

EARTH SCIENCES

AI for atmosphere–ocean sciences: advancements, challenges and ways forward

Jing-Jia Luo^{1,*}, Jiangjiang Xia^{2,*}, Baoxiang Pan^{3,*}, Yoo-Geun Ham^{4,*}, Xiaofeng Li^{5,*}, Wei Shangguan^{6,*}, Wei Xue^{7,8,*}, Yaqiang Wang^{9,10,*}, Bin Mu^{11,*}, Youngjoon Hong^{12,*}, Hao Li^{13,*}, Xiaohui Zhong^{13,*}, Kan Dai^{14,*}, Lei Bai^{15,*}, Fenghua Ling^{15,*}, Niklas Boers^{16,17}, Christopher Bretherton¹⁸, Bin Chen^{19,20}, Dongjin Cho²¹, Pierre Gentine²², Zijie Guo¹⁵, Xiaomeng Huang²³, Daehyun Kang²⁴, Hyunwoo J. Kim²⁵, Jeong-Hwan Kim²⁴, Lili Lei²⁶, Fan Meng¹, Seol-Hee Oh²⁷, Bo Qin¹¹, Zixiong Shen¹, Qiming Sun¹, Yuheng Tang¹, Xuan Tong¹, Bingcheng Wan¹, Lina Wang¹, Ya Wang²⁸, Yiming Wang⁷, Jiye Wu¹, Yi Xiao⁷, Lina Yao^{29,30}, Song Yang³¹, Chaoxia Yuan¹, Shijin Yuan¹¹, Tingzhao Yu^{10,32} and Mengchu Zhao¹

ABSTRACT

Artificial intelligence (AI) is rapidly transforming Earth science, offering unprecedented capabilities to tackle the most pressing challenges in the field. This work explores significant advances and emerging challenges across the AI for atmosphere–ocean sciences, while outlining critical ways forward. We review deep-learning methods and their application in weather and climate forecasting, which outperforms dynamical models in accuracy and computational efficiency. The role of AI in detecting complex phenomena, enhancing data assimilation and reconstruction, bias correction and downscaling coarse model outputs is also examined. However, the ‘black-box’ nature of complex AI models necessitates a focus on explainable AI to build trust and extract mechanistic insight. The most promising path forward is identified as the development of hybrid physics–AI modeling, which integrates the data-driven power of AI with the foundational constraints of physical laws to ensure generalizability and causal consistency. A new framework for AI-based model intercomparison is essential for rigorous benchmark performance. Finally, we contextualize these technical developments by discussing the usefulness and applicability of AI to society, including the improvement of multi-hazard early-warning systems and green energy production. We conclude by envisioning the future of AI agents for Earth science—autonomous, goal-oriented systems capable of designing and running experiments, generating and testing hypotheses, and learning dynamics from multisource data. This synthesis underscores that AI is not merely a tool, but a paradigm shift, which will significantly improve how we understand and adapt to a changing climate.

Keywords: AI application and challenge, atmosphere–ocean sciences, explainable AI, AI-MIP, AI agent

INTRODUCTION

Atmospheric and oceanic sciences have progressed through the combined supports of mon-

itoring, theoretical analysis and numerical modeling over the past century. Intensified observations reveal major atmosphere and ocean phenomena—

¹State Key Laboratory of Climate System Prediction and Risk Management (CPRM)/Jiangsu Key Laboratory of Intelligent Weather Forecasting and Applications Based on Big Data/ICAR/CIC-FEMD/KLME/ILCEC, School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing 210044, China; ²Temperate East Asia Research Center on Global Change, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China; ³National Key Laboratory of Earth System Numerical Modeling and Application, Institute of Atmospheric Physics, Chinese Academy of Science, Beijing 100029, China;

(Continued on next page)

***Corresponding authors.** E-mails: jjluo@nuist.edu.cn; xiajj@tea.ac.cn; panbaoxiang@lasg.iap.ac.cn; yoogeun@snu.ac.kr; lix@qdio.ac.cn; shgwei@mail.sysu.edu.cn; xuewei@tsinghua.edu.cn; yqwang@cma.gov.cn; binmu@tongji.edu.cn; hongyj@snu.ac.kr; lihao_lh@fudan.edu.cn; zhong_xiaohui@fudan.edu.cn; daikan@cma.gov.cn; baisanshi@gmail.com; lingfenghua@pmlab.org.cn

Received 31 October 2025; **Revised** 14 January 2026; **Accepted** 27 January 2026

(Continued from
previous page)

⁴Department of Environmental Management, Seoul National University, Seoul 08826, Republic of Korea; ⁵Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China; ⁶School of Atmospheric Sciences, Sun Yat-sen University, Guangzhou 510275, China; ⁷Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China; ⁸Qinghai University and Intelligent Computing and Application Laboratory of Qinghai Province, Xining 810016, China; ⁹State Key Laboratory of Severe Weather Meteorological Science and Technology & Institute of Artificial Intelligence for Meteorology, Chinese Academy of Meteorological Sciences, Beijing 100081, China; ¹⁰Hebei Key Laboratory of Meteorological Artificial Intelligence, Meteorological Science Modeling Research and Development Center, Xiong'an Institute of Meteorological Artificial Intelligence, Xiong'an New Area 070001, China; ¹¹School of Computer Science and Technology, Tongji University, Shanghai 201804, China; ¹²Department of Mathematical Sciences, Seoul National University, Seoul 08826, Republic of Korea; ¹³Artificial Intelligence Innovation and Incubation Institute, Fudan University, Shanghai 200433, China; ¹⁴National Meteorological Center of CMA, Beijing 100081, China; ¹⁵Shanghai Artificial Intelligence Laboratory, Shanghai 200030, China;

frontal systems, eddies, waves and large-scale circulations—and make it possible to trace their complete life cycles. Meanwhile, advances in fluid dynamics and thermodynamics have paved the way to a solid physical framework for understanding atmospheric and oceanic patterns [1–4].

To forecast them, numerical weather- and ocean-prediction systems have assembled all available data from *in situ* and remote observations into gridded initial conditions. High-performance computers then solve the governing equations to simulate the evolution of atmospheric and oceanic processes on timescales spanning from hours to decades [5,6]. While this classic paradigm has provided a reliable approach for understanding and predicting atmosphere and ocean phenomena, three long-standing challenges still remain. First, data-assimilation (DA) methods continue to struggle with the sheer volume and heterogeneity of expanded observations. Second, uncertainties in subgrid process parameterizations give rise to systematic biases that worsen in extreme-event simulations. Third, the exponentially increased computational cost for high-resolution ensemble forecasting has prevented further model improvements.

Artificial intelligence (AI)—more precisely, deep learning (DL)—now offers a new promising and complementary approach to mitigate these challenges. By directly mining nonlinear relationships from large datasets, AI models reconstruct missing or finer-resolution observations in seconds [7], detect and track critical phenomena [8] and deliver global forecasts on hourly, days, seasonal and even decadal timescales [9–11], with their performance often exceeding those of current dynamical models.

However, several key issues should be resolved before AI can be fully integrated into atmospheric and oceanic sciences. For instance, existing techniques for incorporating ever larger and more diverse observation streams into AI models are still primitive, limiting their performance in real-time forecasts. Most of the purely data-driven AI architectures lack fundamental conservation laws and dynamical balances, which leads to their low explainability and undermines their reliability in dealing with unprecedented or extreme conditions [12,13]. In addition, global AI forecast models also suffer systematic biases in their temporal and spatial structures [12] and tools to trace forecast errors back to model components or initial inputs have not yet matured. It is also unclear whether a single AI framework can cover the entire spectrum of forecast horizons from immediate nowcasting to seasonal–decadal predictions as

well as climate projections [13], which dynamical models can fully provide.

This review article summarizes recent advances and challenges in broad AI applications in atmosphere–ocean sciences along with potential ways forward. After giving a succinct introduction of DL and its key difference from traditional methods in the second section, the third and fourth sections focus on the AI applications for the detection of atmospheric, land and oceanic phenomena, data reconstruction and assimilation. The fifth section reviews AI performance across different timescales on weather–subseasonal to interannual–decadal forecasts, along with the evolution of large AI models, while the sixth section focuses on the bias correction and downscaling of dynamical model products. Concerning the intrinsic defects of DL, the seventh to ninth sections review possible approaches to improving their transparency, reliability and performance evaluation, focusing on eXplainable AI (XAI), hybrid physics–AI modeling and AI-based model intercomparison. The tenth section briefly discusses how AI-based models can better serve society. The final section presents the latest progress in AI agents for climate science, which provides a profound step toward a humans–AI collaborative paradigm for better confronting our challenges.

AI DL VERSUS PHYSICS-BASED AND CLASSIC STATISTICAL METHODS

Scientists have long relied on two main categories of approaches to understand and predict weather and climate processes as well as their variabilities, i.e. physics-based and statistical models [14]. These approaches have collectively formed a robust analytical framework that has advanced our understanding of weather and climate systems. Given the established efficacy of these classical methods, it has become essential to examine the unique capabilities that DL brings to atmosphere–ocean sciences and how these methods complement or enhance existing tools.

Physics-based models

Physics-based models form the cornerstone of Earth science, grounded in fundamental principles such as conservation laws and thermodynamics. However, they exhibit systematic biases stemming from the closure problem, in which unresolved small-scale processes must be approximated through parameterizations [15]. Current schemes often rely on empirical relationships derived from limited observations or high-resolution simulations, introducing substan-

tial uncertainties and context-dependent biases that vary across weather–climate regimes and scales. These biases propagate through the system, impacting both short-term weather forecasts and long-term climate prediction/projection.

While physics-based models excel at the step-by-step integration of governing equations, deficiencies in representing small-scale and slowly evolving physical processes often limit their ability to capture emergent patterns across multiple spatiotemporal scales. The reducible structures evident in the phase space of these dynamical systems suggest that pure physical integration—despite its mechanistic rigor—may not be optimal for all applications. Data-driven methods can complement traditional modeling by directly identifying and exploiting these patterns, offering enhanced insights into complex phenomena [16].

Statistical models

While the weather and climate are governed by a chaotic, high-dimensional, multiscale dynamical system, the atmosphere–ocean sciences have traditionally relied on several key statistical approaches, focusing on extracting dominant patterns from observed data. Using techniques such as linear regression, compositional analysis and time-series analysis, these models reduce dimensionality and analyse spectra to navigate the complexity of the system. Core to this approach is the incorporation of randomness to represent chaotic natural features, resulting in a framework that is probabilistic and largely agnostic toward causality. However, these statistical approaches typically rely on prior knowledge via a parametric prior (e.g. Gaussian distributions), while the high-dimensional, multiscale and chaotic nature of weather and climate systems often violates these assumptions.

DL

DL represents a particularly powerful branch of machine learning (ML) that employs hierarchical layered transformations and abstractions [17]. When data and computation are abundant, DL demonstrates exceptional advantages through several key innovations in the learning process as follows:

$$\text{Learning} = \text{Representation} + \text{Evaluation} + \text{Optimization}, \quad (1)$$

where ‘Representation’ means to choose how to transform raw data into useful features and select a model architecture that defines the hypothesis

space that the learner can explore [18,19]. ‘Evaluation’ means to specify the function that defines ‘good’ performance and steer the behavior of the model during training. ‘Optimization’ means the process that uses data-driven algorithms to adjust parameters to minimize loss and achieve the defined objective. Note that representation (including both feature representation and function representation), evaluation and optimization are not treated separately in DL. Below we highlight how modern DL combines these aspects synergistically for superior performance.

DL revolutionizes function representation by automating feature discovery, in which models adaptively extract task-relevant patterns that optimize specified loss functions, often uncovering insights beyond human-engineered features [20]. This benefit is particularly evident given abundant, high-quality training data. Hierarchical architectures enable the progressive composition of low-level features into abstract, invariant representations, facilitating the nonlinear modeling of complex phenomena and offering superior predictive power over shallow methods. Depth plays a pivotal role in enhancing inductive biases that leverage inherent data symmetries and structures in deep architectures. For instance, convolutional neural networks (CNNs) embed translational invariance for processing gridded data [21]; recurrent neural networks (RNNs) such as long short-term memory (LSTM) capture temporal dependencies in time-series forecasting [22–24]; graph neural networks (GNNs) are utilized to model the relational topologies inherent in weather and climate networks [16,25]; neural operators incorporate spectral mappings for partial differential equation (PDE)-based modeling [26,27]; transformers facilitate global interactions for teleconnections [28]; and hybrid architectures such as convolutional-long short-term memory (Conv-LSTM) integrate multiple biases for spatiotemporal tasks [29]. Beyond discriminative architectures, generative DL approaches—including variational autoencoders (VAEs) [30], generative adversarial networks (GANs) [31] and diffusion models [32,33]—learn to approximate underlying data distributions from samples. In particular, multistep generative models such as diffusion processes, formulated through stochastic differential equations or flow matching [34], progressively capture multiscale spectral structures inherent in data. This hierarchical generation process offers considerable potential for developing foundation models capable of addressing key challenges in DA and seamless prediction across timescales [35].

Evaluation fundamentally revolves around differentiable objectives that enable gradient-based

(Continued from previous page)

¹⁶Earth System Modelling, School of Engineering and Design, Technical University of Munich, Munich 85521, Germany; ¹⁷Potsdam Institute for Climate Impact Research, Potsdam 14412, Germany; ¹⁸Allen Institute for Artificial Intelligence (AI2), Seattle WA 98105, USA; ¹⁹Key Laboratory for Semi-Arid Climate Change of the Ministry of Education, College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000, China; ²⁰Institute of Meteorological Artificial Intelligence Research, Lanzhou University, Lanzhou 730000, China; ²¹Environmental Planning Institute, Seoul National University, Seoul 08826, Republic of Korea; ²²Learning the Earth with AI and Physics (LEAP) NSF Science and Technology Center, Columbia University, New York NY 10027, USA; ²³Department of Earth System Science, Tsinghua University, Beijing 100084, China; ²⁴Center for Climate and Carbon Cycle Research, Korea Institute of Science and Technology, Seoul 02792, Republic of Korea; ²⁵School of Computing, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea; ²⁶School of Atmospheric Sciences, Nanjing University, Nanjing 210063, China and ²⁷Information & Electronics Research Institute, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea

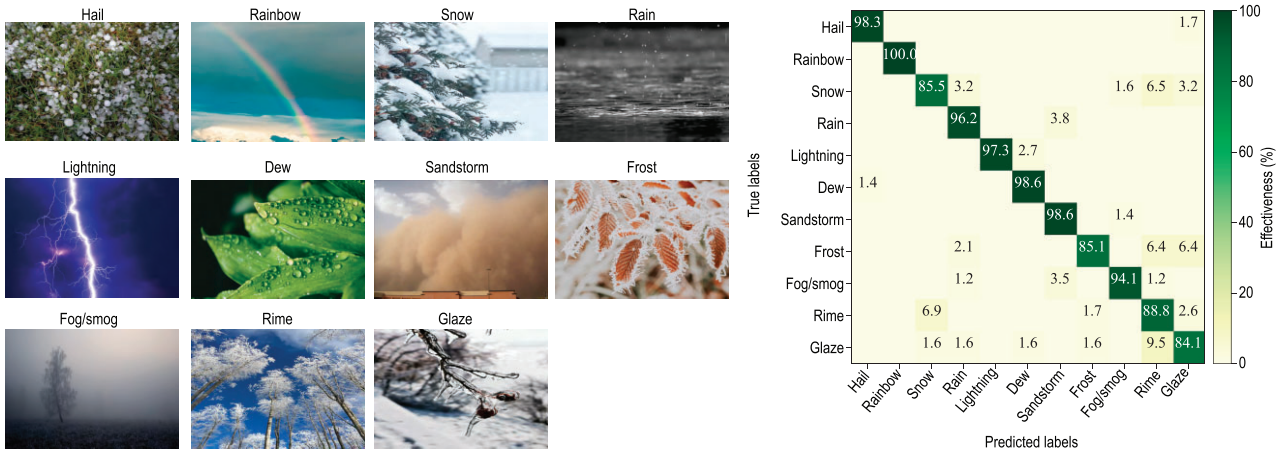


Figure 1. Representative examples of 11 weather-phenomena images (left panel) and the effectiveness of a CNN model evaluated by using the confusion matrix (right panel) [39].

optimization. These include cross-entropy loss for classification tasks; probabilistic losses such as those parameterizing output distributions for regression with uncertainty quantification; generative paradigms such as VAEs [30], GANs [31], diffusion models [32,33] and flow matching [36] for learning high-dimensional distributions; and reward maximization for reinforcement learning [37].

AI-BASED PHENOMENA MONITORING, DETECTION, CLASSIFICATION AND SEGMENTATION

The identification and tracking of phenomena have traditionally relied on physical models or empirically defined threshold-based methods. However, these conventional approaches face significant limitations in multimodal data coupling, fine-scale feature extraction and heterogeneous data fusion. AI-based methods overcome these constraints by data mining representations, eliminating the dependence on manually crafted rules and thresholds. This advancement enables automated, multiscale information extraction—from object-level detection to pixel-level segmentation—and marks a paradigm shift in the dynamic monitoring and mechanistic understanding of multiscale phenomena [38,39].

