

REVIEW

A Review of Recent and Emerging Machine Learning Applications for Climate Variability and Weather Phenomena

MARIA J. MOLINA^{a,b}, TRAVIS A. O'BRIEN,^{c,d} GEMMA ANDERSON,^e MOETASIM ASHFAQ,^f KATRINA E. BENNETT,^g WILLIAM D. COLLINS,^{d,h} KATHERINE DAGON,^b JUAN M. RESTREPO,^{f,i} AND PAUL A. ULLRICH^j

^a University of Maryland, College Park, College Park, Maryland

^b National Center for Atmospheric Research, Boulder, Colorado

^c Indiana University Bloomington, Bloomington, Indiana

^d Lawrence Berkeley National Laboratory, Berkeley, California

^e Lawrence Livermore National Laboratory, Livermore, California

^f Oak Ridge National Laboratory, Oak Ridge, Tennessee

^g Los Alamos National Laboratory, Los Alamos, New Mexico

^h University of California, Berkeley, Berkeley, California

ⁱ University of Tennessee, Knoxville, Tennessee

^j University of California at Davis, Davis, California

(Manuscript received 5 December 2022, in final form 9 June 2023, accepted 9 June 2023)

ABSTRACT: Climate variability and weather phenomena can cause extremes and pose significant risk to society and ecosystems, making continued advances in our physical understanding of such events of utmost importance for regional and global security. Advances in machine learning (ML) have been leveraged for applications in climate variability and weather, empowering scientists to approach questions using big data in new ways. Growing interest across the scientific community in these areas has motivated coordination between the physical and computer science disciplines to further advance the state of the science and tackle pressing challenges. During a recently held workshop that had participants across academia, private industry, and research laboratories, it became clear that a comprehensive review of recent and emerging ML applications for climate variability and weather phenomena that can cause extremes was needed. This article aims to fulfill this need by discussing recent advances, challenges, and research priorities in the following topics: sources of predictability for modes of climate variability, feature detection, extreme weather and climate prediction and precursors, observation–model integration, downscaling, and bias correction. This article provides a review for domain scientists seeking to incorporate ML into their research. It also provides a review for those with some ML experience seeking to broaden their knowledge of ML applications for climate variability and weather.

KEYWORDS: Extreme events; Numerical analysis/modeling; Climate variability; Artificial intelligence; Deep learning; Machine learning

1. Introduction

Climate variability and weather phenomena that cause extremes can significantly stress natural and human resources and result in costly disasters that claim lives (Thornton et al. 2014; Strader and Ashley 2015; Hu et al. 2018; Seneviratne et al. 2021; Ebi et al. 2021). According to the American Meteorological Society (AMS), climate variability refers to “temporal variations of the atmosphere–ocean system around a mean state” (American Meteorological Society 2022), and thus typically comprises phenomena of subseasonal (minimum of 2 weeks) or longer time scales (Fig. 1), which can have natural (e.g., volcanic eruptions and

orbital cycles that impact insolation) and/or anthropogenic sources [e.g., greenhouse gases (GHG) and sulfate/biomass burning aerosols; Ghil 2002; Salinger 2005]. Weather phenomena that can cause extremes include, for example, tropical cyclones (TCs), atmospheric rivers (ARs), and mesoscale convective systems (MCSs; Fig. 1). While the frequency, intensity, and spatial extent of weather features that cause extremes can also be influenced by multiscale modes of climate variability (Ropelewski and Halpert 1986; Molina et al. 2018; Kim et al. 2019; Lin et al. 2020), we purposefully delineate the two to aid in describing the challenges that remain in their understanding. For example, it is not clear whether climate variability is sufficiently captured in Earth system models (ESMs) because of limited observational records. Moreover, ESM grid spacing for climate time scales (i.e., future projections) is usually too coarse (\sim 100 km) to

Corresponding author: Maria J. Molina, mjmolina@umd.edu

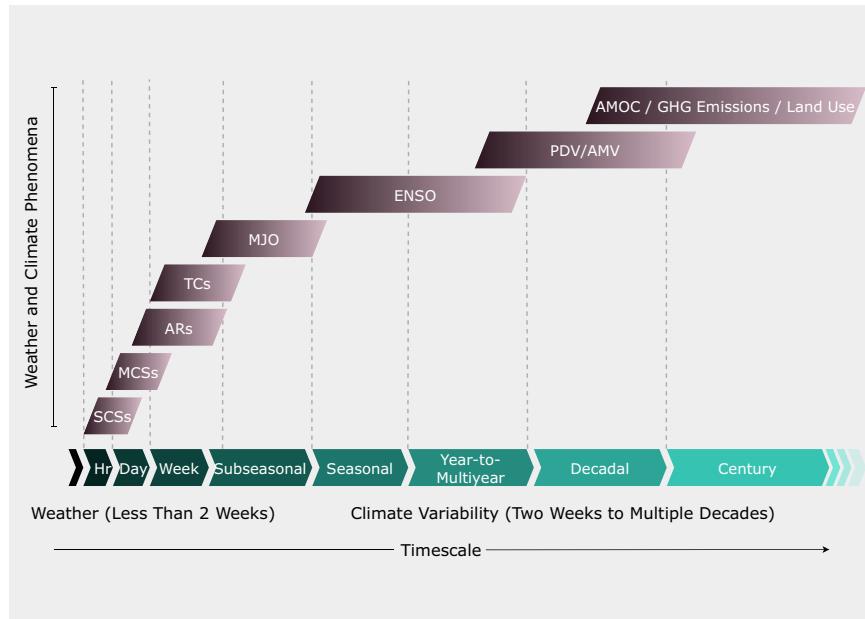


FIG. 1. Time scales of example weather and climate phenomena. Listed example phenomena include severe convective storms (SCSSs), mesoscale convective systems (MCSs), atmospheric rivers (ARs), tropical cyclones (TCs), Madden Julian oscillation (MJO), El Niño-Southern Oscillation (ENSO), Pacific decadal variability (PDV), Atlantic multidecadal variability (AMV), Atlantic meridional overturning circulation (AMOC), and greenhouse gases (GHG). The figure is adapted from [Meehl et al. \(2021\)](#).

capture fine-scale properties of weather phenomena that lead to extremes (~ 50 km or less; see, e.g., [Li et al. 2013](#)), especially considering that the effective resolution of numerical weather prediction (NWP) models and contemporary climate models is several times more coarse than the actual grid spacing due to the effects of diffusion ([Skamarock 2004](#); [Klaver et al. 2020](#)). This paper is not about the stress on natural and human resources, but about the machine learning methods and tools to study those stresses.

Across the broader scientific community, rapid advances in machine learning (ML; [Haupt et al. 2022](#)), and deep learning (DL; [LeCun et al. 2015](#)) in particular, have inspired researchers to consider how these tools might enable science advances that previously would have been unattainable. Since DL is a subset of ML, we use ML to refer to both herein for simplicity. For climate variability and weather, the appeal in ML stems in part from its ability to model complex nonlinear systems and in part from recent algorithmic and computational advances [e.g., graphics processing units (GPUs)] that have improved and accelerated ML model training. ML advances have enabled scientists and software engineers to probe difficult questions that span climate variability and weather, and have been a driving force behind recent national and international workshops (e.g., [Chantry et al. 2021](#)). In particular, the Earth and Environmental Systems Science Division and Advanced Scientific Computing Research of the U.S. Department of Energy Office of Science held a 6-week, large community workshop during the autumn of 2021 to foster research coordination across numerous disciplines and professional

sectors ([Hickmon et al. 2022](#)). This 6-week workshop included an extended session on climate variability and weather phenomena that can cause extremes. One of the planned outcomes included a synthesis of recent progress and key challenges identified during the extended session, which motivated a more comprehensive review of research pertaining to ML applications for climate variability and weather phenomena, contained herein.

This paper targets two primary audiences: 1) domain scientists conducting research on climate variability and weather phenomena that can cause extremes (CV&E) who may have not yet incorporated ML into their research and 2) researchers with some ML experience seeking to broaden their knowledge of ML applications in the field of CV&E. The goals of this paper are to

- 1) help domain scientists understand the range of ML tools that already exist and have been applied to CV&E to thus avoid “reinventing the wheel”,
- 2) discuss challenges that exist in certain applications of ML (and why off-the-shelf ML tools may not always be appropriate for CV&E), and
- 3) outline future directions in ML and CV&E that we believe will have the most impact in advancing the field forward.

2. Overview of ML uses and challenges in CV&E

Here we discuss how ML has been recently used in CV&E and highlight some remaining challenges specific to ML for

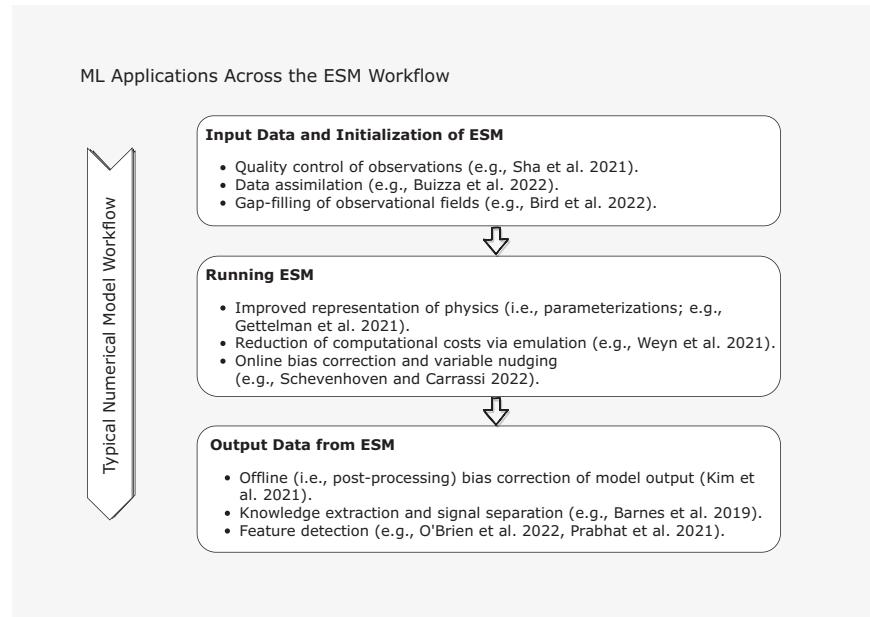


FIG. 2. Example machine learning applications across the typical Earth system modeling workflow include input data and ESM initialization (e.g., Sha et al. 2021; Buizza et al. 2022; Bird et al. 2023), running the ESM (e.g., Gettelman et al. 2021; Weyn et al. 2021; Schevenhoven and Carrassi 2022), and ESM output (e.g., Kim et al. 2021; Barnes et al. 2019; O'Brien et al. 2022; Prabhat et al. 2021).

CV&E. ML applications in CV&E can fall within certain stages of the ESM workflow, which we delineate as initializing the ESM, running the ESM, and postprocessing ESM output (Fig. 2). However, recent studies focusing on extreme weather prediction using large high-resolution observations (on meter to kilometer scales) cannot be placed within stages of the ESM workflow. These openly available large observational datasets, such as from radar (e.g., Multi-Radar Multi-Sensor; Smith et al. 2016; Zhang et al. 2016) and satellite instruments (e.g., NOAA GOES Advanced Baseline Imager; Schmit et al. 2017), made it possible to advance extreme weather prediction using ML (e.g., Lagerquist et al. 2020; Lee et al. 2021). Other topics that have relevance to CV&E but are not discussed within this article, such as representation of subgrid processes (e.g., parameterizations), can be found in the associated workshop report (Hickmon et al. 2022).

a. Sources of predictability for modes of climate variability

There are several climate variability phenomena with strong global influence, such as El Niño–Southern Oscillation (ENSO; Rasmusson and Carpenter 1982; Trenberth 1997) and the Madden–Julian oscillation (MJO; Madden and Julian 1971, 1972), and prediction of these phenomena and associated teleconnections has seen substantial progress in recent years with ML (Table 1; e.g., Ham et al. 2019; Mayer and Barnes 2021; Martin et al. 2022; Gordon and Barnes 2022). For example, Ham et al. (2019) trained a convolutional neural network to predict ENSO at lead times beyond 1 yr with skill that exceeds state-of-the-art dynamical forecast models. Mayer and Barnes

(2021) predicted the sign of geopotential height anomalies on subseasonal time scales (i.e., from 2 weeks to 2 months) using an artificial neural network (ANN). These studies also went beyond generating skillful predictions by using explainable artificial intelligence (XAI) methods, such as layerwise relevance propagation (Toms et al. 2020), to explore the sources of predictability that trained neural networks (NNs) identified as resulting in good forecast performance (McGovern et al. 2019). ML can also be used to assess the representation of modes of climate variability in large climate simulations (e.g., Maher et al. 2022) and to quantify their associated teleconnections in a causal framework (McGraw and Barnes 2018; Runge et al. 2019; Kretschmer et al. 2021). Causal methods include two overarching categories: *causal discovery*, in which models that illustrate cause-and-effect relationships are created (e.g., causal graphs), and *causal inference*, in which impacts to cause-and-effect relationships when a system is modified are explored (Yao et al. 2021; Nogueira et al. 2022).

