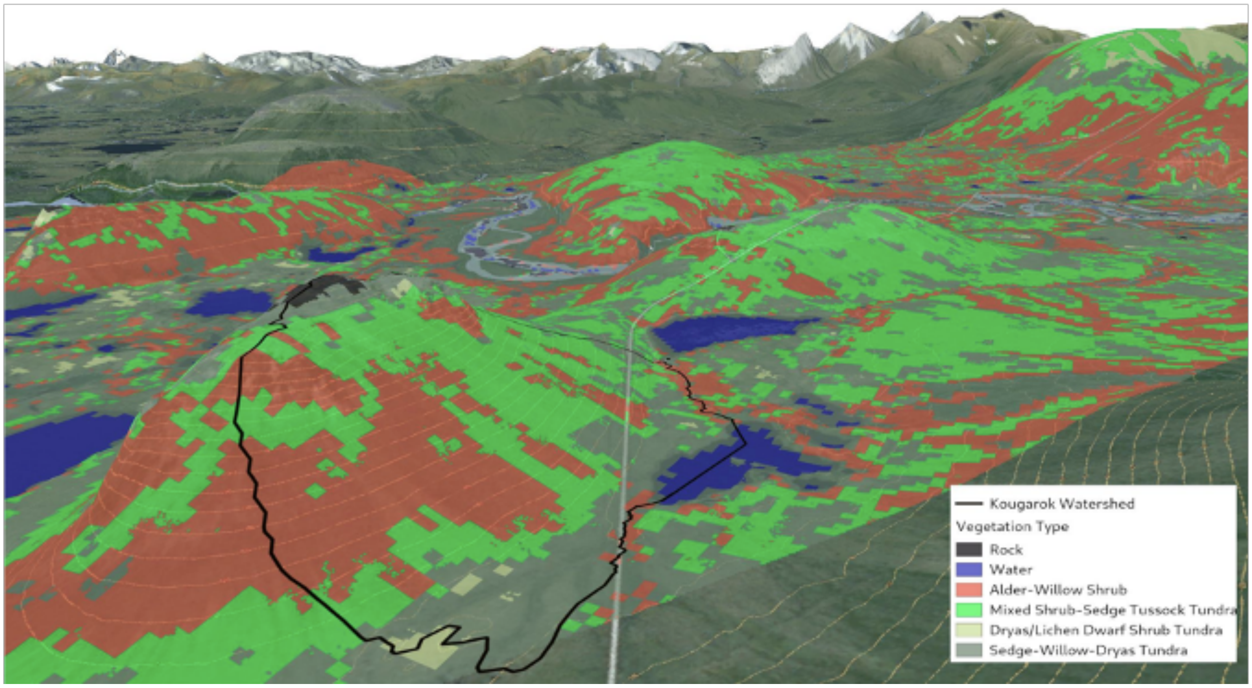


Figure 14-1. Map produced using a combination of linear dimension reduction (principal components analysis), unsupervised classification (parallel k-means clustering), and convolutional neural networks applied to data fused from multisensor sources (Source: Reproduced from Langford et al. 2019 under Creative Commons CC BY 4.0).

Datasets spanning a huge range of variables are being generated at unprecedented rates and at global scales, largely driven by remote sensing, models, and computational power. This includes instrument measurements, numerical simulations using measurements, combining measurements with numerical simulations into products, or combining datasets. However, there is no standard and metadata information that can be easily used to harmonize datasets in general.

Time-dependent data assimilation is now an established forecasting and analysis tool in weather, and with increasing capabilities, in climate. In applications, data assimilation has also been used to estimate parameters (see Lunderman et al. 2020), it is used to incorporate complex heterogeneous physics such as vegetation and hydrological elements, and it is thought to be relevant to working out seamless predictions of weather/climate. In data assimilation models and data, their inherent errors are taken into account to construct Bayesian distributions wherein the likelihood is informed by data and the prior by models. Note that the methods of data assimilation for spatially dependent problems or stationary problems are usually Gauss-Markov estimation. In applications, especially those that combine dynamics at different scales, research into ways to remove biases as well as ways to couple the scales in a Bayesian framework have received a great deal of attention (see Penny et al. 2019; Berry and Harlim 2017; Cych, Morzfeld, and Tauxe 2021). How to deal with parametrizations of phenomena that are not well understood, and the formulation of statistical parametrizations of epistemic errors, are also very current topics of research. A great number of data assimilation methods are variance-minimizing estimators; however, it is becoming clear that other estimators may be useful (e.g., Rosenthal et al. 2017; Restrepo 2017).

Modern reanalysis products such as ERA5, MERRA2, and NCEP-CFSR, produced by computationally intensive data assimilation of observations, are among the most critical resources available to climate change scientists. Often considered as a proxy for actual observations, reanalyses provide highly constrained values of unobserved fields, whether that be in poorly observed regions and even for unobservable quantities. These highly cited datasets serve many purposes, including in model evaluation, detection and attribution of trends, and understanding climate variability, among others. Of particular interest to the DOE are understanding climate change trends. The three state-of-the-art, high-resolution products mentioned above are useful in that regard only in the satellite era (1979–present), as the sudden increase in observations with the introduction of satellites tends to introduce discontinuities. The DOE-supported “20th Century Reanalysis” project (https://psl.noaa.gov/data/20thC_Rean/) circumvents this problem by assimilating only synoptic pressure, sea surface temperature, and sea ice concentrations to produce a homogenous reanalysis product from the early 19th century to nearly present day, but is at a lower resolution than the state of the art due to the much longer period of interest and the high computational costs. How can machine learning help? If a fast proxy AI-based reanalysis tool, trained on ERA5 (and others) using observed synoptic pressure, sea surface temperature, and sea ice concentrations, could be developed, a higher-resolution 20th century reanalysis could be performed aiding the analysis of changes in extreme weather events and other applications that require high spatial resolution.

Current generation atmospheric dynamical cores in ESMs operate at resolutions too coarse to capture fine-scale cloud processes such as convection, and therefore must represent these subgrid-scale processes using statistical parameterizations, which contribute substantially to uncertainty in climate projections. Although high-resolution, physics-based, cloud-resolving models exist, the high computational cost of running multiyear simulations makes this an infeasible option for the foreseeable future (Yuval and O’Gorman 2020). Recently, subgrid modeling strategies employing ML and deep learning methods have been proposed to learn subgrid statistics and predict the unresolved part of quantities based on resolved solutions (Bode et al. 2019; Fukami, Fukagata, and Taira 2019; Rasp, Pritchard, and Gentine 2018). Such data-driven approaches have been shown to be fast and accurate and to greatly enhance the spatial resolution with as few as 50 training data shots (Fukami, Fukagata, and Taira 2019). Simulations generated using such ML methods have been shown to more or less obey the law of energy conservation and reproduce both the mean and variance of the climate system. Machine learning is presently being considered as a means to address the curse of dimensionality via surrogates and by replacing computationally expensive aspects of the dynamics. In addition to surrogates, other reduced-order representation include manifold learning for representing vector fields and an operator acting on vector fields.

Model benchmarking is the process of critically evaluating the performance of ESMs by testing the fidelity of their outputs against a set of defined standards (Luo et al. 2012). ESMs typically simulate several biophysical and biogeochemical variables over large spatiotemporal scales, and obtaining access to gridded data products at such diverse scales can be quite challenging. However, the availability of global-scale remote sensing data spanning multiple decades, as well as data from eddy covariance towers from observation networks such as FLUXNET, open up exciting avenues for creating new benchmark data products. One of the first examples of creating a global gridded benchmark product uses ML methods such as tree-based approaches and ANNs to infer relationships between environmental variables measured by flux towers at the site-level, which were extrapolated and applied at the global-scale using global-grids of satellite-derived vegetation indices and climate variables (Beer et al. 2010).

Existing sensitivity analysis and uncertainty quantification efforts have focused on regional case studies, studies involving a subset of model components, or coarse fully coupled models (Tezaur et al. 2021; Urrego-Blanco et al. 2016; Rasch et al. 2019; Isaac et al. 2015). These restrictions in scope were adopted to reduce the cost of running models, which were needed for Monte Carlo estimation of statistics or sensitivity indices, and/or to reduce the number of uncertain variables needed for practical use of surrogate approaches. Current methods use deep learning (DL) in a Bayesian framework in order to provide uncertainty estimates of the learned model. Bayesian DL methods have also been applied to model-data integration for global parameter optimization to improve the model performance and computational complexity of ESMs (Lu, Liu, and Ricciuto 2019). New approaches will need to investigate state-of-the-art DL approaches applied to time series datasets (e.g., LSTMs) and ways to incorporate DL models that address uncertainties in data products. Deep Bayesian active learning could provide a framework to address uncertainties for efficient data acquisition by selecting points that bias data acquisition toward regions with overlapping support to maximize sample efficiency (Jesson et al. 2021). Normalizing flows (Kobyzev, Prince, and Brubaker 2019) and information-preserving dimensionality reduction are being explored for dealing with non-Gaussian distributions. Moreover, improvements in Gaussian process (GP) model scalability are underway and enable their application to larger datasets (https://gpytorch.ai).

The use of self-supervised models has been prevalent to address some of the implicit data/system reduction challenges. Recently, reinforcement learning has been in the forefront for identifying model/data biases. Along with techniques like Q-learning, these techniques can be used for exploitation of currently available data to aid further data exploration. A great effort has been focused on standardizing the right notion of validation for datasets and reduced order models. This involves identifying the metrics used for validation, making the process independent of ML/DL frameworks, and setting standards for multidisciplinary datasets and models. Furthermore, use of edge computing to aid the data workflow of incoming data from experimental and observational sources has been actively pursued.

14.4 Experimental, Data, and Modeling Opportunities

Experimental, data, and modeling opportunities include improving the quality and ease of the use of data and observations to make ML models informative, workflow tools that can handle larger problems, exploitation of machine learning for data reduction, and ML methods that address physics and computation across scales (see Boukabara, et al. 2021).

14.4.1 Employing ML Techniques to Ease the Use of Data

Much of the rapidly growing body of Earth system science datasets—from sources such as NASA remote sensing products and GCM simulation ensembles—is relatively unexplored, and there is an opportunity to employ clustering and other unsupervised techniques at large scale to effectively sift through these massive and often high-dimensional datasets to discover meaningful patterns and statistical relationships. Furthermore, there is an opportunity to bring some of the recent success with deep neural networks (DNNs) by applying semi-supervised techniques to what are often label-poor datasets.

14.4.2 Utilizing the Model-Experiment (ModEx) Approach

Opportunities exist in developing AI/ML techniques to integrate multiscale, multisource data/observations into physics-based models, using multiscale modeling to inform experimental design and field data collection, and generalizing the methods and tools from a limited set of testbeds to much broader geographical regions. Specific opportunities exist in dissecting the intricate features in the data from noise and leveraging advancements in online data curation from edge devices to aid smart data collection. Opportunities exist in recreating Earth system events, developing probabilistic approximations for validation of reduced order models, and creating robust benchmark datasets, which are updated and improved periodically (e.g., using new data releases, reanalysis data). Identifying outliers in datasets is another big opportunity for effective data and model reduction. Increased collaborations across disciplines under a shift in research culture toward the FAIR-ICON principles are essential for the community to succeed in ModEx.

14.4.3 Development of Surrogate Models

Multifidelity (MF) statistical estimation (Gorodetsky et al. 2020; Peherstorfer, Willcox, and Gunzburger 2018; Giles 2015) and surrogate modeling (Kennedy and O'Hagan 2000; Narayan, Gittelson, and Xiu 2014; Jakeman et al. 2020) offer a powerful balance between cost and accuracy. These methods combine an ensemble of models with varying cost and accuracy of data to produce accurate surrogates or statistics at potentially orders of magnitude smaller cost. The existing DOE simulations with cloud-resolving capacities (~3–4km) (SCREAM (Caldwell et al. 2021) present unique opportunities to develop surrogate models.

14.4.4 Development of ML Techniques to Improve Estimates of Model Predictive Uncertainty

Current state-of-the-art approaches for uncertainty quantification use Bayesian frameworks (e.g., Bayesian networks) to weight ESMs by accounting for model dependencies and changing biases (Sunyer et al. 2014). Future research could focus on ways to incorporate Bayesian deep learning approaches into ESMs for incorporating uncertainties in model predictions (Geer 2021). New UQ methods that do not make assumptions about Gaussian distributions must be developed in order to truthfully capture the data distribution, thus improving the prediction ability of outliers and extremes. Another important need is the access to data uncertainty, which is traditionally not provided with the data products, and so data are commonly used at face value. Close collaborations with data-collecting agencies is needed to provide these uncertainty estimates, thus enabling the improvement of model predictive uncertainty.

14.4.5 Addressing Scalability Issues in Software Libraries

There is a need for software libraries that can scale on large clusters for ESM integration. GPs have been used for model emulation due to their simple formulation and robust uncertainty estimates (Watson-Parris et al. 2021). However, their O(n^3) computation and O(n^2) storage requirements limit GPs to small datasets. GPyTorch is an efficient and general approach to GP inference based on Blackbox Matrix-Matrix multiplication (BBMM) that addresses scalability issues with Cholesky factorization required by GPs (Gardner et al. 2018).

14.5 Research Priorities

One of the critical priorities identified is the development of AI-ready benchmark data for development and testing of different ML methods. Developing a standardized framework for evaluating and validating ML fidelity is another area of priority. Moreover, providing reference implementations of the various knowledge discovery methods for specific Earth and environmental system science applications would be useful for the domain science experts to use and further develop these techniques.

A barrier to exploring environmental datasets with unsupervised learning tools and to developing suitable benchmarks for various ML tasks is that it is difficult to assemble and properly process datasets into suitable form. Google Earth Engine represents a good attempt at making remote sensing data more accessible, but no corresponding platform exists for ESM outputs, and processing is still a challenge because it often requires significant expert knowledge. If we can provide platforms that make it easy for both Earth system scientists and AI/ML researchers to explore the growing body of ESS data with AI/ML techniques, there is potential to enable many new lines of research in both finding undiscovered patterns and relationships in ESS datasets and in the development of new AI/ML techniques suitable for use with large and high-dimensional geo-spatiotemporal data. Enabling the involvement of AI/ML researchers is important, since the development of approaches that can scale to the large size and dimensionality of these datasets is needed. There is also a clear need for developing data collection standards and metadata representation that can be easily used by ML algorithms—including the data provenance information that is critical for establishing confidence intervals—as well as ML methods that can automatically ingest datasets with suitable metadata information. To harvest/digitize historic datasets, we should explore the potential of leveraging NLP algorithms that can extract not only the data points but also ancillary metadata.

Even when barriers to assembling large ESM datasets have been overcome, there is still the problem that, for many applications of interest, labeled training data may be scarce, and labeling by human experts may be an impractical task. In such cases, semi-supervised or active learning techniques (closely related to optimal experimental design in statistics) may provide a practical path forward. Such approaches are relatively unexplored in the geosciences and should be prioritized in light of our current data-rich, label-poor situation era.

As ML or reduced-order models are developed to reduce the cost of running components of ESMs, we must develop algorithms that explicitly quantify the impact of all sources of error and uncertainties including those arising from unknown model parameters, noisy data, missing physics (Harlim, et al. 2021), discretization errors in the solution of model equations, and approximations in reduced-order models or ML models. We then must invest in algorithms that can determine the fidelity needed for each component to balance these sources of error and incorporate experimental data needed to reduce uncertainty. Special attention needs to be paid to the aspect of error and uncertainty quantification when using emulators/surrogate models: what are the errors and uncertainty associated with the emulators and how do they propagate across scales?

Applying statistical learning frameworks to ESMs is not trivial because of the challenges imposed by the computational expense of running large ensembles and the nonlinear interactions of climate system processes (Qian et al. 2016). For example, changing model parameters that improve one model component can have major consequences on another model component (Qian et al. 2016). Next steps to address challenges in statistical learning for AI applications in ESMs include understanding relationships between UQ in data products and model behavior, performing sensitivity analyses that remove unimportant variables, generating fast UQ methods that can scale with ESMs, using DL approximation methods to help accelerate UQ, and integrating DL UQ methods with ESMs. Accounting for uncertainty in the data and the models will allow us to achieve robust and trustworthy predictions. The sensitivity of outcomes to parameters and to dynamic conditions is generally highly dependent on context and the nature of the problem itself for their proper interpretation. ML approaches to tackling the sensitive dependence of outcomes to parameters of typical models for environmental/climate dynamics (see Restrepo and Venkataramani 2016) have yet to be formulated, but they might be useful in circumventing the completely ad-hoc nature of a stochastic approach to the same. There is a need to develop theoretical foundations around the use of stochastic parameterizations (e.g., how to use a physics-based model with stochastic components) through samples or summary statistics. We must also develop statistical and probabilistic models to quantify the intrinsic uncertainties associated with scale-bridging problems.

14.6 Short Term (<5 years), 5-year, and 10-year Goals

Some of the identified **short-term goals (< 5 years)** are:

- *The construction and run of hybrid models within the coming 5 years, and their integration to ESMs*. In the near- and medium-term, developing the modular approaches of these hybrid models is desirable where physics-based models can be easily swap with AI/ML components; and in the long-term, building modularity and transferability across ESMs should be considered.
- *The establishment of Earth system science benchmark datasets that can be used in developing and improving machine learning approaches.* We are thinking of something analogous to the ImageNet database that has enabled rapid advances in the use of AI/ML techniques for visual object recognition but for Earth system science applications. Since ESS datasets are rapidly evolving, it is important that any platform/framework developed for building curated ESS datasets for AI/ML be integrated into existing model benchmarking efforts such as ILAMB (Collier et al. 2018). A standardized ML-friendly metadata format needs to be developed, and novel ML fusion strategies along with NLP methods that can parse advanced metadata formats and generate confidence intervals are on the critical path.
- *The development of methods for out-of-sample, adversarial datasets.* Generating samples and workflows for tractable methods can be useful to tackle this goal. Furthermore, adaptive techniques need to be developed, which could involve a hierarchical flow tracking larger problems first and then concentrating on smaller (finer-scale) problems to identify key features.
- *The reference implementation of the various knowledge discovery techniques.* This can help domain science experts in Earth and environmental sciences start implementing these techniques for their problems.

Short-term goals in statistical learning for AI4ESP include:

- *Good communication between disciplines (e.g., mathematicians and Earth scientists) as UQ crosses disciplinary boundaries*. Exascale algorithms and software must be developed and problems with data products must be addressed by the collaboration of the modeling community and agencies.
- *Development of effective benchmarks for comparing UQ analyses*.

- *The setting of strategies aimed at addressing the curse of dimensionality in current algorithms, which is posed when fusing data from different sources*. Alongside, we need to develop and be able to enforce complete and partial physics constraints, which are poorly addressed by current algorithms.
- *Exploration of new neural network approaches for density estimation of non-Gaussian distributions*.
    - Developing ESM output-like gridded and gap-filled (across different space and time scales) datasets based on multiple observational sources and aligned with resolution and units of ESM outputs. In the same vein, observation-informed interpolation tools should be developed to interpolate model outputs in space, time, and across scales.
    - Developing subgrid-scale parameterizations based on stochastic and deep learning principles, and their extension to physics-informed and physics-constrained as well as scale-aware capabilities, should be a next step to ensure operationality of these models.
    - *Uncertainty quantification tools and techniques* should be developed and should accompany new parameterizations in order to assess errors and uncertainties and their propagations in ESMs running with new parameterizations.
    - *Development of tools for unsupervised learning that are scalable to large datasets* but also interoperable with existing AI/ML workflows in Python, R, and Julia.
    - *Research into techniques to address data sparsity* from missing data or data that are not collocated in the same place and time.

Finally, development of tools for unsupervised learning that are scalable to large datasets but also interoperable with existing AI/ML workflows in Python, R, and Julia is another near- to medium-term goal, as is research into techniques to address data sparsity from missing data or data that are not collocated in the same place and time.

Mid-term goals (5–10 years) aim for a standardized validation framework for the models and datasets and developing robust AI-ready benchmark datasets. It is essential to incorporate UQ in DL models by "default" by developing new algorithms that can handle various versions of data distributions. Automated methods that provide uncertainty estimates with DL model predictions must become the norm. We note that some of this requires development of means to solve difficult (non-Gaussian) probability density estimation problems. In addition, exploring how to use causal learning to gain insight into causal relationships addressing data sparsity challenges are important mid-term targets.

For **a long-term goal**, a completed pipeline and feedback loop should be developed, from data collection to UQ, modeling and data acquisition by actively learning where data should be collected to reduce prediction uncertainty and improve model accuracy. The availability of such a pipeline will make it possible to routinely deploy automated AI methods to exploit full data records. Data sampling workflows built around edge devices can attribute such efforts and can also be a long-term goal. Such opportunities open up a platform where close collaboration between domain experts in Earth science and computer science can take place.

14.7 References

Beer, Christian, Markus Reichstein, Enrico Tomelleri, Philippe Ciais, Martin Jung, Nuno Carvalhais, Christian Rödenbeck, et al. 2010. "Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate." *Science* 329 (5993): 834–38. https://doi.org/10.1126/SCIENCE.1184984/SUPPL_FILE/BEER_SOM.PDF.

Berry, Tyrus, and John Harlim. 2017. "Correcting Biased Observation Model Error in Data Assimilation." *Monthly Weather Review* 145 (7): 2833–53. https://doi.org/10.1175/MWR-D-16-0428.1.

Bode, Mathis, Michael Gauding, Konstantin Kleinheinz, and Heinz Pitsch. 2019. "Deep Learning at Scale for Subgrid Modeling in Turbulent Flows: Regression and Reconstruction." In *High Performance Computing*, edited by Michèle Weiland, Guido Juckeland, Sadaf Alam, and Heike Jagode, 541–60. Cham: Springer International Publishing. https://doi.org/https://doi.org/10.1007/978-3-030-34356-9_41.

Boukabara, Sid Ahmed, Vladimir Krasnopolsky, Stephen G. Penny, Jebb Q. Stewart, Amy McGovern, David Hall, John E. ten Hoeve, et al. 2021. "Outlook for Exploiting Artificial Intelligence in the Earth and Environmental Sciences." *Bulletin of the American Meteorological Society* 102 (5): E1016–32. https://doi.org/10.1175/BAMS-D-20-0031.1.

Caldwell, P. M., C. R. Terai, B. Hillman, N. D. Keen, P. Bogenschutz, W. Lin, H. Beydoun, et al. 2021. "Convection-Permitting Simulations With the E3SM Global Atmosphere Model." *Journal of Advances in Modeling Earth Systems* 13 (11): e2021MS002544. https://doi.org/10.1029/2021MS002544.

Collier, Nathan, Forrest M Hoffman, David M Lawrence, Gretchen Keppel-Aleks, Charles D. Koven, William J. Riley, Mingquan Mu, and James T. Randerson. 2018. "The International Land Model Benchmarking (ILAMB) System: Design, Theory, and Implementation." *Journal of Advances in Modeling Earth Systems* 10 (11): 2731–54. https://doi.org/https://doi.org/10.1029/2018MS001354.

Cych, Brendan, Matthias Morzfeld, and Lisa Tauxe. 2021. "Bias Corrected Estimation of Paleointensity (BiCEP): An Improved Methodology for Obtaining Paleointensity Estimates." *Geochemistry, Geophysics, Geosystems* 22 (8): e2021GC009755. https://doi.org/https://doi.org/10.1029/2021GC009755.

Fukami, Kai, Koji Fukagata, and Kunihiko Taira. 2019. "Super-Resolution Reconstruction of Turbulent Flows with Machine Learning." *Journal of Fluid Mechanics* 870 (July): 106–20. https://doi.org/10.1017/jfm.2019.238.

Gardner, Jacob R., Geoff Pleiss, David Bindel, Kilian Q. Weinberger, and Andrew Gordon Wilson. 2018. "GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration," September. https://doi.org/10.48550/arxiv.1809.11165.

Geer, A. J. 2021. "Learning Earth System Models from Observations: Machine Learning or Data Assimilation?" *Philosophical Transactions of the Royal Society A* 379 (2194). https://doi.org/10.1098/RSTA.2020.0089.

Giles, Michael B. 2015. "Multilevel Monte Carlo Methods." *Acta Numerica* 24 (April): 259–328. https://doi.org/10.1017/S096249291500001X.

Gorodetsky, Alex A., Gianluca Geraci, Michael S. Eldred, and John D. Jakeman. 2020. "A Generalized Approximate Control Variate Framework for Multifidelity Uncertainty Quantification." *Journal of Computational Physics* 408 (May): 109257. https://doi.org/10.1016/J.JCP.2020.109257.

Harlim, John, Shixiao W. Jiang, Senwei Liang, and Haizhao Yang. 2021. "Machine Learning for Prediction with Missing Dynamics." *Journal of Computational Physics* 428 (March): 109922. https://doi.org/10.1016/J.JCP.2020.109922.

Hoffman, Forrest M., Jitendra Kumar, Richard T. Mills, and William W. Hargrove. 2013. "Representativeness-Based Sampling Network Design for the State of Alaska." *Landscape Ecology* 28 (8): 1567–86. https://doi.org/10.1007/S10980-013-9902-0/FIGURES/10.

Isaac, Tobin, Noemi Petra, Georg Stadler, and Omar Ghattas. 2015. "Scalable and Efficient Algorithms for the Propagation of Uncertainty from Data through Inference to Prediction for Large-Scale Problems, with Application to Flow of the Antarctic Ice Sheet." *Journal of Computational Physics* 296 (September): 348–68. https://doi.org/10.1016/J.JCP.2015.04.047.

Jakeman, John D, Michael S Eldred, Gianluca Geraci, and Alex Gorodetsky. 2020. "Adaptive Multi-Index Collocation for Uncertainty Quantification and Sensitivity Analysis." *International Journal for Numerical Methods in Engineering* 121 (6): 1314–43. https://doi.org/https://doi.org/10.1002/nme.6268.

Jesson, Andrew, Panagiotis Tigas, Joost van Amersfoort, Andreas Kirsch, Uri Shalit, and Yarin Gal. 2021. "Causal-BALD: Deep Bayesian Active Learning of Outcomes to Infer Treatment-Effects from Observational Data," November. https://doi.org/10.48550/arxiv.2111.02275.

Kennedy, M C, and A O'Hagan. 2000. "Predicting the Output from a Complex Computer Code When Fast Approximations Are Available." *Biometrika* 87 (1): 1–13. http://www.jstor.org/stable/2673557.

Kim, Seokhyeon, Sajjad Eghdamirad, Ashish Sharma, and Joong Hoon Kim. 2020. "Quantification of Uncertainty in Projections of Extreme Daily Precipitation." *Earth and Space Science* 7 (8): e2019EA001052. https://doi.org/10.1029/2019EA001052.

Kobyzev, Ivan, Simon J. D. Prince, and Marcus A. Brubaker. 2019. "Normalizing Flows: An Introduction and Review of Current Methods," August. https://doi.org/10.1109/tpami.2020.2992934.

Kumar, Jitendra, Richard T. Mills, Forrest M. Hoffman, and William W. Hargrove. 2011. "Parallel K-Means Clustering for Quantitative Ecoregion Delineation Using Large Data

Sets.” *Procedia Computer Science* 4 (January): 1602–11. https://doi.org/10.1016/J.PROCS.2011.04.173.
Langford, Zachary L., Jitendra Kumar, Forrest M. Hoffman, Amy L. Breen, and Colleen M. Iversen. 2019. “Arctic Vegetation Mapping Using Unsupervised Training Datasets and Convolutional Neural Networks.” *Remote Sensing* 11 (1): 69. https://doi.org/10.3390/rs11010069.
Leeuwen, Peter Jan van, Yuan Cheng, and Sebastian Reich. 2015. *Nonlinear Data Assimilation*. 1st ed. Springer Cham. https://doi.org/https://doi.org/10.1007/978-3-319-18347-3.
Lorenz, Edward. 1956. *Empirical Orthogonal Functions and Statistical Weather Prediction*. Cambridge: Massachusetts Institute of Technology Department of Meteorology.
Lu, Dan, Siyan Liu, and Daniel Ricciuto. 2019. “An Efficient Bayesian Method for Advancing the Application of Deep Learning in Earth Science.” *IEEE International Conference on Data Mining Workshops, ICDMW* 2019-November (November): 270–78. https://doi.org/10.1109/ICDMW.2019.00048.
Lunderman, Spencer, Matthias Morzfeld, Franziska Glassmeier, and Graham Feingold. 2020. “Estimating Parameters of the Nonlinear Cloud and Rain Equation from a Large-Eddy Simulation.” *Physica D: Nonlinear Phenomena* 410 (September): 132500. https://doi.org/10.1016/J.PHYSD.2020.132500.
Luo, Y. Q., J. T. Randerson, G. Abramowitz, C. Bacour, E. Blyth, N. Carvalhais, P. Ciais, et al. 2012. “A Framework for Benchmarking Land Models.” *Biogeosciences* 9 (10): 3857–74. https://doi.org/10.5194/bg-9-3857-2012.
Majda, Andrew, and John Harlim. 2012. *Filtering Complex Turbulent Systems*. Cambridge University Press. https://www.cambridge.org/us/academic/subjects/mathematics/mathematical-modelling-and-methods/filtering-complex-turbulent-systems?format=HB&isbn=9781107016668.
Mills, Richard Tran, Jitendra Kumar, Forrest M. Hoffman, William W. Hargrove, Joseph P. Spruce, and Steven P. Norman. 2013. “Identification and Visualization of Dominant Patterns and Anomalies in Remotely Sensed Vegetation Phenology Using a Parallel Tool for Principal Components Analysis.” *Procedia Computer Science* 18 (January): 2396–2405. https://doi.org/10.1016/J.PROCS.2013.05.411.
Mills, Richard Tran, Vamsi Sripathi, Jitendra Kumar, Sarat Sreepathi, Forrest Hoffman, and William Hargrove. 2018. “Parallel K-Means Clustering of Geospatial Data Sets Using Manycore CPU Architectures.” In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 787–94. https://doi.org/10.1109/ICDMW.2018.00118.
Narayan, Akil, Claude Gittelson, and Dongbin Xiu. 2014. “A Stochastic Collocation Algorithm with Multifidelity Models.” *SIAM Journal on Scientific Computing* 36 (2): A495–521. https://doi.org/10.1137/130929461.
Peherstorfer, Benjamin, Karen Willcox, and Max Gunzburger. 2018. “Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization.” *SIAM Review* 60 (3): 550–91. https://doi.org/10.1137/16M1082469.

Penny, S. G., E. Bach, K. Bhargava, C. C. Chang, C. Da, L. Sun, and T. Yoshida. 2019. "Strongly Coupled Data Assimilation in Multiscale Media: Experiments Using a Quasi-Geostrophic Coupled Model." *Journal of Advances in Modeling Earth Systems* 11 (6): 1803–29. https://doi.org/10.1029/2019MS001652.

