BAMS Article

The History and Practice of AI in the Environmental Sciences

Sue Ellen Haupt, David John Gagne, William W. Hsieh, Vladimir Krasnopolsky, Amy McGovern, Caren Marzban, William Moninger, Valliappa Lakshmanan, Philippe Tissot, and John K. Williams

ABSTRACT: Artificial intelligence (AI) and machine learning (ML) have become important tools for environmental scientists and engineers, both in research and in applications. Although these methods have become quite popular in recent years, they are not new. The use of AI methods began in the 1950s and environmental scientists were adopting them by the 1980s. Although an "AI winter" temporarily slowed the growth, a more recent resurgence has brought it back with gusto. This paper tells the story of the evolution of AI in the field through the lens of the AMS Committee on Artificial Intelligence Applications to Environmental Science. The environmental sciences possess a host of problems amenable to advancement by intelligent techniques. We review a few of the early applications along with the ML methods of the time and how their progression has impacted these sciences. While AI methods have changed from expert systems in the 1980s to neural networks and other data-driven methods, and more recently deep learning, the environmental problems tackled have remained similar. We discuss the types of applications that have shown some of the biggest advances due to AI usage and how they have evolved over the past decades, including topics in weather forecasting, probabilistic prediction, climate estimation, optimization problems, image processing, and improving forecasting models. We finish with a look at where AI as employed in environmental science appears to be headed and some thoughts on how it might be best blended with physical/dynamical modeling approaches to further advance our science.

Keywords: Artificial intelligence; Data science; Decision trees; Deep learning; Expert systems; Machine learning

https://doi.org/10.1175/BAMS-D-20-0234.1
Corresponding author: Sue Ellen Haupt, haupt@ucar.edu
In final form 5 December 2021

AFFILIATIONS: Haupt and Gagne—National Center for Atmospheric Research, Boulder, Colorado; **Hsieh**—University of British Columbia, British Columbia, Canada; **Krasnopolsky**—NOAA/Environmental Modeling Center, College Park, Maryland; **McGovern**—University of Oklahoma, Norman, Oklahoma; **Marzban**—University of Washington, Seattle, Washington; **Moninger**—Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, and NOAA/Global Systems Laboratory, Boulder, Colorado; **Lakshmanan**—Google Cloud, Bellevue, Washington; **Tissot**—Texas A&M University–Corpus Christi, Corpus Christi, Texas; **Williams**—The Weather Company, Andover, Massachusetts

Artificial intelligence (AI) and machine learning (ML) have become important tools for environmental scientists and engineers, both in research and in applications. Although these methods have become quite popular in recent years, they are not new; AI and ML have been used in the environmental sciences (ES) for decades. The American Meteorological Society (AMS) Committee on Artificial Intelligence Applications to Environmental Science (AI Committee) has been promoting, advancing, and educating about these techniques since the 1980s. The types of methods used have evolved over this time period and practitioners in the environmental sciences have helped lead the way. Here we tell that story through the lens of this AMS Committee.

Here we use the term AI to encompass any type of machine "intelligence," including expert systems that are typically a set of algorithms that codifies how an expert would make decisions. Machine learning as used here refers to a subset of AI in which an algorithm learns from data. Figure 1 illustrates the AI/ML landscape. Deep learning (DL) is a subset of ML that

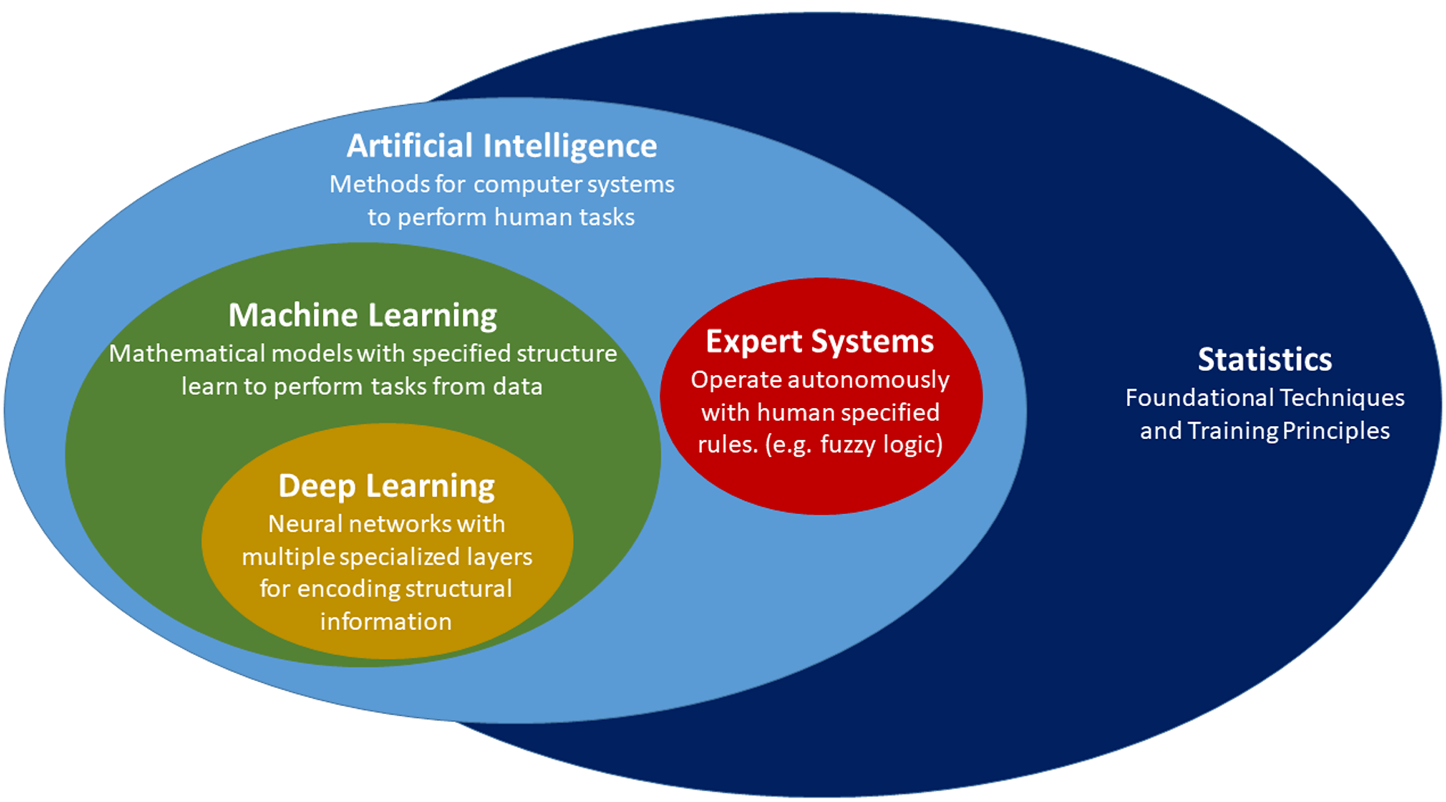

Fig. 1. Venn diagram of the relationship between AI, ML, DL, expert systems, and statistics.

has evolved more recently and typically involves neural networks (NN) with many specialized layers, which allows extracting deeper levels of information at each layer. Foundational to all of these is rigorous application as defined by statistics (see sidebar).

Learning from Statistics

What is the difference between AI/ML and "traditional statistics," hereafter referred to as A and S, respectively? The trouble with such questions is that A and S are assumed to have an essence. Although Aristotle and the Dialogues of Plato dedicate volumes to the definition of "essence," Wikipedia's summary will suffice: "abstract universals logically or ontologically separate from the objects of sense perception" (Wikipedia 2022). With that definition in mind, the question becomes nearly moot, because even the most ardent debaters of A versus S generally agree that A and S do not have a singular essence. And yet, the question arises repeatedly simply because it has practical consequences. For example, the authors of this article have been faced with that question in deciding whether a particular submission to the AMS AI conference is more suitable for the AMS probability and statistics conference.

The question is also important because awareness of the connections between A and S can mutually benefit both. For example, recognizing that feed-forward, multilayered perceptions (i.e., neural networks) are closely related to regression models in statistics behooves one to compare the two. However, at least in the early years of applying AI/ML in ES, it was rare to find such a comparison. Consequently, many A models were put forth for solving problems that could have been solved with much simpler S methods. Even worse, many A models were proposed that were in fact inferior to S models, simply because A models are generally more capable of overfitting data.

As in most debates that fall into this A versus S taxonomy, e.g., wave versus particle nature of light, or frequentist versus Bayesian perspectives in statistics, it often turns out that both are true. As such, it is reasonable to do both. That conclusion may seem obvious, but the fact remains that the majority of A articles make no attempt to compare their results with something on the S side. It is important to point out that doing both A and S has the added benefit of allowing one to answer the A versus S question grounded in the specifics of the problem, rather than in the vacuum of philosophical discourse.

Another reason for clearly highlighting the specific details of A versus S discussions is that such questions require a taxonomy that often does not exist. For example, do NNs (the deep or the shallow variety) belong to the A or the S side? Who invented cross validation—A or S folk? What about general linear models, general additive models, *K*-nearest neighbor, bagging, boosting, *K*-means, and random forests? Again, these questions assume the existence of an essence, which simply does not exist, given the evolutionary nature of the scientific process itself. Ideas gradually evolve, and it is often impossible to draw a border that separates the descendant from an ancestor. For example, it is reasonable to argue that NNs evolved from regression; in fact, some would argue that the evolutionary path is not long enough to consider one to be the descendant of the other; they would argue that regression and neural nets are still the same "species."

To illustrate the ambiguities, consider the following dialogue between an A and an S scientist:

S: Here is a model that we (statisticians) call simple regression, but you A folk call neural net: $y = \alpha + \beta x$.

A: But that has only one input.

S: OK, here is a model we call multiple regression: $y = \alpha + \beta_1 x_1 + \beta_2 x_2$.

A: Yes, but it assumes linear relationships.

S: OK, here is a model we call polynomial regression:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \ldots + \beta x_1^P + \beta x_2^P$$

A: But …

S: Which, by the way, is a special type of multiple regression, concisely written as: $y = \mathbf{X}\beta$, where $\mathbf{X}$ is a matrix containing the data on x_1 and x_2 and powers thereof, and β is a vector of all of what you A types call synaptic weights.

A: But that has only one output, y.

S: OK, here is a model we call multivariate multiple regression: $\mathbf{Y} = \mathbf{X}\beta$, where $\mathbf{Y}$ is now also a matrix whose columns contain the data on each of the outputs.

A: Yes, but as the number of inputs increases, the number of parameters in your model grows much faster than that in my neural net, and therefore your model can lead to overfitting.

S: Yes, but my model does not suffer from local minima, or the opaqueness associated with your neural net. Additionally, every aspect of my model (e.g., its weights and predictions) can be accompanied by confidence or prediction intervals.

