University of Reading

# *Artificial intelligence for climate prediction of extremes: state of the art, challenges, and future perspectives*

Article

Published Version

Creative Commons: Attribution-Noncommercial 4.0

Open Access

Materia, S. ORCID: https://orcid.org/0000-0001-5635-2847, García, L. P., van Straaten, C., O, S., Mamalakis, A., Cavicchia, L., Coumou, D., de Luca, P., Kretschmer, M. ORCID: https://orcid.org/0000-0002-2756-9526 and Donat, M. (2024) Artificial intelligence for climate prediction of extremes: state of the art, challenges, and future perspectives. WIREs Climate Change, 15 (6). e914. ISSN 1757-7799 doi: 10.1002/wcc.914 Available at https://centaur.reading.ac.uk/118423/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1002/wcc.914

Publisher: Wiley

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

University of Reading

www.reading.ac.uk/centaur

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

Check for updates

Received: 24 September 2023 | Revised: 17 July 2024 | Accepted: 24 July 2024

DOI: 10.1002/wcc.914

SYSTEMATIC REVIEW

WIREs CLIMATE CHANGE WILEY

# Artificial intelligence for climate prediction of extremes: State of the art, challenges, and future perspectives

**Stefano Materia**[1] | **Lluís Palma García**[1] | **Chiem van Straaten**[2] | **Sungmin O**[3] | **Antonios Mamalakis**[4] | **Leone Cavicchia**[5] | **Dim Coumou**[2,6] | **Paolo de Luca**[1] | **Marlene Kretschmer**[7,8] | **Markus Donat**[1,9]

[1]Barcelona Supercomputing Center, Barcelona, Spain

[2]Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[3]Ewha Womans University, Seoul, Republic of Korea

[4]Department of Environmental Sciences, School of Data Science, University of Virginia, Charlottesville, Virginia, USA

[5]Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy

[6]Royal Netherlands Meteorological Institute, Utrecht, The Netherlands

[7]Faculty of Physics and Geosciences, University of Leipzig, Leipzig, Germany

[8]University of Reading, Reading, UK

[9]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

**Correspondence**
Stefano Materia, Barcelona Supercomputing Center, Barcelona, Spain.
Email: stefano.materia@bsc.es

**Funding information**
European Union's Horizon 2020, Grant/Award Numbers: 101033654, 101003469, 101065985; European Union's Horizon Europe, Grant/Award Numbers: 101059659, 101137656; European Space Agency's AI4SCIENCE, Grant/Award Number: 4000137110/22/I-EF; National Research Foundation of Korea, Grant/Award Number: RS-2023-00248706

**Edited by:** Eduardo Zorita, Domain Editor and Daniel Friess, Co-Editor-in-Chief

## Abstract

Extreme events such as heat waves and cold spells, droughts, heavy rain, and storms are particularly challenging to predict accurately due to their rarity and chaotic nature, and because of model limitations. However, recent studies have shown that there might be systemic predictability that is not being leveraged, whose exploitation could meet the need for reliable predictions of aggregated extreme weather measures on timescales from weeks to decades ahead. Recently, numerous studies have been devoted to the use of artificial intelligence (AI) to study predictability and make climate predictions. AI techniques have shown great potential to improve the prediction of extreme events and uncover their links to large-scale and local drivers. Machine and deep learning have been explored to enhance prediction, while causal discovery and explainable AI have been tested to improve our understanding of the processes underlying predictability. Hybrid predictions combining AI, which can reveal unknown spatiotemporal connections from data, with climate models that provide the theoretical foundation and interpretability of the physical world, have shown that improving prediction skills of extremes on climate-relevant timescales is possible. However, numerous challenges persist in various aspects, including data curation, model uncertainty, generalizability, reproducibility of methods, and workflows. This review aims at overviewing achievements and

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2024 The Author(s). *WIREs Climate Change* published by Wiley Periodicals LLC.

challenges in the use of AI techniques to improve the prediction of extremes at the subseasonal to decadal timescale. A few best practices are identified to increase trust in these novel techniques, and future perspectives are envisaged for further scientific development.

This article is categorized under:

Climate Models and Modeling > Knowledge Generation with Models

The Social Status of Climate Change Knowledge > Climate Science and Decision Making

**KEYWORDS**

artificial intelligence, climate extreme events, climate forecasting, hybrid modeling, subseasonal to decadal

## 1 | INTRODUCTION

Weather and climate extremes strongly affect many aspects of our society and the natural environment. Being an intrinsic part of a changing climate, there is extensive evidence that the probability and intensity of extreme events have increased and will continue to do so in a warming world (AghaKouchak et al., 2020; Alexander et al., 2006; Orlowsky & Seneviratne, 2012). Therefore, policy-makers and stakeholders urgently need reliable predictions of occurrence probabilities or other aggregated measures of extreme weather on timescales from days to decades. However, the predictive skill of extreme events remains limited, despite recent advancements in weather and climate prediction systems.

Challenges are multiple:

- The anthropogenic climate forcing has accelerated since the beginning of the 21st century, mainly due to growing global economy and reduced absorbing efficiency of land and ocean $CO_2$ sinks (Canadell et al., 2007). Studies based on an older attribution period frequently underestimate the effect of global warming on the probability of unprecedented recent extremes, reflecting the difference between frequencies during the attribution period and out-of-sample verification period (Diffenbaugh, 2020).
- The physical processes driving the occurrence of extreme events differ among timescales, posing unique research questions, and requiring distinct definitions of the event to understand the underlying mechanisms.
- The number of past extreme events is intrinsically small and many may be overlooked due to scarcity of the observation availability (Seneviratne et al., 2021). Therefore, ensembles of dynamical models are often entrusted with the detection and attribution of their drivers, with possible misinterpretations caused by model limitations.
- Poor representation of key processes and feedback mechanisms in state-of-the-art numerical climate models, combined with uncertainties in the initial state, complicates predictions in a complex, and chaotic system like the atmosphere (Faranda et al., 2017).

To face these challenges, the international scientific community has made important steps in the last two decades. On the one hand, the World Climate Research Programme (WCRP) has launched initiatives on near-term climate prediction within the new-born Earth System Modeling core project (https://www.wcrp-climate.org/esmo-overview), like the Lighthouse activity on Explaining and Predicting Earth System Change (EPESC; Findell et al., 2021; see https://www.wcrp-climate.org/epesc). These initiatives aim at developing numerical experiments for subseasonal-to-interdecadal variability, predictability, and predictions, and at delivering, a quantitative understanding of the changes spanning the Earth system through process-based detection and attribution.

Additionally, significant advancements in Earth observation technologies have enhanced the accuracy and scope of collected data (Board, 2019; Guo et al., 2015; Zhang et al., 2022). The launch of the EU Copernicus program, the world's most ambitious program on Earth Observation, the deployment of new satellite systems and sensors (e.g., MODIS, Sentinel) providing high-resolution images of the Earth's surface, and an increasing collaboration between regional space agencies have boosted available information.

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

This era of "big data" has, in turn, fueled the application of artificial intelligence (AI) in many domains of Earth science (Boukabara et al., 2021; Huntingford et al., 2019; Irrgang et al., 2021; Reichstein et al., 2019; Sun et al., 2022). AI here refers to any methodology, including machine learning (ML) and deep learning (DL), in which machines emulate human intelligence to solve tasks based on available data. AI algorithms can learn nonlinear relationships between input and output or to extract spatial and temporal patterns from massive datasets, without prior knowledge of the underlying Earth system processes and dynamics. This makes AI particularly useful for applications lacking have a complete theory. For instance, AI can explore subtle or hidden linkages among Earth system's variables, to uncover relevant processes not yet implemented in physically based models. Additional benefits include AI's flexibility to employ a wider range of input variables, such as novel remote sensing observations, as opposed to physics-based models that use traditionally assumed correlated input variables. AI can thus help exploit the full potential of big data, leading to new insights into Earth system processes that can inform model development and evaluation.

Progress in AI-based forecasting on weather timescales, that is, less than 10 days, has been remarkable in the last few years. In parallel with the rapid rise of AI, forecasting institutes worldwide and Big Tech companies have seized upon the opportunity to improve weather forecasts, gaining skills comparable to that of state-of-the-art dynamical predictions (Bi et al., 2023; Keisler, 2022; Lam et al., 2022; Pathak et al., 2022), even for unprecedented extreme events (Pasche et al., 2024). Recently, a cascade ML system has surpassed the ECMWF high-resolution forecast on a 15-day lead time (Chen et al., 2023).

**TABLE 1** List of acronyms often used in this article.

| | Acronym | Extended name |
|---|---|---|
| Climate | AMV | Atlantic multidecadal variability |
| | ENSO | El Niño southern oscillation |
| | MJO | Madden Julian oscillation |
| | PDO | Pacific decadal oscillation |
| | QBO | Quasi-Biennial oscillation |
| | S2D | Subseasonal to decadal |
| | S2S | Subseasonal to seasonal |
| | SPI | Standardized precipitation index |
| | SSW | Sudden stratospheric warming |
| AIgorithms | ANN | Artificial neural network |
| | CNN | Convolutional neural network |
| | ELM | Extreme learning machine |
| | GAN | Generative adversarial model |
| | LSTM | Long short-term memory |
| | FS | Feature selection |
| | RF | Random forest |
| | SVR | Support vector regression |
| | XAI | eXplainable artificial intelligence |
| | XGBoost | eXtreme gradient boosting |
| Metrics | BSS | Brier skill score |
| | CSI | Critical score index |
| | MAE | Mean absolute value |
| | MCC | Matthews correlation coefficient |
| | ROC | Relative operating characteristics |
| | RMSE | Root mean square error |
| | RPSS | Ranked probability skill score |

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Compared to the short timescales, progress on the subseasonal to decadal (S2D) timescale has been less striking. A fundamental challenge is the limited amount of independent training data, roughly one or two orders of magnitude smaller than for weather timescales. In fact, weather predictions may target individual extreme events, while climate prediction can only aggregate events over time (Meehl et al., 2021), making the number of past observational samples inevitably smaller for climate predictions. This has hampered the development of long-lead forecasts of extremes like drought and warm spells, which likely have some predictability at the S2D scale. The predictability of the climate system beyond the deterministic timescale may be greater than current CMIP6 climate models suggest (Scaife & Smith, 2018; Smith et al., 2019). In fact, an increasing number of articles has been published since the "S2S reboot" opinion paper (Cohen et al., 2019), that claimed that ML techniques developed in computer science could increase the accuracy of predictions at subseasonal to seasonal (S2S) scale.

So far, the development of AI methods for the prediction of extreme events has been overlooked, despite their critical applications and usefulness in real life (Watson, 2022). A recent review (Salcedo-Sanz et al., 2024) addressed the problem of AI for extreme events in terms of attribution, characterization, and prediction, with detailed information on applicable algorithms. Olivetti and Messori (2024) have recently published a perspective paper on the role of DL in weather predictions.

The present survey explores AI's potential to improve the prediction of extremes at the S2D timescale, and to reveal their links to large-scale and local drivers. By reviewing recent literature on AI applications for climate predictions of extreme events and the prospects brought by the combination of empirical and dynamical methods, it discusses the challenges of the data-driven approach and future perspectives, providing climate scientists with a state-of-the-art framework available for rigorous future applications.

Given the frequent appeal to acronyms for climate modes of variability and AI algorithms, we summarize and define those recurrently used in this article in Table 1.

## 2 | PREDICTION AND PREDICTABILITY OF EXTREME EVENTS AT SEASONAL TO DECADAL TIMESCALES

### 2.1 | Definitions of extreme events and their prediction beyond the weather timescale

An event is generally considered extreme if the value of a variable exceeds (or lies below) a given threshold, which can be defined in different ways to focus on specific aspects of the extremes and meet application needs. For example, one definition counts the number of days where temperature, precipitation, etc. exceeds a relative threshold, such as the daily 90th or 99th percentile over a reference period. This can occur at any time of the year, with different seasonal impacts. It is also possible to count events that exceed an absolute threshold, like 35°C, or 50 mm/h of rain, focusing on specific impacts such as health or flood risk. These definitions are standardized in Climpact (https://climpact-sci.org/indices/), a widely used set of indices for identifying and comparing extreme events.

More complex definitions of extremes are based on Extreme Value Theory (EVT; Coles et al., 2001), which differs from methods considering only values exceeding a certain threshold in aiming at fitting extreme values into statistical distributions. The Generalized Pareto Distribution (GPD) and the Generalized Extreme Value (GEV) distribution are the main families that provide information about the probability density function of the extreme values. Both downsize the original time-series to only select extreme data-points. GPD relies on the Peak Over Threshold approach, selecting points exceeding a high threshold as extremes (e.g., Roth et al., 2014; Yiou et al., 2008), while GEV uses the Block Maxima approach, which divides data into temporal blocks (e.g., months or years) to obtain the maximum of each block (e.g., Ben Alaya et al., 2020; Russell & Huang, 2021). A typical application of EVT is to compute return levels of a meteorological events like high precipitation, wind-speed, and temperatures (e.g., Ban et al., 2020; Parey et al., 2019; Zahid et al., 2017).

Extreme weather events development depends on a favorable initial state, the presence of large-scale drivers, positive feedback, and stochastic processes (Sillmann et al., 2017). What marks the distinction between a climate prediction and a weather prediction of these events is the timescale: while specific weather predictions can be made up to 10–15 days in advance, weather as such becomes deterministically unpredictable beyond this time scale (Lorenz, 1969, 1982). Few days ahead of the occurrence of an extreme weather event, it is possible to make predictions of its occurrence and amplitude, with considerable detail about its location, onset, and duration, if the local and remote processes leading to its generation are properly initialized and well predicted (Domeisen et al., 2023). Deterministic forecasts can

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

be made up to 10 days ahead for specific extremes linked with long-lasting atmospheric patterns (e.g., heatwaves; Fragkoulidis et al., 2018).

Climate predictions target longer forecast times (seasons, years, or decades) and are necessarily probabilistic, predicting tendencies of the climate system rather than individual events (Meehl et al., 2021). Beyond the deterministic predictability limit, forecasts include probability distributions of the intensity and duration of extremes (Domeisen et al., 2022), temporal propensity of their occurrence (Prodhomme et al., 2022), or change in their frequency (Delgado-Torres et al., 2023). These characteristics are those potentially captured by a climate prediction of extremes (Figure 1): therefore, in this review, a climate extreme is defined as a temporal aggregation (from weeks to several years) of specific characteristics of an extreme event, like frequency, intensity, and duration. Deterministic weather forecasts will not be discussed since the scientific questions and the approaches to fulfill them can differ.

Verifying predictions of climate extremes at the S2D scale requires metrics beyond the classical anomaly correlation coefficient or root mean squared error (ACC and RMSE, respectively; Wilks, 2011), which may fail to depict the complexity of the phenomena by only considering the "predictable signal" (i.e., the ensemble mean; Kharin & Zwiers, 2002).

Rarity of sample events introduces sampling uncertainty in verification statistics (Casati et al., 2008). By definition, a climate extreme does not happen frequently, and correctly forecasting that an extreme will not occur is often considered a success. On the other hand, a correct forecast of nonoccurrence is easier to achieve than a forecast of occurrence (Jolliffe & Stephenson, 2012), and this should be accounted for in categorical scores. For instance, the Matthews correlation coefficient (MCC; Matthews, 1975) produces a high score only if the prediction is good in all the four confusion matrix categories, accounting for the imbalance between class sizes. The critical score index (CSI) is independent of the number of correct forecasts of nonevents, therefore it has often been used for this task (Schaefer, 1990). A different but related measure of extreme forecast quality is discrimination, visualized by the area below the relative operating characteristic (ROC; Mason & Graham, 2002), whose associated score indicates the forecast's ability to correctly anticipate the

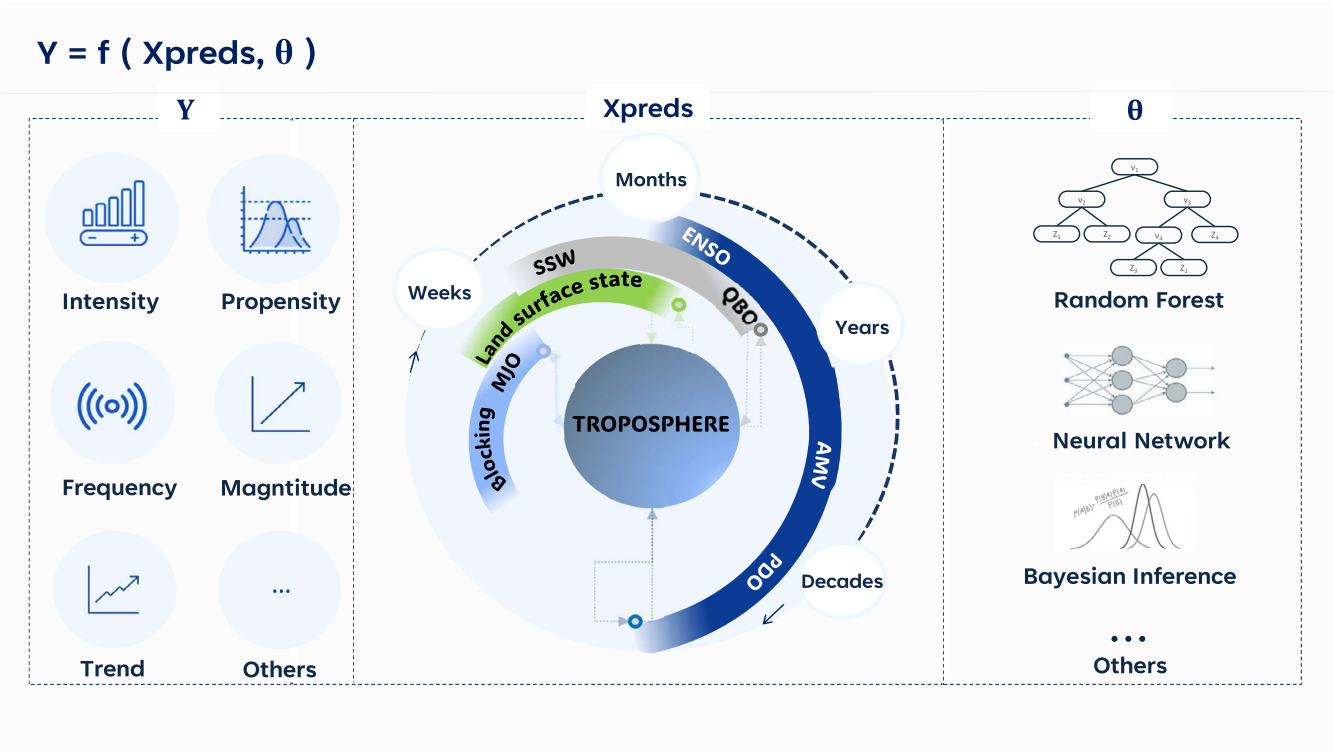


**FIGURE 1** Schematic representation of the AI-based climate prediction function. **Y** represents the *predictand* or *target*, that is a measure of certain aspects characterizing the extremes of interest (e.g., intensity, frequency, etc.). *Xpreds* are the *predictors*, such as modes of variability, or variables describing the state of an Earth System component that affects the troposphere, where the extreme takes place (illustrated as the circle at the center of the figure). These predictors act on the troposphere at different timescales: Sub-days to several days for meteorological drivers (light blue), weeks to ~2 months for land surface drivers (green), weeks to multi-years for stratospheric drivers (gray), months to decades for ocean drivers (dark blue). $\theta$ represents the *parameters* of the AI algorithm used to train the model ($f$).