Qian, Yun, Charles Jackson, Filippo Giorgi, Ben Booth, Qingyun Duan, Chris Forest, Dave Higdon, Z Jason Hou, and Gabriel Huerta. 2016. "Uncertainty Quantification in Climate Modeling and Projection." *Bulletin of the American Meteorological Society* 97 (5): 821–24. https://doi.org/10.1175/BAMS-D-15-00297.1.

Rasch, P. J., S. Xie, P. L. Ma, W. Lin, H. Wang, Q. Tang, S. M. Burrows, et al. 2019. "An Overview of the Atmospheric Component of the Energy Exascale Earth System Model." *Journal of Advances in Modeling Earth Systems* 11 (8): 2377–2411. https://doi.org/10.1029/2019MS001629.

Rasp, Stephan, Michael S. Pritchard, and Pierre Gentine. 2018. "Deep Learning to Represent Subgrid Processes in Climate Models." *Proceedings of the National Academy of Sciences of the United States of America* 115 (39): 9684–89. https://doi.org/10.1073/PNAS.1810286115/SUPPL_FILE/PNAS.1810286115.SAPP.PDF.

Restrepo, J M. 2017. "A Dynamic Likelihood Approach to Filtering." *Quarterly Journal of the Royal Meteorological Society* 143 (708): 2915–24. https://doi.org/https://doi.org/10.1002/qj.3143.

Restrepo, Juan M., and Shankar Venkataramani. 2016. "Stochastic Longshore Current Dynamics." *Advances in Water Resources* 98 (December): 186–97. https://doi.org/10.1016/J.ADVWATRES.2016.11.002.

Rosenthal, W. Steven, Shankar Venkataramani, Arthur J. Mariano, and Juan M. Restrepo. 2017. "Displacement Data Assimilation." *Journal of Computational Physics* 330 (February): 594–614. https://doi.org/10.1016/J.JCP.2016.10.025.

Särkkä, Simo. 2013. *Bayesian Filtering and Smoothing*. Cambridge University Press. https://doi.org/10.1017/CBO9781139344203.

Sreepathi, Sarat, Jitendra Kumar, Richard T Mills, Forrest M Hoffman, Vamsi Sripathi, and William W Hargrove. 2017. "Parallel Multivariate Spatio-Temporal Clustering of Large Ecological Datasets on Hybrid Supercomputers." In *2017 IEEE International Conference on Cluster Computing (CLUSTER)*, 267–77. https://doi.org/10.1109/CLUSTER.2017.88.

Sunyer, Maria Antonia, Henrik Madsen, Dan Rosbjerg, and Karsten Arnbjerg-Nielsen. 2014. "A Bayesian Approach for Uncertainty Quantification of Extreme Precipitation Projections Including Climate Model Interdependency and Nonstationary Bias." *Journal of Climate* 27 (18): 7113–32. https://doi.org/10.1175/JCLI-D-13-00589.1.

Tezaur, Irina Kalashnikova, Kara Peterson, Amy Powell, John Jakeman, and Erika Roesler. 2021. "Global Sensitivity Analysis Using the Ultra-Low Resolution Energy Exascale Earth System Model." *Earth and Space Science Open Archive*, 38. https://doi.org/10.1002/essoar.10508267.2.

Urrego-Blanco, Jorge R, Nathan M Urban, Elizabeth C Hunke, Adrian K Turner, and Nicole Jeffery. 2016. "Uncertainty Quantification and Global Sensitivity Analysis of the Los Alamos Sea Ice Model." *Journal of Geophysical Research: Oceans* 121 (4): 2709–32. https://doi.org/https://doi.org/10.1002/2015JC011558.

Wainwright, Haruko M., Sebastian Uhlemann, Maya Franklin, Nicola Falco, Nicholas J. Bouskill, Michelle E. Newcomer, Baptiste Dafflon, et al. 2022. "Watershed Zonation through Hillslope Clustering for Tractably Quantifying Above- and below-Ground Watershed Heterogeneity and Functions." *Hydrology and Earth System Sciences* 26 (2): 429–44. https://doi.org/10.5194/hess-26-429-2022.

Watson-Parris, Duncan, Andrew Williams, Lucia Deaconu, and Philip Stier. 2021. "Model Calibration Using ESEm v1.1.0-an Open, Scalable Earth System Emulator." *Geoscientific Model Development* 14 (12): 7659–72. https://doi.org/10.5194/GMD-14-7659-2021.

Yuval, Janni, and Paul A O'Gorman. 2020. "Stable Machine-Learning Parameterization of Subgrid Processes for Climate Modeling at a Range of Resolutions." *Nature Communications* 11 (1): 3295. https://doi.org/10.1038/s41467-020-17142-3.

15 Explainable/Interpretable/Trustworthy Artificial Intelligence

Authors: Line Pouchard (BNL), Derek DeSantis (LANL), Ben Brown (LBNL), Bhavya Kailkhura (LLNL), Guang Lin (Purdue), Amy McGovern (Univ. of Oklahoma), Yuhan Rao (NC State Univ.), Byung-Jun Yoon (Texas A&M)

Acknowledgement: Session leaders & workshop organizers would like to acknowledge the significant contribution to report of Maria Glenski & Svitlana Volkova.

15.1 Introduction

Artificial intelligence and machine learning (AI/ML) have emerged as important methods of scientific discovery in Earth and environmental systems (EES). Observation capabilities developed at DOE have enabled the collection of large data archives (Environmental System Science Data Infrastructure for a Virtual Ecosystem [ESS-DIVE]; Earth System Grid Federation [ESGF]) for which these methods can be applied for improved predictions and the discovery of vulnerabilities to extreme events. Increasingly complex methods used for EES, such as deep learning, currently trained on gigabytes of data, include millions of parameters, especially when running on high-performance computing (HPC) systems. Active learning techniques and generative models are also being used for data fusion and the development of hybrid and physics-informed process models (Stevens et al. 2020) for ESP. In spite of these advances, society might not benefit from the full potential in applying these data-driven methods if scientists cannot understand why and how certain outcomes have been predicted and communicate model uncertainty to the general population.

Interpreting the results of ML models in a scientific context and with tools that make sense to domain scientists is a huge challenge across all AI/ML application areas, let alone explaining predictions with wide societal implications to a broader constituency impacted by model results, as could be the case for EES applications (e.g., regarding natural hazards). Alongside the challenge of implementing interpretability and explainability methods, the definitions of *Explainable* and *Interpretable* AI are many, some of which are orthogonal along various dimensions or otherwise conflict with each other (Barredo Arrieta et al. 2020; Arya et al. 2019; Longo et al. 2020; Murdoch et al. 2019), which introduces additional challenges for integration with AI/ML approaches in EES. If methods for explainable/interpretable ML are proposed and classified along specific goals, for instance, Trustworthy AI (Wing 2021), the question becomes, *What is trust*? Are model outcomes reproducible with a given confidence level, and to what extent is it important for the expected goal of model development? Reaching consensus on definitions of AI/ML implies formalizing AI explainable/interpretable methods in a rigorous framework that includes metrics for explainability that quantify why a particular explanation is better than another.

15.2 Grand Challenges

Numerous challenges face the development of Explainable/Interpretable/Trustworthy (EIT) AI that will affect the adoption of promising AI methods in Earth systems science. If not addressed in a timely manner, these challenges will reduce a potentially large, beneficial impact on findings, the mitigation of natural disasters, the management of hazards, and their consequences on human populations.

15.2.1 Grand Challenge: The Discovery of Forward Models from Data Enabling Human Understanding in Domains of Previously Intractable Complexity

Using explainable and interpretable AI for the discovery of first-principles or forward models has been demonstrated in simple physical systems. However, Earth systems are highly complex, and the parameters that are relevant for control – system drivers – are not always known *a priori*. We need techniques that are capable of revealing organizing principles, ecosystem control points, and forward models in systems where the relevant spatiotemporal scales and parameters are unknown. This is a radically more challenging problem than has been solved or even approached by the AI community thus far. Such advancements would enable the data-driven discovery of first principles and forward models for complex systems that have, prior to the emergence of explainable and interpretable AI, remained recalcitrant to concise, human-accessible models and representations. An organized effort at scale is required, one that the DOE (perhaps uniquely) is poised to deliver.

15.2.2 Grand Challenge: Reproducibility, Robustness, and Confidence for AI Models Compatible with Forward Models

While impressive advances have been made in AI/ML techniques, the reproducibility of the prediction outcomes obtained by such data-driven models are often questioned in scientific domains. Concerns regarding robustness and reproducibility of AI-based predictions become more salient when dealing with complex and dynamic systems with substantial uncertainty, especially with Earth science data (e.g., in situ observations, satellite measurements, and citizen science data) for which there may not be sufficient data for reliably training AI/ML models for predictive tasks. As a result, concerns regarding the lack of robustness and reproducibility of AI/ML models are common in ESP and other sciences. This negatively impacts the adoption of AI/ML techniques despite their potential to advance ESP research and accelerate novel scientific discoveries in relevant fields. The lack of reproducibility of AI-based applications, especially in ESP, can also slow the progress of the field since new research is often built on existing research and tools (Peng and Hicks 2021). Reproducibility can also significantly reduce researchers' efforts to recreate previous research and accelerate new development and discovery.

There are various factors affecting the robustness and reproducibility of AI/ML (or lack thereof) in ESP. The biggest challenges include the absence of community-wide guidelines for ensuring the robustness and reproducibility of AI/ML, as well as the lack of standard metrics and computing infrastructure for quantifying the robustness and reproducibility of AI/ML models and their outcomes. Metrics generally applicable across topic areas of Earth and environmental sciences as well as across the multiple data archives are lacking, preventing quantification and understanding of the robustness and reproducibility of AI/ML models. Key components to ensure the reproducibility, including well-preserved data, clearly documented workflow and code (from data preparation to model training and evaluation), and description of the computing environment, are often lacking in AI-based applications, as authors tend to privilege accuracy over trustworthiness in most publications (Gundersen, Gil, and Aha 2018; Haibe-Kains et al. 2020).

15.2.3 Grand Challenge: The Capacity to Invert Prediction Intervals to Obtain Confidence Regions for ESP and EES

We currently lack a rigorous definition of uncertainty quantification (UQ) in ML. Different papers provide diverse and sometimes non-overlapping motivations for UQ and offer several notions of uncertainty and confidence. Existing uncertainty-aware ML methods are mostly designed for and evaluated under ideal scenarios (e.g., noise-free setting, single modal/scale data, etc.). On the other hand, ESP applications face multifidelity, multiresolution, and multimodal data that are high dimensional and spatiotemporal in nature. Existing uncertainty-aware ML methods are not suitable for these scenarios and are expected to face major challenges in terms of accuracy, robustness, and scalability. To make this even more complex, benchmarks and workflows that can identify the shortcomings of existing methods when applied to ESP applications do not exist. Solving this grand challenge will require a close collaboration between domain scientists and ML researchers.

Recent advances in conformal prediction have enabled the ascertainment of prediction intervals with finite sample guarantees for black box AI learning machines, a significant advance. The capacity to invert prediction intervals to obtain confidence regions would have seismic impacts on data science. The capacity to conduct inference has remained the domain of "statistical" and mathematical models with imposed statistical constraints. If the same can be done with highly predictive AI models, the way we think about using such models will fundamentally change.

Combining the accessibility of confidence regions with parameter discovery from representation learning would enable "hypothesis discovery" from AI models with detailed insights into how and why the model functions on par with forward models.

15.2.4 Grand Challenge: Pervasive Environmental Observatories and Trustworthy AI for Impacted Communities

When working to predict many climate and weather phenomena, using observational data is critical. However, there are many areas of the world, even within the United States, where there is a lack of the observational data (a data desert) needed to study the phenomena of interest. For example, air pollution sensors are not well distributed in many cities, and DOE is currently working to expand this network and address the bias of home-based sensors. For this grand challenge that would address the needs of climate and environmental justice, we propose to facilitate the development of sensors and models reliable on a global scale. Then we could responsibly use AI to create synthetic data that would help address the needs of the locations and populations with significant data underrepresentation.

A related challenge is to create trustworthy AI for use by the communities whose needs it addresses. It is crucial that the communities impacted by AI, as well as those who will use it (as they may not necessarily be the same), are involved in the development of the AI from the beginning. As a part of this grand challenge, we need to develop community-centered data collection methods and approaches to explainable AI for a variety of audiences. This will require developing standards of explainability and trust. Finally, there is a need to work on causal discovery, which can be used to identify new scientific knowledge for the impacted communities.

15.2.5 Grand Challenge: A Common Language and Development of Best Practices between Computational, AI, and Domain Scientists

Explaining AI for Earth systems that enables domain scientists to trust and understand predictions and their predicates requires us to find a common language that AI, computational, and domain scientists understand. While experts in their domains, Earth scientists may not be well versed in the technicalities, jargon, and methods of AI. As such, additional effort must be made by all parties to foster communication with patience and open-mindedness. This represents an additional challenge and points to a need for support in the fast-paced research environment common within the DOE. One place to start is to begin to develop best practices of AI implementation within ESP. For example, the use of the appropriate modeling techniques, metrics and diagnostic tools adapted to the various problems and data specificity needs to be normalized. Trustworthy and Explainable AI plays a central role in this development.

15.2.6 Grand Challenge: Domain-informed Development of Interpretable and Explainable AI Methodology

The integration of domain knowledge in the development and research lifecycle for interpretable and explainable AI methodology is another grand challenge facing Interpretable and Explainable AI for Earth systems applications. It is important to understand the context in which an EIT methodology will be used and the questions regarding interpretability or explainability of model behavior that are most critical for EES applications. Currently, EIT methods must be adapted or extended for use in domain applications. By incorporating domain knowledge directly into development of new methodologies, the development of applicable methods can be both optimized for Earth systems' needs, constraints, and opportunities but accelerated.

Domain-informed development can incorporate not only the necessary context for greatest impact of interpretable and explainable methods but also reduce the workload required to introduce interpretability and explainability for the AI used in EES applications. For example, domain-informed methods should be more easily used by domain experts (whose needs the methods were explicitly designed for). This should reduce the barriers to wider adoption within the community.

15.3 State of the Science

The state of the science in Explainable and Interpretable AI and related fields is vast and growing. We present it using the room discussion categories, complemented by a summary table (Table 15-1) at the end of this section.

15.3.1 Interpretable and Explainable AI (XAI)

The current state of interpretable and explainable AI is vast and rapidly evolving. To help organize the field, we divided this section into three subsections. These subsections are: theory (what justifies a good explanation), techniques (methods and tools for evaluating explanations), and XAI uses currently in ESP. The theory and techniques sections purposely cover the entire AI community and are not restricted to environmental sciences.

15.3.1.1 Theory

Theoretical work on XAI applications to science has been explored in the social sciences, where what justifies a "good" explanation is more naturally defined. The survey (Miller 2019) highlights four major findings across these fields:

1. Explanations are contrastive – they are sought in response to a particular counterfactual case, rather than seeking route cause.
2. Explanations are selected in a biased manner – an explanation is not expected to fully explain the full cause of an event.
3. Probabilities do not matter – referring to probabilities as the explanation is not as effective as referring to the route cause.
4. Explanations are social – an explanation is a transfer of knowledge between two individuals and is dependent on the explainer's and explainee's beliefs.

More recently, theory from Mahoney et. al have revealed how and why deep neural networks are able to self-regularize and achieve the extraordinary performance they manage, and this has led to a new class of regularization techniques that is possible even in the absence of access to training data (Martin, Peng, and Mahoney 2021). The revelation of the mechanisms of implicit self-regularization in deep networks is an important result, as it illustrates that even industry-grade deep networks with millions of parameters are not impossibly complex. In parallel, Donoho and colleagues illustrated the form of decision boundaries that a widely applied class of neural networks inevitably learn (Papyan, Han, and Donoho 2020). These are foundations on which we can build techniques to extract locally low-dimensional models – human-understandable models and knowledge – from fitted networks, as Mahoney and colleagues from across the DOE complex pointed out in a recent review, which made use of the meeting report from ASCR's AI4SCI meeting series in 2019 (Pion-Tonachini et al. 2021).

15.3.1.2 Techniques

Techniques within XAI can be broken down into two categories: (1) methods—tools used to provide an explanation, and (2) evaluations—tools used to evaluate the explanation. Here we briefly outline the state of the art for each of these fields.

Methods: The broad state of the science for interpretable and explainable AI encompasses many methods, including Shapley values, integrated gradients, expected hessians, locally interpretable model-agnostic explanations (LIME), counterfactual explanations, and many more. For an extensive review, see Molnar (2020). However, these methods were predominantly constructed to function in regimes with high signal-to-noise ratios, with the majority focus on the impact of individual "variables" or "pixels," without consideration for interactions with other features. A variety of methods, including integrated hessians, multivariate Shapley values, partial dependence plots, and, more recently, Accumulated Local Effects (ALE) plots attempt to capture lower-order interactions, e.g., between pairs of covariates, or, in some cases, higher-order interactions based on prior hypotheses. At present, there is no general approach to discover or map higher-order interactions for arbitrary learning machines – only in special cases, such as Random Forests, is this tractable (Basu et al. 2018).

The pursuit of local marginal importance is largely synonymous with "Saliency Maps," visualizations of important pixels in image data. Radiology has long been held up as an example of where these maps are useful, yet the utility of such maps has recently been called into question by leaders in the field (Arun et al. 2021), and this is not surprising. These methods were developed in regimes where the cost of being wrong is very low. For applications where the cost of being wrong is high, for applications in science and engineering where certainty is valued and essential, new methods are needed.

While saliency maps and related techniques attempt to provide explanations for black box AI/ML models, another class of techniques attempts to build explainability and interpretability directly into the learning framework. The discovery of physics and conservation laws directly from data has been demonstrated in many systems (Liu and Tegmark 2021). However, such demonstrations are thus far confined to low-noise or noiseless systems in very low dimension (e.g., three dimensions of space and one of time—i.e., very few degrees of freedom). Earth systems do not fit this framework, and more advanced methods are needed.

Also intriguing has been the use of AI models for hypothesis discovery and prioritization in mathematics (Davies et al. 2021). Novel theory has been originated by close collaboration between AI models and human mathematicians. This is an important advancement, and perhaps the most compelling example of a "self-driving lab," albeit a purely *in silico* lab, yet advanced. Indeed, the proposal of theory by AI models is precisely what we need in the climate and environmental sciences – the rub, of course, is that Earth systems cannot be written in clean closed forms like much of mathematics.

This is to say there is currently a zoo of methods branded as XAI. However, most of these fall into one of two categories: intrinsically explainable models or post hoc methods for interrogating a "black box" model. Intrinsically explainable models usually include cleverly designed latent spaces that are physically or otherwise interpretable to domain scientists, such as tensor factorization techniques. Interrogation procedures further segregate into three primary classes: (1) saliency maps/feature importance measure, (2) local approximations/local models, and (3) indirect or heuristic explanations. We refer to these as Type 1, 2, and 3 explanations, respectively. There is a strong commonality between methods, often differing only in minor ways (e.g., how a gradient is computed).

- *Type 1 - saliency maps/feature importance*: Saliency/attribution maps and feature importance measures provide methods to describe the decision of a classifier by highlighting regions (features) of "importance." There are many such methods, which are among the oldest and most widely used interpretation techniques. Methods such as Layerwise Relevance Propagation (LRP) and its numerous spin-offs measure importance by propagating information from the prediction back to neurons in earlier layers in a neural network (Montavon et al. 2019). Approximate Shapley values (e.g., SHAP) are widely used, but have recently been shown to be qualitatively inaccurate approximately of actual Shapley values in important use cases (Lundberg and Lee 2017). DeepLift falls into the same category as LRP and SHAP, and indeed the three methods differ principally only in how they center neural activation functions during layer-wise propagation (Shrikumar, Greenside, and Kundaje 2017). Gradient-based methods measure the value or importance of a feature by evaluating the partial derivatives of the network's output with respect to the input variables (Janizek, Sturmfels, and Lee 2021). Integrated and expected gradients are widely used techniques in this family, as are integrated and expected Hessians, which attempt to capture pairwise interactions between features. Work has also been done to identify components of larger networks responsible for classification and prediction problems, such as groups of neurons or entire layers (Olah et al. 2020).

- *Type 2 - local approximations/local models*: Partial dependence plots, ALE plots, and related methods attempt to identify marginal response surfaces, where marginalization is done over all but a few features, yielding low-dimensional response surfaces suitable for human interrogation. These methods have yielded important discoveries in the biological sciences (Basu et al. 2018), and applications in the environmental sciences are promising and on the horizon. H-statistics use the framework of partial dependence plots to develop statistically rigorous tests of dependence between features – but are computationally expensive to compute, and intractable for high-dimensional settings. In some sense, all these methods are "bottom up" – they can be viewed as forward testing procedures analogous to t-tests and the principle of marginality in linear statistical models. "Top down" procedures exist as well – LIME and related procedures attempt to fit local models that approximate the behavior of the method they aim to explain and leverage classical statistical regularization. However, a principal challenge with existing top-down procedures is that they are used to fit to the data, and not the parent learning machine they aim to explain – and therefore do not necessarily capture information about the learned data representation (Pion-Tonachini et al. 2021).
- *Type 3 - indirect or heuristic explanations*: Methods here can be further subdivided into counterfactuals and examples. Given a black-box model, if a perturbation to input is performed in such a way that we know what the consequences should be, analysis can then be performed to see if the model's actions correspond with reality. For example, in Ates et al. (2021), the authors generate counterfactuals on time series data to show which time series needs to be modified, and how, to change the classification result in a desired way. Often models are trained to detect extreme events (e.g., tornadic storms), which are inherently rare. When a less extreme event occurs, these models can classify as a novel observation and relate to the analogous examples the model has seen. Hase et. al. (2019) use a taxonomic organization of classes within a vision model to produce explanations by examples. For instance, if the model has only been trained on images of rifles, and is given an image of a handgun, it can learn from past exposure that the handgun is a "novel object," but should still be classified as a "weapon."

Evaluations: The ability to evaluate an explanation can be as valuable, if not more valuable, than the explanation itself. If an explanation's evaluation is poorly justified, it can prevent the user from post hoc justifying misleading results. Approaches for evaluation include automatic, as well as human study designs. Case studies have been performed to compare the human evaluations and automatic explanations generated (Chu, Roy, and Andreas 2020; Nguyen 2018). Yu and colleagues propose the "Predictability, Computability, Stability" (PCS) Framework (Yu and Kumbier 2020) and suggest study designs and inference procedures to ensure that results are interpretable and useful to domain scientists and therefore for scientific discovery and the furtherance of the scientific method. To test saliency methods, Adebayo et. al. (2018) test the sensitivity of eight different saliency map techniques to the underlying model and input labels. Benchmarks for interpretability methods have been investigated; however, the scope has largely been devoted to image classification and natural language processing (DeYoung et al. 2019; Hooker et al. 2018).

15.3.1.3 XAI ESP Applications

While XAI is itself a relatively new field, it has already found numerous applications within ESP. Methods such as saliency maps have been employed across a range of problems to interrogate the underlying ML model. Examples include tornado predictions (McGovern, Lagerquist, and Gagne 2020) and SST anomalies (Barnes et al. 2020). Feature importance has also been used, for instance in Jergensen et al. (2020) to ensure trust in storm classification. More recently, Mamalakis et. al. (2022) developed a method for evaluating XAI method efficacy on geoscience data.

15.3.2 Robustness and Reproducibility

Currently, there is increasing awareness of the need to establish standardized guidelines and benchmarks across the scientific community to promote robust and reproducible AI/ML, both in their training process as well as their application to scientific predictions. Definitions of Reproducibility and Replicability in science have been adopted by the National Academies of Sciences, Engineering, and Medicine (National Academies of Sciences Engineering and Medicine 2019). Case-by-case and full-factor reproducibility studies are being performed without overall frameworks or guidelines for the evaluation of reproducibility and robustness of AI/ML models, in Earth, environmental, and other sciences (Alahmari et al. 2020; Pouchard, Lin, and van Dam 2020; Plale and Harrell 2021). However, interest in a more principled approach is emerging. For example, the Research Data Alliance (RDA) is forming FAIR4ML, a new Interest Group, whose aim is to establish FAIR (findable, accessible, interoperable, and reusable) principles for machine learning (Katz et al. 2020; Psomopoulos 2021).

The Neural Information Processing System (NeurIPS), one of the leading international ML conferences, established a reproducibility challenge in 2019 and adopted a reproducibility checklist for conference submissions (Pineau et al. 2020). This information checklist is more comprehensive than the ACM Artifact Detection/Artifact Evaluation appendices in use at the SC

(Super Computing) conferences. In particular, in addition to reproducibility, authors are asked questions about ethics and explainability of the AI models discussed in their submissions. To ensure research reproducibility, the American Statistical Association implemented a reproducibility review process for journal article submission to ensure that data, code, and workflow are clear and available (JASA Applications & Case Studies 2019). This review process and guidelines can be adopted by the Earth science community to enhance the reproducibility for ESP. The adoption of FAIR principles for AI/ML, while not the goal in itself, would help toward increasing transparency and establishing a detailed provenance trail for models used in ESP.

Recent work concludes that common ML platforms (e.g., Amazon Sagemaker, Azure ML, Google Cloud ML, Kaggle, etc.) do not support reproducibility out-of-the-box, in an experiment testing the reproducibility of digit classification using a standard CNN (Gundersen, Shamsaliei, and Isdahl 2022). A series of reproducibility metrics quantifying support for reproducibility in these platforms is also proposed, applicable to each documentation type required for Data, Method, and Experiment in publications and platform documentation (Gundersen and Kjensmo 2018). Additionally, recent advancement of open-source science projects, such as Project Jupyter, Pangeo, and Binder Project, can facilitate the documentation, sharing, and preservation of workflow and research output. These projects can be used to enable the reproducibility of AI-based applications with proper curation process. The scientific community is moving in this direction by establishing templates and guidelines on how to share workflow and research outcomes using interactive notebooks (e.g., EarthCube Jupyter notebook template; EarthCube 2021) and AGU's guidelines for authors on Jupyter notebooks (Erdmann and Stall, et al. 2021) and R scripts/markdowns (Erdmann and Meyer, et al. 2021).

Due to the complex nature of AI-based ESP applications, it might be very difficult to reproduce the results to machine precision. While well-developed tools in statistics (e.g., for hypothesis testing) and uncertainty quantification (Ghanem, Owhadi, and Higdon 2017) may be used to assess and enhance the reproducibility of AI/ML predictions, they are not actively utilized in the AI/ML community for verification. Objective-based UQ frameworks based on MOCU (mean objective cost of uncertainty) are promising but will need to be adapted to serve the purpose of reproducibility and use appropriate criteria (Yoon, Qian, and Dougherty 2021; Yoon, Qian, and Dougherty 2013). Statistical reproducibility of model simulations of large ensembles such as E3SM has been measured, and new multivariate tests have been designed that apply to traditional simulations on hybrid architectures, but these have not yet been applied to AI models (Mahajan et al. 2019; Mahajan 2021). Further research is needed for the application to ESP of statistical tools, the design of metrics, and the production of benchmarks.

15.3.3 Benchmarking/Verification/Uncertainty Quantification

Statistical inference methods are important to enable UQ in ML. However, the computationally expensive nature of both training and inference in these methods makes it infeasible to scale these methods to modern ML methods, such as deep neural networks (DNNs) applied to high-dimensional data. Recent advances in approximate Bayesian inference hold significant promise for addressing these concerns. Methods inspired from "Infinitesimal Jackknife" (Giordano et al. 2019) and "Deep Ensembles" (D'Angelo and Fortuin 2021) are pushing the boundaries of Bayesian inference in deep learning. Desiderata for post hoc uncertainty calibration is as follows: accuracy-preserving, data-efficient, and high expressive power. Innovative methods to simultaneously achieve these objectives are proposed in Zhang, Kailkhura, and Han (2020). Reliable methods to evaluate the calibration error are also proposed in Kumar, Liang, and Ma (2019). Please refer to Abdar et al. (2021); Li, Xie, and Li (2020), and Zhang et al. (2019) for more detailed surveys of relevant literature on uncertainty quantification, verification, and benchmarking methods in deep learning, respectively.

15.3.4 Ethics and Responsible AI

The state of science for ethical and responsible AI is new enough that the study of it for ESS is even more nascent. We can look to the study of ethical AI and ethical algorithms (e.g., Kearns and Roth 2019; Benjamin 2019; O'Neil 2016) to influence the development of ethical and responsible AI for ESS. We can also look to the study of ethical behavior in social sciences (e.g., Pidgeon 2021) to inform the development and use of AI for ESS. Other existing approaches for responsible AI include manual solutions, such as model cards (Mitchell et al. 2019), data statements (Bender and Friedman 2018), data sheets (Gebru et al. 2021), and extensions that build on these techniques, such as DAG Cards (Tagliabue et al. 2021). Moreover, responsible AI principles include rigorous model validation and benchmarking, including the evaluation of model robustness, fairness, accountability, and transparency (https://www.pnnl.gov/projects/trusted-and-responsible-ai). Understanding why humans will trust AI for weather and climate predictions and measuring the appropriate level of trust needed to ensure community adoption is a complex task, and the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (ai2es.org) is actively working to understand this (McGovern et al. 2021).

Table 15-1. Synthesis of recent advances in Explainable, Interpretable, and Trustworthy AI.

Topic	**AI Application(s) and Related Work(s)**
Interpretable and Explainable AI	• Attention-based model interpretability techniques, such as saliency maps/. • Functional explainability and exploratory error analysis tools, such as AI Explainability 360 (Arya et al. 2020), Crosscheck (Arendt et al. 2021), IntepretML (Nori et al. 2019, Kaur et al. 2020). • Counterfactual, what-if analyses, e.g., the What-If Tool (Wexler et al. 2019). • Surrogate model-based techniques, such as LIME (Local Interpretable Model-agnostic Explanations) (Ribeiro, Singh, and Guestrin 2016), SHAP (SHapley Additive exPlanations) (Lundberg and Lee 2017), and variants. • Molnar (2020) Interpretable Machine Learning book.
Robustness and Reproducibility	• Case-by-case reproducibility studies (Alahmari et al. 2020; Pouchard, Lin, and van Dam 2020; Plale and Harrell 2021). • Survey of publications and ML platforms for reproducibility (Gundersen, Shamsaliei, and Isdahl 2022; Gundersen, Gil, and Aha 2018; Haibe-Kains et al. 2020; Peng and Hicks 2021). • Multivariate approaches for statistical reproducibility for traditional simulations (but not AI) on hybrid architectures (Mahajan et al. 2019). • Mahajan (2021). • Decision-making frameworks that can be adapted to reproducibility in AI/ML (Yoon, Qian, and Dougherty 2021; Yoon, Qian, and Dougherty 2013).
Uncertainty Quantification	• Ghanem, Roger, David Higdon, and Houman Owhadi, eds. Handbook of uncertainty quantification. Vol. 6. New York: Springer, 2017. • Scalable Bayesian Inference: Methods inspired from "Infinitesimal Jackknife" (Giordano et al. 2019) and "Deep Ensembles" (D'Angelo and Fortuin 2021). • Post hoc Uncertainty Calibration Methods and Metrics: "Mix-n-Match" (Zhang, Kailkhura, and Han 2020) and "Verified Calibration" (Kumar, Liang, and Ma 2019). • Relevant survey papers: uncertainty quantification (Abdar et al. 2021), robustness verification (Li, Xie, and Li 2020), and benchmarking methods (Zhang et al. 2019).
Ethics and Responsible AI	• Study of ethical AI and ethical algorithms (e.g., Kearns and Roth 2019; Benjamin 2019; O'Neil 2016). • Need for ethics in AI for ESS (McGovern et al. 2021). • Documentation frameworks for model and dataset development and bias: model cards (Mitchell et al. 2019), data statements (Bender and Friedman 2018), data sheets (Gebru et al. 2021), and extensions that build on these techniques, such as DAG Cards (Tagliabue et al. 2021). • Responsible AI: fairness, accountability, transparency, security and safety, inclusivity (McGovern et al. 2021).