Atmosphere-phenomena detection

AI-based methods have achieved significant progress in monitoring and analysing various atmosphere phenomena (Fig. 1), including hail, frost, rainbow and smog. In addition, based on remote sensing data, DL models were built to detect and classify tropical cyclones (TCs) and es-

timate their intensities [40], providing pixel-level identification of TCs and atmospheric rivers [41] and accurate precipitation-type classification [42], achieving superior accuracy to those of traditional methods.

Land-phenomena detection

Key applications include the monitoring of land-use and land-cover (LULC) changes, surface albedo and temperature, soil moisture, evapotranspiration, water reservoirs, cryosphere dynamics (glaciers and ice sheets), snow, vegetation and crop growth, as well as related natural hazards such as bushfires, landslides, droughts and river flooding [43]. Among these, LULC is particularly critical as a primary input for the land-surface component of Earth-system models (ESMs), supplying essential information on vegetation types, urban extents and crop land [44]. Despite the abundance of satellite-derived LULC products, inconsistencies in classifications hinder their direct application in ESMs. Addressing these discrepancies through data harmonization remains a key challenge for improving the accuracy and utility of LULC datasets. Another key challenge involves harnessing DL to reconstruct historical trends and project the future dynamics of land-surface variables—such as Land Use/Land Cover Change and the leaf-area index (Fig. 2a)—beyond the temporal coverage of available satellite observations, especially at high resolution.

Ocean-phenomena detection

With AI and multisource remote sensing data—such as infrared, microwave and synthetic aperture radar, complex oceanic processes are

(Continued from previous page)

²⁸State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China; ²⁹The Commonwealth Scientific and Industrial Research Organisation, Sydney 2113, Australia; ³⁰University of New South Wales, Sydney 2052, Australia;

³¹Jiangxi Climate Center, Nanchang 330096, China and ³²Public Meteorological Service Centre, China Meteorological Administration, Beijing 100081, China

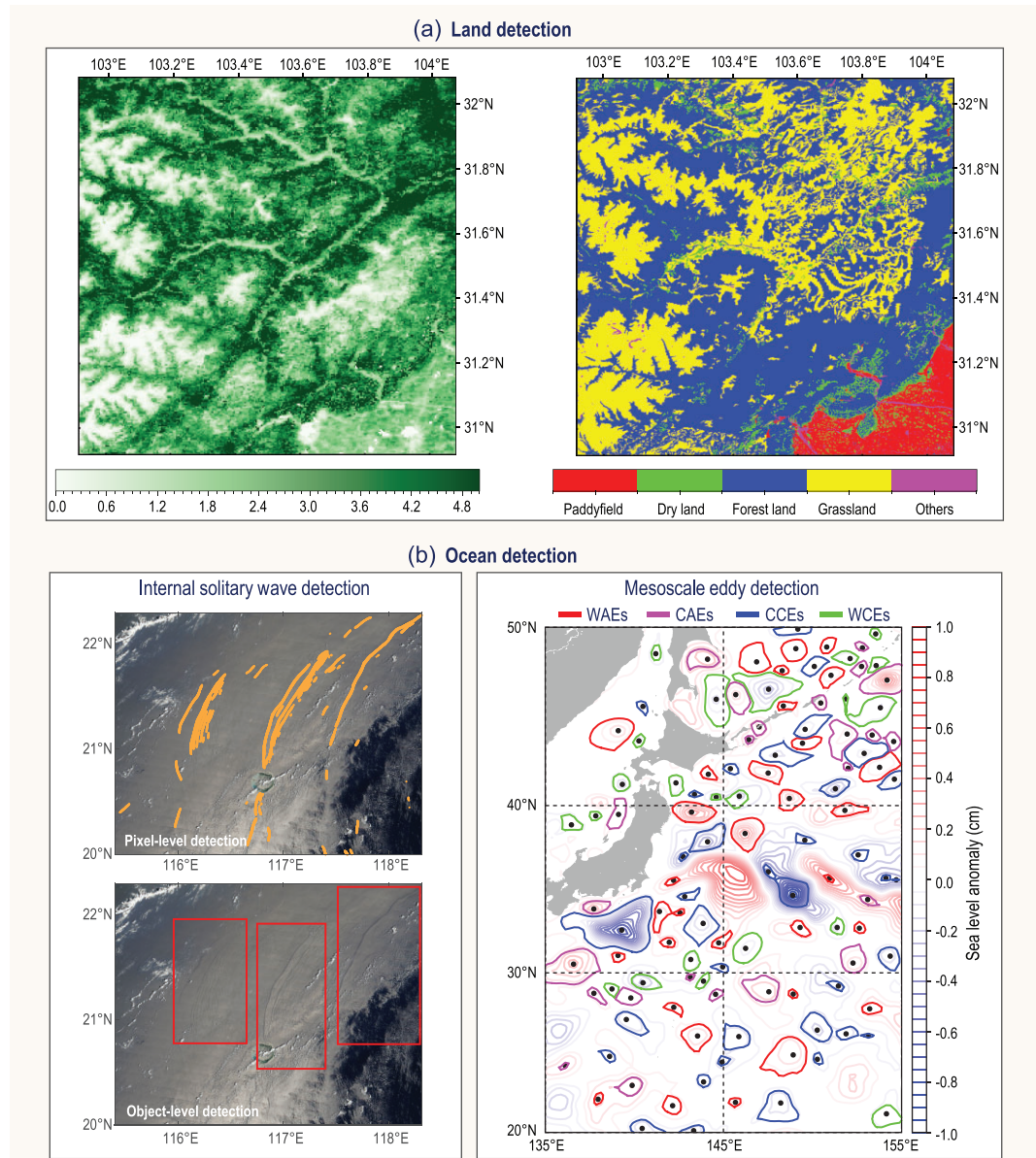


Figure 2. (a) AI-based land detection. Leaf-area index (left) derived by using an artificial neural network from MODIS and LULC classification (right) by using CNNs from Landsat, shown over a region in Sichuan Province, China (30.92°–32.08°N, 102.92°–104.08°E) on 1 June 2020. (b) AI-based ocean detection. Internal solitary waves (ISWs, left) detected at multiple levels: pixel-level and object-level detections. AI-based eddy-detection map (right) categorizes mesoscale eddies into four classes: warm anticyclonic eddies (WAEs, red contours), cold anticyclonic eddies (CAEs, magenta contours), cold cyclonic eddies (CCEs, blue contours) and warm cyclonic eddies (WCEs, green contours).

investigated with unprecedented precision and efficiency [45]. For instance, a number of DL architectures were constructed to successfully detect internal solitary waves at both the object level (bounding boxes) and the pixel level (wave-crest masks) (Fig. 2b, left panel) [46], achieving robust cross-sensor performance and providing a valuable data foundation for forecasting and dynamic analysis. A residual U-Net architecture was also employed to identify four classes of mesoscale eddies—warm-core anticyclones,

cold-core anticyclones, cold-core cyclones and warm-core cyclones [47]—uncovering thermodynamically ‘abnormal’ eddies often missed when using traditional techniques (Fig. 2b, right panel).

DATA RECONSTRUCTION AND ASSIMILATION

Obtaining high-resolution and physically consistent gridded fields is fundamental for atmosphere–ocean sciences. Applications such as regional

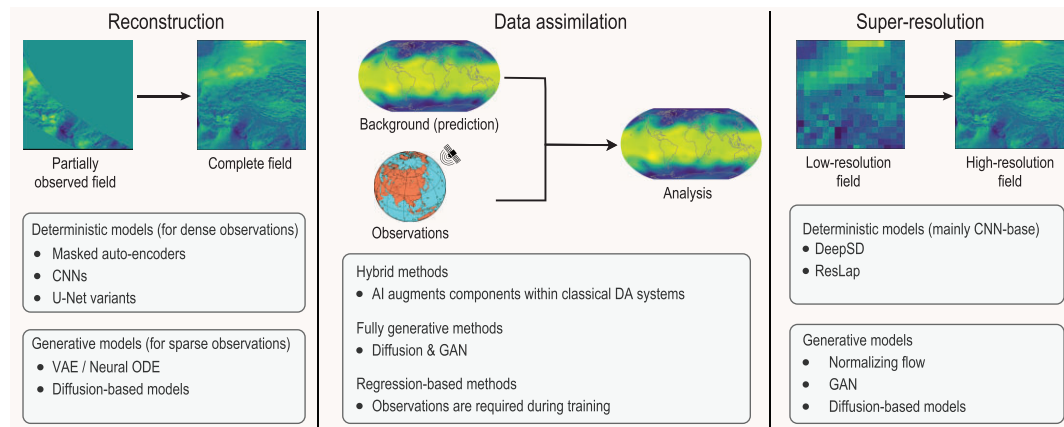


Figure 3. The three core scenarios of AI approaches for atmospheric and oceanic data processing.

weather and climate analyses, Earth-system and extreme-event prediction require accurate spatiotemporal representations of key state variables. Three core scenarios arise in practice: reconstruction from partially observed fields, super-resolution from coarse-grid simulations and DA integrated from model forecasts and observations (Fig. 3). Each presents distinct challenges and has spurred the development of increasingly sophisticated AI approaches.

Data reconstruction

Traditional methods, such as optimal interpolation or empirical orthogonal functions, rely on strong assumptions (e.g. linearity, stationarity, Gaussianity) and therefore often fail to recover spatiotemporally consistent structures. DL offers a more flexible alternative, especially in cases with relatively dense or structured observations. Deterministic supervised models learn direct mappings from masked fields to complete fields by using architectures such as masked autoencoders [48], partial convolutional networks [49] and U-Net variants with uncertainty estimation [50]. GANs have also been applied for inpainting and radar recovery [51,52]. Recent improvements include the use of auxiliary physical variables (e.g. wind, chlorophyll) [50] and transfer learning from pretrained climate models [7,53]. The LaMa DL model has recently been shown to outperform existing statistical and DL approaches for reconstructing global historical temperature fields [54].

More ill-posed settings arise when observations are sparse, irregular or multimodal (e.g. paleoclimate proxies, isolated sensors). They require generative models that approximate the distribution of plausible states. Most admit a two-stage paradigm: learning a prior via VAEs or neural ordinary differential equations, followed

by posterior sampling conditioned on sparse inputs [55]. Diffusion-based models extend this framework by incorporating observations during sampling via score-based guidance, as in S3GM [56] and related approaches [57].

DA

DA aims to generate dynamically consistent analyses by combining background forecasts with observations. Traditional techniques, such as 3DVar, 4DVar and ensemble Kalman filters, are grounded in physical principles, but depend heavily on sample-estimated background-error covariances, limiting the flexibility in non-Gaussian or nonlinear settings. Due to the similarity between optimizing DL networks and solving the analysis equation in variational DA methods—both relying on the gradient calculation of the loss function—it has been widely discussed that DL networks could mimic the DA process [58–60].

AI-based DA approaches fall into two paradigms: hybrid and fully generative. In hybrid methods, AI augments specific components within classical DA systems. A first class of hybrid approaches focuses on shared DA components that are common to both the 3DVar and the 4DVar frameworks. For instance, neural networks have been used to model background-error covariances [61] and satellite radiance observation operators [62], enabling more flexible and nonlinear representations while remaining compatible with variational DA formulations. A second class of hybrid approaches specifically targets challenges that are more prominent in 4DVar, such as flow-dependent background errors, adjoint models and the high computational cost of optimization over assimilation windows. AI-based methods have been proposed to improve flow-dependency modeling [63,64],

accelerate optimization schemes [63] and approximate adjoint models [65], thereby alleviating the computational and optimization difficulties inherent in 4DVar. In ensemble-based settings, AI has also been applied to improve Kalman gain estimation [66], covariance localization [67] and particle filter efficiency through score-based sampling [68,69]. Latent-space DA, which embeds system states in lower-dimensional manifolds, is also gaining traction [70–72]. However, fully data-driven DA methods are currently more often explored in 3DVar-like settings and their extension to 4DVar, in which flow-dependent dynamics play a central role, remains an open research challenge.

In the fully generative paradigm, AI models learn the posterior distribution of analysis fields directly from the paired training samples of forecasts, observations and analysis states. The U-Net architecture using fusion modules successfully assimilates satellite observations [73–75]. Diffusion models have become the dominant approach due to their ability to capture complex priors and incorporate observations through conditioning or score-based guidance [57,76,77]. GAN-based methods have also been explored [74,78,79], though diffusion models are now preferred due to their greater stability and physical consistency.

By directly using the automatic differentiation algorithm embedded in the AI-based DA, targeted observations that identify the most sensitive regions for improving forecast skill can be detected [80–82], thereby enhancing the overall quality of the AI-based DA.

AI PREDICTIONS AT DIFFERENT TIMESCALES

DL-based forecast models have begun to outperform current dynamical models, particularly in predicting atmospheric variations along with the occurrence/intensity of extreme events in weather and subseasonal timescales by learning complex, high-dimensional and nonlinear relationships, thereby mitigating forecast errors that often arise from simplifications in traditional physical parameterizations. By modeling land, ocean and sea-ice components and incorporating their relationship with the anthropogenic factors, DL models are anticipated to yield substantial improvements at longer climate timescales. Further advancements in DL methodologies, combined with the integration of physics-informed and meta-learning approaches, can help to address the data-scarcity and ‘black-box’ issues in building data-driven weather and climate models. These

innovations enable learning from limited data and contribute to the more effective optimization of DL models.

Nowcasting

Precipitation nowcasting—high-resolution forecasting of precipitation up to several hours ahead—supports the socioeconomic needs of many sectors. To leverage the advantages of DL models in terms of both predictive accuracy and computational efficiency, DL approaches have been introduced that directly predict future rainfall rates from radar observations with high spatiotemporal frequencies (1-km horizontal resolution with a temporal range from 5 min to 1 h). Although early stages of DL models for precipitation nowcasting accurately predict low-intensity rainfall, their operational utility remains limited due to the blurry nowcasts at longer lead times, resulting in poor performance in predicting heavy-precipitation events [83]. Recent DL models for probabilistic precipitation nowcasting from radar data tend to address these challenges by adopting stochastic generative models [84]. The unnatural motion and intensity in the predicted precipitation are mitigated by introducing an advection-based evolution network that incorporates horizontal-wind information [85]. In MetNet-3, precipitation-related variables from *in situ* observations are further utilized [86]. It achieves a realistic high-resolution structure of precipitation from sparse observations by introducing a process called densification. Namely, certain portions of station data are randomly dropped from the input of the network during training, while keeping these observations as targets, and the model is evaluated both through spatial generalization to unseen stations and through hyperlocal performance at observational locations.

Weather to subseasonal forecasts

In addition to regional DL-based models with a few kilometers’ resolution for precipitation nowcasting [87], various DL-based global models have been developed and have achieved superior accuracy in forecasting air temperature, humidity, geopotential heights and horizontal winds, by leveraging historical reanalysis datasets [9,88–98] (Fig. 4). Pangu-Weather based on Swin-Transformer architecture reported 10% lower root mean square errors (RMSEs) in predicting key tropospheric and surface atmospheric variables than the operational integrated forecasting system (IFS) of European Centre for Medium-

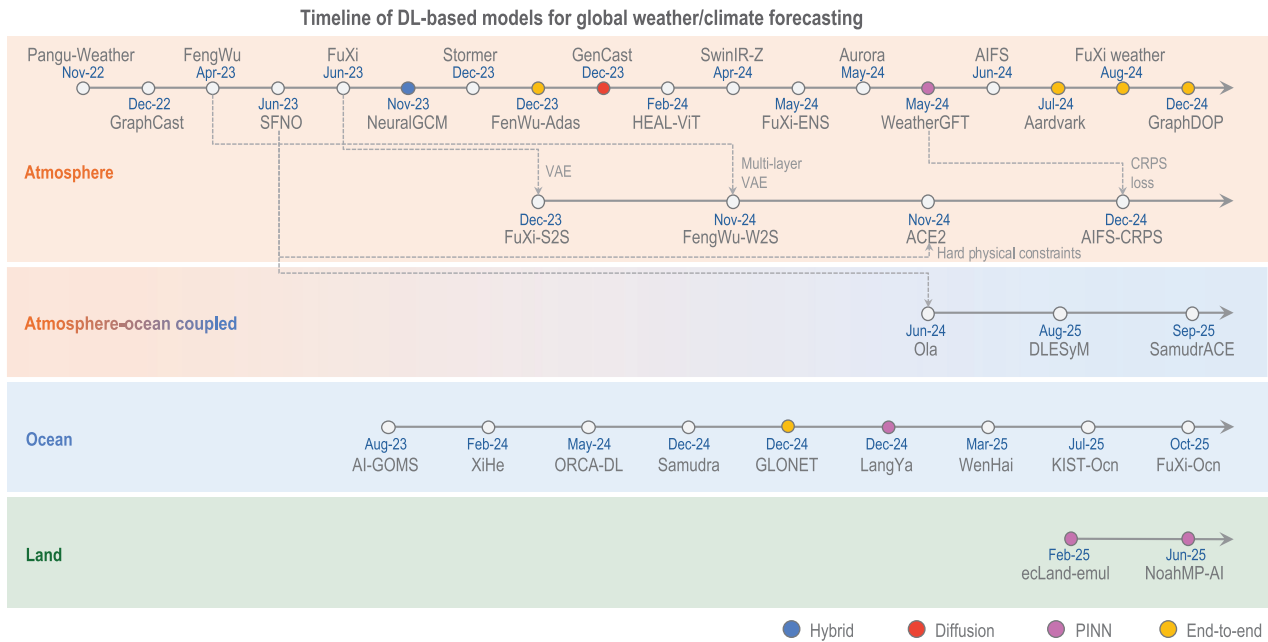


Figure 4. Timeline of DL-based state-of-the-art models for global weather and climate forecasts. Dates indicate the initial release date of each model, and AE, CRPS and PINN are abbreviations for autoencoder, continuous ranked probability score and physics-informed neural network (including physics-guided neural network), respectively.