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

occurrence or non-occurrence of a predefined event. Reliability and resolution are assessed by metrics like the Brier and the ranked probability skill scores (BSS; Brier, 1950; RPSS, Epstein, 1969), that use the entire forecast ensemble to predict the likelihood of an aggregated period (e.g., a season or many years) exceeding a certain extreme threshold (e.g., Delgado-Torres et al., 2023; Torralba et al., 2024). A shortcoming of categorical metrics is that rarity can lead to small or zero counts of extremes, and statistical methods to deal with such sparseness may not be always effective (Agresti, 2002).

The recent adoption of AI for extreme predictions has made common metrics often insufficient for human interpretation of results. Climate scientists generally use ML models as black boxes with no understanding of the model learning process and the justifications behind its decisions, making interpretability an important passage between model verification and the decision based on model's prediction. The interpretability tool mainly used in climate science is SHAP (SHapley Additive exPlanations; Lundberg & Lee, 2017), which assigns each feature of the ML/DL model an importance score, ultimately creating a ranking of importance among features.

## 2.2 | Predictability at different timescales

On (multi-)weekly to decadal timescales, both local and remote physical processes can contribute to the predictability of climate extremes (Figure 1). These mechanisms act seamlessly across timescales, but their relative contribution varies across forecast times and relates to the degree and timescale of interaction between the troposphere (target of this review) and the slower-evolving climate components (Mariotti et al., 2018).

In general, land–atmosphere coupled processes convey predictability over weeks to a few months, primarily modulating the occurrence, duration, and intensity of droughts and heat waves (Materia et al., 2022; Miralles et al., 2019). Stratospheric variability and stratosphere-troposphere coupling provide potential predictability from subseasonal to multi-annual timescales (Scaife et al., 2022 and references therein). For example, sudden stratospheric warmings may influence the tropospheric midlatitude jet stream (Kidston et al., 2015) for many weeks, causing, prolonged cold extremes (Domeisen et al., 2020), while the teleconnection between the QBO and northern hemisphere circulation has shown to affect precipitation amounts in the western Pacific on multi-annual scale (Gray et al., 2018). Ocean–atmosphere coupled mechanisms act on a wide range of timescales, from weeks to multiple years (see Santoso et al., 2017 for extreme ENSO events; Hardiman et al., 2020 for S2S drivers of record warm European winters, Ding et al., 2022 for multi-year predictability of ENSO warm/cold phases), while the effects of slow modes such as the Atlantic Multi-decadal Variability span decades (Zhang et al., 2019).

These variations and their interlinks with the troposphere act as boundary conditions for the atmospheric circulation (Shukla, 1998), contributing to oscillations like MJO (Zhang, 2005; Zhang et al., 2020) or ENSO (Capotondi et al., 2015; Rasmusson & Wallace, 1983), and patterns like blocking and quasi-stationary waves (Reinhold & Pierrehumbert, 1982), able to give the atmosphere persistence characteristics.

These modes of variability and couplings drive the insurgence of climate extremes, with different timescales potentially interfering with each other and affecting the amplitude and characteristics of extremes. For example, a negative PDO phase is associated with prolonged wet phases in the southeastern US winter (Fuentes-Franco et al., 2016). Seasonal damp anomalies are often reduced during El Niño years, while they are reinforced in winters marked by La Niña (Wang et al., 2014). However, during a phase of an active MJO, the extratropical response can amplify or mask the interannual ENSO signal for a few weeks in the southeastern US, potentially resulting in precipitation extremes of the opposite sign than anticipated by the ENSO phase (Arcodia et al., 2020).

Initialization with observed climate conditions allows coupled climate models to capture mechanisms of internally generated climate variations. Large trends in boundary conditions, such as greenhouse gases and aerosols, may also represent a source of predictability (Meehl et al., 2009) increasingly important with longer lead times, with the impact of background warming becoming evident in decadal predictions (Bellucci et al., 2015; Smith et al., 2019). However, the contribution of trends to predictability is also detected at subseasonal and seasonal timescales (Butler et al., 2019; Patterson et al., 2022; Wulff et al., 2022), particularly in summer when year-to-year variability is lower, potentially increasing the skill of extremes forecasts (Prodhomme et al., 2022).

AI has significant potential to enhance the predictive skill of climate extremes across timescales, possibly advancing our scientific understanding and societal preparedness (Huntingford et al., 2019). Sophisticated data-driven approaches offer the opportunity to harness the interconnections of climate drivers and processes, including local and remote interactions, stratosphere–troposphere couplings, and ocean–atmosphere interactions (see Figure 1). AI algorithms can

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

effectively analyze vast climate data and identify complex patterns that influence extreme events nonlinearly, without using potentially biased numerical Earth system models. Moreover, methods of explainable AI and causal discovery algorithms (e.g., Lundberg & Lee, 2017; Section 2.1) can elucidate the relative contribution of various climate drivers, such as PDO, ENSO, and the AMV, and their interaction in modulating climate extremes.

## 3 | AI-BASED PREDICTIONS OF CLIMATE EXTREMES: AN OVERVIEW

We overview the three main families of climate extremes holding predictability at the S2D timescale: extreme temperatures (heat waves and cold spells), droughts, and cyclones (storms and heavy rains). The use of different AI approaches and meteorological datasets, the lack of well-established benchmarks, the specificity in the definition of extremes, and the geographical variability of the applications, make a systematic comparison of the different studies virtually impossible. Many examples of extremes predicted using ML/DL algorithms are discussed in this section and summarized in Table 2, together with an overview of the perspective of combining AI with numerical models in a so-called hybrid approach.

### 3.1 | Extreme temperatures

Hot extremes are becoming more frequent, intense, and longer because of anthropogenic climate change (Dunn et al., 2020), with heatwaves being their most common manifestation (Barriopedro et al., 2023; Thompson et al., 2023). Cold spells are expected to become less frequent, durable, and intense in a warming world, but still pose significant challenges especially in the northern mid-latitudes during boreal winter (Matthias & Kretschmer, 2020; Tomassini et al., 2012).

Classifying temperature extremes can be challenging, since definitions differ according to the specific scientific questions, sectoral application, and timescale. Temperature extremes are often detected as (consecutive) days exceeding a certain threshold (e.g., Perkins & Alexander, 2013; Russo et al., 2014; Sillmann et al., 2013), but approaches based on cumulative metrics, are also used (Russo & Domeisen, 2023). Extreme indices cover a wide range of attributes, such as amplitude, intensity, duration, and frequency (Zhang et al., 2011), but defining extreme events is complicated when linked to climate predictions, since definitions are on one hand related to impacts, on the other hand, limited by S2D predictability. Prodhomme et al. (2022) introduced the concept of *heat wave propensity* to assess the predisposition of a season to heat waves, claiming that a seasonal forecast may more easily predict such a characteristic. Ragone et al. (2018) used a definition that merges temperature and geopotential height anomalies to detect long-lasting heat waves as temporal and spatial averages, improving the statistics of extremely rare events (Ragone & Bouchet, 2021).

AI's ability to forecast various aspects of temperature extremes on S2D timescale is demonstrated in several studies using various methods. Decision-tree-based ensemble methods like RF and XGBoost (He et al., 2021; Kiefer et al., 2023; van Straaten et al., 2022; Weirich-Benet et al., 2023) are often chosen for their robustness against overfitting. He et al. (2021), found that a simpler XGBoost trained on both reanalysis and observational data compares favorably to DL methods, outperforming climatology and least square models based on climate indices and persistence of previous weeks' anomalies. The presence of a well-defined test set and out-of-a-bag cross-validation provides a comprehensive benchmark for the comparison of AI methods for subseasonal forecasts, although stricter definitions of extremes are needed. van Straaten et al. (2022) and Kiefer et al. (2023) used RFs with explainability tools to investigate the influence of atmospheric, oceanic, and land surface states on the occurrence of heat and cold extremes. AI algorithms learn from data relationships that are either well-known or have traditionally not been considered part of the relevant processes and outperform climatology and the background trend. However, the lack of an out-of-sample test set makes results sensitive to possible overfitting. A clear distinction between validation and test is present in Weirich-Benet et al. (2023), whose RF outperforms persistence and climatology up to 6 weeks ahead of a heatwave, competing well with the ECMWF dynamical forecast.

Neural networks (NN) like CNNs, LSTMs, and transformers benefit from directly taking in spatiotemporal information (Jacques-Dumas et al., 2022; Miloshevich et al., 2023). Given the limited occurrence of extreme temperature events in historical data, NNs are often trained on data from numerical climate model simulations (Chattopadhyay et al., 2020). Given the imbalance between extremes and nonextremes, many studies undersampled the training set using only a fraction of the nonheatwave samples. Jacques-Dumas et al. (2022) achieve relevant forecast quality for

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 2** List of publications using AI for the predictions of extremes at the S2D timescale.

| Type of extreme | Reference (timescale) | AI methods | Datasets | Benchmark and metrics |
|---|---|---|---|---|
| Heat extreme | Ham et al., 2019 (Multiseasonal) | CNN | Large ensemble of CMIP5 historical runs and reanalysis<br>Training: >100 model and reanalysis years<br>Validation/test: reanalysis 1976–2017 | Initialized climate model<br>ACC |
| | He et al., 2021 (Subseasonal) | Multitask FS<br>K-nearest neighbor<br>Lasso regression<br>XGBoost<br>CNN-LSTM | Daily observations<br>Training 1986–2016<br>5-day CV over the last 5 years<br>Test: 2017–2018 | Reanalysis climatology<br>Spatial cosine similarity |
| | Jacques-Dumas et al., 2022 (Subseasonal) | CNN, Transfer learning | Multi-daily climate model outputs for the summer months<br>Training: 900 model years<br>Test: 100 model years | No benchmark provided<br>MCC for different levels of heat extremes magnitude |
| | Kiefer et al., 2023 (Subseasonal) | RF + shap XAI;<br>QRF + shap XAI | Daily Nov-Apr temperature observations from 1950 to 2020.<br>Training: every year until 2000 and each year excluding the predicted year from 2001 to 2020 | Reanalysis climatological ensemble<br>CRPSS, BSS |
| | Lopez-Gomez et al., 2023 (Subseasonal) | CNN | All days from 43 years of reanalysis.<br>Training: 34 years<br>Validation: 4 years<br>Test: 5 years | Persistence; initialized model prediction<br>Categorical scores |
| | Miloshevich et al., 2023 (Seasonal) | CNN | Dynamical model of 8000 years<br>Stratified 10-fold cross-validation | Model climatology |
| | Pyrina et al., 2021 | Multilinear regression based on PCA and CCA | Several reanalysis/observational products<br>Training ∼60 years<br>Test ∼30 years | Reanalysis of persistence and climatology<br>ACC |
| | Polkova et al., 2021 (Subseasonal to seasonal) | NN + causality algorithm (Causal Effect Network) | Reanalysis<br>10-day averages over the winter period<br>Training ∼32 years<br>Test ∼8 years | Initialized climate model<br>ACC, ROC |
| | Trenary & DelSole, 2023 (Subseasonal) | Laplacian eigenvectors + Lasso | Dynamical model(s) 3000 years<br>Reanalysis | Climatology, ENSO regression |
| | Van Straaten et al., 2022 (Subseasonal) | RF + shap XAI | Reanalysis<br>Training and test: 1981–2019, 5-fold CV | Reanalysis climatology; Trend<br>BSS |
| | Weirich-Benet et al., 2023 (Subseasonal) | RF + linear model | Weekly observation and reanalysis.<br>Training: 1981–2000<br>Validation: 2001–2009<br>Test: 2010–2018 | Initialized climate model; persistence, climatology<br>Categorical scores, BSS and RMSE |
| Droughts | Danandeh Mehr et al., 2022 (Subseasonal) | ANN, CNN, LSTM, CNN–LSTM | Observation during 45 years (1971–2016), 70/30 training and test | Statistics scores including RMSE, MAE, Nash-Sutcliffe coeff. |
| | Deo and Şahin (2015) | ELM | Monthly historical observations (1958–2008 for training, 2009–2011 for testing) | ANN<br>MAE, RMSE, Coeff of Determination, Willmott Index |

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 2** (Continued)

| Type of extreme | Reference (timescale) | AI methods | Datasets | Benchmark and metrics |
|---|---|---|---|---|
| | Dikshit et al., 2021 (Subseasonal to seasonal) | LSTM | Gridded observations during 120 years (1901–2018) 85/15 training and test | No benchmark specified ROC score, MAE, RMSE, R2 over observations |
| | Dikshit & Pradhan, 2021 (Seasonal) | LSTM + XAI | Observations during 120 years (1901–2018) 85/15 training and test | No benchmark provided RMSE, R2, Nash-Sutcliffe coeff. |
| | Felsche & Ludwig, 2021 (Subseasonal) | ANN + XAI | Regional climate model, 50 ens members Training: 42 years (2150 samples) Test: 6 years | Model generated droughts Simple counting of model drought |
| | Li et al., 2021 (Subseasonal to seasonal) | RF, SVR, ELM | Monthly observations 30 moving training periods of 88 years for 30 sample tests | Comparison of the three ML methods BS and BSS |
| | Mokhtarzad et al., 2017 (Seasonal) | ANN, ANFIS, SVM | ~30 years of meteorological observations (1984–2012) 85/15 training/test | Observed SPI R, RMSE, and cumulative distribution function |
| | Poornima & Pushpalatha, 2019 (Seasonal) | LSTM | Reanalysis daily data Training 55 years Test 1 year | No benchmark provided RMSE, mean absolute error, R2 (unclear observational ground) |
| | Raza et al., 2022 (Seasonal) | ELM, multi-layer perceptron | Monthly observation Training 1951–2013 Test 2014–2016 | Autoregressive moving average RMSE, MAE, Willmott index, Pearson correlation |
| | Rhee & Im, 2017 (Subseasonal to seasonal) | DT, RF, Extremely randomized trees) | Observation, remote sensing data, model forecast data (2003–2015), leave-1 year-out cross-validation | Climatology; Numerical forecasts Accuracy measure for drought category classification |
| | Sahoo et al., 2019 (Subseasonal) | LSTM-RNN, RNN | Observation during 50 years (1971–2020) | Persistence RMSE, Efficiency coefficient, R, MAE |
| | Zhang et al., 2019 (Seasonal) | XGBoost; ANN | Observations during ~60 years (1961–2017) Training/test in a 10-fold cross-validation | Distributed lag nonlinear model Cross-validated R2, RMSE, MAE |
| | Fu et al., 2023 (Seasonal) | CNN; Transfer learning | CESM climate simulations, HighResMIP simulations, and decadal prediction simulations, ERA5 reanalysis ~3000 years for NH ~1500 years for SH | Initialized seasonal forecasts Simple count of the seasonal number of cyclones in comparison to observations |
| Storms and heavy precipitation | Polkova et al., 2021 (Subseasonal to seasonal) | NN + causality algorithm (Causal Effect Network) | Reanalysis 10-day averages over the winter period Training ~32 years Test ~8 years | Initialized climate model BSS, ROC, ACC |
| | Richman et al. (2017) (Seasonal) | SVR | Observations 1984–2015 | Statistical model TCs count, MAE, RMSEt |
| | Scheuerer et al., 2020 (Subseasonal) | ANN/CNN | Weekly precipitation amounts over 20 cool seasons Training: 19*61 samples (LOOCV) Test: 1*61 samples | Empirically corrected initialized forecast RPSS |

(Continues)

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 (Continued)

| Type of extreme | Reference (timescale) | AI methods | Datasets | Benchmark and metrics |
|---|---|---|---|---|
| | Tan et al., 2018 (Seasonal) | RF + Lasso | Reanalysis<br>Training: 1978–2011<br>Test: 2012–2016 | Lasso + RF with no feature selection, MLR<br>MAE, R2 |
| | Vosper et al., 2023 | GAN | IBTrACS + MSWEP | Bilinear interpolation of low-resolution data |
| | Zhang et al., 2023 | RF | NASA's Goddard Earth Observation System model version five (GEOS5) | Observed precipitation on regular grid (PRISM precipitation dataset) |

long-lasting heatwaves, although the lack of a generalizability test makes their result conditional on the model output used for training. Miloshevich et al. (2023) extensively discuss the problem of data scarcity when predicting extremes, and show significant forecast skill for heatwaves in France, identifying in a combination between geopotential and soil moisture as the best predictor. Their results might be even undermined by the lack of ocean drivers, whose initial state is known to impart predictability at monthly and longer timescales. Lopez-Gomez et al. (2023), instead, use all the available information without explicit undersampling, finding that their NN architecture performs better than ECMWF's forecast system for extremely hot days beyond the 2-week forecast time.

Multiple linear regression approaches with tailored filtering procedures also show success (Miller & Wang, 2019; Pyrina et al., 2021; Trenary & DelSole, 2023). These studies are possibly affected by information leakage from training to the verification samples, typical of the standard use of leave-one-out cross-validation (von Storch & Zwiers, 1999), but link to the AI important goal of identifying and disentangling drivers to supplement incomplete theory. Suarez-Gutierrez et al. (2020) uses Multiple Linear Regression to understand dynamical and thermodynamical contributions to European heat extremes: they use moving threshold definition of extremes to account for the increasing trend, finding atmospheric blocking and soil moisture initial conditions crucial for realistic prediction.

## 3.2 | Droughts

Drought has wide-ranging impacts on the environment and society, and its risk is expected to increase in a warmer future climate (Chiang et al., 2021; Wilhite, 2000). Skillful drought predictions with sufficient lead time remain a challenge, due to multiple driving factors across different spatial and temporal scales (Hao et al., 2018) and the complex nature of drought. Various indices based on individual or multiple hydro-climatic factors have been developed to meet the need for applications. These include standardized precipitation index (SPI; McKee et al., 1993) and the Palmer drought severity index (PDSI; Palmer, 1965) for meteorological droughts, the Streamflow Drought Index (SDI; Nalbantis & Tsakiris, 2009) for hydrologic droughts, and the SPI including the effect of evapotranspiration (SPEI; Vicente-Serrano et al., 2010) for agricultural applications.

Droughts are primarily triggered by anomalies in hydrologic and meteorological conditions, including precipitation deficit, high temperature, or evapotranspiration. Land–atmosphere interaction and remote processes such as sea surface temperature variations and short-term atmospheric variability (e.g., MJO) also play an important role (Schubert et al., 2007). Furthermore, long-term droughts are associated with the mechanisms that control hydroclimate at multi-year scales, such as decadal ocean variability, deep soil moisture variability, and the impact of climate change and human activities (Esit et al., 2021; Schubert et al., 2007).

While improved weather and climate predictions using AI can potentially enhance drought condition estimations, the verification of these AI approaches focusing on drought events is still lacking. In general, AI in drought prediction involves predicting drought severity using several input data, including hydrometeorological variables and teleconnection indices (Hao et al., 2018).