Despite progress, numerous challenges remain. For example, it remains unclear whether climate variability is sufficiently captured in ESMs due to artificial trends in observational data (e.g., improvements in observing networks over time), data gaps, and scale differences among observational products and ESMs. Uncertainty and consistency across XAI methods is a topic that remains underexplored, but highly relevant to knowledge discovery (Fayaz et al. 2022). For example, Mamalakis et al. (2022) illustrated that inconsistent results can be gleaned from different XAI methods, suggesting that a variant of layerwise relevance propagation and inputs multiplied by their respective gradients produced the correct XAI results for their study as compared with other methods. Thus, identifying potential failure modes of XAI, and uncertainty

TABLE 1. Synthesis of recent ML-based advances for sources of predictability for modes of climate variability.

| Topic | ML application(s) and related work(s) |
|--|--|
| Causal analysis | PCMCI: PC-based (Spirtes et al. 2000) multivariate causal discovery for time series datasets (Runge et al. 2019) Quantifying the effect of soil moisture on precipitation using causal inference and random forests (Li et al. 2020) Causal inference and regressions for quantification of teleconnection pathways (Kretschmer et al. 2021) |
| Climate prediction and sources of predictability | Multiyear ENSO prediction using a CNN (Ham et al. 2019) Identification of subseasonal forecasts of opportunity using explainable ANNs (Mayer and Barnes 2021) ANN-based prediction of the MJO (Martin et al. 2022) ANN with uncertainty for detecting decadal state-dependent predictability (Gordon and Barnes 2022) CNN to predict ENSO with 9-month lead time and uncover sources of predictability with contribution maps and sensitivity (Shin et al. 2022) Viewing forced climate signals using an ANN (Barnes et al. 2019) |
| Signal separation and anomaly detection | Unsupervised physics-based representation learning of spatiotemporal coherent structures (Rupe et al. 2019b) Explainable AI methods, including feature importance, partial-dependence plots, saliency maps, and layerwise relevance propagation (McGovern et al. 2019; Toms et al. 2020) Anomaly detection for physics analysis including supervised, unsupervised, and weakly supervised methods (Nachman 2022) Physical causes of climate change on Midwest extreme precipitation using a CNN (Davenport and Diffenbaugh 2021) ENSO representation in large ensembles using ensemble voting classifier (consisting of an ANN, random forest, and nearest neighbor; Maher et al. 2022) |

quantification pertaining to different types of XAI methods, are both crucial to establish confidence levels in XAI output and determine whether ML predictions are “right for the right reasons.” Collinearity among input variables also presents challenges for XAI attribution studies, limiting our ability to quantify variable importance or contributions to ML model predictions. This issue is particularly acute in the Earth sciences, where many physical variables are interrelated. For example, [Kamangir et al. \(2022\)](#) showed that their input variables, wind and temperature, were correlated, which could affect fidelity of their attribution results. Variants to the permutation feature importance test have been used to assess sensitivity to collinearity (e.g., multipass forward algorithm, single-pass backward algorithm, multipass backward algorithm), although they are computationally expensive ([Lagerquist 2020](#)). Additionally, while XAI can reveal correlations between input features and outputs, the statistics adage states: “correlation does not imply causation.” Currently, many statistical and explainable ML methods do not consider causality, and if used to infer predictability sources, may result in spurious or overconfident sources of predictability for modes of climate variability and teleconnectivity of natural modes of climate variability to regional climate or weather extremes. Thus, sources of predictability need to be better quantified using causal frameworks (e.g., [Kretschmer et al. 2021](#)). Human biases can also leak into the scientific process when using ML for knowledge discovery, which can lead to the discarding of new information that may be in conflict with past established work ([Molnar et al. 2021; Molina et al. 2021; Kumar and Sharma 2022](#)).

b. Feature detection

Over the last decade, there has been a considerable increase in research associated with the detection of mesoscale and synoptic-scale weather phenomena (i.e., features) that can cause extremes, such as tropical cyclones (e.g., [Walsh et al. 2015](#)), atmospheric rivers (e.g., [O'Brien et al. 2022](#)), and mesoscale convective systems (e.g., [Feng et al. 2021](#)). The spatial extent of the mesoscale is usually defined as a few kilometers to several hundred kilometers, in contrast to the spatial extent of the synoptic scale, which is usually defined as a few hundred to several thousand kilometers. The majority of feature detection research to date uses objective, heuristic (i.e., threshold based) detection algorithms; however, in the past several years, objective approaches using ML have appeared (see a summary in Table 2).

Recently, for example, [Liu et al. \(2016\)](#) trained a convolutional neural network (CNN) to detect tropical cyclones, atmospheric rivers, and fronts in ESM output, using a combination of heuristic detection algorithm outputs and hand-labeled data for training. [Lagerquist et al. \(2019\)](#) likewise used a CNN to detect fronts in reanalysis data, using human-labeled fronts as the training dataset. [Prabhat et al. \(2021\)](#) built on the work of [Liu et al. \(2016\)](#) and trained a deep CNN to detect atmospheric rivers and tropical cyclones in a climate simulation using a large dataset of expert-labeled fields. [Quinting and Grams \(2022\)](#) trained a deep CNN to detect warm conveyor belts in NWP output, using an analysis of Lagrangian tracer advection as the training dataset, showing that the model works on a range of

TABLE 2. Synthesis of recent ML-based advances for feature detection in CV&E. In addition to NN, ANN, and CNN, which are defined in the main text, several other abbreviations are used for brevity: recurrent NN (RNN), residual NN (ResNet), generative adversarial network (GAN), long short-term memory network (LSTM), and U-Net (a type of CNN; [Ronneberger et al. 2015](#)).

| Topic | ML application(s) and related work(s) |
|--|--|
| Extremes and related-feature detection | CNNs for detection of extremes (Liu et al. 2016) Self-organizing maps for climate extremes (Gibson et al. 2017) A CNN and RNN approach for simultaneous tracking of weather phenomena (Mudigonda et al. 2017) Detection of fronts with CNNs and U-Nets (Biard and Kunkel 2019; Lagerquist et al. 2019; Dagon et al. 2022; Niebler et al. 2022) Statistical-heuristic sensitivity analysis and Bayesian framework for detecting atmospheric rivers and tropical cyclones (Zarzycki and Ullrich 2017; O'Brien et al. 2020) ClimateNet: CNN/ResNet architecture for extremes detection using hand-labeled targets (Prabhat et al. 2021) Unsupervised nonnegative matrix/tensor factorization coupled with k -means clustering for characterizing drought (Talsma et al. 2022) North American weather regime detection and predictability analysis using k -means clustering (Molina et al. 2023) Physics constraints and normalization for NNs (Beucler et al. 2020) Physics-informed ML for weather and climate using GANs, CNNs, LSTMs, and others (Kashinath et al. 2021) Assessment of CNN classification robustness to nonstationarity for severe thunderstorms (Molina et al. 2021) |
| Out-of-sample generalization | Physics constraints and normalization for NNs (Beucler et al. 2020) Physics-informed ML for weather and climate using GANs, CNNs, LSTMs, and others (Kashinath et al. 2021) Assessment of CNN classification robustness to nonstationarity for severe thunderstorms (Molina et al. 2021) |

input datasets ([Quinting et al. 2022](#)). There have also been several attempts to use unsupervised approaches to detect weather phenomena. [Radić et al. \(2015\)](#) used self-organizing maps to characterize changes in different types of extreme atmospheric rivers in climate model projections. [Rupe et al. \(2019a,b\)](#) demonstrated that computational mechanics (specifically, ϵ machines) can objectively and automatically detect eddy-like features in fluid simulations. There have also been some efforts to combine heuristic detection approaches with optimization based on expert-labeled datasets: a form of statistical-heuristic machine learning that has been used for tropical cyclones ([Zarzycki and Ullrich 2017](#)) and for atmospheric rivers ([O'Brien et al. 2020](#)). Among the advances outlined in this paragraph, detection of localized extremes at the microscale (such as hail and tornadoes) that pose significant societal danger remains a gap in feature detection work at climate scales due to several reasons, such as (i) insufficient spatial resolution to explicitly capture these events (e.g., ERA5 is \sim 32-km horizontal grid spacing; [Hersbach et al. 2020](#)), (ii) lack of labels globally (i.e., in situ observations to use as predictands), and (iii) limitations in label quality (e.g., hail reports are clustered on roads; [Allen and Tippett 2015](#)). However, severe weather hazard prediction and probabilistic guidance using ML has made strides, having been recently adopted by the National Weather Service and included in experimental NWP (e.g., the NOAA Warn-on-Forecast; [Flora et al. 2021; Clark and Loken 2022](#)). Extreme weather prediction is discussed further in section 2c.

ML approaches to detect or track weather phenomena have not yet been adopted widely, which may be partly due to research questions that still need to be addressed. Key questions are related to transferability across ESMs and the lack of interpretability of results, for example: will an ML model trained to

detect features in one specific region perform well in another; will an ML model trained to detect features on the present climate perform well in a different (e.g., future) climate, and will an ML model trained to detect features on a specific dataset or ESM perform well on different ESMs? Uncertainty in detecting weather phenomena also stems in part from ambiguities in the definitions of these phenomena as provided by domain experts. For example, while much effort was placed in creating a formal definition for “atmospheric rivers” that the scientific community could agree upon ([Ralph et al. 2018](#)), ambiguities regarding how to delineate the MJO from atmospheric rivers ([Toride and Hakim 2022](#)) and how to define Antarctic atmospheric rivers ([Pohl et al. 2021](#)) persist. Uncertainty in the definition of weather phenomena can also arise during transitional stages (e.g., tropical cyclone to extratropical cyclone). With that said, by employing techniques from explainable ML ([Mamalakis et al. 2022](#)) and considering how ML methods identify features, these studies could yield insights that would allow us to better disambiguate these features. Additionally, lack of training and evaluation data challenges many areas of ML research in Earth system prediction. Some weather phenomena have extensive publicly available databases of expert labels (e.g., IBTrACs for tropical cyclones; [Knapp et al. 2010](#)) whereas other phenomena have no existing standardized datasets (e.g., global monsoon or Antarctic atmospheric rivers; [Geen et al. 2020; Wille et al. 2021](#)). This lack of labeled data is a significant barrier for training supervised algorithms and it inhibits rigorous evaluation of unsupervised methods. There is a major opportunity for the academic community to engage students systematically in the development of such databases. Initiatives such as ClimateNet ([Prabhat et al. 2021](#)), which led to the development of a web interface for labeling tropical cyclones and atmospheric rivers in climate snapshots, will be invaluable for enabling communities of climate experts

TABLE 3. Synthesis of recent ML-based advances for prediction and precursors in CV&E. Here, XGBoost is extreme gradient boosting.

| Topic | ML application(s) and related work(s) |
|---|---|
| Climate and weather extremes prediction | Random forests for prediction of extremes, including precipitation or severe hazards (McGovern et al. 2017 ; Herman and Schumacher 2018 ; Loken et al. 2019 ; Hill and Schumacher 2021) Using a CNN for near-term hail prediction and explainable AI for extraction of prediction signals (Gagne et al. 2019) XGBoost for lightning prediction (Mostajabi et al. 2019) Analog forecasting of extreme-causing weather patterns using capsule neural networks (Chattopadhyay et al. 2020) CNNs for near-term tornado prediction (Lagerquist et al. 2020) Random forests for probabilistic short-term forecasts of severe weather hazards (Loken et al. 2020, 2022 ; Hill et al. 2020) GANs for short-term prediction of precipitation (Ravuri et al. 2021) Severe weather hazard probabilistic guidance derived using tree-based models (Flora et al. 2021 ; Clark and Loken 2022) Rainfall-runoff prediction following extremes using LSTMs and a mass-conserving LSTM variant (Frame et al. 2022) Seasonal probabilistic prediction with conditional generative forecasting (Pan et al. 2022) Predictability of daily southeastern U.S. precipitation using CNNs and layerwise relevance propagation (Pegion et al. 2022) 3D CNN for coastal fog prediction (Kamangir et al. 2022) Extreme heat prediction using CNNs and custom loss functions (Lopez-Gomez et al. 2023) |
| Conditional stationarity | Local causal states and discrete coherent structures (Rupe and Crutchfield 2018) Discovering causal structure with reproducing-kernel Hilbert space ϵ machines (Brodu and Crutchfield 2022) |

to develop such datasets for other climate variability and weather phenomena.