Range Weather Forecasts (ECMWF), one of the most accurate operational dynamical model-based forecast systems [99]. GraphCast, a GNNs-based model, significantly outperforms the IFS on 90% of the verification targets and supports better prediction of severe events, including tropical cyclone tracking and atmospheric rivers [89]. A spherical multigrid neural operator that integrates spherical harmonic convolution and a spherical Fourier neural operator (SFNO) achieves a significant improvement in predicting lower and middle tropospheric variables with a lead of ≤ 7 days [100].

The capabilities of DL-based atmospheric models are further extended to cover broader forecasting objectives and longer time horizons. To estimate the uncertainty induced by various atmospheric instabilities, which allows a nuanced interpretation of model outputs and enables users to facilitate robust and risk-aware strategies, DL-based models such as FengWu, FuXi, ArchesWeather and GenCast incorporate diffusion-based generative algorithms into the deterministic DL architectures, harnessing the strengths of generative models [101–104]. Diffusion-based approaches and a noise-injection strategy with model ensembles [105] also effectively mitigate the blurry issue, enhance their ability to predict extreme-weather events (e.g. heatwaves) and improve uncertainty estimation. An integration of diverse data sources enhances the prediction

accuracy of the data-driven model by utilizing *in situ* and remote sensing measurements as well as reanalysis [106,107]. Diffusion-based approaches such as generative assimilation and prediction (GAP) have also been shown to be able to unify DA and weather prediction from medium range to seasonal cycle in one framework [35].

To extend forecast durations to subseasonal timescales, a common strategy is to reduce temporal resolution to prevent the accumulation of errors during the iterative forecasts [104,108]. Another way is to construct a loss as a function of forecasts on the subseasonal lead times. There is a multistep loss method in which predictions are performed repeatedly over multiple steps and the loss at all steps is calculated to update the model parameters [89]. For example, the training process of FuXi-S2S follows an autoregressive training regime and a curriculum training schedule, incrementally increasing the number of autoregressive steps from 1 to 17 days' lead [109]. As a result, FuXi-S2S achieves skillful daily mean predictions for ≤ 42 days and significantly extends the predictive skill for Madden-Julian oscillation (MJO). In a similar manner, FengWu applied a replay buffer mechanism, which stores the predicted results from previous optimization iterations and uses them as the input for the model for subsequent iterations. As it mimics the intermediate input error during the autoregressive

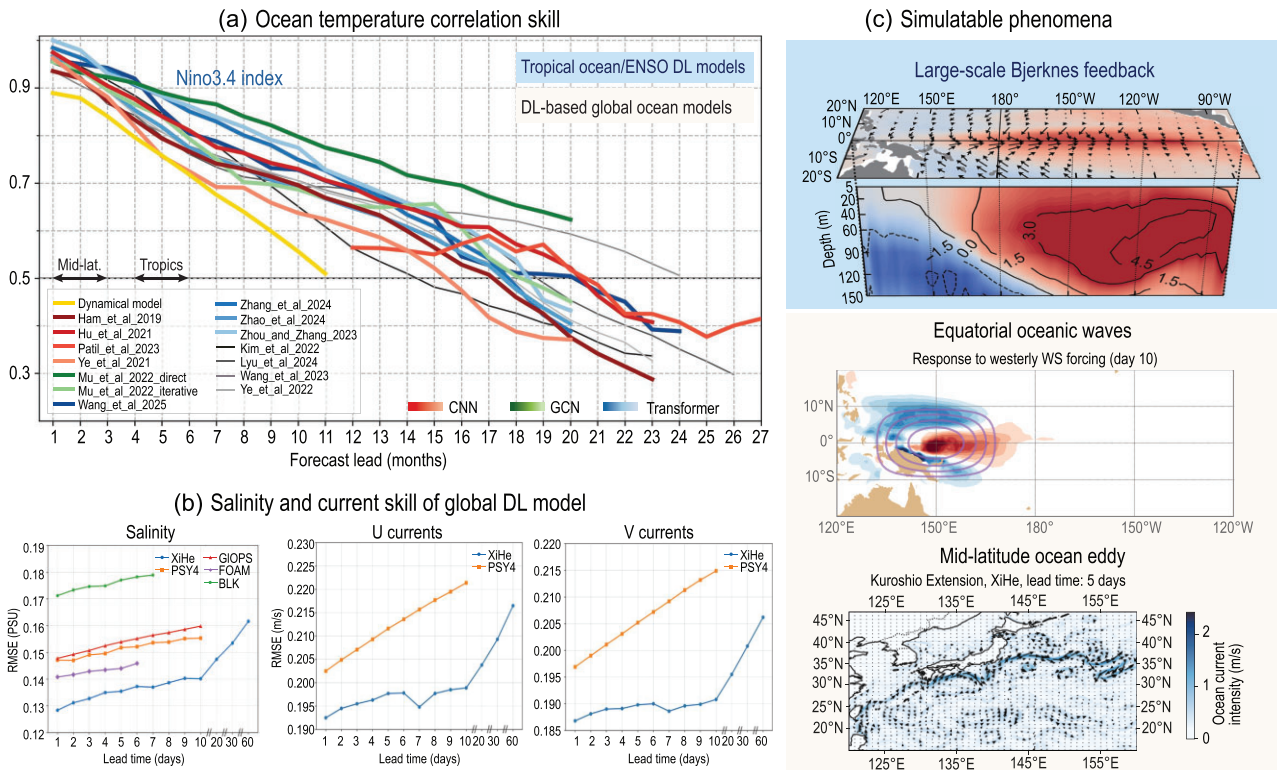


Figure 5. (a) Correlation skill of sea-surface temperature (SST) using various DL-based models. Each model is represented by a line showing the prediction skill of the Niño3.4 SST index, and models are grouped by their base DL architecture (CNN-based, graph convolutional network-based, or transformer-based). Arrows at the correlation threshold of 0.5 indicate the lead-time ranges over which the correlation skill exceeds 0.5 for predictions over tropical and mid-latitude regions using DL-based global ocean models. (b) RMSE of the global salinity and currents of the DL-based XiHe model with four operational numerical models (adopted from [111]). (c) Example of simulatable phenomena of DL-based model targeted to predict tropical ocean variabilities by using monthly fields (upper panel) and DL-based global ocean models (lower panel).

inference stage, it is known to be beneficial for long-lead forecasts with efficient computation and memory [101]. The probability-based loss function in Artificial Intelligence Forecasting System-Continuous Ranked Probability Score contributes to capturing the subseasonal variability in the atmospheric system while preserving the original temporal frequency [96]. The GAP framework also shows remarkable skill at subseasonal scales, including MJO prediction [35].

Enhancing the generalization and physical consistency of DL-based models, NeuralGCM combines a simplified, differentiable atmospheric dynamical core with AI-based physical parameterizations [13]. This hybrid architecture preserves the integrity of large-scale fluid dynamics while allowing the learned components to simulate sub-grid processes. If these DL-based parameterizations are modularized and categorized (e.g. NeuralGCM with a precipitation scheme [110]), the entire modeling becomes more transparent and interpretable, thus providing clearer guidance for future model development.

Seasonal-to-interannual forecasts

Given that oceanic processes are the primary drivers of climate variability on interannual to decadal timescales, data-driven global ocean models are expected to extend the predictive capabilities of current DL models to longer timescales [11,111]. Various DL models that aimed to predict dominant interannual variability in the tropical Pacific, namely El Niño Southern Oscillation (ENSO), by using architectures such as Vision-Transformer (ViT) [112], convolutional block attention modules [113] and graph convolutions [114] have extended predictive lead times by ≤ 2 years, thereby surpassing the performance of most advanced dynamical models and a pioneering CNN-based DL model [10] (Fig. 5a). A modified version of the aforementioned DL models has shown promising results in predicting dominant oceanic variabilities in the Indian Ocean [115], the tropical Atlantic Ocean [116,117] and Arctic sea ice [118,119], despite these regions being traditionally considered as less predictable

due to shorter-lived and weaker variability [115].

To improve the explainability and applicability of DL models, 3D data-driven ocean models have been developed, enabling more accurate simulation and prediction of dominant interannual and decadal variability with systematically reduced biases compared with those of the state-of-the-art dynamical models (Fig. 5b). DL ocean models, based on Swin-Transformer architectures [28], can resolve oceanic eddies [76] and tropical basin-scale atmosphere–ocean coupled processes [112], capturing both local and global atmospheric [120,121] and oceanic connections [111] (Fig. 5c). It is demonstrated that an SFNO-based [100] coupled atmosphere–ocean model successfully reproduces ENSO-related large-scale Bjerknes feedbacks and associated global atmospheric teleconnection patterns [122]. Through idealized experiments, it is further demonstrated that the DL ocean model is able to simulate realistic 3D thermodynamic responses associated with the equatorial oceanic waves (i.e. Kelvin and Rossby waves) driven by atmospheric wind forcings [11].

By taking the leverage of the DL model to keep any constraints by minimizing the loss function with the differential algorithm, a specified loss can be designed to seek the fast-growing perturbations, which is known to be optimal for ensemble forecasts [123]. As the fast-growing perturbations represent the initial patterns that most sensitively influence the target phenomenon, they can also be interpreted as a powerful tool for identifying optimal precursors and for conducting predictability studies [123–126].

To compensate for limited samples of data, DL models often incorporate synthetic training data from multidecadal simulations of dynamical climate models [10,11,127]. However, similar approaches have proven to be less effective in predicting mid-latitude climate variability [128,129]. This is primarily due to the limited reliability of dynamical model simulations in these regions, stemming from the complexity of their driving mechanisms [130], resulting in a lower signal-to-noise ratio in forecasts [131]. Instead of using simulations of dynamical climate models, meta-learning algorithms facilitate the rapid adaptation of DL models to new tasks by using limited data [132], making it possible to perform skillful mid-latitude climate forecasting based solely on the limited number of observational records [133,134].

Decadal prediction and climate projection

Climate projection simulates the impacts of varying external forcings, such as greenhouse gases and aerosols, over several decades under various scenarios to inform future policy decisions. AI-based global models typically adopt an autoregressive forecasting approach that performs iterative predictions at intervals ranging from as short as 1 h to as long as 1 month [11,13,100,135]. Consequently, conducting a 100-year simulation requires an autoregressive forecast that repeats $\geq 10\,000$ steps. Although data-driven models can qualitatively reproduce realistic dynamical properties across a certain number of time steps [136], errors can accumulate over successive prediction steps if fundamental physical constraints (e.g. conservation of energy) are violated. Moreover, concerns remain about whether AI models can capture climate responses to radiative forcing changes over multidecadal periods that far exceed their prediction timesteps.

To prevent physical violations and enable stable long-term simulations, AI models have been augmented with domain knowledge (i) by embedding physical constraints directly into their architecture and (ii) by creating a hybrid physics–AI model. Physics-informed neural networks apply soft physical constraints through loss-function penalties [137,138], while hard physical constraints enforce conservation laws through architectural design [135,138,139]. Examples include constraints from the momentum equation, continuity equation [140], thermodynamic relations [139], mass-conservation equation [140] and moisture conservation [135]. Sensitivity experiments have identified moisture processes as the primary cause of collapse in long-term simulations, suggesting that constraining moisture conservation is possibly more critical than constraining other physical relations [141–143]. Accordingly, enforcing global moisture conservation and condensation constraints was shown to substantially enhance the stability of long-term climate simulations [135,144]. Notably, it was demonstrated that, even when inference intervals (e.g. 6 h or 1 day) are short relative to the timescales of CO₂ forcing, AI models can still learn and reflect CO₂–induced climate responses [135].

Hybrid physics–AI models further address physical inconsistencies and support stable long-term integrations by deploying AI primarily in a supportive role. Key implementations include: integrating a dynamical core or process-based model with a neural network-based learnable physics

Table 1. Bias correction for dynamical model prediction at different timescales.

Timescale	Variables	Methods	Advantage	References
Weather forecast	Wind	ML and stacking methods	Effectively combine different foundational models for bias correction	[157]
	Precipitation	DL	Correct the bias in the precipitation prediction through considering different variables	[158]
	Temperature	DL	The performance of bias correction through the DL is better than that of the traditional method	[159]
Climate prediction	Temperature	Adaptive bias correction	Combines state-of-the-art dynamical forecasts with observations using ML	[163]
	Precipitation	DL	Reduced bias in Western Pacific Subtropical High (WPSH) and SST is critical to enhancing forecast accuracy	[160]
	MJO	DL	Blend the dynamical forecasts and observations to improve the accuracy of MJO prediction	[164]
Climate projection	Precipitation	GNNs	Distinguish historical climate simulation samples and observation samples	[167]

module [13,141,144–146], adapting differentiable modeling frameworks to jointly optimize physical and neural network components [43] and employing AI for systematic bias correction in dynamical model forecasts [147].

Despite these advances, NeuralGCM still yields unrealistic outputs under extreme-warming scenarios (+4 K) [13], indicating limits to the extrapolation beyond training regimes. Climate-invariant transformations (e.g. from specific humidity to relative humidity) to stabilize probability density functions across regimes [138] and deep stable learning techniques that remove spurious correlations to focus on intrinsic features under distribution shifts [148] can help mitigate the climate extrapolation challenge under unseen extreme conditions.

Decadal climate prediction bridges seasonal forecasting and long-term projections by targeting 1–10 years of variability and hence requires the representation of internal modes such as the ENSO, Pacific Decadal Oscillation (PDO) and Atlantic Multi-decadal Oscillation (AMO) [149–151]. Map-to-index AI models targeting specific decadal variability, such as PDO and AMO, have achieved skillful multiyear to multidecadal predictions [152,153]. Beyond this, recent map-to-map AI ocean models have demonstrated robust predictive skills for decadal variabilities for up to multiple years [11].

POST-PROCESSING OF DYNAMICAL MODEL PRODUCTS

To mitigate the limitations of numerical weather predictions (NWP) and climate forecasts of dynamical models induced by coarse resolutions,

non-perfect parameterization schemes and initializations [154], post-processing approaches have been widely applied, such as correcting bias and the downscaling of temperature, precipitation and wind forecasts [155]. Recently, AI has played a key role in post-processing due to its ability to depict nonlinear relationships and it efficiently improves the prediction skills of dynamical models at different scales (Table 1).

AI bias correction for dynamical weather forecasts

By distilling patterns from vast archives of observations and model outputs, machine-learning techniques have been utilized to compensate for dynamical model biases and generalize across diverse weather and climate regimes, delivering noticeable gains in both accuracy and reliability. Early studies demonstrated that random forests and support-vector regressors, when trained by using a direct-mapping strategy, outperform classical statistical post-processing schemes for short-term wind-speed forecasts [156]. Furthermore, ensemble strategies such as stacking generalization have attracted growing interest for their ability to blend complementary strengths and suppress residual errors [157]. Similar AI-driven advances are now being extended to precipitation and temperature prediction, signaling a broader paradigm shift in operational forecasts [158,159] (Table 1).

AI bias correction for dynamical climate simulation and prediction

AI approaches have also been applied to correct biases in dynamical climate prediction. Compared

with weather forecasts, bias correction with AI in climate prediction is more difficult due to the lack of sufficient samples [155]. Nevertheless, ML and DL have helped to improve the seasonal prediction skills for precipitation over China [160], displaying obvious advantages over the conventional quantile mapping method [161]. The results suggest that reduced biases in predicting the Western Pacific Subtropical High and sea-surface temperatures is critical to enhancing the precipitation forecast accuracy [160], consistently with the well-recognized mechanism. Data-driven methods also help to optimize the weights of ensemble prediction [162]. Regarding the subseasonal to seasonal (S2S) prediction, in which MJO provides a primary predictability source but has not yet been well predicted in dynamical forecasting systems, the ML and DL bias-correction methods have been applied to blend the dynamical forecasts and observations to improve the accuracy of temperature, precipitation and MJO prediction [163,164] (Table 1).

The ML and DL methods have also been utilized to improve long-term climate simulations made by dynamical models. A spatial-based 2D Conv-LSTM network was proposed to correct the bias in spatial dynamics and precipitation products from regional climate models (RCMs) [165]. Furthermore, online corrections can be integrated into a model running cycle as well, allowing subsequent cycles to benefit from previous improvements. For instance, CNN online bias-reduction models were developed to create a corrective model parameterization, improving the accuracy of precipitation and MJO simulations [166]. To mitigate the weakness that existing approaches usually focus on limited low-order statistics or break the spatiotemporal consistency of the target variable, a regularized adversarial domain adaptation methodology was developed to overcome these deficiencies, enhancing the efficient identification and correction of climate model biases [167] (Table 1).

AI downscaling for dynamical model outputs

Traditional approaches include interpolation (e.g. bilinear, spline), statistical disaggregation and downscaling based on climatological regression and analog matching. However, these methods either assume fixed spatial relationships and thus exhibit limited generalizability or suffer from high computational costs. AI-based downscaling aims to enhance the spatial resolution of geophysical fields by learning mappings from

coarse-resolution simulations to high-resolution representations.