ANNs often outperform traditional statistical models, such as regression models and autoregressive moving average models, as shown by Poornima and Pushpalatha (2019) in a comparative study. However, their study's test set is only 1 year, and the work fails to provide details about the observational data chosen for verification, therefore its results are

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

difficult to reproduce. Mokhtarzad et al. (2017) show how complex networks may handle large amounts of nonlinear data-providing SPI and SPEI indices very close to those observed. DL algorithms such as Multi-Layer Perceptrons, LSTM, and CNN have also shown effective performance for drought prediction. Dikshit et al. (2021) uses an LSTM architecture to forecast spatial variations of SPI and SPEI at the subseasonal timescale with promising results, though lacking a proper benchmark. Sahoo et al. (2019) work on a different but related variable, that is the low flow at a gauge station in the Mahanadi River (India). Their LSTM NN, with a very clear separation between training, validation, and test sets, beats the persistence benchmark.

Tree-based models like RF or XGBoost, less prone to overfitting problems, are useful for drought prediction with multiple input predictors. These models effectively process multi-source data such as ground observation, remote sensing data, and climate model output. Rhee and Im (2017) find that tree-based algorithms perform approximately as well as climatology for droughts in South Korea 1–6 months in advance and better than numerical forecasts. Zhang et al. (2019) show that an XGBoost decision tree outperforms a lagged nonlinear model and an ANN for predicting SPEI in central China up to 6 months ahead.

Identifying the most effective ML algorithm for drought prediction is challenging, but the Extreme Learning Machine (ELM) often proves highly skillful. ELM implements a single-hidden layer feedforward neural network, similar to a multi-layer perceptron ANN, and generally provides good generalization performance at a fast-learning speed (Huang et al., 2006). Li et al. (2021) developed three ML models using Random Forest (RF), support vector regression (SVR), and ELM to predict the Standardized Precipitation-Evapotranspiration Index (SPEI) with 1- and 3-month lead times, using previous sea surface temperatures (SST) as predictors. These models are tested on four river basins with different climates and frequent droughts (Colorado, Danube, Orange, and Pearl River), and ELM shows the best prediction skill for all of them. Deo and Şahin (2015) compare predictions of the monthly effective drought index in Australia using ELM and conventional ANN, finding that ELM outperformed ANN in predicting drought duration and severity while also demonstrating faster learning and training speeds. Raza et al. (2022) compare ELM's performance in predicting droughts across several weather stations in Pakistan against another ANN, the multi-layer perceptron, and an autoregressive stochastic model. Results show that ELM produced the best drought predictions for nearly all meteorological stations at different forecast times, while also being the most efficient during training and the fastest in generating forecasts. The authors suggest that ELM should be used as an early warning tool for drought forecasting.

More recent studies have suggested combining AI models to take advantage of different algorithms. For instance, CNN merged with RNNs better captures time series dependence (Danandeh Mehr et al., 2022), outperforming all compared ML models for 1-month-ahead drought forecasting. Extensive examples of AI applications in drought prediction are found in AghaKouchak et al. (2022) or Prodhan et al. (2022). A wide range of AI algorithms have shown progress in predicting drought occurrence and related characteristics. However, most studies are geographically limited to areas with available ground data, making generalization complicated. Recent efforts focus on understanding drivers' contribution to drought prediction by applying explainable AI (Dikshit & Pradhan, 2021; Felsche & Ludwig, 2021).

AI has also demonstrated potential in predicting drought impacts. For instance, Sutanto et al. (2019) show that RF can forecast drought-affected sectors such as agriculture, energy, or wildfire with several months' lead time. AghaKouchak et al. (2022) noted that while current studies mainly focus on prediction accuracy, quantifying uncertainties in AI models is needed. This study also suggests using AI as more than just a drought prediction model, proposing it as a tool to discover unknown drought drivers

## 3.3 | Cyclones and heavy precipitation

Synoptic scale cyclones, both in the tropics and the mid-latitudes, cause considerable economic damage (Mendelsohn et al., 2012) due to associated heavy precipitation, strong winds, and storm surges. Climate change could exacerbate the severity of such extremes, but not necessarily their frequency (Knutson et al., 2020), and predicting their variability on S2D timescales remains challenging (Befort et al., 2022). Heavy precipitation events are not always linked to large-scale weather systems such as cyclones or fronts: many impactful events are instead linked to short-lived, small-scale severe convective events. These extremes are even more challenging for operational climate prediction systems, as their spatial resolution is too coarse to allow the explicit representation of convection. Indeed, numerical climate prediction systems' skill for extreme precipitation substantially decreases beyond a few days in most regions where it has been analyzed (e.g., King et al., 2020; Rivoire et al., 2022).

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Characterizing synoptic-scale cyclones is in principle relatively straightforward, as they are associated with macroscopic features in large-scale atmospheric fields such as mean sea level pressure or atmospheric vorticity. Metrics used to describe such extremes range from simple counting to more sophisticated indices aggregating frequency and intensity (e.g., Emanuel, 2005). In contrast, characterizing small-scale heavy precipitation is more sensitive to the method used to define the extreme, for example, exceedance of an absolute value or of a given climatological percentile (Scoccimarro et al., 2013).

AI techniques have been successfully applied to improve the prediction of both cyclones and heavy precipitation events. Examples of ML/DL methods used to predict tropical cyclone (TC) occurrence include CNNs and RFs. RF provides more interpretability on the role of different drivers, and outperforms nonlinear regression models in predicting TCs in the Western North Pacific (Tan et al., 2018). CNN shows higher prediction skill compared to numerical models, particularly when leveraging transfer learning techniques to exploit larger training sets from modeling data (Fu et al., 2023). On the other hand, increased skill from complex architecture often reduces interpretability. de Burgh-Day and Leeuwenburg (2023) propose ablation studies to overcome DL models' interpretability issues while retaining their good skill.

Richman et al. (2017) used SVR to predict the number of seasonal TCs in the North Atlantic and their spatial distribution, outperforming the statistical forecast in place, even if they lack an out-of-sample test. Regression-based approaches have also shown promising results for subseasonal predictions of precipitation (Zhang et al., 2023). A major weakness for localized precipitation extremes is the absence of uniform benchmarking datasets, which hinders the comparison of different methodologies. The representation of precipitation extremes is less reliable in global reanalyzes, leading to the choice of localized observational reference datasets, which are often limited to single countries or regions (Scheuerer et al., 2020; Zhang et al., 2023). The use of high-resolution, high-frequency global precipitation datasets (e.g., Beck et al., 2017) as a benchmarking standard could improve the robustness of comparisons across studies.

A promising framework for S2D prediction of cyclones is hybrid statistical dynamical predictions. This approach improves the skill of numerical prediction systems in representing weather extremes by finding relationships between large-scale drivers (generally well predicted by dynamical models) and extreme events. It has been applied to precipitation fields, providing probabilistic forecast of precipitation at the subseasonal timescale (Scheuerer et al., 2020). Their ANN/CNN architecture, using predictors from the ECMWF subseasonal forecast, improves forecast of high precipitation accumulation, in week 2, but degrades in weeks 3–4, suggesting the need for additional large-scale predictors.

Several AI applications enhance prediction of climate extremes by improving the way climate forecast model outputs are processed. Polkova et al. (2021) used causal discovery algorithms to understand and improve numerical seasonal forecasts of marine cold air outbreaks, identifying atmospheric circulation patterns and local sea surface temperature as valuable predictors. Applications aimed at improving the representation of wind and precipitation fields exploit DL algorithms trained on high-resolution observations to improve the representation of precipitation (Vosper et al., 2023) or wind patterns (Yang, Zhang, et al., 2022) associated with cyclones, effectively acting as a form of downscaling.

## 3.4 | Hybrid climate prediction of extremes

Hybrid climate predictions, which combine numerical forecasting techniques with AI methods, have emerged as a promising approach for improving the accuracy and reliability of climate predictions. This approach leverages the physical consistency of dynamical models and the flexibility and adaptability of ML/DL models. This section reviews recent advances in hybrid modeling for predicting weather and climate extremes. We follow the typology developed for hydrological forecasting by Slater et al. (2023) and focus on three main areas: (1) a coupled approach where AI improves the parametrization and initialization of climate models, (2) a serial approach where climate model outputs are post-processed or blended with ML, and (3) a statistical–dynamical approach where climate models are used to train a data-driven model.

The first area addresses the major challenge of accurately representing small-scale processes in climate models, such as cloud formation and convective storms, critical for predicting some types of weather and climate extremes. AI can improve the representation of these processes leading to enhanced predictions of extremes. For instance, Rasp et al. (2018) and Gentine et al. (2018) employ NNs to parametrize convection and cloud processes within the atmospheric column of a climate model. The NNs are trained in multi-scale cloud-resolving simulations to emulate fine-scale modeled processes for coarser-scale forecasts (Figure 2a). Steps towards coupling are also being made for land-surface

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

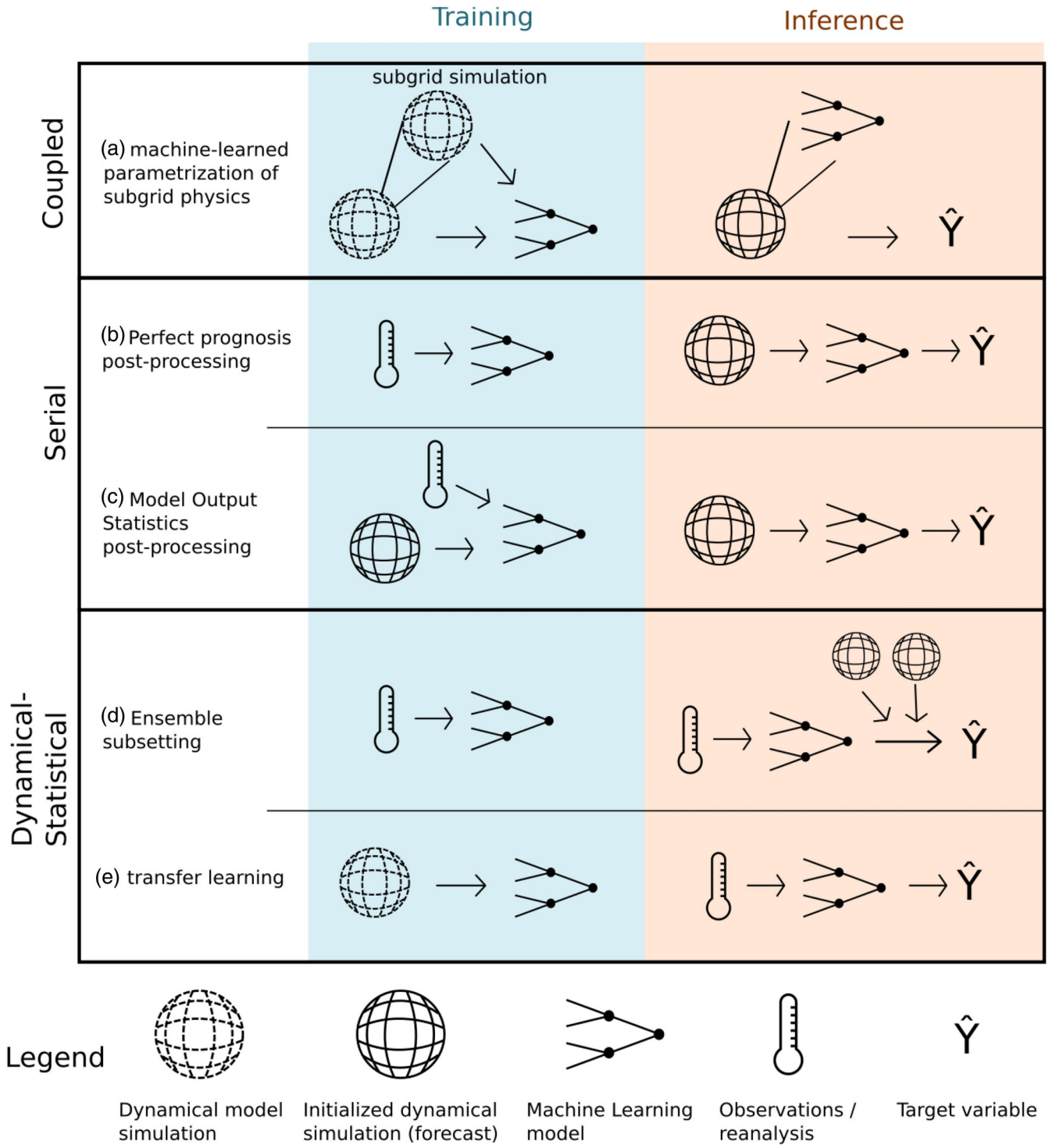


**FIGURE 2** Schematic representation of methods to combine dynamical models with machine or deep learning to create hybrid predictions.

parametrizations (ElGhawi et al., 2023). Stability of the AI algorithm and generalization to unseen situations are a prerequisite for skillful coupled performance, which is why these algorithms are designed to adhere to known physical relations.

The second area of hybrid prediction uses ML techniques to post-process climate model outputs, correct model bias, and/or downscale model outputs. Statistical models learn systematic biases between forecasts and observations, improving the reliability of extremes' predictions by correcting the probability distribution. For example, errors in model-

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

generated heavy precipitation are corrected by learning the precipitation patterns induced by ENSO and applying those to forecasted ENSO (perfect prognosis approach; Doss-Gollin et al., 2018, Strazzo et al., 2019, Figure 2b). Systematic errors between forecasts and observations can also be learned with Model Output Statistics algorithms (Figure 2c), where links between predictors and predictands are those of the models (Glahn & Lowry, 1972). Current ML techniques can learn non-linear relations (Haupt et al., 2021; Schulz & Lerch, 2022; Vannitsem et al., 2021) and thus apply corrections depending on the weather conditions in which an extreme is occurring. On subseasonal time scales, studies have employed fully connected NNs (Fan et al., 2023; van Straaten et al., 2023), convolutional and UNet-type NNs (Horat & Lerch, 2023; Scheuerer et al., 2020), RFs (Zhang et al., 2023), regression models (Hwang et al., 2019), and Bayesian methods (Schepen et al., 2014; Specq & Batté, 2020; Strazzo et al., 2019). These post-processing methods produce probabilistic forecasts of weekly, bi-weekly, or monthly accumulated precipitation or average temperature.

Statistical post-processing models can also evaluate the dynamical models they correct, relating error characteristics to the weather circumstances under which they occur, thus expanding the physical understanding of extremes that would be limited by only using numerical models (Mouatadid et al., 2023; Silva et al., 2022; van Straaten et al., 2023).

The third category of hybrid predictions, statistical-dynamical forecasting (Slater et al., 2023) combines dynamical predictions with different empirical, purely data-driven approaches (Figure 2d). One way is to use statistical methods to provide first-guess prediction of important state variables (e.g., NAO) related to the extreme, and then weigh or select dynamical simulations accordingly (Dobrynin et al., 2018; Neddermann et al., 2019; Polkova et al., 2021). The first-guess prediction can be based on expert-informed regression models (Dobrynin et al., 2018), or using causal discovery algorithms (Polkova et al., 2021).

Training a ML model on climate model simulations, and integrating data from large climate model ensembles, is also possible. Known as transfer learning (Figure 2e; Weiss et al., 2016), this approach expands the size of available training sets (Andersson et al., 2021; Gibson et al., 2021; Ham et al., 2019), offering a more statistically robust version of the climate system and its future trends. Combining multiple models or ensemble realizations can further expand the dataset by some orders of magnitude, allowing better generalization under future climate conditions. Additionally, data-driven models based on causal discovery algorithms can be fitted to dynamical simulations to evaluate the presence of known links in the climate system (di Capua et al., 2022).

## 4 | CHALLENGES IN THE APPLICATION OF ARTIFICIAL INTELLIGENCE APPROACHES

The vehemence with which AI has irrupted in the climate prediction discussion comes with numerous unresolved challenges that must be addressed to build trust in this emerging technology at the service of climate applications. We identify five major areas of challenges (Figure 3) and propose a few best practices that should be acknowledged and carried out in future studies.

### 4.1 | Data and processing

Extreme events are rare by definition, and their increasing likelihood in a non-stationary climate (White et al., 2023) poses important scientific challenges for improving climate predictions using AI. In fact, the scarcity of historical data and the absence of unprecedented events in the past, make any statistical approach based on observations prone to failure (Miloshevich et al., 2023). The physical processes driving climate extremes have time cycles ranging from weeks to years, with different seasons exhibiting varying predictive relations. Additionally, many climate variables are temporally correlated at multiple time scales (He et al., 2021), complicating the acquisition of sufficient (effective) samples to learn from. This problem is more pronounced with longer forecast periods, limiting the verification of time-independent forecasts: multi-annual (5–10 years) predictions, whose training sample relies on current atmospheric reanalysis (e.g., ERA5, Hersbach et al., 2020), may only have a dozen samples for training.

As discussed in Section 3, transfer learning can potentially enlarge the learning sample using climate model data, provided relevant processes are realistically represented and model systematic errors are characterized through process-based studies (Eyring et al., 2019). Thus, the climate model selection becomes a crucial part of the pipeline with significant implications for the ML models' final performance. Recent studies have followed various approaches to ameliorate this issue. Some choose a single model that accurately represents the physical processes behind the targeted tasks

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

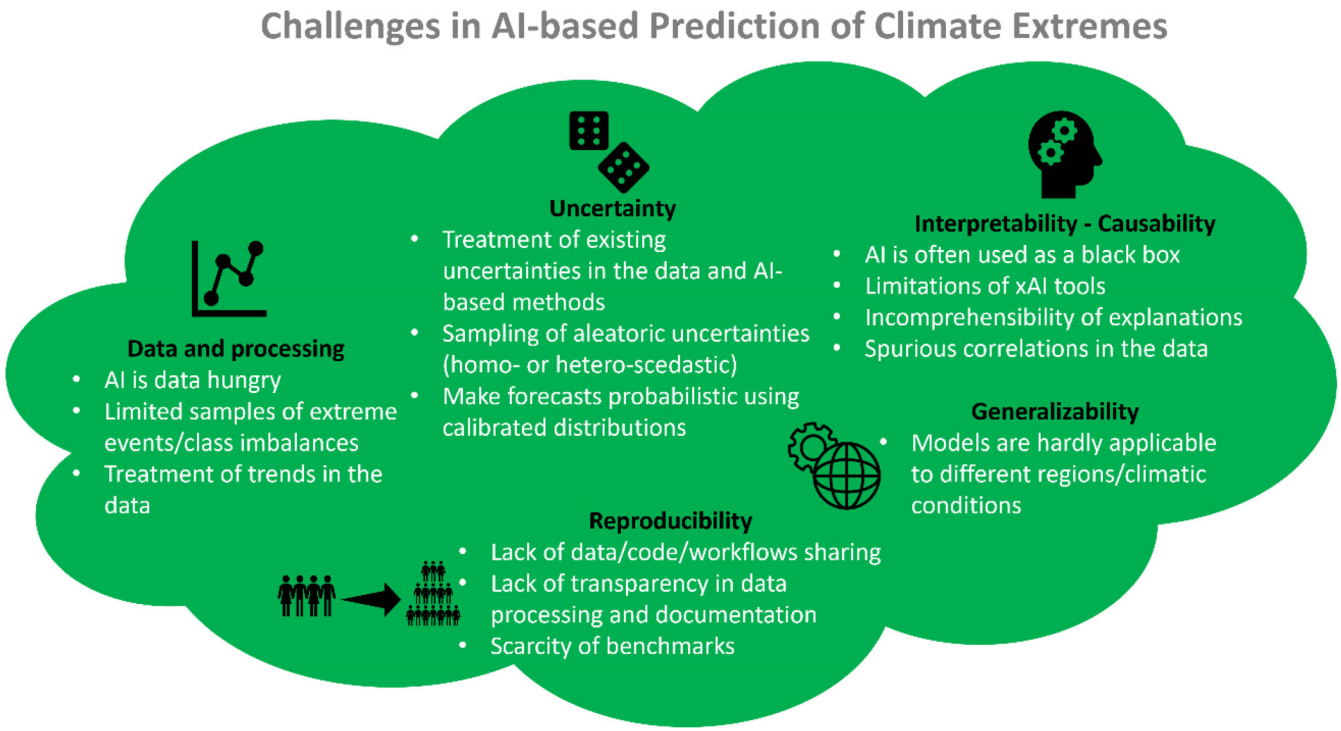


**FIGURE 3** The current challenges facing the prediction of climate extremes with AI. Major areas of difficulty (Data and processing, Uncertainty, Interpretability and Causability, Generalizability, and Reproducibility) are shown in black, and accompanied by a complementary icon, while related tasks discussed in the text (Section 4) are shown in white.