Here, we gave the last word to the S fellow, but the dialogue is not likely to end on a productive point. Two, more constructive scenarios would be

A: You do S, and I'll do A, and then let us compare. (scenario 1)

or

S: You do A, and I'll do S, and then let us compare. (scenario 2)

Superficially, the two scenarios may appear identical. But the difference between the scenarios is in the "culture" and expectations underlying the respective proposals. In the first scenario, A is usually confident that their model will emerge as the winner. By contrast, in the second scenario, chances are that S suspects that by the time they do their comparisons *correctly* (see the verification section), they will not be able to say which model is better.

The story of AI in ES over the last four decades has been driven by a rapid evolution of Earth observations, communications bandwidth, compute capabilities, and AI/ML methodologies. Applying AI to real-world problems has become more urgent as the negative impacts of weather events have grown worldwide with a changing climate, growing populations in vulnerable areas, and continuing unsustainable practices. AI has entered the cultural mainstream as the social media, entertainment, and retail industries have seen companies, such as Google, Amazon, and Facebook, built on innovative AI become the most valuable in the world. Those companies, as well as academic researchers, have developed improved ML methodologies and an increasingly mature set of software tools, workflows, and best practices. Faced with a volume and velocity of environmental information far beyond the ability of humans alone to manage, ES practitioners have naturally turned to automation infused with AI and ML to derive new knowledge and provide real-time actionable predictions, augmenting the capabilities of researchers and forecasters for the benefit of society. While early activities of the AMS AI Committee included many focused on building knowledge bases and expert systems to encode and automate the thought processes of human experts, the advent of increasingly robust, fast, and explainable ML methods—including DL—has caused an inevitable transition toward leveraging ML methods. While expert knowledge remains essential to appropriately formulating the learning problem, including the data and "features" (quantities derived from the raw data) provided as input and the learning objective, these methods are free to discover and exploit relationships not previously known or easily articulated. As a result, over time, the orientation of AMS AI Committee conferences has been increasingly data driven.

While interest and progress in AI applications to ES have been largely driven by enhanced performance, computational efficiency (Chevallier et al. 2000; Krasnopolsky et al. 2002), and modeling flexibility, these gains also reflect its primary strength—the ability of ML to model complex and nonlinear systems. The development of ML methods has been accompanied at times with the ability to bring new understanding to the underlying physics. The development of ES expert systems was also seen as an occasion to codify the knowledge of the expert sources (Moninger et al. 1987). Neural networks were used as a nonlinear version of principal component analysis (NLPCA) as early as the late 1990s (Monahan 2000; Hsieh 2001), which are actually an ancestor to the present DL autoencoders. NLPCA brought new understanding to the dynamics of monthly tropical Pacific sea surface temperatures (Hsieh 2009). Random forests (RFs) (Breiman 2001) inherently allowed the estimation of the relative importance of each predictor to the performance of the model. The approach was generalized and other methods, either inherent to the ML technique or model agnostic, have been developed (such as permutation techniques; Mielke et al. 1981) to bring new insights into these nonlinear systems, including drivers of the systems' extreme events. Such methods are now part of the field of Explainable AI (XAI) further discussed below. While AI can help scientists better understand a system, scientific knowledge of the physics of the problem is important to configuring the method to select predictive features (see feature identification below), inform model architecture, group physically correlated features, or enforce other constraints guiding the calibration of the models toward physically consistent solutions.

A brief history of AI in environmental science

Although the concept of codifying human intelligence in programmable machines goes back centuries, the term "artificial intelligence" was coined by John McCarthy as he convened a conference at Dartmouth in 1956 to advance the use of machines to emulate human thought (AAAI 2017). The discussions between these researchers led to Department of Defense funding in areas such as language translation using primarily empirical approaches (Poole and Mackworth 2017). The early hype did not yield the anticipated results, leading to an "AI winter" in the 1970s where the funding agencies were not receptive to AI projects (Drew 1973). AI began to rebound in the 1980s with interest in applications, primarily using expert system approaches, where computers were supplied with rules to make decisions that mimic what an expert would choose. It was against this backdrop that atmospheric scientists began to make forays into applying AI to their problems as discussed below. But again, within the greater community, more was promised than could be delivered and a second AI winter ensued. Interest was rekindled when International Business Machines (IBM)'s Deep Blue beat Gary Kasparov at chess in 1997 (Smith et al. 2006). The most recent surge in AI was spurred by the success of DL methods (see below) to beat a world champion at the complex game of Go (Silver et al. 2016, 2017). Figure 2 illustrates these ebbs and flows in AI activity.

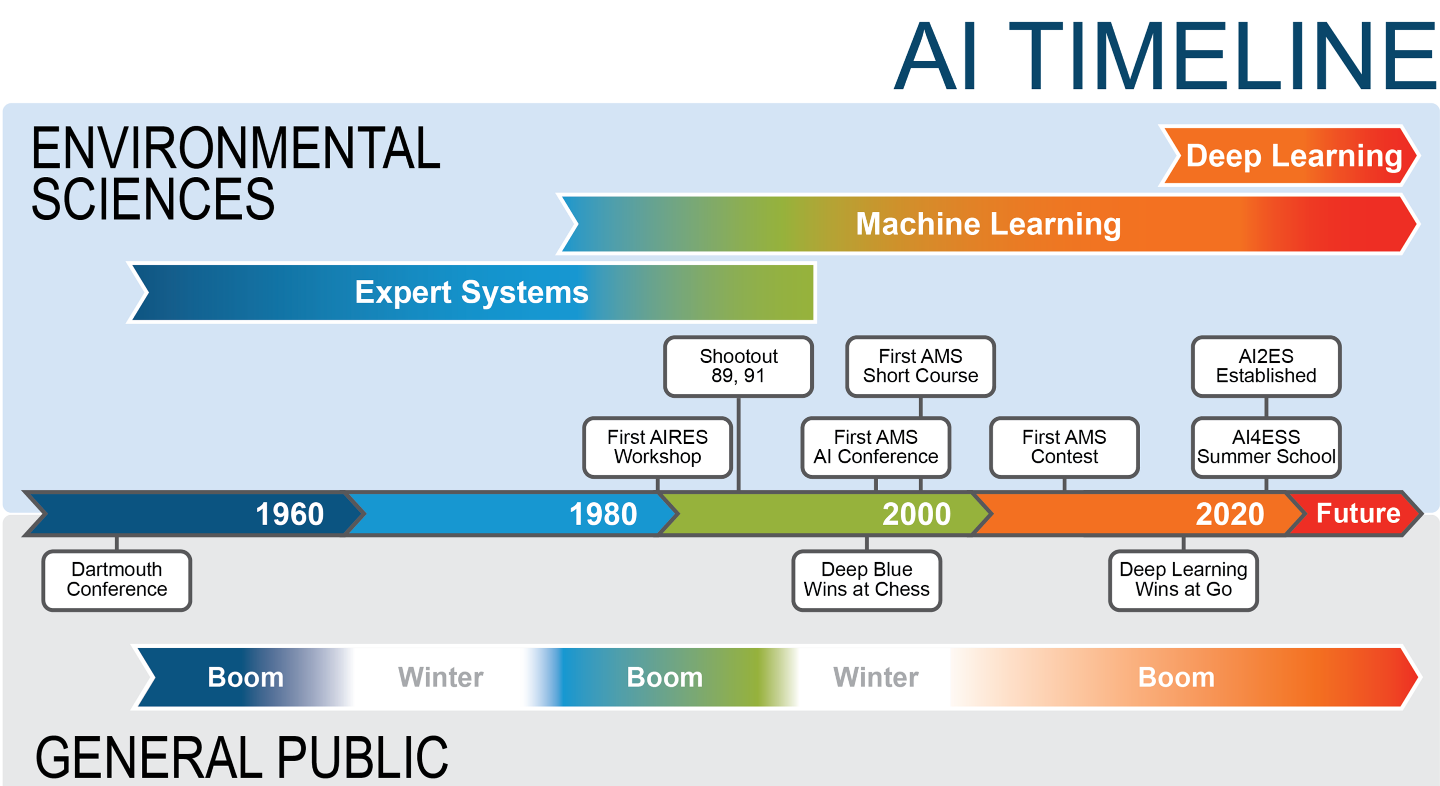

Fig. 2. Timeline of the progression of AI. The top portion indicates the timeline in the environmental science community, including popular methods employed in the colored arrows. The bottom portion traces the history of use within the greater AI community, including the progression of AI booms and winters along the arrow.

In 1984, the Environmental Research Laboratories (ERL) of the National Oceanic and Atmospheric Administration (NOAA) established the position of Special Advisor in Artificial Intelligence (coauthor Bill Moninger) to look into the potential of AI to improve weather forecasting and other aspects of NOAA's work. The primary and most well-understood implementations of AI in those days were expert systems, in which expert human intelligence is encapsulated into if–then rules that can be executed automatically. Such expert systems had shown efficacy in medical and other fields, and seemed to offer potential benefits to meteorology. The year 1984 also marks the start of the Artificial Intelligence Research in the Environmental Sciences (AIRES) workshops. The first AIRES workshops were organized in 1986 and 1987 in Boulder, Colorado. The records of these first workshops (Moninger et al. 1987; Dyer and Moninger 1988) and the ensuing other workshops and AMS conferences track the progression of AI methods in ES.

The AIRES workshops morphed into the AMS AI conferences, which began in 1998, organized by the same group of pioneers. While AI methods have changed considerably over time, the fields of their applications have been relatively constant. During this first phase of the development of AI for ES, during the mid-1980s and into the 1990s, several expert systems for weather forecasting were devised and tested. Developing the systems required extensive "knowledge engineering"—eliciting knowledge from experts in a form that could be encoded as executable rules (Dyer and Moninger 1988). Unfortunately, the limitations of expert systems, at least as applied to meteorology, became increasingly evident by the early 1990s. It is apparent that not all useful knowledge can be expressed in words, or can be described in if–then statements; indeed, a number of forecaster tasks seem more akin to visual pattern recognition than to analytic processing. Also, running the AI systems—answering the large number of questions posed by the systems—often took as much time for reasonably skilled forecasters as generating forecasts in more traditional ways. In addition, there was no obvious way for traditional expert systems to take advantage of rapidly improving computer capabilities for storing and analyzing large amounts of data—this is in contrast to other AI approaches, such as neural networks, that can naturally take advantage of greater computer power. These limitations of expert systems impacted other fields similarly and the initial glow of AI to revolutionize decision-making dimmed in many fields in the early 1990s, which is what ushered in the second AI winter (Fig. 2), when many AI-based computer corporations and AI startups failed, such as Symbolics, LISP Machines Inc., and Lucid Inc. (all companies building special-purpose AI hardware; Wikipedia 2021; Newquist 1994).