Recent advances in AI have enabled more flexible and data-driven resolution enhancement by learning nonlinear mappings from coarse to fine scales (Fig. 3). Deterministic supervised models, such as DeepSD [168] and ResLap [169], employ CNNs for efficient upscaling, but often underestimate fine-scale variability and associated uncertainties. To address these limitations, generative models have been increasingly adopted to sample from the conditional distribution of high-resolution fields. Representative approaches include normalizing flows [170], GAN-based architectures [171,172], progressive and multiscale GAN variants [173,174], as well as more recent diffusion and consistency models [175–178]. Additional architectural components, such as attention mechanisms, have further improved performance, particularly in capturing spatial heterogeneity and extreme events [179]. These AI-based super-resolution and downscaling methods have been successfully applied across a wide range of weather and climate applications, including precipitation and temperature downscaling, climate field reconstruction and the refinement of dynamical model-generated products [180–182].

EXPLAINABLE AI

Though AI has been widely used in atmosphere–ocean sciences, the intrinsic ‘black-box’ nature of AI raises an indispensable double crisis [183], i.e. it completely shields the decision-making process from inferences and shakes the foundation of the scientific community’s trust in AI to a large extent. To address this challenge, XAI technology [184] attempts to answer fundamental questions, such as ‘How does AI effectively extract physical laws from data?’, by systematically dismantling the key components of AI models. In principle, applications of XAI in atmosphere–ocean modeling can be classified into three types (Fig. 6) [185,186].

First, inherently interpretable AI models such as linear regression, generalized additive models, decision trees and rule-based systems are commonly used due to their ability to display the way in which decisions are made (Fig. 6a). However, they often sacrifice predictive performance compared with more complex models such as deep neural networks [187]. Note that these types of methods are sometimes considered not as XAI, but as so-called interpretable AI.

Second, well-developed geoscience theories and physical mechanisms can guide AI modeling for better feature extraction from large datasets,

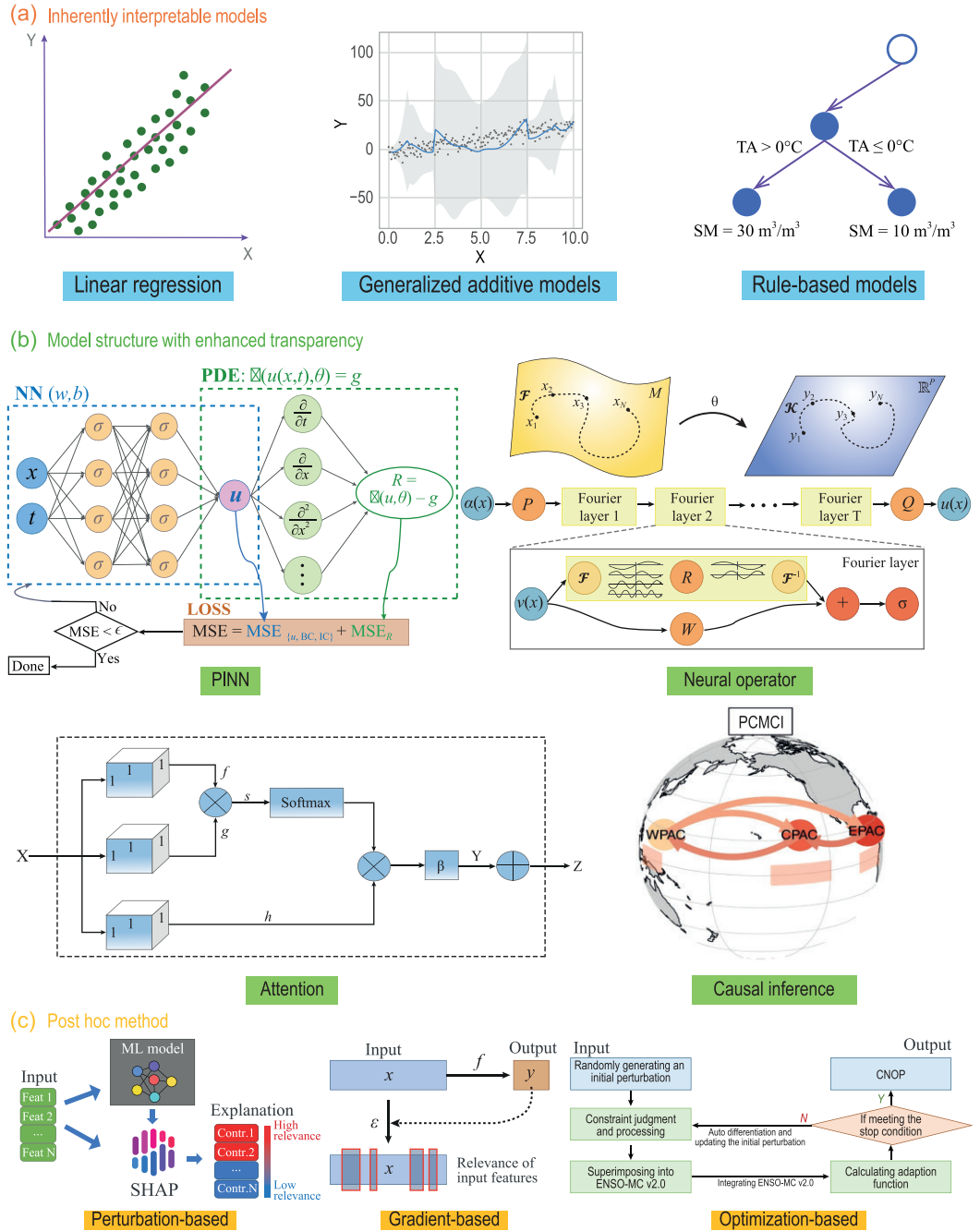


Figure 6. Three types of XAI technologies, including (a) inherently interpretable models, (b) model structures with enhanced transparency and (c) post-hoc method.

which makes the model more explainable. Based on many successes, this has gradually become a modeling statute (Fig. 6b). Physics-informed neural networks are recognized as one of the most explainable architectures and they directly incorporate PDEs into neural layers and loss functions [188]. Attention layers are beneficial in revealing element-level characteristics, such as spatiotemporal and variable domains [189–191], and neural

operators [192,193] facilitate the capture of variability at different temporal frequencies. It is also noteworthy that causal neural networks, which combine the rigorous causal inference theory with AI modeling [194,195], are effective tools for mining knowledge from data. Time-lag or teleconnection information between multiple physical variables and processes can be efficiently expressed in the form of causal diagrams.

Third, utilizing mature post-hoc explainable methods in a well-trained AI model can derive significant scientific findings, which helps to make black-box models transparent (Fig. 6c). These methods can be generally grouped into four categories. The first two types of methods are called model-agnostic, while the other two are model-specific. The first type is a surrogate approach. Global surrogates approximate the entire behavior of a black-box model by using inherently interpretable AI models such as decision trees, while local surrogates explain the predictions for individual instances (e.g. local interpretable model-agnostic explanations [196]). The second type is a perturbation-based attribution approach. Shapley additive explanation [197] is the most representative, which effectively detects the contributions of individual features or well-designed feature groups by AI model input modifications [198]. The third type is a gradient-based attribution approach [199] and is the most widely used method for DL models. With the help of the automatic derivation of AI models, one can easily explore the gradient of various objective functions with respect to the input to generalize physical meanings [115,200]. The fourth type is the optimization-based attribution approach, which overcomes the strong assumption of linearity and small perturbation magnitude in the first two types of approaches [201]. Conditional nonlinear optimal perturbation [202] is the most effective method, which takes advantage of the strong nonlinear characteristics of AI models and uses optimization to reveal the initial perturbations that most affect the model skill [125].

It should be emphasized that the effectiveness of the XAI technology is strictly dependent on the high performance of AI models [201]. Physical inconsistency may arise if there are fatal systematic biases in AI models. For example, AI weather-forecasting models trained on present-day climate data exhibit systematic biases, such as a global-mean cold drift, when applied to future warmer climate scenarios, causing physically inconsistent interpretations of climate dynamics [203]. In addition to the structure of AI models, the construction and selection of high-quality data used in AI modeling—such as ensuring relevance, representativeness, physic consistency and fairness—are critical for enhancing the robustness and reliability of explanations derived from XAI. It should be noted that different AI models and different XAI methods may offer inconsistent explanations. As a result, the derived explanations should be carefully examined among different approaches. The application of XAI in Earth-system science is still in its infancy and tremendous efforts are needed

to develop co-designed frameworks that integrate domain knowledge, causal inference and human-centered interfaces [186].

HYBRID PHYSICS-AI MODELING

Traditional NWP systems rely on PDEs that describe atmospheric dynamics, thermodynamics and various physical processes. However, these models are often computationally expensive and sensitive to both initial conditions and model resolution. Meanwhile, purely data-driven models, although fast and flexible, often lack fundamental physical constraints, making their extrapolation to out-of-sample conditions questionable. To address these limitations, hybrid physics-AI methods integrate the best of both worlds: the interpretability and rigor of physical models with the adaptability and predictive power of modern ML.

Physics-informed machine learning (PIML) exemplifies this synergy by directly incorporating physical equations into the neural network training process [204,205]. In weather and climate contexts, PIML can embed dynamical and thermodynamical equations into the loss function so that the model learns to satisfy physical constraints while fitting the observed or simulated data [206]. Alternatively, methods to directly modify learned dynamics use stabilization techniques to ensure that conservation laws have been introduced for chaotic dynamics [207]. This explicit blending of data and physics aims to reduce the amount of data required and improve generalization, as the network is guided not just by statistical correlations, but also by known laws of fluid dynamics and thermodynamics (Fig. 7).

AI is now transforming our capability to rapidly and accurately simulate complex multi-scale processes, particularly in weather and climate prediction. Previous AI methodologies, such as sparse representation, RNNs, reservoir computing and related approaches, have demonstrated significant potential for modeling dynamical systems [208,209]. Recent advancements, particularly neural operators, further enhance this potential by providing a generalized framework for interpreting and modeling these complex phenomena at finer scales, resulting in high-fidelity predictions. Neural operators generalize traditional neural networks by learning operators—mappings between input and output functions defined over continuous domains [27,152]. Unlike traditional numerical approaches, neural operators are not limited to explicitly solving PDEs; instead, they learn representations directly from data, capturing underlying physics implicitly through obser-

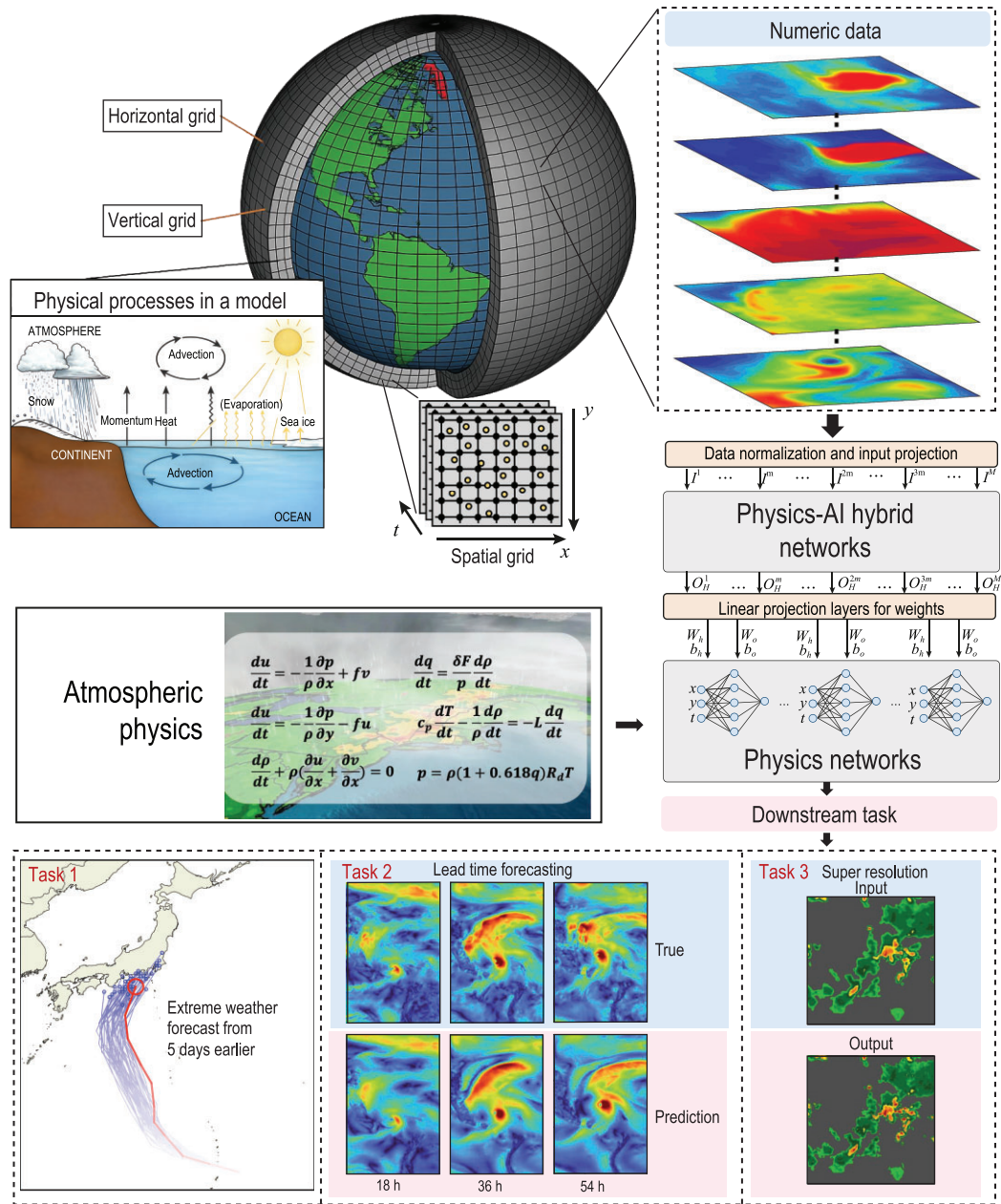


Figure 7. Workflow of the hybrid physics-AI system and results. This figure presents a composite overview of recent state-of-the-art hybrid physics-AI weather-forecasting models, including GenCast, WeatherGFM, DeepPhysiNet and FourCastNet. It integrates representative workflows and selected results from these models to effectively illustrate the architectural structure and core operators of the hybrid modeling approaches. Subfigures are adapted from GenCast [105], DeepPhysiNet [205], and FourCastNet [192] with modifications.

vations or simulations. That is, neural operators encompass many data-driven weather and climate models, inherently learning mappings between physical states across various temporal and spatial scales, and therefore can avoid inaccuracies resulting from simplified equations or incomplete physical representations [210–212]. This capability enhances their applicability in practical forecasting scenarios, in which observational data

often include complex phenomena beyond the scope of numerical models. Thus, neural operators generalize better to out-of-sample conditions and unforeseen extreme events, improving predictive robustness. Many data-driven forecasting models can be effectively understood within this broader neural operator framework, highlighting their flexibility and efficacy in tackling complex weather- and climate-prediction challenges.

Among neural operators, Fourier neural operators (FNOs) are widely utilized [8]. This technique leverages the powerful representation capabilities of Fourier transforms, capturing spatial correlations across multiple scales. In weather and climate forecasting, FourCastNet utilizes the framework of FNOs to achieve a highly efficient, data-driven approach while leveraging physical insights [192]. FourCastNet uses spherical convolutions or spectral transformations to ensure that predictions respect global properties on Earth's spherical geometry. WeatherGFT proposes a hybrid physics–AI model that generalizes weather forecasts to finer-grained temporal scales beyond the training dataset [213]. More precisely, they employ a carefully designed PDE kernel to simulate physical evolution on a small timescale (e.g. 300 s) and use a parallel neural network with a learnable router for bias correction. Furthermore, WeatherGFT introduces a lead-time-aware training framework to promote the generalization of the model at different lead times. The weight analysis of physics–AI modules indicates that physics conducts major evolution while AI performs corrections adaptively.

NeuralGCM and ClimODE are further examples of hybrid frameworks that integrate physical principles into DL architectures [13,214]. These models often start with a structure inspired by general circulation equations—describing fluid motion on rotating Earth—and embed trainable components that learn the subgrid-scale processes or parameterizations. By respecting conservation laws and large-scale atmospheric balances, NeuralGCM can focus their learning capacity on complex physical parameterizations (e.g. cloud microphysics, radiation or ocean–atmosphere coupling). Similarly, ClimODE seeks to merge neural ordinary differential equation structures with convective dynamics, ensuring that the physical backbone remains intact. These strategies reduce the risk of unphysical results while harnessing AI for tasks such as parameter estimation, downscaling or DA.

A crucial motivation for these hybrid physics–AI approaches is that purely data-driven methods may fail when encountering conditions beyond their training distribution. Extreme-weather events, long-term climate shifts and complex interactions between land, ocean and atmosphere processes are difficult to capture without explicit physical constraints [41]. By integrating domain knowledge directly into AI architectures, one can reduce overfitting and ensure more robust extrapolation. Furthermore, hybrid physics–AI methods often require fewer labeled data points than purely

data-driven approaches, as the physical components guide or regularize the learning process [215]. This synergy can be particularly beneficial in regions or at timescales with sparse observations.