c. Extreme weather and climate prediction and precursors

There has been recent success in the ML-based prediction of weather and climate extremes, along with the identification of associated precursors (Table 3; e.g., [Herman and Schumacher 2018](#); [Sønderby et al. 2020](#); [Ravuri et al. 2021](#); [Frame et al. 2022](#)). At the weather time scale, for example, [Gagne et al. \(2019\)](#) used a convolutional neural network to predict hail with 1-h lead time from convection-permitting (3-km horizontal grid spacing) model output and then identified associated precursors using ML explainability tools. Permutation feature importance and saliency maps ([McGovern et al. 2019](#)) yielded insights into what the convolutional neural network learned for skillful hail prediction, including storm morphology (e.g., supercell) and environmental parameters that contribute to storm intensification and hail growth ([Gagne et al. 2019](#)). At short, medium, and sub-seasonal time scales, [Lopez-Gomez et al. \(2023\)](#) trained convolutional neural networks with custom loss functions to predict surface temperature and extreme heat waves with lead time of 1–28 days on a global coarse grid (about 200 km). On longer climate time scales (decades to centuries), [Dagon et al. \(2022\)](#) detected weather fronts (e.g., cold, warm, occluded, and stationary fronts) using a convolutional neural network and then quantified their association with extreme precipitation over North America along with projected changes in a future climate; thus, feature detection was used to extract precursors to extreme

precipitation events. Skillful ML-based prediction of weather and climate has garnered broad interest from private sector industry as well, although not necessarily focused on extremes. Recent examples include Fourier forecasting neural network (named FourCastNet) for short to medium-range predictions ([Pathak et al. 2022](#)), an autoregressive model using graph neural networks (named GraphCast) with 10-day lead time ([Lam et al. 2022](#)), a 3D Earth Specific Transformer (3DEST) model (named Pangu-Weather) with 1-h to 1-week lead time ([Bi et al. 2022](#)), and ClimaX based on transformers with encoding and aggregation blocks for both weather and climate time scales ([Nguyen et al. 2023](#)).

There are challenges that remain in the prediction and precursor identification of extremes on weather and climate time scales. Extremes by definition are rare events and, therefore, separation of signal from noise in extremes statistics remains demanding due to limited observational records. The limited observational record also presents challenges to understanding covariability and compounding extremes (e.g., [Mukherjee et al. 2020](#)), in addition to long-term trends of extremes. The issue of limited samples of extremes can also manifest as class imbalance, where the minority class may represent the extreme phenomena and be undersampled as compared with the majority class(es). Several studies have overcome class sample size issues within extreme weather prediction or classification as follows: (i) undersampling of the majority class (e.g., [Lagerquist et al. 2019](#)), (ii) robust model and/or hyperparameter grid search to find ML architectures that can

TABLE 4. Synthesis of recent ML-based advances for observation–model integration in CV&E.

| Topic | ML application(s) and related work(s) |
|------------------|---|
| Emulation | Random forests for predicting output of high-resolution climate models (Anderson and Lucas 2018) ANNs for emulation and parameter estimation (Dagon et al. 2020) |
| Hybrid modeling | ANNs for clouds and associated processes in general circulation models (Gettelman et al. 2021) Hybrid modeling: combining physics and data-driven approaches (Reichstein et al. 2019) |
| Synthetic data | Application of physics-guided LSTM for rainfall-runoff and extremes (Xie et al. 2021) MJO-index time series reconstruction using one-dimensional CNNs (Dasgupta et al. 2020) Reconstruction of missing climate observations using partial convolutions (Kadow et al. 2020) Synthetic precipitation generation using ResNets (Bird et al. 2023) |
| Sensor placement | Convolutional Gaussian neural process to maximize measurement informativeness and place sensors efficiently (Andersson et al. 2022) Random forests for soil moisture sensor placement in a future climate (Bessenbacher et al. 2023) |

overcome class imbalance limitations (e.g., Mostajabi et al. 2019; Kamangir et al. 2022), (iii) use of custom class-weighted loss functions (e.g., Ebert-Uphoff et al. 2021), and (iv) others found that the respective ML task was performed skillfully despite severe class imbalance (e.g., Flora et al. 2021; Molina et al. 2021). Characterization of extreme events (e.g., extreme precipitation and floods) also need to be improved given that different predefined thresholds used to identify extremes can result in very different properties and disparate research results (Pendergrass 2018; Leung et al. 2022). Despite existing identified features, precursors to extreme events are still poorly understood and require further identification and characterization. For example, while ML-based detection of tropical cyclones has seen much progress (e.g., Prabhat et al. 2021), predicting which tropical disturbances (e.g., African easterly waves) will intensify into a tropical cyclone or not necessitates further exploration (e.g., Núñez Ocasio et al. 2021).

d. Observation–model integration

The use of ML for observation–model integration (e.g., ModEx approach¹) is motivated by the need to improve the fidelity of ESMs using observations, perform reduced-order modeling, overcome data input/output (I/O) challenges with exascale, and improve assimilation of observations into modeling systems (e.g., Geer 2021; Schevenhoven and Carrassi 2022). Some progress has been recently made on these topics (see Table 4) and examples include the reconstruction of missing climate observations (Kadow et al. 2020), generation of a synthetic MJO-index time series from a past climate (Dasgupta et al. 2020), ESM emulation and parameter estimation (e.g., Lu et al. 2018; Dagon et al. 2020), and the blending of ML and data assimilation via a Bayesian perspective for specific tasks, such as parameterizations (e.g., Bocquet et al. 2020; Brajard et al. 2021; Cleary et al. 2021; Buizza et al. 2022).

Verification and validation of climate model simulations (i.e., a class of ESMs used to create projections of Earth’s future climate) using observations are challenging, particularly for longer time scales (e.g., decadal prediction; Meehl et al. 2021), future projections, and rare events. Challenges related

to verification and validation of climate models are due in part to limited or lack of data, which is a barrier to subsequent development. Observations are often inhomogeneous across space and time, which presents a challenge when gap filling observations and models. ML could help determine where to deploy observations by evaluating how much uncertainty can be reduced with a new data point (Andersson et al. 2022) or projecting where more observations will be needed in a future climate (Bessenbacher et al. 2023). A risk of overparameterization when integrating observations into models also exists and thus there is a need for methods that indicate when observation and model agreement is suboptimal. Uncertainty quantification is at the intersection of observations and model simulations and is an important tool 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 (e.g., Bennett et al. 2022) and Antarctic (e.g., DuVivier et al. 2023). On weather time scales, data latency, which is the length of time between the measurement of an observation and its transmission to a forecasting center, can pose a limit to model skill; earlier incorporation of data as they are collected is needed (e.g., Casey and Cucurull 2022; Slivinski et al. 2022). Dynamic approaches to filtering can help address issues with temporal lags in the assimilation of data (Restrepo 2017; Foster and Restrepo 2022). Additionally, stronger communication and collaboration between observationalists, theoreticians, and process and climate modelers can help further advance our physical understanding, tool development, and fidelity of ESMs [as demonstrated by Legg et al. (2009)]. Moreover, a heavy emphasis has been placed on the development of data-driven ML methods, which may not obey underlying physical principles of Earth systems nor generalize well to unseen events (Reichstein et al. 2019; Kashinath et al. 2021); more development of physics-informed ML methods is needed in observation–model integration, but with consideration of their possible failure modes (e.g., Krishnapriyan et al. 2021; Wang et al. 2022).

e. Downscaling and bias correction

A spatial refinement and/or bias correction of climate models (i.e., a class of ESMs used to create projections of Earth’s future climate) is warranted due their coarse horizontal grid

¹ Iterative model-experiment (ModEx) approach (<https://ess.science.energy.gov/modex/>).

TABLE 5. Synthesis of recent ML-based advances for downscaling and bias correction in CV&E.

| Topic | ML 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) Nudged ESMs using random forests and ANNs (Watt-Meyer et al. 2021 ; Bretherton et al. 2022 ; Clark et al. 2022) One-class support vector machine novelty detector to identify out-of-sample data (Sanford et al. 2022) Unpaired image-to-image translation networks for offline bias correction (Fulton et al. 2023) |
| Downscaling | Downscaling climate model projections using generalized stacked superresolution CNNs (Vandal et al. 2017) CNNs for statistical downscaling (Baño-Medina et al. 2020) Unsupervised, generative downscaling using normalizing flows for variational inference (Groenke et al. 2020) Superresolution GAN (SRGAN; Ledig et al. 2017) for downscaling ESM wind and solar outputs (Stengel et al. 2020) Enhanced SRGANs and content adaptive resampler for downscaled precipitation (Watson et al. 2020) Downscaling precipitation over complex topography using U-Nets and modified U-Nets (Sha et al. 2020a,b) Spatiotemporal downscaling (3D U-Nets) of turbulence with partial differential equations as regularization (Jiang et al. 2020) CNN-based precipitation downscaling (Wang et al. 2021) Spatial and temporal downscaling using CNNs (Serifi et al. 2021) Hard-constrained NNs for physically consistent downscaling (Harder et al. 2022) |

spacing (~ 100 km) that contributes to inaccuracies in representing subgrid-scale Earth system component processes and interactions (e.g., land–atmosphere). Bias correction of ESMs has seen progress in recent years (see Table 5), with corrections learned from data applied after the ESM has been run (i.e., offline) or while the ESM is run (i.e., online). Applications include offline bias correction of weather observations for initial state quality control (e.g., [Sha et al. 2021](#)), offline bias correction of subseasonal climate oscillation predictions, such as the MJO (e.g., [Kim et al. 2021](#)), and online bias correction of global coarse-grid (i.e., not convection-permitting) weather and climate models (e.g., [Watt-Meyer et al. 2021](#); [Bretherton et al. 2022](#); [Clark et al. 2022](#)). There is also some success in the development of ML models to generate higher-resolution output of regional domains (e.g., [Baño-Medina et al. 2020](#)) and in the downscaling of climate model outputs (e.g., [Vandal et al. 2017](#); [Pan et al. 2021](#)). However, most of these efforts are currently limited to snapshot modeling or downscaling (e.g., [Sha et al. 2020a](#)), with fewer studies focusing on downscaling in both space and time (e.g., [Jiang et al. 2020](#); [Serifi et al. 2021](#)).

ML models can break physical laws if constraints or equations are not incorporated during the training process ([Beucler et al. 2020](#); [Kashinath et al. 2021](#)), where for example, a data-driven approach may not capture statistical tails in subgrid-scale physics, which could lead to inaccurate downscaling or bias correction, and numerical instabilities if subsequently run online in an ESM (e.g., [Gettelman et al. 2021](#)). [Yuval et al. \(2021\)](#) achieves numerical stability by using an artificial neural network to predict fluxes rather than tendencies, which serves as a physical constraint in the respective ML-based parameterization. Open-source microscale (1 km or less) benchmarking datasets at high temporal frequencies (e.g., 1 h or less) are also very limited or lacking (e.g., clouds and aerosols), which could be used for standardized evaluations, albeit with a large computational cost at the global scale. There are also challenges in the selection and

availability of reasonable priors for uncertainty quantification from a Bayesian perspective, which is partly because priors can be highly subjective and partly because many computational scientists do not receive formal training in Bayesian methods as part of their standard curricula ([Hoegh 2020](#)). Future advancements in downscaling and bias correction using ML should consider both data driven and physics-based approaches (e.g., [de Wolff et al. 2021](#)), downscaling in both space and time, and blending traditional domain science methods (e.g., fluid mechanics) with ML to fully exploit the strengths of multiple data science approaches. Sufficiently long and high-resolution (e.g., convection-permitting) data streams would be vital to exploiting ML as a tool to emulate physical models representing weather and climate extremes.