One exception to this trend was in the area of "fuzzy logic" (FL). Rather than attempting to encode human expertise into decision trees populated by "yes–no" branches, FL attempts to more closely model human reasoning's ability to draw on diverse, ambiguous sources of evidence to arrive at an informed assessment. FL extends classical dichotomous logic by allowing statements to have truth values along a continuum from 0 (false) to 1 (true). This continuum provides a powerful formalism for representing concepts and implications via "membership functions," allowing ambiguity and uncertainty to be propagated through an inference process and only resolved to a discrete answer at the final stage. FL was widely used at MIT Lincoln Laboratory (e.g., Delanoy and Troxel 1993) and the National Center for Atmospheric Research beginning in the early 1990s for a wide range of applications involving image processing, data quality control, and consensus prediction. Early successes included a microburst detection system for airports that ended a spate of aircraft crashes due to the phenomenon (Albo 1994). FL continues to be used in a number of applications, including remote sensing and hazard forecasting (Williams et al. 2009).

Tracking the AIRES and AI workshops and conferences tells the story of the development of AI in ES. As part of the 1987 AIRES II summary, neural networks are mentioned only once. When the ongoing series of AMS AI conferences started in 1998, expert systems were no longer mentioned and more than half of the talks were based on neural networks. The further evolution of AI techniques used in ES is illustrated in Fig. 3, which splits AI methods into seven categories: neural networks (shallow), DL, tree-based methods including random forests and boosted regression trees, genetic algorithms, FL, and support vector machines. When two AI methods are applied, half credit is tallied for each. The category "other AI methods" includes less frequently used techniques, comparison of more than two AI methods, or contributions discussing in more general terms the role and progression of AI in ES. The latter includes ethical discussion, risk communication, and other such topics. Figure 3 displays the percentage of presentations focused on each method as compared to the total number of AI presentations at each conference. From 1998 through 2005, neural networks are the focus of more than half of the presented works. Shallow neural networks are defined here as networks with three or fewer layers, (i.e., at most two hidden layers and an output layer). FL is the second-most-used method during the late 1990s through the mid-2000s followed by genetic algorithms. Work based on support vector machines is presented consistently starting in 2000 until the late 2010s. The use of tree-based methods became significant in the mid-2000s and continues to this day. These methods are currently the second-most-prevalent ML methods behind neural networks when shallow and DL models are combined. While as a percentage of works presented at the conferences the use of tree-based methods has been relatively stable during the late 2010s, this is relative to the total number of AI-focused presentations, which has grown dramatically.

Figure 4 estimates the work on AI in the ES field by displaying the number of presentations at AMS AI conferences since 1998, excluding works presented as part of joint sessions not directly focused on AI. The number of presentations increased from about 24 between 2014 and 2016 to 162 in 2020. For example, while on a percentage basis tree-based methods presentations were relatively stable around 20%, the number of works increased substantially from between 4 and 9 presentations per year to 30 in 2020. Figure 4 also illustrates the dramatic increase in overall AI works focused on ES starting around 2017–18. Prior to 2018 the AMS AI conference hosted between 19 and 47 presentations per year with an average of 31. DL methods emerged in significant numbers during the 2018 conference and increased rapidly from 5 presentations in 2018 to 50 in 2020. During the same time the total number of AI-focused

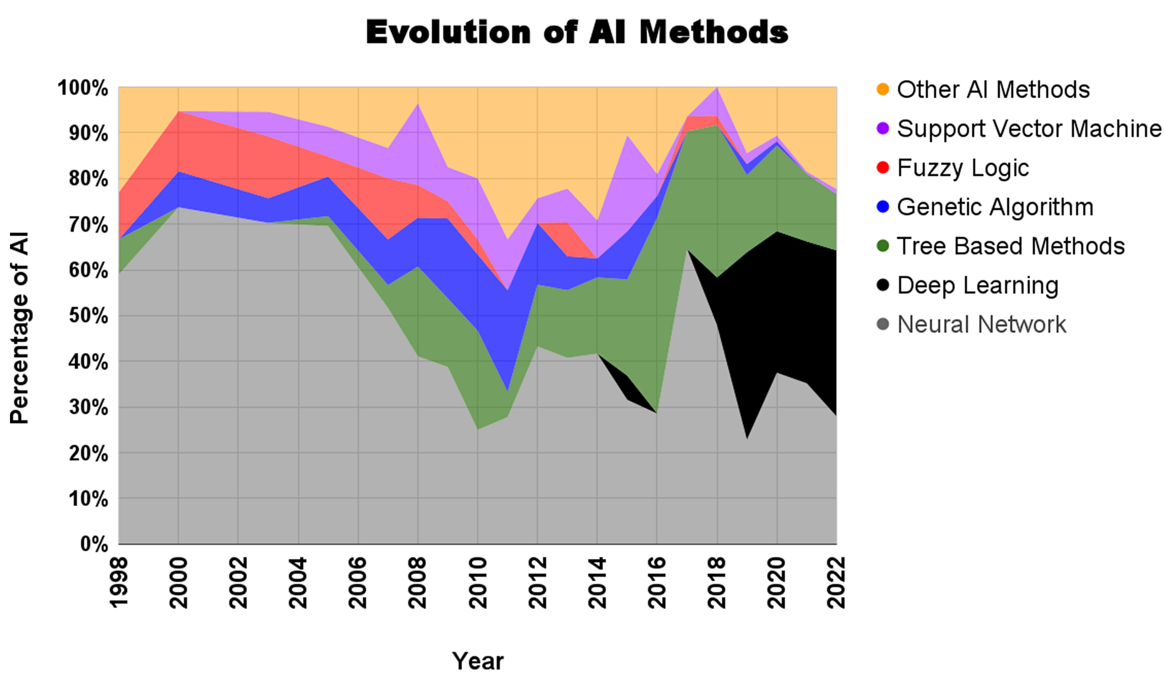

Fig. 3. **Evolution of the AI methods used for works presented at the AMS AI conferences through the years.**

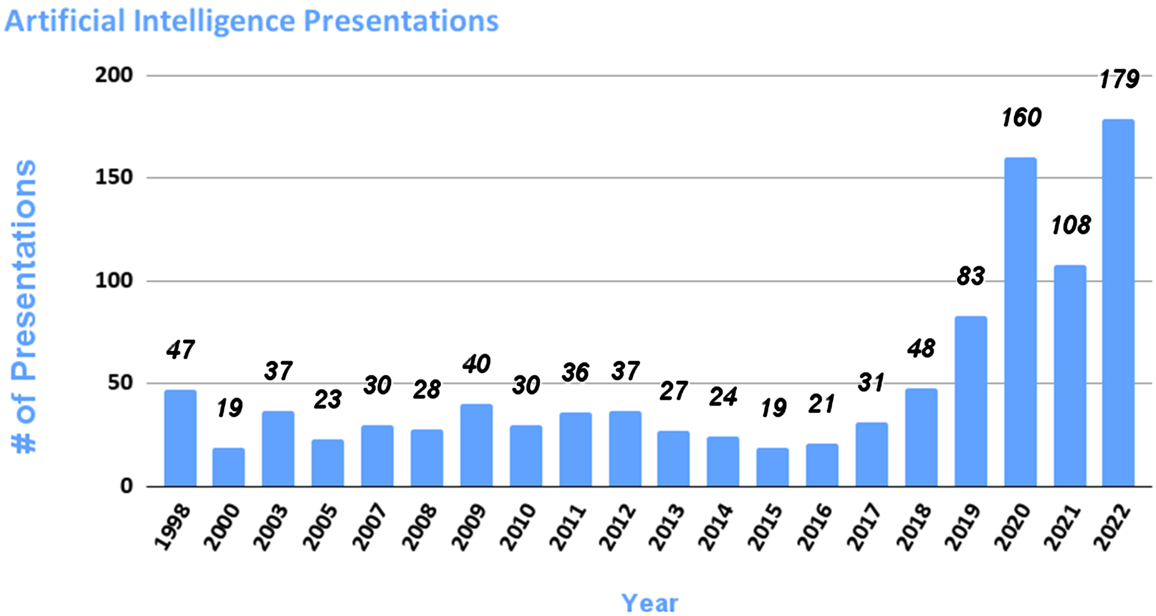

Fig. 4. Number of AI presentations at the AMS AI conferences through time including a large increase starting in 2018. Non-AI-focused presentations are not included. The 2021 conference was impacted by the COVID-19 pandemic.

presentations increased by a factor greater than 5, from 32 in 2017 to 162 in 2020.[1] This rapid increase in interest in AI was previously observed in computer science conferences with an inflection point about 4 years earlier around 2014 (Perrault et al. 2019). This dramatic rise in the use of AI clearly coincides with the DL revolution but is not strictly associated with DL. Shallow neural networks including self-organizing maps and tree-based methods represented about 33% and 20% of the presentations, respectively, in 2020 as compared to 35% for the DL presentations. During the previous AI transition from expert systems to neural networks and other techniques, research in the use of expert systems in ES all but ceased. This time, while DL is adding significantly to the AI toolbox, the new methods have not replaced the prior techniques, but rather have increased overall interest and use of AI for ES problems.

[1] While only 109 AI-focused works were presented in 2021, this conference was entirely virtual due to the COVID-19 pandemic.

This interest in DL for ES problems that skyrocketed in the mid-2010s led to the AI Conference hosting dedicated DL sessions that covered the wide variety of applications where it has been applied. One remarkable aspect of DL adoption is how quickly some highly complex methods were widely adopted across the research community after being introduced the previous year. One particularly notable example is the U-Net image segmentation method (Ronneberger et al. 2015). The U-Net is a neural network that encodes and then decodes images to other images at different spatial scales and can be applied across arbitrarily sized grids. Because of its strong fit with many geospatial problems, and the availability of U-Net implementations through open source DL libraries, the network was quickly adopted for a wide range of domain applications (e.g., Kumler-Bonfanti et al. 2020).

The role of short courses, contests, and books in advancing the use of AI in ES

The AMS AI Committee has been running short courses on how to use AI/ML for environmental science (including in Long Beach in 2000, Orlando in 2001, Seattle in 2004, Atlanta in 2006, Corpus Christi in 2007, Seattle in 2011, Seattle in 2017, Austin in 2018, Phoenix in 2019, and virtually in 2020 and 2021). We have also collaborated on short courses with the NOAA Workshops on Leveraging AI for Environmental Sciences in 2019 and virtually in 2020. At the beginning, only a handful of scientists registered. As an indication of how quickly excitement is building around AI/ML, the past few years of short courses have sold out within a few days of opening. Even with the shift to online in 2020 and 2021, the courses have filled to capacity quickly. NCAR's AI for Earth System Science Summer School (AI4ESS) held in June 2020 was originally planned for about 50 students in Boulder, Colorado, but when it became a virtual conference, it saw more than 8,000 individual logins.

The AMS AI Committee used the lectures compiled for the 2007 short course as a springboard for compiling an edited book on *Artificial Intelligence Methods in the Environmental Sciences* (Haupt et al. 2009), which documented various successful applications. Books by several other AMS AI Committee members also helped expose the utility of these techniques (Hsieh 2009; Krasnopolsky 2013; Haupt and Haupt 2004).