However, hybrid physics–AI models face several challenges. Foremost among them is the complexity of integrating high-fidelity physical equations into trainable architectures without making them computationally unwieldy. Most PIML methods may struggle when the domain is complex or when the underlying dynamics involve significant turbulence or stochasticity [216]. Balancing accuracy and efficiency often requires the careful design of model architectures, solvers and DA schemes. Another issue is interpretability: while physics-based equations are transparent, neural network representations are not. One must ensure that the hybrid designs remain explainable enough to build trust with end users [217].

AI-MIP (AI-BASED MODEL INTERCOMPARISON PROJECT)

In recent years, AI-based weather and climate foundation models have demonstrated advanced forecasting capabilities, but with critical challenges remaining unresolved, as discussed above [218]. Standardized benchmarks with high-quality data, such as ImageNet [219] and COCO [220], have driven progress by providing structured datasets, well-defined objectives and rigorous evaluation protocols. Similarly, comprehensive benchmarks are also crucial to the development of weather and climate AI models by (i) broadening participation, (ii) enabling systematic comparison, (iii) ensuring reproducibility, (iv) assessing model applicability, (v) tackling model limitations and (vi) guiding future directions (Fig. 8).

AI-AMIP (AI-based atmospheric model intercomparison project)

Current benchmarks for AI-based atmosphere models can be classified by forecast horizon, forecasting approach, data modality and application focus.

Regarding forecast horizon, benchmarks cover nowcasting, medium-range, subseasonal and climate-scale forecasting. Weather4cast targets precipitation nowcasting by using satellite data [221], while the AI for Nowcasting Pilot Project (AINPP) of the World Meteorological Organization (WMO) focuses on operational nowcasting

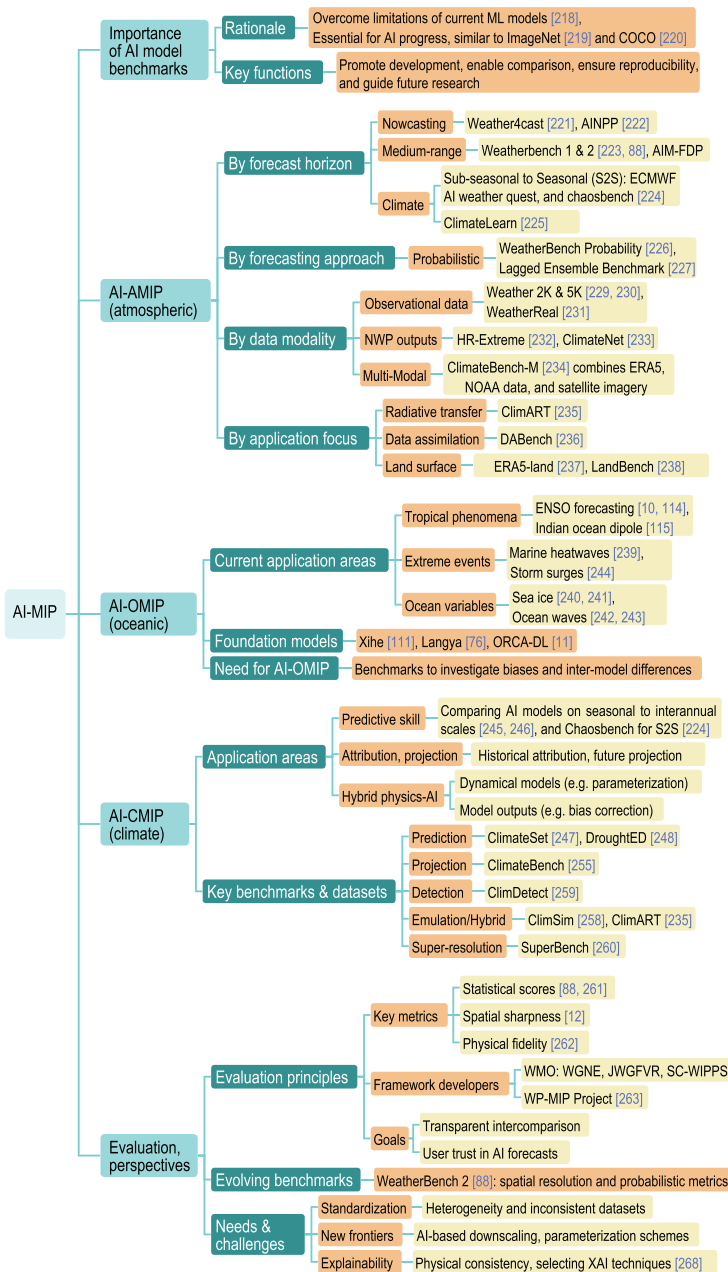


Figure 8. Schematic diagram of the AI-MIP (Model Intercomparison Project).

[222]. For medium-range (1–15 days) forecasting, WeatherBench [223] and WeatherBench2 [88] provide standardized datasets and evaluation metrics. The AI Model Forecast Demonstration Project (AIM-FDP) of China Meteorological Administration supports the operational deployment of AI-based medium-range weather-forecasting models. For S2S forecasting, ECMWF's AI Weather Quest and ChaosBench [224] foster AI-based S2S advancements. ClimateLearn provides an open-source library for climate AI research [225].

Forecasting approaches include deterministic and probabilistic methods. While early benchmarks focused on deterministic metrics, recent benchmarks, such as WeatherBench2, WeatherBench Probability [226] and Lagged Ensemble Benchmark [227], incorporate probabilistic scores such as the continuous ranked probability score, Brier score and spread-skill ratio.

In terms of data modality, benchmarks utilize reanalysis (e.g. ERA5), NWP outputs, satellite observations and *in situ* measurements. Despite its popularity, ERA5 has known biases, particularly in precipitation [228] and tropical cyclone intensity [102]. Benchmarks such as Weather2K [229], WEATHER-5K [230] and WeatherReal [231] integrate observational datasets, while HR-Extreme [232] and ClimateNet [233] include NWP outputs. ClimateBench-M [234] combines ERA5, NOAA extreme-weather-event data and satellite imagery.

Application encompasses general-purpose forecasting and specialized tasks, including extreme-event prediction, radiative transfer (e.g. ClimART [235]), DABench [236]) and land-surface prediction (ERA5-land [237], LandBench [238]).

AI-OMIP (AI-based ocean model intercomparison project)

AI has also been applied to predict tropical air–sea coupled modes [10,114,115], mesoscale eddies, marine heatwaves [239], sea ice [240,241], ocean waves [242,243] and storm surges [244]. Recent foundation models [11] outperform conventional statistical and dynamical models [239]. The growing diversity of AI-based ocean-forecasting models underscores the need for an AI-ocean-model benchmark, with unified metrics and standardized datasets to systematically assess model performance and biases as well as forecast uncertainties.

AI-CMIP (AI-Climate Model Intercomparison Project)

The AI-based Climate Model Intercomparison Project (AI-CMIP) can be grouped into predictive-skill assessment, climate attribution/projection and a hybrid physics–AI model.

For predictive-skill assessment on seasonal-to-interannual timescales [245,246], model heterogeneity in architectures, training data and evaluation protocols complicates universal benchmarking. ClimateSet [247] offers large and consistent datasets derived from 36 Input4MIPs and CMIP6

models. Specialized datasets further support applications for drought [248] and extreme-weather events [249].

For attribution and projection, AI offers efficient alternatives to computationally intensive ESMs [250–254]. ClimateBench [255] establishes a unified framework for evaluating both dynamical and AI models, and is applied broadly in climate tasks [256,257]. ClimSim [258] and ClimDetect [259] provide large-scale datasets for detection and attribution.

Hybrid physics–AI models are increasingly applied for physical parameterization, bias correction, downscaling and post-processing. SuperBench [260] provides high-resolution datasets, though comparable climate downscaling datasets remain limited. ClimART [235] offers large-scale benchmark data for atmospheric radiative transfer.

Guiding principles for evaluation protocol and metrics

In contrast to conventional physics-based models, AI models are trained on data rather than governed by physical equations. This fundamental paradigm shift introduces distinct challenges for model evaluation and intercomparison, particularly in assessing the robustness, generalization, stability, physical consistency, dynamical balance and conservation properties of AI models [261]. Therefore, evaluating AI-based weather and climate foundation models must go beyond standardized benchmarks and include a broader set of verification diagnostics. These should combine traditional statistical skill scores [88,223], measures of spatial sharpness or smoothness [12] and targeted tests of physical consistency and process fidelity [262]. This multidimensional evaluation framework is critical to ensure that AI models are not only skillful, but also physically credible and reliable for scientific and operational use.

To address these needs, WMO expert groups, including the Working Group on Numerical Experimentation (WGNE), the Joint Working Group on Forecast Verification Research (JWGFVR) and the Standing Committee on WMO Integrated Processing and Prediction System (SC-WIPPS), are developing evaluation frameworks through the Weather Prediction Model Intercomparison Project (WP-MIP) [263] to enable fair and systematic comparison of AI-based, physics-based and hybrid forecasting systems, while explicitly accounting for the unique characteristics and limitations of AI models. Establishing such standards is essential for transparent intercomparison with traditional

approaches and for building user trust in AI-based forecasts.

Perspective

As AI-based foundation models advance, benchmarks must evolve accordingly. For instance, WeatherBench2 improves upon its predecessor by enhancing the spatial resolution to 0.25° and incorporating probabilistic metrics, reflecting the superior probabilistic forecasting performance of AI relative to NWP models [13,104,264]. Evaluating AI-based models requires domain-specific expertise [265,266]. To promote broader participation, open-source evaluation platforms such as WeatherBench2 promote broader participation by collecting models for standardized evaluation.

Despite progress, ocean and climate AI models remain highly heterogeneous, complicating unified benchmarking. While ChaosBench and ClimateBench provide important foundations, further efforts are needed to rigorously assess model skill, physical consistency and explainability [267], particularly for downscaling and physical parameterization.

USEFULNESS AND APPLICABILITY OF AI MODELS TO SOCIETY

AI is emerging as a disruptive technology, profoundly reshaping various societal sectors by offering high-resolution forecasts at significantly lower computational costs [268]. AI-driven systems excel at handling high-dimensional datasets and identifying subtle patterns among weather variables, which enables them to deliver faster and more energy-efficient forecasts. These advancements provide direct benefits to critical sectors. In the energy industry, AI offers precise generation forecasts for wind and solar installations, meeting their unique operational needs [269], and demonstrates superior performance in both accuracy and efficiency [270]. In agriculture, AI assists farmers in optimizing planting schedules and crop selection based on weather outlooks, thereby enhancing productivity and sustainability. Furthermore, leveraging the integration of meteorological and geospatial foundation models through AI can effectively enhance the development of multi-hazard early-warning systems [271]. The global market for AI-based climate modeling is expanding rapidly, projected to grow at a compound annual growth rate of > 23% between 2025 and 2034, reflecting increasing investment and adoption by governments and industries seeking data-driven strategies for climate resilience.

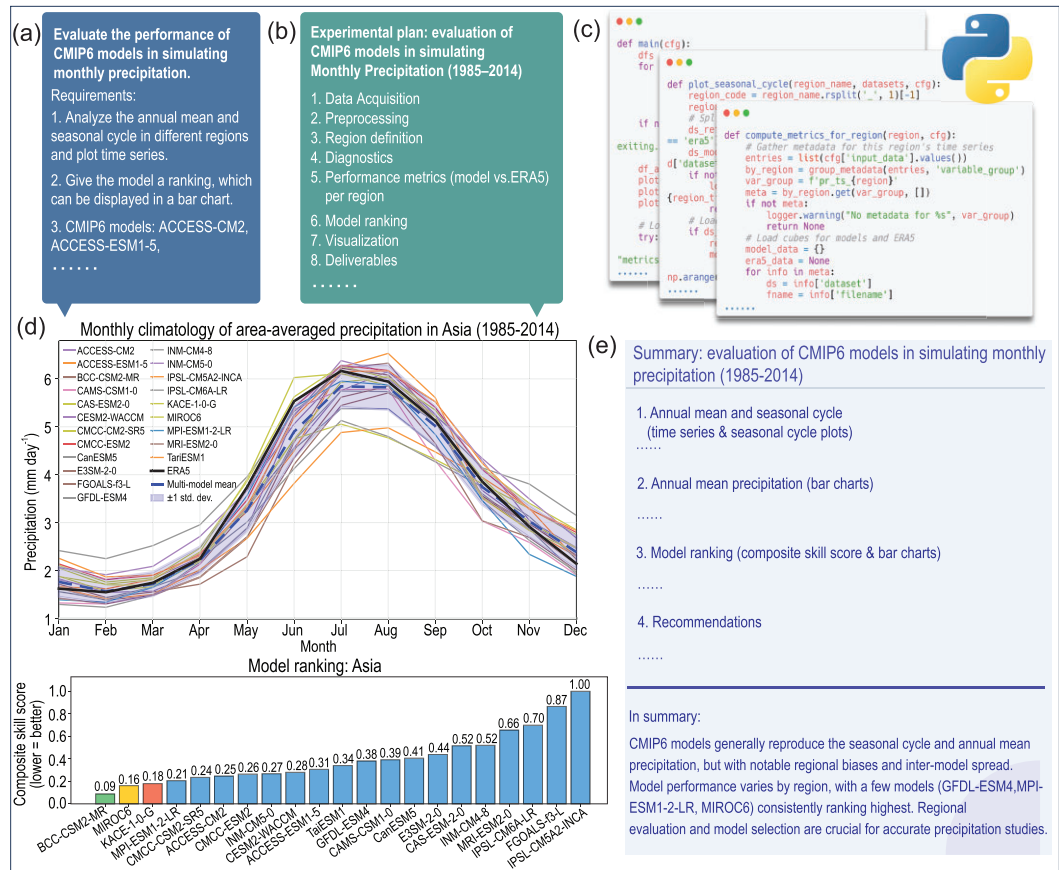


Figure 9. Evaluation of CMIP6 models in simulating seasonal cycles of monthly precipitation in Asia by EarthLink. (a) Task definition and diagnostic requirements, including annual mean and seasonal cycles and regional time-series analyses. (b) Automated planning output from EarthLink, detailing end-to-end workflow from data acquisition to deliverables. (c) Example code snippet generated by the system for data processing and analysis. (d) Modeled and observed precipitation seasonal cycles over the selected region, with annual mean ranking and model performance comparison. (e) Automated textual interpretation of the results, providing a plain-language summary generated by the system.

However, the widespread application of AI models in atmosphere–ocean sciences faces significant challenges that temper their operational adoption. A primary concern is the reliability of the models in forecasting unprecedented extreme events or ‘gray swans’ [272]. Furthermore, specific applications face data limitations. For instance, many radar-data-based AI nowcasting systems cannot provide a comprehensive, high-resolution view of all relevant factors. Ensuring equitable access and governance also remains a critical challenge, as gaps in digital infrastructure and the presence of algorithmic bias can limit the effectiveness of these models, particularly in the southern hemisphere [273].

To fully unlock the societal potential of AI, future development must address these challenges. A critical step is the systematic comparison of traditional NWP and AI models to better understand how their fundamental differences in effi-

ciency and methodology impact forecast outcomes [272,274]. Moreover, it is crucial to move beyond statistical accuracy and quantify the tangible value that AI improvements provide for end users. A more statistically accurate forecast does not automatically translate into a better decision [272], highlighting the need to co-design and evaluate AI solutions with stakeholders in sectors such as energy, transportation and agriculture to measure their real-world impact. This involves developing robust, accountable and fair AI pipelines that support evidence-based policymaking and operational use. Despite these hurdles, the market trend shows a clear and growing integration of AI into operational weather services worldwide, signaling that their immense potential is recognized. This trajectory points toward a new era of intelligent forecasting, in which data-driven insights and physical laws are deeply intertwined to build a more resilient and sustainable society [269].

AI AGENTS FOR CLIMATE SCIENCE (AS ONE GROUND-BREAKING WAY FORWARD)

AI is transforming science, not only by accelerating simulations and predictions, but also by acting as a genuine partner in discovery. In materials science, autonomous laboratories guided by intelligent agents are discovering new compounds and catalysts [275]. In life sciences, general-purpose biomedical agents and virtual laboratories of AI systems are designing experiments, proposing hypotheses and even discovering novel therapeutics such as nanobodies [276,277]. These advances highlight a broader trend toward the emergence of the ‘AI scientist’, in which machines evolve from analytical tools into collaborative partners that support and extend human intellect.

Earth science is beginning to follow this trajectory. The central challenge in this domain is not simply to improve prediction from ever larger datasets, but to achieve a deeper, human-like understanding of the physical world, especially when working with sparse and heterogeneous observations. Meeting this challenge requires moving beyond pattern recognition toward genuine scientific reasoning. A step in this direction is represented by EarthLink—an AI agent system designed as an interactive copilot for climate scientists [278]. Researchers can pose questions in natural language and EarthLink interprets intent, generates experimental plans, writes analysis code and synthesizes results into transparent and traceable research reports, such as evaluating the annual precipitation cycle in Asia (Fig. 9). By learning from successful cases through a practice–learning–improvement cycle, EarthLink continuously evolves to address increasingly complex problems. Rather than replacing researchers, it elevates their role from data processors to strategic directors of inquiry, enabling them to focus on hypothesis generation and creative exploration.