15.4 Experimental, Data, and Modeling Opportunities

In this section, we introduce the experimental, data, and modeling opportunities for advancing the development and use of AI/ML approaches in Earth and environmental sciences, and in particular to address the grand challenges described above. In Figure 15-1, we present a summary of responses from workshop participants to the question "What are some short-term opportunities (low-hanging fruit)?" for Explainable, Interpretable, and Trustworthy AI for ESP. Several key themes emerge in the word cloud: *effective usage* of EIT methodology and the *definition of community standards*, the benefit of *(narrow) use cases,* and the *scale of data*.

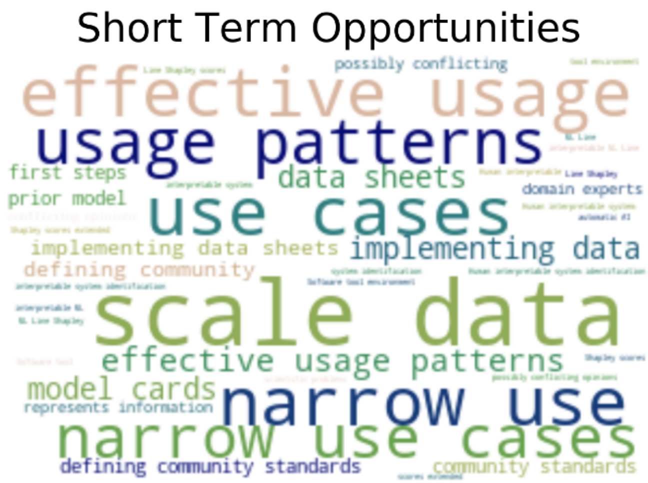


Figure 15-1. Word cloud summarizing the pre-workshop responses for "What are some short-term opportunities (low-hanging fruit)?" for Explainable, Interpretable, and Trustworthy AI for ESP.

15.4.1 Interpretable and Explainable AI

AI models currently predict environmental processes for which we have no first-principle models: biomass production; couplings across scales; the impact of rare events; and cloud physics, to name a few. Do these AI models contain within them knowledge that humans at present lack? Explainable and interpretable methods are needed to understand what it is that such models have "understood." There is an opportunity to build upon XAI methods in regimes where the physics does have first-principle models, and leverage these developments into new, uncharted scientific discoveries.

Earth systems datasets are sparse. In the Watershed SFA, we have hundreds, not thousands, of soil organic matter and microbiome datasets. Weather stations, snowfall measurements, and groundwater measurements are sparse. Earth systems datasets have been, heretofore, designed to meet the needs of mechanistic models by humans with a limited understanding of systems for which they develop hypotheses. AI built to be explainable and interpretable constitutes a parallel and equally valid system for understanding the natural world—though it requires far more data, and far more complex experimental designs. Our experience underscores the importance of supporting, in parallel, this new paradigm, in that it stands to augment and enhance human intuition with insights derived from data and "superhuman" intuition and insight.

Further, ESP provides a valuable opportunity to provide benchmarks for AI models leveraging domain knowledge about complex, nontrivial systems. Such benchmarks would be extremely valuable as we develop new and more advanced tools, just as classical physics has provided useful benchmarks for explainable AI tools such as AI Feynman, AI Poincare, and inductive graph networks (Udrescu and Tegmark 2020; Liu and Tegmark 2021). Can explainable and interpretable AI tools recover known relationships about hydrological forcings, or the couplings between reaction-diffusion and reactive transport processes across spatiotemporal scales? Deep hidden physics models have already demonstrated the recoverability of fluid dynamics. PFLOTRAN, ATS, and related codes provide exciting next-generation benchmarks for these methods.

In short, the "ModEx" paradigm should be expanded to include experimental design based on AI models. Interpretable and explainable AI methods can leverage existing concepts for ESP-focused development of AI/ML approaches to increase the usability of explanations or simplify explanations, which should increase the adoption of XAI methods. Building on existing approaches such as OSSE (observing system simulation experiment), interpretable and explainable AI approaches could be used to guide the identification of where more data are needed for ESP applications, for example, due to poor sampling or undersampling of most informative features.

15.4.2 Robustness and Reproducibility

The creation of community-wide standards, guidelines, and benchmarks to promote and verify the robustness and reproducibility of AI/ML models and their predictions in ESP can have significant and lasting impacts on improving the robustness and reproducibility of AI/ML research in the field, as well as other scientific fields that deal with real-world systems with immense complexity and uncertainty. Not only will these efforts improve the reliability of AI/ML outcomes in ESP, they are expected to enhance the trustworthiness of AI/ML-based scientific predictions and conclusions, thereby facilitating the adoption of AI/ML in ESP applications.

The envisioned standards and guidelines should inform researchers who develop and/or adopt AI/ML models in ESP research of the minimum requirements (1) for training/testing AI/ML models to ensure the robustness of the trained models as well as the predictions obtained from the trained models, as well as (2) for making models and the predictions reproducible. These include, but are not limited to, guidelines for dataset size/split for training/testing, AI/ML training procedure, performance metrics to be used, and sharing of data/metadata. Unlike typical performance benchmarks, robustness/reproducibility benchmarks should be designed to assess the robustness and reproducibility of AI/ML training procedure, training outcomes, and the predictions made by the trained AI/ML models. Furthermore, these benchmarks should help end-users better understand the limitations of the models, procedures, and predictions and point out the primary factors affecting robustness and reproducibility (or the lack thereof).

15.4.3 Benchmarking/Verification/Uncertainty Quantification

Solving these grand challenges will likely require drawing inspirations from many diverse research areas and a close collaboration between domain scientists and ML researchers. Three specific research areas that will play a major role in achieving this are: (1) verification, (2) statistics, and (3) ML benchmarking.

DNN verification (Albarghouthi 2021) is a process of verifying whether for every possible input, the neural network output satisfies the desired properties. Popular verification approaches draw motivations from formal methods and can now verify robustness to additive perturbations (Xu et al. 2020) or semantic changes in the data (Li et al. 2021).

Probabilistic modeling and inference are critical for achieving uncertainty-awareness in ML/deep learning. Normalizing flows (Papamakarios et al. 2019) provide a general mechanism for defining expressive probability distributions, only requiring the specification of a (usually simple) base distribution and a series of bijective transformations. There has been much recent progress on normalizing flows, ranging from improving their expressive power to expanding their application.

Properly curated datasets and rigorous benchmarks are critical for evaluating the progress of a research field and identifying gaps and shortcomings. A limited number of efforts have been carried out in traditional computer vision applications to achieve this (Nado et al. 2021). Similar efforts and evaluation practices are desired for ESP applications.

15.4.4 Ethics/Responsible AI

Solving the grand challenges presented here from a perspective of ethical and responsible AI requires a community-centered approach to data collection and model creation (Renn et al. 1995; National Research Council 1996; Chilvers 2009; Voinov et al. 2016; Pidgeon, 2021). The data must be collected in conjunction with local leaders, and there must be clear guidelines on what data are collected, how it is to be used for AI, and how the data must be stored or shared. Data are key to the success of AI models, and there must be clear guidelines on what to collect and store. Data collection should take into account underrepresented populations and geographic regions to reduce model biases and generalizability. AI problems and tasks must be driven by diverse populations and local communities to ensure models developed address the needs of diverse communities.

There must also be a focus on transparency and communication of the AI models with a variety of end-users. While this includes the need to develop and enhance explainable and interpretable AI models (e.g., Molnar 2020), we also need to develop a standard for sharing AI models and data. This will facilitate reproducible research and also help to minimize biases in the models. However, easy and broad access to AI models might also trigger misuse by adversaries.

15.5 Research Priorities

In this section, we outline what research must be conducted as ***next steps*** for addressing the Grand Challenges, and what the research priorities are.

15.5.1 Interpretable and Explainable AI

We have identified several key steps to be taken for XAI within ESP. These can be summarized as developing XAI libraries; training, integrating, and developing new XAI methods; developing explainable models; and involving stakeholders:

1. ***Development of easy-to-use XAI libraries.*** The mass adoption, automation, and integration of XAI into the ESP framework will require a standard library of tools for the scientific community.
2. ***Training XAI ethos for Earth Scientists.*** Alongside the development of standard libraries, workforce development should be emphasized. New scientists entering the sphere need to be trained in these practices, while the old guard needs to have this framework updated as standard.
3. ***XAI exploration for ESP.*** While many XAI methods have been tested within ESP, the tools used thus far are far from exhaustive. Scientists need to explore options already developed by the other sciences to open the possibility of improvement within the Earth sciences.
4. ***ESP-Specific XAI tools***. Most implementations of XAI methods for ESP have arisen from directly importing tools developed for other purposes. The Earth sciences need to produce more XAI methods designed for problems within the Earth sciences. Examples here could include ML that attempts to reveal relevant spatiotemporal scales such as in DeSantis et al. (2020).
5. ***Explainable models to learn first-principles.*** This can only be accomplished by assembling relevant spatiotemporal representations and the physics that bind them to measurable Earth system outcomes.
6. ***Get stakeholders involved.*** What is deemed important and worth understanding within a model is largely a function of those who are impacted by its success. It will be important to leverage domain knowledge and engage stakeholders early and often in the development of novel approaches that can take advantage of the opportunities provided by ESP applications.

15.5.2 Robustness and Reproducibility

The next steps for addressing the challenges to develop robust and reproducible AI/ML methods in ESP include the following.

1. ***Establish reproducibility standards for AI-based ESP research***. The ESP research community should develop a common standard of reproducible ESP research by incorporating the unique challenge of the complexity of ESP and dynamic Earth science data.
2. ***Develop community-wide guidelines and recommended practices to promote robust and reproducible AI/ML in ESP***. Inspired by FAIR guiding principles, these guidelines and recommended practices should include, but are not limited to, clear documentation of the availability and provenance of data, code availability, instructions on how to implement the code with critical information (e.g., version of software used, model hyperparameters, and model training strategy), and expected outcomes. These guidelines could also be leveraged to address the challenge of preserving large Earth science data for robust and reproducible ESP.
3. ***Create benchmarks for robust and reproducible ESP research***. The ESP community should develop benchmark frameworks including data, community-adopted metrics, and assessment tools that can be used to assess and validate the robustness and reproducibility of an AI/ML model and its predictions. Especially, the tools and procedures for evaluating the robustness and reproducibility of AI/ML should be based on the intended scientific objectives of ESP research. To support the aforementioned next steps, it would also be important to develop standard metrics for quantifying the reproducibility of AI/ML models, i.e., both their training, as well as their predictive outcomes based on its impact on the scientific goals, similar to unit testing in software development. Statistical tools for hypothesis testing and objective-based UQ techniques—e.g., based on the concept of mean objective cost of uncertainty (MOCU)—for quantifying the impact of uncertainties on robustness and reproducibility may provide a useful means for achieving this goal.

15.5.3 Benchmarking/Verification/Uncertainty Quantification

To solve the grand challenge of uncertainty quantification in AI4ESP, cross-pollination of ideas between verification, statistics, and ML benchmarking communities is highly desired.

First, we need to extend existing uncertainty-aware ML methods, such as Bayesian deep learning, to support the complexity of ESP application data. This will bring unique modeling and scalability challenges for UQ methods due to the multifidelity, multiresolution, and multimodal nature of the data that are high dimensional and spatiotemporal. Next, in such a setup, supporting accurate and robust uncertainty quantification will require the development of reliable evaluation metrics and efficient design/training methods. Another potentially worthwhile priority direction is to leverage DOE advances in the areas of quantum/neuromorphic/exascale computing and explore software-hardware co-design strategies for next-generation, uncertainty-aware AI methods for ESP. Finally, the development of reproducible workflows and UQ benchmarks (inspired from FAIR principles) will likely be a worthwhile investment advancing the AI4ESP field. To facilitate the trust in AI4ESP, there is also an urgent need to understand the user interpretation of prediction uncertainty and improve how to communicate uncertainty to the public and data users.

15.5.4 Ethics/Responsible AI

To support the development of ethical and responsible AI, we need a focus on both policy and research. For example, we need to support the development of standards for what it means for an AI to be *trustworthy*. This is something that the research community should work on and in conjunction with agencies who can set standards, such as the National Institute of Standards and Technology (NIST). Likewise, the research community needs to develop standards for ethical and responsible development and use of AI. Funding agencies need to ensure that all of their calls for AI-related proposals also require that the AI be developed and used in an ethical and responsible manner.

15.6 Short-term (<5 years), 5-year, and 10-year Goals

In this section, we order the priorities outlined in the previous section, starting with the low-hanging fruit. However, we believe that many long-term goals should be attacked as soon as possible. For instance, efforts to address the lack of a rigorous framework for Explainability, while a long-term goal, should start right away to increase chances of success of XAI in ESP.

15.6.1 Interpretable and Explainable AI

Short term: We have narrowed down two key short-term goals for XAI.

1 ***Explainable, interpretable, and trustworthy AI systems and codes for multiscale Earth systems science.*** Although developing easy-to-use frameworks with accurate UQ (with finite sample guarantees, e.g., via conformal prediction) for the integration of models across scales (and where boundary conditions are intractable or computationally expensive to evaluate) will require a significant effort; all of the pieces are present. Focused efforts could produce codes for the specific task of model coupling across scales with sufficient generality to be applied broadly across the Earth and environmental sciences. Demonstration projects might include improved coupling between reaction diffusion and reactive transport models or improved use of biological data (plants and microbial communities) and metabolic processes to augment Earth systems models. A few distinct pilot projects would be useful to reveal broader commonalities and potential synergies across multiscale (and multidomain) modeling problems. Such projects could lead to a new foundational capability with broad benefits across the BER mission space.

2 ***Proof-of-principle, explainable and interpretable models for complex systems along with benchmarks.*** Well-understood Earth systems processes should be used as initial targets for the discovery of complex physics from observational data. Especially useful would be systems that require multiple modes of observation at more than one scale to define physical processes, e.g., seasonal water quality in a well-studied watershed. These systems would then be integrated into a standard ML workflow, whereby the practitioner would develop XAI methods to measure against expert interpretations. It is imperative that these models cover the wide spectrum of processes observed within the Earth system across multitudes of scales to ensure trust within the benchmarks.

Long term: While the following two goals may only be achieved in the long-term, their importance and complexity warrants starting to address them as soon as possible.

1. ***The development of domain-informed guidelines and discoveries that directly inform and transform domain science – on where, when, and how interpretable and explainable methodology can and should be used***. At what point in the process should this methodology be suggested versus required? Evolving stakeholder input at evolving stages of the interpretable and explainable AI/ML methodology development will necessitate shared understanding of ESP applications and incorporation of interpretable/explainable AI/ML practices.
2. ***AI frameworks.*** We currently lack a rigorous framework for explainability in machine learning. For example, there are no methods to measure the value of explanations, leading one to be unable to falsify the hypothesis of any individual method. Consequently, it is not possible to judge the value of one interpretability method over another. Furthermore, explanations are hard to falsify and can be post hoc justified, resulting in misleading results. While some progress has been made clarifying weakly defined terminology on the philosophy, psychology, and cognitive science front, there is still plenty of debate about the ultimate purpose and function of interpretability in ESP and ML broadly. Those working in ESP cannot tackle this task alone, and it will require joint effort across STEM and the social sciences.

15.6.2 Robustness and Reproducibility

Short term: One of the near-term goals for robust and reproducible AI/ML is establishing community-wide standards and guidelines for better provenance of AI/ML models developed for ESP, which will promote the development and application of more robust and reproducible AI/ML techniques in ESP applications. Furthermore, creating benchmarks for assessing the robustness and reproducibility of AI/ML models for ESP, developing standard metrics for evaluation, and designing software tools that make assessing the robustness and reproducibility of AI/ML in ESP applications streamlined and relatively straightforward would lead to meaningful improvements in the robustness and reproducibility of AI/ML models for ESP in the short term.

Long term: An important long-term goal would be the development of highly reproducible and trustworthy AI/ML models for scientific research in ESP that will lead to AI-driven acceleration of scientific discoveries in ESP applications. To facilitate such development, sustained investment will be required in interlinked data facilities across labs and relevant agencies and accessible computing infrastructures that can enable robust and reproducible AI4ESP research. Such long-term investment can also facilitate equitable participation from historically underrepresented communities and institutions in ESP research. Ultimately, we desire to develop automated and streamlined training/testing procedures for AI/ML models based on standardized, quantifiable, distribution-based metrics that can verify and enhance the robustness and reproducibility of AI/ML models and outcomes in ESP applications.

15.6.3 Benchmarking/Verification/Uncertainty Quantification

Short term: Efforts toward standardization of definitions and metrics to evaluate the quality of UQ are much needed. The community should extend existing UQ methods in ML to support the unique features of the ESP application data. These features include the multifidelity, multiresolution, and multimodal nature of the data, along with high dimensional and spatiotemporal challenges. Development of reliable benchmarks and autonomous workflows should also be carried out in the near future.

Long term: The community should exploit DOE advances in the areas of quantum/neuromorphic and exascale computing and explore software-hardware co-design strategies for next-generation, uncertainty-aware AI methods for ESP. Application of uncertainty-aware ML methods to understand extreme events and seasonal to decadal predictions will likely be a potentially worthwhile direction. Development of uncertainty propagation schemes applied to complex ESP workflows where ML is not a stand-alone component but tightly integrated to the ESP pipeline is another challenge that needs to be overcome in the long term.

15.6.4 Ethics/Responsible AI

Near term: In the near term, we need to ensure that the community is involved in the development of the AI development and lifecycle from the beginning. We also need to immediately begin the development of standards for sharing data and AI models; understand the trade-off between data and model sharing and potential misuse; and develop an open-source approach that facilitates easy communication of models, their data, and any limitations.

Short term: Within the five-year timeframe, there needs to be a publicly accepted definition of trustworthy, ethical, and responsible AI for environmental sciences, and all AI researchers should be focusing on ensuring that their AI projects follow the accepted standards. Research should be more convergent, including both AI researchers, social scientists, and environmental scientists. Additional students are needed to measure and identify areas where there is a lack of data, and approaches to address the missing data should be developed. Finally, there needs to be a continued focus on the development of explainable and interpretable AI methods.

Long term: By the end of the 10-year time frame, most AI methods should be either explainable or interpretable. This will facilitate additional trust in the methods as well as the ethical and responsible use of AI for ESP. Given the changing climate and associated extreme weather events, AI tools should be facilitating decision-making around climate resiliency and climate and environmental justice.

15.7 References

Abdar, Moloud, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, et al. 2021. "A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges." *Information Fusion* 76 (December): 243–97. https://doi.org/10.1016/j.inffus.2021.05.008.

Adebayo, Julius, Justin Gilmer, Michael Muelly, Ian Goodfellow, Moritz Hardt, and Been Kim. 2018. "Sanity Checks for Saliency Maps," October. https://doi.org/10.48550/arxiv.1810.03292.

Alahmari, Saeed S., Dmitry B. Goldgof, Peter R. Mouton, and Lawrence O. Hall. 2020. "Challenges for the Repeatability of Deep Learning Models." *IEEE Access* 8: 211860–68. https://doi.org/10.1109/ACCESS.2020.3039833.

Albarghouthi, Aws. 2021. "Introduction to Neural Network Verification," September. https://doi.org/10.48550/arxiv.2109.10317.

Alexander, Francis J., Tammie Borders, Angie Sheffield, and Marc Wonders. 2020. "Workshop Report for Next-Gen AI for Proliferation Detection: Accelerating the Development and Use of Explainability Methods to Design AI Systems Suitable for Nonproliferation Mission Applications." United States. https://doi.org/10.2172/1768761.

Arendt, Dustin, Zhuanyi Shaw, Prasha Shrestha, Ellyn Ayton, Maria Glenski, and Svitlana Volkova. 2021. "CrossCheck: Rapid, Reproducible, and Interpretable Model Evaluation." In *Proceedings of the Second Workshop on Data Science with Human in the Loop: Language Advances*, 79–85. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.dash-1.13.

Arun, Nishanth, Nathan Gaw, Praveer Singh, Ken Chang, Mehak Aggarwal, Bryan Chen, Katharina Hoebel, et al. 2021. "Assessing the Trustworthiness of Saliency Maps for Localizing Abnormalities in Medical Imaging." https://doi.org/10.1148/RYAI.2021200267.

Arya, Vijay, Rachel K. E. Bellamy, Pin-Yu Chen, Amit Dhurandhar, Michael Hind, Samuel C. Hoffman, Stephanie Houde, et al. 2019. "One Explanation Does Not Fit All: A Toolkit and

Taxonomy of AI Explainability Techniques," September. https://doi.org/10.48550/arxiv.1909.03012.

Arya, Vijay, Rachel K E Bellamy, Pin-Yu Chen, Amit Dhurandhar, Michael Hind, Samuel C Hoffman, Stephanie Houde, et al. 2020. "AI Explainability 360: An Extensible Toolkit for Understanding Data and Machine Learning Models." *Journal of Machine Learning Research* 21 (130): 1–6. http://jmlr.org/papers/v21/19-1035.html.

Ates, Emre, Burak Aksar, Vitus J Leung, and Ayse K Coskun. 2021. "Counterfactual Explanations for Multivariate Time Series." In *2021 International Conference on Applied Artificial Intelligence (ICAPAI)*, 1–8. https://doi.org/10.1109/ICAPAI49758.2021.9462056.

Barnes, Elizabeth A., Kirsten Mayer, Benjamin Toms, Zane Martin, and Emily Gordon. 2020. "Identifying Opportunities for Skillful Weather Prediction with Interpretable Neural Networks," December. https://doi.org/10.48550/arxiv.2012.07830.

Barredo Arrieta, Alejandro, Natalia Díaz-Rodríguez, Javier del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, et al. 2020. "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI." *Information Fusion* 58 (June): 82–115. https://doi.org/10.1016/J.INFFUS.2019.12.012.

Basu, Sumanta, Karl Kumbier, James B Brown, and Bin Yu. 2018. "Iterative Random Forests to Discover Predictive and Stable High-Order Interactions." *Proceedings of the National Academy of Sciences* 115 (8): 1943–48. https://doi.org/10.1073/pnas.1711236115.

Bender, Emily M, and Batya Friedman. 2018. "Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science." *Transactions of the Association for Computational Linguistics* 6: 587–604. https://doi.org/10.1162/tacl_a_00041.

Benjamin, Ruha. 2019. *Race after Technology : Abolitionist Tools for the New Jim Code*. Cambridge: Polity. https://www.wiley.com/en-us/Race+After+Technology:+Abolitionist+Tools+for+the+New+Jim+Code-p-9781509526437.

Chilvers, Jason. 2009. "Deliberative and Participatory Approaches in Environmental Geography." In *A Companion to Environmental Geography*, 400–417. John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.1002/9781444305722.ch24.

Chu, Eric, Deb Roy, and Jacob Andreas. 2020. "Are Visual Explanations Useful? A Case Study in Model-in-the-Loop Prediction," July. https://doi.org/10.48550/arxiv.2007.12248.

D'Angelo, Francesco, and Vincent Fortuin. 2021. "Repulsive Deep Ensembles Are Bayesian," June. https://doi.org/10.48550/arxiv.2106.11642.

Davies, Alex, Petar Veličković, Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomašev, Richard Tanburn, et al. 2021. "Advancing Mathematics by Guiding Human Intuition with AI." *Nature 2021 600:7887* 600 (7887): 70–74. https://doi.org/10.1038/s41586-021-04086-x.

DeSantis, Derek, Phillip J Wolfram, Katrina Bennett, and Boian Alexandrov. 2020. "Coarse-Grain Cluster Analysis of Tensors with Application to Climate Biome Identification."

Machine Learning: Science and Technology 1 (4): 045020. https://doi.org/10.1088/2632-2153/abb676.

DeYoung, Jay, Sarthak Jain, Nazneen Fatema Rajani, Eric Lehman, Caiming Xiong, Richard Socher, and Byron C. Wallace. 2019. "ERASER: A Benchmark to Evaluate Rationalized NLP Models," November. https://doi.org/10.48550/arxiv.1911.03429.

EarthCube. 2021. "Notebook Templates." 2021. https://github.com/earthcube/NotebookTemplates/.

Erdmann, Christopher, Michael F. Meyer, John R. Little, Kelly Hondula, Jemma Stachelek, Isabella Oleksy, Matthew R. Brousil, Kelly Claborn, Jorrit Mesman, and Tim Dennis. 2021. "Guidance for AGU Authors: R Script(s)/Markdown," November. https://doi.org/10.5281/ZENODO.5647998.

Erdmann, Christopher, Shelley Stall, Chelle Gentemann, Chris Holdgraf, Filipe P. A. Fernandes, Karsten Peters-von Gehlen, and Marianne Corvellec. 2021. "Guidance for AGU Authors – Jupyter Notebooks," May. https://doi.org/10.5281/ZENODO.5651648.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2021. "Datasheets for Datasets." *Commun. ACM* 64 (12): 86–92. https://doi.org/10.1145/3458723.

Ghanem, Roger, Houman Owhadi, and David Higdon. 2017. "Handbook of Uncertainty Quantification." *Handbook of Uncertainty Quantification*, June, 1–2053. https://doi.org/10.1007/978-3-319-12385-1/COVER.

Giordano, Ryan, William Stephenson, Runjing Liu, Michael Jordan, and Tamara Broderick. 2019. "A Swiss Army Infinitesimal Jackknife." In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, edited by Kamalika Chaudhuri and Masashi Sugiyama, 89:1139–47. Proceedings of Machine Learning Research. PMLR. https://doi.org/https://doi.org/10.48550/arXiv.1806.00550.

Gundersen, Odd Erik, Yolanda Gil, and David W Aha. 2018. "On Reproducible AI: Towards Reproducible Research, Open Science, and Digital Scholarship in AI Publications." *AI Magazine* 39 (3): 56–68. https://doi.org/10.1609/aimag.v39i3.2816.

Gundersen, Odd Erik, and Sigbjørn Kjensmo. 2018. "State of the Art: Reproducibility in Artificial Intelligence." *Proceedings of the AAAI Conference on Artificial Intelligence* 32 (1). https://doi.org/10.1609/aaai.v32i1.11503.

Gundersen, Odd Erik, Saeid Shamsaliei, and Richard Juul Isdahl. 2022. "Do Machine Learning Platforms Provide Out-of-the-Box Reproducibility?" *Future Generation Computer Systems* 126 (January): 34–47. https://doi.org/10.1016/J.FUTURE.2021.06.014.

Haibe-Kains, Benjamin, George Alexandru Adam, Ahmed Hosny, Farnoosh Khodakarami, Thakkar Shraddha, Rebecca Kusko, Susanna Assunta Sansone, et al. 2020. "Transparency and Reproducibility in Artificial Intelligence." *Nature 2020 586:7829* 586 (7829): E14–16. https://doi.org/10.1038/s41586-020-2766-y.

Hase, Peter, Chaofan Chen, Oscar Li, and Cynthia Rudin. 2019. "Interpretable Image Recognition with Hierarchical Prototypes," June. https://doi.org/10.48550/arxiv.1906.10651.

Hooker, Sara, Dumitru Erhan, Pieter-Jan Kindermans, and Been Kim. 2018. "A Benchmark for Interpretability Methods in Deep Neural Networks," June. https://doi.org/10.48550/arxiv.1806.10758.

Janizek, Joseph D, Pascal Sturmfels, and Su-In Lee. 2021. "Explaining Explanations: Axiomatic Feature Interactions for Deep Networks." *Journal of Machine Learning Research* 22 (104): 1–54. http://jmlr.org/papers/v22/20-1223.html.

JASA Applications & Case Studies. 2019. "JASA Reproducibility Guide | Reproducibility Guide for Authors and Reviewers Publishing in the Journal of the American Statistical Association (JASA)." 2019. https://jasa-acs.github.io/repro-guide/pages/review-form.

Jergensen, G. Eli, Amy McGovern, Ryan Lagerquist, and Travis Smith. 2020. "Classifying Convective Storms Using Machine Learning." *Weather and Forecasting* 35 (2): 537–59. https://doi.org/10.1175/WAF-D-19-0170.1.

Katz, Daniel S., Tom Pollard, Fotis Psomopoulos, Eliu Huerta, Chris Erdmann, and Ben Blaiszik. 2020. "FAIR Principles for Machine Learning Models," November. https://doi.org/10.5281/ZENODO.4271996.

Kaur, Harmanpreet, Harsha Nori, Samuel Jenkins, Rich Caruana, Hanna Wallach, and Jennifer Wortman Vaughan. 2020. "Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning." In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14. CHI '20. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3313831.3376219.

Kearns, Michael, and Aaron Roth. 2019. *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. USA: Oxford University Press, Inc. https://dl.acm.org/doi/book/10.5555/3379082.

Kumar, Ananya, Percy Liang, and Tengyu Ma. 2019. "Verified Uncertainty Calibration," September. https://doi.org/10.48550/arxiv.1909.10155.

Li, Linyi, Maurice Weber, Xiaojun Xu, Luka Rimanic, Bhavya Kailkhura, Tao Xie, Ce Zhang, and Bo Li. 2020. "TSS: Transformation-Specific Smoothing for Robustness Certification," February. https://doi.org/10.48550/arxiv.2002.12398.

Li, Linyi, Tao Xie, and Bo Li. 2020. "SoK: Certified Robustness for Deep Neural Networks," September. https://doi.org/10.48550/arxiv.2009.04131.

Liu, Ziming, and Max Tegmark. 2021. "Machine Learning Conservation Laws from Trajectories." *Physical Review Letters* 126 (18): 180604. https://doi.org/10.1103/PhysRevLett.126.180604.

Longo, Luca, Randy Goebel, Freddy Lecue, Peter Kieseberg, and Andreas Holzinger. 2020. "Explainable Artificial Intelligence: Concepts, Applications, Research Challenges and Visions." In *Machine Learning and Knowledge Extraction*, edited by Andreas Holzinger, Peter Kieseberg, A Min Tjoa, and Edgar Weippl, 1–16. Cham: Springer International Publishing. https://doi.org/https://doi.org/10.1007/978-3-030-57321-8_1.