(Gibson et al., 2021; Miloshevich et al., 2023), while others pool many models into the training set to capture robust signals (Ham et al., 2019; Ling et al., 2022; Pan et al., 2022). Furthermore, this procedure can be extended by a second training loop, known as fine-tuning, where the ML model is further trained using available observations (Andersson et al., 2021; Ham et al., 2019; Pan et al., 2022), to account for the biases in the first training.

Despite methods to extend the training dataset, the so-called imbalanced learning problem (He & Ma, 2013) is inherent to extreme forecasting. Being rare events, extremes (events in the distribution's tails) result in a large ratio between extreme and nonextreme samples. Imbalanced learning leads to models that are less confident in predicting extreme states resulting in unskilful predictions if not adequately addressed. Proposed solutions often employ resampling techniques, where either the minority class is oversampled or the majority class gets undersampled either through random sampling of the currently available samples (Miloshevich et al., 2023), or by generating new samples using interpolation [Synthetic Minority Oversampling Technique (SMOTE); Chawla et al., 2002]. However, re-sampling do not add extra information from the few available extra samples, and some studies point out that resampling may result in unreliable probabilities (Fissler et al., 2022).

Several extreme events are clearly affected by the background global warming trend. An increase in heat wave frequency and magnitude (and the concurrent decrease of cold spells) is occurring virtually everywhere in the world, while heavy precipitation, storm intensity, and drought frequency/duration only locally show increments likely linked to a warming climate. This raises questions on how to handle this anthropogenic constraint to separate its induced predictability from the natural predictability of the system. Many studies simply perform an out-of-sample pre-processing by detrending the training time-series before applying the learning algorithm (e.g., Weirich-Benet et al., 2023). While this approach allows to isolate of natural variability, it excludes a significant source of predictability across multiple timescales (Bellucci et al., 2015; Prodhomme et al., 2022; van Straaten et al., 2023). There is no general solution, and the treatment of trends mostly depends on the specific goals and research questions AI-based predictions are designed for. Improving prediction skills for climate services would benefit from including the trend during learning, while removing it makes sense if the aim is to uncover potential drivers for the studied extreme and separate human-made contribution from natural variability (Zeder & Fischer, 2023). This approach requires an additional choice of trend removal methods (Frankcombe et al., 2015), variables to be detrended, and dealing with variables indirectly affected by

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the background trend (e.g., soil moisture). Efforts in this sense have been limited within the ML framework for climate predictions of extremes.

## 4.2 | Uncertainties

Due to the inherent complexity and chaoticity of the climate system, the relationship between predictors and predictands is intrinsically probabilistic, lacking a one-to-one correspondence. Thus, climate predictions of extreme events are subject to significant uncertainties, particularly aleatoric uncertainties that significantly impact the potential predictability of an extreme event (Lucente et al., 2022). Aleatoric uncertainty can be further split into homoscedastic (invariant to different inputs), and heteroscedastic (varying over different inputs), therefore noisier. This result in over-/under-confident predictions, therefore understanding the statistical properties of the extreme and modeling them accordingly is essential. Likewise, uncertainties in data and ML methods (i.e., epistemic uncertainties) also affect prediction. Hence, probabilistic models are best suited to express the combined uncertainties (Miloshevich et al., 2023; van Straaten et al., 2022), providing a distribution of possible future states that better characterize the nature of the prediction problem. Such distributions are most informative to decision-makers when calibrated, ensuring issued probabilities match observed occurrences (Gneiting et al., 2007). Many ML methods produce probabilistic forecasts, but not all explicit output uncertainty distributions (Luo et al., 2022). Thus, reliable probabilistic ML modeling remains a vital research line in climate prediction of extremes.

Regarding probabilistic approaches, one specific method for ANN is the dropout Monte Carlo approach, during training, a small fraction of neurons are randomly "dropped out" (i.e., deactivated) in each iteration making the trained network more robust (Scher & Messori, 2021). During inference (after training is completed), uncertainty can be quantified by sampling different predictions, each time deactivating random neurons. In contrast, methods like variational autoencoders (VAEs; Kingma & Welling, 2019) or diffusion models (Yang, Lee, et al., 2022), learn the data's underlying probability distribution to sample an ensemble of predictions. Other methods applicable to any ML model train an ensemble of models on data subsets of the data or using different random seeds (e.g., Weyn et al., 2021), or retaining multiple "best estimators" to optimize the hyperparameters. In all these cases uncertainty is treated in a post hoc fashion with no guarantee for calibration.

## 4.3 | Interpretability and causability

One of the key pitfalls of using AI for predicting climate extremes is the opacity of the model's decision-making process. With few exceptions (e.g., see linear/logistic regression, decision trees, and newly introduced DL; see Agarwal et al., 2021; Barnes et al., 2022; Chen et al., 2019), most ML/DL algorithms are highly complex and non-transparent, making their predictions difficult to interpret. Although high accuracy might be sufficient and interpretability might be less of an issue for some applications (e.g., machine translation or text generation), for high-stakes climate applications and extreme event prediction interpretability is crucial. Trust in the forecast becomes more essential when verification data is limited verification data, as in the S2S/S2D context. There are examples of AI models predicting correctly for the wrong reasons (known as "clever Hans" models; see Lapuschkin et al., 2019). Thus, interpretability becomes fundamental to test against "clever Hans" models and to ensure that the model has learned relevant processes rather than spurious associations.

To address interpretability, the computer science community has developed tools that can be used to explain predictions of black-box AI models in a post-prediction setting, known as eXplainable AI (XAI; Buhrmester et al., 2019; Tjoa & Guan, 2019; Das & Rad, 2020). XAI has attracted attention in numerous fields, including geosciences (e.g., Barnes et al., 2020; Ebert-Uphoff & Hilburn, 2020; McGovern et al., 2019; Toms et al., 2020), by making black box models more transparent, building trust, fine-tuning poor performing models and providing scientific insights (Mamalakis, Barnes, & Ebert, 2022; McGovern et al., 2019). Given the importance of instilling interpretability XAI methods have been successfully applied to forecast climate extremes as well (see Pegion et al., 2022; Salcedo-Sanz et al., 2024; van Straaten et al., 2022).

Despite XAI potential, challenges remain in climate prediction of extremes and geosciences in general. First, XAI tools are imperfect, and their representation of AI models may depend on the application and the prediction setting (Mamalakis, Ebert-Uphoff, & Barnes, 2022). Some studies have pointed out issues with faithfulness to the ML/DL

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

model, comprehensibility of their results, and reproducibility of the XAI methods (Mamalakis et al., 2022, 2023). Due to these pitfalls, some argue for developing inherently interpretable AI models instead of XAI (Rudin, 2019; Rudin et al., 2022). Even assuming XAI tools are faithful, their insights should only be used to highlight sources of predictability and not to infer causality, as AI models might be using nonphysical relationships (spurious correlations) to make predictions. The limitation in drawing causal conclusions from XAI applications is a significant challenge (Holzinger et al., 2021; Mamalakis, Ebert-Uphoff, & Barnes, 2022; Silva & Keller, 2024). Physics-guided AI (also known as knowledge-guided or physics-informed AI) is a promising approach to impose physical realism in the prediction algorithms and limit spurious correlations during training (Section 5), but this area of research is still in its infancy.

## 4.4 | Generalizability

Generalizability or generalization refers to the model's ability to make accurate predictions beyond the spatio-temporal boundaries of the training datasets. Traditional ML/DL algorithms assume that training and testing (unseen) data are identically distributed and that relationships between inputs and targets learned during training are valid for testing data. However, in climate science applications, this assumption often fails when models predict extreme values lying outside the climatological distribution of the training data. This out-of-distribution generalization issue can considerably degrade model performance, especially under a warming climate that shifts spatial and temporal distributions of Earth system variables, as current relationships may no longer be valid in the future (D'Amour et al., 2020; O'Gorman & Dwyer, 2018; Rasp et al., 2018). Predicting extreme events across diverse climatic regimes also poses a challenge, as ML/DL model accuracy can vary substantially when applied to contrasting climate zones, such as training in humid regions and predicting in arid ones (Meyer & Pebesma, 2021; O et al., 2020). Therefore, understanding model performance on unseen conditions without labeled data is a fundamental challenge for enhancing the robustness of ML/DL models.

Recent studies show that large and diverse training data from various climate regimes are crucial for robust model performance, even when a model is used over limited geographic regions. AI can infer temporal variabilities of extreme events from contemporary spatial variabilities (space-for-time approach; O et al., 2020; Wi & Steinschneider, 2022). Physics-informed ML/DL also shows a promising step forward for enhancing model robustness (Wi & Steinschneider, 2022). However, out-of-distribution generalization has not been sufficiently explored yet in the context of climate extreme predictions.

## 4.5 | Reproducibility

Reproducibility refers to the ability of researchers to independently replicate and verify study results given access to data, methods, and procedures employed. This is a minimal prerequisite for ensuring that findings are reliable and trustworthy, particularly in innovative and highly transformative tasks as for ML applications in climate predictions, where new techniques, algorithms, and workflows are published or at an accelerating rate.

However, more than 60% of earth scientists have failed to reproduce other researchers' work, and over 40% could not reproduce their own experiments (Baker, 2016), raising concerns about a "reproducibility crisis" exacerbated in AI literature (Hutson, 2018). A significant issue is the frequent lack of shared source code, due to reasons such as reluctance of disclosing ongoing work, eagerness for competitive advantage, or discomfort with scripting skills (Gundersen & Kjensmo, 2018). Additionally, datasets used for training are often not made available to the community.

Sharing codes and data is crucial, but detailed documentation of the conducted investigation and experiments is equally important. Comprehensive documentation facilitates independent replication of results, increases trust, reduces the effort, and lowers barriers for others to reproduce the experiments. In drafting this review, several papers were excluded because the description of methodology was poor or inaccurate, making the findings questionable.

The scarcity of common datasets and evaluation metrics hinders the intercomparison of climate extreme prediction studies. Benchmarks can make algorithms quantitatively inter-comparable and foster competition. Well-curated benchmark datasets enable collaboration between researchers with different expertise, like climate and computer scientists (Rasp et al., 2020). However, designing standardized datasets is complicated due to the high-dimensional and multifaceted nature of climate problems (Dueben et al., 2022). Efforts to build benchmark datasets are ongoing for weather-scale forecasts where the atmosphere is still deterministic (WeatherBENCH; Rasp et al., 2020, 2023), or multi-decadal

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**BOX 1 Best practices to improve trust in AI-based forecast of extremes.**

While artificial intelligence provides tools to target potential windows of predictability and eventually improve prediction skill of extreme events, this skill by itself is insufficient. Trust is essential for any early action, which is the ultimate goal of forecasts. Currently, AI-based forecasting generally suffers from a lack of trust for multiple reasons: (1) the exact data processing is often nontransparent and nonreproducible, (2) there are many technical pitfalls that have resulted in exaggerated claims on ML-based skill, and (3) methods are often used as black-boxes with the sources of predictability unexplained.

To overcome this lack of trust, we recommend the following "good practices":

1. Data, workflows, and analyses should be transparent and easily reproducible across different big-volume datasets. This can technically be achieved by linking open-source software to big climate data platforms, then studies should provide access to the source code, the actual AI model (via appropriate repository, e.g., Github), and exact data used, including preprocessing and postprocessing (on publicly accessible data platforms, e.g., Climate Data Store).
2. Studies should use standardized benchmark datasets and multiple skill-metrics. The use of single and/or uncritical skill metrics (e.g., correlation or area under the ROC curve) can easily lead to inflated skill estimates.
3. Validation should be described step by step, and preferably multiple cross-validation approaches should be tested (Sweet et al., 2023), being aware of the possibility of information leakage from train to test data (Risbey et al., 2021). Ideally, all pre-processing (deseasonalizing, standardizing, etc.) is performed out of sample, though in practice this can be challenging due to lack of independent data samples.
4. Proper and suitable quantification of uncertainties should be prioritized in order to minimize epistemic uncertainties and sample all possible aleatoric uncertainties.
5. An effort in understanding sources of predictability and underlying physical mechanisms is required. Interpreting machine learning models should be a top priority, with interpretability focusing on causality instead of association. Explainable AI can provide insights into the sources of predictability, but a commitment towards interpretable models is highly encouraged (Rudin, 2019).

timescale where the climate response is largely driven by socioeconomic scenarios (ClimateBENCH; Watson-Parris et al., 2022). However, no such benchmarks exist yet for climate predictions, which involve a wide range of process timescales and intricate interactions within and across scales (Box 1).

## 5 | FUTURE PERSPECTIVES

Artificial intelligence has proven powerful to target potential windows of predictability and improve forecast skill at S2D time scale. In the last 5 years, the presence of ML/DL algorithms has exponentially grown in studies targeting the prediction of extreme events weeks and seasons ahead. Evolution and progress in this topic have been extraordinarily quick, and we expect even faster-growing development (Figure 4).

For predictions on inter-annual to decadal time scales, the observational record of rare events provides relatively few independent samples, which impedes robust training and therefore the proliferation of studies on extremes at such a timescale. To be applicable for real-world climate predictions, sufficient useful information must be learned from numerical model simulations (Section 4.1). Applications of such transfer learning implementations include IOD, sea ice, and precipitation seasonal forecasts (Andersson et al., 2021; Gibson et al., 2021; Ling et al., 2022), medium-range weather forecasts (Rasp & Thuerey, 2021), or reconstruction of climate observations (Kadow et al., 2020). Increasing the training sample size is promising for climate predictions of extremes (Miloshevich et al., 2023), especially very rare ones for which only a vast ensemble may cover enough samples.

Recent methods constraining or sub-selecting simulations from large climate model ensembles improved the skill of decadal and multi-decadal climate projections by aligning the phases of internal variability modes with the observed

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

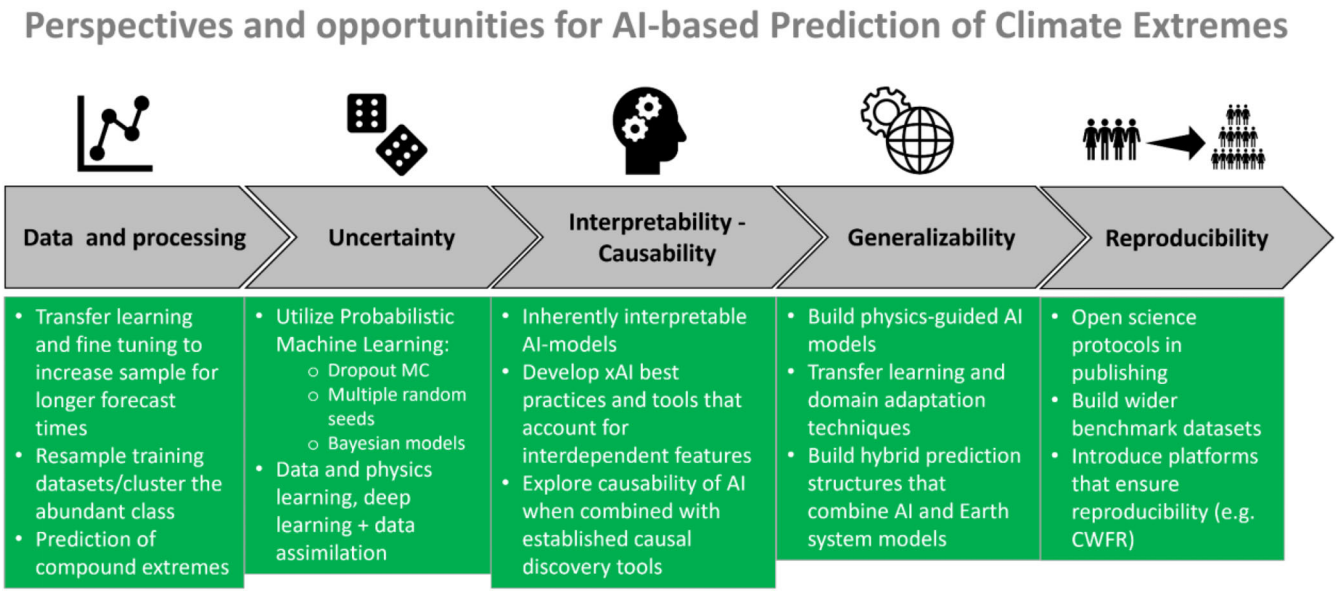


**FIGURE 4** Perspectives and opportunities in AI-based prediction of extremes.

climate (De Luca et al., 2023; Mahmood et al., 2022). These constraints involve numerous choices, implying sensitivities of the results to specific prediction targets both in space and time. We suggest that ML/DL can be useful to further optimize these methods, for example, by learning the most effective constraining criteria or identifying optimal analogs leading to the genesis of extreme events, to select those simulations providing the highest skill at a specific region and time.

Whether data are representative and comprehensive enough for the ML/DL model to be generalizable and finding the boundary between physical-knowledge and data-learning are major challenges (Balaji, 2021). Discarding the underlying structure of equations seems impossible without incurring in several issues. As learning is only as good as the training data, the resulting NN may not generalize well or violate some basic physics such as the conservation laws. In hydrological modeling, these issues have been addressed by employing an end-to-end hybrid modeling approach based on a ML/DL algorithm constrained by energy (Zhao et al., 2019) or water (Kraft et al., 2022) conservation. Beucler et al. (2021) successfully emulated convective processes using NN while enforcing conservation laws. So-called physics-constrained ML thrived at extracting information in observations while maintaining model interpretability and physical consistency. Ideally, it is possible to venture into learning the fundamental physics by learning the underlying equations for well-known systems (Brunton et al., 2016) and the structure of parameterizations from data, with the advantage of intrinsic interpretability (Zanna & Bolton, 2020). In this context, "learning the physics" means solve closed-form equations for unresolved physics on resolved-scale tendencies, using relations in the data. Resolving the turbulent vertical mixing in the atmospheric boundary layer, for example, may be key to fully understand the atmosphere–land process able to modulate heat waves at subseasonal and seasonal scales.