Looking ahead, the most powerful advances will come from integrating predictive models with reasoning agents. State-of-the-art forecasting systems, which already achieve remarkable accuracy, can be coupled with agents such as EarthLink to form a forecast-to-action framework in which AI not only predicts events, but also investigates their underlying drivers, evaluates potential impacts and proposes mitigation strategies. Such a synthesis would represent a profound step toward truly collaborative science, allowing humans and machines to jointly confront the complex challenges of a changing climate.

ACKNOWLEDGEMENTS

We thank Prof. Huijun Wang for his kind recommendation for this review article and three anonymous reviewers for their constructive comments and suggestions that improved the original manuscript. We also acknowledge the High Performance Computing Center of Nanjing University of Information Science & Technology for providing computational support.

FUNDING

J.-J.L. was supported by the National Natural Science Foundation of China (42088101), the Jing-Jin-Ji Regional Integrated Environmental Improvement-National Science and Technology (2025ZD1201900) and the State Key Laboratory of Climate System Prediction and Risk Management (CPRM) initiative project (CPRM-2025-NUIST-012). Y.-G.H. was supported by the ASTRA Project through the National Research Foundation (NRF) funded by the Ministry of Science and ICT (RS-2024-00440482). W.X. was supported by the National Natural Science Foundation of China (U2242210). Y.-Q.W. was supported by the National Natural Science Foundation of China (U2342219). M.Z. was supported by the National Natural Science Foundation of China (42505033). Y.T. was supported by the National Natural Science Foundation of China (42505031). Y. Hong was supported by the ASTRA Project through the National Research Foundation (NRF) funded by the Ministry of Science and ICT (RS-2024-00440063).

Conflict of interest statement. The authors declare that they have no conflict of interest.

REFERENCES

- Hoskins BJ and Karoly DJ. The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J Atmos Sci* 1981; **38**: 1179–96.
- Hoskins BJ, McIntyre ME, Robertson AW. On the use and significance of isentropic potential vorticity maps. *Quart J Royal Meteorol Soc* 1985; **111**: 877–946.
- Simmons AJ, Wallace JM, Branstator GW. Barotropic wave propagation and instability, and atmospheric teleconnection patterns. *J Atmos Sci* 1983; **40**: 1363–92.
- Held IM, Ting M, Wang H. Northern winter stationary waves: theory and modeling. *J Clim* 2002; **15**: 2125–44.
- Charney JG, Fjörtoft R, Neumann JV. Numerical integration of the barotropic vorticity equation. *Tellus A Dyn Meteorol Oceanogr* 1950; **2**: 237–54.
- Richardson LF. *Weather Prediction by Numerical Process*. Cambridge: Cambridge University Press, 2007.
- Kadow C, Hall DM, Ulbrich U. Artificial intelligence reconstructs missing climate information. *Nat Geosci* 2020; **13**: 408–13.
- Pickering E, Guth S, Karniadakis GE et al. Discovering and forecasting extreme events via active learning in neural operators. *Nat Comput Sci* 2022; **2**: 823–33.
- Bi K, Xie L, Zhang H et al. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* 2023; **619**: 533–8.

10. Ham Y-G, Kim J-H, Luo J-J. Deep learning for multi-year ENSO forecasts. *Nature* 2019; **573**: 568–72.
11. Guo Z, Lyu P, Ling F *et al*. Data-driven global ocean modeling for seasonal to decadal prediction. *Sci Adv* 2025; **11**: eadu2488.
12. Bonavita M. On some limitations of current machine learning weather prediction models. *Geophys Res Lett* 2024; **51**: e2023GL107377.
13. Kochkov D, Yuval J, Langmore I *et al*. Neural general circulation models for weather and climate. *Nature* 2024; **632**: 1060–6.
14. Jeevanjee N, Hassanzadeh P, Hill S *et al*. A perspective on climate model hierarchies. *J Adv Model Earth Syst* 2017; **9**: 1760–71.
15. Stensrud DJ. *Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models*. Cambridge: Cambridge University Press, 2007.
16. Wu Z, Pan S, Chen F *et al*. A comprehensive survey on graph neural networks. *IEEE Trans Neural Netw Learning Syst* 2021; **32**: 4–24.
17. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436–44.
18. Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 2013; **35**: 1798–828.
19. Bronstein MM, Bruna J, LeCun Y *et al*. Geometric deep learning: going beyond Euclidean data. *IEEE Signal Process Mag* 2017; **34**: 18–42.
20. Amari S-I. Backpropagation and stochastic gradient descent method. *Neurocomputing* 1993; **5**: 185–96.
21. LeCun Y and Bengio Y. Convolutional networks for images, speech, and time series. In: Arbib MA (ed.). *The Handbook of Brain Theory and Neural Networks*. Cambridge: MIT Press, 1995, 255–8.
22. McCulloch WS and Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 1943; **5**: 115–33.
23. Cho K, Merriënboer B, Gulcehre C *et al*. Learning phrase representations using RNN encoder-decoder for statistical machine translation [Preprint]. arXiv: 1406.1078. 2014.
24. Graves A. Long short-term memory. In: Graves A (ed.). *Supervised Sequence Labelling with Recurrent Neural Networks*. Berlin, Heidelberg: Springer, 2012, 37–45.
25. Scarselli F, Gori M, Tsoi AC *et al*. The graph neural network model. *IEEE Trans Neural Netw* 2009; **20**: 61–80.
26. Li Z, Kovachki N, Azizzadenesheli K *et al*. Fourier neural operator for parametric partial differential equations [Preprint]. arXiv: 2010.08895. 2021.
27. Kovachki N, Li Z, Liu B *et al*. Neural operator: learning maps between function spaces with applications to PDEs. *J Mach Learn Res* 2023; **24**: 89.
28. Liu Z, Lin Y, Cao Y *et al*. Swin transformer: hierarchical vision transformer using shifted windows. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Los Alamitos: IEEE Computer Society, 2021, 10012–22.
29. Shi X, Chen Z, Wang H *et al*. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. *Adv Neural Inform Process Syst* 2015; **28**: 802–10.
30. Kingma DP and Welling M. Auto-encoding variational Bayes [Preprint]. arXiv: 1312.6114. 2013.
31. Goodfellow I, Pouget-Abadie J, Mirza M *et al*. Generative adversarial networks. *Commun ACM* 2020; **63**: 139–44.
32. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. *Adv Neural Inform Process Syst* 2020; **33**: 6840–51.
33. Sohl-Dickstein J, Weiss E, Maheswaranathan N *et al*. Deep unsupervised learning using nonequilibrium thermodynamics. In: *International Conference on Machine Learning*. PMLR, 2015, 2256–65.
34. Song Y, Sohl-Dickstein J, Kingma DP *et al*. Score-based generative modeling through stochastic differential equations. In: *International Conference on Learning Representations*. Virtual, 3–7 May 2021.
35. Yang S, Nai C, Liu X *et al*. Generative assimilation and prediction for weather and climate [Preprint]. arXiv: 2503.03038. 2025.
36. Lipman Y, Chen RTQ, Ben-Hamu H *et al*. Flow matching for generative modeling [Preprint]. arXiv: 2210.02747. 2023.
37. Mousavi SS, Schukat M, E. H. Deep reinforcement learning: an overview. In: *Proceedings of SAI Intelligent Systems Conference*. Cham: Springer, 2016, 426–40.
38. Zhu XX, Tuia D, Mou L *et al*. Deep learning in remote sensing: a comprehensive review and list of resources. *IEEE Geosci Remote Sens Mag* 2017; **5**: 8–36.
39. Xiao H, Zhang F, Shen Z *et al*. Classification of weather phenomenon from images by using deep convolutional neural network. *Earth Space Sci* 2021; **8**: e2020EA001604.
40. Lam L, George M, Gardoll S *et al*. Tropical cyclone detection from the thermal infrared sensor IASI data using the deep learning model YOLOv3. *Atmosphere* 2023; **14**: 215.
41. Kashinath K, Mustafa M, Albert A *et al*. Physics-informed machine learning: case studies for weather and climate modelling. *Phil Trans R Soc A* 2021; **379**: 20200093.
42. Pu K, Liu X, Xian M *et al*. Machine learning classification of rainfall types based on the differential attenuation of multiple frequency microwave links. *IEEE Trans Geosci Remote Sens* 2020; **58**: 6888–99.
43. Shen C, Appling AP, Gentile P *et al*. Differentiable modelling to unify machine learning and physical models for geosciences. *Nat Rev Earth Environ* 2023; **4**: 552–67.
44. Dai Y, Zeng X, Dickinson RE *et al*. The common land model. *Bull Am Meteorol Soc* 2003; **84**: 1013–24.
45. Li X, Liu B, Zheng G *et al*. Deep-learning-based information mining from ocean remote-sensing imagery. *Natl Sci Rev* 2020; **7**: 1584–605.
46. Zhang S, Li X, Zhang X. Internal wave signature extraction from SAR and optical satellite imagery based on deep learning. *IEEE Trans Geosci Remote Sens* 2023; **61**: 4203216.
47. Liu Y, Zheng Q, Li X. Characteristics of global ocean abnormal mesoscale eddies derived from the fusion of sea surface height and temperature data by deep learning. *Geophys Res Lett* 2021; **48**: e2021GL094772.
48. Goh E, Yepremyan A, Wang J *et al*. MAESSTRO: masked autoencoders for sea surface temperature reconstruction under occlusion. *Ocean Sci* 2024; **20**: 1309–23.
49. Appel M. Efficient data-driven gap filling of satellite image time series using deep neural networks with partial convolutions. *Arti Intellig Earth Syst* 2024; **3**: 220055.
50. Barth A, Alvera-Azcárate A, Troupin C *et al*. DINCAE 2.0: multivariate convolutional neural network with error estimates to reconstruct sea surface temperature satellite and altimetry observations. *Geosci Model Dev* 2022; **15**: 2183–96.
51. Dong J, Yin R, Sun X *et al*. Inpainting of remote sensing SST images with deep convolutional generative adversarial network. *IEEE Geosci Remote Sens Lett* 2019; **16**: 173–7.
52. Wang H, Zhou Y, Li X. GDCM: generalized data completion model for satellite observations. *Remote Sens Environ* 2025; **324**: 114760.

53. Plésiat É, Dunn RJH, Donat MG *et al.* Artificial intelligence reveals past climate extremes by reconstructing historical records. *Nat Commun* 2024; **15**: 9191.
54. Bochow N, Poltronieri A, Rypdal M *et al.* Reconstructing historical climate fields with deep learning. *Sci Adv* 2025; **11**: eadp0558.
55. Benaïchouche S, Legoff C, Guichoux Y *et al.* Unsupervised reconstruction of sea surface currents from AIS maritime traffic data using trainable variational models. *Remote Sens* 2021; **13**: 3162.
56. Li Z, Han W, Zhang Y *et al.* Learning spatiotemporal dynamics with a pretrained generative model. *Nat Mach Intell* 2024; **6**: 1566–79.
57. Rozet F and Louppe G. Score-based data assimilation. *Adv Neural Inform Process Syst* 2023; **36**: 40521–41.
58. Hsieh WW and Tang B. Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bull Am Meteor Soc* 1998; **79**: 1855–70.
59. Vijaykumar NL, Stephany S, Preto AJ *et al.* *Optimized Neural Network Code for Data Assimilation*. In: Brazilian Congress on Meteorology, Foz do Iguaçu: Sociedade Brasileira de Meteorologia, 2002. pp. 3841–49.
60. Härter FP, de C, Velho HF. New approach to applying neural network in nonlinear dynamic model. *Appl Math Modell* 2008; **32**: 2621–33.
61. Xiao Y, Jia Q, Chen K *et al.* VAE-Var: variational autoencoder-enhanced variational methods for data assimilation in meteorology. In: *The Thirteenth International Conference on Learning Representations*. Singapore, 24–28 April 2025.
62. Liang J, Terasaki K, Miyoshi T. A machine learning approach to the observation operator for satellite radiance data assimilation. *J Meteorol Soc Japan* 2023; **101**: 79–95.
63. Fablet R, Chapron B, Drumetz L *et al.* Learning variational data assimilation models and solvers. *J Adv Model Earth Syst* 2021; **13**: e2021MS002572.
64. Xiao Y, Bai L, Xue W *et al.* Towards a self-contained data-driven global weather forecasting framework. In: *Forty-First International Conference on Machine Learning*. PMLR. 2024.
65. Hatfield S, Chantry M, Dueben P *et al.* Building tangent-linear and adjoint models for data assimilation with neural networks. *J Adv Model Earth Syst* 2021; **13**: e2021MS002521.
66. Lu F. U-Net Kalman Filter (UNetKF): an example of machine learning-assisted data assimilation. *J Adv Model Earth Syst* 2025; **17**: e2023MS003979.
67. Wang Z, Lei L, Anderson JL *et al.* Convolutional neural network-based adaptive localization for an ensemble Kalman filter. *J Adv Model Earth Syst* 2023; **15**: e2023MS003642.
68. Bao F, Zhang Z, Zhang G. An ensemble score filter for tracking high-dimensional nonlinear dynamical systems. *Comput Meth Appl Mech Eng* 2024; **432**: 117447.
69. Xiao P, Si P, Chen P. LD-EnSF: synergizing latent dynamics with ensemble score filters for fast data assimilation with sparse observations [Preprint]. arXiv: 2411.19305. 2024.
70. Peyron M, Fillion A, Gürol S *et al.* Latent space data assimilation by using deep learning. *Quart J Royal Meteor Soc* 2021; **147**: 3759–77.
71. Melinc B and Zaplotnik Ž. 3D-Var data assimilation using a variational autoencoder. *Quart J Royal Meteor Soc* 2024; **150**: 2273–95.
72. Mack J, Arcucci R, Molina-Solana M *et al.* Attention-based convolutional autoencoders for 3D-variational data assimilation. *Comput Meth Appl Mech Eng* 2020; **372**: 113291.
73. Xu X, Sun X, Han W *et al.* FuXi-Da: a generalized deep learning data assimilation framework for assimilating satellite observations. *npj Clim Atmos Sci* 2025; **8**: 156.
74. Chen K, Bai L, Ling F *et al.* Towards an end-to-end artificial intelligence driven global weather forecasting system [Preprint]. arXiv: 2312.12462. 2024.
75. Xiang Y, Jin W, Dong H *et al.* ADAF: an artificial intelligence data assimilation framework for weather forecasting. *J Adv Model Earth Syst* 2025; **17**: e2024MS004839.
76. Yang N, Wang C, Zhao M *et al.* LangYa: revolutionizing cross-spatiotemporal ocean forecasting [Preprint]. arXiv: 2412.18097. 2025.
77. Qu Y, Nathaniel J, Li S *et al.* Deep generative data assimilation in multimodal setting. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos, CA: IEEE Computer Society, 2024, pp. 449–59.
78. Ham Y-G, Joo Y-S, Kim J-H *et al.* Partial-convolution-implemented generative adversarial network for global oceanic data assimilation. *Nat Mach Intell* 2024; **6**: 834–43.
79. Wang W, Duan B, Ni W *et al.* A four-dimensional variational informed generative adversarial network for data assimilation. *J Adv Model Earth Syst* 2025; **17**: e2024MS004437.
80. Zhou Z, Duan W, Yang R *et al.* Uncertainty of ‘Pangu-Weather’ in Bay of Bengal storm track forecasts: target observation perspective. *Atmos Res* 2025; **326**: 108313.
81. Li Y, Duan W, Han W *et al.* Improving tropical cyclone track forecast skill through assimilating target observation achieved by AI-based conditional nonlinear optimal perturbation. *JGR Atmos* 2025; **130**: e2024JD043261.
82. Niu Z, Wang D, Mu M *et al.* Machine-learning (ML)-physics fusion model accelerates the paradigm shift in typhoon forecasting with a CNOP-based assimilation framework. *Geophys Res Lett* 2025; **52**: e2025GL115926.
83. Agrawal S, Barrington L, Bromberg C *et al.* Machine learning for precipitation nowcasting from radar images [Preprint]. arXiv: 1912.12132. 2019.
84. Ravuri S, Lenc K, Willson M *et al.* Skilful precipitation nowcasting using deep generative models of radar. *Nature* 2021; **597**: 672–7.
85. Zhang Y, Long M, Chen K *et al.* Skilful nowcasting of extreme precipitation with NowcastNet. *Nature* 2023; **619**: 526–32.
86. Andrychowicz M, Espeholt L, Li D *et al.* Deep learning for day forecasts from sparse observations [Preprint]. arXiv: 2306.06079. 2023.
87. Espeholt L, Agrawal S, Sønderby C *et al.* Deep learning for twelve hour precipitation forecasts. *Nat Commun* 2022; **13**: 5145.
88. Rasp S, Hoyer S, Meroze A *et al.* WeatherBench 2: a benchmark for the next generation of data-driven global weather models. *J Adv Model Earth Syst* 2024; **16**: e2023MS004019.
89. Lam R, Sanchez-Gonzalez A, Willson M *et al.* Learning skillful medium-range global weather forecasting. *Science* 2023; **382**: 1416–21.
90. Han T, Guo S, Ling F *et al.* FengWu-GHR: learning the kilometer-scale medium-range global weather forecasting [Preprint]. arXiv: 2402.00059. 2024.
91. Nguyen T, Shah R, Bansal H *et al.* Scaling transformer neural networks for skillful and reliable medium-range weather forecasting. *Adv Neural Inform Process Syst* 2024; **37**: 68740–71.
92. Ramavajjala V. HEAL-ViT: vision transformers on a spherical mesh for medium-range weather forecasting [Preprint]. arXiv: 2403.17016. 2024.