3. Prospects for future contributions to CV&E with ML

Previously we detailed ML applications and challenges in CV&E (for a summary, see Fig. 3). Next, we outline research directions and outstanding problems that span across CV&E and were identified as priorities (also summarized in Table 6).

a. ML-detected features and model catalogs

Massive catalogs from feature trackers, such as the Atmospheric River Tracking Method Intercomparison Project (ARTMIP) catalog ([Shields et al. 2018](#)), already exist in consistent format and with uniform standards, and could be used as an example for cataloging existing ML-detected features for benchmarking or research ([Dueben et al. 2022](#)). Already-trained ML models for feature detection, such as ClimateNet ([Prabhat et al. 2021](#)), can also be catalogued, enabling the prioritization of other features with comparatively limited attention (e.g., monsoon depressions).

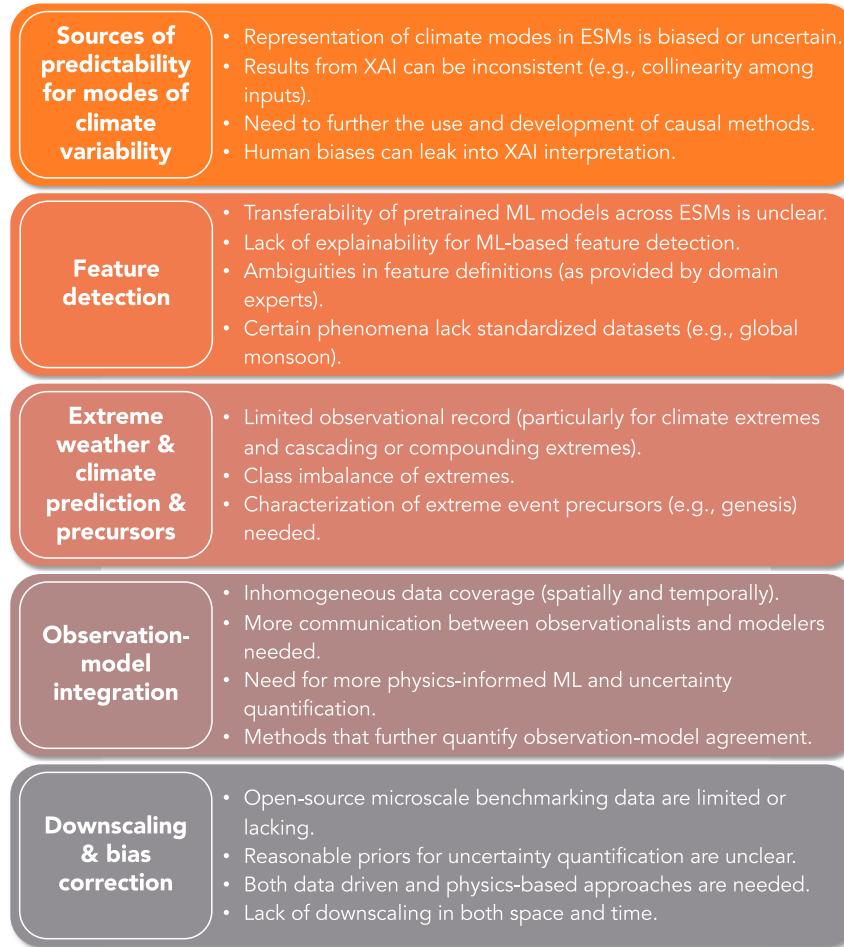


FIG. 3. Synthesis of challenges in CV&E and/or possible needs that ML can address (the full text is given in section 2).

b. Robustness assessment

Further development and benchmarking of XAI methods (e.g., [Mamalakis et al. 2022](#)), assessment of the transferability of trained ML models from one dataset to another (e.g., [Dagon et al. 2022](#)), and whether knowledge extracted is consistent across models and data products, especially in a non-stationary system (e.g., new water and energy pathways due to a changing climate), were all identified as research priorities. Assessing and ensuring robustness of trained ML models to adversarial data is also a research priority due to security and societal implications.

c. Beyond detection of extremes

ML for feature detection has been successful (e.g., [Prabhat et al. 2021](#)) and research priorities should lie in gaining a better understanding of the characteristics of extremes (such as their inception, e.g., cyclogenesis) both within observations and ESMs. Distinguishing extremes from outliers that may be due to erroneous data also remains difficult and is of priority (e.g., anomaly detection; [Sinha et al. 2020](#)).

d. Uncertainty for teleconnection pathways

The teleconnections of ENSO and other climate modes have complex and varying pathways (e.g., [Mehmood et al. 2022](#); [Molina et al. 2022](#)), and thus a clear understanding of these pathways is important, including identification of ESM-specific signals and those that apply in observations. ML could be used to quantify uncertainty ([Haynes et al. 2023](#)), such as with large ML-based ensembles (e.g., [Weyn et al. 2021](#)), ML model hierarchy, and models with uncertainty built in (e.g., [Barnes and Barnes 2021a,b](#)).

e. Leveraging existing ML for precursor discovery

Already existing and/or trained ML models, such as artificial neural networks and linear inverse models, present opportunities to discover new signals in the Earth system and should be further used with XAI to interrogate data (e.g., [Shin et al. 2020](#); [Toms et al. 2020](#)). Causality should be further incorporated into ML workflows to help with the quantification of physical drivers and the assessment of robust sources

TABLE 6. Summary of ML-based priorities for CV&E.

| Topic | Priority task |
|--|--|
| ML-detected features and model catalogs | Catalog existing ML-derived features and models using uniform standards to enable benchmarking and identification of research gaps |
| Robustness assessment | Further assess the robustness of XAI methods, ML transferability, and sensitivity to adversarial data |
| Beyond detection of extremes | Better understand physics of detected features (e.g., genesis drivers) |
| Uncertainty for teleconnection pathways | Further distinguish ESM-specific signals from observations and quantify associated uncertainty |
| ML for precursor discovery | Apply XAI to pretrained ML models to uncover precursors; further incorporate causality into these workflows |
| Gap filling and synthetic data | Extend existing datasets; identify where new observing sensors would provide the greatest value |
| Improving subgrid scales and ESM development | Uncover manifolds between coarse and subgrid scales to better capture extremes; use ML for ESM development |
| Benchmarking metrics | Develop ML-based metrics that better capture extremes or nontraditional quantities (e.g., scalability or trust) |
| Data compression and online learning | Leverage ML to compress data with minimal loss and for variable reconstruction, or use online learning |

of predictability (e.g., [Kretschmer et al. 2021](#)), while also considering their spatiotemporal nonstationarity.

f. Gap filling and synthetic data creation

ML also presents 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 other observational technologies were unavailable (e.g., [Dasgupta et al. 2020](#)). ML can also be used to identify regions that need more observations (e.g., [Bessenbacher et al. 2023](#)), such as with techniques borrowed from adaptive sampling ([Fer et al. 2018](#)) and active learning ([Ren et al. 2021](#)).

g. Improving subgrid-scale representation and ESM development

Climate models cannot resolve or accurately represent processes associated with extremes given their relatively coarse grid spacing (~100 km) and thus better understanding connections to extremes using, for example, ML-based downscaling, inverse modeling, or symbolic relationships, are additional research priorities. Incorporating ML during model development in order to diagnose errors, rather than applying it after the fact, was identified as a research priority, along with implementing ML to identify numerical model deficiencies (e.g., Bayesian calibration; [Cleary et al. 2021](#)).

h. Benchmarking metrics for CV&E

Metrics are scalar measures of skill that are used in model development to benchmark performance. Various existing metrics do not describe climate extremes well, such as generalized extreme value theory (e.g., [Majumdar et al. 2020](#); [Tabari 2021](#)), partly due to heavily skewed distributions, and thus ML could be potentially used to expand distributions or create more useful metrics that can better capture process

complexities and nonlinearities. Development of metrics that measure more than ML model performance in prediction tasks, such as scalability across hardware and software systems, in addition to trustworthiness ([McGovern et al. 2022](#)), were also identified as priorities.

i. Data compression tools and online learning techniques

Data compression tools present an opportunity to store large data, enabling subsequent training of data-hungry ML models ([Pinard et al. 2020](#)). ML methods could also 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 with variable reconstruction. Online learning techniques, such as during a numerical model simulation ([Partee et al. 2022](#)), also present opportunities for observation–model integration and reducing the need for large model outputs.

4. Conclusions

The major motivation driving the use of ML for climate variability and weather extremes is related to the ability of ML to learn nonlinear, complex relationships in large datasets and across multiple variables, which has enabled improved predictive skill across time scales (e.g., [Gagne et al. 2019](#); [Ravuri et al. 2021](#)) and process understanding (e.g., [Toms et al. 2020](#); [Kretschmer et al. 2021](#)). Recent applications of ML for climate variability and extremes include uncovering sources of predictability for modes of climate variability (e.g., [Toms et al. 2020](#); [Mayer and Barnes 2021](#)), feature detection (e.g., [Prabhat et al. 2021](#); [Dagon et al. 2022](#)), identifying extreme weather precursors and improving their prediction (e.g., [Gagne et al. 2019](#); [Huang et al. 2021](#)), observation–model integration (e.g., [Dasgupta et al. 2020](#); [Kadow et al. 2020](#)), and downscaling and bias correction (e.g., [Sha et al. 2020b](#); [Kim et al. 2021](#)). However, numerous algorithmic and statistical challenges must be addressed and overcome to robustly and reliably answer pressing questions with ML within these subject areas. Challenges include nonstationarity within the climate system (e.g., [Beucler et al. 2020](#);

Molina et al. 2021), uncertainty quantification associated with ML explainability (e.g., Abdar et al. 2021; Mamalakis et al. 2022), and limited sampling of extremes (e.g., Sillmann et al. 2017), among many others discussed throughout this article.

A long-term vision (e.g., during the next 10 years) for climate variability and weather extremes includes the further advancement and application of ML that is unaware of human-imposed labels (e.g., unsupervised learning; Celebi and Aydin 2016), which allows for the relaxation of a priori criteria and can potentially lead to the discovery of new modes of climate variability, climate signals, and sources of predictability. In a changing climate, extracting coherent structures from model output will require robust strategies that can separate long-term forcing signals from short-term variability. Efficient detection and attribution of the compound and cascading extremes, robust to a changing climate, and the creation/emulation of factual and counterfactual scenarios that incorporate uncertainty for climate change attribution (e.g., Easterling et al. 2016) will become increasingly important, particularly as applications extend into socioeconomic impacts (Perkins-Kirkpatrick et al. 2022). Overcoming observational limitations could be achieved using transfer operators with climate models to map a history of observations to predictions, potentially helping to generate ultrahigh-resolution simulations (e.g., 1 km or less) of extreme phenomena (tornadoes, lightning, etc.) at some future time. Critically, continuing to build trust between researchers, operational scientists, and the public is also of utmost importance, by increasing ML explainability and enabling robust identification of reasons for prediction failures (McGovern et al. 2022).

Beyond climate variability and weather extremes, collaboration and codevelopment by computer scientists, domain experts, and software engineers is important to make visionary changes and push scientific boundaries. Training and development of the future and the current workforce, 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, were all seen as imperative. Focus on stakeholder and end-user engagement is also emphasized for ML and predictive analytics. High-risk/high-reward research was also identified as critical, with needed support by funding agencies and the broader scientific community to target creativity and risk-taking in research to create effective and transformative change.