Contests were another method used to help popularize applications of AI in ES. To encourage the development and evaluation of AI systems for weather forecasting, NOAA's ERL sponsored two forecast contests, which became known as Shootout-89 (Moninger et al. 1991) and Shootout-91 (Walker et al. 1992). Both of these contests addressed the question of forecasting severe weather over the U.S. High Plains. Six systems participated in Shootout-89: three traditional expert systems, a hybrid system including a linear model augmented by a small expert system, an analog-based system, and a system developed using methods from the cognitive science/judgment analysis tradition. Forecast skill of the systems was generally low—not surprising given that the forecast task (forecasting afternoon convective storms at 1030 local time) is known to be a very difficult one. And because of the low scores, skill differences among the forecast models were difficult to determine.

Because of these limitations, Shootout-91 was designed and took place in Colorado and Oklahoma in the spring and summer of 1991. Unfortunately, the results were not more impressive than those from 1989. Indeed, among the conclusions published is that "it is not likely, however, that the systems will be able to perform as more than second opinions in the foreseeable future" (Walker et al. 1992), suggesting that the limitations of expert systems in weather forecasting were becoming evident by 1992.

Later forecasting contests, such as those reported in Lakshmanan et al. (2010) and McGovern et al. (2015), include a variety of AI methods beyond expert systems, such as neural networks, decision trees, random forests, and gradient boosted forests. Those contests addressed more restricted tasks. One tested the use of polarimetric radar data to develop a hydrometeor classification algorithm to distinguish between frozen and liquid hydrometeors, or none. Others challenged the participants to provide solar energy forecasts, wind power forecasts, or en route aviation turbulence encounter predictions given a large number of features derived from an NWP model, satellite, and radar. However, even in this case, significant skill differences among the distinct ML methods (when well applied) were not found: "The lack of statistical significance is not so much due to shortcomings of particular methods as it is to natural variability in the dataset" (Lakshmanan et al. 2010). Here we can summarize the results of these several forecast contests, along with the experiences of many system developers:

- Forecast skill depends far more on the quality of the input data (training data, and in the case of real-time forecasts, data on current meteorological conditions) than on the quality of the particular forecast algorithms. This also means that much of the effort spent on "winning" the contests is spent on data engineering. In the DL era, hand-engineering of features has been substituted with hand-engineering and automated searches of neural network architectures, which is feature engineering under another guise.
- Determining significant differences in skill requires a great deal of data spread over many forecast situations.
- For a related lesson, if you make any unintended exploits in a scoring method for a contest, expect someone to find it and use it.

These contests provided positive practical experience with AI and meteorological data but did not bring in additional existing AI practitioners. They did, however, help identify promising talented students and introduced them to a broader network of mentors and opportunities. To bring additional AI talent to our problems, the committee hosted the 2014 contest on Kaggle, a platform that invites AI practitioners to test their methods on archived data. That contest drew competitors external to the field. To our surprise, the top three performing algorithms were all from outside the ES community and based on gradient boosted regression trees (GBRT), a method that had not been widely used in ES applications previously. The community learned from this experience and GBRTs are now commonly used in modern applications.

AI usage in environmental science

AI/ML methods have been applied to a plethora of problems and for a variety of use cases in ES. The list below only scratches the surface, but demonstrates the wide applicability of these methods.

Postprocessing for improved forecasting. Using AI to postprocess NWP output is one of the most prevalent applications of AI in ES because it typically works well to improve forecasts of both basic and derived variables (Haupt et al. 2020a). This usage is particularly important for forecasting events that are not resolved by the NWP model, such as clouds, hail, and tornadoes. Although NWP has advanced in speed and accuracy over the decades, particularly with the advance of computing technology, accuracy is still imperfect and decreases with lead time as memory is lost from the initial condition, which is based on assimilating observations. Additionally, models have biases, which when determined can be corrected. Although some errors can be corrected with multilinear regression [model output statistics (MOS); Glahn and Lowry 1972], AI/ML methods are now doing a better job of capturing the nonlinearities.

It has long been understood that expert consensus is almost always more skillful over time than any single forecaster, which is why stock market index funds are so difficult to beat (Timmermann 2006). AI approaches to both NWP ensemble postprocessing for forecasting and data fusion of NWP and observations for nowcasting have proven pragmatic and successful, penetrating deeply into the weather enterprise. For example, the Dynamical Integrated Forecast (DICast) system developed at the National Center for Atmospheric Research (Gerding and Myers 2003; Myers and Linden 2011; Myers et al. 2011) dynamically optimizes bias corrections and weights for the correction and combination of a collection of input model forecasts, using stochastic gradient descent or ridge regression. This approach was adopted and extended by The Weather Company (TWC; Koval et al. 2015; Williams et al. 2016), where it is used to provide billions of individual forecasts daily around the globe. TWC also makes use of an “augmented intelligence” human-in-the-loop capability that allows forecasters to add filters to constrain or nudge the automatically updating AI predictions before they are delivered to users.

Statistical and AI methods have also advanced probabilistic forecasting. Such methods are used to derive calibrated probabilities and scenarios suitable for quantitative decision-making from multimodel NWP ensembles (e.g., Delle Monache et al. 2013; Williams et al. 2018). AI methods have also been applied to derive nonlinear averages of single-model and multimodel ensembles (Krasnopolsky and Lin 2012; Campos et al. 2020; Fan et al. 2022). Those works show that AI methods can improve upon standard conservative ensemble averaging and linear regression averaging techniques because the AI approach takes into account nonlinear correlations between ensemble members that standard statistical approaches miss.

Use in the private sector. The advent of increasingly user-friendly AI tools, cloud computing, and Internet bandwidth for moving large amounts of both open and proprietary environmental data have enabled a plethora of private sector applications. With low barriers to entry, startup companies can quickly add value to public forecasts by tuning them to local conditions, client sensors, or for specific applications. Established corporations provide bespoke weather and climate solutions, self-service cloud/AI/data platforms and industry-specific alerting, monitoring and decision support tools for agriculture, ground transportation, aviation, energy and utilities, insurance, financial services, retail, supply chain optimization, media and entertainment, tourism, and more. Many of these solutions are never published in academic journals, but have been presented in AMS conferences. AI appears uniquely positioned to help map weather data to predictions, impacts, outcomes, and decisions, helping to extract more of the value from weather information (e.g., Williams and Neilley 2020). The conditions seem ripe for private sector AI weather applications to continue rapid growth.

Feature engineering. Much of the early work in machine learning in the environmental sciences involved crafting the feature inputs into relatively simple models. A good example is the problem of identifying ground clutter and anomalous propagation (AP) from weather radar. The initial approaches were expert systems based on hardcoded rules. For example, Johnson et al. (1998) applied rules around changes in reflectivity value along a radial and treated extremely large changes between gates as AP. The 1D rules had lots of false positives, and could be made more sophisticated in three ways—by improving the statistics of what constituted unusually large changes (e.g., 2D fractal methods by Charalampidis et al. 2002), by looking for unusual changes in the 3D radar volume (e.g., Steiner and Smith 2002), or by incorporating an FL system consisting of many piecewise linear rules (e.g., Kessinger et al. 2003). These piecewise linear rules could be more effectively combined by means of a neural network (e.g., Lakshmanan et al. 2007) but operational forecasters preferred the interpretability offered by decision trees; thus, operational systems (e.g., Hubbert et al. 2007) tended to use readily interpretable rules combined using simple, nonstatistical approaches.

Similarly, the early hydrometeor classification algorithms (e.g., Park et al. 2009) consisted of a large number of rules to identify different hydrometeor types using statistical analysis of the distribution of hydrometeors as observed by both radars and disdrometers. However, a neural network based on optimally combining features of the same form as the expert rules (Lakshmanan et al. 2014) is used operationally in the Multi-Radar Multi-Sensor (MRMS) system. To aid with operational adoption, the 2014 paper used model explainability through an analysis of feature importance to aid with forecaster confidence in the model.

The advent of DL has made the approach of building structured data neural networks to process images obsolete. With the development of convolutional filters and much more sophisticated ideas, it is possible to do away with feature engineering altogether and directly train on the raw images using DL models, as demonstrated for lightning prediction (Lakshmanan et al. 2019) and precipitation nowcasting (Agrawal et al. 2019). There remain two open questions. One is whether this approach will meet the oft-expressed desire of operational forecasters to avoid relying on "black box" models. Explainability methods such as XRAI (Bartelt et al. 2020) might help in this regard. The second open question is whether we can obtain sufficient labeled data to carry out DL. Lightning prediction and precipitation nowcasting were convenient in that the observed data itself functions as the labels. This is not always true. One potential approach could be to use transfer learning or fine-tuning of large models (such as EfficientNet; Tan and Le 2019).

Hybrid approaches in NWP and climate modeling systems. Due to their growing complexity (e.g., increase of horizontal and vertical resolutions, the number of ensemble members, and the sophistication of model physics), both global and regional numerical weather and climate modeling activities consume a tremendous amount of computational resources, presenting model developers and users with a significant challenge despite the availability of expanding computing capabilities. Ensemble forecasting systems are particularly impacted by the limits on computational resources that constrain their resolution and/or the number of the ensemble members. Also, despite significant progress in understanding of physical phenomena and improvement in observing systems, substantial uncertainties remain in the representation of many processes in numerical models, e.g., effects of clouds in general circulation models of the atmosphere. New flexible and powerful numerical techniques are required to speed up model calculations and learn underlying physics from available data.

Almost from the beginning of the computer-based numerical weather forecast era in the 1950s, the numerical weather and climate models contained the equations of motion codifying physical processes (i.e., deterministic or physically based approach) in the atmosphere and the ocean as well as statistically derived coefficients and simple dependencies. In a sense, numerical models have always been hybrids of deterministic and statistical approaches. However, the deterministic approach has dominated, and statistical components played a supporting role.

The advent of ML approaches changed the distribution of power between deterministic and statistical/ML constituents of numerical models, significantly increasing the incorporation of statistical/ML components in new hybrid NWP models. ML applications developed in this field during the last several decades are focused on two major tasks: (i) improving computational performance of NWP models by replacing time consuming physically based parts of the model with their accurate and fast ML emulations (e.g., Chevallier et al. 2000; Krasnopolsky et al. 2002) and (ii) developing improved (as compared with physically based parameterizations) ML parameterizations based on learning physical processes from data observed or/and simulated by higher-resolution models like large-eddy simulations (e.g., Krasnopolsky et al. 2013; Brenowitz and Bretherton 2018; Rasp et al. 2018). Attempts to completely replace physically based numerical weather and climate models with the state-of-the-art black-box ML-based ones have usually met limited acceptance in scientific domains due to their inability to provide a meaningful physical interpretation of underlying processes, their large data requirements, and their limited generalizability to out-of-sample scenarios. Given that neither an ML-only nor a physically based–only approach can be considered sufficient for complex scientific and engineering applications, the research community has been exploring the continuum of hybrids of physically based and ML-based models, where both scientific knowledge and data are integrated in a synergistic manner (Krasnopolsky and Fox-Rabinovitz 2006; Rai and Sahu 2020). This paradigm is fundamentally different from mainstream practices in the ML community where domain-specific knowledge is often used in secondary roles, e.g., feature engineering or postprocessing. In contrast to these practices that can only work with simpler forms of heuristics and constraints, hybrid NWP models incorporate a tighter coupling of ML-derived model components with scientific knowledge embedded into physically based model components (e.g., Beucler et al. 2021).