Lundberg, Scott, and Su-In Lee. 2017. “A Unified Approach to Interpreting Model Predictions,” May. https://doi.org/10.48550/arxiv.1705.07874.

Mahajan, Salil. 2021. “Ensuring Statistical Reproducibility of Ocean Model Simulations in the Age of Hybrid Computing.” *Proceedings of the Platform for Advanced Scientific Computing Conference, PASC 2021*, July. https://doi.org/10.1145/3468267.3470572.

Mahajan, Salil, Katherine J. Evans, Joseph H. Kennedy, Min Xu, and Matthew R. Norman. 2019. “A Multivariate Approach to Ensure Statistical Reproducibility of Climate Model Simulations.” *Proceedings of the Platform for Advanced Scientific Computing Conference, PASC 2019*, June. https://doi.org/10.1145/3324989.3325724.

Mamalakis, Antonios, Imme Ebert-Uphoff, and Elizabeth A. Barnes. 2022. “Neural Network Attribution Methods for Problems in Geoscience: A Novel Synthetic Benchmark Dataset.” *Environmental Data Science* 1: 8–9. https://doi.org/10.1017/EDS.2022.7.

Martin, Charles H., Tongsu S. Peng, and Michael W. Mahoney. 2021. “Predicting Trends in the Quality of State-of-the-Art Neural Networks without Access to Training or Testing Data.” *Nature Communications 2021 12:1* 12 (1): 1–13. https://doi.org/10.1038/s41467-021-24025-8.

McGovern, Amy, Imme Ebert-Uphoff, David John Gagne II, and Ann Bostrom. 2021. “The Need for Ethical, Responsible, and Trustworthy Artificial Intelligence for Environmental Sciences,” December. https://doi.org/10.1017/eds.2022.5.

McGovern, Amy, Ryan A Lagerquist, and David John II Gagne. 2020. “Using Machine Learning And Model Interpretation And Visualization Techniques To Gain Physical Insights In Atmospheric Science.” *ICLR*. https://ai4earthscience.github.io/iclr-2020-workshop/papers/ai4earth16.pdf.

Miller, Tim. 2019. “Explanation in Artificial Intelligence: Insights from the Social Sciences.” *Artificial Intelligence* 267 (February): 1–38. https://doi.org/10.1016/J.ARTINT.2018.07.007.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Deborah Raji, and Timnit Gebru. 2019. “Model Cards for Model Reporting.” *Proceedings of the Conference on Fairness, Accountability, and Transparency*. https://doi.org/10.1145/3287560.

Molnar, C. 2020. *Interpretable Machine Learning*. Lulu.com. https://books.google.com/books?id=jBm3DwAAQBAJ.

Montavon, Grégoire, Alexander Binder, Sebastian Lapuschkin, Wojciech Samek, and Klaus-Robert Müller. 2019. “Layer-Wise Relevance Propagation: An Overview.” In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, edited by Wojciech Samek, Grégoire Montavon, Andrea Vedaldi, Lars Kai Hansen, and Klaus-Robert Müller, 193–209. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-28954-6_10.

Murdoch, W James, Chandan Singh, Karl Kumbier, Reza Abbasi-Asl, and Bin Yu. 2019. "Definitions, Methods, and Applications in Interpretable Machine Learning." *Proceedings of the National Academy of Sciences* 116 (44): 22071–80. https://doi.org/10.1073/pnas.1900654116.

Nado, Zachary, Neil Band, Mark Collier, Josip Djolonga, Michael W. Dusenberry, Sebastian Farquhar, Qixuan Feng, et al. 2021. "Uncertainty Baselines: Benchmarks for Uncertainty & Robustness in Deep Learning," June. https://doi.org/10.48550/arxiv.2106.04015.

National Academies of Sciences Engineering and Medicine. 2019. *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25303.

National Research Council. 1996. *Understanding Risk: Informing Decisions in a Democratic Society*. Edited by Paul C Stern and Harvey v Fineberg. Washington, DC: The National Academies Press. https://doi.org/10.17226/5138.

Nguyen, Dong. 2018. "Comparing Automatic and Human Evaluation of Local Explanations for Text Classification." In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 1069–78. New Orleans, Louisiana: Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-1097.

Nori, Harsha, Samuel Jenkins, Paul Koch, and Rich Caruana. 2019. "InterpretML: A Unified Framework for Machine Learning Interpretability," September. https://doi.org/10.48550/arxiv.1909.09223.

Olah, Chris, Nick Cammarata, Ludwig Schubert, Gabriel Goh, Michael Petrov, and Shan Carter. 2020. "Zoom In: An Introduction to Circuits." *Distill*. https://doi.org/10.23915/distill.00024.001.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. USA: Crown Publishing Group. https://doi.org/https://dl.acm.org/doi/10.5555/3002861.

Papamakarios, George, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. 2019. "Normalizing Flows for Probabilistic Modeling and Inference," December. https://doi.org/10.48550/arxiv.1912.02762.

Papyan, Vardan, X Y Han, and David L Donoho. 2020. "Prevalence of Neural Collapse during the Terminal Phase of Deep Learning Training." *Proceedings of the National Academy of Sciences* 117 (40): 24652–63. https://doi.org/10.1073/pnas.2015509117.

Peng, Roger D, and Stephanie C Hicks. 2021. "Reproducible Research: A Retrospective." *Annual Review of Public Health* 42 (1): 79–93. https://doi.org/10.1146/annurev-publhealth-012420-105110.

Pidgeon, Nick. 2021. "Engaging Publics about Environmental and Technology Risks: Frames, Values and Deliberation." *Journal of Risk Research* 24 (1): 28–46. https://doi.org/10.1080/13669877.2020.1749118.

Pineau, Joelle, Philippe Vincent-Lamarre, Koustuv Sinha, Vincent Larivière, Alina Beygelzimer, Florence d'Alché-Buc, Emily Fox, and Hugo Larochelle. 2020. "Improving Reproducibility in Machine Learning Research (A Report from the NeurIPS 2019 Reproducibility Program)," March. https://doi.org/10.48550/arxiv.2003.12206.

Pion-Tonachini, Luca, Kristofer Bouchard, Hector Garcia Martin, Sean Peisert, W. Bradley Holtz, Anil Aswani, Dipankar Dwivedi, et al. 2021. "Learning from Learning Machines: A New Generation of AI Technology to Meet the Needs of Science," November. https://doi.org/10.48550/arxiv.2111.13786.

Plale, Beth, and Stephen Lien Harrell. 2021. "Transparency and Reproducibility Practice in Large-Scale Computational Science: A Preface to the Special Section." *IEEE Transactions on Parallel and Distributed Systems* 32 (11): 2607–8. https://doi.org/10.1109/TPDS.2021.3058393.

Pouchard, Line, Yuewei Lin, and Hubertus van Dam. 2020. "Replicating Machine Learning Experiments in Materials Science." *Advances in Parallel Computing* 36: 743–55. https://doi.org/10.3233/APC200105.

Psomopoulos, Fotis. 2021. "Steps towards Defining FAIR Principles for Machine Learning (ML)." RDA. July 28, 2021. https://www.rd-alliance.org/steps-towards-defining-fair-principles-machine-learning-ml.

Renn, Ortwin, Thomas Webler, Peter Wiedemann, ed. 1995. *Fairness and Competence in Citizen Participation*. *Fairness and Competence in Citizen Participation*. Springer Netherlands. https://doi.org/10.1007/978-94-011-0131-8.

Ribeiro, Marco, Sameer Singh, and Carlos Guestrin. 2016. *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*. https://doi.org/10.18653/v1/N16-3020.

Shrikumar, Avanti, Peyton Greenside, and Anshul Kundaje. 2017. "Learning Important Features through Propagating Activation Differences." In *Proceedings of the 34th International Conference on Machine Learning – Volume 70*, 3145–53. ICML'17. JMLR.org. https://doi.org/https://doi.org/10.48550/arXiv.1704.02685.

Stevens, Rick, Valerie Taylor, Jeff Nichols, Arthur Barney Maccabe, Katherine Yelick, and David Brown. 2020. "AI for Science: Report on the Department of Energy (DOE) Town Halls on Artificial Intelligence (AI) for Science." United States. https://doi.org/10.2172/1604756.

Tagliabue, Jacopo, Ville Tuulos, Ciro Greco, and Valay Dave. 2021. "DAG Card Is the New Model Card," October. https://doi.org/10.48550/arxiv.2110.13601.

Udrescu, Silviu-Marian, and Max Tegmark. 2020. "AI Feynman: A Physics-Inspired Method for Symbolic Regression." *Science Advances* 6 (16): eaay2631. https://doi.org/10.1126/sciadv.aay2631.

Voinov, Alexey, Nagesh Kolagani, Michael K. McCall, Pierre D. Glynn, Marit E. Kragt, Frank O. Ostermann, Suzanne A. Pierce, and Palaniappan Ramu. 2016. "Modelling with Stakeholders – Next Generation." *Environmental Modelling & Software* 77 (March): 196–220. https://doi.org/10.1016/J.ENVSOFT.2015.11.016.

Wexler, J., M. Pushkarna, T. Bolukbasi, M. Wattenberg, F. Viégas, and J. Wilson. 2020. "The What-If Tool: Interactive Probing of Machine Learning Models." *IEEE Transactions on Visualization and Computer Graphics* 26 (1): 56–65. https://doi.org/10.1109/TVCG.2019.2934619.

Wing, Jeannette M. 2021. "Trustworthy AI." *Communications of the ACM* 64 (10): 64–71. https://doi.org/10.1145/3448248.

Xu, Kaidi, Zhouxing Shi, Huan Zhang, Yihan Wang, Kai-Wei Chang, Minlie Huang, Bhavya Kailkhura, Xue Lin, and Cho-Jui Hsieh. 2020. "Automatic Perturbation Analysis for Scalable Certified Robustness and Beyond." In *Advances in Neural Information Processing Systems*, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, 33:1129–41. Curran Associates, Inc. https://doi.org/https://doi.org/10.48550/arXiv.2002.12920.

Yoon, B.-J., X Qian, and E. R. Dougherty. 2013. "Quantifying the Objective Cost of Uncertainty in Complex Dynamical Systems." *IEEE Transactions on Signal Processing* 61 (9): 2256–66. https://doi.org/10.1109/TSP.2013.2251336.

Yoon, Byung-Jun, Xiaoning Qian, and Edward R. Dougherty. 2021. "Quantifying the Multi-Objective Cost of Uncertainty." *IEEE Access* 9: 80351–59. https://doi.org/10.1109/ACCESS.2021.3085486.

Yu, Bin, and Karl Kumbier. 2020. "Veridical Data Science." *Proceedings of the National Academy of Sciences* 117 (8): 3920–29. https://doi.org/10.1073/pnas.1901326117.

Zhang, Jize, Bhavya Kailkhura, and T. Yong-Jin Han. 2020. "Mix-n-Match: Ensemble and Compositional Methods for Uncertainty Calibration in Deep Learning," March. https://doi.org/10.48550/arxiv.2003.07329.

Zhang, Qin, Li Zha, Jian Lin, Dandan Tu, Mingzhe Li, Fan Liang, Ren Wu, and Xiaoyi Lu. 2019. "A Survey on Deep Learning Benchmarks: Do We Still Need New Ones?" In *Benchmarking, Measuring, and Optimizing*, edited by Chen Zheng and Jianfeng Zhan, 36–49. Cham: Springer International Publishing. https://doi.org/https://doi.org/10.1007/978-3-030-32813-9_5.

16 Hybrid Modeling

Authors: Sivasankaran Rajamanickam (SNL), Prasanna Balaprakash (Argonne), Peishi Jiang (PNNL), Jiali Wang (Argonne), Nathan Urban (BNL)

16.1 Grand Challenges

16.1.1 Use of the Diverse and Multiscale Data for Training AI/ML Methods

Crucial to the effectiveness of many AI/ML methods is the data. A major challenge that is specific for modeling in general and hybrid modeling in particular is *the diversity and multiscale nature of the data sources*. For example, there are a wide range of simulation and observational data that can differ in spatial-temporal scales and are available at various fidelity and resolution levels. The observational data are often sparse, and they are not available at the spatial-temporal scale required for various predictive modeling tasks. This difficulty is further exacerbated by the fact that the observation instruments are noisy and thus provide noisy data. Consequently, there is an inherent noise in the data that needs to be carefully handled when used for training AI/ML methods.

16.1.2 Data Preparation for AI/ML Training

Given the advancement of AI/ML software stack and open-source libraries, AI/ML model training is increasingly becoming an easier task. However, the most challenging aspect is the *data preparation*. Often overlooked is how much effort is required to prepare the data that are ready for AI/ML training. In fact, in many AI/ML tasks, the data preparation is the bottleneck and can take a significant amount of time and prevent scientists from leveraging the AI/ML methods for various datasets. While standard data preprocessing methods do exist, they cannot be used out of the box for the climate and weather datasets. On the other hand, there is not a single data processing method that will work for all the settings. One cannot make advances in hybrid modeling without making progress in domain-informed data processing methods.

16.1.3 Combining Data from Multiple Sources

Combining data from multiple sources is one of the central aspects of hybrid modeling. However, developing such multimodal fusion strategies in the face of Earth system data is a challenging task. First, the heterogeneity and nature of the data sources (simulations, observations, derived products such as satellites) make it harder to learn representations that are common across different modalities. The datasets can exhibit long-range dependencies among different modalities and learning teleconnections are difficult across such spatial-temporal scales. We have to bridge the scaling and resolution for both models and data, thus accounting for the heterogeneity needed to simulate processes accurately. Moreover, hybrid models include hyperresolution, very fine-scale modeling (e.g., soil respiration, moisture, etc., and scale differences in microbial processes require multiple steps to connect to climate). There can be different levels of interactions based on hierarchies that take advantage of both high-resolution simulation and ensembles of lower-resolutions runs. To that end, developing a hybrid-model for all these fine scales, transferring information between scales, and identifying where high-resolution (regional) model runs are needed for informing coarse-resolution (global) ones, including short- and long-term human impacts, are open research avenues. Noise, resolution, and missing data can vary across the modalities. Creation and curation of diverse and multiscale datasets create additional difficulties in preparing combined data for hybrid modeling.

16.1.4 Bridging the Gaps between Short- and Long-term Forecasting

While data-driven forecasting is typically short term and physics-based models are for long term, there is a *huge gap in the medium-range forecasting*. Hybrid modeling methods that can combine AI/ML models with the PDE-based models are promising, but the best ways to perform interpolation between these models is open research. To that end, real-time data assimilation for short-term models, online hybrid model tuning, and calibration need significant advancement.

16.1.5 Uncertainty Quantification and Propagation

Uncertainty quantification will be key to the effective use and deployment of hybrid models. There are various forms of uncertainty that the hybrid models need to face. The uncertainty in the data due to noise, measurement errors, and low resolution needs to be captured in the predictions appropriately. The AI/ML-based hybrid models are dependent on the training data, and therefore it is critical to capture the model uncertainty or epistemic uncertainty. Moreover, the modeled uncertainties need to be propagated across models, spatial-temporal scales, and resolutions. This is quite challenging in many ways.

16.1.6 Out of Distribution

Out of distribution will be a key challenge for hybrid modeling. Despite simulation, modeling, and observational capabilities, the hybrid models that we train will slowly become obsolete as the state of the Earth system changes. For example, the data and simulation for a hydrologic regime in a particular region can change over time, and thus hybrid models trained with simulation and observational data have to deal with out-of-sample/distribution phenomena. Consequently, hybrid models require tuning and calibration based on observations. It has been acknowledged that such tuning and calibration are performed manually and in an ad hoc way. Currently, there is no systematic approach for dealing with the out of distribution, tuning, and calibrating the hybrid models to reduce thousands of compute core-hours' waste and the sheer time required for tuning the simulation.

16.1.7 Physics-informed AI/ML and Computational Cost

Physics-informed AI/ML rely on expert knowledge to incorporate physics into loss functions and constraints. Under these settings, the AI/ML models are trained by enforcing known physics and data. The challenges are associated with high computational cost. Realistic applications require large DNNs to approximate space-time varying states, parameters, and constitutive relationships. Physics-informed training of these DNNs requires a large number of "collocation" points where the PDEs' residuals are minimized. The large number of collocation points and coefficients in DNNs leads to large minimization problems that are also nonconvex, and, therefore, non-unique, leading to additional model uncertainty. Moreover, there are multiple scenarios in which the physics is unknown (e.g., artificial viscosity, heterogeneous permeability fields). To that end, we need to advance our ability to extract unknown physics from the data.

16.1.8 Hybrid Model Stability

Hybrid model stability will be a crucial factor in the use and deployment of the models. Specifically, when ML surrogates, which are trained offline, are plugged within the simulation, due to distribution shifts the efforts can propagate and make the models unstable. Offline training can potentially induce more issues as they are not informed by physics constraints.

16.1.9 Trustworthiness and Validation of Hybrid Models

Trustworthiness and validation of hybrid models will be a major challenge to wider adoption. Most of the ML models are either overfit or underfit. Although physical constraints can improve, we need domain-specific methods to establish trustworthiness and validation. The standard metrics from the ML literature can help to an extent, but model validation and trust for timescales of interest (climatology) require development of metrics from the domain experts. Moreover, extracting and verifying underlying causal relations in the surrogate mode are critical for both trustworthiness and validation.

An ongoing challenge, even for simple models, is to maintain the stability and forecast skill for a sufficiently long time period. Integrating emulators (representing small scales) into existing PDE dynamic solvers (representing large scales) is a substantial software architecture research challenge. First, the representation of data in the PDE models and in the emulators may differ greatly (e.g., processes, scales). Second, there is a concurrency management problem as emulators compete for resources with existing PDE models, and the relative computation times and resource requirements may change dynamically (e.g., time-stepping). Third, monitoring and debugging models that mix intermediate results from PDE models and emulators will be difficult (e.g., exponential growth of errors). Overall, the development of a hybrid modeling system will face similar challenges when compared to the development of conventional models, such as the complexity of modeling the Earth system, with its nonlinear interactions between model components, scale interactions, exponential growth of errors in initial conditions, numerical instabilities, and many more.

16.2 State-of-the-Science

In this section, we summarize the state-of-the-science in hybrid modeling with regard to (1) the state of science in hybrid modeling for Earth science, and (2) the state of science in ML for supporting hybrid modeling.

16.2.1 The State of Science in Hybrid Modeling for Earth Science

Broadly speaking, hybrid modeling has been applied in the following three scenarios in Earth science: (1) learning PDEs through AI; (2) coupling simulation from physics-based models with observation/experiment data; and (3) coupling AI-based emulators with physics-based models.

Extensive efforts exist in employing AI as surrogates in representing PDE-based dynamics. For instance, physics-informed neural networks (PINNs) learn PDE by incorporating physics-based laws in the loss function of a DL model (Karniadakis et al. 2021). PINN has been employed in different studies in Earth science, particularly in subsurface hydrological processes (Tartakovsky et al. 2020), where observations are usually hard to obtain. Also, neural operators started gaining attention recently (Anandkumar et al. 2019). This technique learns PDE through a predefined operator such as Fourier operator (Li et al. 2020) and has experienced some successes in emulating sea surface height (Jiang et al. 2021) and belowground multiphase flow (Wen et al. 2021).

Another scenario of hybrid modeling is to couple model simulations with observations data. One widely used application is to employ AI as error corrector to adjust the model simulations (Willard et al. 2020). Such a corrector can be deployed in the following two ways. The first approach is to directly construct the AI-based error corrector using the difference between simulation and observations (Karpatne et al. 2017). The second approach first pre-trains a surrogate using model simulations, and then fine-tunes the pre-trained emulator using an observation dataset (Read et al. 2019).

One good example of integrating AI-based emulators with physics-based models is coupling a DL-based emulator for turbulent heat fluxes with a process-based hydrological model framework (Bennett and Nijssen 2021). However, compared with the previous two scenarios, there are limited applications on such hybrid modeling. The main reason probably lies in the technical bottleneck of integrating python-based AI codes (e.g., pytorch) with C++/Fortran-based physical models (see Grand Challenges, section 16.1, for details).

16.2.2 The State of Science in ML to Support Hybrid Modeling

In general, the workflows and computational infrastructure for hybrid modeling are still in their early stages. While deep learning frameworks (e.g., pytorch and tensorflow) have gain their popularity rapidly in the past decade, a more universal framework for ML-based scientific programming is still in its infancy phase. Nevertheless, emerging tools exist, such as Jax – a python library for differential programming serving as a more mature alternative to numpy autograd; and Julia – a relatively new programming language that is specifically designed for high-performance computing. These examples of progress have enabled new modeling activities in Earth science. An ocean scientist developed Veros (Häfner et al. 2018), a JAX-based ocean model that supports both CPU and GPU-based parallel computing and shows comparable performance with the corresponding Fortran codes. These developments would eventually not only change the "legacy way" of physical modeling using C++/Fortran and but also greatly ease the process of integrating physics-based models and DL models.

16.3 Experimental, Data, and Modeling Opportunities

Opportunities lay ahead owing to the increasing availability of observation/simulations in the Earth system; the potential computational efficiency gain through hybrid modeling; the development of computing software/resources; and the potential funding avenues.

Over the last decade, the advance of observation technology and computation power enables the generation of numerous datasets including both observation and model simulations, which would greatly benefit the development of hybrid models or AI models in general. First, the increasing observations, which range from point-based (e.g., AmeriFlux https://ameriflux.lbl.gov/]) to remote sensing product (e.g., NASA MODIS [https://modis.gsfc.nasa.gov/], NASA SMAP [https://smap.jpl.nasa.gov/]), provide extensive training datasets to refine DL models that can be used as either data-driven modeling or error correction. In particular, these observations are usually available across different spatiotemporal scales. Future research can focus on how to efficiently utilize these heterogeneous data sources for multiscale modeling study. Second, open model simulation is another source for developing AI surrogate models. For instance, National Water Model (NWM, https://water.noaa.gov/about/nwm) regularly updates its reanalysis product for hydrological simulation across the CONUS. These model simulations can facilitate surrogate model development. Third, these days many datasets come online, including both remote sensing and NWM reanalysis products. Developing a systematic way to integrate these datasets into the E3SM using technology such as online learning would be a great opportunity.

Similar to surrogate modeling, hybrid modeling provides significant speedup for ensemble-based modeling and analysis. It would benefit the performance of ensemble-based forecasting (e.g., Kay et al. 2015; Weyn, Durran, and Caruana 2020), uncertainty quantification, and sensitivity analysis. In particular, the computation efficiency achieved by hybrid modeling can further facilitate uncertainty quantification on high-dimensional space. Another advantage of hybrid modeling is in multiscale modeling (Schneider et al. 2017). Instead of developing a big surrogate model for a physical model, which is extremely hard and at risk of losing the embedded physics, one can keep the dynamics of the scale of interest (e.g., high resolution) unchanged and replace the remaining (e.g., low-resolution) dynamics with an AI surrogate, such that both model stability and physics are obtained.

Advances in high-performance computing and deep learning tools provide opportunities for Earth scientists to develop a powerful, user-friendly software framework for conducting research on hybrid modeling. Such new software frameworks can leverage the following advances in deep learning and computer science in general. Automated ML (e.g., H2O's AutoML [https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html]) can be used to ease the development of an ML component such as hyperparameter tuning/ensemble machine learning. In addition, SmartSim (https://www.craylabs.org/docs/overview.html) and Fortran-Keras Bridge (https://github.com/scientific-computing/FKB) are powerful tools to connect machine learning codes with physics modeling codes programmed on legacy Fortran/C++ codes. Furthermore, the new framework should systematically archive existing modeling simulation (e.g., from E3SM) such that these simulations can be readily accessed and used for hybrid modeling development.

16.4 Research Priorities

16.4.1 Use In-situ Data to Improve Surrogates for Hybrid Modeling

Previous studies have developed emulators for complex climate models based on purely numerical simulations (e.g., Krasnopolsky, Fox-Rabinovitz, and Chalikov 2005; Wang, Balaprakash, and Kotamarthi 2019). This could be problematic, especially for small-scale and fast-moving processes, if the physical processes are still not or only partially resolved (e.g., the missing interaction between neighboring cells through cold pools; the three-dimensional nature of radiation) in these numerical models. These processes may be directly trained by in situ observations, remote sensing data, and super-resolution simulations. But much remains unknown about how much data from the past and present could be used to train models for future prediction and how to make sure these models conserve important physical properties. The use of observational data in developing these emulators will provide the ability to identify critical observations and develop numerical/observational and computational strategies to improve both their representation and the predictability of the model. Data examples include the Atmospheric Radiation Measurement's observatory and routine high-resolution modeling at the Southern Great Plains. For certain locations, the improvement of using both observations and model simulations most likely can also help the neighboring locations, for example, using advances in transfer learning and continual learning methods.

16.4.2 Explore Parameter Space for Optimizing Emulator's Performance and for Uncertainty Quantifications

This research priority is to better understand the source of model errors and uncertainties through perturbation of model initial conditions, physical parameterizations and parameters, and external forcings. This requires a fast, accurate, and stable modeling system that couples AI-based emulators with PDE-based dynamic solvers using advanced computational techniques (e.g., software integration, coupling, workflows) to achieve long time-period and/or large ensemble simulations. Such emulators need to be able to sample physical parameters (not hyperparameter, but the parameter used in the physics parameterizations) within the parameter space. It should also work on both regular grids as well as non-regular mesh like the E3SM or MPAS-Ocean models do (Shi et al. 2022) and should allow interactive post-hoc exploration and analysis, so the scientist can be analyzing the sensitivity of different parameters in real time during the training (He et al. 2020).

16.4.3 Continual Learning Hybrid Models

Most of the hybrid models that leverage AI/ML are developed for the given snapshot of training data. These models become less effective when the data on which they are trained on becomes less relevant due to distribution shifts, which often happens in the spatial-temporal climate data. To that end, we need to develop hybrid models that can quickly adapt to the change in the data. Specifically, continual learning algorithms seek to train models without losing the learning from the past (stability) and adapt rapidly to the new information from the newly collected data (plasticity). Hybrid models with continual learning capabilities are key to rapid training and fast adaptation.

16.4 4 Combine Online and Offline Training

While most of the emulators are developed and tested "offline," there is a need to plug these offline emulators back to the PDE solves, because offline training is not informed by physics constraints. Therefore, to make certain that the hybrid model can be stably run, approaches for effective online training are needed. There are studies that train the emulators "online" while coupled with the dynamical solvers. However, these studies use very simple models and ignore important aspects (e.g., cloud water and ice) for simplicity (Brenowitz and Bretherton 2019; Maulik et al. 2019). A new compute pipeline is needed that can help prevent instability due to separate compute flows of PDE versus AI for complex Earth system models, such as E3SM.

16.4.5 Automated Machine Learning

Design and development of AI/ML-based hybrid models open up several design choices and hyperparameters, which are often set manually and/or by using trial and error. Given the expected complexity and volume of the data, this manually intensive or ad hoc approach will not scale and will result in models with poor performance. A promising approach to overcome these issues is scalable automated machine learning. This will include a hyperparameter and neural architecture search to tune the hyperparameters of the models and the components of the neural architectures. In the context of hybrid models, we can expand these tuning to the components of the physical models and tune all of the components holistically. Another area where automated machine learning can help is to develop model ensembles. Specifically, automated machine learning can be used to develop an ensemble of hybrid models and use them for uncertainty quantification.

16.4.6 Software for Hybrid Model, Coupling between AI and PDE

To couple AI/ML-based emulators and complex PDE-based climate models and respond to the software architecture research challenges, software development is an urgent need. This includes, for example, the development of data adapters that allow in-memory, distributed data to be presented to the emulators; lightweight workflow-like components to distribute work on demand; and decision-tracking components to capture the data that go into and out of the emulator. The DOE Leadership Computing Facility is preparing for CPU and GPU hybrid supercomputers, such as the exascale Aurora, which could be beneficial for coupling AI and numerical simulations. One big advantage of replacing process models with ML-based emulators is that ML models tend to run efficiently on GPUs. The GPU will also significantly speed up data generation from numerical models that can provide the training data for AI emulator development.

16.4.7 MLOps for Hybrid Modeling

In traditional model/software development, the underlying physical phenomena that need to be modeled and implemented do not change rapidly; therefore, the software development, test, validation and verification, and scaling are often stable and static. This is often not the case in hybrid models because of the iterative nature of data-driven science, wherein the ML models drive the data collection, which in turn changes the models. Consequently, the ML models need to be constantly retrained, validated, verified, and deployed. MLOps (akin to DevOps in traditional software development) seeks to automate, orchestrate, and manage various stages of ML model development and deployment. MLOps applies to the entire hybrid model development lifecycle: data collection, data processing, feature engineering, data labeling, model design, model training, optimization, deployment, and monitoring.

16.5 Short-term (<5 years), 5-year, and 10-year Goals

Table 16-1 describes the short-term, 5-year, and 10+-year goals for several topic areas or targets. Specifically, advances expected in hybrid modeling algorithms, data, UQ, and software frameworks for simulations are all shown in a concise manner.

Table 16-1. Aspirational Goals for Hybrid Modeling for Earth System Predictions

Topic Areas / Target	Short-Term Goals	5-Year Goals	10+-Year Goals
Expected Algorithmic Speedup	10×	10× – 100×	1000×
Hybrid Modeling Algorithmic Advances	• PINNs and other SciML method development • Identifying reduced order hybrid modeling opportunities • Integrated framework to handle multiscale and cross-physics • Progress on stability, physics constraints • Formulating ML models so that projections are interpolated between training data • Offline training inference on need	• Formulating ML models for rare events (different distributions than training data); few-shot or zero-shot learning • Enhancing explainability of hybrid models; same problem as physics models • Stable, production hybrid modeling in climate science • Continuous learning / transfer learning for hybrid models • Identify data gaps based on hybrid models	• Automated model development and new physics discovery • Online/coupled, federated training on climatological timescales and close to data • Developing learning approaches to enable machine reasoning (> 10 years)
Data Advances	• Benchmark datasets for standardized comparison of methods and tools (MNIST for ESP) • Common data pre-processing for data engineering	• Smart practices for data augmentation, e.g., for extremes • Augments, large scale, production, open datasets for hybrid models	• Automatic data curation and pre-processing for multiple data streams • Build infrastructure to stream data into ESGF to build ML models.

Table 16-1. (Cont.)