Acknowledgments. We acknowledge the following for their contributions as active session participants: Steven Klein (LLNL), Scott Collis (ANL), Alex Cannon (Environment and Climate Change Canada), Nicola Maher (CIRES/CU Boulder), Da Yang (UC Davis, LBNL), John Allen (Central Michigan University), Manos Anagnostou (University of Connecticut), Jim Ang (PNNL), Marian Anghel (LANL), Kenneth Ball (Geometric Data Analytics), Karthik Balaguru (PNNL), Antara Banerjee (CIRES, NOAA/CSL), Elizabeth Barnes (Colorado State University), Carolyn Begeman (LANL), Emily Bercos-Hickey (LBNL), Celine Bonfils (LLNL), Andrew Bradley (Sandia National Laboratories), Antonietta Capotondi (CIRES, NOAA/PSL), William Chapman (Scripps Institution of Oceanography),

Nan Chen (University of Wisconsin-Madison), Paul J. Durack (LLNL), Zhe Feng (PNNL), Andrew Geiss (PNNL), Carlo Grazian (ANL), Gary Geernaert (DOE), Huanping Huang (LBNL), Whitney Huang (Clemson University), Brian Hunt (University of Maryland), Robert Jacob (ANL), Renu Joseph (DOE), Karthik Kashinath (Nvidia), Grace E Kim (Booz Allen Hamilton), Gu Lianhong (ORNL), Jian Lu (PNNL), Valerio Lucarini (University of Reading), Hsi-Yen Ma (LLNL), Kirsten Mayer (Colorado State University), Amy McGovern (University of Oklahoma), Gerald Meehl (NCAR), Balu Nadiga (LANL), Christina Patricola (Iowa State University), Steve Penny (CIRES, NOAA/PSL), Stephen Po-Chedley (LLNL), Philip Rasch (PNNL), Deeksha Rastogi (ORNL), Adam Rupe (LANL), Christine Shields (NCAR), Jeff Stehr (DOE), Aneesh Subramanian (CU Boulder), Bob Vallario (DOE), Charuleka Varadharajan (LBNL), Hailong Wang (PNNL), Jiali Wang (ANL), Michael Wehner (LBNL), Brian White (UNC Chapel Hill, UC Berkeley), and Mark Zelinka (LLNL). This material is based upon work supported by the U.S. Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research (BER), Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program under Award DE-SC0022070 and National Science Foundation IA 1947282. The National Center for Atmospheric Research is sponsored by the National Science Foundation. A portion of this work was performed under the auspices of the U.S. DOE by Lawrence Livermore National Laboratory (LLNL) under Contract DE-AC52-07NA27344. Author Anderson was supported by LLNL Laboratory Directed Research and Development project 22-SI-008. This paper has been authored by UT-Battelle, LLC, under Contract DE-AC05-00OR22725 with the U.S. Department of Energy. The publisher, by accepting the article for publication, acknowledges that the U.S. government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of the paper, or allow others to do so, for U.S. government purposes. The DOE will provide public access to these results in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

Data availability statement. No datasets were generated or analyzed during the current study.

REFERENCES

- Abdar, M., and Coauthors, 2021: A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Inf. Fusion*, **76**, 243–297, <https://doi.org/10.1016/j.inffus.2021.05.008>.
- Allen, J. T., and M. K. Tippett, 2015: The characteristics of United States hail reports: 1955–2014. *Electron. J. Severe Storms Meteor.*, **10** (3), <https://doi.org/10.55599/ejssm.v10i3.60>.
- American Meteorological Society, 2022: Climate variability. Glossary of Meteorology, https://glossary.ametsoc.org/wiki/Climate_variability.
- Anderson, G. J., and D. D. Lucas, 2018: Machine learning predictions of a multiresolution climate model ensemble. *Geophys. Res. Lett.*, **45**, 4273–4280, <https://doi.org/10.1029/2018GL077049>.
- Andersson, T. R., and Coauthors, 2022: Active learning with convolutional Gaussian neural processes for environmental sensor placement. arXiv, 2211.10381v1, <https://doi.org/10.48550/arXiv.2211.10381>.

- Baño-Medina, J., R. Manzanas, and J. M. Gutierrez, 2020: Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geosci. Model Dev.*, **13**, 2109–2124, <https://doi.org/10.5194/gmd-13-2109-2020>.
- Barnes, E. A., and R. J. Barnes, 2021a: Controlled abstention neural networks for identifying skillful predictions for classification problems. *J. Adv. Model. Earth Syst.*, **13**, e2021MS002573, <https://doi.org/10.1029/2021MS002573>.
- , and —, 2021b: Controlled abstention neural networks for identifying skillful predictions for regression problems. *J. Adv. Model. Earth Syst.*, **13**, e2021MS002575, <https://doi.org/10.1029/2021MS002575>.
- , J. W. Hurrell, I. Ebert-Uphoff, C. Anderson, and D. Anderson, 2019: Viewing forced climate patterns through an AI lens. *Geophys. Res. Lett.*, **46**, 13 389–13 398, <https://doi.org/10.1029/2019GL084944>.
- Bennett, K. E., and Coauthors, 2022: Spatial patterns of snow distribution in the sub-Arctic. *Cryosphere*, **16**, 3269–3293, <https://doi.org/10.5194/tc-16-3269-2022>.
- Bessenbacher, V., L. Gudmundsson, and S. I. Seneviratne, 2023: Optimizing soil moisture station networks for future climates. *Geophys. Res. Lett.*, **50**, e2022GL101667, <https://doi.org/10.1029/2022GL101667>.
- Beucler, T., M. Pritchard, P. Gentine, and S. Rasp, 2020: Towards physically-consistent, data-driven models of convection. *2020 IEEE Int. Geoscience and Remote Sensing Symp.*, Waikoloa, HI, IEEE, 3987–3990, <https://doi.org/10.1109/IGARSS39084.2020.9324569>.
- Bi, K., L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, 2022: Pangu-weather: A 3D high-resolution model for fast and accurate global weather forecast. arXiv, 2211.02556v1, <https://doi.org/10.48550/arXiv.2211.02556>.
- Biard, J. C., and K. E. Kunkel, 2019: Automated detection of weather fronts using a deep learning neural network. *Adv. Stat. Climatol. Meteor. Oceanogr.*, **5**, 147–160, <https://doi.org/10.5194/ascmo-5-147-2019>.
- Bird, L., M. Walker, G. Bodeker, I. Campbell, G. Liu, S. J. Sam, J. Lewis, and S. Rosier, 2023: Deep learning for stochastic precipitation generation—Deep SPG v1.0. *Geosci. Model Dev.*, **16**, 3785–3808, <https://doi.org/10.5194/gmd-16-3785-2023>.
- Bocquet, M., J. Brajard, A. Carrassi, and L. Bertino, 2020: Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization. arXiv, 2001.06270v2, <https://doi.org/10.48550/arXiv.2001.06270>.
- Brajard, J., A. Carrassi, M. Bocquet, and L. Bertino, 2021: Combining data assimilation and machine learning to infer unresolved scale parametrization. *Philos. Trans. Roy. Soc.*, **A379**, 20200086, <https://doi.org/10.1098/rsta.2020.0086>.
- Bretherton, C. S., and Coauthors, 2022: Correcting coarse-grid weather and climate models by machine learning from global storm-resolving simulations. *J. Adv. Model. Earth Syst.*, **14**, e2021MS002794, <https://doi.org/10.1029/2021MS002794>.
- Brodu, N., and J. P. Crutchfield, 2022: Discovering causal structure with reproducing-kernel Hilbert space ϵ -machines. *Chaos*, **32**, 023103, <https://doi.org/10.1063/5.0062829>.
- Buizza, C., and Coauthors, 2022: Data learning: Integrating data assimilation and machine learning. *J. Comput. Sci.*, **58**, 101525, <https://doi.org/10.1016/j.jocs.2021.101525>.
- Casey, S. P. F., and L. Cucurull, 2022: The impact of data latency on operational global weather forecasting. *Wea. Forecasting*, **37**, 1211–1220, <https://doi.org/10.1175/WAF-D-21-0149.1>.
- Celebi, M. E., and K. Aydin, 2016: *Unsupervised Learning Algorithms*. Springer, 558 pp.
- Chantry, M., H. Christensen, P. Dueben, and T. Palmer, 2021: Opportunities and challenges for machine learning in weather and climate modelling: Hard, medium and soft AI. *Philos. Trans. Roy. Soc.*, **A379**, 20200083, <https://doi.org/10.1098/rsta.2020.0083>.
- Chattopadhyay, A., E. Nabizadeh, and P. Hassanzadeh, 2020: Analog forecasting of extreme-causing weather patterns using deep learning. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001958, <https://doi.org/10.1029/2019MS001958>.
- Clark, A. J., and E. D. Loken, 2022: Machine learning-derived severe weather probabilities from a warn-on-forecast system. *Wea. Forecasting*, **37**, 1721–1740, <https://doi.org/10.1175/WAF-D-22-0056.1>.
- Clark, S. K., and Coauthors, 2022: Correcting a 200 km resolution climate model in multiple climates by machine learning from 25 km resolution simulations. *J. Adv. Model. Earth Syst.*, **14**, e2022MS003219, <https://doi.org/10.1029/2022MS003219>.
- Cleary, E., A. Garbuno-Ingó, S. Lan, T. Schneider, and A. M. Stuart, 2021: Calibrate, emulate, sample. *J. Comput. Phys.*, **424**, 109716, <https://doi.org/10.1016/j.jcp.2020.109716>.
- 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. *Adv. Stat. Climatol. Meteor. Oceanogr.*, **6**, 223–244, <https://doi.org/10.5194/ascmo-6-223-2020>.
- , J. Truesdale, J. C. Biard, K. E. Kunkel, G. A. Meehl, and M. J. Molina, 2022: Machine learning-based detection of weather fronts and associated extreme precipitation in historical and future climates. *J. Geophys. Res. Atmos.*, **127**, e2022JD037038, <https://doi.org/10.1029/2022JD037038>.
- Dasgupta, P., A. Metya, C. V. Naidu, M. Singh, and M. K. Roxy, 2020: Exploring the long-term changes in the Madden Julian Oscillation using machine learning. *Sci. Rep.*, **10**, 18567, <https://doi.org/10.1038/s41598-020-75508-5>.
- Davenport, F. V., and N. S. Diffenbaugh, 2021: Using machine learning to analyze physical causes of climate change: A case study of U.S. Midwest extreme precipitation. *Geophys. Res. Lett.*, **48**, e2021GL093787, <https://doi.org/10.1029/2021GL093787>.
- de Wolff, T., H. Carrillo, L. Martí, and N. Sanchez-Pi, 2021: Assessing physics informed neural networks in ocean modelling and climate change applications. *AI: Modeling Oceans and Climate Change Workshop*, Santiago, Chile, ICLR, <https://oceania.inria.cl/assets/pdfs/aimocc/paper-13.pdf>.
- Dueben, P. D., M. G. Schultz, M. Chantry, D. J. Gagne II, D. M. Hall, and A. McGovern, 2022: Challenges and benchmark datasets for machine learning in the atmospheric sciences: Definition, status, and outlook. *Artif. Intell. Earth Syst.*, **1**, e21002, <https://doi.org/10.1175/AIES-D-21-0002.1>.
- DuVivier, A. K., M. J. Molina, M. M. Holland, L. Landrum, K. Krumhardt, and S. Jenouvrier, 2023: Projections of winter polynyas and their biophysical impacts in the Ross Sea Antarctica. Research Square, <https://doi.org/10.21203/rs.3.rs-2575015/v1>.
- Easterling, D. R., K. E. Kunkel, M. F. Wehner, and L. Sun, 2016: Detection and attribution of climate extremes in the observed record. *Wea. Climate Extremes*, **11**, 17–27, <https://doi.org/10.1016/j.wace.2016.01.001>.
- Ebert-Uphoff, I., R. Lagerquist, K. Hilburn, Y. Lee, K. Haynes, J. Stock, C. Kumler, and J. Q. Stewart, 2021: CIRA guide to custom loss functions for neural networks in environmental sciences—Version 1. arXiv, 2106.09757v1, <https://doi.org/10.48550/arXiv.2106.09757>.
- Ebi, K. L., and Coauthors, 2021: Extreme weather and climate change: Population health and health system implications.