During the last 20 years several hybridization approaches have been developed and applied that blend ML and physically based components of NWP and climate models at different hierarchical system levels. The approach formulated by Chevallier et al. (1998, 2000) and applied to longwave radiation introduced a hybrid parameterization, i.e., applied hybridization inside of the parameterization. This approach combined, in the hybrid longwave radiation parameterization, calculations of cloudiness based on first principle equations with NN approximations for a partial or individual flux at each vertical level. Recently, this approach was applied by Veerman et al. (2021), with the opposite combination of cloudiness computed with an NN and first principle equations to produce a hybrid parameterization based on a standard radiation parameterization (Pincus et al. 2019). Those authors applied an NN to emulate atmospheric optical properties and relied on radiative transfer equations to calculate the outputs of the RRTMGP parameterization.

Krasnopolsky et al. (2002, 2005) and Krasnopolsky and Fox-Rabinovitz (2006) introduced a hybrid model physics approach that combines ML parameterizations (e.g., shortwave and longwave NN radiation) with moisture and other parameterizations based on first principles. This approach was applied by many authors (Goldstein and Coco 2015; O'Gorman and Dwyer 2018; Brenowitz and Bretherton 2018; Rasp et al. 2018; Bolton and Zanna 2019; Pal et al. 2019; Wang et al. 2019; Beucler et al. 2021) who combined NN and physically based parameterizations in different ways inside the model physics suite. The aforementioned approaches usually use data simulated by numerical models.

A recent development has been to use observational data to build an ML model to replace the prior parameterization. Both NN and RF approaches have been shown to better model the surface layer than does the semiempirical Monin–Obukhov similarity theory (McCandless et al. 2022). More recently, hybrid GCMs have been introduced that combine the NN dynamics with a physically based model physics suite or the physically based dynamics with an NN model physics suite. Several attempts have been made to create ML emulations of regional and simplified GCMs (Van der Merwe et al. 2007; Scher 2018; Dueben and Bauer 2018; Willard et al. 2021; Weyn et al. 2020). It was shown by Van der Merwe et al. (2007) that an NN can provide very fast (1,000 times faster) and accurate emulation of a coupled large-scale circulation model for the Columbia River, its estuary, and near-ocean regions. Scher (2018) used a convolutional NN to emulate a dry hydrostatic aquaplanet at ~500-km horizontal resolution with 10 vertical levels, without either diurnal or seasonal cycles and with idealized physics. Dueben and Bauer (2018) used a toy model for global weather prediction to identify challenges and fundamental design choices for a forecast system based on neural networks. Weyn et al. (2020) built a data-driven global weather forecasting framework using a deep convolutional NN. Such fast and accurate emulations (if successful) will enable significant advances in developing new geophysical modeling systems. They may serve as an improved incarnation of statistical models widely used before the numerical weather prediction era for short term and local forecasting.

Extreme weather prediction. As pointed out for postprocessing above, although operational NWP gives a good basic weather forecast, it is not able to resolve local details and error often increases with time from forecast issue. Providing forecasts for severe weather requires improvements in both resolution and short-term predictive skill: we need information to support decisions at detailed locales and warnings issued early enough to take action. In recent years, AI/ML have proven to be quite skillful at improving forecasts of severe weather and its associated hazards (McGovern et al. 2017). Some of the early work focused on predicting specific hazards, including hail (Gagne et al. 2017) and severe winds (Lagerquist et al. 2017). In both cases, traditional machine learning methods including RFs and gradient boosted forests proved most adept at forecasting the hazards. More recently, the ability of DL to provide detailed spatial information of weather fields has further improved prediction for tornado (Lagerquist et al. 2020) and hail forecasting (Gagne et al. 2019). The methods used in these studies additionally allow adding elements of interpretability through back-propagation to determine the patterns that cause the severe weather. As the methods have gained visibility with their predictive power, it has become clear that it is important to provide insight into how the methods perform internally (McGovern et al. 2019) and to work directly with forecasters to understand their needs (Burke et al. 2020).

Hydrological prediction. Hydrology has important applications in many areas—water supply, hydropower generation, drought/flood and landslide risks, irrigation, sediment transport, etc. The relation between streamflow and precipitation is complex—water from precipitation is affected by the type of soil and vegetation in the watershed, or locked into snow or ice, before it eventually feeds into the streamflow. Hence "conceptual" or physical models that try to model the physical mechanisms of the hydrological processes are not very skillful in forecasting streamflow from precipitation data. NN models adapted by hydrologists to predict streamflow (Crespo and Mora 1993; Karunanithi et al. 1994) quickly became popular, as they are much simpler to develop than the physical models and offer better skill in predicting streamflow. Review papers show the growth of NN usage, such as Maier and Dandy (2000) and Dawson and Wilby (2001), which cite 43 and 51 papers, respectively, and more recently, Abrahart et al. (2012), Zounemat-Kermani et al. (2020), and Sit et al. (2020).

As outputs from general circulation models are of relatively coarse resolution, NN models have been used to downscale to local streamflow (Cannon and Whitfield 2002). Besides input from local observations and NWP, adding climate indices of atmospheric teleconnections to NN models improved daily streamflow forecasts during longer lead times of 5–7 days (Rasouli et al. 2012). Online learning, which provides for continually updating the ML model as new data arrive, allows streamflow forecast models to be continually updated inexpensively (Lima et al. 2016, 2017).

Snowpack on mountains is important for storing winter precipitation and releasing the water during spring/summer. The decline of mountain snowpack from a warmer climate, e.g., in western North America (Mote et al. 2005), negatively impacts water supply and hydropower generation during the drier months. NNs allow retrieval of snow water equivalent (SWE) and snow depth from microwave satellite data (Tedesco et al. 2004). Over mountainous terrain, as coarse-resolution gridded products of SWE are inaccurate, NN models can provide downscaled SWE estimates (Snauffer et al. 2016).

Climate applications. In general, there have been far fewer applications of AI/ML methods to climate problems than weather problems, as the much longer time scales in climate problems lead to much smaller effective sample size for observed data. "Transfer learning" has helped to overcome the limitation of small effective sample size by utilizing the almost unlimited amount of simulation data from climate models. Ham et al. (2019) used a convolutional neural network (CNN) model to learn El Niño–Southern Oscillation (ENSO) behavior from coupled dynamical models (CMIP5 climate model data for 2,961 months). The CNN model was further trained with 103 months of reanalysis data, i.e., the CNN first learned from the large coupled model dataset, then transferred the learning to the small reanalysis dataset. The resulting CNN model had better accuracy in ENSO prediction than the dynamical models.

Another interesting climate application was accomplished by Pasini et al. (2017), who used both anthropogenic and natural environmental variables to build a NN model of climate over the past 160 years. They then used sensitivity analysis by fixing certain variables while allowing others to change in order to determine changes in the model under differing assumptions. When they fixed anthropogenic forcing at preindustrial levels, the NN results differed from those observed, indicating that anthropogenic forcings were associated with the observed changes in temperature.

ML can also aid in providing actionable information for climate-related decisions. The energy industry requires information on projected changes in the resource for both wind and solar. Haupt et al. (2016) addressed this issue using model output from current reanalyses as well as from regional climate models and a series of statistical and ML methods to estimate current and projected climate with similar patterns. This process leveraged NN-based self-organizing maps (SOMs) to recognize patterns and to project the future climate changes. Stengel et al. (2020) use generative adversarial networks to downscale coarse-resolution (100 km) climate model data to resolutions (2 km) required for renewable energy planning.

The importance of verification

Verification generally refers to the assessment of the quality of forecasts, often in comparison to some benchmark (e.g., random forecasts, or forecasts from a competing model). As such, the early years of AI/ML in environmental sciences witnessed a great deal of comparisons. Although comparisons still abound, a shortage of "error bars" exists. For example, an analysis of whether the mean squared error (MSE) of model A is lower than that of model B, in a sense that generalizes beyond the specific data/sample used for estimating the MSE is incomplete without some sort of a two-sample test or confidence interval. Similarly, properly comparing the temperature forecast on a given day from two competing models requires computing prediction intervals for both.

The essence of p values, confidence, and prediction intervals first acknowledges, then quantifies, variability. For the statistician, variability is a good thing: something to be explained; after all, would anyone consider a repeated sequence of constant numbers "data"? But a sequence of varying numbers is "real data" and begs to be modeled. Without this appreciation for variability, much of the initial work on AI/ML in the environmental sciences treated variability as a bad thing—noise—and consequently, the belief was propagated that by throwing a sufficiently complex AI/ML model at a problem, one can, and should, eliminate that noise. The result was a large set of models that were overfitting data, unbeknownst to their developers. Although use of resampling methods such as cross validation and bootstrapping has mostly eliminated overfitting concerns, the fact remains that in some AI/ML circles variability is not adequately taken into account.

For example, resampling methods address sampling variability. But there exist other sources of variability that make verification (model comparison) difficult especially in AI/ML. Most AI/ML models rely on optimization algorithms for parameter estimation, and all such algorithms can get stuck in local minima. How is this a source of variability? The answer becomes obvious upon noting that optimization algorithms generally require an initial guess for the values of the parameters and a different initial guess can lead to a different local minimum. Just train your favorite NN starting from a different set of random weights and you will end up with a trained network that has different weight values, and therefore, different training and validation errors. Most practitioners witness this source of "computational variability," but few take it into account when performing model comparison. Why is this important? Because without accounting for this source of variability one may conclude that model A is better than model B, where in fact there is no evidence from the data to favor one model over the other. In such cases, it would be prudent to adopt the simpler model regardless of what the validation errors may suggest. Work in progress performed by some of the authors has shown that variability due to local minima in neural networks can in fact be larger than that due to cross validation, suggesting that more attention to computational variability is warranted. Note that traditional statistical methods that do not rely on optimization algorithms, e.g., multivariate multiple regression, have no sources of computational variability, and are, therefore, easier to compare.

These (and other) issues point to the fact that to apply AI for ES correctly, one must first be steeped in methods of proper statistical analysis (see sidebar). There are appropriate methods for training and testing methods and for comparing their accuracy to other methods. Details are beyond the scope of this paper, but we refer the reader to Chase et al. (2022, manuscript submitted to *Wea. Forecasting*) as well as many excellent statistics and ML textbooks.