As pointed out in the sidebar, exploiting the potential of those techniques requires addressing issues related to trust in the AI models. Introducing open-source benchmark datasets can enhance the community's confidence by providing a framework to compare different models on common grounds (O et al., 2020; Mamalakis, Ebert-Uphoff, & Barnes, 2022; see also Section 4.5). Yet, similar examples for climate prediction of extreme events are not available. In particular, introducing a platform to ensure reproducibility according to the FAIR (Findable, Accessible, Interoperable, Reusable; Wilkinson et al., 2016) approach toward ML applications in weather and climate appears undelayable. The Canonical Workflow Framework for Research (CWFR) has been proposed to ensure the FAIRness and reproducibility of these practices (Mozaffari et al., 2022), targeting data, algorithms, tools, and workflows.

Most AI-based climate prediction models are developed for deterministic predictions, but providing predictions in a probabilistic framework is beneficial for robust estimation of uncertainties and skill improvement. AI-based models offer a larger spectrum of approaches to predict probabilities, but well-calibrated estimates of uncertainties should be ensured, such as introducing perturbations in the initial conditions or creating model ensembles. Calibration can be better achieved when methods are directly trained to output distributions, and when probabilistic loss is used.

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Common methods range from distributional regression (networks; e.g., Hu, Ghazvinian, et al., 2023; Schulz & Lerch, 2022), to (implicit) quantile networks (e.g., Bremnes, 2020; Dabney et al., 2018), to histogram-estimation networks (e.g., Scheuerer et al., 2020), or Generalized Linear Models (GLMs) and Generalized Additive Models (GAMs; e.g., Tuel & Martius, 2022).

Numerous unexplored probabilistic ML methods might suit the prediction of climate extremes, like Bayesian NNs (Polson & Sokolov, 2017) or nonparametric models like Gaussian Processes (Rasmussen & Williams, 2005), that have yet not received much attention. Generative models, including Variational Auto Encoders (VAEs; Kingma & Welling, 2019), Generative Adversarial Networks (GANs; Goodfellow, 2016), Normalizing Flows models (Papamakarios et al., 2019), and Diffusion Models (Yang, Lee, et al., 2022), have only very rec explored for climate predictions of extremes (Spuler et al., 2024), while for weather extremes these techniques have been employed in several applications (Lam et al., 2022; Lessig et al., 2023; Price et al., 2023; Thuemmel et al., 2023). Conformal prediction (Vovk et al., 2005) is another interesting probabilistic approach focusing on distribution-free uncertainty quantification and reliable probabilities, essential to decision-making applications of climate predictions.

While AI research has examined individual extreme events (Section 2), work on predictions of compound extremes is still in its infancy (Zhang et al., 2020). Examples include flooding caused by co-occurrence of high sea level and precipitation, causing substantial runoff in coastal areas (Bevacqua et al., 2019; Wahl et al., 2015); compound hot-dry events linked to persistent anticyclonic weather systems (Bevacqua et al., 2022; De Luca & Donat, 2023; Yin et al., 2023); and heavy precipitation-high wind speed events during cyclonic weather (De Luca et al., 2020; Martius et al., 2016; Zscheischler et al., 2021).

At the time of writing this article, few studies have investigated AI for compound extreme events prediction, particularly at the S2D scale. Park and Lee (2020) assessed coastal flooding as the compound effect of high tides and heavy rainfall in South Korea, developing a future (2030–2080) compound risk map using ML algorithms like k-nearest neighbor, RF, and support vector machine. Sampurno et al. (2022) used a hydrodynamic model trained with similar ML models, to predict compound flooding over the Kapuas River delta (Indonesia) at the weather timescale. In addition, AI-based techniques help quantify relationships between the extremes of two variables (Zhang et al., 2022). Bayesian networks (Sanuy et al., 2020) and ANN (Huang et al., 2021) have been used to understand compound extremes, while complex networks are found capable to drive causal relationships between two or more variables (Sun et al., 2022). In conclusion, scientific knowledge for AI-based skillful predictions of compound extremes exists, and as co-occurrence and interaction of climate extremes often generate more severe socioeconomic impacts (Zscheischler et al., 2018), implementing this knowledge at the S2D timescale may prove useful for planning climate adaptation strategies.

As the community refines AI technologies, we stand to gain invaluable insights to prepare, mitigate, and adapt to climate extremes with greater precision and foresight. The integration of AI into climate prediction of extremes holds immense potential for building more resilient and sustainable societies in the face of an increasingly variable and changing climate.

## AUTHOR CONTRIBUTIONS

**Stefano Materia:** Conceptualization (lead); data curation (equal); funding acquisition (lead); investigation (equal); visualization (equal); writing – original draft (lead). **Lluís Palma García:** Conceptualization (supporting); data curation (equal); investigation (equal); visualization (equal); writing – original draft (supporting). **Chiem van Straaten:** Conceptualization (supporting); data curation (equal); investigation (equal); visualization (equal); writing – original draft (equal). **Sungmin O:** Conceptualization (equal); data curation (equal); investigation (equal); writing – original draft (equal). **Antonios Mamalakis:** Conceptualization (supporting); data curation (equal); investigation (equal); visualization (equal); writing – original draft (equal). **Leone Cavicchia:** Conceptualization (equal); data curation (equal); investigation (equal); writing – original draft (equal). **Dim Coumou:** Conceptualization (equal); supervision (equal); writing – original draft (supporting). **Paolo de Luca:** Conceptualization (supporting); investigation (equal); writing – original draft (supporting). **Marlene Kretschmer:** Conceptualization (equal); data curation (supporting); investigation (equal); writing – original draft (equal). **Markus Donat:** Conceptualization (equal); investigation (equal); supervision (equal); writing – original draft (supporting).

## ACKNOWLEDGMENTS

The authors thank Antonia Frangeskou for her technical support in drawing Figure 1. The authors thank two anonymous reviewers and the editor, Dr Zorita, whose in-depth contribution helped improve the original version of the manuscript.

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

## FUNDING INFORMATION

Stefano Materia has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement 101033654 (ARTIST). Lluís Palma García and Markus Donat have received funding from the European Space Agency's project AI4Drought under the AI4SCIENCE call contract No. 4000137110/22/I-EF. Stefano Materia and Markus Donat also acknowledge funding from the European Union's Horizon Europe program grant agreement 101137656 (EXPECT). Chiem van Straaten, Dim Coumou, and Marlene Kretschmer acknowledge funding from the European Union's Horizon 2020 program grant agreement 101003469 (XAIDA). Sungmin O acknowledges the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2023-00248706). Leone Cavicchia has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement 101065985 (CYCLOPS) and under the CLINT project (grant agreement 101003876). Paolo de Luca has received funding from the European Union's Horizon Europe research and innovation program under the Marie Skłodowska-Curie grant agreement 101059659.

## CONFLICT OF INTEREST STATEMENT

The authors have declared no conflicts of interest for this article.

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## ORCID

*Stefano Materia* https://orcid.org/0000-0001-5635-2847

## RELATED WIREs ARTICLE

S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts

## FURTHER READING

Adikari, K. E., Shrestha, S., Ratnayake, D. T., Budhathoki, A., Mohanasundaram, S., & Dailey, M. N. (2021). Evaluation of artificial intelligence models for flood and drought forecasting in arid and tropical regions. *Environmental Modelling & Software*, *144*, 105136. https://doi.org/10.1016/j.envsoft.2021.105136

Al Kafy, A., Bakshi, A., Saha, M., Faisal, A. A., Almulhim, A. I., Rahaman, Z. A., & Mohammad, P. (2023). Assessment and prediction of index based agricultural drought vulnerability using machine learning algorithms. *Science of the Total Environment*, *867*, 161394. https://doi.org/10.1016/j.scitotenv.2023.161394

Bolton, T., & Zanna, L. (2019). Applications of deep learning to ocean data inference and subgrid parameterization. *Journal of Advances in Modeling Earth Systems*, *11*, 376–399. https://doi.org/10.1029/2018MS001472

Bose, R., Pintar, A. L., & Simiu, E. (2023). Simulation of Atlantic hurricane tracks and features: A coupled machine learning approach. *Artificial Intelligence for the Earth Systems*, *2*(2), 220060. https://doi.org/10.1175/AIES-D-22-0060.1

Cai, W., Borlace, S., Lengaigne, M., van Rensch, P., Collins, M., Vecchi, G., Timmermann, A., Santoso, A., McPhaden, M. J., Wu, L., England, M. H., Wang, G., Guilyardi, E., & Jin, F. F. (2014). Increasing frequency of extreme El Niño events due to greenhouse warming. *Nature Climate Change*, *4*(2), 111–116.

Fischer, E. M., Beyerle, U., Bloin-Wibe, L., Gessner, C., Humphrey, V., Lehner, F., Pendergrass, A. G., Sippel, S., Zeder, J., & Knutti, R. (2023). Storylines for unprecedented heatwaves based on ensemble boosting. *Nature Communications*, *14*(1), 4643.

Ganguli, P., & Reddy, M. J. (2014). Ensemble prediction of regional droughts using climate inputs and the SVM–copula approach. *Hydrological Processes*, *28*, 4989–5009. https://doi.org/10.1002/hyp.9966

Gordon, E. M., Barnes, E. A., & Hurrell, J. W. (2021). Oceanic harbingers of Pacific decadal oscillation predictability in CESM2 detected by neural networks. *Geophysical Research Letters*, *48*(21), e2021GL095392. https://doi.org/10.1029/2021GL095392

Hu, Y., Chen, L., Wang, Z., & Li, H. (2023). SwinVRNN: A data-driven ensemble forecasting model via learned distribution perturbation. *Journal of Advances in Modeling Earth Systems*, *15*, e2022MS003211. https://doi.org/10.1029/2022MS003211

Jiang, T., Su, X., Zhang, G., Zhang, T., & Wu, H. (2023). Estimating propagation probability from meteorological to ecological droughts using a hybrid machine learning copula method. *Hydrology and Earth System Sciences*, *27*, 559–576. https://doi.org/10.5194/hess-27-559-2023

Mamalakis, A., Ebert-Uphoff, I., & Barnes, E. A. (2022). Explainable artificial intelligence in meteorology and climate science: Model fine-tuning, calibrating trust and learning new science. In A. Holzinger, R. Goebel, R. Fong, T. Moon, K.-R. Müller, & W. Samek (Eds.), *Beyond explainable artificial intelligence. Springer Lecture Notes on Artificial Intelligence (LNAI)*. Springer.

Mukherjee, S., & Mishra, A. K. (2021). Increase in compound drought and heatwaves in a warming world. *Geophysical Research Letters*, *48*, e2020GL090617. https://doi.org/10.1029/2020GL090617

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Nath, S., Kotal, S., Kundu, P. J. M., & Physics, A. (2016). Seasonal prediction of tropical cyclone activity over the north Indian Ocean using three artificial neural networks. *Meteorology and Atmospheric Physics*, *128*, 751–762.

Pirone, D., Cimorelli, L., del Giudice, G., & Pianese, D. (2023). Short-term rainfall forecasting using cumulative precipitation fields from station data: A probabilistic machine learning approach. *Journal of Hydrology*, *617*, 128949.

Qian, Q. F., Jia, X. J., & Lin, H. (2020). Machine learning models for the seasonal forecast of winter surface air temperature in North America. *Earth and Space Science*, *7*(8), e2020EA001140.

Sonnewald, M., & Lguensat, R. (2021). Revealing the impact of global heating on North Atlantic circulation using transparent machine learning. *Journal of Advances in Modeling Earth Systems*, *13*(8), e2021MS002496.

Stevens, A., Willett, R., Mamalakis, A., Foufoula-Georgiou, E., Tejedor, A., Randerson, J. T., Smyth, P., & Wright, S. (2021). Graph-guided regularized regression of Pacific Ocean climate variables to increase predictive skill of southwestern U.S. winter precipitation. *Journal of Climate*, *34*, 737–754. https://doi.org/10.1175/JCLI-D-20-0079.1

## REFERENCES

Agarwal, R., Melnick, L., Frosst, N., Zhang, X., Lengerich, B., Caruana, R., & Hinton, G. (2021). Neural additive models: Interpretable machine learning with neural nets. arXiv Preprint, arXiv:2004.13912.

AghaKouchak, A., Chiang, F., Huning, L. S., Love, C. A., Mallakpour, I., Mazdiyasni, O., Mofthakari, H., Papalexiou, S. M., Ragno, E., & Sadegh, M. (2020). Climate extremes and compound hazards in a warming world. *Annual Review of Earth and Planetary Sciences*, *48*, 519–548.

AghaKouchak, A., Pan, B., Mazdiyasni, O., Sadegh, M., Jiwa, S., Zhang, W., Love, C. A., Madadgar, S., Papalexiou, S. M., Davis, S. J., Hsu, K., & Sorooshian, S. (2022). Status and prospects for drought forecasting: Opportunities in artificial intelligence and hybrid physical–statistical forecasting. *Philosophical Transactions of the Royal Society A*, *380*, 20210288. https://doi.org/10.1098/rsta.2021.0288

Agresti, A. (2002). *Categorical data analysis* (2nd ed., p. 710). Wiley.

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, A., Rupa, K. K., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, B., Burn, J., Aguilar, E., Brunet, M., … Vazquez-Aguirre, J. L. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research: Atmospheres*, *111*(D5), 1–22. https://doi.org/10.1029/2005JD006290

Andersson, T. R., Hosking, J. S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., Law, S., Jones, D. C., Wilkinson, J., Phillips, T., Byrne, J., Tietsche, S., Sarojini, B. B., Blanchard-Wrigglesworth, E., Aksenov, Y., Downie, R., & Shuckburgh, E. (2021). Seasonal Arctic sea ice forecasting with probabilistic deep learning. *Nature Communications*, *12*(1), 5124.

Arcodia, M. C., Kirtman, B. P., & Siqueira, L. S. (2020). How MJO teleconnections and ENSO interference impacts US precipitation. *Journal of Climate*, *33*(11), 4621–4640.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*, 452–454. https://doi.org/10.1038/533452a

Balaji, V. (2021). Climbing down Charney's ladder: Machine learning and the post-Dennard era of computational climate science. *Philosophical Transactions of the Royal Society A*, *379*(2194), 20200085. https://doi.org/10.1098/rsta.2020.0085

Ban, N., Rajczak, J., Schmidli, J., & Schär, C. (2020). Analysis of alpine precipitation extremes using generalized extreme value theory in convection-resolving climate simulations. *Climate Dynamics*, *55*, 61–75. https://doi.org/10.1007/s00382-018-4339-4

Barnes, E. A., Barnes, R. J., Martin, Z. K., & Rader, J. K. (2022). This looks like that there: Interpretable neural networks for image tasks when location matters. *Artificial Intelligence for the Earth Systems*, *1*, e220001. https://doi.org/10.1175/AIES-D-22-0001.1

Barnes, E. A., Toms, B., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C., & Anderson, D. (2020). Indicator patterns of forced changes learned by an artificial neural network. *Journal of Advances in Modeling Earth Systems*, *12*, e2020MS002195. https://doi.org/10.1029/2020MS002195

Barriopedro, D., García–Herrera, R., Ordóñez, C., Miralles, D. G., & Salcedo-Sanz, S. (2023). Heat waves: Physical understanding and scientific challenges. *Reviews of Geophysics*, *61*(2), e2022RG000780.

Beck, H. E., van Dijk, A. I., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., & de Roo, A. (2017). MSWEP: 3-hourly 0.25 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, *21*(1), 589–615.

Befort, D. J., Hodges, K. I., & Weisheimer, A. (2022). Seasonal prediction of tropical cyclones over the North Atlantic and Western North Pacific. *Journal of Climate*, *35*(5), 1385–1397.

Bellucci, A., Haarsma, R., Gualdi, S., Athanasiadis, P. J., Caian, M., Cassou, C., Fernandez, E., Germe, A., Jungclaus, J., Kröger, J., Matei, D., Müller, W., Pohlmann, H., Salas y Melia, D., Sanchez, E., Smith, D., Terray, L., Wyser, K., & Yang, S. (2015). An assessment of a multimodel ensemble of decadal climate predictions. *Climate Dynamics*, *44*, 2787–2806.

Ben Alaya, M. A., Zwiers, F., & Zhang, X. (2020). An evaluation of block-maximum-based estimation of very long return period precipitation extremes with a large ensemble climate simulation. *Journal of Climate*, *33*, 6957–6970. https://doi.org/10.1175/JCLI-D-19-0011.1

Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., & Gentine, P. (2021). Enforcing analytic constraints in neural networks emulating physical systems. *Physical Review Letters*, *126*(9), 098302.

Bevacqua, E., Maraun, D., Vousdoukas, M. I., Voukouvalas, E., Vrac, M., Mentaschi, L., & Widmann, M. (2019). Higher probability of compound flooding from precipitation and storm surge in Europe under anthropogenic climate change. *Science Advances*, *5*, eaaw5531.

Bevacqua, E., Zappa, G., Lehner, F., & Zscheischler, J. (2022). Precipitation trends determine future occurrences of compound hot–dry events. *Nature Climate Change*, *12*, 350–355. https://doi.org/10.1038/s41558-022-01309-5

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, *619*, 533–538. https://doi.org/10.1038/s41586-023-06185-3

Board, S. S., & National Academies of Sciences, Engineering, and Medicine. (2019). *Thriving on our changing planet: A decadal strategy for earth observation from space*. National Academies Press.

Boukabara, S., Krasnopolsky, V., Penny, S. G., Stewart, J. Q., McGovern, A., Hall, D., Ten Hoeve, J. E., Hickey, J., Huang, H.-L. A., Williams, J. K., Ide, K., Tissot, P., Haupt, S. E., Casey, K. S., Oza, N., Geer, A. J., Maddy, E. S., & Hoffman, R. N. (2021). Outlook for exploiting artificial intelligence in the earth and environmental sciences. *Bulletin of the American Meteorological Society*, *102*, E1016–E1032. https://doi.org/10.1175/BAMS-D-20-0031.1

Bremnes, J. B. (2020). Ensemble postprocessing using quantile function regression based on neural networks and Bernstein polynomials. *Monthly Weather Review*, *148*(1), 403–414.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, *78*(1), 1–3.

Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 3932–3937. https://doi.org/10.1073/pnas.1517384113

Buhrmester, V., Münch, D., & Arens, M. (2019). Analysis of explainers of black box deep neural networks for computer vision: A survey. *Machine Learning and Knowledge Extraction*, *3*, 966–989. https://doi.org/10.3390/make3040048

Butler, A., Charlton-Perez, A., Domeisen, D. I., Garfinkel, C., Gerber, E. P., Hitchcock, P., Karpechko, A. Y., Maycock, A. C., Sigmond, M., Simpson, I., & Son, S. W. (2019). Subseasonal predictability and the stratosphere. In A. W. Robertson & F. Vitart (Eds.), *Subseasonal to seasonal prediction. The gap between weather and climate forecasting* (pp. 223–241). Elsevier.