Topic Areas / Target	Short-Term Goals	5-Year Goals	10+-Year Goals
UQ Advances	• Build UQ methods focused on hybrid physics + ML models • Identify the source of uncertainty and identify key gaps of uncertainty identification for UQ	• Scientifically useful uncertainty information in predictions	• Established mathematical UQ framework for hybrid physics and ML models, error propagation, optimal decision-making
Community Development	• Building new multidisciplinary communities around data, ESP, and ML models • Basic understanding of other areas within ESP	• Integrated teams of ML, mathematics, data, and domain scientists with knowledge in other areas • A pipeline of students, postdocs, and staff members with knowledge across areas of interest	• A tightly integrated AI4ESP community with a community-wide inter-disciplinary understanding of problems and methods in other areas
Software Frameworks, Infrastructure for Simulation Advances	• Infrastructure to allow data sitting next to the computer • Demonstration of methods with existing frameworks	• Learning with distributed physics/observation datasets • Workflow for MLOps and AI/ML-focused computing infrastructure • Development of new differentiable programming languages and packages with built-in support for scientific applications (ODEs, PDEs) and integration with DOE codes • Adjoint/derivative-enhanced training	• Integrated software frameworks for routine hybrid model training, coupling, validation, hardware acceleration • Differentiable Climate Modeling

16.6 References

Anandkumar, Anima, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Nikola Kovachki, Zongyi Li, Burigede Liu, and Andrew Stuart. 2019. "Neural Operator: Graph Kernel Network for Partial Differential Equations." In *ICLR 2020 Workshop on Integration of*

Deep Neural Models and Differential Equations. https://openreview.net/forum?id=fg2ZFmXFO3.

Bennett, Andrew, and Bart Nijssen. 2021. “Deep Learned Process Parameterizations Provide Better Representations of Turbulent Heat Fluxes in Hydrologic Models.” *Water Resources Research* 57 (5): e2020WR029328. https://doi.org/https://doi.org/10.1029/2020WR029328.

Brenowitz, Noah D, and Christopher S Bretherton. 2019. “Spatially Extended Tests of a Neural Network Parametrization Trained by Coarse-Graining.” *Journal of Advances in Modeling Earth Systems* 11 (8): 2728–44. https://doi.org/https://doi.org/10.1029/2019MS001711.

Häfner, Dion, René Løwe Jacobsen, Carsten Eden, Mads R. B. Kristensen, Markus Jochum, Roman Nuterman, and Brian Vinter. 2018. “Veros v0.1 – a Fast and Versatile Ocean Simulator in Pure Python.” *Geoscientific Model Development* 11 (8): 3299–3312. https://doi.org/10.5194/gmd-11-3299-2018.

He, Wenbin, Junpeng Wang, Hanqi Guo, Ko-Chih Wang, Han-Wei Shen, Mukund Raj, Youssef S G Nashed, and Tom Peterka. 2020. “InSituNet: Deep Image Synthesis for Parameter Space Exploration of Ensemble Simulations.” *IEEE Transactions on Visualization and Computer Graphics* 26 (1): 23–33. https://doi.org/10.1109/TVCG.2019.2934312.

Jiang, Peishi, Nis Meinert, Helga Jordão, Constantin Weisser, Simon Holgate, Alexander Lavin, Björn Lütjens, et al. 2021. “Digital Twin Earth – Coasts: Developing a Fast and Physics-Informed Surrogate Model for Coastal Floods via Neural Operators,” October. https://doi.org/10.48550/arxiv.2110.07100.

Karniadakis, George Em, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. 2021. “Physics-Informed Machine Learning.” *Nature Reviews Physics 2021 3:6* 3 (6): 422–40. https://doi.org/10.1038/s42254-021-00314-5.

Karpatne, Anuj, Gowtham Atluri, James H. Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, and Vipin Kumar. 2017. “Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data.” *IEEE Transactions on Knowledge and Data Engineering* 29 (10): 2318–31. https://doi.org/10.1109/TKDE.2017.2720168.

Kay, J. E., C. Deser, A. Phillips, A. Mai, C. Hannay, G. Strand, J. M. Arblaster, et al. 2015. “The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability.” *Bulletin of the American Meteorological Society* 96 (8): 1333–49. https://doi.org/10.1175/BAMS-D-13-00255.1.

Krasnopolsky, Vladimir M., Michael S. Fox-Rabinovitz, and Dmitry v Chalikov. 2005. “New Approach to Calculation of Atmospheric Model Physics: Accurate and Fast Neural Network Emulation of Longwave Radiation in a Climate Model.” *Monthly Weather Review* 133 (5): 1370–83. https://doi.org/10.1175/MWR2923.1.

Li, Zongyi, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. 2020. “Fourier Neural Operator for Parametric Partial Differential Equations,” October. https://doi.org/10.48550/arxiv.2010.08895.

Maulik, R., O. San, A. Rasheed, and P. Vedula. 2019. "Subgrid Modelling for Two-Dimensional Turbulence Using Neural Networks." *Journal of Fluid Mechanics* 858 (January): 122–44. https://doi.org/10.1017/jfm.2018.770.

Read, Jordan S, Xiaowei Jia, Jared Willard, Alison P Appling, Jacob A Zwart, Samantha K Oliver, Anuj Karpatne, et al. 2019. "Process-Guided Deep Learning Predictions of Lake Water Temperature." *Water Resources Research* 55 (11): 9173–90. https://doi.org/https://doi.org/10.1029/2019WR024922.

Schneider, Tapio, Shiwei Lan, Andrew Stuart, and João Teixeira. 2017. "Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations." *Geophysical Research Letters* 44 (24): 12,396-12,417. https://doi.org/10.1002/2017GL076101.

Shi, Neng, Jiayi Xu, Skylar W. Wurster, Hanqi Guo, Jonathan Woodring, Luke P. van Roekel, and Han-Wei Shen. 2022. "GNN-Surrogate: A Hierarchical and Adaptive Graph Neural Network for Parameter Space Exploration of Unstructured-Mesh Ocean Simulations," February. https://doi.org/10.48550/arxiv.2202.08956.

Tartakovsky, A M, C Ortiz Marrero, Paris Perdikaris, G D Tartakovsky, and D Barajas-Solano. 2020. "Physics-Informed Deep Neural Networks for Learning Parameters and Constitutive Relationships in Subsurface Flow Problems." *Water Resources Research* 56 (5): e2019WR026731. https://doi.org/https://doi.org/10.1029/2019WR026731.

Wang, Jiali, Prasanna Balaprakash, and Rao Kotamarthi. 2019. "Fast Domain-Aware Neural Network Emulation of a Planetary Boundary Layer Parameterization in a Numerical Weather Forecast Model." *Geoscientific Model Development* 12 (10): 4261–74. https://doi.org/10.5194/gmd-12-4261-2019.

Wen, Gege, Zongyi Li, Kamyar Azizzadenesheli, Anima Anandkumar, and Sally M. Benson. 2021. "U-FNO – An Enhanced Fourier Neural Operator-Based Deep-Learning Model for Multiphase Flow," September. https://doi.org/10.48550/arxiv.2109.03697.

Weyn, Jonathan A, Dale R Durran, and Rich Caruana. 2020. "Improving Data-Driven Global Weather Prediction Using Deep Convolutional Neural Networks on a Cubed Sphere." *Journal of Advances in Modeling Earth Systems* 12 (9): e2020MS002109. https://doi.org/https://doi.org/10.1029/2020MS002109.

Willard, Jared, Xiaowei Jia, Shaoming Xu, Michael Steinbach, and Vipin Kumar. 2020. "Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems," March. https://doi.org/10.48550/arxiv.2003.04919.

17 AI Architectures and Co-design

Authors: Jim Ang (PNNL), Simon Hammond (SNL), Maya Gokhale (LLNL), James Hoe (CMU), Tushar Krishna (GA-Tech), Sarat Sreepathi (ORNL), Matt Norman (ORNL), Ivy Peng (LLNL), Maruti Mudunuru (PNNL), Phil Jones (LANL), Mahantesh Halappanavar (PNNL)

17.1 Future System Concepts

Participants in the breakout room focus groups discussed several plausible future system concepts. The first group addressed the evolution of DOE's Leadership Computing Facility (LCF) systems for HPC and AI. These large-scale heterogeneous computing systems provide a foundation for advanced concepts with the potential to provide a radically different approach to future Earth system modeling and AI-enabled integration with measurement and observation data.

17.1.1 Centralized Large-Scale HPC Concept

The baseline system concept is the future evolution of large-scale HPC and Cloud computing systems. This next step will extend post-exascale architectures beyond the first generation of DOE exascale heterogeneous systems that integrate CPUs and GPUs. As the HPC and Cloud computing communities increasingly rely on hardware specialization to improve performance, co-design approaches will support the development of accelerators for frequently used kernels in scientific modeling and AI/ML methods. New specialized accelerators may arise to support additional data science capabilities such as uncertainty quantification, streaming analytics, or graph analysis. These future large-scale computing systems with extreme heterogeneity will need to be co-designed to support the increased computational and dataset sizes associated with Earth science predictability and scientific machine reasoning.

17.1.2 Edge Sensors with Centralized HPC/Cloud Resources Concept

In the second system concept, environmental data are recorded from a broad collection of sensors spread across the globe and located at points of interest. The sensors are designed to monitor specific items of interest (e.g., river flow, temperature, light, etc.) and to communicate these data back to a centralized location. At this centralized facility, large HPC or Cloud computing environments will process the incoming data streams for integration into online simulations of extreme weather events, climate, hydrology, etc.

AI/ML capabilities could be utilized at multiple points within this system concept. First, the excessive volume of data coming into the system will have the potential to overrun even the largest processing capabilities and is unlikely to be able to be stored in memory or even temporary storage resources (such as file systems or object stores). AI/ML models could be trained and tailored to either summarize or select relevant features from the incoming data streams so that a significantly reduced amount of data needs to be kept and integrated into ongoing simulations. One other potential is for AI/ML models to identify anomalies from the incoming data streams that might suggest areas of interest for simulations to be focused on – for instance, the start of a hurricane, or the high likelihood of significant snowfall.

Due to the distributed nature and inhospitable environments where sensors may need to be placed or roam, it is unlikely that a reliable stream of data will reach the centralized location for all possible inputs. One common use case is the smart city scenario, Figure 17-1, that describes a wide range of sensor, computing, and data storage capabilities. AI/ML models could be used in such an environment to help patch measurement gaps and present a more consistent view of real-world data to a future simulation run on a large-compute resource.

Figure 17-1. A Smart City scenario with a large number of sites for fixed sensor deployments, e.g., temperature, wind, CO_2, precipitation, etc., plus a variety of mobile devices that can also be used to intermittently augment collection of measurement and observation data. An urban setting will support advanced wireless communications like 5G and eventually 6G (Source: Pacific Northwest National Laboratory).

17.1.3 Federated Processing from the Edge to the Data Center Concept

The third potential system design extends the second concept by leveraging a much higher degree of processing located either in, or near to, the distributed sensor network. In this model, sensor data can be processed either directly on the sensor itself or in a nearby edge server with processing elements that may stream a small collection of sensor data into it. Local processing stations can then send on either their raw or locally processed data to a centralized HPC and/or Cloud resource for inclusion in simulation models and centralized AI/ML models as in the first system concept.

The advantage of this approach is that data down-selection and feature extraction can be performed locally, significantly reducing the volume of data that must be transmitted to a centralized resource. Assuming that a sufficiently performant local network among sensors can be established, model parameters and partial results, perhaps even AI/ML model updates can be exchanged within a locale, allowing for a truly federated aspect to the design.

Initially, this concept takes advantage of existing gateways and local area networks serving sensors in the field. Through co-design collaborations, it is possible to expand that service to include application/sensor-specific processing to filter, analyze, compress, encrypt, and unify multiple sensor streams transmitting measurements through the network.

17.1.4 Dynamic and Adaptive Federated Processing Concept

The last system concept builds on the previous three by augmenting feedback and control paths within distributed networks of sensor-local resources. Local control offers lower latency decision-making to dynamically control what information is observed, measured, recorded, and relayed by the sensor network. Such a design has powerful implications: by dynamically controlling sensors online, simulations of the Earth's weather and climate can essentially focus sensor inputs on specific quantities or geographic locations of interest. Examples might include where severe weather events are expected or whether climate scientists identify where specific information is needed to help improve the quality of their models. This concept expands to potentially multiple HPC and/or Cloud data centers for federated AI/ML modeling. AI/ML models can play a crucial part in this system by performing continuous, autonomous online inspection of evolving simulations or of recorded data to identify areas of data insufficiency or statistical weakness. Furthermore, a dynamic and adaptive system may be able to carefully obtain and select data to improve the quality of its training, reducing the need for vast, potentially intractable, datasets to be collected over long periods.

17.2 Grand Challenges

These two system concepts that integrate federated processing are beyond the capabilities of affordable technologies today and will require a significant investment both in foundational technology systems and co-design programs so that climate scientists, mathematicians, AI/ML experts, computer scientists, and hardware engineers can collaborate to balance the competing performance, energy, cost, and security challenges associated with AI/ML-accelerated Earth system modeling and observation/measurement capabilities.

Significant technical challenges will arise in the following areas:

17.2.1 Programmability and Usability

The current and near-term challenge is the integration of scientific modeling and simulation applications with AI/ML methods. This drives the need to integrate Earth system HPC applications written in C/C++ and/or Fortran with AI/ML methods that use Python ML frameworks. Programming models are under development to support the convergence of applications and workflows onto heterogeneous computing systems. Many AI/ML architectures provide hardware support for reduced or mixed precision, and tools will be required to analyze which specific model components can use these capabilities. Protocols and tools for ESP data-sharing and data federation on Cloud must be created. The usability challenge is to manage the complexity of mapping converged application workloads to future heterogeneous computing architectures that will integrate specialized hardware accelerators with commodity CPU/GPU processors.

Domain scientists[1] are interested in exploring the capabilities of new heterogeneous advanced architecture computing systems, but there are challenges to understand how to map AI4ESP workflows to the diverse collection of computing system options. It is especially important to understand how AI/ML capabilities originally developed for generic commercial workloads may or may not be applicable for ESP hybrid modeling applications, or observation and measurement capabilities. From centralized large-scale modeling and training to edge computing inferencing and federated learning, new challenges arise for composition and distribution of applications, algorithms, and methods. This is an important opportunity for the AI4ESP community to develop a new generation of proxy applications and benchmarks for both modeling and observation capabilities[2] and to facilitate communication and co-design collaborations with hardware designers, system software developers, algorithm developers, and domain scientists.

17.2.2 Data Movement

The expected volume of data associated with a full, coordinated Earth sensor capability will be unprecedented. Not only will such a network generate a previously unimaginable quantity and diversity of data, but the computing and network load for processing, transmitting, and subsequent storage of this volume will be orders of magnitude higher than any system available today. Data movement costs in terms of energy and latency motivates the interest in federation and distribution of computing across the AI4ESP scientific ecosystem. AI/ML technologies could help reduce such volumes through identification of patterns and anomalies, and in summarization of sub-volumes. Significant investment will be required in AI/ML approaches to ensure that the modeling capabilities will be compatible and efficient for the types of data being recorded, especially where this may deviate from commercial photo or video capabilities. Technologies that may assist in energy-efficient data transfers could include investment in silicon photonic network capabilities, as well as wide-area 5G- or 6G-like communication networks enabling sensors to communicate over short/medium distances without the need for physical wiring. On the storage side, Cloud technologies such as high-performance, large-volume data object stores could likely provide a capability to address increased sparse data storage volumes, although at present this would pose a significant cost barrier using contemporary commercial Cloud pricing. AI/ML may also be used to enable smart compression

---

[1] Interfacing with sensors and AI analytics at the Edge will allow domain scientists to extract actionable information needed for improved modeling of disturbances and extreme events. This type of co-design is needed for most ESP applications. For example, watershed science, hydrology, ecohydrology, climate variability and extremes, aerosols and clouds, and atmospheric modeling are cross-cutting themes where AI@Edge has the highest impact. Co-design approaches that interface with distributed sensor networks will allow us to (1) collect reliable and relevant watershed data under disturbances, (2) monitor land-atmosphere-coastal interactions by embedding intelligence on the Atmospheric Radiation Measurement (ARM) instruments, (3) understand wildfire events and their impact on ecosystems in near-real-time, and (4) assess critical infrastructure impacted by extreme events (e.g., see Human Systems and Dynamics, chapter 9).

[2] For example, AI-enabled co-design will enable us to emulate and deploy DOE codes such as PFLOTRAN, ATS, and E3SM at sensor edge for empowering ARM instruments and EMSL user facilities.

techniques on Earth system data to increase information density without increasing storage costs. Additionally, DOE HPC centers could incorporate concepts and methods from Cloud storage systems into future parallel file and storage systems to slowly move toward such capability.

17.2.3 Energy Efficiency

Large-scale networks with integrated sensors and federated processing, as well as wide area communication networks to handle data transmissions, are likely to be very expensive in terms of energy consumption. While this was a lower-priority focus for exascale computing, processing and communication of data remain power-expensive operations. Co-design has the potential to help improve this situation through the use of novel materials, devices, and processing techniques (e.g., neuromorphic-based accelerators to analyze images/video). However, significant investment will still be required in foundational technologies if large-scale, power-efficient sensing networks are to be realized. Co-design to balance performance and energy efficiency will also address key questions of how the modeling, machine learning, uncertainty quantification, and other streaming analytics capabilities are partitioned across the AI4ESP scientific ecosystem that integrates DOE's heterogeneous HPC systems with Cloud computing, edge servers, and sensors with IoT devices.

17.2.4 Privacy and Security of Data

As Earth systems modeling becomes increasingly integrated with a distributed network of observation and measurement sensors, and perhaps federated processing capabilities, the quality, accuracy, and robustness of the information will become more important. This is increased significantly if the information generated from modeling and measurement capabilities is used to support high-consequence national or international scientific policy decisions. The implications of potential data tampering or nefarious modification are clear: a national or international resource for accurate scientific prediction could be severely affected. As well as the security implications of such a system, data privacy concerns will also need to be addressed. This is particularly true of a system where individual human subject images or videos may be captured, or where their behavior could be discerned from the data. Imagine sensor capabilities that were able to identify patterns such as individual schedules going to and from work, etc. Co-design has a potential role to play in this space—by including security experts in cyber-physical designs from the outset, secure data transmission and processing can be integrated as a first-level citizen rather than as a later, software-derived additional layer. In addition, data privacy may be afforded if local artifacts associated with specific individuals can be aggregated into a larger, federated model with individual patterns obfuscated or redacted into the full model of the system.

17.3 State of the Science

17.3.1 DOE HPC

The HPC user facilities build on DOE's investments in exascale computing. The first generation of these systems is based on the integration of CPU and GPU processors into heterogeneous systems. The impact on AI4ESP is through DOE's development of scientific modeling capabilities to achieve energy-efficient performance on GPUs, while also leveraging the commercial drivers for GPU-based AI/ML performance. With the slowing of Moore's Law, the computing community recognizes the increased need for architectural specialization, so the next generation of HPC systems is likely to extend beyond the current designs for CPU/GPU heterogeneity. DOE's efforts in AI for Science are exploring capabilities that provide a foundation for integration of HPC applications with data science and AI/ML frameworks.

17.3.2 Cloud Computing

While Cloud providers have user-friendly tools to run AI/ML workloads, there is incompatibility among tool capabilities and user interfaces among different providers that make it difficult to achieve interoperability in a federation of Clouds. While some ESM data are presently stored on Cloud storage systems, the data stores are associated with a patchwork of individual groups and projects, lacking a federated view. For data storage, cloud providers can presently accommodate petabytes to exabytes of data. The commercial cloud cost is based on accessing and computing or analyzing the data, and can become an extreme cost if data transmission into/out of the Cloud becomes a frequent operation. Commercial AI/ML cloud infrastructure and services are predominately motivated by text and image data. Cloud providers have demonstrated AI-at-Scale for these applications. The largest NLP models approaching of 1 trillion parameters has been demonstrated on Selene (the #6 machine on the November 2021 Top 500 list). Workflow services exist on the cloud for certain applications, including many AI/ML methods, and raw materials are available on cloud platforms to create more complex workflows. ESM workflows currently do not exist that combine external data sources or coordinate with HPC simulation. Computer science expertise would be required to create such workflows in a form suitable for domain scientists.

17.3.3 Edge Computing

There is a broad portfolio of AI methods for classifying patterns, anomaly detection, unsupervised learning for data compression, inference at the edge, and continuous learning with streaming sensor data. The integration of edge computing sensors has a number of distinct deployment scenarios including NOAA and NASA Earth-observing satellite imagery with edge processing in space or at dedicated ground stations. Edge computing can also be integrated with the diverse collection of distributed sensors that collect observation and measurements for the

DOE's Atmospheric Radiation Measurement (ARM) user facility. Adaptive sensors with embedded hardware accelerators are now emerging. New concepts for distributed applications are also under development such as geomorphic computing where weather research and forecasting models are distributed, federated, and able to dynamically adapt to the environment.

17.4 Experimental, Data, and Modeling Opportunities

17.4.1 DOE HPC

There is an immediate opportunity to support scientific machine learning with higher level scientific application language APIs to allow domain scientists to develop and deploy AI4ESP modeling capabilities. Leveraging the experiences and expertise in DOE exascale co-design, there are opportunities to develop a new generation of mini-apps and benchmarks that reflect the needs of AI4ESP workflows from both the modeling and observation/measurement perspectives. AI architecture advances are driving the development of new, lower-precision computer number formats and high arithmetic intensity tensor operations. While clearly targeting deep learning methods that can be directly applied to ESP measurement data, there are also opportunities to develop tools to map ESP modeling and simulation capabilities to these advanced architectures.

17.4.2 Cloud Computing

Cloud can provide virtually unlimited compute cycles, enabling collections of individual simulations running concurrently on the distributed network. Use of the Cloud in ESM may facilitate data sharing due to ease of the merging of distributed data sources. Most commercial Cloud providers offer HPC environments with dedicated servers. Cloud resources include a variety of heterogeneous compute resources with CPUs, GPUs, and other accelerators. However, it should be noted that the cost model of commercial Cloud providers is based on use of the compute resources, and thus the cost for analyses and simulations may be potentially unbounded. An attractive direction for the future is for DOE facilities to offer Cloud environments within the HPC data center to accommodate Cloud storage models, workflow, and job scheduling co-existing with traditional HPC approaches. Commercial Cloud providers also have pioneered support for integration of data center computing with a diverse portfolio of commercial Internet of Things (IoT) devices like smart speakers, thermostats, and other consumer home appliances. There is a significant opportunity for the ESM community to identify and prioritize how to leverage this commercial IoT infrastructure to support scientific observation and measurement instruments.

17.4.3 *Edge Computing*

The most immediate opportunity for edge computing is the integration of processing with sensors to integrate inferencing, streaming analytics, or other data science capabilities to enhance the observations and measurements for Earth sciences. Edge computing can also provide local control of sensors to support real-time, autonomous sensor control to improve the quality and value of measurements for subsequent data assimilation workflows. There are a number of interesting co-design opportunities associated with the partitioning of computing workflows that can be distributed among DOE HPC user facilities, commercial Cloud computing data centers, and these distributed edge sensing instruments. The co-design of this modeling and experiment approach could also apply AI/ML to optimize some of the data communication challenges present in the integration of large-scale models/simulations with smart sensors. The data movement and energy efficiency challenges may also drive innovative approaches to distribute weather simulations across the AI4ESP scientific ecosystem that could integrate large-scale E3SM models with widely distributed "E3SM-Lite" edge computing simulations.

17.5 Research Priorities: Short-term (<5 years), 5-year, and 10-year Goals

17.5.1 *Short Term*

Short-term goals include efforts to:

- Develop a suite of AI4ESP workflows and benchmarks that share common tools and building blocks with commercial AI/ML activities for economy of scale.
- Develop network simulators to provide quantitative analysis of distributed AI4ESP workflows that integrate computing and data capabilities for HPC + Cloud + Edge.
- Develop high-level frameworks and APIs that support convergence of scientific modeling (e.g., PFLOTRAN, ATS, E3SM) and measurement with AI/ML and UQ methods.
- Develop service-level abstractions to support scientists by hiding the scale and complexity (heterogeneity) of the underlying resources and execution on a range of different systems from DOE LCFs, to Cloud computing, to Edge computing capabilities for ARM instruments.
- Support much closer communication and collaboration of domain scientists with data scientists, computer scientists, applied mathematicians, and computer architects.

17.5.2 *Medium Term*

Medium-term goals include efforts to:

- Create data formats and structures to support interoperable, federation, and data sharing.
- Understand how AI/ML can define better metrics for model and sensor uncertainty quantification.
- Create AI/ML surrogates from high-resolution simulation and data.
- Develop automated control of edge sensors.

17.5.3 Long Term

Long-term goals include efforts to:

- Improve efficiency by co-design across many currently disparate technical domains (ML/traditional, compute/storage/communication, SW/CPU/GPU/accelerator, and HPC+Cloud+Edge.
- Demonstrate AI-at-scale opportunities for the envisioned concepts (HPC+Cloud+Edge) with secure storage, authentication, and provenance.
- Develop federated learning that integrates streaming analytics and AI at edge sensors with HPC modeling.

17.5.4 Vision of Co-design Opportunities

- Anomaly analysis for weather and climate VERY early/extreme weather, real-time analysis needed, early prediction?
    - Wildfire predictions, predict propagation of fires, then direct resources to areas of concern.
    - Hurricane tracking/prediction: More accurate determination of evacuation zones, pre-positioning of infrastructure restoration assets.
- Climate simulations have a longer timeframe, need archival records, larger datasets (long time series)? Can AI help fill in gaps in data, limited data/missing, etc.?
- Create an AI4ESP Digital Twin of the planet.
- Partitioning of processing and data capabilities over widely distributed AI4ESP scientific ecosystem that spans DOE supercomputers, Cloud computing and data centers, Edge servers, and sensors with integrated IoT processor capabilities.

Appendix A: Acronyms

1D, 3D, 4D	one-, three-, four-dimension(al)
ABM	agent-based modeling
AI	artificial intelligence
AI2	Allen Institute for Artificial Intelligence
AI4ESP	Artificial Intelligence for Earth System Predictability
ALE	Accumulated Local Effect
ANN	artificial neural network
API	application programming interface
ARD	ARM Data Center
Argonne	Argonne National Laboratory
ARM	Atmospheric Radiation Measurement
ASCR	Advanced Scientific Computing Research (DOE Office of Science)
ASU	Arizona State University
BER	Biological and Environmental Research (DOE)
BGC	biogeochemistry
BNL	Brookhaven National Laboratory
CAM	community atmospheric model
CESM	Community Earth System Model
CI	cyberinfrastructure
CIRES	Cooperative Institute for Research in Environmental Sciences
CMIP	Coupled Model Intercomparison Project
CMU	Carnegie-Mellon University
CNN	convolutional neural network
CPU	central processing unit
CRM	cloud-resolving model
CU	Colorado University
DA	data assimilation
DAAC	Distributed Active Archive Center (NASA)
DAQ	data acquisition
DDROM	data-driven reduced order modeling
DL	deep learning
DNN	deep neural network
DNS	direct numerical simulation
DOE	U.S. Department of Energy

E3SM	Energy Exascale Earth System Model
EES	Earth and Environmental Systems
EESSD	Earth and Environmental Systems Science Division
eMBB	enhanced mobile broadband
EMSL	Environmental Molecular Science Laboratory
ENSO	El Niño Southern Oscillation
ERF	effective radiative forcing
ESGF	Earth System Grid Federation
ESM	Earth system model
ESP	Earth system predictability/prediction
ESS	Earth system science
ESS-DIVE	Environmental Systems Science Data Infrastructure for a Virtual Ecosystem
ET	evapotranspiration
FACE	Free Air Carbon Dioxide Enrichment
FAIR	findable, accessible, interoperable, and reproducible
GAN	generative adversarial network
GC	Grand Challenge
GCAM	Global Change Analysis Model
GCM	global climate model
GHG	greenhouse gas
GP	Gaussian process
GPP	gross primary productivity
GPU	graphical processing unit
GSA	global sensitivity analysis
HPC	high-performance computing
I/O	input/output
IAI	interpretable AI
ILAMB	International Land Model Benchmarking
IoT	Internet of Things
IPCC	Intergovernmental Panel on Climate Change
JPL	Jet Propulsion Laboratory (NASA)
KIML	knowledge-informed machine learning
LANL	Los Alamos National Laboratory

LBL, LBNL	Lawrence Berkeley National Laboratory
LCF	Leadership Computing Facility (DOE)
LES	large eddy simulation
LIME	local interpretable model-agnostic
LRP	Layerwise Relevance Propagation
LSTM	long short-term memory
LTER	Long-Term Ecological Research
MJO	Madden Julian Oscillation
ML	machine learning
MLP	multilayer perceptron
MOCU	mean objective cost of uncertainty
ModEx	MOdel Driven EXperiment
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research (NSF)
NCICS	North Carolina Institute for Climate Studies
NCSU	North Carolina State University
NEON	National Ecological Observatory Network
NERSC	National Energy Research Scientific Computing Center
NETL	National Energy Technology Laboratory
NeurIPS	Neural Information Processing System
NGEE	Next-Generation Ecosystem Experiment
NLP	natural language processing
NN	neural network
NOAA	National Oceanic and Atmospheric Administration
NSF	National Science Foundation
NWM	National Water Model
ODE	ordinary differential equation
ODML	on-demand machine learning
ORNL	Oak Ridge National Laboratory
OSSE	Observing System Simulation Experiment
PDE	partial differential equation
PE	parameter estimation
PINN	physics-informed neural network
PNNL	Pacific Northwest National Laboratory
PPE	perturbed physics ensemble
PSL	Physical Sciences Laboratory (NOAA)

PSU	Pennsylvania State University
PUBS	prediction in unmonitored basins
QA/QC	quality assurance and quality control
QoI	quantity of interest
R&D	research and development
RF	random forest
RNN	recurrent neural network
ROI	return on investment
ROM	reduced order modeling
RS	remote sensing
SAR	synthetic aperture radar
SFA	Science Focus Area
SH	sensible heat
SHAP	Shapley
SNL	Sandia National Laboratories
SMAP	soil moisture active passive
SOC	soil organic carbon
SPEEDY	Simplified Parameterization, primitive Equation Dynamics
SPRUCE	Spruce and Peatland Responses Under Changing Environments
SSL	self-supervised learning
SST	sea surface temperature
SVM	support vector machine
TC	tropical cyclone
UAV	unmanned aerial vehicle
UC	University of California
UIUC	University of Illinois at Urbana-Champaign
UQ	uncertainty quantification
USACE	U.S. Army Corps of Engineers
USGS	U.S. Geological Survey
UT	University of Texas
VAE	variational autoencoder
WHONDRS	Worldwide Hydrobiogeochemistry Observation Network for Dynamic River Systems
XAI	explainable AI

Appendix B: Workshop Agenda

Video and additional information is linked on the online version of the agenda.
Times Are North American Eastern Time Zone

October 25: Day 1 (Week 1)