Future directions

AI is in the midst of its fastest and most productive developmental phase, including in the environmental sciences. Ongoing and foreseen advances are discussed below while continuing to provide historical context. The fast growth of AI and the related growth of the AMS AI conference have also brought us to a new phase in AI progression. AI/ML is no longer a niche specialty but is an approach embedded in a growing number of ES disciplines and, as such, is now part of most AMS conferences. The AMS AI conference will need to evolve synergistically with other conferences toward core AI developments benefiting the community broadly while progressively moving away from more discipline-specific applications. This transition has been ongoing through a large number of joint sessions and requires continued and deliberate progression to minimize the pitfalls of this rapid development. The growth in application of AI to ES problems and its importance to the field has led to the advent of a new AMS journal in the field, *Artificial Intelligence for the Earth Systems*. In addition, one of the largest AMS constituents, NOAA, has embraced AI with over 200 applications (Sims et al. 2020), an AI strategy (NOAA 2021), and initiated a series of annual workshop on the use of AI in weather and climate applications (Boukabara et al. 2021).

All of this points to the need for continued collaboration and advances in how we work together to train and support new AI practitioners. AI specialists must include in their research guidelines on how to develop and use AI models; how to make AI more understandable, explainable, and trusted by stakeholders; and provide ethical guidance. Although a strength of the AMS AI Committee has long been an integration of computational specialists with physicists who use AI in pioneering ways, we expect membership to evolve to include those who seek to use AI as a tool in their research and applications. In particular, the Committee will seek a diverse membership to assure equity in the way we approach AI/ML.

Enhancing forecasting. While AI/ML has demonstrated its efficacy at improving NWP-generated forecasts, there is much left to be done. First, although most methods have focused on making improvements of point forecasts and sometimes interpolating to gridded forecasts, for the most part, not much has been accomplished on improvements in all dimensions of the problem simultaneously. The newer DL methods promise that will change in the coming years. We are now demonstrating that it is possible to correct for phase errors in the forecast (Chapman et al. 2019). Additionally, methods to enhance probabilistic forecasting using ML are improving forecasts while saving substantial computational costs, such as the analog ensemble method discussed above, and are becoming more widely adopted. As discussed above, AI/ML is also showing its usefulness at both emulating and building new parameterizations for implementation within the physics models that promise to both speed and improve our NWP and climate modeling capabilities.

Enabling better use of observations. The weather enterprise currently makes use of only a tiny fraction of environmental observations, and AI offers one of the most promising avenues to derive value from the deluge. For instance, AI has begun to play a key role in quality-controlling satellite, radar, and other observation data, including the diverse and rapidly growing set of internet-of-things data. AI has proved useful for model identification; for instance, genetic algorithms can optimize parameters in heuristic or physical models, such as estimating transport and diffusion model parameters from plume observations (Haupt et al. 2013), and is being explored as a tool for increasing the use of data in NWP data assimilation (e.g., Boukabara et al. 2019). AI/ML are also used for inferring the environmental state from remote sensing data, such as algorithms developed under the GOES-R program that retrieve or diagnose various state variables, from precipitation rates to clear-air turbulence. Moreover, AI and ML are uniquely useful in learning skillful "data fusion" of NWP model output and observations for nowcasting (McCandless et al. 2016, 2020; Haupt et al. 2020b).

Interpretable/explainable AI. Given the explosion in the use of AI/ML, many forecasters want to know what is happening inside the AI methods, to improve their trust and use of the methods. Model interpretation and visualization has become a growing field, primarily now called Explainable AI (XAI) (Flora et al. 2022a,b, manuscripts submitted to *Artif. Intell. Earth Syst.*).

XAI methods vary based on the underlying machine learning method being used but they primarily focus on identifying the most important variables in a model (this is of particular importance for selective models such as random forests) and interactions among variables. With DL growing in popularity across many scientific fields, new XAI techniques are being developed to peer inside the black box of deep networks. These include techniques that highlight the most important regions of input images [see McGovern et al. (2019) and Gagne et al. (2019) for examples in severe weather]. Recent work (McGovern et al. 2020) has demonstrated that such techniques also need to use physics-based constraints to improve the interpretability of the results as well as improving model performance.

Trustworthy AI. As AI/ML continues to grow in popularity in the atmospheric sciences, the practitioners and creators of the AI/ML methods need to focus on the trust of the end-users of these methods. When AI began being applied to the atmospheric sciences, it was for research purposes and not for operational use. To transition to operations, AI must prove trustworthy. Forecasters are already bombarded with information and will not pay attention to additional input if they do not trust it. Part of the goal of XAI is to improve trust by scientific end-users by providing them with a better understanding of how a model generates its predictions. Additionally, knowledge of the scenarios in which the model tends to do well or poorly enables the user to more confidently calibrate their confidence in the model predictions in individual cases. Studying trust from the social science aspect is also critical as it is important to understand why different end users will choose different sources of information and what can be done to improve trust in AI. This is one of the goals of the newly formed NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography.

Summary

AI/ML is here for good. Despite several AI winters, the current push to apply AI in many fields is certainly occurring in the environmental sciences. As we have seen herein, it is not a new trend, but rather has been building in the field since the 1980s. The AMS AI Committee has fostered this growth in interest over several decades now. Over that period, the use of AI/ML in ES has evolved from the more heuristic expert systems to the full realm of available methods, including applications using the various types of DL. There have been many successes in the field, and we expect those to continue. The field has progressed to the point where applications of AI/ML to build, postprocess, assimilate data, and improve our modeling capabilities have become nearly as ubiquitous as those of more traditional statistics. In fact, many of the practitioners of AI/ML began their foray from statistics. As we have built from that field, we will continue to grow. To avoid more AI winters, we must avoid overpromising results, further develop the XAI methods in a trustworthy manner, foster collaboration across disciplines, include physics practitioners in formulating AI solutions that are consistent with known physics, quantify uncertainty in our solutions, and apply best practices in training to avoid overfitting or nonrepresentative models. The community of practice of AI in ES is growing rapidly and the AMS AI Committee is pleased to welcome this new cadre of users. As AI becomes a common tool, much like statistics, those best practices will become better articulated and disseminated. The AMS AI Committee looks forward to being part of disseminating those best practices and encouraging their use. The promise of AI is here and the applications have begun.

Acknowledgments. SEH and DJG note that their contributions are based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement 1852977. Thanks to Christian Duff for reviewing the majority of the AMS AI abstracts to build Figs. 3 and 4. We thank four anonymous reviewers for their perceptive comments, which helped us to improve the manuscript.

References

AAIA, 2017: A brief history of AI. AITopics, accessed 8 December 2017, https://aitopics.org/misc/brief-history.

Abrahart, R. J., and Coauthors, 2012: Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. *Prog. Phys. Geogr.*, **36**, 480–513, https://doi.org/10.1177/0309133312444943.

Agrawal, S., L. Barrington, C. Bromberg, J. Burge, C. Gazen, and J. Hickey, 2019: Machine learning for precipitation nowcasting from radar images. arXiv, 6 pp., https://arxiv.org/abs/1912.12132.

Albo, D., 1994: Microburst detection using fuzzy logic. Terminal Area Surveillance System Program Project Report, Federal Aviation Administration, 49 pp. [Available from NCAR, P.O. Box 3000, Boulder, CO 80307.]

Bartelt, C., S. Marton, and H. Stuckenschmidt, 2020: xRAI: Explainable representations through AI. arXiv, 8 pp., https://arxiv.org/abs/2012.06006.

Beucler, T., M. Pritchard, S. Rasp, J. Ott, P. Baldi, and P. Gentine, 2021: Enforcing analytic constraints in neural networks emulating physical systems. *Phys. Rev. Lett.*, **126**, 098302, https://doi.org/10.1103/PhysRevLett.126.098302.

Bolton, T., and L. Zanna, 2019: Applications of deep learning to ocean data inference and subgrid parameterization. *J. Adv. Model. Earth Syst.*, **11**, 376–399, https://doi.org/10.1029/2018MS001472.

Boukabara, S.-A., and Coauthors, 2019: Exploring the use of Artificial Intelligence (AI) to optimize the exploitation of big satellite data in NWP and nowcasting. *Ninth Symp. on Advances in Modeling and Analysis Using Python/18th Conf. on Artificial and Computational Intelligence and its Applications to the Environmental Sciences*, Phoenix, AZ, Amer. Meteor. Soc., J4.2, https://ams.confex.com/ams/2019Annual/webprogram/Paper353226.html.

——, and Coauthors, 2021: Outlook for exploiting artificial intelligence in the Earth and environmental sciences. *Bull. Amer. Meteor. Soc.*, **102**, E1016–E1032, https://doi.org/10.1175/BAMS-D-20-0031.1.

Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, https://doi.org/10.1023/A:1010933404324.

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

Burke, A., N. Snook, D. J. Gagne II, S. McCorkle, and A. McGovern, 2020: Calibration of machine learning–based probabilistic hail predictions for operational forecasting. *Wea. Forecasting*, **35**, 149–168, https://doi.org/10.1175/WAF-D-19-0105.1.

Campos, R. M., V. Krasnopolsky, J.-H. Alves, and S. G. Penny, 2020: Improving NCEP's global-scale wave ensemble averages using neural networks. *Ocean Modell.*, **149**, 101617, https://doi.org/10.1016/j.ocemod.2020.101617.

Cannon, A. J., and P. H. Whitfield, 2002: Downscaling recent streamflow conditions in British Columbia, Canada using ensemble neural network models. *J. Hydrol.*, **259**, 136–151, https://doi.org/10.1016/S0022-1694(01)00581-9.

Chapman, W. E., A. C. Subramanian, L. Delle Monache, S. P. Xie, and F. M. Ralph, 2019: Improving atmospheric river forecasts with machine learning. *Geophys. Res. Lett.*, **46**, 10 627–10 635, https://doi.org/10.1029/2019GL083662.

Charalampidis, D., T. Kasparis, and W. Jones, 2002: Removal of nonprecipitation echoes in weather radar using multifractals and intensity. *IEEE Trans. Geosci. Remote Sens.*, **40**, 1121–1131, https://doi.org/10.1109/TGRS.2002.1010899.

Chevallier, F., F. Chéruy, N. A. Scott, and A. Chedin, 1998: A neural network approach for a fast and accurate computation of longwave radiative budget. *J. Appl. Meteor.*, **37**, 1385–1397, https://doi.org/10.1175/1520-0450(1998)037<1385:ANNAFA>2.0.CO;2.

——, J.-J. Morcrette, F. Chéruy, and N. A. Scott, 2000: Use of a neural-network-based longwave radiative transfer scheme in the ECMWF atmospheric model. *Quart. J. Roy. Meteor. Soc.*, **126**, 761–776, https://doi.org/10.1002/qj.49712656318.

Crespo, J. L., and E. Mora, 1993: Drought estimation with neural networks. *Adv. Eng. Software*, **18**, 167–170, https://doi.org/10.1016/0965-9978(93)90064-Z.

Dawson, C. W., and R. L. Wilby, 2001: Hydrological modelling using artificial neural networks. *Prog. Phys. Geogr.*, **25**, 80–108, https://doi.org/10.1177/030913330102500104.

Delanoy, R. L., and S. W. Troxel, 1993: Machine intelligent gust front detection. *Lincoln Lab. J.*, **6**, 187–211.

Delle Monache, L., F. A. Eckel, D. L. Rife, B. Nagarajan, and K. Searight, 2013: Probabilistic weather prediction with an analog ensemble. *Mon. Wea. Rev.*, **141**, 3498–3516, https://doi.org/10.1175/MWR-D-12-00281.1.