Canadell, J. G., Le Quéré, C., Raupach, M. R., Field, C. B., Buitenhuis, E. T., Ciais, P., Conway, T. J., Gillett, N. P., Houghton, R. A., & Marland, G. (2007b). Contributions to accelerating atmospheric $CO_2$ growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proceedings of the national academy of sciences*, *104*(47), 18866–18870.

Capotondi, A., Wittenberg, A. T., Newman, M., di Lorenzo, E., Yu, J. Y., Braconnot, P., Cole, J., Dewitte, B., Giese, B., Guilyardi, E., Jin, F.-F., Karnauskas, K., Kirtman, B., Lee, T., Schneider, N., Xue, Y., & Yeh, S. W. (2015). Understanding ENSO diversity. *Bulletin of the American Meteorological Society*, *96*(6), 921–938.

Casati, B., Wilson, L. J., Stephenson, D. B., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E. E., Brown, B. G., & Mason, S. (2008b). Forecast verification: Current status and future directions. *Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling*, *15*(1), 3–18.

Chattopadhyay, A., Nabizadeh, E., & Hassanzadeh, P. (2020). Analog forecasting of extreme-causing weather patterns using deep learning. *Journal of Advances in Modeling Earth Systems*, *12*(2), e2019MS001958.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. In arXiv [cs. AI]. http://arxiv.org/abs/1106.1813

Chen, C., Li, O., Tao, C., Barnett, A. J., Su, J., & Rudin, C. (2019). This looks like that: Deep learning for interpretable image recognition, arXiv Preprint, arXiv:1806.10574.

Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., & Li, H. (2023). FuXi: A cascade machine learning forecasting system for 15-day global weather forecast. *NPJ Climate and Atmospheric Science*, *6*, 190. https://doi.org/10.1038/s41612-023-00512-1

Chiang, F., Mazdiyasni, O., & AghaKouchak, A. (2021). A. Evidence of anthropogenic impacts on global drought frequency, duration, and intensity. *Nature Communications*, *12*, 2754. https://doi.org/10.1038/s41467-021-22314-w

Cohen, J., Coumou, D., Hwang, J., Mackey, L., Orenstein, P., Totz, S., & Tziperman, E. (2019). S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts. *WIREs Climate Change*, *10*(2), e00567.

Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001). *An introduction to statistical modeling of extreme values* (Vol. 208, p. 208). Springer.

D'Amour, A., Heller, K., Moldovan, D., Adlam, B., Alipanahi, B., Beutel, A., Chen, C., Deaton, J., Eisenstein, J., Hoffman, M. D., Hormozdiari, F., Houlsby, N., Hou, S., Jerfel, G., Karthikesalingam, A., Lucic, M., Ma, Y., McLean, C., Mincu, D., ... Sculley, D. (2020). Underspecification presents challenges for credibility in modern machine learning. *Journal of Machine Learning Research*, *23*, 1–61.

Dabney, W., Ostrovski, G., Silver, D., & Munos, R. (2018). Implicit quantile networks for distributional reinforcement learning. In *International conference on machine learning* (pp. 1096–1105). PMLR.

Danandeh Mehr, A., Rikhtehgar Ghiasi, A., Yaseen, Z. M., Sorman, A. U., & Abualigah, L. (2022). A novel intelligent deep learning predictive model for meteorological drought forecasting. *Journal of Ambient Intelligence and Humanized Computing*, *14*, 10441–10455. https://doi.org/10.1007/s12652-022-03701-7

Das, A., & Rad, P. (2020). Opportunities and challenges in explainable artificial intelligence (XAI): A survey. arXiv, 2006.11371v2. https://doi.org/10.48550/arXiv.2006.11371

de Burgh-Day, C. O., & Leeuwenburg, T. (2023). Machine learning for numerical weather and climate modelling: A review. *EGUsphere*, *2023*, 1–48.

de Luca, P., Delgado-Torres, C., Mahmood, R., Samsó, M., & Donat, M. (2023). Constraining decadal variability regionally improves near-term projections of hot, cold and dry extremes. *Environmental Research Letters*, *18*(9), 094054. https://doi.org/10.1088/1748-9326/acf389

de Luca, P., Messori, G., Pons, F. M. E., & Faranda, D. (2020). Dynamical systems theory sheds new light on compound climate extremes in Europe and eastern North America. *Journal of the Royal Meteorological Society*, *146*, 1636–1650. https://doi.org/10.1002/qj.3757

de Luca, P., & Donat, M. G. (2023). Projected changes in hot, dry, and compound hot-dry extremes over global land regions. *Geophysical Research Letters*, *50*, e2022GL102493. https://doi.org/10.1029/2022GL102493

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Delgado-Torres, C., Donat, M. G., Soret, A., González-Reviriego, N., Bretonnière, P. A., Ho, A. C., Pérez-Zanón, N., Cabré, M. S., & Doblas-Reyes, F. J. (2023). Multi-annual predictions of the frequency and intensity of daily temperature and precipitation extremes. *Environmental Research Letters*, *18*(3), 034031.

Deo, R. C., & Şahin, M. (2015). Application of the extreme learning machine algorithm for the prediction of monthly effective drought index in eastern Australia. *Atmospheric Research*, *153*, 512–525.

di Capua, G., Coumou, D., van den Hurk, B., Weisheimer, A., Turner, A. G., & Donner, R. V. (2022). Validation of boreal summer tropical-extratropical causal links in seasonal forecasts. *Weather and Climate Dynamics Discussions*, *2022*, 1–40.

Diffenbaugh, N. S. (2020). Verification of extreme event attribution: Using out-of-sample observations to assess changes in probabilities of unprecedented events. *Science Advances*, *6*, eaay2368. https://doi.org/10.1126/sciadv.aay2368

Dikshit, A., Pradhan, B., & Alamri, A. M. (2021). Long lead time drought forecasting using lagged climate variables and a stacked long short-term memory model. *Science of the Total Environment*, *755*, 142638. https://doi.org/10.1016/j.scitotenv.2020.142638

Dikshit, A., & Pradhan, B. (2021). Explainable AI in drought forecasting. *Machine Learning with Applications*, *6*, 100192. https://doi.org/10.1016/j.mlwa.2021.100192

Ding, R., Tseng, Y., di Lorenzo, E., Shi, L., Li, J., Yu, J.-Y., Wang, C., Sun, C., Luo, J.-J., Ha, K.-J., Hu, Z.-Z., & Li, F. (2022). Multi-year El Niño events tied to the North Pacific oscillation. *Nature Communications*, *13*, 3871. https://doi.org/10.1038/s41467-022-31516-9

Dobrynin, M., Domeisen, D. I. V., Müller, W. A., Bell, L., Brune, S., Bunzel, F., Düsterhus, A., Fröhlich, K., Pohlmann, H., & Baehr, J. (2018). Improved teleconnection-based dynamical seasonal predictions of boreal winter. *Geophysical Research Letters*, *45*, 3605–3614. https://doi.org/10.1002/2018GL077209

Domeisen, D. I., Eltahir, E. A., Fischer, E. M., Knutti, R., Perkins-Kirkpatrick, S. E., Schär, C., Seneviratne, S. I., Weisheimer, A., & Wernli, H. (2023). Prediction and projection of heatwaves. *Nature Reviews Earth and Environment*, *4*(1), 36–50.

Domeisen, D. I., White, C. J., Afargan-Gerstman, H., Muñoz, Á. G., Janiga, M. A., Vitart, F., Wulff, C. O., Antoine, S., Ardilouze, C., Batté, L., Bloomfield, H. C., Brayshaw, D. J., Camargo, S. J., Charlton-Pérez, A., Collins, D., Cowan, T., del Mar Chaves, M., Ferranti, L., Gómez, R., ... Tian, D. (2022). Advances in the subseasonal prediction of extreme events: Relevant case studies across the globe. *Bulletin of the American Meteorological Society*, *103*(6), E1473–E1501.

Domeisen, D. I. V., Butler, A. H., Charlton-Perez, A. J., Ayarzaguena, B., Baldwin, M. P., Dunn-Sigouin, E., Furtado, J. C., Garfinkel, C. I., Hitchcock, P., Karpechko, A. Y., Kim, H., Knight, J., Lang, A. L., Lim, E.-P., Marshall, A., Roff, G., Schwartz, C., Simpson, I. R., Son, S.-W., & Taguchi, M. (2020). The role of stratosphere-troposphere coupling in sub-seasonal to seasonal prediction. 2. Predictability arising from stratosphere-troposphere coupling. *Journal of Geophysical Research*, *125*, e2019JD030923. https://doi.org/10.1029/2019JD030923

Doss-Gollin, J., Muñoz, Á. G., Mason, S. J., & Pastén, M. (2018). Heavy rainfall in Paraguay during the 2015/16 austral summer: Causes and subseasonal-to-seasonal predictive skill. *Journal of Climate*, *31*(17), 6669–6685.

Dueben, P. D., Schultz, M. G., Chantry, M., Gagne, D. J., Hall, D. M., & McGovern, A. (2022). Challenges and benchmark datasets for machine learning in the atmospheric sciences: Definition, status, and outlook. *Artificial Intelligence for the Earth Systems*, *1*(3), e210002.

Dunn, R. J., Alexander, L. V., Donat, M. G., Zhang, X., Bador, M., Herold, N., Lippmann, T., Allan, R., Aguilar, E., Barry, A. A., Brunet, M., Caesar, J., Chagnaud, G., Cheng, V., Cinco, T., Durre, I., de Guzman, R., Htay, T. M., Ibadullah, W. M. W., ... Yussof, M. N. B. H. (2020). Development of an updated global land in situ-based data set of temperature and precipitation extremes: HadEX3. *Journal of Geophysical Research: Atmospheres*, *125*(16), e2019JD032263.

Ebert-Uphoff, I., & Hilburn, K. (2020). Evaluation, tuning, and interpretation of neural networks for working with images in meteorological applications. *Bulletin of the American Meteorological Society*, *101*, E2149–E2170. https://doi.org/10.1175/BAMS-D-20-0097.1

ElGhawi, R., Kraft, B., Reimers, C., Reichstein, M., Körner, M., Gentine, P., & Winkler, A. J. (2023). Hybrid modeling of evapotranspiration: Inferring stomatal and aerodynamic resistances using combined physics-based and machine learning. *Environmental Research Letters*, *18*(3), 034039.

Emanuel, K. (2005). Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, *436*(7051), 686–688.

Epstein, E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, *8*, 985–987.

Esit, M., Kumar, S., Pandey, A., Lawrence, D. M., Rangwala, I., & Yeager, S. (2021). Seasonal to multi-year soil moisture drought forecasting. *NPJ Climate and Atmospheric Science*, *4*, 16. https://doi.org/10.1038/s41612-021-00172-z

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier, B. K., Hall, A. D., Hoffman, F. M., Hurtt, G. C., Jahn, A., Jones, C. D., Klein, S. A., Krasting, J. P., Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl, G. A., ... Williamson, M. S. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, *9*(2), 102–110.

Fan, Y., Krasnopolsky, V., van den Dool, H., Wu, C. Y., & Gottschalck, J. (2023). Using artificial neural networks to improve CFS Week-3–4 precipitation and 2-m air temperature forecasts. *Weather and Forecasting*, *38*(5), 637–654.

Faranda, D., Messori, G., Alvarez-Castro, M. C., & Yiou, P. (2017). Dynamical properties and extremes of northern hemisphere climate fields over the past 60 years. *Nonlinear Processes in Geophysics*, *24*, 713–725. https://doi.org/10.5194/npg-24-713-2017

Felsche, E., & Ludwig, R. (2021). Applying machine learning for drought prediction in a perfect model framework using data from a large ensemble of climate simulations. *Natural Hazards and Earth System Sciences*, *21*, 3679–3691. https://doi.org/10.5194/nhess-21-3679-2021

Fissler, T., Lorentzen, C., & Mayer, M. (2022). Model comparison and calibration assessment: User guide for consistent scoring functions in machine learning and actuarial practice. ArXiv [Stat.ML]. https://doi.org/10.48550/ARXIV.2202.12780

Fragkoulidis, G., Wirth, V., Bossmann, P., & Fink, A. H. (2018). Linking Northern Hemisphere temperature extremes to Rossby wave packets. *Quarterly Journal of the Royal Meteorological Society*, *144*(711), 553–566.

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Frankcombe, L. M., England, M. H., Mann, M. E., & Steinman, B. A. (2015). Separating internal variability from the externally forced climate response. *Journal of Climate*, *28*(20), 8184–8202.

Fu, D., Chang, P., & Liu, X. (2023). Using convolutional neural network to emulate seasonal tropical cyclone activity. *Journal of Advances in Modeling Earth Systems*, *15*(10), e2022MS003596.

Fuentes-Franco, R., Giorgi, F., Coppola, E., & Kucharski, F. (2016). The role of ENSO and PDO in variability of winter precipitation over North America from twenty first century CMIP5 projections. *Climate Dynamics*, *46*, 3259–3277.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, *45*(11), 5742–5751.

Gibson, P. B., Chapman, W. E., Altinok, A., Delle Monache, L., DeFlorio, M. J., & Waliser, D. E. (2021). Training machine learning models on climate model output yields skillful interpretable seasonal precipitation forecasts. *Communications Earth & Environment*, *2*(1), 159.

Glahn, H. R., & Lowry, D. A. (1972). The use of model output statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology and Climatology*, *11*(8), 1203–1211.

Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *69*(2), 243–268.

Goodfellow, I. (2016). NIPS 2016 tutorial: Generative adversarial networks. arXiv [cs.LG]. http://arxiv.org/abs/1701.00160

Gray, L. J., Anstey, J. A., Kawatani, Y., Lu, H., Osprey, S., & Schenzinger, V. (2018). Surface impacts of the quasi biennial oscillation. *Atmospheric Chemistry and Physics*, *18*, 8227–8247. https://doi.org/10.5194/acp-18-8227-2018

Gundersen, O. E., & Kjensmo, S. (2018). State of the art: Reproducibility in artificial Intelligence. In *Proceedings of the Thirty-Second Association for the Advancement of Artificial Intelligence (AAAI) Conference*. Palo Alto, CA: Association for the Advancement of Artificial Intelligence (AAAI) Press.

Guo, H. (2017). Big Earth data: A new frontier in Earth and information sciences. *Big Earth Data*, *1*(1-2), 4–20.

Ham, Y. G., Kim, J. H., & Luo, J. J. (2019). Deep learning for multi-year ENSO forecasts. *Nature*, *573*(7775), 568–572.

Hao, Z., Singh, V. P., & Xia, Y. (2018). Seasonal drought prediction: Advances, challenges and future prospects rev. *Geophysics*, *56*, 108–141. https://doi.org/10.1002/2016RG000549

Hardiman, S. C., Dunstone, N. J., Scaife, A. A., Smith, D. M., Knight, J. R., Davies, P., Claus, M., & Greatbatch, R. J. (2020). Predictability of European winter 2019/20: Indian Ocean dipole impacts on the NAO. *Atmospheric Science Letters*, *21*(12), e1005.

Haupt, S. E., Chapman, W., Adams, S. V., Kirkwood, C., Hosking, J. S., Robinson, N. H., Lerch, S., & Subramanian, A. C. (2021). Towards implementing artificial intelligence post-processing in weather and climate: Proposed actions from the Oxford 2019 workshop. *Philosophical Transactions of the Royal Society A*, *379*(2194), 20200091.

He, H., & Ma, Y. (2013). *Imbalanced learning: Foundations, algorithms, and applications*. Wiley-IEEE Press.

He, S., Li, X., DelSole, T., Ravikumar, P., & Banerjee, A. (2021). Subseasonal climate forecasting via machine learning: Challenges, analysis, and advances. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(1), 169–177. https://doi.org/10.1609/aaai.v35i1.16090

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., ... Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*, 1999–2049. doi:10.1002/qj.3803

Holzinger, A., Malle, B., Saranti, A., & Pfeifer, B. (2021). Towards multi-modal causability with graph neural networks enabling information fusion for explainable AI. *Information Fusion*, *71*, 28–37. https://doi.org/10.1016/j.inffus.2021.01.008

Horat, N., & Lerch, S. (2023). Deep learning for post-processing global probabilistic forecasts on subseasonal time scales. arXiv Preprint arXiv:2306.15956.

Hu, W., Ghazvinian, M., Chapman, W. E., Sengupta, A., Ralph, F. M., & Delle Monache, L. (2023). Deep learning forecast uncertainty for precipitation over the Western United States. *Monthly Weather Review*, *151*(6), 1367–1385.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, *70*(1–3), 489–501.

Huang, W. K., Monahan, A. H., & Zwiers, F. W. (2021). Estimating concurrent climate extremes: A conditional approach. *Weather and Climate Extremes*, *33*, 100332. https://doi.org/10.1016/j.wace.2021.100332

Huntingford, C., Jeffers, E. S., Bonsall, M. B., Christensen, H. M., Lees, T., & Yang, H. (2019). Machine learning and artificial intelligence to aid climate change research and preparedness. *Environmental Research Letters*, *14*(12), 124007.

Hutson, M. (2018). Artificial intelligence faces reproducibility crisis. *Science*, *359*, 725–726. https://doi.org/10.1126/science.359.6377.725

Hwang, J., Orenstein, P., Cohen, J., Pfeiffer, K., & Mackey, L. (2019). Improving subseasonal forecasting in the western U.S. with machine learning. In *Proceedings of the 25th ACM SIGKDD International Conference on knowledge discovery & data mining* (pp. 2325–2335). Association for Computing Machinery. https://doi.org/10.1145/3292500.3330674

Irrgang, C., Boers, N., Sonnewald, M., Barnes, E. A., Kadow, C., Staneva, J., & Saynisch-Wagner, J. (2021). Towards neural earth system modelling by integrating artificial intelligence in earth system science. *Nature Machine Intelligence*, *3*, 667–674. https://doi.org/10.1038/s42256-021-00374-3

Jacques-Dumas, V., Ragone, F., Borgnat, P., Abry, P., & Bouchet, F. (2022). Deep learning-based extreme heatwave forecast. *Frontiers in Climate*, *4*, 789641.

Jolliffe, I. T., & Stephenson, D. B. (Eds.). (2012). *Forecast verification: A practitioner's guide in atmospheric science*. John Wiley & Sons.

Kadow, C., Hall, D. M., & Ulbrich, U. (2020). Artificial intelligence reconstructs missing climate information. *Nature Geoscience*, *13*, 408–413. https://doi.org/10.1038/s41561-020-0582-5

Keisler, R. (2022). Forecasting global weather with graph neural networks. *arXiv preprint arXiv:2202.07575*.

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Kharin, V. V., & Zwiers, F. W. (2002). Climate predictions with multimodel ensembles. *Journal of Climate*, *15*(7), 793–799.

Kidston, J., Scaife, A. A., Hardiman, S. C., Mitchell, D. M., Butchart, N., Baldwin, M. P., & Gray, L. J. (2015). Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. *Nature Geoscience*, *8*(6), 433–440.

Kiefer, S. M., Lerch, S., Ludwig, P., & Pinto, J. G. (2023). Can machine learning models be a suitable tool for predicting central European cold winter weather on subseasonal to seasonal timescales? *Artificial Intelligence for the Earth Systems*, *2*(4), e230020. https://doi.org/10.1175/AIES-D-23-0020.1

King, A. D., Hudson, D., Lim, E. P., Marshall, A. G., Hendon, H. H., Lane, T. P., & Alves, O. (2020). Sub-seasonal to seasonal prediction of rainfall extremes in Australia. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 2228–2249.