12:00 Welcome - Nicki Hickmon
Deputy Director for Science Programs, DOE Office of Science - Harriet Kung
Introduction to AI4ESP initiative - Nicki Hickmon
Earth & Environmental Systems Sciences Division (EESSD) - Gary Geernaert
Advanced Scientific Computing Research (ASCR) - Barb Helland
13:15 AI4ESP Workshop Structure & Charge - Haruko Wainwright
AI4ESP State-of-the-Science - Haruko Wainwright, Forrest Hoffman, Scott Collis
14:00 Break
14:15 Panel Discussion
Panel Chair: Rick Stevens
Panel: Grace E. Kim, Prabhat Ram, Kirk Borne
15:00 Earth System Predictability Session - Atmospheric Modeling (Invited Only)
Session Chair: Ruby Leung - Presentation
17:00 Adjourn

October 26: Day 2 (Week 1)

12:00 Plenary talk - Amy McGovern
12:15 Plenary talk - Pierre Gentine
12:30 Break
12:45 Earth Science Topic Session - Land Modeling (Invited Only)
Session Chair: Beth Drewniak
14:45 Break
15:00 Cross-cut Session - Data Acquisition to Distribution (Invited Only)
Session Chair: Giri Prakash
17:00 Adjourn

November 1: Day 3 (Week 2)

12:00 Reports From Previous Sessions
12:30 Break
12:45 Earth Science Topic Session - Human Systems & Dynamics (Invited Only)
Session Chair: Christa Brelsford
14:45 Break
15:00 Earth Science Topic Session - Hydrology (Invited Only)
Session Chair: Charuleka Varadharajan
17:00 Adjourn

November 2: Day 4 (Week 2)

12:00 Plenary talk - Chaopeng Shen
12:15 Plenary talk - Rob Ross
12:30 Break
12:45 Earth Science Topic Session - Watershed Science (Invited Only)
Session Chair: Mavrik Zavarin
14:45 Break
15:00 Cross-cut Session - Neural Networks (Invited Only)
Session Chair: Auroop Ganguly
17:00 Adjourn

November 8: Day 5 (Week 3)

12:00 Reports From Previous Sessions
12:30 Break
12:45 Earth Science Session - Ecohydrology (Invited Only)
Session Chair: Forrest Hoffman
14:45 Break
15:00 Cross-cut Session - Surrogate Models & Emulators (Invited Only)
Session Chair: Nathan Urban
17:00 Adjourn

November 9: Day 6 (Week 3)

12:00 Plenary Talk - Tapio Schneider
12:15 Plenary Talk - Alison Appling
12:30 Break
12:45 Earth Science Session - Aerosols & Clouds (Invited Only)
Session Chair: Po-Lun Ma
14:45 Break
15:00 Cross-cut Session - Knowledge-Informed Machine Learning (Invited Only)
Session Chair: Frank Alexander
17:00 Adjourn

November 29 Day 7 (Week 4)

12:00 Reports From Previous Sessions
12:30 Break
12:45 Earth Science Session - Coastal Dynamics, Oceans & Ice (Invited Only)
Session Chair: Matt Hoffman
14:45 Break
15:00 Cross-cut Session - Knowledge Discovery & Statistical Learning (Invited Only)
Session Chair: Xingyuan Chen
17:00 Adjourn

November 30: Day 8 (Week 4)
12:00 Plenary talk - Laure Zanna
12:15 Plenary talk - Katie Dagon
12:30 Break
12:45 Earth Science Topic Session - Climate Variability & Extremes (Invited Only)
Session Chair: Maria Molina
14:45 Break
15:00 Cross-cut Session - Explainable/Interpretable/Trustworthy AI (Invited Only)
Session Chair: Line Pouchard
17:00 Adjourn

December 2: Day 9 (Week 5)
12:00 Reports From Previous Sessions
12:30 Break
12:45 Cross-cut Session - Hybrid Modeling (Invited Only)
Session Chair: Sivasankaran Rajamanickam
14:45 Break
15:00 Cross-cut Session - AI Architecture Co-Design (Invited Only)
Session Chair: Jim Ang
17:00 Adjourn

December 3: Day 10 (Week 5)
12:00 Reports From Previous Sessions
12:15 Workshop session wrap-up and discussion motivation
12:30 Break
12:45 Panel/Open Discussion (Invited Only)
Common challenges & opportunities
Resources, capabilities, and facilities (DOE + Multi-agency)
14:45 Break
15:00 Panel/Open discussion (Invited Only)
Short-term, 5-year, 10-year goals
Earth system predictability and applied math and computer science research priorities
17:00 Adjourn

December 6
12:00 - 16:00 Editing Meeting (Invited Only)

December 20
12:00 - 16:00 Editing Meeting (Invited Only)

Appendix C: Call for AI4ESP Papers

White Paper Purpose, Structure, and Submittal

Purpose

Submitted white papers will be used to inform the design of three sequential workshops (conducted in 2021-2022) focused on answering the following overarching question of: How can DOE directly leverage artificial intelligence (AI) to engineer a substantial (paradigm-changing) improvement in Earth System Predictability?

Structure of white papers

White papers should be prepared using the following outline and may be up to a maximum of 3 pages long (12-point font, not including the optional Suggested Partners and References sections). Use of the template provided is optional.

1. **Title**
2. **Authors/Affiliations:** List in order of largest contribution
3. **Focal Area(s):** One or two sentences only
4. **Science Challenge:** Short statement describing the area addressed by the white paper
5. **Rationale**: Description of the research needs/gaps, the barriers to progress, and the justification for and benefits associated with the proposed approach
6. **Narrative**: Scientific and technical description of the opportunities and approach; activities that will advance the science; and specific field, laboratory, model, synthesis, and/or analysis examples
7. **Suggested Partners/Experts (Optional)**: Laboratory and university partners or experts in the field who may be able to present a related webinar or plenary presentation at a workshop
8. **References (Optional)**

Authors are limited to two submissions as lead author but may participate as a co-author in other submissions. Teaming is encouraged to reduce the reviewing workload. Multi-institutional responses are welcome; however, a clear lead who can speak authoritatively on the white paper contents should be identified. [Note: Protected information should not be included in white papers, but instead should be shared directly with the appropriate U.S. Department of Energy (DOE) program manager(s).]

Submittal

White papers must be submitted as PDF files by **5:00 p.m. EST on February 15, 2021**, using this Google Form. After the submission date, white papers will be posted publicly on a website (active Feb 2021) and be made available in advance of the workshops for review by participant and general public engagement.

Questions: Prospective authors are welcome to seek clarification on any part of this announcement through the AI4ESP slack workspace #white-paper-call-questions slack channel.

Background

Throughout its history, the U.S. Department of Energy (DOE) has tackled some of the world's most difficult scientific and technical challenges. In response to changing scientific needs, DOE's national laboratories manage some of the most sophisticated facilities and observatories on the planet; develop multi-scale, multi-physics Earth system models (ESMs); and apply artificial intelligence and machine learning, advanced visualization, and cutting-edge computing assets in innovative ways to solve scientific challenges. Major scientific breakthroughs are seeded when individual and distinct investments can be integrated in novel ways, for example, where facilities, models, experiments, and artificial intelligence (AI) are partnered.

Nearly a decade ago, DOE recognized that an acceleration was needed in the transition of basic science into new predictive capabilities to meet the needs of scientists and stakeholders. To help meet this challenge, DOE's Earth and Environmental System Sciences Division (EESSD) incorporated a novel model-experiment (ModEx) approach, linking interdependent observation and model development into its management philosophy and strategic planning (Figure C-1).

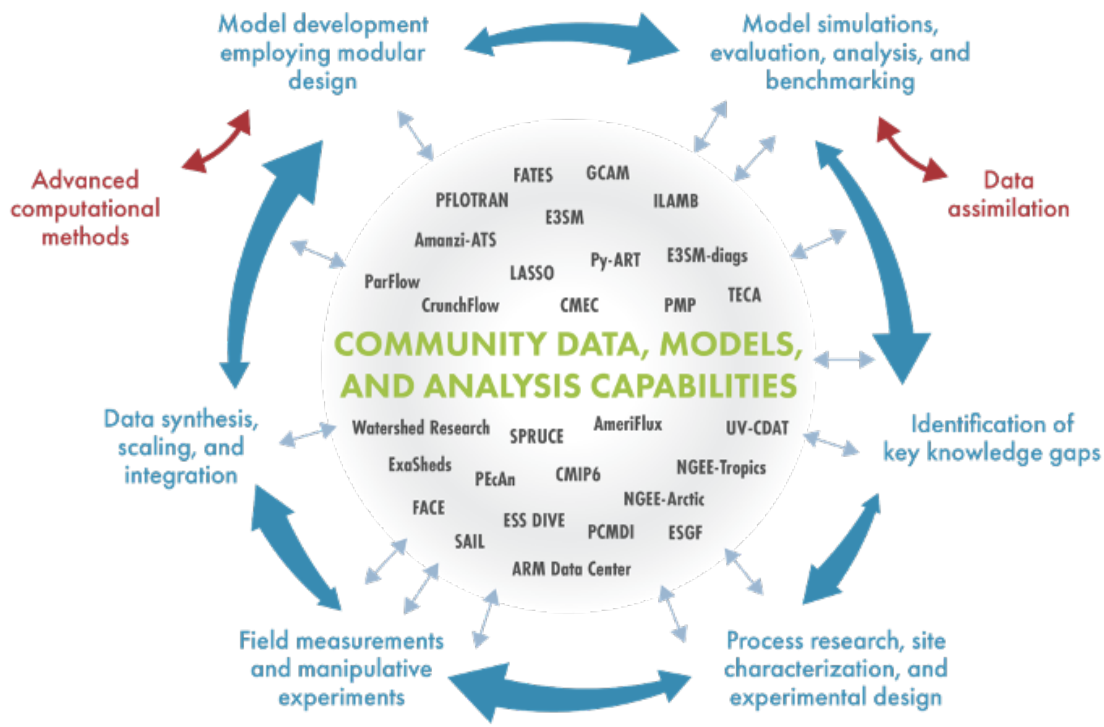


Figure C-1. Schematic of the ModEx approach to scientific discovery (outer ring) and various DOE data, models and analysis capabilities that should be linked as community resources based on open science principles (inner sphere).

Despite advances in high-resolution modeling and better observational capabilities, scientific and decision-making communities have increasingly sought predictive capabilities that exceed current knowledge. Some examples include an urgent need for more accurate prediction of extreme events, more complete characterization of uncertainty in models and data necessary to constrain scientific findings, and effective bridges between observational designs and useful predictions.

To meet these pressing needs, a paradigm shift is necessary to build the integrative research framework of the future. Building such a framework will require attention to the design of integrative research approaches that take advantage of recent scientific and technological advances – such as artificial intelligence and exascale computing, etc. – that are not widely incorporated in EESSD research. A successful shift has the potential to result in an entirely new framework that will integrate new observational strategies with capabilities in automated data quality validation; edge computing; new nonlinear and multiscale data assimilation methodologies; model parameter estimation and feature detection using AI; and hybrid prediction models that combine physics with AI.

The hierarchies of models that must be considered in such a paradigm shift include the hybridization of one-dimensional and multi-dimensional ESMs, large eddy simulations, and agent-based models; AI-generated surrogates for such models; and scale-aware, AI-based analytics to complement traditional physical approaches to evaluating uncertainties. With advanced computing and AI as leadership disciplines within DOE, we envision new and heretofore unforeseen possibilities for EESSD to use these new frameworks to discover next generation science, revolutionizing our Nation's predictive capabilities and advancing its scientific agenda.

This exercise to explore a paradigm shift in prediction science follows the July 2020 release of the AI for Science technical report, prepared by a consortium of DOE national laboratories to identify scientific opportunities for AI in the upcoming decade. Two weeks later, the White House issued a joint OSTP-OMB President's S&T memorandum, that highlighted Earth system predictability as a national science priority and identified AI and edge computing as areas that the federal agencies should continue to develop. More recently, DOE collaborated with other agencies to identify science capabilities that the research community could rapidly and aggressively apply to advance prediction science in Earth science research.

The DOE vision is to radically improve predictive capabilities by applying AI methods to build a new integrative system that spans the continuum from observations to predictive modeling. This effort will require the exploration of AI across the ModEx enterprise (Figure C-1) to determine the most impactful applications along the observation-modeling continuum.

Call for White Papers

All interested researchers are asked to read the current version of the Earth and Environmental Systems Sciences Division (EESSD) Strategic Plan. This plan enumerates five grand challenges that frame the Division's investments through 2023: integrated water cycle, biogeochemistry, high-latitude science, drivers and responses, and data-model integration. Since the development of the plan, the following cross-cutting areas of interest have emerged: predictability of extreme events, terrestrial aquatic interfaces, regions of high gradients (e.g., coastal zones and watersheds), and integration of AI and other new technologies into scientific research.

The purpose of this announcement is to solicit white papers from the scientific community that focus on development and application of AI methods in areas relevant to EESSD research with an emphasis on quantifying and improving Earth system predictability, particularly related to the integrative water cycle and associated water cycle extremes. White papers must describe novel and innovative approaches to improving the predictability of the Earth system and explain why these approaches are expected to succeed. We expect that a novel framework, derived from white paper concepts and a series of workshops, will improve capabilities for knowledge capture and distillation that provide future computational constructs across the EESSD research enterprise. Authors should consider novel approaches, needed resources, technologies available and/or in the pipeline, and unforeseen developments that have the potential to transform the Earth system research enterprise out to FY2030 and beyond. We anticipate that exascale computing, edge computing, 5G/6G, and use of quantum computing and quantum sensors will be further developed during this period.

White papers should adhere to the following criteria and constraints:

- White papers should identify a transformational science question that involves the water cycle and associated water cycle extremes as a centerpiece in conjunction with the EESSD's cross-cutting areas described above. This science question may include, for example, extreme precipitation, extreme drought, strong perturbations on surface water or groundwater systems, and/or extreme flooding or inundation, and the impacts of these events on biogeochemistry, terrestrial aquatic interfaces and high latitude/gradient regions.
- Responses should be framed around one or more of the following focal area, ensuring that technology and techniques are incorporated as a critical component of a development pathway:
    1. Data acquisition and assimilation enabled by machine learning, AI, and advanced methods including experimental/network design/optimization, unsupervised learning (including deep learning), and hardware-related efforts involving AI (e.g., edge computing)
    2. Predictive modeling through the use of AI techniques and AI-derived model components; the use of AI and other tools to design a prediction system comprising of a hierarchy of models (e.g., AI driven model/component/parameterization selection)
    3. Insight gleaned from complex data (both observed and simulated) using AI, big data analytics, and other advanced methods, including explainable AI and physics- or knowledge-guided AI
- White papers should identify the primary relevant focal area (1,2, or 3) - the workshops will be arranged around these focal areas. Responses should include how unique DOE capabilities would be brought to bear on the scientific question, including e.g., exascale computing, existing data holdings, community modeling programs and unique observational capabilities. White papers must provide sufficient detail to show how the conceptual idea can more rapidly and fully address the scientific question.
- White papers must address the data-model integration grand challenge presented in the EESSD Strategic Plan. Ideas that incorporate data generated by the Atmospheric Radiation Measurement (ARM) Facility, Environmental Molecular Science Laboratory (EMSL), Next Generation Ecosystem Experiments (NGEEs), Science Focus Area (SFA) observatories, and/or other DOE-generated information are encouraged. Use of data provided by other agencies is also encouraged, as long as it is supplemented by and enhances DOE-supported datasets and software.
- Develop a high-level approach that incorporates one or more of the following as a critical component of a development pathway (if unknown, state any known barriers to understanding the pathway):
    - Topics must employ machine learning or another AI techniques or technologies
    - Topics are encouraged to employ other emerging technologies, e.g., edge computing, nonlinear data assimilation, 5G/6G and advanced wireless
    - Topics should consider a level of difficulty that demands advanced computing capabilities that are likely to be available over the next decade
    - Topics that are more forward looking and explore the application of quantum sensors and/or quantum computing related to AI are also encouraged
- Responses should outline how code and other tools generated in the process of addressing the author-selected science question will be made reusable and findings reproducible. Such tools can include (but are not limited to) open source code, packaging to allow deployment on a variety of hardware, outreach and workforce development. White papers should outline how FAIR (Findable, Accessible, Interoperable, Reusable) principles will be incorporated into products (e.g., data, networks, models, tools, etc.).

Appendix D: Participants

AI4ESP would like to extend our utmost gratitude to the white paper call and workshop participants. Our goal was to provide a forum and avenue for your great ideas to be heard by other scientists, agency managers and administration, and U.S. government decision makers. Your selfless participation and open communication of ideas invigorated hope in the scientific process and community to drive innovation and transformation for humankind. We feel confident to express, on behalf of all authors of this report, that we hope to have done justice to the information provided and discussed in the workshop and beyond. Thank you again for the hard work and sacrifices you made to participate.

Participants (listed alphabetically by role and session; name and institution provided by participants on workshop registration form):

Session POCs and Chairs

Name	Affiliation	Sessions
Po-Lun Ma	Pacific Northwest National Laboratory	Aerosols & Clouds
Jim Ang	Pacific Northwest National Laboratory	AI Architecture Co-Design
Ruby Leung	Pacific Northwest National Laboratory	Atmospheric Modeling
Maria Molina	University of Maryland, College Park	Climate Variability & Extremes
Matthew Hoffman	Los Alamos National Laboratory	Coastal Dynamics, Oceans & Ice
Giri Prakash	Oak Ridge National Laboratory	Data Acquisition to Distribution
Forrest Hoffman	Oak Ridge National Laboratory	Ecohydrology
Line Pouchard	Brookhaven National Laboratory	Explainable/Interpretable/Trustworthy AI
Christa Brelsford	Oak Ridge National Laboratory	Human Systems & Dynamics
Sivasankaran Rajamanickam	Sandia National Laboratories	Hybrid Modeling
Charuleka Varadharajan	Lawrence Berkeley National Laboratory	Hydrology
Xingyuan Chen	Pacific Northwest National Laboratory	Knowledge Discovery & Statistical Learning
Frank Alexander	Brookhaven National Laboratory	Knowledge-Informed Machine Learning
Beth Drewniak	Argonne National Laboratory	Land Modeling
Auroop Ganguly	Northeastern University	Neural Networks
Nathan Urban	Brookhaven National Laboratory	Surrogate Models & Emulators
Mavrik Zavarin	Lawrence Livermore National Laboratory	Watershed Science

Session Co-chairs

Name	Affiliation	Sessions
Pavlos Kollias	Brookhaven National Laboratory	Aerosols & Clouds

Yunyan Zhang	Lawrence Livermore National Laboratory	Aerosols & Clouds
Salil Mahjan	Oak Ridge National Laboratory	Aerosols & Clouds
Adam Varble	Pacific Northwest National Laboratory	Aerosols & Clouds
Sam Silva	Pacific Northwest National Laboratory	Aerosols & Clouds
Paloma Borque	Pacific Northwest National Laboratory	Aerosols & Clouds
James Hoe	Carnegie Mellon University	AI Architecture Co-Design
Tushar Krishna	Georgia Institute of Technology	AI Architecture Co-Design
Maya Gokhale	Lawrence Livermore National Laboratory	AI Architecture Co-Design
Simon Hammond	NNSA	AI Architecture Co-Design
Sarat Sreepathi	Oak Ridge National Laboratory	AI Architecture Co-Design
Travis O'Brien	University of Indiana	Climate Variability & Extremes
Will Pringle	Argonne National Laboratory	Coastal Dynamics, Oceans & Ice
Steven Brus	Argonne National Laboratory	Coastal Dynamics, Oceans & Ice
Rao Kotamarthi	Argonne National Laboratory	Coastal Dynamics, Oceans & Ice
Paul Durack	Lawrence Livermore National Laboratory	Coastal Dynamics, Oceans & Ice
Carolyn Begeman	Los Alamos National Laboratory	Coastal Dynamics, Oceans & Ice
Andrew Roberts	Los Alamos National Laboratory	Coastal Dynamics, Oceans & Ice
David Judi	Pacific Northwest National Laboratory	Coastal Dynamics, Oceans & Ice
Rob Hetland	Pacific Northwest National Laboratory	Coastal Dynamics, Oceans & Ice
Irina Tezaur	Sandia National Laboratory	Coastal Dynamics, Oceans & Ice
Shawn Serbin	Brookhaven National Laboratory	Data Acquisition to Distribution
Jinyun Tang	Lawrence Berkeley National Laboratory	Ecohydrology
Zheng Shi	University of Oklahoma	Ecohydrology
Byung-Jun Yoon	BNL/Texas A&M	Explainable/Interpretable/Trustworthy AI
Line Pouchard	Brookhaven National Laboratory	Explainable/Interpretable/Trustworthy AI
Timo Bremer	Lawrence Livermore National Laboratory	Explainable/Interpretable/Trustworthy AI
Bhavya Kailkhura	Lawrence Livermore National Laboratory	Explainable/Interpretable/Trustworthy AI
Ben Brown	LBNL	Explainable/Interpretable/Trustworthy AI
Derek DeSantis	Los Alamos National Laboratory	Explainable/Interpretable/Trustworthy AI
Maria Glenski	Pacific Northwest National Laboratory	Explainable/Interpretable/Trustworthy AI
Svlitana Volkova	Pacific Northwest National Laboratory	Explainable/Interpretable/Trustworthy AI
Guang Lin	Purdue University	Explainable/Interpretable/Trustworthy AI
Amy McGovern	University of Oklahoma	Explainable/Interpretable/Trustworthy AI
Donatella Pasqualini	Los Alamos National Laboratory	Human Systems & Dynamics

Melissa R. Allen-Dumas	Oak Ridge National Laboratory	Human Systems & Dynamics
Nathalie Voisin	Pacific Northwest National Laboratory	Human Systems & Dynamics
Abigail Snyder	Pacific Northwest National Laboratory	Human Systems & Dynamics
Thushara Gunda	Sandia National Laboratories	Human Systems & Dynamics
Prasanna Balaprakash	Argonne National Laboratory	Hybrid Modeling
Jiali Wang	Argonne National Laboratory	Hybrid Modeling
Nathan Urban	Brookhaven National Laboratory	Hybrid Modeling
Peishi Jiang	Pacific Northwest National Laboratory	Hybrid Modeling
Jitendra Kumar	Oak Ridge National Laboratory	Hydrology
Scott Painter	Oak Ridge National Laboratory	Hydrology
Dan Lu	Oak Ridge National Laboratory	Hydrology
Chaopeng Shen	Penn State Univ	Hydrology
Richard Mills	Argonne National Laboratory	Knowledge Discovery & Statistical Learning
Vipin Kumar	University of Minnesota	Knowledge Informed Machine Learning
Romit Maulik	ANL and IIT-Chicago	Neural Networks
Juliane Mueller	Lawrence Berkeley National Laboratory	Neural Networks
Kate Duffy	NASA Bay Area Environmental Research Institute	Neural Networks
Richard Archibald	Oak Ridge National Laboratory	Neural Networks
Nathan Hodas	Pacific Northwest National Laboratory	Neural Networks
Khachik Sargsyan	Sandia National Laboratories	Neural Networks
Dan Lu	Oak Ridge National Laboratory	Surrogate Models & Emulators
Carl Steefel	Lawrence Berkeley National Laboratory	Watershed Science
Dipankar Dwivedi	Lawrence Berkeley National Laboratory	Watershed Science
David Moulton	Los Alamos National Laboratory	Watershed Science
Scott Painter	Oak Ridge National Laboratory	Watershed Science
Xingyuan Chen	Pacific Northwest National Laboratory	Watershed Science
Li Li	Penn State University	Watershed Science

Breakout Room Leads and Rapporteurs

Name	Affiliation	Sessions	Contribution
Pavlos Kollias	Brookhaven National Laboratory	Aerosols & Clouds	BR lead
Yunyan Zhang	Lawrence Livermore National Laboratory	Aerosols & Clouds	BR lead
Salil Mahjan	Oak Ridge National Laboratory	Aerosols & Clouds	BR lead
Min Xu	Oak Ridge National Laboratory	Aerosols & Clouds	BR lead
Adam Varble	Pacific Northwest National Laboratory	Aerosols & Clouds	BR lead
Sam Silva	Pacific Northwest National Laboratory	Aerosols & Clouds	BR lead
Paloma Borque	Pacific Northwest National Laboratory	Aerosols & Clouds	BR lead
Po-Lun Ma	Pacific Northwest National Laboratory	Aerosols & Clouds	BR lead
Baoxiang Pan	Lawrence Livermore National Laboratory	Aerosols & Clouds	Rapporteur

Murali M. Gopalakrishnan	Oak Ridge National Laboratory	Aerosols & Clouds	Rapporteur
Andrew Geiss	Pacific Northwest National Laboratory	Aerosols & Clouds	Rapporteur
Colleen Kaul	Pacific Northwest National Laboratory	Aerosols & Clouds	Rapporteur
Ivy Peng	Lawrence Livermore National Laboratory	AI Architecture Co-Design	Rapporteur
Matthew Norman	Oak Ridge National Laboratory	AI Architecture Co-Design	Rapporteur
Roberto Gioiosa	Pacific Northwest National Laboratory	AI Architecture Co-Design	Rapporteur
Maruti Mudunuru	Pacific Northwest National Laboratory	AI Architecture Co-Design	Rapporteur
Antonino Tumeo	Pacific Northwest National Laboratory	AI Architecture Co-Design	Rapporteur
Noah Brenowitz	Allen Institute for AI	Atmospheric Modeling	BR lead
Chris Bretherton	Allen Institute for AI and University of Washington	Atmospheric Modeling	BR lead
Marcus van Lier-Walqui	Columbia University	Atmospheric Modeling	BR lead
Gregory Elsaessser	Columbia University & NASA Goddard Institute for Space Studies	Atmospheric Modeling	BR lead
Istvan Szunyogh	Texas A&M University	Atmospheric Modeling	BR lead
Amy McGovern	University of Oklahoma	Atmospheric Modeling	BR lead
Chris Fletcher	University of Waterloo	Atmospheric Modeling	BR lead
Scott Collis	Argonne National Laboratory	Atmospheric Modeling	Rapporteur
Yangang Liu	Brookhaven National Laboratory	Atmospheric Modeling	Rapporteur
Peter Caldwell	Lawrence Livermore National Laboratory	Atmospheric Modeling	Rapporteur
Andy Salinger	Sandia National Laboratory	Atmospheric Modeling	Rapporteur
Andrew Bradley	Sandia National Laboratory	Atmospheric Modeling	Rapporteur
Paul Ullrich	University of California, Davis	Atmospheric Modeling	Rapporteur
David Mechem	University of Kansas	Atmospheric Modeling	Rapporteur
Gemma Anderson	Lawrence Livermore National Laboratory	Climate Variability & Extremes	BR lead
Steve Klein	Lawrence Livermore National Laboratory	Climate Variability & Extremes	BR lead
Katrina Bennett	Los Alamos National Laboratory	Climate Variability & Extremes	BR lead
Bill Collins	University of California at Berkeley; Lawrence Berkeley National Laboratory	Climate Variability & Extremes	BR lead
Scott Collis	Argonne National Laboratory	Climate Variability & Extremes	Rapporteur
Katie Dagon	National Center for Atmospheric Research	Climate Variability & Extremes	Rapporteur
Moet Ashfaq	Oak Ridge National Laboratory	Climate Variability & Extremes	Rapporteur
Paul Ullrich	University of California at Davis	Climate Variability & Extremes	Rapporteur
William Pringle	Argonne National Laboratory	Coastal Dynamics, Oceans & Ice	BR lead

Steven Brus	Argonne National Laboratory	Coastal Dynamics, Oceans & Ice	Rapporteur
Derek DeSantis	Los Alamos National Laboratory	Coastal Dynamics, Oceans & Ice	Rapporteur
Luke Van Roekel	Los Alamos National Laboratory	Coastal Dynamics, Oceans & Ice	Rapporteur
Elena Reinisch	Los Alamos National Laboratory	Coastal Dynamics, Oceans & Ice	Rapporteur
Sanjib Sharma	Pennsylvania State University/Now at McGill University	Coastal Dynamics, Oceans & Ice	Rapporteur
Scott Giangrande	Brookhaven National Laboratory	Data Acquisition to Distribution	BR lead
Chongai Kuang	Brookhaven National Laboratory	Data Acquisition to Distribution	BR lead
Line Pouchard	Brookhaven National Laboratory	Data Acquisition to Distribution	BR lead
Baptiste Dafflon	Lawrence Berkeley National Laboratory	Data Acquisition to Distribution	BR lead
Giri Prakash	Oak Ridge National Laboratory	Data Acquisition to Distribution	BR lead
Kyle Pressel	Pacific Northwest National Laboratory	Data Acquisition to Distribution	BR lead
Prassana		Data Acquisition to Distribution	BR lead
Beth Drewniak	Argonne National Laboratory	Data Acquisition to Distribution	Rapporteur
Shawn Serbin	Brookhaven National Laboratory	Data Acquisition to Distribution	Rapporteur
Cory Stuart	Oak Ridge National Laboratory	Data Acquisition to Distribution	Rapporteur
Mallory Barnes	Indiana University	Ecohydrology	BR lead
Matthias Sprenger	Lawrence Berkeley National Laboratory	Ecohydrology	BR lead
Erica Woodburn	Lawrence Berkeley National Laboratory	Ecohydrology	BR lead
Chonggang Xu	Los Alamos National Laboratory	Ecohydrology	BR lead
Jiafu Mao	Oak Ridge National Laboratory	Ecohydrology	BR lead
Umakant Mishra	Sandia National Laboratories	Ecohydrology	BR lead
Sarah Scott	Sandia National Laboratories	Ecohydrology	BR lead
Richard Mills	Argonne National Laboratory	Ecohydrology	Rapporteur
James Dennedy-Frank	Lawrence Berkeley National Laboratory	Ecohydrology	Rapporteur
Paul Levine	NASA Jet Propulsion Laboratory	Ecohydrology	Rapporteur
Sagar Gautam	Pacific Northwest National Laboratory	Ecohydrology	Rapporteur
Elias Massoud	University of California, Berkeley	Ecohydrology	Rapporteur
Yaoping Wang	University of Tennessee	Ecohydrology	Rapporteur
Claire Zarakas	University of Washington	Ecohydrology	Rapporteur
Aric Hagberg	Los Alamos National Laboratory	Human Systems & Dynamics	Rapporteur
Kuldeep Kurte	Oak Ridge National Laboratory	Human Systems & Dynamics	Rapporteur