Drew, L., 1973: Environmental council. *Phys. Bull.*, **24**, 411, https://doi.org/10.1088/0031-9112/24/7/006.

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

Dyer, R., and W. Moninger, 1988: Summary report on the second workshop on Artificial Intelligence Research in the Environmental Sciences (AIRES). 15–17 September 1987, Boulder, Colorado. *Bull. Amer. Meteor. Soc.*, **69**, 508–514, https://doi.org/10.1175/1520-0477-69.5.508.

Fan, Y., V. Krasnopolsky, H. van den Dool, Ch.-Y. Wu, and J. Gottschalck, 2022: Using artificial neural networks to improve CFS week 3–4 precipitation and 2-meter air temperature forecasts. *Wea. Forecasting*, https://doi.org/10.1175/WAF-D-20-0014.1, in press.

Gagne, D. J., II, A. McGovern, S. E. Haupt, R. A. Sobash, J. K. Williams, and M. Xue, 2017: Storm-based probabilistic hail forecasting with machine learning applied to convection-allowing ensembles. *Wea. Forecasting*, **32**, 1819–1840, https://doi.org/10.1175/WAF-D-17-0010.1.

——, 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.

Gerding, S., and B. Myers, 2003: Adaptive data fusion of meteorological forecast modules. *3rd Conf. on Artificial Intelligence Applications*, Long Beach, CA, Amer. Meteor. Soc., 4.8, https://ams.confex.com/ams/annual2003/techprogram/paper_55795.htm.

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211, https://doi.org/10.1175/1520-0450(1972)011%3C1203:TUOMOS%3E2.0.CO;2.

Goldstein, E. B., and G. Coco, 2015: Machine learning components in deterministic models: Hybrid synergy in the age of data. *Front. Environ. Sci.*, **3**, 33, https://doi.org/10.3389/fenvs.2015.00033.

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.

Haupt, R. L., and S. E. Haupt, 2004: *Practical Genetic Algorithms*. 2nd ed. John Wiley and Sons, 255 pp.

Haupt, S. E., A. Pasini, and C. Marzban, Eds., 2009: *Artificial Intelligence Methods in the Environmental Sciences*. Springer, 424 pp.

——, A. J. Annunzio, and K. J. Schmehl, 2013: Evolving turbulence realizations of atmospheric flow. *Bound.-Layer Meteor.*, **149**, 197–217, https://doi.org/10.1007/s10546-013-9845-7.

——, J. Copeland, W. Y. Y. Cheng, C. Amman, Y. Zhang, and P. Sullivan, 2016: Quantifying the wind and solar power resource and their inter-annual variability over the united states under current and projected future climate. *J. Appl. Meteor. Climatol.*, **55**, 345–363, https://doi.org/10.1175/JAMC-D-15-0011.1.

——, W. Chapman, S. V. Adams, C. Kirkwood, J. S. Hosking, N. H. Robinson, S. Lerch, and A. C. Subramanian, 2020a: Towards implementing AI post-processing in weather and climate: Proposed actions from the Oxford 2019 workshop. *Philos. Trans. Roy. Meteor. Soc.*, **A379**, 20200091, https://doi.org/10.1098/rsta.2020.0091.

——, and Coauthors, 2020b: Combining artificial intelligence with physics-based methods for probabilistic renewable energy forecasting. *Energies*, **13**, 1979, https://doi.org/10.3390/en13081979.

Hsieh, W. W., 2001: Nonlinear principal component analysis by neural networks. *Tellus*, **53A**, 599–615, https://doi.org/10.3402/tellusa.v53i5.12230.

——, 2009: *Machine Learning Methods in the Environmental Sciences*. Cambridge University Press, 349 pp.

Hubbert, J., M. Dixon, and C. Kessinger, 2007: Real time clutter identification and mitigation for NEXRAD. *23rd Int. Conf. on IIPS*, San Antonio, TX, Amer. Meteor. Soc., 5B.6, https://ams.confex.com/ams/87ANNUAL/techprogram/paper_117270.htm.

Johnson, J., P. MacKeen, A. Witt, E. Mitchell, G. Stumpf, M. Eilts, and K. Thomas, 1998: The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm. *Wea. Forecasting*, **13**, 263–276, https://doi.org/10.1175/1520-0434(1998)013<0263:TSCIAT>2.0.CO;2.

Karunanithi, N., W. J. Grenney, D. Whitley, and K. Bovee, 1994: Neural networks for river flow prediction. *J. Comput. Civ. Eng.*, **8**, 201–220, https://doi.org/10.1061/(ASCE)0887-3801(1994)8:2(201).

Kessinger, C., S. Ellis, and J. Van Andel, 2003: The radar echo classifier: A fuzzy logic algorithm for the WSR-88D. *3rd Conf. on Artificial Applications to the Environmental Sciences*, Long Beach, CA, Amer. Meteor. Soc., P1.6, https://ams.confex.com/ams/annual2003/techprogram/paper_54946.htm.

Koval, J. P., and Coauthors, 2015: 1-15 day weather forecast guidance at the weather company. *27th Conf. on Weather Analysis and Forecasting*, Chicago, IL, Amer. Meteor. Soc., 12B.8, https://ams.confex.com/ams/27WAF23NWP/webprogram/Paper273550.html.

Krasnopolsky, V. M., 2013: *The Application of Neural Networks in the Earth System Sciences*. Springer, 189 pp.

——, and M. S. Fox-Rabinovitz, 2006: Complex hybrid models combining deterministic and machine learning components for numerical climate modeling and weather prediction. *Neural Networks*, **19**, 122–134, https://doi.org/10.1016/j.neunet.2006.01.002.

——, and Y. Lin, 2012: A neural network nonlinear multimodel ensemble to improve precipitation forecasts over continental US. *Adv. Meteor.*, **2012**, 649450, https://doi.org/10.1155/2012/649450.

——, D. Chalikov, and H. L. Tolman, 2002: A neural network technique to improve computational efficiency of numerical oceanic models. *Ocean Modell.*, **4**, 363–383, https://doi.org/10.1016/S1463-5003(02)00010-0.

——, M. S. Fox-Rabinovitz, and D. V. Chalikov, 2005: New approach to calculation of atmospheric model physics: Accurate and fast neural network emulation of long wave radiation in a climate model. *Mon. Wea. Rev.*, **133**, 1370–1383, https://doi.org/10.1175/MWR2923.1.

——, ——, and A. A. Belochitski, 2013: Using ensemble of neural networks to learn stochastic convection parameterization for climate and numerical weather prediction models from data simulated by cloud resolving model. *Adv. Artif. Neural Syst.*, **2013**, 485913, https://doi.org/10.1155/2013/485913.

Kumler-Bonfanti, C., J. Stewart, D. Hall, and M. Govett, 2020: Tropical and extratropical cyclone detection using deep learning. *J. Appl. Meteor. Climatol.*, **59**, 1971–1985, https://doi.org/10.1175/JAMC-D-20-0117.1.

Lagerquist, R., A. McGovern, and T. Smith, 2017: Machine learning for real-time prediction of damaging straight-line convective wind. *Wea. Forecasting*, **32**, 2175–2193, https://doi.org/10.1175/WAF-D-17-0038.1.

——, ——, 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.

Lakshmanan, V., A. Fritz, T. Smith, K. Hondl, and G. J. Stumpf, 2007: An automated technique to quality control radar reflectivity data. *J. Appl. Meteor.*, **46**, 288–305, https://doi.org/10.1175/JAM2460.1.

——, K. L. Elmore, and M. B. Richman, 2010: Reaching scientific consensus through a competition. *Bull. Amer. Meteor. Soc.*, **91**, 1423–1427, https://doi.org/10.1175/2010BAMS2870.1.

——, C. Karstens, J. Krause, and L. Tang, 2014: Quality control of weather radar data using polarimetric variables. *J. Atmos. Oceanic Technol.*, **31**, 1234–1249, https://doi.org/10.1175/JTECH-D-13-00073.1.

——, J. Hickey, C. Gazen, and S. Hoyer, 2019: Nowcasting lightning events using a cloud-based deep learning approach. *24th Conf. on Applied Climatology*, Phoenix, AZ, Amer. Meteor. Soc., 3.5, https://ams.confex.com/ams/2019Annual/webprogram/Paper354046.html.

Lima, A. R., A. J. Cannon, and W. W. Hsieh, 2016: Forecasting daily streamflow using online sequential extreme learning machines. *J. Hydrol.*, **537**, 431–443, https://doi.org/10.1016/j.jhydrol.2016.03.017.

——, W. W. Hsieh, and A. J. Cannon, 2017: Variable complexity online sequential extreme learning machine, with applications to streamflow prediction. *J. Hydrol.*, **555**, 983–994, https://doi.org/10.1016/j.jhydrol.2017.10.037.

Maier, H. R., and G. C. Dandy, 2000: Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. *Environ. Modell. Software*, **15**, 101–124, https://doi.org/10.1016/S1364-8152(99)00007-9.

McCandless, T. C., G. S. Young, S. E. Haupt, and L. M. Hinkelman, 2016: Regime-dependent short-range solar irradiance forecasting. *J. Appl. Meteor. Climatol.*, **55**, 1599–1613, https://doi.org/10.1175/JAMC-D-15-0354.1.

——, S. Dettling, and S. E. Haupt, 2020: Comparison of implicit vs explicit regime identification in machine learning methods for solar irradiance prediction. *Energies*, **13**, 689, https://doi.org/10.3390/en13030689.

——, D. J. Gagne, B. Kosovic, S. E. Haupt, B. Yang, C. Becker, and J. Schreck, 2022: Machine learning for improving surface layer flux estimates. *Bound. Layer Meteor.,* in press.

McGovern, A., D. J. Gagne II, J. Basara, T. M. Hamill, and D. Margolin, 2015: Solar energy prediction: An international contest to initiate interdisciplinary research on compelling meteorological problems. *Bull. Amer. Meteor. Soc.*, **96**, 1388–1395, https://doi.org/10.1175/BAMS-D-14-00006.1.

——, 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, 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.

——, ——, and ——, 2020: Using machine learning and model interpretation and visualization techniques to gain physical insights in atmospheric science. *Eighth Int. Conf. on Learning Representations*, Addis Ababa, Ethiopia, International Conference on Learning Representations, 12 pp., https://ai4earthscience.github.io/iclr-2020-workshop/papers/ai4earth16.pdf.

Mielke, P. W., K. J. Berry, and G. W. Brier, 1981: Application of multi-response permutation procedures for examining seasonal changes in monthly mean sea-level pressure patterns. *Mon. Wea. Rev.*, **109**, 120–126, https://doi.org/10.1175/1520-0493(1981)109<0120:AOMRPP>2.0.CO;2.

Monahan, A. H., 2000: Nonlinear principal component analysis by neural networks: Theory and application to the Lorenz system. *J. Climate*, **13**, 821–835, https://doi.org/10.1175/1520-0442(2000)013<0821:NPCABN>2.0.CO;2.