Kingma, D. P., & Welling, M. (2019). An introduction to variational autoencoders. arXiv [cs.LG]. http://arxiv.org/abs/1906.02691

Knutson, T., Camargo, S. J., Chan, J. C., Emanuel, K., Ho, C. H., Kossin, J., Mohapatra, M., Satoh, M., Sugi, M., Walsh, K., & Wu, L. (2020). Tropical cyclones and climate change assessment: Part II: Projected response to anthropogenic warming. *Bulletin of the American Meteorological Society*, *101*(3), E303–E322.

Kraft, B., Jung, M., Körner, M., Koirala, S., & Reichstein, M. (2022). Towards hybrid modeling of the global hydrological cycle. *Hydrology and Earth System Sciences*, *26*, 1579–1614. https://doi.org/10.5194/hess-26-1579-2022

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., Battaglia, P., & Battaglia, P. (2022). GraphCast: Learning skillful medium-range global weather forecasting. arXiv Preprint arXiv:2212.12794.

Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., & Müller, K. R. (2019). Unmasking Clever Hans predictors and assessing what machines really learn. *Nature communications*, *10*(1), 1096.

Lessig, C., Luise, I., Gong, B., Langguth, M., Stadtler, S., & Schultz, M. (2023). AtmoRep: A stochastic model of atmosphere dynamics using large scale representation learning (arXiv:2308.13280). *arXiv*. https://doi.org/10.48550/arXiv.2308.13280

Li, J., Wang, Z., Wu, X., Xu, C. Y., Guo, S., Chen, X., & Zhang, Z. (2021). Robust meteorological drought prediction using antecedent SST fluctuations and machine learning. *Water Resources Research*, *57*(8), e2020WR029413.

Ling, F., Luo, J. J., Li, Y., Tang, T., Bai, L., Ouyang, W., & Yamagata, T. (2022). Multi-task machine learning improves multi-seasonal prediction of the Indian Ocean dipole. *Nature Communications*, *13*, 7681. https://doi.org/10.1038/s41467-022-35412-0

Lopez-Gomez, I., McGovern, A., Agrawal, S., & Hickey, J. (2023). Global extreme heat forecasting using neural weather models. *Artificial Intelligence for the Earth Systems*, *2*(1), e220035. https://doi.org/10.1175/AIES-D-22-0035.1

Lorenz, E. N. (1969). The predictability of a flow which possesses many scales of motion. *Tellus*, *21*, 289–307.

Lorenz, E. N. (1982). Atmospheric predictability experiments with a large numerical model. *Tellus*, *34*, 505–513. https://doi.org/10.3402/tellusa.v34i6.10836

Lucente, D., Herbert, C., & Bouchet, F. (2022). Committor functions for climate phenomena at the predictability margin: The example of El Niño–southern oscillation in the Jin and Timmermann model. *Journal of the Atmospheric Sciences*, *79*(9), 2387–2400. https://doi.org/10.1175/jas-d-22-0038.1

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4768–4777.

Luo, X., Nadiga, B. T., Park, J. H., Ren, Y., Xu, W., & Yoo, S. (2022). A Bayesian deep learning approach to near-term climate prediction. *Journal of Advances in Modeling Earth Systems*, *14*(10), e2022MS003058.

Mahmood, R., Donat, M. G., Ortega, P., Doblas-Reyes, F. J., Delgado-Torres, C., Samsó, M., & Bretonnière, P.-A. (2022). Constraining low-frequency variability in climate projections to predict climate on decadal to multi-decadal timescales—A poor man's initialized prediction system. *Earth System Dynamics*, *13*, 1437–1450. https://doi.org/10.5194/esd-13-1437-2022

Mamalakis, A., Barnes, E. A., & Ebert-Uphoff, I. (2022). Investigating the fidelity of explainable artificial intelligence methods for applications of convolutional neural networks in geoscience. *Artificial Intelligence for the Earth Systems*, *1*(4), e220012.

Mamalakis, A., Barnes, E. A., & Ebert-Uphoff, I. (2023). Carefully choose the baseline: Lessons learned from applying XAI attribution methods for regression tasks in geoscience. *Artificial Intelligence for the Earth Systems*, *2*(1), e220058.

Mamalakis, A., Ebert-Uphoff, I., & Barnes, E. A. (2022). Neural network attribution methods for problems in geoscience: A novel synthetic benchmark dataset. *Environmental Data Science*, *1*, e8. https://doi.org/10.1017/eds.2022.7

Mariotti, A., Ruti, P. M., & Rixen, M. (2018). Progress in subseasonal to seasonal prediction through a joint weather and climate community effort. *npj Climate and Atmospheric Science*, *1*(1), 4.

Martius, O., Pfahl, S., & Chevalier, C. (2016). A global quantification of compound precipitation and wind extremes. *Geophysical Research Letters*, *43*, 7709–7717. https://doi.org/10.1002/2016GL070017

Mason, S. J., & Graham, N. E. (2002). Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society: A Journal of the Atmospheric Sciences, Applied Meteorology and Physical Oceanography*, *128*(584), 2145–2166.

Materia, S., Ardilouze, C., Prodhomme, C., Donat, M. G., Benassi, M., Doblas-Reyes, F. J., Peano, D., Caron, L. P., Ruggieri, P., & Gualdi, S. (2022). Summer temperature response to extreme soil water conditions in the Mediterranean transitional climate regime. *Climate Dynamics*, *58*(7–8), 1943–1963.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta (BBA)-Protein. *Structure*, *405*(2), 442–451.

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Matthias, V., & Kretschmer, M. (2020). The influence of stratospheric wave reflection on north American cold spells. *Monthly Weather Review*, *148*, 1675–1690. https://doi.org/10.1175/MWR-D-19-0339.1

McGovern, A., Lagerquist, R., John Gagne, D., Jergensen, G. E., Elmore, K. L., Homeyer, C. R., & Smith, T. (2019). Making the Black box more transparent: Understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society*, *100*(11), 2175–2199.

McKee, T. B., Doesken, N. J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on applied climatology* (Vol. 17(22), pp. 179-183).

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., Dixon, K., Giorgetta, M., Greene, A., Hawkins, E., Hegerl, G., Karoly, D., Keenlyside, N., Kimoto, M., Kirtman, B., Navarra, A., Pulwarty, R., Smith, D., Stammer, D., & Stockdale, T. (2009). Decadal prediction: Can it be skillful? *Bulletin of the American Meteorological Society*, *90*(10), 1467–1486.

Meehl, G. A., Richter, J. H., Teng, H., Capotondi, A., Cobb, K., Doblas-Reyes, F., Donat, M. G., England, M. H., Fyfe, J. C., Han, W., Kim, H., Kirtman, B. P., Kushnir, Y., Lovenduski, N. S., Mann, M. E., Merryfield, W. J., Nieves, V., Pegion, K., Rosenbloom, N., … Xie, S. P. (2021). Initialized earth system prediction from subseasonal to decadal timescales. *Nature Reviews Earth and Environment*, *2*(5), 340–357.

Mendelsohn, R., Emanuel, K., Chonabayashi, S., & Bakkensen, L. (2012). The impact of climate change on global tropical cyclone damage. *Nature Climate Change*, *2*(3), 205–209.

Meyer, H., & Pebesma, E. (2021). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*, *12*, 1620–1633. https://doi.org/10.1111/2041-210X.13650

Miller, D. E., & Wang, Z. (2019). Skillful seasonal prediction of Eurasian winter blocking and extreme temperature frequency. *Geophysical Research Letters*, *46*(20), 11530–11538.

Miloshevich, G., Cozian, B., Abry, P., Borgnat, P., & Bouchet, F. (2023). Probabilistic forecasts of extreme heatwaves using convolutional neural networks in a regime of lack of data. *Physical Review Fluids*, *8*(4), 040501.

Miralles, D. G., Gentine, P., Seneviratne, S. I., & Teuling, A. J. (2019). Land–atmospheric feedbacks during droughts and heatwaves: State of the science and current challenges. *Annals of the New York Academy of Sciences*, *1436*(1), 19–35.

Mokhtarzad, M., Eskandari, F., Jamshidi Vanjani, N., & Arabasadi, A. (2017). Drought forecasting by ANN, ANFIS, and SVM and comparison of the models. *Environment and Earth Science*, *76*, 729. https://doi.org/10.1007/s12665-017-7064-0

Mouatadid, S., Orenstein, P., Flaspohler, G., Cohen, J., Oprescu, M., Fraenkel, E., & Mackey, L. (2023). Adaptive bias correction for improved subseasonal forecasting. *Nature Communications*, *14*(1), 3482.

Mozaffari, A., Langguth, M., Gong, B., Ahring, J., Campos, A. R., Nieters, P., Escobar, O. J. C., Wittenbrink, M., Baumann, P., & Schultz, M. G. (2022). HPC-oriented canonical workflows for machine learning applications in climate and weather prediction. *Data Intelligence*, *4*(2), 271–285.

Nalbantis, I., & Tsakiris, G. (2009). Assessment of hydrological drought revisited. *Water Resources Management*, *23*, 881–897.

Neddermann, N. C., Müller, W. A., Dobrynin, M., Düsterhus, A., & Baehr, J. (2019). Seasonal predictability of European summer climate reassessed. *Climate Dynamics*, *53*, 3039–3056.

O'Gorman, P. A., & Dwyer, J. G. (2018). Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, *10*, 2548–2563. https://doi.org/10.1029/2018MS00135

O, S., Dutra, E., & Orth, R. (2020). Robustness of process-based versus data-driven modeling in changing climatic conditions. *Journal of Hydrometeorology*, *21*(9), 1929–1944. https://doi.org/10.1175/JHM-D-20-0072.1

Olivetti, L., & Messori, G. (2024). Advances and prospects of deep learning for medium-range extreme weather forecasting. *Geoscientific Model Development*, *17*, 2347–2358. https://doi.org/10.5194/gmd-17-2347-2024

Orlowsky, B., & Seneviratne, S. I. (2012). Global changes in extreme events: Regional and seasonal dimension. *Climatic Change*, *110*, 669–696.

Palmer, W. C. (1965). *Meteorological drought* (Vol. 30). US Department of Commerce, Weather Bureau.

Pan, B., Anderson, G. J., Goncalves, A., Lucas, D. D., Bonfils, C. J., & Lee, J. (2022). Improving seasonal forecast using probabilistic deep learning. *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002766.

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., & Lakshminarayanan, B. (2019). Normalizing flows for probabilistic modelling and inference. In arXiv [Stat.ML]. http://arxiv.org/abs/1912.02762

Parey, S., Hoang, T. T. H., & Dacunha-Castelle, D. (2019). Future high-temperature extremes and stationarity. *Natural Hazards*, *98*, 1115–1134. https://doi.org/10.1007/s11069-018-3499-1

Park, S.-J., & Lee, D.-K. (2020). Prediction of coastal flooding risk under climate change impacts in South Korea using machine learning algorithms *environ*. *Research Letters*, *15*, 094052. https://doi.org/10.1088/1748-9326/aba5b3

Pasche, O. C., Wider, J., Zhang, Z., Zscheischler, J., & Engelke, S. (2024). Validating deep-learning weather forecast models on recent high-impact extreme events. arXiv Preprint arXiv:2404.17652.

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., Hall, D., Miele, A., Kashinath, K., & Anandkumar, A. (2022). Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. arXiv Preprint arXiv: 2202.11214.

Patterson, M., Weisheimer, A., Befort, D. J., & O'Reilly, C. H. (2022). The strong role of external forcing in seasonal forecasts of European summer temperature. *Environmental Research Letters*, *17*(10), 104033.

Pegion, K., Becker, E. J., & Kirtman, B. P. (2022). Understanding predictability of daily southeast U.S. precipitation using explainable machine learning. *Artificial Intelligence for the Earth Systems*, *1*, e220011. https://doi.org/10.1175/AIES-D-22-0011.1

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Perkins, S. E., & Alexander, L. V. (2013). On the measurement of heat waves. *Journal of Climate*, *26*, 4500–4517. https://doi.org/10.1175/JCLI-D-12-00383.1

Polkova, I., Afargan-Gerstman, H., Domeisen, D. I., King, M. P., Ruggieri, P., Athanasiadis, P., Dobrynin, M., Aarnes, Ø., Kretschmer, M., Baehr, J., & Baehr, J. (2021). Predictors and prediction skill for marine cold-air outbreaks over the Barents Sea. *Quarterly Journal of the Royal Meteorological Society*, *147*(738), 2638–2656.

Polson, N. G., & Sokolov, V. (2017). Deep learning: A Bayesian perspective. *Bayesian Analysis*, *12*(4), 1275–1304.

Poornima, S., & Pushpalatha, M. (2019). Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network. *Soft Computing*, *23*, 1–14. https://doi.org/10.1007/s00500-019-04120-1

Price, I., Sanchez-Gonzalez, A., Alet, F., Ewalds, T., El-Kadi, A., Stott, J., Mohamed, S., Battaglia, P., Lam, R., & Willson, M. (2023). GenCast: Diffusion-based ensemble forecasting for medium-range weather (arXiv:2312.15796; version 1). arXiv. http://arxiv.org/abs/2312.15796

Prodhan, F. A., Zhang, J., Hasan, S. S., Sharma, T. P. P., & Mohana, H. P. (2022). A review of machine learning methods for drought hazard monitoring and forecasting: Current research trends, challenges, and future research directions. *Environmental Modelling & Software*, *149*, 105327. https://doi.org/10.1016/j.envsoft.2022.105327

Prodhomme, C., Materia, S., Ardilouze, C., White, R. H., Batté, L., Guemas, V., Fragkoulidis, G., & García-Serrano, J. (2022). Seasonal prediction of European summer heatwaves. *Climate Dynamics*, *58*, 2149–2166. https://doi.org/10.1007/s00382-021-05828-3

Pyrina, M., Nonnenmacher, M., Wagner, S., & Zorita, E. (2021). Statistical seasonal prediction of European summer mean temperature using observational, reanalysis, and satellite data. *Weather and Forecasting*, *36*(4), 1537–1560.

Ragone, F., Wouters, J., & Bouchet, F. (2018). Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings. National Academy of Sciences. United States of America*, *115*, 24–29. https://doi.org/10.1073/pnas.1712645115

Ragone, F., & Bouchet, F. (2021). Rare event algorithm study of extreme warm summers and heatwaves over europe. *Geophysical Research Letters*, *48*, e2020GL091197. https://doi.org/10.1029/2020GL091197

Rasmussen, C. E., & Williams, C. K. I. (2005). *Gaussian processes for machine learning*. The MIT Press.

Rasmusson, E. M., & Wallace, J. M. (1983). Meteorological aspects of the El Nino/southern oscillation. *Science*, *222*(4629), 1195–1202.

Rasp, S., Dueben, P. D., Scher, S., Weyn, J. A., Mouatadid, S., & Thuerey, N. (2020). WeatherBench: A benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, *12*, e2020MS002203. https://doi.org/10.1029/2020MS002203

Rasp, S., Hoyer, S., Merose, A., Langmore, I., Battaglia, P., Russel, T., Sanchez-Gonzalez, A., Yang, V., Carver, R., Agrawal, S., Chantry, M., Bouallegue, Z. B., Dueben, P., Bromberg, C., Sisk, J., Barrington, L., Bell, A., & Sha, F. (2023). WeatherBench 2: A benchmark for the next generation of data-driven global weather models. arXiv Preprint arXiv:2308.15560.

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(39), 9684–9689.

Rasp, S., & Thuerey, N. (2021). Data-driven medium-range weather prediction with a resnet pretrained on climate simulations: A new model for weatherbench. *Journal of Advances in Modeling Earth Systems*, *13*(2), e2020MS002405.

Raza, M. A., Almazah, M. M., Ali, Z., Hussain, I., & Al-Duais, F. S. (2022). Application of extreme learning machine algorithm for drought forecasting. *Complexity*, *2022*(1), 4998200.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (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

Reinhold, B. B., & Pierrehumbert, R. T. (1982). Dynamics of weather regimes: Quasi-stationary waves and blocking. *Monthly Weather Review*, *110*(9), 1105–1145.

Rhee, J., & Im, J. (2017). Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data. *Agricultural and Forest Meteorology*, *237–238*, 105–122. https://doi.org/10.1016/j.agrformet.2017.02.011

Richman, M. B., Leslie, L. M., Ramsay, H. A., & Klotzbach, P. J. (2017). Reducing tropical cyclone prediction errors using machine learning approaches. *Procedia computer science*, *114*, 314–323.

Risbey, J. S., Squire, D. T., Black, A. S., DelSole, T., Lepore, C., Matear, R. J., Monselesan, D. P., Moore, T. S., Richardson, D., Schepen, A., Tippett, M. K., & Tozer, C. R. (2021). Standard assessments of climate forecast skill can be misleading. *Nature Communications*, *12*(1), 1–14. https://doi.org/10.1038/s41467-021-23771-z

Rivoire, P., Martius, O., Naveau, P., & Tuel, A. (2023). Assessment of subseasonal-to-seasonal (S2S) ensemble extreme precipitation forecast skill over Europe. *Natural Hazards and Earth System Sciences*, *23*(8), 2857–2871.

Roth, M., Buishand, T. A., Jongbloed, G., Tank, A. K., & Van Zanten, J. H. (2014). Projections of precipitation extremes based on a regional, non-stationary peaks-over-threshold approach: A case study for The Netherlands and north-western Germany. *Weather and Climate Extremes*, *4*, 1–10.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*, 206–215. https://doi.org/10.1038/s42256-019-0048-x

Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statistics Surveys*, *16*, 1–85. https://doi.org/10.1214/21-SS133

Russell, B. T., & Huang, W. K. (2021). Modeling short-ranged dependence in block extrema with application to polar temperature data. *Environmetrics*, *32*, e2661. https://doi.org/10.1002/env.2661

Russo, E., & Domeisen, D. I. (2023). Increasing intensity of extreme heatwaves: The crucial role of metrics. *Geophysical Research Letters*, *50*(14), e2023GL103540.

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Russo, S., Dosio, A., Graversen, R. G., Sillmann, J., Carrao, H., Dunbar, M. B., Singleton, A., Montagna, P., Barbola, P., & Vogt, J. V. (2014). Magnitude of extreme heat waves in present climate and their projection in a warming world. *Journal of Geophysical Research: Atmospheres*, *119*(22), 12–500.

Sahoo, B. B., Jha, R., Singh, A., & Kumar, D. (2019). Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophysica*, *67*, 1471–1481. https://doi.org/10.1007/s11600-019-00330-1

Salcedo-Sanz, S., Pérez-Aracil, J., Ascenso, G., del Ser, J., Casillas-Pérez, D., Kadow, C., Fister, D., Barriopedro, D., García-Herrera, R., Giuliani, M., & Castelletti, A. (2024). Analysis, characterization, prediction, and attribution of extreme atmospheric events with machine learning and deep learning techniques: A review. *Theoretical and Applied Climatology*, *155*(1), 1–44.