Joe Tuccillo	Oak Ridge National Laboratory	Human Systems & Dynamics	Rapporteur
Chris Vernon	Pacific Northwest National Laboratory	Human Systems & Dynamics	Rapporteur
Jim Yoon	Pacific Northwest National Laboratory	Human Systems & Dynamics	Rapporteur
Nicole Jackson	Sandia National Laboratories	Human Systems & Dynamics	Rapporteur
Julie Bessac	Argonne National Laboratory	Hybrid Modeling	Rapporteur
Thushara Gunda	Sandia National Laboratories	Hybrid Modeling	Rapporteur
Hongkyu Yoon	Sandia National Laboratories	Hybrid Modeling	Rapporteur
Andrew Bradley	Sandia National Laboratory	Hybrid Modeling	Rapporteur
David Moulton	Los Alamos National Laboratory	Hydrology	BR lead
Xingyuan Chen	Pacific Northwest National Laboratory	Hydrology	BR lead
Julie Bessac	Argonne National Laboratory	Knowledge Discovery & Statistical Learning	BR lead
Emil Constantinescu	Argonne National Laboratory	Knowledge Discovery & Statistical Learning	BR lead
Michael Wehner	Lawrence Berkeley National Laboratory	Knowledge Discovery & Statistical Learning	BR lead
Juliane Mueller	Lawrence Berkeley National Laboratory	Knowledge Discovery & Statistical Learning	BR lead
Chandrika Kamath	Lawrence Livermore National Laboratory	Knowledge Discovery & Statistical Learning	BR lead
Shashank Konduri	NASA Goddard Space Flight Center	Knowledge Discovery & Statistical Learning	BR lead
Juan Restrepo	Oak Ridge National Laboratory	Knowledge Discovery & Statistical Learning	BR lead
Forrest Hoffman	Oak Ridge National Laboratory	Knowledge Discovery & Statistical Learning	BR lead
Murali M. Gopalakrishnan	Oak Ridge National Laboratory	Knowledge Discovery & Statistical Learning	BR lead
Zachary Langford	Oak Ridge National Laboratory	Knowledge Discovery & Statistical Learning	BR lead
John Jakeman	Sandia National Laboratories	Knowledge Discovery & Statistical Learning	BR lead
Elias Massoud	University of California, Berkeley	Knowledge Discovery & Statistical Learning	BR lead
Youzuo Lin		Knowledge Discovery & Statistical Learning	BR lead
Shinjae Yoo	Brookhaven National Laboratory	Knowledge-Informed Machine Learning	BR lead
Jayaraman Thiagarajan	Lawrence Livermore National Laboratory	Knowledge-Informed Machine Learning	BR lead
Mahantesh Halappanavar	Pacific Northwest National Laboratory	Knowledge-Informed Machine Learning	BR lead
Karthik Kashinath	Lawrence Berkeley National Laboratory	Knowledge-Informed Machine Learning	Rapporteur
Arvind Thanam Mohan	Los Alamos National Laboratory	Knowledge-Informed Machine Learning	Rapporteur
Sutanay Choudhury	Pacific Northwest National Laboratory	Knowledge-Informed Machine Learning	Rapporteur

Pamela Weisenhorn	Argonne National Laboratory	Land Modeling	BR lead
Qing Zhu	Lawrence Berkeley National Laboratory	Land Modeling	BR lead
Anthony Walker	Oak Ridge National Laboratory	Land Modeling	BR lead
Dan Ricciuto	Oak Ridge National Laboratory	Land Modeling	BR lead
Xiaojuan Yang	Oak Ridge National Laboratory	Land Modeling	BR lead
Beth Drewniak	Argonne National Laboratory	Land Modeling	Rapporteur
Charlie Koven	Lawrence Berkeley National Laboratory	Land Modeling	Rapporteur
William Riley	LBNL/Now at the University of Southern California	Land Modeling	Rapporteur
Jingfeng Xiao	University of New Hampshire	Land Modeling	Rapporteur
Abigail Swann	University of Washington	Land Modeling	Rapporteur
Vishwas Rao	Argonne National Laboratory	Neural Networks	Rapporteur
Jack Watson	Northeastern University	Neural Networks	Rapporteur
Sebastian Ruf	Northeastern University	Neural Networks	Rapporteur
Puja Das	Northeastern University	Neural Networks	Rapporteur
Jong Youl Choi	Oak Ridge National Laboratory	Neural Networks	Rapporteur
Craig Bakker	Pacific Northwest National Laboratory	Neural Networks	Rapporteur
Cosmin Safta	Sandia National Laboratories	Neural Networks	Rapporteur
Vanessa Lopez-Marrero	Brookhaven National Laboratory	Surrogate Models & Emulators	BR lead
Shinjae Yoo	Brookhaven National Laboratory	Surrogate Models & Emulators	BR lead
Panos Stinis	Pacific Northwest National Laboratory	Surrogate Models & Emulators	BR lead
Draguna Vrabie	Pacific Northwest National Laboratory	Surrogate Models & Emulators	BR lead
Kenny Chowdhary	Sandia National Laboratory	Surrogate Models & Emulators	BR lead
Earl Lawrence		Surrogate Models & Emulators	BR lead
Carl Steefel	Lawrence Berkeley National Laboratory	Watershed Science	BR lead
Dipankar Dwivedi	Lawrence Berkeley National Laboratory	Watershed Science	BR lead
Mavrik Zavarin	Lawrence Livermore National Laboratory	Watershed Science	BR lead
David Moulton	Los Alamos National Laboratory	Watershed Science	BR lead
Scott Painter	Oak Ridge National Laboratory	Watershed Science	BR lead
Xingyuan Chen	Pacific Northwest National Laboratory	Watershed Science	BR lead
Vanessa Bailey	Pacific Northwest National Laboratory	Watershed Science	Rapporteur
Umakant Mishra	Sandia National Laboratory	Watershed Science	Rapporteur

Plenary Speakers

First Name	Last Name	Institution
Alison	Appling	U.S. Geological Survey

Kirk	Borne	DataPrime Inc.
Katie	Dagon	National Center for Atmospheric Research
Pierre	Gentine	Columbia University
Grace	Kim	Booz Allen Hamilton
Amy	McGovern	University of Oklahoma
Pabhat	Ram	University of California, Berkeley
Rob	Ross	Argonne National Laboratory
Tapio	Schneider	California Institute of Technology
Chaopeng	Shen	Pennsylvania State University
Rick	Stevens	Argonne National Laboratory
Laure	Zanna	New York University

Other Participants

First Name	Last Name	Institution
Erin	Acquesta	Sandia National Laboratories
Deb	Agarwal	Lawrence Berkeley National Laboratory
Osinachi	Ajoku	Howard University
John	Allen	Central Michigan University
Steven	Allison	University of California, Irvine
Philipe	Ambrozio Dias	Oak Ridge National Laboratory
Emmanouil	Anagnostou	University of Connecticut
Animashree	Anandkumar	California Institute of Technology
Valentine	Anantharaj	Oak Ridge National Laboratory
Marian	Anghel	Los Alamos National Laboratory
Jeff	Arnold	U.S. Army Engineer Climate Change Program
Bhavna	Arora	Lawrence Berkeley National Laboratory
Xylar	Asay-Davis	Los Alamos National Laboratory
Hessam	Babaee	University of Pittsburgh
David	Bader	Lawrence Livermore National Laboratory
Karthik	Balaguru	Pacific Northwest National Laboratory
Jerad	Bales	Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI)
Kenneth	Ball	Geometric Data Analytics, Inc.
Antara	Banerjee	University of Colorado and National Oceanic and Atmospheric Administration, Physical Sciences Laboratory
Feng	Bao	Florida State University
Jie	Bao	Pacific Northwest National Laboratory
David	Barajas-Solano	Pacific Northwest National Laboratory
John	Bargar	SLAC National Accelerator Laboratory
Elizabeth	Barnes	Colorado State University

Ana	Barros	University of Illinois at Urbana-Champaign
Kanad	Basu	University of Texas at Dallas
Jennifer	Bauer	National Energy Technology Laboratory
Susanne	Bauer	NASA Goddard Institute for Space Studies
Paul	Bayer	U.S. Department of Energy
James	Benedict	Los Alamos National Laboratory
Andrew	Bennett	University of Arizona
Russell	Bent	Los Alamos National Laboratory
Emily	Bercos-Hickey	Lawrence Berkeley National Laboratory
Mira	Berdahl	University of Washington
Larry	Berg	Pacific Northwest National Laboratory
Karianne	Bergen	Brown University
Max	Berkelhammer	University of Illinois at Chicago
Judith	Berner	National Center for Atmospheric Research
Tom	Beucler	University of Lausanne
Budhu	Bhaduri	Oak Ridge National Laboratory
Soumendra	Bhanja	Oak Ridge National Laboratory
Derek	Bingham	Simon Fraser University
Sebastien	Biraud	Lawrence Berkeley National Laboratory
Ayan	Biswas	Los Alamos National Laboratory
Cecilia	Bitz	University of Washington
Pavel	Bochev	Sandia National Laboratories
Peter	Bogenschutz	Lawrence Livermore National Laboratory
Ben	Bond-Lamberty	Pacific Northwest National Laboratory
Celine	Bonfils	Lawrence Livermore National Laboratory
Kevin	Booth	Radiant Earth Foundation
Nick	Bouskill	Lawrence Berkeley National Laboratory
Renato	Braghiere	NASA Jet Propulsion Laboratory
Marcia	Branstetter	Oak Ridge National Laboratory
Amy	Braverman	NASA Jet Propulsion Laboratory
Douglas	Brinkerhoff	University of Montana
Eoin	Brodie	Lawrence Berkeley National Laboratory
Tamas	Budavari	Johns Hopkins University
Diana	Bull	Sandia National Laboratories
Maria	Cadeddu	Argonne National Laboratory
Mariana	Cains	National Center for Atmospheric Research
Kate	Calvin	Pacific Northwest National Laboratory
Alex	Cannon	Climate Research Division, Environment and Climate Change Canada
Zhendong	Cao	University of Delaware
Antonietta	Capotondi	University of Colorado and National Oceanic and Atmospheric Administration, Physical Sciences Laboratory
Zoe	Cardon	Marine Biological Laboratory, The Ecosystems Center, Woods Hole, MA
Julie	Caron	National Center for Atmospheric Research
Ilene	Carpenter	Hewlett Packard Enterprise

Arijit	Chakraborty	Columbia University
Srija	Chakraborty	NASA Goddard Space Flight Center
Chandra V	Chandrasekar	Colorado State University
Won	Chang	University of Cincinnati
Will	Chapman	Scripps Institution of Oceanography
Samrat	Chatterjee	Pacific Northwest National Laboratory
Nan	Chen	University of Wisconsin, Madison
Yunxiang	Chen	Pacific Northwest National Laboratory
Gyan	Chhipi Shrestha	University of British Columbia
Christine	Chiu	Colorado State University
Youngsoo	Choi	Lawrence Livermore National Laboratory
Shreyas	Cholia	Lawrence Berkeley National Laboratory
Hannah	Christensen	University of Oxford
Matthew	Christensen	Pacific Northwest National Laboratory
Danielle	Christianson	Lawrence Berkeley National Laboratory
Carter	Christopher	Oak Ridge National Laboratory
Maria	Chu	University of Illinois at Urbana-Champaign
Seungwon	Chung	University of Texas at Austin
Deborah	Coen	Yale University
Nathan	Collier	Oak Ridge National Laboratory
Jennifer	Comstock	Pacific Northwest National Laboratory
Laura	Condon	University of Arizona
Ethan	Coon	Oak Ridge National Laboratory
Bruce	Cornuelle	University of California, San Diego
Louise	Criscenti	Sandia National Laboratories
Octavia	Crompton	Duke University
Robert	Crystal-Ornelas	Lawrence Berkeley National Laboratory
Marta	D'Elia	Sandia National Laboratories
Angela	Dalton	Advanced Micro Devices, Inc.
Joan	Damerow	Lawrence Berkeley National Laboratory
Gokhan	Danabasoglu	National Center for Atmospheric Research
Maggie	Davis	Oak Ridge National Laboratory
Warren	Davis	Sandia National Laboratories
Xujing	Davis	U.S. Department of Energy
Clint	Dawson	University of Texas at Austin
Martin	De Kauwe	University of Bristol
Belay	Demoz	University of Maryland, Baltimore County
Yi	Deng	Georgia Institute of Technology
Ranjeet	Devarakonda	Oak Ridge National Laboratory
Alan	Di Vittorio	Lawrence Berkeley National Laboratory
Michael	Dietze	Boston University
Jan	Drgona	Pacific Northwest National Laboratory
Shiheng	Duan	University of California, Davis
Manvendra	Dubey	Los Alamos National Laboratory
Ravindra	Duddu	Vanderbilt University
Peter	Dueben	European Centre for Medium-Range Weather Forecasts

Edward	Dumas	National Oceanic and Atmospheric Administration, Atmospheric Turbulence and Diffusion Division
Imme	Ebert-Uphoff	Colorado State University
Scott	Elliott	Los Alamos National Laboratory
Darren	Engwirda	Los Alamos National Laboratory
Nicola	Falco	Lawrence Berkeley National Laboratory
Caroline	Famiglietti	Stanford University
Jiwen	Fan	Pacific Northwest National Laboratory
Yilin	Fang	Pacific Northwest National Laboratory
Boris	Faybishenko	Lawrence Berkeley National Laboratory
Graham	Feingold	National Oceanic and Atmospheric Administration
Daniel	Feldman	Lawrence Berkeley National Laboratory
Dapeng	Feng	Pennsylvania State University
Yan	Feng	Argonne National Laboratory
Zhe	Feng	Pacific Northwest National Laboratory
Carlos	Fernandez-Granda	New York University
Darren	Ficklin	Indiana University
Laura	Fierce	Pacific Northwest National Laboratory
Hal	Finkel	U.S. Department of Energy
Rosie	Fisher	Center for International Climate Research
Ian	Foster	Argonne National Laboratory
Efi	Foufoula-Georgiou	University of California, Irvine
Meg	Fowler	National Center for Atmospheric Research
Baylor	Fox-Kemper	Brown University
James	Franke	University of Chicago
Ari	Frankel	Sandia National Laboratories
Christian	Frankenberg	California Institute of Technology & NASA Jet Propulsion Laboratory
Ann	Fridlind	NASA Goddard Institute for Space Studies
Xiaojing	Fu	California Institute of Technology
Kwun Yip	Fung	University of Texas at Austin
David John	Gagne	National Center for Atmospheric Research
Sudershan	Gangrade	Oak Ridge National Laboratory
Peiyuan	Gao	Pacific Northwest National Laboratory
Alfredo	Garbuno-Inigo	Instituto Tecnológico de México (ITAM)
John	Gardner	University of Pittsburgh
Arti	Garg	Hewlett Packard Enterprise
Sagar	Gautam	Sandia National Laboratories
Gary	Geernaert	U.S. Department of Energy
Katerina	Georgiou	Lawrence Livermore National Laboratory
Andrew	Gettelman	National Center for Atmospheric Research
Stu	Geza	South Dakota School of Mines
Roger	Ghanem	University of Southern California
Virendra	Ghate	Argonne National Laboratory
Daniel	Gilford	Climate Central
Sarah	Gille	Scripps Institution of Oceanography & University of California, San Diego
David	Gochis	National Center for Atmospheric Research

Humberto	Godinez	Los Alamos National Laboratory
Chris	Golaz	Lawrence Livermore National Laboratory
Evan	Goldstein	University of North Carolina, Greensboro
Miquel	Gonzalez-Meler	University of Illinois at Chicago
Michael	Gooseff	University of Colorado
Jennifer	Granholm	U.S. Secretary of Energy
Carlo	Graziani	Argonne National Laboratory
Julia	Green	University of California, Berkeley
Lianhong	Gu	Oak Ridge National Laboratory
Kaiyu	Guan	University of Illinois
Yawen	Guan	University of Nebraska-Lincoln
Oksana	Guba	Sandia National Laboratories
Abhinav	Gupta	Desert Research Institute
Abhishek	Gupta	Montreal AI Ethics Institute, Microsoft, and Green Software Foundation
Hoshin	Gupta	University of Arizona
Ritwik	Gupta	University of California, Berkeley
William	Gustafson	Pacific Northwest National Laboratory
George	Guthrie	Los Alamos National Laboratory
Antonia	Hadjimichael	Cornell University
Samson	Hagos	Pacific Northwest National Laboratory
Alex	Hall	University of California, Los Angeles
Hendrik	Hamann	IBM Research
Dorit	Hammerling	Colorado School of Mines
Glenn	Hammond	Pacific Northwest National Laboratory
Erin	Hanan	University of Nevada, Reno
Carly	Hansen	Oak Ridge National Laboratory
Stijn	Hantson	University of California, Irvine
Joseph	Hardin	ClimateAI
John	Harlim	Pennsylvania State University
Bryce	Harrop	Pacific Northwest National Laboratory
Paul	Hartke	Xilinx
Pedram	Hassanzadeh	Rice University
Sue Ellen	Haupt	National Center for Atmospheric Research
Linnia	Hawkins	Oregon State University
Qizhi	He	San Diego State University
Erin	Hestir	University of California, Merced
Nicki	Hickmon	Argonne National Laboratory
Magnus	Hieronymus	Swedish Meteorological and Hydrological Institute
Benjamin	Hillman	Sandia National Laboratories
Andy	Hock	Cerebras Systems
Ben	Hodges	University of Texas at Austin
Alma	Hodzic	National Center for Atmospheric Research
Marvin	Höge	Eawag, Dübendorf, Switzerland
Jennifer	Holm	Lawrence Berkeley National Laboratory
Tianzhen	Hong	Lawrence Berkeley National Laboratory
Julia	Hopkins	Northeastern University
Z. Jason	Hou	Pacific Northwest National Laboratory
Kuolin	Hsu	University of California, Irvine

Aixue	Hu	National Center for Atmospheric Research
Huanping	Huang	Lawrence Berkeley National Laboratory
Whitney	Huang	Clemson University
Yuanyuan	Huang	Commonwealth Scientific and Industrial Research Organisation (CSIRO)
Susan	Hubbard	Lawrence Berkeley National Laboratory
Gabriel	Huerta	Sandia National Laboratories
Elizabeth	Hunke	Los Alamos National Laboratory
Brian	Hunt	University of Maryland
Kayo	Ide	University of Maryland
Adele	Igel	University of California, Davis
Jennifer	Irish	Virginia Tech University
Robert	Jacob	Argonne National Laboratory
Daniel	Jacobson	Oak Ridge National Laboratory
Kolby	Jardine	Lawrence Berkeley National Laboratory
Julie	Jastrow	Argonne National Laboratory
Shantenu	Jha	Brookhaven National Laboratory
Xiaowei	Jia	University of Pittsburgh
Yufang	Jin	University of California, Davis
Zhenong	Jin	University of Minnesota
Jennifer	Johnson	Carnegie Institution
Andrew	Jones	Lawrence Berkeley National Laboratory
Philip	Jones	Los Alamos National Laboratory
Renu	Joseph	U.S. Department of Energy
Erik	Kabela	Oak Ridge National Laboratory
Ramakrishnan	Kannan	Oak Ridge National Laboratory
Shih-Chieh	Kao	Oak Ridge National Laboratory
Ulas	Karaoz	Lawrence Berkeley National Laboratory
Ian	Karlin	Intel Corporation
Anuj	Karpatne	Virginia Tech University
Satish	Karra	Los Alamos National Laboratory
Trevor	Keenan	Lawrence Berkeley National Laboratory & University of California, Berkeley
Michael	Kelleher	Oak Ridge National Laboratory
Klaus	Keller	Pennsylvania State University
Maxwell	Kelley	NASA Goddard Institute for Space Studies
Ken	Kemner	Argonne National Laboratory
Rajkumar	Kettimuthu	Argonne National Laboratory
Yannis	Kevrekidis	Johns Hopkins University
Been	Kim	Google Brain
Hyungjun	Kim	Korea Advanced Institute of Science and Technology
Kerstin	Kleese van Dam	Brookhaven National Laboratory
James	Knighton	University of Connecticut
Ryan	Knox	Lawrence Berkeley National Laboratory
Goutam	Konapala	NASA Goddard Space Flight Center
Alexandra	Konings	Stanford University
Gabriel	Kooperman	University of Georgia

Bob	Kopp	Rutgers University
Olivera	Kotevska	Oak Ridge National Laboratory
Christopher	Krapu	Oak Ridge National Laboratory
Chongai	Kuang	Brookhaven National Laboratory
Paul	Kuberry	Sandia National Laboratories
Lara	Kueppers	University of California, Berkeley & Lawrence Berkeley National Laboratory
Saurabh	Kulkarni	Graphcore
Neeraj	Kumar	Pacific Northwest National Laboratory
Peeyush	Kumar	Microsoft
Praveen	Kumar	University of Illinois
Takuya	Kurihana	University of Chicago
Nathan	Kutz	University of Washington
Mikael	Kuusela	Carnegie Mellon University
Ching-Yao	Lai	Princeton University
Mark	Lara	University of Illinois
Vince	Larson	University of Wisconsin, Milwaukee & Pacific Northwest National Laboratory
Alexander	Lavin	Institute for Simulation Intelligence
David	Lawrence	National Center for Atmospheric Research
Phong	Le	University of California, Irvine
Allan	Leal	ETH Zurich
Zachary	Lebo	University of Wyoming
Jiwoo	Lee	Lawrence Livermore National Laboratory
Temple	Lee	National Oceanic and Atmospheric Administration, Atmospheric Turbulence and Diffusion Division & Cooperative Institute for Severe and High-Impact Weather Research and Operations
Amanda	Lenzi	Argonne National Laboratory
Elena	Leonarduzzi	Princeton University
Sven	Leyffer	Argonne National Laboratory
Xueke	Li	Brown University
Seung-Hwan	Lim	Oak Ridge National Laboratory
Aranildo	Lima	Aquatic Informatics
Meifeng	Lin	Brookhaven National Laboratory
Ning	Lin	Princeton University
Wuyin	Lin	Brookhaven National Laboratory
Frank	Liu	Oak Ridge National Laboratory
Gavin	Liu	National Energy Technology Laboratory
Siyan	Liu	Oak Ridge National Laboratory
Xiaohong	Liu	Texas A&M University
Yan	Liu	Oak Ridge National Laboratory
Danica	Lombardozzi	National Center for Atmospheric Research
Marcos	Longo	Lawrence Berkeley National Laboratory
Lauren	Lowman	Wake Forest University
Jian	Lu	Pacific Northwest National Laboratory
Valerio	Lucarini	University of Reading
Don	Lucas	Lawrence Livermore National Laboratory
Dalton	Lunga	Oak Ridge National Laboratory
Xiangzhong	Luo	National University of Singapore

Yiqi	Luo	Northern Arizona University
Bethany	Lusch	Argonne National Laboratory
Nicholas	Lutsko	Scripps Institution of Oceanography
Hsi-Yen	Ma	Lawrence Livermore National Laboratory
Shuang	Ma	NASA Jet Propulsion Laboratory
Natasha	MacBean	Indiana University
Katharine	Mach	University of Miami
Stefanie	Mack	British Antarctic Survey
Mickey	MacKie	University of Florida
Gavin D.	Madakumbura	University of California, Los Angeles
Sandeep	Madireddy	Argonne National Laboratory
Salil	Mahajan	Oak Ridge National Laboratory
Kate	Maher	Stanford University
Nicola	Maher	University of Colorado
Michael	Mahoney	University of California, Berkeley
Avni	Malhotra	University of Zurich
Adam	Mangel	Pacific Northwest National Laboratory
Andrew	Margenot	University of Illinois at Urbana-Champaign
James	Marquis	Pacific Northwest National Laboratory
Gianna	Marschmann	Lawrence Berkeley National Laboratory
Lucy	Marshall	University of New South Wales
Zane	Martin	Colorado State University
Roser	Matamala	Argonne National Laboratory
James	Mather	Pacific Northwest National Laboratory
Alyssa	Matthews	Pacific Northwest National Laboratory
Kathryn	Maupin	Sandia National Laboratories
Kirsten	Mayer	Colorado State University
Matthew	Mazloff	University of California, San Diego
Mahmoud	Mbarak	University of Texas at Austin
Renata	McCoy	Lawrence Livermore National Laboratory
Nate	McDowell	Pacific Northwest National Laboratory
Michael	McKerns	Los Alamos National Laboratory
Gerald	Meehl	National Center for Atmospheric Research
Rachel	Meyer	University of California, Santa Cruz
Ariel	Miara	National Renewable Energy Laboratory
Brent	Minchew	Massachusetts Institute of Technology
Utkarsh	Mital	Lawrence Berkeley National Laboratory
Yara	Mohajerani	University of Washington
Claire	Monteleoni	University of Colorado
Paige	Morkner	Support Contractor at the National Energy Technology Laboratory
Hugh	Morrison	National Center for Atmospheric Research
Matthias	Morzfeld	Scripps Institution of Oceanography & University of California, San Diego
Liz	Moyer	University of Chicago
Johannes	Muelmenstaedt	Pacific Northwest National Laboratory
Surya Karthik	Mukkavilli	Neura Lode & Vayuh & Lawrence Berkeley National Laboratory
Srideep	Musuvathy	Sandia National Laboratories
Allison	Myers-Pigg	Pacific Northwest National Laboratory

Balu	Nadiga	Los Alamos National Laboratory
Habib	Najm	Sandia National Laboratories
Kalyana	Nakshatrala	University of Houston
Asmeret	Naugle	Sandia National Laboratories
Grey	Nearing	University of California, Davis
Cristina	Negri	Argonne National Laboratory
Robinson	Negron-Juarez	Lawrence Berkeley National Laboratory
Michelle	Newcomer	Lawrence Berkeley National Laboratory
Jake	Nichol	Sandia National Laboratories
Peter	Nico	Lawrence Berkeley National Laboratory
Veronica	Nieves	University of Valencia
Bart	Nijssen	University of Washington
Dev	Niyogi	University of Texas at Austin
Fearghal	O'Donncha	IBM Research
Daniel	O'Malley	Los Alamos National Laboratory
Mohammed	Ombadi	Lawrence Berkeley National Laboratory
Edward	Ott	University of Maryland
Mariko	Oue	Stony Brook University
Houman	Owhadi	California Institute of Technology
Christopher	Paciorek	University of California, Berkeley
Piya	Pal	University of California, San Diego
Giuliana	Pallotta	Lawrence Livermore National Laboratory
Nishant	Panda	Los Alamos National Laboratory
Gilchan	Park	Brookhaven National Laboratory
Valerio	Pascucci	University of Utah
Linsey	Passarella	Oak Ridge National Laboratory
Shabaz	Patel	One Concern
Tapasya	Patki	Lawrence Livermore National Laboratory
Abani	Patra	Tufts University
Christina	Patricola	Iowa State University
Sean	Peisert	Lawrence Berkeley National Laboratory
Bin	Peng	University of Illinois
Steve	Penny	University of Colorado Boulder
Talita	Perciano	Lawrence Berkeley National Laboratory
Paris	Perdikaris	University of Pennsylvania
Mauro	Perego	Sandia National Laboratories
Kara	Peterson	Sandia National Laboratories
Matt	Peterson	Sandia National Laboratories
Eric	Pierce	Oak Ridge National Laboratory
Suzanne A	Pierce	University of Texas at Austin
Anastasia	Piliouras	Los Alamos National Laboratory
Charity	Plata	Brookhaven National Laboratory
Stephen	Po-Chedley	Lawrence Livermore National Laboratory
Derek	Posselt	NASA Jet Propulsion Laboratory & California Institute of Technology
Jonathan	Poterjoy	University of Maryland
Amy	Powell	Sandia National Laboratories
Evan	Poworoznek	Booz Allen Hamilton
Kyle	Pressel	Pacific Northwest National Laboratory

Stephen	Price	Los Alamos National Laboratory
Mike	Pritchard	University of California, Irvine
Pranav	Puthan	Scripps Institution of Oceanography
Nik	Qafoku	Pacific Northwest National Laboratory
Yun	Qian	Pacific Northwest National Laboratory
Johannes	Quaas	Leipzig University
Tristan	Quaife	University of Reading/UK National Centre for Earth Observation
Christopher	Rackauckas	Massachusetts Institute of Technology
Lavanya	Ramakrishnan	Lawrence Berkeley National Laboratory
James	Randerson	University of California, Irvine
Yuhan (Douglas)	Rao	North Carolina State University
Philip	Rasch	Pacific Northwest National Laboratory
Stephan	Rasp	ClimateAI
Deeksha	Rastogi	Oak Ridge National Laboratory
Patrick	Reed	Cornell University
Peter	Regier	Pacific Northwest National Laboratory
Peter	Reichert	Eawag, Dübendorf, Switzerland
Huiying	Ren	Pacific Northwest National Laboratory
Alexandre	Renchon	Argonne National Laboratory
Kristofer	Reyes	Brookhaven National Laboratory
Alan	Rhoades	Lawrence Berkeley National Laboratory
Jadwiga	Richter	National Center for Atmospheric Research
Bryan	Riel	Massachusetts Institute of Technology
Nicole	Riemer	University of Illinois at Urbana-Champaign
William	Riley	Lawrence Berkeley National Laboratory
Lettie	Roach	University of Washington
Alexander	Robel	Georgia Institute of Technology
David	Roberts	Micron Technology, Inc.
Erika	Roesler	Sandia National Laboratories
Alistair	Rogers	Brookhaven National Laboratory
Kelly	Rose	U.S. Department of Energy, National Energy Technology Laboratory
Andrew	Ross	New York University
David	Rounce	Carnegie Mellon University
Barry	Rountree	Lawrence Livermore National Laboratory
Joel	Rowland	Los Alamos National Laboratory
Ben	Ruddell	Northern Arizona University
Jakob	Runge	German Aerospace Center and TU Berlin
Adam	Rupe	Los Alamos National Laboratory
Stuart	Russell	University of California, Berkeley
Sheila	Saia	State Climate Office of North Carolina
Stephan	Sain	Jupiter Intelligence
Vidya	Samadi	Clemson University
Sean	Santos	Columbia University
Bridget	Scanlon	University of Texas at Austin
Timothy	Scheibe	Pacific Northwest National Laboratory
Ann	Scheliga	University of California, Berkeley
Gavin	Schmidt	NASA Goddard Institute for Space Studies
Hauke	Schulz	Max Planck Institute for Meteorology