- Annu. Rev. Public Health*, **42**, 293–315, <https://doi.org/10.1146/annurev-publhealth-012420-105026>.
- Fayaz, S. A., M. Zaman, and M. A. Butt, 2022: Knowledge discovery in geographical sciences—A systematic survey of various machine learning algorithms for rainfall prediction. *International Conference on Innovative Computing and Communications*, A. Khanna et al., Eds., Advances in Intelligent Systems and Computing, Vol. 1388, 593–608, Springer, https://doi.org/10.1007/978-981-16-2597-8_51.
- Feng, Z., F. Song, K. Sakaguchi, and L. R. Leung, 2021: Evaluation of mesoscale convective systems in climate simulations: Methodological development and results from MPAS-CAM over the United States. *J. Climate*, **34**, 2611–2633, <https://doi.org/10.1175/JCLI-D-20-0136.1>.
- Fer, I., R. Kelly, P. R. Moorcroft, A. D. Richardson, E. M. Cowdery, and M. C. Dietze, 2018: Linking big models to big data: Efficient ecosystem model calibration through Bayesian model emulation. *Biogeosciences*, **15**, 5801–5830, <https://doi.org/10.5194/bg-15-5801-2018>.
- Flora, M. L., C. K. Potvin, P. S. Skinner, S. Handler, and A. McGovern, 2021: Using machine learning to generate storm-scale probabilistic guidance of severe weather hazards in the warn-on-forecast system. *Mon. Wea. Rev.*, **149**, 1535–1557, <https://doi.org/10.1175/MWR-D-20-0194.1>.
- Foster, D., and J. M. Restrepo, 2022: An improved framework for the dynamic likelihood filtering approach to data assimilation. *Chaos*, **32**, 053118, <https://doi.org/10.1063/5.0083071>.
- Frame, J. M., and Coauthors, 2022: Deep learning rainfall-runoff predictions of extreme events. *Hydrol. Earth Syst. Sci.*, **26**, 3377–3392, <https://doi.org/10.5194/hess-26-3377-2022>.
- Fulton, D. J., B. J. Clarke, and G. C. Hegerl, 2023: Bias correcting climate model simulations using unpaired image-to-image translation networks. *Artif. Intell. Earth Syst.*, **2**, e220031, <https://doi.org/10.1175/AIES-D-22-0031.1>.
- Gagne, D. J., II, S. E. Haupt, D. W. Nychka, and G. Thompson, 2019: Interpretable deep learning for spatial analysis of severe hailstorms. *Mon. Wea. Rev.*, **147**, 2827–2845, <https://doi.org/10.1175/MWR-D-18-0316.1>.
- Geen, R., S. Bordoni, D. S. Battisti, and K. Hui, 2020: Monsoons, ITCZs, and the concept of the global monsoon. *Rev. Geophys.*, **58**, e2020RG000700, <https://doi.org/10.1029/2020RG000700>.
- Geer, A. J., 2021: Learning Earth system models from observations: Machine learning or data assimilation? *Philos. Trans. Roy. Soc.*, **A379**, 20200089, <https://doi.org/10.1098/rsta.2020.0089>.
- 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. *J. Adv. Model. Earth Syst.*, **13**, e2020MS002268, <https://doi.org/10.1029/2020MS002268>.
- Ghil, M., 2002: Natural climate variability. *Encyclopedia of Global Environmental Change*, M. MacCracken and J. Perry, Eds., Wiley and Sons, 544–549.
- Gibson, P. B., S. E. Perkins-Kirkpatrick, P. Uotila, A. S. Pepler, and L. V. Alexander, 2017: On the use of self-organizing maps for studying climate extremes. *J. Geophys. Res. Atmos.*, **122**, 3891–3903, <https://doi.org/10.1002/2016JD026256>.
- Gordon, E. M., and E. A. Barnes, 2022: Incorporating uncertainty into a regression neural network enables identification of decadal state-dependent predictability in CESM2. *Geophys. Res. Lett.*, **49**, e2022GL098635, <https://doi.org/10.1029/2022GL098635>.
- Groenke, B., L. Madaus, and C. Monteleoni, 2020: ClimAlign: Unsupervised statistical downscaling of climate variables via normalizing flows. *Proc. 10th Int. Conf. on Climate Informatics*, Online, Association for Computing Machinery, 60–66, <https://doi.org/10.1145/3429309.3429318>.
- Ham, Y. G., J. H. Kim, and J. J. Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature*, **573**, 568–572, <https://doi.org/10.1038/s41586-019-1559-7>.
- Harder, P., Q. Yang, V. Ramesh, P. Sattigeri, A. Hernandez-Garcia, C. Watson, D. Szwarcman, and D. Rolnick, 2022: Generating physically-consistent high-resolution climate data with hard-constrained neural networks. arXiv, 2208.05424v1, <https://doi.org/10.48550/arXiv.2208.05424>.
- Haupt, S. E., and Coauthors, 2022: The history and practice of AI in the environmental sciences. *Bull. Amer. Meteor. Soc.*, **103**, E1351–E1370, <https://doi.org/10.1175/BAMS-D-20-0234.1>.
- Haynes, K., R. Lagerquist, M. McGraw, K. Musgrave, and I. Ebert-Uphoff, 2023: Creating and evaluating uncertainty estimates with neural networks for environmental-science applications. *Artif. Intell. Earth Syst.*, **2**, 220061, <https://doi.org/10.1175/AIES-D-22-0061.1>.
- Herman, G. R., and R. S. Schumacher, 2018: Money doesn't grow on trees, but forecasts do: Forecasting extreme precipitation with random forests. *Mon. Wea. Rev.*, **146**, 1571–1600, <https://doi.org/10.1175/MWR-D-17-0250.1>.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Hickmon, N. L., C. Varadharajan, F. M. Hoffman, S. Collis, and H. M. Wainwright, 2022: Artificial Intelligence for Earth System Predictability (AI4ESP) workshop report. Argonne National Laboratory Tech. Rep. ANL-22/54, 413 pp., <https://doi.org/10.2172/1888810>.
- Hill, A. J., and R. S. Schumacher, 2021: Forecasting excessive rainfall with random forests and a deterministic convection-allowing model. *Wea. Forecasting*, **36**, 1693–1711, <https://doi.org/10.1175/WAF-D-21-0026.1>.
- , G. R. Herman, and R. S. Schumacher, 2020: Forecasting severe weather with random forests. *Mon. Wea. Rev.*, **148**, 2135–2161, <https://doi.org/10.1175/MWR-D-19-0344.1>.
- Hoegh, A., 2020: Why Bayesian ideas should be introduced in the statistics curricula and how to do so. *J. Stat. Educ.*, **28**, 222–228, <https://doi.org/10.1080/10691898.2020.1841591>.
- Hu, P., Q. Zhang, P. Shi, B. Chen, and J. Fang, 2018: Flood-induced mortality across the globe: Spatiotemporal pattern and influencing factors. *Sci. Total Environ.*, **643**, 171–182, <https://doi.org/10.1016/j.scitotenv.2018.06.197>.
- Huang, H., C. M. Patricola, E. Bercos-Hickey, Y. Zhou, A. Rhoades, M. D. Risser, and W. D. Collins, 2021: Sources of subseasonal-to-seasonal predictability of atmospheric rivers and precipitation in the western United States. *J. Geophys. Res. Atmos.*, **126**, e2020JD034053, <https://doi.org/10.1029/2020JD034053>.
- Jiang, C., and Coauthors, 2020: MESHFREEFLOWNET: A physics-constrained deep continuous space-time super-resolution framework. *Int. Conf. for High Performance Computing, Networking, Storage and Analysis*, Atlanta, GA, IEEE, 1–15, <https://doi.org/10.1109/SC41405.2020.00013>.
- Kadow, C., D. M. Hall, and U. Ulbrich, 2020: Artificial intelligence reconstructs missing climate information. *Nat. Geosci.*, **13**, 408–413, <https://doi.org/10.1038/s41561-020-0582-5>.
- Kamangir, H., E. Krell, W. Collins, S. A. King, and P. Tissot, 2022: Importance of 3D convolution and physics on a deep learning coastal fog model. *Environ. Model. Software*, **154**, 105424, <https://doi.org/10.1016/j.envsoft.2022.105424>.
- Kashinath, K., and Coauthors, 2021: Physics-informed machine learning: Case studies for weather and climate modelling.

- Philos. Trans. Roy. Soc.*, **A379**, 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. *Nat. Commun.*, **12**, 3087, <https://doi.org/10.1038/s41467-021-23406-3>.
- Kim, H.-M., Y. Zhou, and M. A. Alexander, 2019: Changes in atmospheric rivers and moisture transport over the northeast Pacific and western North America in response to ENSO diversity. *Climate Dyn.*, **52**, 7375–7388, <https://doi.org/10.1007/s00382-017-3598-9>.
- Klaver, R., R. Haarsma, P. L. Vidale, and W. Hazeleger, 2020: Effective resolution in high resolution global atmospheric models for climate studies. *Atmos. Sci. Lett.*, **21**, e952, <https://doi.org/10.1002/asl.952>.
- Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS) unifying tropical cyclone data. *Bull. Amer. Meteor. Soc.*, **91**, 363–376, <https://doi.org/10.1175/2009BAMS2755.1>.
- Kretschmer, M., S. V. Adams, A. Arribas, R. Prudden, N. Robinson, E. Saggioro, and T. G. Shepherd, 2021: Quantifying causal pathways of teleconnections. *Bull. Amer. Meteor. Soc.*, **102**, E2247–E2263, <https://doi.org/10.1175/BAMS-D-20-0117.1>.
- Krishnapriyan, A., A. Gholami, S. Zhe, R. Kirby, and M. W. Mahoney, 2021: Characterizing possible failure modes in physics-informed neural networks. *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, M. Ranzato et al., Eds., Curran Associates, Inc., 26548–26560, <https://proceedings.neurips.cc/paper/2021/hash/df438e5206f31600e6ae4af72f2725f1-Abstract.html>.
- Kumar, P., and M. Sharma, 2022: Data, machine learning, and human domain experts: None is better than their collaboration. *Int. J. Hum. Comput. Interact.*, **38**, 1307–1320, <https://doi.org/10.1080/10447318.2021.2002040>.
- Lagerquist, R., 2020: Using deep learning to improve prediction and understanding of high-impact weather. Ph.D. dissertation, University of Oklahoma, 290 pp.
- , A. M. McGovern, and D. J. Gagne II, 2019: Deep learning for spatially explicit prediction of synoptic-scale fronts. *Wea. Forecasting*, **34**, 1137–1160, <https://doi.org/10.1175/WAF-D-18-0183.1>.
- , A. McGovern, C. R. Homeyer, D. J. Gagne II, and T. Smith, 2020: Deep learning on three-dimensional multiscale data for next-hour tornado prediction. *Mon. Wea. Rev.*, **148**, 2837–2861, <https://doi.org/10.1175/MWR-D-19-0372.1>.
- Lam, R., and Coauthors, 2022: GraphCast: Learning skillful medium-range global weather forecasting. arXiv, 2212.12794v1, <https://doi.org/10.48550/arXiv.2212.12794>.
- LeCun, Y., Y. Bengio, and G. Hinton, 2015: Deep learning. *Nature*, **521**, 436–444, <https://doi.org/10.1038/nature14539>.
- Ledig, C., and Coauthors, 2017: Photo-realistic single image super-resolution using a generative adversarial network. *2017 IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, IEEE, 105–114, <https://doi.org/10.1109/CVPR.2017.19>.
- Lee, Y., C. D. Kummerow, and I. Ebert-Uphoff, 2021: Applying machine learning methods to detect convection using Geostationary Operational Environmental Satellite-16 (GOES-16) Advanced Baseline Imager (ABI) data. *Atmos. Meas. Tech.*, **14**, 2699–2716, <https://doi.org/10.5194/amt-14-2699-2021>.
- Legg, S., and Coauthors, 2009: Improving oceanic overflow representation in climate models: The gravity current entrainment climate process team. *Bull. Amer. Meteor. Soc.*, **90**, 657–670, <https://doi.org/10.1175/2008BAMS2667.1>.
- Leung, L. R., and Coauthors, 2022: Exploratory precipitation metrics: Spatiotemporal characteristics, process-oriented, and phenomena-based. *J. Climate*, **35**, 3659–3686, <https://doi.org/10.1175/JCLI-D-21-0590.1>.
- Li, F., W. D. Collins, M. F. Wehner, and L. R. Leung, 2013: Hurricanes in an aquaplanet world: Implications of the impacts of external forcing and model horizontal resolution. *J. Adv. Model. Earth Syst.*, **5**, 134–145, <https://doi.org/10.1002/jame.20020>.
- Li, L., and Coauthors, 2020: A causal inference model based on random forests to identify the effect of soil moisture on precipitation. *J. Hydrometeor.*, **21**, 1115–1131, <https://doi.org/10.1175/JHM-D-19-0209.1>.
- Lin, I.-I., and Coauthors, 2020: ENSO and tropical cyclones. *El Niño Southern Oscillation in a Changing Climate*, *Geophys. Monogr.*, Vol. 253, Amer. Geophys. Union, 377–408, <https://doi.org/10.1002/9781119548164.ch17>.
- Liu, Y., and Coauthors, 2016: Application of deep convolutional neural networks for detecting extreme weather in climate datasets. arXiv, 1605.01156v1, <https://doi.org/10.48550/arXiv.1605.01156>.
- Loken, E. D., A. J. Clark, A. McGovern, M. Flora, and K. Knopfmeier, 2019: Postprocessing next-day ensemble probabilistic precipitation forecasts using random forests. *Wea. Forecasting*, **34**, 2017–2044, <https://doi.org/10.1175/WAF-D-19-0109.1>.
- , —, and C. D. Karstens, 2020: Generating probabilistic next-day severe weather forecasts from convection-allowing ensembles using random forests. *Wea. Forecasting*, **35**, 1605–1631, <https://doi.org/10.1175/WAF-D-19-0258.1>.
- , —, and A. McGovern, 2022: Comparing and interpreting differently designed random forests for next-day severe weather hazard prediction. *Wea. Forecasting*, **37**, 871–899, <https://doi.org/10.1175/WAF-D-21-0138.1>.
- Lopez-Gomez, I., A. McGovern, S. Agrawal, and J. Hickey, 2023: Global extreme heat forecasting using neural weather models. *Artif. Intell. Earth Syst.*, **2**, e220035, <https://doi.org/10.1175/AIES-D-22-0035.1>.
- Lu, D., D. Ricciuto, M. Stoyanov, and L. Gu, 2018: Calibration of the E3SM land model using surrogate-based global optimization. *J. Adv. Model. Earth Syst.*, **10**, 1337–1356, <https://doi.org/10.1002/2017MS001134>.
- Madden, R. A., and P. R. Julian, 1971: Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28**, 702–708, [https://doi.org/10.1175/1520-0469\(1971\)028<0702:DOADOI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2).
- , —, 1972: Description of global-scale circulation cells in the tropics with a 40–50 day period. *J. Atmos. Sci.*, **29**, 1109–1123, [https://doi.org/10.1175/1520-0469\(1972\)029<1109:DGSCC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1972)029<1109:DGSCC>2.0.CO;2).
- Maher, N., T. P. Tabarin, and S. Milinski, 2022: Combining machine learning and SMILEs to classify, better understand, and project changes in ENSO events. *Earth Syst. Dyn.*, **13**, 1289–1304, <https://doi.org/10.5194/esd-13-1289-2022>.
- Majumdar, S. N., A. Pal, and G. Schehr, 2020: Extreme value statistics of correlated random variables: A pedagogical review. *Phys. Rep.*, **840**, 1–32, <https://doi.org/10.1016/j.physrep.2019.10.005>.
- Mamalakis, A., E. A. Barnes, and I. Ebert-Uphoff, 2022: Investigating the fidelity of explainable artificial intelligence methods for applications of convolutional neural networks in geoscience.