93. Alexe M, Boucher E, Lean P *et al*. GraphDOP: towards skilful data-driven medium-range weather forecasts learnt and initialised directly from observations [Preprint]. arXiv: 2412.15687. 2024.
94. Sun X, Zhong X, Xu X *et al*. A data-to-forecast machine learning system for global weather. *Nat Commun* 2025; **16**: 6658.
95. Bodnar C, Bruinsma WP, Lucic A *et al*. A foundation model for the Earth system. *Nature* 2025; **641**: 1180–7.
96. Lang S, Alexe M, Clare MC *et al*. AIFS-CRPS: ensemble forecasting using a model trained with a loss function based on the continuous ranked probability score [Preprint]. arXiv: 2412.15832. 2024.
97. Ling F, Chen K, Wu J *et al*. FengWu-W2S: a deep learning model for seamless weather-to-subseasonal forecast of global atmosphere [Preprint]. arXiv: 2411.10191. 2024.
98. Xiong W, Xiang Y, Wu H *et al*. AI-GOMS: large AI-driven global ocean modeling system [Preprint]. arXiv: 2308.03152. 2023.
99. Bougeault P, Toth Z, Bishop C *et al*. The THORPEX interactive grand global ensemble. *Bull Am Meteor Soc* 2010; **91**: 1059–72.
100. Bonev B, Kurth T, Hundt C *et al*. Spherical Fourier neural operators: learning stable dynamics on the sphere. In: *Proceedings of the 40th International Conference on Machine Learning* 2023; **202**: 2806–23.
101. Chen K, Han T, Ling F *et al*. The operational medium-range deterministic weather forecasting can be extended beyond a 10-day lead time. *Commun Earth Environ* 2025; **6**: 518.
102. Zhong X, Chen L, Liu J *et al*. FuXi-Extreme: improving extreme rainfall and wind forecasts with diffusion model. *Sci China Earth Sci* 2024; **67**: 3696–708.
103. Couairon G, Singh R, Charantonis A *et al*. ArchesWeather & ArchesWeatherGen: a deterministic and generative model for efficient ML weather forecasting [Preprint]. arXiv: 1312.6114. 2024.
104. Price I, Sanchezgonzalez A, Alet F *et al*. Probabilistic weather forecasting with machine learning. *Nature* 2025; **637**: 84–90.
105. Alet F, Price I, El-Kadi A *et al*. Skillful joint probabilistic weather forecasting from marginals [Preprint]. arXiv: 2506.10772. 2025.
106. Andrychowicz M. Deep learning for day forecasts from sparse observations [Preprint]. arXiv: 2306.06079. 2023.
107. Allen A, Markou S, Tebbutt W *et al*. End-to-end data-driven weather prediction. *Nature* 2025; **641**: 1172–9.
108. Liu Y, Zheng Z, Cheng J *et al*. CirT: global subseasonal-to-seasonal forecasting with geometry-inspired transformer [Preprint]. arXiv: 2502.19750. 2025.
109. Chen L, Zhong X, Li H *et al*. A machine learning model that outperforms conventional global subseasonal forecast models. *Nat Commun* 2024; **15**: 6425.
110. Yuval J, Langmore I, Kochkov D *et al*. Neural general circulation models optimized to predict satellite-based precipitation observations [Preprint]. arXiv: 2412.11973. 2024.
111. Wang X, Wang R, Hu N *et al*. XiHe: a data-driven model for global ocean eddy-resolving forecasting [Preprint]. arXiv: 2402.02995. 2024.
112. Zhou L and Zhang R-H. A self-attention-based neural network for three-dimensional multivariate modeling and its skillful ENSO predictions. *Sci Adv* 2023; **9**: eadf2827.
113. Wang H, Hu S, Li X. An interpretable deep learning ENSO forecasting model. *Ocean-Land-Atmos Res* 2023; **2**: 0012.
114. Mu B, Qin B, Yuan S. ENSO-GTC: ENSO deep learning forecast model with a global spatial-temporal teleconnection coupler. *J Adv Model Earth Syst* 2022; **14**: e2022MS003132.
115. Ling F, Luo J-J, Li Y *et al*. Multi-task machine learning improves multi-seasonal prediction of the Indian Ocean Dipole. *Nat Commun* 2022; **13**: 7681.
116. Nikolaev A, Richter I, Sadowski P. Deep learning for climate models of the Atlantic Ocean. In: *AAAI Spring Symposium: MLPS*. Washington, DC: AAAI Press. 2020.
117. Bachélery ML, Brajard J, Patacchiola M *et al*. Predicting Atlantic and Benguela Niño events with deep learning. *Sci Adv* 2025; **11**: eads5185.
118. Kim YJ, Kim HC, Han D *et al*. Prediction of monthly Arctic sea ice concentrations using satellite and reanalysis data based on convolutional neural networks. *The Cryosphere* 2020; **14**: 1083–104.
119. Andersson TR, Hosking JS, Pérez-Ortiz M *et al*. Seasonal Arctic sea ice forecasting with probabilistic deep learning. *Nat Commun* 2021; **12**: 5124.
120. Tang Y and Duan A. Using deep learning to predict the East Asian summer monsoon. *Environ Res Lett* 2021; **16**: 124006.
121. Sharma D, Das S, Goswami BN. Seasonal prediction of Indian summer monsoon extreme rainfall frequency. *npj Clim Atmos Sci* 2025; **8**: 141.
122. Wang C, Pritchard MS, Brenowitz N *et al*. Coupled ocean-atmosphere dynamics in a machine learning Earth system model [Preprint]. arXiv: 2406.08632. 2024.
123. Pu J, Mu M, Feng J *et al*. A fast physics-based perturbation generator of machine learning weather model for efficient ensemble forecasts of tropical cyclone track. *npj Clim Atmos Sci* 2025; **8**: 128.
124. Vonich PT and Hakim GJ. Predictability limit of the 2021 Pacific Northwest heatwave from deep-learning sensitivity analysis. *Geophys Res Lett* 2024; **51**: e2024GL110651.
125. Qin B, Yang Z, Mu M *et al*. The first kind of predictability problem of El Niño predictions in a multivariate coupled data-driven model. *Quart J Royal Meteor Soc* 2024; **150**: 5452–71.
126. Wang Q, Qian J, Mu M *et al*. The predictability study of oceanic deep learning models: taking Kuroshio intrusion into South China Sea as an example. *Ocean Modell* 2025; **198**: 102622.
127. Chen Y, Jin Y, Liu Z *et al*. Combined dynamical-deep learning ENSO forecasts. *Nat Commun* 2025; **16**: 3845.
128. Jin W, Luo Y, Wu T *et al*. Deep learning for seasonal precipitation prediction over China. *J Meteorol Res* 2022; **36**: 271–81.
129. Scher S and Messori G. Weather and climate forecasting with neural networks: using general circulation models (GCMs) with different complexity as a study ground. *Geosci Model Dev* 2019; **12**: 2797–809.
130. Min Y-M, Kryjov VN, Oh SM *et al*. Skill of real-time operational forecasts with the APCC multi-model ensemble prediction system during the period 2008–2015. *Clim Dyn* 2017; **49**: 4141–56.
131. Stockdale TN, Anderson DLT, Balmaseda MA *et al*. ECMWF seasonal forecast system 3 and its prediction of sea surface temperature. *Clim Dyn* 2011; **37**: 455–71.
132. Hospedales T, Antoniou A, Micaelli P *et al*. Meta-learning in neural networks: a survey. *IEEE Trans Pattern Anal Mach Intell* 2022; **44**: 5149–69.
133. Li Z, He J, Ni T *et al*. Numerical computation based few-shot learning for intelligent sea surface temperature prediction. *Multimedia Syst* 2023; **29**: 3001–13.
134. Oh SH and Ham YG. Few shot learning for Korean winter temperature forecasts. *npj Clim Atmos Sci* 2024; **7**: 279.
135. Watt-Meyer O, Henn B, McGibbon J *et al*. ACE2: accurately learning sub-seasonal to decadal atmospheric variability and forced responses. *npj Clim Atmos Sci* 2025; **8**: 205.

136. Hakim GJ and Masanam S. Dynamical tests of a deep learning weather prediction model. *Artif Intellig Earth Syst* 2024; **3**: e230090.
137. Chen Z, Liu Y, Sun H. Physics-informed learning of governing equations from scarce data. *Nat Commun* 2021; **12**: 6136.
138. Beucler T. Enforcing analytic constraints in neural networks emulating physical systems. *Phys Rev Lett* 2021; **126**: 098302.
139. González-Abad J, Hernández-García Á, Harder P et al. Multi-variable hard physical constraints for climate model downscaling. In: *AAAI-SS* 2023; **2**: 62–7.
140. Tartakovsky AM, Marrero CO, Perdikaris P et al. Physics-informed deep neural networks for learning parameters and constitutive relationships in subsurface flow problems. *Water Resour Res* 2020; **56**: e2019WR026731.
141. Brenowitz ND and Bretherton CS. Spatially extended tests of a neural network parametrization trained by coarse-graining. *J Adv Model Earth Syst* 2019; **11**: 2728–44.
142. Fuchs D, Sherwood SC, Prasad A et al. TorchClim v1.0: a deep-learning plugin for climate model physics. *Geosci Model Dev* 2024; **17**: 5459–75.
143. Wang X, Yang J, Adie J et al. CondensNet: enabling stable long-term climate simulations via hybrid deep learning models with adaptive physical constraints [Preprint]. arXiv: 2502.13185, 2025.
144. He Y, Wang B, Fang J et al. Constraint of air–sea interaction significant to skillful predictions of North Pacific climate variations. *J Clim* 2023; **36**: 5941–62.
145. Liu L, Zhou W, Guan K et al. Knowledge-guided machine learning can improve carbon cycle quantification in agroecosystems. *Nat Commun* 2024; **15**: 357.
146. Zhu Y, Zhang R-H, Moum JN et al. Physics-informed deep-learning parameterization of ocean vertical mixing improves climate simulations. *Natl Sci Rev* 2022; **9**: nwac044.
147. François B, Thao S, Vrac M. Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks. *Clim Dyn* 2021; **57**: 3323–53.
148. Zhang X, Cui P, Xu R et al. Deep stable learning for out-of-distribution generalization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos, CA: IEEE Computer Society, 2021, pp. 5372–82.
149. Meehl GA, Goddard L, Murphy J et al. Decadal prediction: can it be skillful?. *Bull Am Meteor Soc* 2009; **90**: 1467–86.
150. Kushnir Y, Scaife AA, Arritt R et al. Towards operational predictions of the near-term climate. *Nat Clim Change* 2019; **9**: 94–101.
151. Zhu Y, Zhang R-H, Wang F et al. Projection of ENSO using observation-informed deep learning. *Nat Commun* 2025; **16**: 7736.
152. Lu L, Jin P, Pang G et al. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nat Mach Intell* 2021; **3**: 218–29.
153. Liu G, Wang P, Kwon Y-O. Physical insights from the multidecadal prediction of North Atlantic sea surface temperature variability using explainable neural networks. *Geophys Res Lett* 2023; **50**: e2023GL106278.
154. Bonavita M and Laloyaux P. Machine learning for model error inference and correction. *J Adv Model Earth Syst* 2020; **12**: e2020MS002232.
155. Liang Z and Hu Z. A Bayes-based approach against sample imbalance to improving the potential forecasts of gale. *Geophys Res Lett* 2022; **49**: e2022GL100019.
156. Sun L, Lan Y, Sun X et al. Deterministic forecasting and probabilistic post-processing of short-term wind speed using statistical methods. *JGR Atmos* 2024; **129**: e2023JD040134.
157. Zeng Z, Wu H, Liu Z et al. Enhancing short-term wind speed prediction capability of numerical weather prediction through machine learning methods. *JGR Atmos* 2024; **129**: e2024JD041822.
158. Hu YF, Yin FK, Zhang WM. Deep learning-based precipitation bias correction approach for Yin–He global spectral model. *Meteorol Appl* 2021; **28**: e2032.
159. Han L, Chen M, Chen K et al. A deep learning method for bias correction of ECMWF 24–240 h forecasts. *Adv Atmos Sci* 2021; **38**: 1444–59.
160. Tong X and Zhou W. Assessing predictive attribution in NMME forecasts of summer precipitation in eastern China using deep learning. *npj Clim Atmos Sci* 2024; **7**: 304.
161. Xu L, Chen N, Zhang X et al. Improving the North American multi-model ensemble (NMME) precipitation forecasts at local areas using wavelet and machine learning. *Clim Dyn* 2019; **53**: 601–15.
162. Xu L, Chen N, Zhang X et al. A data-driven multi-model ensemble for deterministic and probabilistic precipitation forecasting at seasonal scale. *Clim Dyn* 2020; **54**: 3355–74.
163. Mouatadid S, Orenstein P, Flaspohler G et al. Adaptive bias correction for improved subseasonal forecasting. *Nat Commun* 2023; **14**: 3482.
164. Kim H, Ham YG, Joo YS et al. Deep learning for bias correction of MJO prediction. *Nat Commun* 2021; **12**: 3087.
165. Nourani V, Bahghanam AH, Pourali H et al. Bias correcting the precipitation dynamics of regional climate models via kernel-aware 2D convolutional-long short-term memory. *J Hydrol* 2025; **657**: 133068.
166. Chapman WE and Berner J. Improving climate bias and variability via CNN-based state-dependent model-error corrections. *Geophys Res Lett* 2025; **52**: e2024GL114106.
167. Pan B, Anderson GJ, Goncalves A et al. Learning to correct climate projection biases. *J Adv Model Earth Syst* 2021; **13**: e2021MS002509.
168. Vandal T, Kodra E, Ganguly S et al. DeepSD: generating high resolution climate change projections through single image super-resolution. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery, 2017, pp. 1663–72.
169. Cheng J, Kuang Q, Shen C et al. ResLap: generating high-resolution climate prediction through image super-resolution. *IEEE Access* 2020; **8**: 39623–34.
170. Groenke B, Madaus L, Monteleoni C. ClimAlign: unsupervised statistical downscaling of climate variables via normalizing flows. In: *Proceedings of the 10th International Conference on Climate Informatics*. New York: Association for Computing Machinery, 2020, pp. 60–6.
171. Wan ZY, Lopez-Gomez I, Carver R et al. Regional climate risk assessment from climate models using probabilistic machine learning [Preprint]. arXiv: 2412.08079. 2025.
172. Stengel K, Glaws A, Hettinger D et al. Adversarial super-resolution of climatological wind and solar data. *Proc Natl Acad Sci USA* 2020; **117**: 16805–15.
173. Karras T, Aila T, Laine S et al. Progressive growing of GANs for improved quality, stability, and variation [Preprint]. arXiv: 1710.10196. 2018.
174. Karnewar A and Wang OMSG-GAN.: Multi-scale gradients for generative adversarial networks. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos, CA: IEEE Computer Society, 2020, pp. 7799–808.
175. Mardani M, Brenowitz N, Cohen Y et al. Residual corrective diffusion modeling for km-scale atmospheric downscaling. *Commun Earth Environ* 2025; **6**: 124.