Sampurno, J., Vallaeys, V., Ardianto, R., & Hanert, E. (2022). Integrated hydrodynamic and machine learning models for compound flooding prediction in a data-scarce estuarine delta. *Nonlinear Processes in Geophysics*, *29*, 301–315. https://doi.org/10.5194/npg-29-301-2022

Santoso, A., Mcphaden, M. J., & Cai, W. (2017). The defining characteristics of ENSO extremes and the strong 2015/2016 El Niño. *Reviews of Geophysics*, *55*(4), 1079–1129.

Sanuy, M., Rigo, T., Jiménez, J. A., & Llasat, M. C. (2020). Classifying compound coastal storm and heavy rainfall events in the north-western Spanish Mediterranean, Hydrol. *Earth Systems Science*, *25*, 3759–3781. https://doi.org/10.5194/hess-25-3759-2021

Scaife, A. A., Baldwin, M. P., Butler, A. H., Charlton-Perez, A. J., Domeisen, D. I. V., Garfinkel, C. I., Hardiman, S. C., Haynes, P., Karpechko, A. Y., Lim, E.-P., Noguchi, S., Perlwitz, J., Polvani, L., Richter, J. H., Scinocca, J., Sigmond, M., Shepherd, T. G., Son, S.-W., & Thompson, D. W. J. (2022). Long-range prediction and the stratosphere. *Atmospheric Chemistry and Physics*, *22*, 2601–2623. https://doi.org/10.5194/acp-22-2601-2022

Scaife, A. A., & Smith, D. (2018). A signal-to-noise paradox in climate science. *npj Climate and Atmospheric Science*, *1*, 28. https://doi.org/10.1038/s41612-018-0038-4

Schaefer, J. T. (1990). The critical success index as an indicator of warning skill. *Weather and Forecasting*, *5*(4), 570–575.

Schepen, A., Wang, Q. J., & Robertson, D. E. (2014). Seasonal forecasts of Australian rainfall through calibration and bridging of coupled GCM outputs. *Monthly Weather Review*, *142*(5), 1758–1770.

Scher, S., & Messori, G. (2021). Ensemble methods for neural network-based weather forecasts. *Journal of Advances in Modeling Earth Systems*, *13*(2), e2020MS002331. https://doi.org/10.1029/2020MS002331

Scheuerer, M., Switanek, M. B., Worsnop, R. P., & Hamill, T. M. (2020). Using artificial neural networks for generating probabilistic subseasonal precipitation forecasts over California. *Monthly Weather Review*, *148*, 3489–3506. https://doi.org/10.1175/MWR-D-20-0096.1

Schubert, S., Koster, R., Hoerling, M., Seager, R., Lettenmaier, D., Kumar, A., & Gutzler, D. (2007). Predicting drought on seasonal-to-decadal time scales. *Bulletin of the American Meteorological Society*, *88*, 1625–1630. https://doi.org/10.1175/BAMS-88-10-1625

Schulz, B., & Lerch, S. (2022). Machine learning methods for postprocessing ensemble forecasts of wind gusts: A systematic comparison. *Monthly Weather Review*, *150*(1), 235–257.

Scoccimarro, E., Gualdi, S., Bellucci, A., Zampieri, M., & Navarra, A. (2013). Heavy precipitation events in a warmer climate: Results from CMIP5 models. *Journal of Climate*, *26*(20), 7902–7911.

Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S. M., Wehner, M., & Zhou, B. (2021). Weather and climate extreme events in a changing climate. In V. Mason-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou (Eds.), *Climate change 2021: The physical science basis. Contribution of Working Group I to the sixth assessment report of the intergovernmental panel on climate change* (pp. 11.1–11.345). Cambridge University Press.

Shukla, J. (1998). Predictability in the midst of chaos: A scientific basis for climate forecasting. *Science*, *282*(5389), 728–731.

Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., & Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research – Atmospheres*, *118*, 1716–1733. https://doi.org/10.1002/jgrd.50203

Sillmann, J., Thorarinsdottir, T., Keenlyside, N., Schaller, N., Alexander, L. V., Hegerl, G., Seneviratne, S. I., Vautard, R., Zhang, X., & Zwiers, F. W. (2017). Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities. *Weather and Climate Extremes*, *18*, 65–74.

Silva, S. J., Keller, C. A., & Hardin, J. (2022). Using an explainable machine learning approach to characterize earth system model errors: Application of SHAP analysis to modeling lightning flash occurrence. *Journal of Advances in Modeling Earth Systems*, *14*(4), e2021MS002881.

Silva, S. J., & Keller, C. A. (2024). Limitations of XAI methods for process-level understanding in the atmospheric sciences. *Artificial Intelligence for the Earth Systems*, *3*, e230045. https://doi.org/10.1175/AIES-D-23-0045.1

Slater, L. J., Arnal, L., Boucher, M.-A., Chang, A. Y.-Y., Moulds, S., Murphy, C., Nearing, G., Shalev, G., Shen, C., Speight, L., Villarini, G., Wilby, R. L., Wood, A., & Zappa, M. (2023). Hybrid forecasting: Blending climate predictions with AI models. *Hydrology and Earth System Sciences*, *27*, 1865–1889. https://doi.org/10.5194/hess-27-1865-2023

Smith, D. M., Eade, R., Scaife, A. A., Caron, L. P., Danabasoglu, G., DelSole, T. M., Delworth, T., Doblas-Reyes, F. J., Dunstone, N. J., Hermanson, L., Kharin, V., Kimoto, M., Merryfield, W. J., Mochizuki, T., Müller, W. A., Pohlmann, H., Yeager, S., & Yang, X. (2019). Robust skill of decadal climate predictions. *Npj Climate and Atmospheric Science*, *2*(1), 13. https://doi.org/10.1038/s41612-019-0071-y

Specq, D., & Batté, L. (2020). Improving subseasonal precipitation forecasts through a statistical–dynamical approach: Application to the southwest tropical Pacific. *Climate Dynamics*, *55*(7–8), 1913–1927.

Spuler, F. R., Kretschmer, M., Kovalchuck, Y., Balmaseda, M. A., & Shepherd, T. G. (2024). Identifying probabilistic weather regimes targeted to a local-scale impact variable. arXiv Preprint arXiv:2402.15379.

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Strazzo, S., Collins, D. C., Schepen, A., Wang, Q. J., Becker, E., & Jia, L. (2019). Application of a hybrid statistical–dynamical system to seasonal prediction of north American temperature and precipitation. *Monthly Weather Review*, *147*(2), 607–625.

Suarez-Gutierrez, L., Müller, W. A., Li, C., & Marotzke, J. (2020). Dynamical and thermodynamical drivers of variability in European summer heat extremes. *Climate Dynamics*, *54*, 4351–4366.

Sun, Z., Sandoval, L., Crystal-Ornelas, R., Mousavi, S. M., Wang, J., Lin, C., Cristea, N., Tong, D., Carande, W. H., Ma, X., Rao, Y., Bednar, J. A., Tan, A., Wang, J., Purushotham, S., Gill, T. E., Chastang, J., Howard, D., Holt, B., … John, A. (2022). A review of earth artificial intelligence. *Computational Geosciences*, *159*, 105034. https://doi.org/10.1016/j.cageo.2022.105034

Sutanto, S. J., van der Weert, M., Wanders, N., Blauhut, V., & van Lanen, H. A. J. (2019). Moving from drought hazard to impact forecasts. *Nature Communications*, *10*, 4945. https://doi.org/10.1038/s41467-019-12840-z

Sweet, L. B., Müller, C., Anand, M., & Zscheischler, J. (2023). Cross-validation strategy impacts the performance and interpretation of machine learning models. *Artificial Intelligence for the Earth Systems*, *2*, 1–35.

Tan, J., Liu, H., Li, M., & Wang, J. (2018). A prediction scheme of tropical cyclone frequency based on lasso and random forest. *Theoretical and Applied Climatology*, *133*, 973–983.

Thompson, V., Mitchell, D., Hegerl, G. C., Collins, M., Leach, N. J., & Slingo, J. M. (2023). The most at-risk regions in the world for high-impact heatwaves. *Nature Communications*, *14*, 2152. https://doi.org/10.1038/s41467-023-37554-1

Thuemmel, J., Karlbauer, M., Otte, S., Zarfl, C., Martius, G., Ludwig, N., Scholten, T., Friedrich, U., Wulfmeyer, V., Goswami, B., & Butz, M. V. (2023). Inductive biases in deep learning models for weather prediction (arXiv:2304.04664). arXiv. https://doi.org/10.48550/arXiv.2304.04664

Tjoa, E., & Guan, C. (2019). A survey on explainable artificial intelligence (xai): Towards medical xai. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(11), 4793–4813. https://doi.org/10.48550/arXiv.1907.07374

Tomassini, L., Gerber, E. P., Baldwin, M. P., Bunzel, F., & Giorgetta, M. (2012). The role of stratosphere-troposphere coupling in the occurrence of extreme winter cold spells over northern Europe. *Journal of Advances in Modeling Earth Systems*, *4*, M00A03. https://doi.org/10.1029/2012MS000177

Toms, B. A., Barnes, E. A., & Ebert-Uphoff, I. (2020). Physically interpretable neural networks for the geosciences: Applications to earth system variability. *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS002002. https://doi.org/10.1029/2019MS002002

Torralba, V., Materia, S., Cavicchia, L., Álvarez-Castro, M. C., Prodhomme, C., McAdam, R., Scoccimarro, E., & Gualdi, S. (2024). Nighttime heat waves in the Euro-Mediterranean region: Definition, characterisation, and seasonal prediction. *Environmental Research Letters*, *19*(3), 034001.

Trenary, L., & DelSole, T. (2023). Skillful statistical prediction of subseasonal temperature by training on dynamical model data. *Environmental Data Science*, *2*, E7. https://doi.org/10.1017/eds.2023.2

Tuel, A., & Martius, O. (2022). The influence of modes of climate variability on the subseasonal temporal clustering of extreme precipitation. *IScience*, *25*(3), 103855. https://doi.org/10.1016/j.isci.2022.103855

van Straaten, C., Whan, K., Coumou, D., van den Hurk, B., & Schmeits, M. (2022). Using explainable machine learning forecasts to discover subseasonal drivers of high summer temperatures in western and central Europe. *Monthly Weather Review*, *150*(5), 1115–1134.

van Straaten, C., Whan, K., Coumou, D., van den Hurk, B., & Schmeits, M. (2023). Correcting subseasonal forecast errors with an explainable ANN to understand misrepresented sources of predictability of European summer temperatures. *Artificial Intelligence for the Earth Systems*, *2*, 1–49.

Vannitsem, S., Bremnes, J. B., Demaeyer, J., Evans, G. R., Flowerdew, J., Hemri, S., Lerch, S., Roberts, N., Theis, S., Atencia, A., Ben Bouallègue, Z., Bhend, J., Dabernig, M., de Cruz, L., Hieta, L., Mestre, O., Moret, L., Plenković, I. O., Schmeits, M., … Ylhaisi, J. (2021). Statistical postprocessing for weather forecasts: Review, challenges, and avenues in a big data world. *Bulletin of the American Meteorological Society*, *102*(3), E681–E699. https://doi.org/10.1175/BAMS-D-19-0308.1

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *Journal of climate*, *23*(7), 1696–1718.

von Storch, H., & Zwiers, F. W. (1999). Misuses of statistical analysis in climate research. In *Statistical analysis in climate research*. Cambridge University Press.

Vosper, E., Watson, P., Harris, L., McRae, A., Santos-Rodriguez, R., Aitchison, L., & Mitchell, D. (2023). Deep learning for downscaling tropical cyclone rainfall to hazard-relevant spatial scales. *Journal of Geophysical Research: Atmospheres*, *128*, e2022JD038163. https://doi.org/10.1029/2022JD038163

Vovk, V., Gammerman, A., & Shafer, G. (2005). *Algorithmic learning in a random world*. Springer.

Wahl, T., Jain, S., Bender, J., Meyers, S. D., & Luther, M. E. (2015). Increasing risk of compound flooding from storm surge and rainfall for major US cities. *Nature Climate Change*, *5*, 1093–1097. https://doi.org/10.1038/nclimate2736

Wang, S., Huang, J., He, Y., & Guan, Y. (2014). Combined effects of the Pacific decadal oscillation and El Nino-southern oscillation on global land dry–wet changes. *Scientific Reports*, *4*(1), 6651.

Watson-Parris, D., Rao, Y., Olivié, D., Seland, Ø., Nowack, P., Camps-Valls, G., Stier, P., Bouabid, S., Dewey, M., Fons, E., Gonzalez, J., Harder, P., Jeggle, K., Lenhardt, J., Manshausen, P., Novitasari, M., Ricard, L., & Roesch, C. (2022). ClimateBench v1. 0: A benchmark for data-driven climate projections. *Journal of Advances in Modeling Earth Systems*, *14*(10), e2021MS002954.

Watson, P. A. G. (2022). Machine learning applications for weather and climate need greater focus on extremes. *Environmental Research Letters*, *17*, 111004. https://doi.org/10.1088/1748-9326/ac9d4e

Weirich-Benet, E., Pyrina, M., Jiménez-Esteve, B., Fraenkel, E., Cohen, J., & Domeisen, D. I. (2023). Subseasonal prediction of central European summer heatwaves with linear and random Forest machine learning models. *Artificial Intelligence for the Earth Systems*, *2*(2), e220038.

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*, 9. https://doi.org/10.1186/s40537-016-0043-6

Weyn, J. A., Durran, D. R., Caruana, R., & Cresswell-Clay, N. (2021). Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. *Journal of Advances in Modeling Earth Systems*, *13*(7), e2021MS002502.

White, R. H., Anderson, S., Booth, J. F., Braich, G., Draeger, C., Fei, C., Harley, C. D. G., Henderson, S. B., Jakob, M., Lau, C.-A., Admasu, L. M., Narinesingh, V., Rodell, C., Roocroft, E., Weinberger, K. R., & West, G. (2023). The unprecedented Pacific northwest heatwave of June 2021. *Nature Communications*, *14*(1), 727.

Wi, S., & Steinschneider, S. (2022). Assessing the physical realism of deep learning hydrologic model projections under climate change. *Water Resources Research*, *58*, e2022WR032123. https://doi.org/10.1029/2022WR032123

Wilhite, D. A. (2000). Drought as a natural hazard: Concepts and definitions. In D. A. Wilhite (Ed.), *Drought: A Global Assessment* (Vol. 1, pp. 1–18). Routledge.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 1–9.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*. Academic press.

Wulff, C. O., Vitart, F., & Domeisen, D. I. (2022). Influence of trends on subseasonal temperature prediction skill. *Quarterly Journal of the Royal Meteorological Society*, *148*(744), 1280–1299.

Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., & Yang, M.-H. (2022). Diffusion models: A comprehensive survey of methods and applications. arXiv [cs.LG]. http://arxiv.org/abs/2209.00796

Yang, Q., Lee, C., Tippett, M. K., Chavas, D. R., & Knutson, T. R. (2022). Machine learning-based hurricane wind reconstruction. *Weather and Forecasting*, *37*(4), 477–493. https://doi.org/10.1175/WAF-D-21-0077.1

Yin, J., Gentine, P., Slater, L., Gu, L., Pokhrel, Y., Hanasaki, N., Guo, S., Xiong, L., & Schlenker, W. (2023). Future socio-ecosystem productivity threatened by compound drought–heatwave events. *Nature Sustainability*, *6*, 259–272. https://doi.org/10.1038/s41893-022-01024-1

Yiou, P., Goubanova, K., Li, Z. X., & Nogaj, M. (2008). Weather regime dependence of extreme value statistics for summer temperature and precipitation. *Nonlinear Processes in Geophysics*, *15*(3), 365–378.

Zahid, M., Blender, R., Lucarini, V., & Bramati, M. C. (2017). Return levels of temperature extremes in southern Pakistan. *Earth System Dynamics*, *8*, 1263–1278. https://doi.org/10.5194/esd-8-1263-2017

Zanna, L., & Bolton, T. (2020). Data-driven equation discovery of ocean mesoscale closures. *Geophysical Research Letters*, *47*(17), e2020GL088376.

Zeder, J., & Fischer, E. M. (2023). Quantifying the statistical dependence of mid-latitude heatwave intensity and likelihood on prevalent physical drivers and climate change. *Advances in Statistical Climatology, Meteorology and Oceanography*, *9*, 83–102. https://doi.org/10.5194/ascmo-9-83-2023

Zhang, B., Wu, Y., Zhao, B., Chanussot, J., Hong, D., Yao, J., & Gao, L. (2022). Progress and challenges in intelligent remote sensing satellite systems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 1814–1822.

Zhang, C. (2005). Madden-Julian oscillation. *Reviews of Geophysics*, *43*(2), RG2003. https://doi.org/10.1029/2004RG000158

Zhang, C., Adames, Á. F., Khouider, B., Wang, B., & Yang, D. (2020). Four theories of the madden-Julian oscillation. *Reviews of Geophysics*, *58*(3), e2019RG000685.

Zhang, L., Yang, T., Gao, S., Hong, Y., Zhang, Q., Wen, X., & Cheng, C. (2023). Improving subseasonal-to-seasonal forecasts in predicting the occurrence of extreme precipitation events over the contiguous US using machine learning models. *Atmospheric Research*, *281*, 106502.

Zhang, R., Chen, Z. Y., Xu, L. J., & Ou, C. Q. (2019). Meteorological drought forecasting based on a statistical model with machine learning techniques in Shaanxi province. *China. Science of the Total Environment*, *665*, 338–346. https://doi.org/10.1016/j.scitotenv.2019.01.431

Zhang, R., Sutton, R., Danabasoglu, G., Kwon, Y. O., Marsh, R., Yeager, S. G., Amrhein, D. E., & Little, C. M. (2019). A review of the role of the Atlantic meridional overturning circulation in Atlantic multidecadal variability and associated climate impacts. *Reviews of Geophysics*, *57*(2), 316–375.

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., Trewin, B., & Zwiers, F. W. (2011). Indices for monitoring changes in extremes based on daily temperature and precipitation data. *WIREs Climate Change*, *2*(6), 851–870.

Zhao, W. L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Lin, C., Li, X., & Qiu, G. Y. (2019). Physics-constrained machine learning of evapotranspiration. *Geophysical Research Letters*, *46*(24), 14496–14507.

Zscheischler, J., Naveau, P., Martius, O., Engelke, S., & Raible, C. C. (2021). Evaluating the dependence structure of compound precipitation and wind speed extremes. *Earth System Dynamics*, *12*, 1–16. https://doi.org/10.5194/esd-12-1-2021

Zscheischler, J., Westra, S., van den Hurk, B. J. J. M., Ward, P. J., Pitman, A., Aghakouchak, A., Bresch, D. N., Leonard, M., Wahl, T., & Zhang, X. (2018). Future climate risk from compound events. *Nature Climate Change*, *8*, 469–477. https://doi.org/10.1038/s41558-018-0156-3

**How to cite this article:** Materia, S., García, L. P., van Straaten, C., O, S., Mamalakis, A., Cavicchia, L., Coumou, D., de Luca, P., Kretschmer, M., & Donat, M. (2024). Artificial intelligence for climate prediction of extremes: State of the art, challenges, and future perspectives. *WIREs Climate Change*, e914. https://doi.org/10.1002/wcc.914

17577799, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.914 by Test, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License