- Artif. Intell. Earth Syst.*, **1**, e220012, <https://doi.org/10.1175/AIES-D-22-0012.1>.
- Martin, Z. K., E. A. Barnes, and E. Maloney, 2022: Using simple, explainable neural networks to predict the Madden-Julian Oscillation. *J. Adv. Model. Earth Syst.*, **14**, e2021MS002774, <https://doi.org/10.1029/2021MS002774>.
- Mayer, K. J., and E. A. Barnes, 2021: Subseasonal forecasts of opportunity identified by an explainable neural network. *Geophys. Res. Lett.*, **48**, e2020GL092092, <https://doi.org/10.1029/2020GL092092>.
- McGovern, A., K. L. Elmore, D. J. Gagne II, S. E. Haupt, C. D. Karstens, R. Lagerquist, T. Smith, and J. K. Williams, 2017: Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bull. Amer. Meteor. Soc.*, **98**, 2073–2090, <https://doi.org/10.1175/BAMS-D-16-0123.1>.
- , R. Lagerquist, D. J. Gagne II, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith, 2019: Making the black box more transparent: Understanding the physical implications of machine learning. *Bull. Amer. Meteor. Soc.*, **100**, 2175–2199, <https://doi.org/10.1175/BAMS-D-18-0195.1>.
- , I. Ebert-Uphoff, D. J. Gagne II, and A. Bostrom, 2022: Why we need to focus on developing ethical, responsible, and trustworthy artificial intelligence approaches for environmental science. *Environ. Data Sci.*, **1**, E6, <https://doi.org/10.1017/eds.2022.5>.
- McGraw, M. C., and E. A. Barnes, 2018: Memory matters: A case for Granger causality in climate variability studies. *J. Climate*, **31**, 3289–3300, <https://doi.org/10.1175/JCLI-D-17-0334.1>.
- Meehl, G. A., and Coauthors, 2021: Initialized Earth system prediction from subseasonal to decadal timescales. *Nat. Rev. Earth Environ.*, **2**, 340–357, <https://doi.org/10.1038/s43017-021-0015-x>.
- Mehmood, S., and Coauthors, 2022: Dominant controls of cold-season precipitation variability over the high mountains of Asia. *npj Climate Atmos. Sci.*, **5**, 65, <https://doi.org/10.1038/s41612-022-00282-2>.
- Molina, M. J., J. T. Allen, and V. A. Gensini, 2018: The Gulf of Mexico and ENSO influence on subseasonal and seasonal CONUS winter tornado variability. *J. Appl. Meteor. Climatol.*, **57**, 2439–2463, <https://doi.org/10.1175/JAMC-D-18-0046.1>.
- , D. J. Gagne, and A. F. Prein, 2021: A benchmark to test generalization capabilities of deep learning methods to classify severe convective storms in a changing climate. *Earth Space Sci.*, **8**, e2020EA001490, <https://doi.org/10.1029/2020EA001490>.
- , A. Hu, and G. A. Meehl, 2022: Response of global SSTs and ENSO to the Atlantic and Pacific meridional overturning circulations. *J. Climate*, **35**, 49–72, <https://doi.org/10.1175/JCLI-D-21-0172.1>.
- , J. H. Richter, A. A. Glanville, K. Dagon, J. Berner, A. Hu, and G. A. Meehl, 2023: Subseasonal representation and predictability of North American weather regimes using cluster analysis. *Artif. Intell. Earth Syst.*, **2**, e220051, <https://doi.org/10.1175/AIES-D-22-0051.1>.
- Molnar, C., G. Casalicchio, and B. Bischl, 2021: Interpretable machine learning—A brief history, state-of-the-art and challenges. *ECML PKDD 2020 Workshops*, I. Koprinska et al., Eds., Communications in Computer and Information Science, Vol. 1323, Springer, 417–431, https://doi.org/10.1007/978-3-030-65965-3_28.
- Mostajabi, A., D. L. Finney, M. Rubinstein, and F. Rachidi, 2019: Nowcasting lightning occurrence from commonly available meteorological parameters using machine learning techniques. *npj Climate Atmos. Sci.*, **2**, 41, <https://doi.org/10.1038/s41612-019-0098-0>.
- Mudigonda, M., and Coauthors, 2017: Segmenting and tracking extreme climate events using neural networks. *31st Conf. on Neural Information Processing Systems*, Long Beach, CA, Association for Computing Machinery, https://dl4physicalsciences.github.io/files/nips_dlps_2017_20.pdf.
- Mukherjee, S., M. Ashfaq, and A. K. Mishra, 2020: Compound drought and heatwaves at a global scale: The role of natural climate variability-associated synoptic patterns and land-surface energy budget anomalies. *J. Geophys. Res. Atmos.*, **125**, e2019JD031943, <https://doi.org/10.1029/2019JD031943>.
- Nachman, B., 2022: Anomaly detection for physics analysis and less than supervised learning. *Artificial Intelligence for High Energy Physics*, P. Calafuria, D. Rousseau, and K. Terao, Eds., World Scientific, 85–112.
- Nguyen, T., J. Brandstetter, A. Kapoor, J. K. Gupta, and A. Grover, 2023: ClimaX: A foundation model for weather and climate. arXiv, 2301.10343v2, <https://doi.org/10.48550/arXiv.2301.10343>.
- Niebler, S., A. Miltenberger, B. Schmidt, and P. Spichtinger, 2022: Automated detection and classification of synoptic-scale fronts from atmospheric data grids. *Wea. Climate Dyn.*, **3**, 113–137, <https://doi.org/10.5194/wcd-3-113-2022>.
- Nogueira, A. R., A. Pugnana, S. Ruggieri, D. Pedreschi, and J. Gama, 2022: Methods and tools for causal discovery and causal inference. *Wiley Interdiscip. Rev.: Data Mining Knowl. Discovery*, **12**, e1449, <https://doi.org/10.1002/widm.1449>.
- Núñez Ocasio, K. M., A. Brammer, J. L. Evans, G. S. Young, and Z. L. Moon, 2021: Favorable monsoon environment over eastern Africa for subsequent tropical cyclogenesis of African easterly waves. *J. Atmos. Sci.*, **78**, 2911–2925, <https://doi.org/10.1175/JAS-D-20-0339.1>.
- O'Brien, T. A., and Coauthors, 2020: Detection of atmospheric rivers with inline uncertainty quantification: TECA-BARD v1.0.1. *Geosci. Model Dev.*, **13**, 6131–6148, <https://doi.org/10.5194/gmd-13-6131-2020>.
- , and Coauthors, 2022: Increases in future AR count and size: Overview of the ARTMIP Tier 2 CMIP5/6 experiment. *J. Geophys. Res. Atmos.*, **127**, e2021JD036013, <https://doi.org/10.1029/2021JD036013>.
- Pan, B., G. J. Anderson, A. Goncalves, D. D. Lucas, C. J. Bonfils, J. Lee, Y. Tian, and H.-Y. Ma, 2021: Learning to correct climate projection biases. *J. Adv. Model. Earth Syst.*, **13**, e2021MS002509, <https://doi.org/10.1029/2021MS002509>.
- , —, —, —, —, and —, 2022: Improving seasonal forecast using probabilistic deep learning. *J. Adv. Model. Earth Syst.*, **14**, e2021MS002766, <https://doi.org/10.1029/2021MS002766>.
- Partee, S., M. Ellis, A. Rigazzi, A. E. Shao, S. Bachman, G. Marques, and B. Robbins, 2022: Using machine learning at scale in numerical simulations with SmartSim: An application to ocean climate modeling. *J. Comput. Sci.*, **62**, 101707, <https://doi.org/10.1016/j.jocs.2022.101707>.
- Pathak, J., and Coauthors, 2022: FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. arXiv, 2202.11214v1, <https://doi.org/10.48550/arXiv.2202.11214>.
- Pegion, K., E. J. Becker, and B. P. Kirtman, 2022: Understanding predictability of daily southeast U.S. precipitation using explainable machine learning. *Artif. Intell. Earth Syst.*, **1**, e220011, <https://doi.org/10.1175/AIES-D-22-0011.1>.