Moninger, W. R., J. Davis, R. Dyer, R. Kittredge, R. McArthur, A. Murphy, and I. Racer, 1987: Summary of the first conference on Artificial Intelligence Research in Environmental Sciences (AIRIES). *Bull. Amer. Meteor. Soc.*, **68**, 793–800, https://doi.org/10.1175/1520-0477-68.7.783.

——, and Coauthors, 1991: Shootout-89, a comparative evaluation of knowledge-based systems that forecast severe weather. *Bull. Amer. Meteor. Soc.*, **72**, 1339–1354, https://doi.org/10.1175/1520-0477(1991)072<1339:SACEOK>2.0.CO;2.

Mote, P. W., A. F. Hamlet, M. P. Clark, and D. P. Lettenmaier, 2005: Declining mountain snowpack in western North America. *Bull. Amer. Meteor. Soc.*, **86**, 39–50, https://doi.org/10.1175/BAMS-86-1-39.

Myers, W., and S. Linden, 2011: A turbine hub height wind speed consensus forecasting system. *Ninth Conf. on Artificial Intelligence and Its Applications to the Environmental Sciences*, Seattle, WA, Amer. Meteor. Soc., 1.2, https://ams.confex.com/ams/91Annual/webprogram/Paper187355.html.

——, G. Wiener, S. Linden, and S. E. Haupt, 2011: A consensus forecasting approach for improved turbine hub height wind speed predictions. *Proc. WindPower 2011*, Anaheim, CA, American Wind Energy Association, 6 pp., https://opensky.ucar.edu/islandora/object/conference:3296.

Newquist, H. P., 1994: *The Brain Makers: The History of Artificial Intelligence – Genius, Ego, and Greed in the Quest for Machines that Think.* SAMS, 300 pp.

NOAA, 2021: NOAA artificial intelligence: Analytics for next-generation Earth science, strategic plan 2021–2015. National Oceanic and Atmospheric Administration, 16 pp., https://sciencecouncil.noaa.gov/Portals/0/Artificial%

20Intelligence%20Strategic%20Plan_Final%20Signed.pdf?ver=2021-01-19-114254-380.

O'Gorman, P. A., and J. G. Dwyer, 2018: Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change and extreme events. *J. Adv. Model. Earth Syst.*, **10**, 2548–2563, https://doi.org/10.1029/2018MS001351.

Pal, A., S. Mahajan, and M. R. Norman, 2019: Using deep neural networks as cost-effective surrogate models for super-parameterized E3SM radiative transfer. *Geophys. Res. Lett.*, **46**, 6069–6079, https://doi.org/10.1029/2018GL081646.

Park, H., A. Ryzhkov, D. Zrnić, and K. Kim, 2009: The hydrometeor classification algorithm for the polarimetric WSR-88D: Description and application to a MCS. *Wea. Forecasting*, **24**, 730–748, https://doi.org/10.1175/2008WAF2222205.1.

Pasini, A., P. Racca, S. Amendola, G. Cartocci, and C. Cassardo, 2017: Attribution of recent temperature behavior reassessed by a neural-network method. *Sci. Rep.*, **7**, 17681, https://doi.org/10.1038/s41598-017-18011-8.

Perrault, R., and Coauthors, 2019: The AI index 2019 annual report. AI Index Steering Committee, Human-Centered AI Institute, Stanford University, 291 pp., https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf.

Pincus, R., E. J. Mlawer, and J. S. Delamere, 2019: Balancing accuracy, efficiency, and flexibility in radiation calculations for dynamical models. *J. Adv. Model. Earth Syst.*, **11**, 3074–3089, https://doi.org/10.1029/2019MS001621.

Poole, D. L., and A. K. Mackworth, 2017: *Artificial Intelligence: Foundations of Computational Agents.* 2nd ed. Cambridge University Press, 820 pp., http://artint.info/2e/html/ArtInt2e.html.

Rai, R., and C. K. Sahu, 2020: Driven by data or derived through physics? A review of hybrid physics guided machine learning techniques with Cyber-Physical System (CPS) focus. *IEEE Access*, **8**, 71050–71073, https://doi.org/10.1109/ACCESS.2020.2987324.

Rasouli, K., W. W. Hsieh, and A. J. Cannon, 2012: Daily streamflow forecasting by machine learning methods with weather and climate inputs. *J. Hydrol.*, **414–415**, 284–293, https://doi.org/10.1016/j.jhydrol.2011.10.039.

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

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., Eds., Lecture Notes in Computer Science, Vol. 9351, Springer, 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.

Scher, S., 2018: Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning. *Geophys. Res. Lett.*, **45**, 12616–12622, https://doi.org/10.1029/2018GL080704.

Silver, D., and Coauthors, 2016: Mastering the game of Go with deep neural networks and tree search. *Nature*, **529**, 484–489, https://doi.org/10.1038/nature16961.

——, and Coauthors, 2017: Mastering the game of Go without human knowledge. *Nature,* **550**, 354–359, https://doi.org/10.1038/nature24270.

Sims, J., and Coauthors, 2020: NOAA's Artificial Intelligence (AI) strategy. *19th Conf. on Artificial Intelligence for Environmental Science*, Boston, MA, Amer. Meteor. Soc., 10.3, https://ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/364942.

Sit, M., B. Z. Demiray, Z. Xiang, G. J. Ewing, Y. Sermet, and I. Demir, 2020: A comprehensive review of deep learning applications in hydrology and water resources. *Water Sci. Technol.*, **82**, 2635–2670, https://doi.org/10.2166/wst.2020.369.

Smith, C., B. McGuire, T. Huang, and G. Yang, 2006: The history of artificial intelligence. 28 pp., accessed 8 December 2017, https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf.

Snauffer, A. M., W. W. Hsieh, and A. J. Cannon, 2016: Comparison of gridded snow water equivalent products with in situ measurements in British Columbia, Canada. *J. Hydrol.*, **541**, 714–726, https://doi.org/10.1016/j.jhydrol.2016.07.027.

Steiner, S., and J. Smith, 2002: Use of three-dimensional reflectivity structure for automated detection and removal of non-precipitating echoes in radar data. *J. Atmos. Oceanic Technol.*, **19**, 673–686, https://doi.org/10.1175/1520-0426(2002)019<0673:UOTDRS>2.0.CO;2.

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**, 16805–16815, https://doi.org/10.1073/pnas.1918964117.

Tan, M., and Q. V. Le, 2019: EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv, 11 pp., https://arxiv.org/abs/1905.11946.

Tedesco, M., J. Pulliainen, M. Takala, M. Hallikainen, and P. Pampaloni, 2004: Artificial neural network-based techniques for the retrieval of SWE and snow depth from SSM/I data. *Remote Sens. Environ.*, **90**, 76–85, https://doi.org/10.1016/j.rse.2003.12.002.

Timmermann, A., 2006: Forecast combinations. *Handbook of Economic Forecasting*, Vol. 1, G. Elliott, C. W. J. Granger, and A. Timmermann, Eds., Elsevier, 135–196, https://doi.org/10.1016/S1574-0706(05)01004-9.

Van der Merwe, R., T. K. Lee, Z. Lu, S. Frolov, and A. M. Baptista, 2007: Fast neural network surrogates for very high dimensional physics-based models in computational oceanography. *Neural Networks*, **20**, 462–478, https://doi.org/10.1016/j.neunet.2007.04.023.

Veerman, M. A., R. Pincus, R. Stoffer, C. M. van Leeuwen, D. Podareanu, and Ch. C. van Heerwaarden, 2021: Predicting atmospheric optical properties for radiative transfer computations using neural networks. *Philos. Trans. Roy. Soc.*, **A379**, 20200095, https://doi.org/10.1098/rsta.2020.0095.

Walker, D. C., W. R. Moninger, and T. R. Stewart, 1992: Shootout-91: A strategy for Integrating Computer Assistance into the Operational Environment. Preprints, *Fourth AES/CMOS Workshop on Operational Meteorology*, Whistler, BC, Canada, Atmospheric Environment Service (Canada) and Canadian Meteorological and Oceanographic Society, 49–58.

Wang, J., P. Balaprakash, and R. Kotamarthi, 2019: Fast domain-aware neural network emulation of a planetary boundary layer parameterization in a numerical weather forecast model. *Geosci. Model Dev.*, **12**, 4261–4274, https://doi.org/10.5194/gmd-12-4261-2019.

Weyn, J. A., D. R. Durran, and R. Caruana, 2020: Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *J. Adv. Model. Earth Syst.,* **12**, e2020MS002109, https://doi.org/10.1029/2020MS002109.

Wikipedia, 2021: AI winter: The collapse of the LISP machine market. Accessed 9 November 2021, https://en.wikipedia.org/wiki/AI_winter#The_collapse_of_the_LISP_machine_market.

——, 2022: Essence. Accessed 23 February 2022, https://en.wikipedia.org/wiki/Essence.

Willard, J., X. Jia, Sh. Xu, M. Steinbach, and V. Kumar, 2021: Integrating scientific knowledge with machine learning for engineering and environmental systems. arXiv, 35 pp., https://arxiv.org/abs/2003.04919.

Williams, J. K., and P. Neilley, 2020: From decision support to decision services: An expanded role for AI in the weather enterprise. *19th Conf. on Artificial Intelligence for Environmental Science/15th Symp. on Societal Applications: Policy, Research and Practice*, Boston, MA, Amer. Meteor. Soc., J61.1, https://ams.confex.com/ams/2020Annual/webprogram/Paper370186.html.

——, C. Kessinger, J. Abernethy, and S. Ellis, 2009: Fuzzy logic applications. *Artificial Intelligence Methods in the Environmental Sciences*, S. E. Haupt, C. Marzban, and A. Pasini, Eds., Springer, 347–377, https://doi.org/10.1007/978-1-4020-9119-3_17.

Williams, J. K., P. P. Neilley, J. P. Koval, and J. McDonald, 2016: Adaptable regression method for ensemble consensus forecasting. *Proc. AAAI Conf. Artif. Intell.*, **30** (1), 3915–3921, https://ojs.aaai.org/index.php/AAAI/article/view/9913.

——, P. Neilley, J. Belanger, J. Koval, A. Kalmikov, S. Marshall, and J. McDonald, 2018: "Prototypes" from The Weather Company's Probability Forecast Platform: A tool for forecasting uncertainty in weather-dependent outcomes. *29th Conf. on Weather Analysis and Forecasting*, Denver, CO, Amer. Meteor. Soc., 12A.6, https://ams.confex.com/ams/29WAF25NWP/webprogram/Paper345810.html.

Zounemat-Kermani, M., E. Matta, A. Cominola, X. Xia, Q. Zhang, Q. Liang, and R. Hinkelmann, 2020: Neurocomputing in surface water hydrology and hydraulics: A review of two decades retrospective, current status and future prospects. *J. Hydrol.*, **588**, 125085, https://doi.org/10.1016/j.jhydrol.2020.125085.