176. Wang S, Li X, Zhu X *et al.* Spatial downscaling of sea surface temperature using diffusion model. *Remote Sens* 2024; **16**: 3843.
177. Hess P, Aich M, Pan B *et al.* Fast, scale-adaptive and uncertainty-aware downscaling of Earth system model fields with generative machine learning. *Nat Mach Intell* 2025; **7**: 363–73.
178. Ling F, Lu Z, Luo J-J *et al.* Diffusion model-based probabilistic downscaling for 180-year East Asian climate reconstruction. *npj Clim Atmos Sci* 2024; **7**: 131.
179. Chen A and Yuan C. Deep learning-based spatial downscaling and its application for tropical cyclone detection in the western North Pacific. *Front Earth Sci* 2024; **12**: 1345714..
180. Harris L, McRae ATT, Chantry M *et al.* A generative deep learning approach to stochastic downscaling of precipitation forecasts. *J Adv Model Earth Syst* 2022; **14**: e2022MS003120.
181. Baño-Medina J, Manzanar R, Gutiérrez JM. Configuration and inter-comparison of deep learning neural models for statistical downscaling. *Geosci Model Dev* 2020; **13**: 2109–24.
182. Pan B, Hsu K, AghaKouchak A *et al.* Improving precipitation estimation using convolutional neural network. *Water Resour Res* 2019; **55**: 2301–21.
183. Holm EA. In defense of the black box. *Science* 2019; **364**: 26–7.
184. Minh D, Wang HX, Li YF *et al.* Explainable artificial intelligence: a comprehensive review. *Artif Intell Rev* 2022; **55**: 3503–68.
185. O'Loughlin RJ, Li D, Neale R *et al.* Moving beyond post hoc explainable artificial intelligence: a perspective paper on lessons learned from dynamical climate modeling. *Geosci Model Dev* 2025; **18**: 787–802.
186. Wu S and Easterbrook SM. Bridging the Gap: climate scientists' concerns and expectations for machine learning. In: *Proceedings of the ACM SIG-CAS/SIGCHI Conference on Computing and Sustainable Societies*. New York: Association for Computing Machinery, 2025, pp. 284–96.
187. Gunning D and Aha D. DARPA's explainable artificial intelligence (XAI) program. *AI Magazine* 2019; **40**: 44–58.
188. Li X, Wang H, Yang Y *et al.* Deep learning-based solution for the KdV-family governing equations of ocean internal waves. *Ocean Modell* 2025; **194**: 102493.
189. Vaswani A, Shazeer N, Parmar N *et al.* Attention is all you need. *Adv Neural Inform Process Syst* 2017; **30**: 5998–6008.
190. Mu B, Qin B, Yuan S. ENSO-ASC 1.0.0: ENSO deep learning forecast model with a multivariate air–sea coupler. *Geosci Model Dev* 2021; **14**: 6977–99.
191. Mu B, Cui Y, Yuan S *et al.* Simulation, precursor analysis and targeted observation sensitive area identification for two types of ENSO using ENSO-MC v1.0. *Geosci Model Dev* 2022; **15**: 4105–27.
192. Pathak J, Subramanian S, Harrington P *et al.* FourCastNet: a global data-driven high-resolution weather model using adaptive Fourier neural operators [Preprint]. arXiv: 2202.11214. 2022.
193. Xiong W, Huang X, Zhang Z *et al.* Koopman neural operator as a mesh-free solver of non-linear partial differential equations. *J Comput Phys* 2024; **513**: 113194.
194. Tesch T, Kollet S, Garcke J. Causal deep learning models for studying the Earth system. *Geosci Model Dev* 2023; **16**: 2149–66.
195. Qin B, Meng F, Yuan S *et al.* CAU: a causality attention unit for spatial-temporal sequence forecast. *IEEE Trans Multimedia* 2024; **26**: 4749–63.
196. Huang Q, Yamada M, Tian Y *et al.* GraphLIME: local interpretable model explanations for graph neural networks. *IEEE Trans Knowl Data Eng* 2023; **35**: 6968–72.
197. GvD B, Lykov A, Schleich M *et al.* On the tractability of SHAP explanations. *J Art Intellig Res* 2022; **74**: 851–86.
198. Abdollahi A and Pradhan B. Explainable artificial intelligence (XAI) for interpreting the contributing factors feed into the wildfire susceptibility prediction model. *Sci Total Environ* 2023; **879**: 163004.
199. Jacovi A, Hadash G, Kermany E *et al.* Neural network gradient-based learning of black-box function interfaces [Preprint]. arXiv: 1901.03995. 2019.
200. Jiang S, Sweet L-b, Blougouras G *et al.* How interpretable machine learning can benefit process understanding in the geosciences. *Earth's Future* 2024; **12**: e2024EF004540.
201. Mu M, Qin B, Dai G. Predictability study of weather and climate events related to artificial intelligence models. *Adv Atmos Sci* 2025; **42**: 1–8.
202. Mu M, Duan WS, Wang B. Conditional nonlinear optimal perturbation and its applications. *Nonlin Processes Geophys* 2003; **10**: 493–501.
203. Rackow T, Koldunov N, Lessig C *et al.* Robustness of AI-based weather forecasts in a changing climate [Preprint]. arXiv: 2409.18529. 2024.
204. Karniadakis GE, Kevrekidis IG, Lu L *et al.* Physics-informed machine learning. *Nat Rev Phys* 2021; **3**: 422–40.
205. Li W, Liu Z, Chen K *et al.* DeepPhysiNet: bridging deep learning and atmospheric physics for accurate and continuous weather modeling [Preprint]. arXiv: 2401.04125. 2024.
206. Zhu Z, Wang Z, Dong C *et al.* Physics informed neural network modelling for storm surge forecasting: a case study in the Bohai Sea. *Coast Eng* 2025; **197**: 104686.
207. White A, Kilbertus N, Gelbrecht M *et al.* Stabilized neural differential equations for learning dynamics with explicit constraints. *Adv Neural Inform Process Syst* 2023; **36**: 12929–50.
208. Brunton SL, Proctor JL, Kutz JN. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc Natl Acad Sci USA* 2016; **113**: 3932–7.
209. Penny SG, Smith TA, Chen TC *et al.* Integrating recurrent neural networks with data assimilation for scalable data-driven state estimation. *J Adv Model Earth Syst* 2022; **14**: e2021MS002843.
210. Zhong X, Ma Z, Yao Y *et al.* WRF–ML v1.0: a bridge between WRF v4.3 and machine learning parameterizations and its application to atmospheric radiative transfer. *Geosci Model Dev* 2023; **16**: 199–209.
211. Ji C, Mu M, Qin B *et al.* Toward skillful forecasting of super El Niño events using a diffusion-based westerly wind burst parameterization. *npj Clim Atmos Sci* 2025; **8**: 273.
212. Mu B, Chen L, Yuan S *et al.* A radiative transfer deep learning model coupled into WRF with a generic fortran torch adaptor. *Front Earth Sci* 2023; **11**: 1149566.
213. Xu W, Ling F, Zhang W *et al.* Generalizing weather forecast to fine-grained temporal scales via physics-AI hybrid modeling. *Adv Neural Inform Process Syst* 2024; **37**: 23325–51.
214. Verma Y, Heinonen M, ClimODE GV. Climate and weather forecasting with physics-informed neural ODEs [Preprint]. arXiv: 2404.10024. 2024.
215. Slater LJ, Arnal L, Boucher M-A *et al.* Hybrid forecasting: blending climate predictions with AI models. *Hydrol Earth Syst Sci* 2023; **27**: 1865–89.
216. Bracco A, Brajard J, Dijkstra HA *et al.* Machine learning for the physics of climate. *Nat Rev Phys* 2025; **7**: 6–20.
217. Lee JY, Ko S, Hong Y. Finite element operator network for solving elliptic-type parametric PDEs. *SIAM J Sci Comput* 2025; **47**: C501–28.

218. Bonavita M. On some limitations of data-driven weather forecasting models [Preprint]. arXiv: 2309.08473. 2023.
219. Deng J, Dong W, Socher R *et al*. ImageNet: a large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009, pp. 248–55.
220. Lin T-Y, Maire M, Belongie S *et al*. Microsoft COCO: common objects in context. In: *European Conference on Computer Vision*. Cham: Springer International Publishing, Los Alamitos, CA: IEEE Computer Society, 2014, 740–55.
221. Gruca A, Rípodas P, Calbet X *et al*. Multi-task challenges for rain movie prediction on the road to hi-res foundation models. In: *NeurIPS 2024 Competition Track*. Red Hook, NY: Curran Associates, 2024.
222. WMO AINPP Steering group. AI for Nowcasting Pilot Project. <https://www.ainpp.org> (26 February 2026, date last accessed).
223. Rasp S, Dueben PD, Scher S *et al*. WeatherBench: a benchmark data set for data-driven weather forecasting. *J Adv Model Earth Syst* 2020; **12**: e2020MS002203.
224. Nathaniel J, Qu Y, Nguyen T *et al*. ChaosBench: a multi-channel, physics-based benchmark for subseasonal-to-seasonal climate prediction. *Adv Neural Inform Process Syst* 2024; **37**: 43715–29.
225. Nguyen T, Jewik J, Bansal H *et al*. ClimateLearn: benchmarking machine learning for weather and climate modeling. *Adv Neural Inform Process Syst* 2023; **36**: 75009–25.
226. Garg S, Rasp S, Thuerey N. WeatherBench probability: a benchmark dataset for probabilistic medium-range weather forecasting along with deep learning baseline models [Preprint]. arXiv: 2205.00865. 2022.
227. Brenowitz ND, Cohen Y, Pathak J *et al*. A practical probabilistic benchmark for AI weather models. *Geophys Res Lett* 2025; **52**: e2024GL113656.
228. Lavers DA, Simmons A, Vamborg F *et al*. An evaluation of ERA5 precipitation for climate monitoring. *Quart J Royal Meteorol Soc* 2022; **148**: 3152–65.
229. Zhu X, Xiong Y, Wu M *et al*. Weather2K: a multivariate spatio-temporal benchmark dataset for meteorological forecasting based on real-time observation data from ground weather stations [Preprint]. arXiv: 2302.10493. 2023.
230. Han T, Guo S, Chen Z *et al*. How far are today's time-series models from real-world weather forecasting applications? [Preprint]. arXiv: 2406.14399, 2024.
231. Jin W, Weyn J, Zhao P *et al*. WeatherReal: a benchmark based on in-situ observations for evaluating weather models [Preprint]. arXiv: 2409.09371. 2024.
232. Ran N, Xiao P, Wang Y *et al*. HR-Extreme: a high-resolution dataset for extreme weather forecasting [Preprint]. arXiv: 2409.18885. 2025.
233. Prabhat KK and Mudigonda M *et al*. ClimateNet: an expert-labeled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather. *Geosci Model Dev* 2021; **14**: 107–24.
234. Fu D, Zhu Y, Liu Z *et al*. ClimateBench-M: a multi-modal climate data benchmark with a simple generative method. In: *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*. New York: Association for Computing Machinery, 2025, pp. 6367–71.
235. Cachay SR, Ramesh V, Cole JNS *et al*. ClimART: a benchmark dataset for emulating atmospheric radiative transfer in weather and climate models [Preprint]. arXiv: 2111.14671. 2021.
236. Wang W, Ni W, Han T *et al*. A benchmark for AI-based weather data assimilation [Preprint]. arXiv: 2408.11438. 2024.
237. Muñoz-Sabater J, Dutra E, Agustí-Panareda A *et al*. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth Syst Sci Data* 2021; **13**: 4349–83.
238. Li Q, Zhang C, Shangquan W *et al*. LandBench 1.0: a benchmark dataset and evaluation metrics for data-driven land surface variables prediction. *Expert Syst Appl* 2024; **243**: 122917.
239. Shu R, Wu H, Gao Y *et al*. Advanced forecasts of global extreme marine heatwaves through a physics-guided data-driven approach. *Environ Res Lett* 2025; **20**: 044030.
240. Wei J, Hang R, Luo J-J. Prediction of pan-Arctic sea ice using attention-based LSTM neural networks. *Front Mar Sci* 2022; **9**: 860403.
241. Wang W, Yang W, Wang L *et al*. Seasonal forecasting of Pan-Arctic sea ice with state space model. *npj Clim Atmos Sci* 2025; **8**: 172.
242. Ouyang L, Ling F, Li Y *et al*. Wave forecast in the Atlantic Ocean using a double-stage ConvLSTM network. *Atmos Ocean Sci Lett* 2023; **16**: 100347.
243. Ling F, Ouyang L, Larbi BR *et al*. Improving global weather and ocean wave forecast with large artificial intelligence models. *Sci China Earth Sci* 2024; **67**: 3641–54.
244. Xie W, Xu G, Zhang H *et al*. Developing a deep learning-based storm surge forecasting model. *Ocean Modell* 2023; **182**: 102179.
245. Gao Z, Shi X, Wang H *et al*. Earthformer: exploring space-time transformers for Earth system forecasting. *Adv Neural Inform Process Syst* 2022; **35**: 25390–403.
246. Lyu P, Tang T, Ling F *et al*. ResoNet: robust and explainable ENSO forecasts with hybrid convolution and transformer networks. *Adv Atmos Sci* 2024; **41**: 1289–98.
247. Kaltenborn J, Lange C, Ramesh V *et al*. ClimateSet: a large-scale climate model dataset for machine learning. *Adv Neural Inform Process Syst* 2023; **36**: 21757–92.
248. Minixhofer C, Swan M, McMeekin C *et al*. DroughtED: a dataset and methodology for drought forecasting spanning multiple climate zones. In: *Tackling Climate Change with Machine Learning: Workshop at ICML 2021*. 2021.
249. Racah E, Beckham C, Maharaj T *et al*. ExtremeWeather: a large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. In: *Proceedings of the 31st International Conference on Neural Information Processing System*, 2017, 3405–16.
250. O'Neill BC, Tebaldi C, van Vuuren DP *et al*. The Scenario Model Inter-comparison Project (ScenarioMIP) for CMIP6. *Geosci Model Dev* 2016; **9**: 3461–82.
251. Barnes EA, Hurrell JW, Ebert-Uphoff I *et al*. Viewing forced climate patterns through an AI lens. *Geophys Res Lett* 2019; **46**: 13389–98.
252. Rader JK, Barnes EA, Ebert-Uphoff I *et al*. Detection of forced change within combined climate fields using explainable neural networks. *J Adv Model Earth Syst* 2022; **14**: e2021MS002941.
253. Ham Y-G, Kim J-H, Min S-K *et al*. Anthropogenic fingerprints in daily precipitation revealed by deep learning. *Nature* 2023; **622**: 301–7.
254. Watt-Meyer O, Dresdner G, McGibbon J *et al*. ACE: a fast, skillful learned global atmospheric model for climate prediction [Preprint]. arXiv: 2310.02074. 2023.
255. Watson-Parris D, Rao Y, Olivé D *et al*. ClimateBench v1.0: a benchmark for data-driven climate projections. *J Adv Model Earth Syst* 2022; **14**: e2021MS002954.

256. Lütjens B, Ferrari R, Watson-Parris D *et al.* The impact of internal variability on benchmarking deep learning climate emulators. *J Adv Model Earth Syst* 2025; **17**: e2024MS004619.
257. Yik W, Silva SJ, Geiss A *et al.* Exploring randomly wired neural networks for climate model emulation. *Artif Intellig Earth Syst* 2023; **2**: 220088.
258. Yu S, Hannah W, Peng L *et al.* ClimSim: a large multi-scale dataset for hybrid physics-ML climate emulation. *Adv Neural Inform Process Syst* 2023; **36**: 22070–84.
259. Yu S, White BL, Bhiwandiwala A *et al.* ClimDetect: a benchmark dataset for climate change detection and attribution [Preprint]. arXiv: 2408.15993. 2025.
260. Ren P, Erichson NB, Guo J *et al.* SuperBench: a super-resolution benchmark dataset for scientific machine learning [Preprint]. arXiv: 2306.14070. 2025.
261. Sha Y, Schreck JS, Chapman W *et al.* Improving AI weather prediction models using global mass and energy conservation schemes. *J Adv Model Earth Syst* 2025; **17**: e2025MS005138.
262. Magnusson L, McTaggart-cowan R, Polichtchouk I *et al.* WP-MIP White Paper. <https://www.wcrp-esmo.org/activities/wp-mip> (26 February 2026, date last accessed).
263. Jean M, Chan PW, Vice President I *et al.* Forecasting the future: the role of artificial intelligence in transforming weather prediction and policy. *Boletín-Organización Meteorológica Mundial* 2025; **74**: 2–9.
264. Zhong X, Chen L, Li H *et al.* FuXi-ENS: a machine learning model for efficient and accurate ensemble weather prediction. *Sci Adv* 2025; **11**: eadu2854.
265. Epstein ES. A scoring system for probability forecasts of ranked categories. *J Appl Meteor* 1969; **8**: 985–7.
266. Wilks DS. *Statistical Methods in the Atmospheric Sciences*. Cambridge: Academic Press, 2011.
267. Bommer PL, Kretschmer M, Hedström A *et al.* Finding the right XAI method: a guide for the evaluation and ranking of explainable AI methods in climate science. *Artif Intellig Earth Syst* 2024; **3**: e230074.
268. Ben Bouallègue Z, Clare M, Magnusson L *et al.* The rise of data-driven weather forecasting: a first statistical assessment of machine learning-based weather forecasts in an operational-like context. *Bull Am Meteorol Soc* 2024; **105**: E864–83.
269. Shobanke M, Bhatt M, Shittu E. Advancements and future outlook of artificial intelligence in energy and climate change modeling. *Adv Appl Energy* 2025; **17**: 100211.
270. Zhang Z, Lin L, Gao S *et al.* A machine learning model for hub-height short-term wind speed prediction. *Nat Commun* 2025; **16**: 3195.
271. Reichstein M, Benson V, Blunk J *et al.* Early warning of complex climate risk with integrated artificial intelligence. *Nat Commun* 2025; **16**: 2564.
272. Camps-Valls G, Fernández-Torres M-Á, Cohrs K-H *et al.* Artificial intelligence for modeling and understanding extreme weather and climate events. *Nat Commun* 2025; **16**: 1919.
273. Cao Y, Chen L, Wu J *et al.* Enhancing Nowcasting with multi-resolution inputs using deep learning: exploring model decision mechanisms. *Geophys Res Lett* 2025; **52**: e2024GL113699.
274. Charlton-Perez AJ, Dacre HF, Driscoll S *et al.* Do AI models produce better weather forecasts than physics-based models? A quantitative evaluation case study of Storm Ciarán. *npj Clim Atmos Sci* 2024; **7**: 93.
275. Boiko DA, Reschützegger T, Sanchez-Lengeling B *et al.* Advancing molecular machine learning representations with stereoelectronics-infused molecular graphs. *Nat Mach Intell* 2025; **7**: 771–81.
276. Boiko DA, MacKnight R, Kline B *et al.* Autonomous chemical research with large language models. *Nature* 2023; **624**: 570–8.
277. Swanson K, Wu W, Bulaong NL *et al.* The virtual lab of AI agents designs new SARS-CoV-2 nanobodies. *Nature* 2025; **646**: 716–23.
278. Guo Z, Wang J, Yue X *et al.* EarthLink: a self-evolving AI agent for climate science [Preprint]. arXiv: 2507.17311. 2025.