- Pendergrass, A. G., 2018: What precipitation is extreme? *Science*, **360**, 1072–1073, <https://doi.org/10.1126/science.aat1871>.
- Perkins-Kirkpatrick, S. E., and Coauthors, 2022: On the attribution of the impacts of extreme weather events to anthropogenic climate change. *Environ. Res. Lett.*, **17**, 024009, <https://doi.org/10.1088/1748-9326/ac44c8>.
- Pinard, A., D. M. Hammerling, and A. H. Baker, 2020: Assessing differences in large spatio-temporal climate datasets with a new python package. *2020 IEEE Int. Conf. on Big Data*, Atlanta, GA, IEEE, 2699–2707, <https://doi.org/10.1109/BigData50022.2020.9378100>.
- Pohl, B., and Coauthors, 2021: Relationship between weather regimes and atmospheric rivers in East Antarctica. *J. Geophys. Res. Atmos.*, **126**, e2021JD035294, <https://doi.org/10.1029/2021JD035294>.
- Prabhat, and Coauthors, 2021: ClimateNet: An expert-labeled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather. *Geosci. Model Dev.*, **14**, 107–124, <https://doi.org/10.5194/gmd-14-107-2021>.
- Quinting, J. F., and C. M. Grams, 2022: EuLerian identification of ascending AirStreams (ELIAS 2.0) in numerical weather prediction and climate models—Part 1: Development of deep learning model. *Geosci. Model Dev.*, **15**, 715–730, <https://doi.org/10.5194/gmd-15-715-2022>.
- , —, A. Oertel, and M. Pickl, 2022: EuLerian Identification of ascending AirStreams (ELIAS 2.0) in numerical weather prediction and climate models—Part 2: Model application to different datasets. *Geosci. Model Dev.*, **15**, 731–744, <https://doi.org/10.5194/gmd-15-731-2022>.
- Radić, V., A. J. Cannon, B. Menounos, and N. Gi, 2015: Future changes in autumn atmospheric river events in British Columbia, Canada, as projected by CMIP5 global climate models. *J. Geophys. Res. Atmos.*, **120**, 9279–9302, <https://doi.org/10.1002/2015JD023279>.
- Ralph, F. M., M. D. Dettinger, M. M. Cairns, T. J. Galarneau, and J. Eylander, 2018: Defining “atmospheric river”: How the glossary of meteorology helped resolve a debate. *Bull. Amer. Meteor. Soc.*, **99**, 837–839, <https://doi.org/10.1175/BAMS-D-17-0157.1>.
- Rasmusson, E. M., and T. H. Carpenter, 1982: Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. *Mon. Wea. Rev.*, **110**, 354–384, [https://doi.org/10.1175/1520-0493\(1982\)110<354:VITSST>2.0.CO;2](https://doi.org/10.1175/1520-0493(1982)110<354:VITSST>2.0.CO;2).
- Ravuri, S., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, and N. Carvalhais, 2021: Skillful precipitation nowcasting using deep generative models of radar. *Nature*, **597**, 672–677, <https://doi.org/10.1038/s41586-021-03854-z>.
- Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, and N. Carvalhais, 2019: Deep learning and process understanding for data-driven Earth system science. *Nature*, **566**, 195–204, <https://doi.org/10.1038/s41586-019-0912-1>.
- Ren, P., Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, B. B. Gupta, X. Chen, and X. Wang, 2021: A survey of deep active learning. *ACM Comput. Surv.*, **54** (9), 1–40, <https://doi.org/10.1145/3472291>.
- Restrepo, J. M., 2017: A dynamic likelihood approach to filtering. *Quart. J. Roy. Meteor. Soc.*, **143**, 2915–2924, <https://doi.org/10.1002/qj.3143>.
- Ronneberger, O., P. Fischer, and T. Brox, 2015: U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, N. Navab et al, Lecture Notes in Computer Science, Vol. 9351, Springer, 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.
- Ropelewski, C. F., and M. S. Halpert, 1986: North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **114**, 2352–2362, [https://doi.org/10.1175/1520-0493\(1986\)114<2352:NAPATP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1986)114<2352:NAPATP>2.0.CO;2).
- Runge, J., P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, 2019: Detecting and quantifying causal associations in large non-linear time series datasets. *Sci. Adv.*, **5**, eaau4996, <https://doi.org/10.1126/sciadv.aau4996>.
- Rupe, A., and J. P. Crutchfield, 2018: Local causal states and discrete coherent structures. *Chaos*, **28**, 075312, <https://doi.org/10.1063/1.5021130>.
- , K. Kashinath, N. Kumar, V. Lee, Prabhat, and J. P. Crutchfield, 2019a: Towards unsupervised segmentation of extreme weather events. arXiv, 1909.07520v1, <https://doi.org/10.48550/arXiv.1909.07520>.
- , and Coauthors, 2019b: DisCo: Physics-based unsupervised discovery of coherent structures in spatiotemporal systems. *2019 IEEE/ACM Workshop on Machine Learning in High Performance Computing Environments*, Denver, CO, IEEE, 75–87, <https://doi.org/10.1109/MLHPC49564.2019.00013>.
- Salinger, M. J., 2005: Climate variability and change: Past, present and future—An overview. *Increasing Climate Variability and Change*, J. Salinger, M. Sivakumar, and R. P. Motha, Eds., Springer, 9–29.
- Sanford, C., A. Kwa, O. Watt-Meyer, S. Clark, N. Brenowitz, J. McGibbon, and C. Bretherton, 2022: Improving the predictions of ML-corrected climate models with novelty detection. arXiv, 2211.13354v1, <https://doi.org/10.48550/arXiv.2211.13354>.
- Schevenhoven, F., and A. Carrassi, 2022: Training a supermodel with noisy and sparse observations: A case study with CPT and the synch rule on SPEEDO—v.1. *Geosci. Model Dev.*, **15**, 3831–3844, <https://doi.org/10.5194/gmd-15-3831-2022>.
- Schmit, T. J., P. Griffith, M. M. Gunshor, J. M. Daniels, S. J. Goodman, and W. J. Lebar, 2017: A closer look at the ABI on the GOES-R series. *Bull. Amer. Meteor. Soc.*, **98**, 681–698, <https://doi.org/10.1175/BAMS-D-15-00230.1>.
- Seneviratne, S., and Coauthors, 2021: Weather and climate extreme events in a changing climate. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al, Eds., Cambridge University Press, 1513–1766, https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter11.pdf.
- Serifi, A., T. Günther, and N. Ban, 2021: Spatio-temporal downscaling of climate data using convolutional and error-predicting neural networks. *Front. Climate*, **3**, 656479, <https://doi.org/10.3389/fclim.2021.656479>.
- Sha, Y., D. J. Gagne II, G. West, and R. Stull, 2020a: Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. Part I: Daily maximum and minimum 2-m temperature. *J. Appl. Meteor. Climatol.*, **59**, 2057–2073, <https://doi.org/10.1175/JAMC-D-20-0057.1>.
- , —, —, and —, 2020b: Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. Part II: Daily precipitation. *J. Appl. Meteor. Climatol.*, **59**, 2075–2092, <https://doi.org/10.1175/JAMC-D-20-0058.1>.
- , —, —, and —, 2021: Deep-learning-based precipitation observation quality control. *J. Atmos. Oceanic Technol.*, **38**, 1075–1091, <https://doi.org/10.1175/JTECH-D-20-0081.1>.
- Shields, C. A., and Coauthors, 2018: Atmospheric River Tracking Method Intercomparison Project (ARTMIP): Project goals

- and experimental design. *Geosci. Model Dev.*, **11**, 2455–2474, <https://doi.org/10.5194/gmd-11-2455-2018>.
- Shin, J., S. Park, S.-I. Shin, M. Newman, and M. A. Alexander, 2020: Enhancing ENSO prediction skill by combining model-analog and linear inverse models (MA-LIM). *Geophys. Res. Lett.*, **47**, e2019GL085914, <https://doi.org/10.1029/2019GL085914>.
- Shin, N.-Y., Y.-G. Ham, J.-H. Kim, M. Cho, and J.-S. Kug, 2022: Application of deep learning to understanding ENSO dynamics. *Artif. Intell. Earth Syst.*, **1**, e210011, <https://doi.org/10.1175/AIES-D-21-0011.1>.
- Sillmann, J., and Coauthors, 2017: Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities. *Wea. Climate Extremes*, **18**, 65–74, <https://doi.org/10.1016/j.wace.2017.10.003>.
- Sinha, S., S. Giffard-Roisin, F. Karbou, M. Deschaires, A. Karas, N. Eckert, C. Coléou, and C. Monteleoni, 2020: Variational autoencoder anomaly-detection of avalanche deposits in satellite SAR imagery. *Proc. 10th Int. Conf. on Climate Informatics*, Online, Association for Computing Machinery, 113–119, <https://doi.org/10.1145/3429309.3429326>.
- Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032, <https://doi.org/10.1175/MWR2830.1>.
- Slivinski, L. C., D. E. Lippi, J. S. Whitaker, G. Ge, J. R. Carley, C. R. Alexander, and G. P. Compo, 2022: Overlapping windows in a global hourly data assimilation system. *Mon. Wea. Rev.*, **150**, 1317–1334, <https://doi.org/10.1175/MWR-D-21-0214.1>.
- Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 1617–1630, <https://doi.org/10.1175/BAMS-D-14-00173.1>.
- Sønderby, C. K., and Coauthors, 2020: MetNet: A neural weather model for precipitation forecasting. arXiv, 2003.12140v2, <https://doi.org/10.48550/arXiv.2003.12140>.
- Spirites, P., C. N. Glymour, and R. Scheines, 2000: *Causation, Prediction, and Search*. MIT Press, 543 pp.
- Stengel, K., A. Glaws, D. Hettinger, and R. N. King, 2020: Adversarial super-resolution of climatological wind and solar data. *Proc. Natl. Acad. Sci. USA*, **117**, 16 805–16 815, <https://doi.org/10.1073/pnas.1918964117>.
- Strader, S. M., and W. S. Ashley, 2015: The expanding bull's-eye effect. *Weatherwise*, **68**, 23–29, <https://doi.org/10.1080/00431672.2015.1067108>.
- Tabari, H., 2021: Extreme value analysis dilemma for climate change impact assessment on global flood and extreme precipitation. *J. Hydrol.*, **593**, 125932, <https://doi.org/10.1016/j.jhydrol.2020.125932>.
- Talsma, C. J., K. E. Bennett, and V. V. Vesselino, 2022: Characterizing drought behavior in the Colorado River Basin using unsupervised machine learning. *Earth Space Sci.*, **9**, e2021EA002086, <https://doi.org/10.1029/2021EA002086>.
- Thornton, P. K., P. J. Erickson, M. Herrero, and A. J. Challinor, 2014: Climate variability and vulnerability to climate change: A review. *Global Change Biol.*, **20**, 3313–3328, <https://doi.org/10.1111/gcb.12581>.
- Toms, B. A., E. A. Barnes, and I. Ebert-Uphoff, 2020: Physically interpretable neural networks for the geosciences: Applications to Earth system variability. *J. Adv. Model. Earth Syst.*, **12**, e2019MS002002, <https://doi.org/10.1029/2019MS002002>.
- Toride, K., and G. J. Hakim, 2022: What distinguishes MJO events associated with atmospheric rivers? *J. Climate*, **35**, 6135–6149, <https://doi.org/10.1175/JCLI-D-21-0493.1>.
- Trenberth, K. E., 1997: The definition of El Niño. *Bull. Amer. Meteor. Soc.*, **78**, 2771–2778, [https://doi.org/10.1175/1520-0477\(1997\)078<2771:TDOENO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2).
- Vandal, T., E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, 2017: DeepSD: Generating high resolution climate change projections through single image super-resolution. *Proc. 23rd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, Halifax, NS, Association for Computing Machinery, 1663–1672, <https://doi.org/10.1145/3097983.3098004>.
- Walsh, K. J. E., and Coauthors, 2015: Hurricanes and climate: The U.S. CLIVAR Working Group on hurricanes. *Bull. Amer. Meteor. Soc.*, **96**, 997–1017, <https://doi.org/10.1175/BAMS-D-13-00242.1>.
- Wang, J., Z. Liu, I. Foster, W. Chang, R. Kettimuthu, and V. R. Kotamarthi, 2021: Fast and accurate learned multiresolution dynamical downscaling for precipitation. *Geosci. Model Dev.*, **14**, 6355–6372, <https://doi.org/10.5194/gmd-14-6355-2021>.
- Wang, S., S. Sankaran, and P. Perdikaris, 2022: Respecting causality is all you need for training physics-informed neural networks. arXiv, 2203.07404v1, <https://doi.org/10.48550/arXiv.2203.07404>.
- Watson, C. D., C. Wang, T. Lynar, and K. Weldemariam, 2020: Investigating two super-resolution methods for downscaling precipitation: ESRGAN and CAR. arXiv, 2012.01233v1, <https://doi.org/10.48550/arXiv.2012.01233>.
- Watt-Meyer, O., N. D. Brenowitz, S. K. Clark, B. Henn, A. Kwa, J. McGibbon, W. A. Perkins, and C. S. Bretherton, 2021: Correcting weather and climate models by machine learning nudged historical simulations. *Geophys. Res. Lett.*, **48**, e2021GL092555, <https://doi.org/10.1029/2021GL092555>.
- Weyn, J. A., D. R. Durran, R. Caruana, and N. Cresswell-Clay, 2021: Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. *J. Adv. Model. Earth Syst.*, **13**, e2021MS002502, <https://doi.org/10.1029/2021MS002502>.
- Wille, J. D., and Coauthors, 2021: Antarctic atmospheric river climatology and precipitation impacts. *J. Geophys. Res. Atmos.*, **126**, e2020JD033788, <https://doi.org/10.1029/2020JD033788>.
- Xie, K., P. Liu, J. Zhang, D. Han, G. Wang, and C. Shen, 2021: Physics-guided deep learning for rainfall-runoff modeling by considering extreme events and monotonic relationships. *J. Hydrol.*, **603**, 127043, <https://doi.org/10.1016/j.jhydrol.2021.127043>.
- Yao, L., Z. Chu, S. Li, Y. Li, J. Gao, and A. Zhang, 2021: A survey on causal inference. *ACM Trans. Knowl. Discovery Data*, **15** (5), 1–46, <https://doi.org/10.1145/3444944>.
- Yuval, J., P. A. O'Gorman, and C. N. Hill, 2021: Use of neural networks for stable, accurate and physically consistent parameterization of subgrid atmospheric processes with good performance at reduced precision. *Geophys. Res. Lett.*, **48**, e2020GL091363, <https://doi.org/10.1029/2020GL091363>.
- Zarzycki, C. M., and P. A. Ullrich, 2017: Assessing sensitivities in algorithmic detection of tropical cyclones in climate data. *Geophys. Res. Lett.*, **44**, 1141–1149, <https://doi.org/10.1002/2016GL071606>.
- Zhang, J., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 621–638, <https://doi.org/10.1175/BAMS-D-14-00174.1>.