Jon	Schwenk	Los Alamos National Laboratory
Axel	Seifert	Deutscher Wetterdienst
Harold	Shanafield	Oak Ridge National Laboratory
Lyndsay	Shand	Sandia National Laboratories
Andrew	Shao	Canadian Centre for Climate Modeling and Analysis
Bharat	Sharma	Northeastern University
Xinyi	Shen	University of Connecticut
Xiaoyinng	Shi	Oak Ridge National Laboratory
Christine	Shields	National Center for Atmospheric Research
Alexey	Shiklomanov	NASA Goddard Space Flight Center
ManishKumar	Shrivastava	Pacific Northwest National Laboratory
Pin	Shuai	Pacific Northwest National Laboratory
Debjani	Sihi	Emory University
Debjani	Singh	Oak Ridge National Laboratory
Sameer	Singh	University of California, Irvine
Anirban	Sinha	Los Alamos National Laboratory
Jared	Smith	U.S. Geological Survey
Nick	Smith	Texas Tech University
Timothy	Smith	University of Colorado/National Oceanic and Atmospheric Administration, Pacific Marine Environmental Laboratory
Balwinder	Snigh	Pacific Northwest National Laboratory
Chad	Sockwell	Sandia National Laboratories
Michael	Sohn	Lawrence Berkeley National Laboratory
Kurt	Solander	Los Alamos National Laboratory
Guillem	Sole Mari	Lawrence Berkeley National Laboratory
Kyongho	Son	Pacific Northwest National Laboratory
Hyun-Seob	Song	University of Nebraska-Lincoln
Xuehang	Song	Pacific Northwest National Laboratory
Yang	Song	University of Arizona
Andrea	Staid	Sandia National Laboratories
James	Stegen	Pacific Northwest National Laboratory
Jeff	Stehr	U.S. Department of Energy
Jebb	Stewart	National Oceanic and Atmospheric Administration, Global Systems Laboratory
Philip	Stier	University of Oxford
Christopher	Still	Oregon State University
Aneesh	Subramanian	University of Colorado Boulder
Ryan	Sullivan	Argonne National Laboratory
Deborah	Sulsky	University of New Mexico
Alex	Sun	University of Texas at Austin
Christine	Sweeney	Los Alamos National Laboratory
Laura	Swiler	Sandia National Laboratories
Sabiha	Tabassum	University of Texas at Austin
Christina (Naomi)	Tague	University of California, Santa Barbara
Nathan	Tallent	Pacific Northwest National Laboratory
Zeli	Tan	Pacific Northwest National Laboratory
Qi	Tang	Lawrence Livermore National Laboratory
Feng	Tao	Tsinghua University

Alexandre	Tartakovsky	University of Illinois at Urbana-Champaign
Daniel	Tartakovsky	Stanford University
Mark	Taylor	Sandia National Laboratories
Claudia	Tebaldi	Pacific Northwest National Laboratory
Joao	Teixeira	NASA Jet Propulsion Laboratory
Chris	Terai	Lawrence Livermore National Laboratory
Teklu	Tesfa	Pacific Northwest National Laboratory
Christopher	Tessum	University of Illinois
Savannah	Thais	Princeton University
Adam	Theisen	Argonne National Laboratory
Erin	Thomas	Los Alamos National Laboratory
Peter	Thornton	Oak Ridge National Laboratory
Eric	Tidd	Amazon Web Services
Katherine	Todd-Brown	University of Florida
Simon	Topp	U.S. Geological Survey
Hoang	Tran	Princeton University
Nhan	Tran	Fermilab
Derek	Tucker	Sandia National Laboratories
Adrian	Turner	Los Alamos National Laboratory
Olga	Tweedy	U.S. Department of Energy
Bob	Vallario	U.S. Department of Energy
Eric	Van Hensbergen	ARM
Thomas	Vandal	NASA Ames Research Center
Kush	Varshney	IBM Research
Lav	Varshney	Brookhaven National Laboratory
Tarun	Verma	Princeton University
Roelof	Versteeg	Subsurface Insights
Velimir	Vesselinov	Los Alamos National Laboratory
Jeffrey	Vetter	Oak Ridge National Laboratory
Daniel	Vimont	University of Wisconsin, Madison
Andrew	Vogelmann	Brookhaven National Laboratory
Sam	Voisin	Geometric Data Analytics, Inc.
Benjamin	Wagman	Sandia National Laboratories
Gregory	Wagner	Massachusetts Institute of Technology
Thomas	Wahl	University of Central Florida
Cheng	Wang	Argonne National Laboratory
Dali	Wang	Oak Ridge National Laboratory
Hailong	Wang	Pacific Northwest National Laboratory
Sally	Wang	Pacific Northwest National Laboratory
Nicholas	Ward	Pacific Northwest National Laboratory
Jeff	Warren	Oak Ridge National Laboratory
Jerry	Watkins	Sandia National Laboratories
Peter	Watson	Bristol University
Duncan	Watson-Parris	University of Oxford
Helen	Weierbach	Lawrence Berkeley National Laboratory
Anthony	Wexler	University of California, Davis
Brian	White	University of North Carolina, Chapel Hill & University of California, Berkeley

Laurent	White	Advanced Micro Devices, Inc.
Stijn	Wielandt	Lawrence Berkeley National Laboratory
Stefan	Wild	Argonne National Laboratory
Jared	Willard	University of Minnesota
Daniel	Williamson	University of Exeter
Jeannette	Wing	Columbia University
Christopher	Wirz	National Center for Atmospheric Research
David	Womble	Oak Ridge National Laboratory
Justin	Wozniak	Argonne National Laboratory
John	Wu	Lawrence Berkeley National Laboratory
Yuxin	Wu	Lawrence Berkeley National Laboratory
Weiming	Xiang	Augusta University
Shaocheng	Xie	Lawrence Livermore National Laboratory
Tianfang	Xu	Arizona State University
Wei	Xu	Brookhaven National Laboratory
Wenwei	Xu	Pacific Northwest National Laboratory
Xiaofeng	Xu	San Diego State University
Zexuan	Xu	Lawrence Berkeley National Laboratory
Eugene	Yan	Argonne National Laboratory
Cheng-En	Yang	University of Tennessee
Da	Yang	University of California, Davis & Lawrence Berkeley National Laboratory
Fan	Yang	Brookhaven National Laboratory
Wenming	Ye	Amazon Web Services
Steve	Yeager	National Center for Atmospheric Research
Byung-Jun	Yoon	Texas A&M University
Rose	Yu	University of California, San Diego
Tianle	Yuan	NASA Goddard Space Flight Center & UMBC JCET
Xingqiu	Yuan	Argonne National Laboratory
Janni	Yuval	Massachusetts Institute of Technology
Mark	Zelinka	Lawrence Livermore National Laboratory
Chengzhu (Jill)	Zhang	Lawrence Livermore National Laboratory
Chidong	Zhang	National Oceanic and Atmospheric Administration, Pacific Marine Environmental Laboratory
Guannan	Zhang	Oak Ridge National Laboratory
Tao	Zhang	Brookhaven National Laboratory
Yu	Zhang	Los Alamos National Laboratory
Yulong	Zhang	University of Tennessee
Lei	Zhao	University of Illinois at Urbana-Champaign
Jianqiu	Zheng	Pacific Northwest National Laboratory
Xue	Zheng	Lawrence Livermore National Laboratory
Wei	Zhi	Pennsylvania State University
Xin	Zhou	Brookhaven National Laboratory
Margaret	Zimmer	University of California, Santa Cruz
Paquita	Zuidema	University of Miami
Jacob	Zwart	U.S. Geological Survey

Appendix E: AI4ESP Workshop White Papers

Digital copies of the white papers can be downloaded from the https://ai4esp.org website.

Number	Title	Lead Author	DOI
AI4ESP1001	A Spatiotemporal Sequence Forecasting Platform to Advance the Prediction of Changing Spatiotemporal Patterns of CO_2 Concentration by Incorporating Human Activity and Hydrological Extremes	Melissa R. Allen-Dumas	10.2172/1769653
AI4ESP1002	Robust data-driven uncertainty quantification in water cycle extreme predictions	Gemma J. Anderson	10.2172/1769775
AI4ESP1003	A Multi-Scale Inference, Estimation, and Prediction Engine for Earth System Modeling	Marian Anghel	10.2172/1769648
AI4ESP1004	Open-source AI-ready data for prediction of coastal water and carbon budgets under a changing climate	Bhavna Arora	10.2172/1769758
AI4ESP1005	Event-scale predictions of water and nitrogen exports in coastal watersheds	Bhavna Arora	10.2172/1769706

AI4ESP1006	Predictive Understanding of Compound and Cascading Extremes and Their Impacts	Moetasim Ashfaq	10.2172/1769764
AI4ESP1007	Leveraging machine learning to improve understanding and predictability of weather/climate extremes and the resilience of human systems	Karhik Balaguru	10.2172/1769737
AI4ESP1008	AI-enabled MODEX and edge-computing over 5G for improving the predictability of water cycle extremes	Prasanna Balaprakash	10.2172/1769672
AI4ESP1009	Automation Is All You Need: Faster Earth Systems Models with AI/ML	Kenneth Ball	10.2172/1769679
AI4ESP1010	Interpretable Deep Learning for the Earth System with Fractal Nets	Carolyn Begeman	10.2172/1769730
AI4ESP1011	Characterization of Extreme Hydroclimate Events in Earth System Models using ML/AI	Katrina E. Bennett	10.2172/1769685
AI4ESP1012	Past and Future Trends of Severe Storms	Emily Bercos-Hickey	10.2172/1769759

AI4ESP1013	AI-Automated Detection of Subgrid-scale Processes for Adaptivity Guidance	Julie Bessac	10.2172/1769664
AI4ESP1014	Enhanced prediction of terrestrial feedbacks to the coastal carbon cycle: using machine learning to improve sub-grid biogeochemical processes	Nick Bouskill	10.2172/1769704
AI4ESP1015	EAM-HLR: Enhancing the low-resolution E3SM Atmosphere Model with an ML model of high-low-resolution residual in convective processes	Andrew M. Bradley	10.2172/1769697
AI4ESP1016	AI-Improved Resolution Projections of Population Characteristics and Imperviousness Can Improve Resolution and Accuracy of Urban Flood Predictions	Christa Brelsford	10.2172/1769673
AI4ESP1017	Transforming ESM Physical Parameterization Development Using Machine Learning Trained on Global Cloud-Resolving Models and Process Observations	Christopher S. Bretherton	10.2172/1769790

AI4ESP1018	Elucidating and predicting the dynamic evolution of water and land systems due to natural and energy-related forcings	Grant Bromhal	10.2172/1769701
AI4ESP1019	Learning from learning machines: improving the predictive power of energy-water-land nexus models with insights from complex measured and simulated data	James Bentley Brown	10.2172/1769736
AI4ESP1020	Climate Intervention Assessment and Attribution	Diana Bull	10.2172/1769687
AI4ESP1021	AI-Assisted Parameter Tuning Will Speed Development and Clarify Uncertainty in E3SM	Peter M. Caldwell	10.2172/1769663
AI4ESP1022	Building Intelligent Cyberinfrastructure to Learn Iteratively from both Observations and Models for Understanding Watershed Dynamics	Xingyuan Chen	10.2172/1769684
AI4ESP1023	Upscaling cross-scale flow and respiration interactions at river sediment interface leveraging observation, numerical models, and machine learning	Yunxiang Chen	10.2172/1769792

AI4ESP1024	Enhancing Resilience of Urban Systems Against Climate-Induced Floods Using Advanced Data-Driven and Computing Techniques: A Driver-Pressure-State-Impact-Response (DPSIR) Framework	Gyan Chhipi-Shrestha	10.2172/1769705
AI4ESP1025	Integrating Models with Real-time Field Data for Extreme Events: From Field Sensors to Models and Back with AI in the Loop	Shreyas Cholia	10.2172/1769727
AI4ESP1026	Model Hierarchy for Mountainous Hydrological Observatories (MH2O)	William D. Collins	10.2172/1769748
AI4ESP1027	Tracking Extremes in Exascale Simulations Utilizing Exascale Platforms	William D. Collins	10.2172/1769788
AI4ESP1028	Framework for an adaptive integrated observation system using a hierarchy of machine learning approaches	Jennifer Comstock	10.2172/1769712
AI4ESP1029	An AI-Assisted Approach to Represent Human Influence on Surface and Subsurface Hydrology	Ethan Coon	10.2172/1769674

AI4ESP1030	A library of AI-assisted FAIR water cycle and related disturbance datasets to enable model training, parameterization and validation	Robert Crystal-Ornelas	10.2172/1769646
AI4ESP1031	Revolutionizing observations and predictability of Arctic system dynamics through next-generation dense, heterogeneous and intelligent wireless sensor networks with embedded AI	Baptiste Dafflon	10.2172/1769774
AI4ESP1032	Machine learning to extend and understand the sources and limits of water cycle predictability on subseasonal-to-decadal timescales in the Earth system	Katherine Dagon	10.2172/1769744
AI4ESP1033	AI-Driven Data Discovery to Improve Earth System Predictability	Ranjeet Devarakonda	10.2172/1769671
AI4ESP1034	Semi-automated Design of Artificial Intelligence Earth Systems Models	Philipe Dias	10.2172/1769777
AI4ESP1035	Using AI to build a hydrobiogeochemical soil model	Beth A. Drewniak	10.2172/1769793

AI4ESP1036	AI for Extreme Volcanic Climate Forcing and Feedback Forecasting in the 21st century	Manvendra Dubey	10.2172/1769659
AI4ESP1037	Knowledge-Guided Machine Learning (KGML) Platform to Predict Integrated Water Cycle and Associated extremes	Dipankar Dwivedi	10.2172/1769733
AI4ESP1038	A Modular System for Increasing Predictiveness for Extreme Climate Predictions	Christopher Rakauckas	10.2172/1769647
AI4ESP1039	Jaynesian Analysis of Environmental Chemistry: Systems Model Component Integration via the Arctic Aquatic Carbon Cycle	Scott Elliott	10.2172/1769731
AI4ESP1040	Rapid assimilation and analysis of a suit of remote sensing data for predicting extreme events and their impact on ecological-human systems	Nicola Falco	10.2172/1769770
AI4ESP1041	Develop a weather-aware climate model to understand and predict extremes and associated power outages and renewable energy shortages with uncertainty-aware and physics-informed machine learning	Jiwen Fan	10.2172/1769695

AI4ESP1042	On AI Prediction of Hydrological Processes Based on Integration of Retrospective and Forecasting ML Techniques	Boris Faybishenko	10.2172/1769756
AI4ESP1043	Reliable modeling and prediction of precipitation & radiation for mountainous hydrology	Daniel Feldman	10.2172/1769771
AI4ESP1044	Characterization of Extremes and Compound Impacts: Applications of Machine Learning and Interpretable Neural Networks	Yan Feng	10.2172/1769686
AI4ESP1045	Land Surface Modeling 2.0 for agricultural climate change impact assessments	James A. Franke	10.2172/1769734
AI4ESP1046	A Grand Challenge "Uncertainty Project" to Accelerate Advances in Earth System Predictability: AI-Enabled Concepts and Applications	Ann Fridlind	10.2172/1769643
AI4ESP1047	Science-integrated Artificial-intelligence for Flooding and precipitation Extremes (SAFE)	Auroop R. Ganguly	10.2172/1769776
AI4ESP1048	Deep Learning for Ensemble Forecasting	Andrew Geiss	10.2172/1769692

AI4ESP1049	Toward Hybrid Physics -Machine Learning to improve Land Surface Model predictions	Mangistu (Stu) Geza	10.2172/1769785
AI4ESP1050	Geophysical Retrievals in an Artificial Intelligence (AI) Framework for Illuminating Processes Controlling Water Cycle	Virendra P. Ghate	10.2172/1769714
AI4ESP1051	AI Automated Discovery of New Climate Water System Knowledge from Models and Observations	André Goncalves	10.2172/1769658
AI4ESP1052	Autonomous reinforcement learning agents for improving predictions and observations of extreme climate events	André Goncalves	10.2172/1769680
AI4ESP1053	Data-Driven Exploration of Climate Attractor Manifolds For Long-Term Predictability	Carlo Graziani	10.2172/1769691
AI4ESP1054	Feature Detection	Yawen Guan	10.2172/1769711
AI4ESP1055	Modeling Noise: Paths toward AI-Enabled Stochastic Earth System Models and Parameterizations	Samson Hagos	10.2172/1769749

AI4ESP1056	Making Atmospheric Convective Parameterizations Obsolete with Machine Learning Emulation	Walter Michael Hannah	10.2172/1769746
AI4ESP1057	The Usage of Observing System Simulation Experiments and Reinforcement Learning to Optimize Experimental Design and Operation	Joseph C. Hardin	10.2172/1769782
AI4ESP1058	Machine Learned Radiative Transport for Enhanced Resolution Earth System Modeling	Benjamin Hillman	10.2172/1769738
AI4ESP1059	Integrating AI with physics-based hydrological models and observations for insight into changing climate and anthropogenic impacts	Ben R. Hodges	10.2172/1769725
AI4ESP1060	AI-Constrained Bottom-Up Ecohydrology and Improved Prediction of Seasonal, Interannual, and Decadal Flood and Drought Risks	Forrest M. Hoffman	10.2172/1769668
AI4ESP1061	Deep learning techniques to disentangle water use efficiency, climate change, and carbon sequestration across ecosystem scales1	Jennifer A. Holm	10.2172/1769694

AI4ESP1062	Multi-scale Multi-physics Scientific Machine Learning for Water Cycle Extreme Events Identification, Labelling, Representation, and Characterization	Zhangshuan (Jason) Hou	10.2172/1769751
AI4ESP1063	Subseasonal-to-seasonal Prediction of Atmospheric Rivers in the Western United States	Huanping Huang	10.2172/1769780
AI4ESP1064	The use of soil moisture and Standardized Evaporative Stress Ratio (SESR) anomalies for increased lead time of the development flash drought and heat waves	Eric Hunt	10.2172/1769783
AI4ESP1065	Towards Trustworthy and Interpretable Deep Learning-assisted Ecohydrological Models	Peishi Jiang	10.2172/1769787
AI4ESP1066	Combining artificial intelligence, Earth observations, and climate models to improve predictability of ice-biogeochemistry interactions	Grace E. Kim	10.2172/1769689
AI4ESP1067	A Quantum-Ai Framework for Extreme Weather Prediction	Grace E. Kim	10.2172/1769650

AI4ESP1068	Improving Short Term Predictability of Hydrologic Models with Deep Learning	Ryan King	10.2172/1769722
AI4ESP1069	Advancing the Predictability of Water Cycle Phenomena via the Application of AI to Model Ensemble Simulations and Observations	Stephen A. Klein	10.2172/1769656
AI4ESP1070	Multisensor Agile Adaptive Sampling of Convective Storms Driven by Real-time Analytics	Pavlos Kollias	10.2172/1769753
AI4ESP1071	Modular hybrid modeling to increase efficiency, explore structural uncertainty, and allow multidimensional complexity scaling in land surface models.	Charles Koven	10.2172/1769750
AI4ESP1072	End-to-End Differentiable Modeling and Management of the Environment	Christopher Krapu	10.2172/1769703
AI4ESP1073	Representing the Unrepresented Impact of River Ice on Hydrology, Biogeochemistry, Vegetation, and Geomorphology: A Hybrid Physics-Machine Learning Approach	Jitendra Kumar	10.2172/1769772

AI4ESP1074	In Situ Inference for Earth System Predictability	Earl Lawrence	10.2172/1769723
AI4ESP1075	Toward the Development of New Parameterizations for Surface Fluxes	Temple R. Lee	10.2172/1769786
AI4ESP1076	Physics-Informed Learning for Predictive Multi-Scale Modeling of Water Cycle and Extreme Events	Lai-Yung (Ruby) Leung	10.2172/1769761
AI4ESP1077	Deep Learning for Hydro-Biogeochemistry Processes	Li Li	10.2172/1769693
AI4ESP1078	A Self-Evolution Data Fusion Platform for Large-Scale Water Models	Xinya Li	10.2172/1769652
AI4ESP1079	Structurally flexible cloud microphysics, observationally constrained at all scales via ML-accelerated Bayesian inference	Marcus Van Lier-Walqui	10.2172/1769779
AI4ESP1080	Building an AI-enhanced modeling framework to address multiscale predictability challenges	Yangang Liu	10.2172/1769683

AI4ESP1081	A Bayesian Neural Network Ensemble Approach for Improving Large-Scale Streamflow Predictability	Dan Lu	10.2172/1769641
AI4ESP1082	An AI-Enabled MODEX Framework for Improving Predictability of Subsurface Water Storage across Local and Continental Scales	Dan Lu	10.2172/1769675
AI4ESP1083	Advancing Regional Climate Predictability through ML-enabled Dynamical System Approach	Jian Lu	10.2172/1769654
AI4ESP1084	Machine Learning for Surrogate Modeling of the Upper Ocean and Heat Exchange Between the Ocean and Atmosphere	Nicholas Lutsko	10.2172/1769742
AI4ESP1085	Facilitating better and faster simulations of aerosol-cloud interactions in Earth system models	Po-Lun Ma	10.2172/1769709
AI4ESP1086	Assessing Teleconnections-Induced Predictability of Regional Water Cycle on Seasonal to Decadal Timescales Using Machine Learning Approaches	Salil Mahajan	10.2172/1769676

AI4ESP1087	Identifying precursors of daily to seasonal hydrological extremes over the USA using deep learning techniques and climate model ensembles	Nicola Maher	10.2172/1769719
AI4ESP1088	Separating Climate Signals with Machine Learning	Ankur Mahesh	10.2172/1769778
AI4ESP1089	AI-Based Integrated Modeling and Observational Framework for Improving Seasonal to Decadal Prediction of Terrestrial Ecohydrological Extremes	Jiafu Mao	10.2172/1769666
AI4ESP1090	Surrogate multi-fidelity data and model fusion for scientific discovery and uncertainty quantification in Earth System Models	Romit Maulik	10.2172/1769781
AI4ESP1091	Trustworthy AI for Extreme Event Prediction and Understanding	Amy McGovern	10.2172/1769791
AI4ESP1092	Computationally Tractable High-Fidelity Representation of Global Hydrology in ESMs via Machine Learning Approaches to Scale-Bridging	Richard Tran Mills	10.2172/1769690

AI4ESP1093	Rethink hydrologic modeling framework with AI integrating multi-processes across scales	Eugene Yan	10.2172/1769773
AI4ESP1094	New Understanding of Cloud Processes via Unsupervised Cloud Classification in Satellite Images	Elisabeth J. Moyer	10.2172/1769754
AI4ESP1095	EdgeAI: How to Use AI to Collect Reliable and Relevant Watershed Data	Maruti Kumar Mudunuru	10.2172/1769700
AI4ESP1096	Machine Learning for Adaptive Model Refinement to Bridge Scales	Juliane Mueller	10.2172/1769741
AI4ESP1097	Machine Learning to Enable Efficient Uncertainty Quantification, Data Assimilation, and Informed Data Acquisition	Juliane Mueller	10.2172/1769743
AI4ESP1098	Co-Evolving Climate Models under Uncertainty to Improve Predictive Skill	Balu T. Nadiga	10.2172/1769688
AI4ESP1099	A Fire Community Observatory: Interdisciplinary, AI-informed Post-Fire Rapid Response for Improved Water Cycle Science at Watershed Scale	Michelle E. Newcomer	10.2172/1769642

AI4ESP1100	Physics-Informed Deep Learning for Multiscale Water Cycle Prediction	Brenda Ng	10.2172/1769760
AI4ESP1101	AI-Directed Adaptive Multifidelity Modeling of Water Availability and Quality at River Basin Scales	Scott L. Painter	10.2172/1769669
AI4ESP1102	Integration of AI/ML with Data Assimilation for Earth System Prediction2	Stephen G. Penny	10.2172/1769728
AI4ESP1103	Hybrid (PDE+ML) models in the context of land ice modeling	Mauro Perego	10.2172/1769717
AI4ESP1104	Advancing Sea Ice Predictability in E3SM with Machine Learning	Kara Peterson	10.2172/1769655
AI4ESP1105	FAIR data infrastructure and tools for AI-assisted streamflow prediction	Line Pouchard	10.2172/1769710
AI4ESP1106	Water Cycle-Driven Infectious Diseases as Multiscale, Reliable, Continuously Updating Water Cycle Sensors	Amy Powell	10.2172/1769797
AI4ESP1107	AI-Based Upgrades to Observatories Enabling Data Interoperability	Giri Prakash	10.2172/1769667

AI4ESP1108	Early detection and uncertainty quantification of rapid sea-level rise from Antarctica	Stephen Price	10.2172/1769698
AI4ESP1109	Machine learning and artificial intelligence for wildfire prediction	James Tremper Randerson	10.2172/1769739
AI4ESP1110	Probabilistic Machine Learning and Data Assimilation	Vishwas Rao	10.2172/1769766
AI4ESP1111	AI-Based Approach for Advancing the Understanding of Spatiotemporal Drought Characteristics	Deeksha Rastogi	10.2172/1769665
AI4ESP1112	Predictability and feedbacks of the ocean-soil-plant-atmosphere water cycle: deep learning water conductance in Earth System Model	Alexandre A. Renchon	10.2172/1769763
AI4ESP1113	Machine learning to generate gridded extreme precipitation data sets for global land areas with limited in situ measurements	Mark Risser	10.2172/1769784

AI4ESP1114	Integrating Applied Energy and BER Smart Data Capabilities to Develop a DOE Data Fabric for Energy-Water R&DÂ	Kelly Rose	10.2172/1769726
AI4ESP1115	GANpiler	Barry Rountree	10.2172/1769713
AI4ESP1116	Transfer Operator Framework for Earth System Predictability and Water Cycle Extremes	Adam Rupe	10.2172/1769789
AI4ESP1117	Earth System Model Improvement Pipeline via Uncertainty Attribution and Active Learning	Khachik Sargsyan	10.2172/1769699
AI4ESP1118	A new era of observationally-infused E3SM: GANs for unifying imagery archives	Jon Schwenk	10.2172/1769649
AI4ESP1119	AI to Automate ModEx for Optimal Predictive Improvement and Scientific Discovery	Shawn P. Serbin	10.2172/1769662
AI4ESP1120	Integrative data-driven approaches for characterization & prediction of aerosol-cloud processes	Lyndsay Shand	10.2172/1769729

AI4ESP1121	Integrated parameter and process learning for hydrologic and biogeochemical modules in Earth System Models	Chaopeng Shen	10.2172/1769724
AI4ESP1122	Improved Understanding of Coupled Water and Carbon Cycle Processes through Machine Learning Approaches	Debjani Sihi	10.2172/1769721
AI4ESP1123	AI predicted shifts in watershed hydrodynamics driven by extreme weather and fire	Erica Siirila-Woodburn	10.2172/1769660
AI4ESP1124	Automated Discovery of DOminaNt physics Informed Surrogates (ADDONIS) Framework for Improving Water Cycling Predictability	Kenneth (Chad) Sockwell	10.2172/1769678
AI4ESP1125	Preferential flow in subsurface hydrology: From a century of denial to a decade of addressing it via ML?	Matthias Sprenger	10.2172/1769765
AI4ESP1126	On Demand Machine Learning for Multi-Fidelity Biogeochemistry in River Basins Impacted by Climate Extremes	Carl I Steefel	10.2172/1769757

AI4ESP1127	Emergent Concepts from a Community Ideation on AI4ESP	James Stegen	10.2172/1769702
AI4ESP1128	Bridging Multiscale Processes in Earth System Models with Physics-Guided Hierarchical Machine Learning	Alexander Y. Sun	10.2172/1769682
AI4ESP1129	Machine-Learning-Assisted Hybrid Earth System Modelling	Istvan Szunyogh	10.2172/1769745
AI4ESP1130	Using machine learning and artificial intelligence to improve model-data integrated earth system model predictions of water and carbon cycle extremes	Jinyun Tang	10.2172/1769794
AI4ESP1131	Machine Learning for a-posteriori model-observed data fusion to enhance predictive value of ESM output	Claudia Tebaldi	10.2172/1769740
AI4ESP1132	Learned implicit representations of aerosol chemistry and physics for enhancing the predictability of water cycle extreme events	Christopher W. Tessum	10.2172/1769735

AI4ESP1133	Physics-Informed Machine Learning from Observations for Clouds, Convection, and Precipitation Parameterizations and Analysis	Paul Ullrich	10.2172/1769762
AI4ESP1134	Black-Box Neural System Identification and Differentiable Programming to Improve Earth System Model Predictions	Nathan Urban	10.2172/1769681
AI4ESP1135	Using Machine Learning to Develop a Predictive Understanding of the Impacts of Extreme Water Cycle Perturbations on River Water Quality	Charuleka Varadharajan	10.2172/1769795
AI4ESP1136	Observational Capabilities to Capture Water Cycle Event Dynamics and Impacts in the Age of AI	Charuleka Varadharajan	10.2172/1769755
AI4ESP1137	Using machine learning to improve land use/cover characterization and projection for scenario-based global modeling	Alan Di Vittorio	10.2172/1769796

AI4ESP1138	A science paradigm shift is needed for Earth and Environmental Systems Sciences (EESS) to integrate Knowledge-Guided Artificial Intelligence (KGAI) and lead new EESS-KGAI theories	Nathalie Voisin	10.2172/1769651
AI4ESP1139	AI-Driven Cross-Domain Knowledge Discovery and Hypotheses Generation for Enhanced Earth System Predictability	Svitlana Volkova	10.2172/1769670
AI4ESP1140	Automated Custom Calibration for E3SM	Benjamin M. Wagman	10.2172/1769677
AI4ESP1141	Development of Explainable, Knowledge-Guided AI Models to Enhance the E3SM Land Model Development and Uncertainty Quantification	Dali Wang	10.2172/1769696
AI4ESP1142	A Hybrid Climate Modeling System Using AI-assisted Process Emulators	Jiali Wang	10.2172/1769645
AI4ESP1143	Exploring variability in seasonal average and extreme precipitation using unsupervised machine learning	Michael Wehner	10.2172/1769708

AI4ESP1144	High-Accuracy Module Emulators from Physically-Constrained AI Algorithms	Anthony S. Wexler	10.2172/1769715
AI4ESP1145	Quality Data Essential for Modeling Water Cycles Effectively	John Wu	10.2172/1769769
AI4ESP1146	How AI Predicts the Untrained and Unseen	Yuxin Wu	10.2172/1769716
AI4ESP1147	Process-based Neural Network to Forecast Vegetation Dynamics	Chonggang Xu	10.2172/1769768
AI4ESP1148	A HPC Theory-Guided Machine Learning Cyberinfrastructure for Communicating Hydrometeorological Data Across Scales	Haowen Xu	10.2172/1769644
AI4ESP1149	Mapping hydrologic and biogeochemical information flows to improve predictive models and understand climate influence	Zexuan Xu	10.2172/1769747
AI4ESP1150	Multiscale Reduced Order Modeling and Parameter Estimation for Climate Sciences	Arvind Mohan	10.2172/1769752

AI4ESP1151	AI Scaling Laws for Extremes (AISLE)	Da Yang	10.2172/1769661
AI4ESP1152	Process Discovery through Assimilation of Complex Biogeochemical Datasets	Mavrik Zavarin	10.2172/1769767
AI4ESP1153	AI as a Bridge between ARM Observations and E3SM for improving Clouds and Precipitation	Yunyan Zhang	10.2172/1769657
AI4ESP1154	Improve wildfire predictability driven by extreme water cycle with interpretable physically-guided ML/AI	Qing Zhu	10.2172/1769720
AI4ESP1155	Hybridizing Machine Learning and Physically-based Earth System Models to Improve Prediction of Multivariate Extreme Events (AI Exploration of Wildland Fire Prediction)Â	Yufei Zou	10.2172/1769718
AI4ESP1156	Represent precipitation-induced geological hazards in Earth system models using artificial intelligence	Zeli Tan	10.2172/1784543

Environmental Science Division

Argonne National Laboratory
9700 South Cass Avenue, Bldg. 240
Lemont, IL 60439

www.anl.gov