Towards Neural Earth System Modelling by integrating Artificial Intelligence in Earth System Science

Christopher Irrgang[1], Niklas Boers[2,3,4], Maike Sonnewald[5,6,7], Elizabeth A. Barnes[8], Christopher Kadow[9], Joanna Staneva[10], and Jan Saynisch-Wagner[1]

[1]Helmholtz Centre Potsdam, German Research Centre for Geosciences GFZ, Potsdam, Germany
[2]Department of Mathematics and Computer Science, Free University of Berlin, Germany
[3]Potsdam Institute for Climate Impact Research, Potsdam, Germany
[4]Department of Mathematics and Global Systems Institute, University of Exeter, UK
[5]Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ 08540, USA
[6]NOAA/OAR Geophysical Fluid Dynamics Laboratory, Ocean and Cryosphere Division, Princeton, NJ 08540, USA
[7]University of Washington, School of Oceanography, Seattle, WA, USA
[8]Colorado State University, Fort Collins, USA
[9]German Climate Computing Center DKRZ, Hamburg, Germany
[10]Helmholtz-Zentrum Geesthacht, Center for Material and Coastal Research HZG, Geesthacht, Germany

June 25, 2021

Abstract

Earth System Models (ESMs) are our main tools to quantify the physical state of the Earth and to predict how it might change in the future under ongoing anthropogenic forcing. In recent years, however, artificial intelligence (AI) methods have been increasingly used to augment or even replace classical ESM tasks, raising hopes that AI could solve some of the grand challenges of climate science. In this perspective, we survey the recent achievements and limitations of both process-based models and AI in Earth system and climate research and propose a methodological transformation, in which deep neural networks and ESMs are dismantled as individual approaches and reassembled as learning, self-validating, and interpretable Earth system model-network hybrids. Following this path, we coin the term "Neural Earth System Modelling". We examine the concurrent potential and

pitfalls of Neural Earth System Modelling and discuss the open question whether AI can infuse ESMs or, ultimately, even render them obsolete.

For decades, scientists have utilized mathematical equations to describe geophysical and climate processes and to construct deterministic computer simulations that allow for the analysis of such processes. Until recently, process-based models had been considered irreplaceable tools that helped to understand the complex interactions in the coupled Earth system and that provided the only tools to predict the Earth system's response to anthropogenic climate change.

Earth system models (ESMs)[1] combine process-based models of the different sub-systems of the Earth system into an integrated numerical model that yields for a given state of the coupled system at time t a prediction of the system state for time $t+1$. The individual model components, or modules, describe sub-systems including the atmosphere, the oceans, the carbon and other biogeochemical cycles, radiation processes, as well as land surface and vegetation processes and marine ecosystems. These modules are then combined by a dynamic coupler to obtain a consistent state of the full system for each time step.

The inclusion of a vastly increasing number of processes, together with continuously rising spatial resolution, have led to the development of comprehensive ESMs to analyse and predict the state of the Earth system. From the first assessment report of the Intergovernmental Panel on Climate Change (IPCC) in 1990 to the fifth phase of the Climate Model Intercomparison Project (CMIP5)[2] and the associated fifth IPCC assessment report in 2014, the spatial resolution has increased from around 500 km to up to 70 km. In accordance, the CMIP results show that the models have, over the course of two decades, greatly improved in their accuracy to reproduce crucial characteristics of the Earth system, such as the evolution of the global mean temperatures (GMT) since the beginning of instrumental data in the second half of the 19th century, or the average present-day spatial distribution of temperature or precipitation[3,4].

The provocative thought that ESMs might lose their fundamental importance in the advent of novel artificial intelligence (AI) tools has sparked both a gold-rush feeling and caution in the scientific communities. On the one hand, deep neural networks have been developed that complement and aim to match the skill of process-based models in various applications, ranging from numerical weather prediction to climate research. On the other hand, most neural networks are trained for isolated applications under simplified conditions and lack true process knowledge. Regardless, the daily increasing data streams from Earth system observation (ESO), increasing computational resources, and the availability and accessibility of powerful AI tools, particularly in machine learning (ML), have led to numerous innovative developments that aim to resolve persistent shortcomings of current Earth system models.

In the following, we survey the current state, recent achievements, and recognised limitations of both process-based modelling and AI in Earth and climate research. Based on this survey, we draw a perspective on an imminent and profound methodological transformation, hereafter named Neural Earth System Modelling (NESYM), that aims for a deep and interpretable integration of AI into Earth system modelling. We discuss emerging challenges of this perspective and highlight the necessity of new transdisciplinary collaborations between the involved communities.

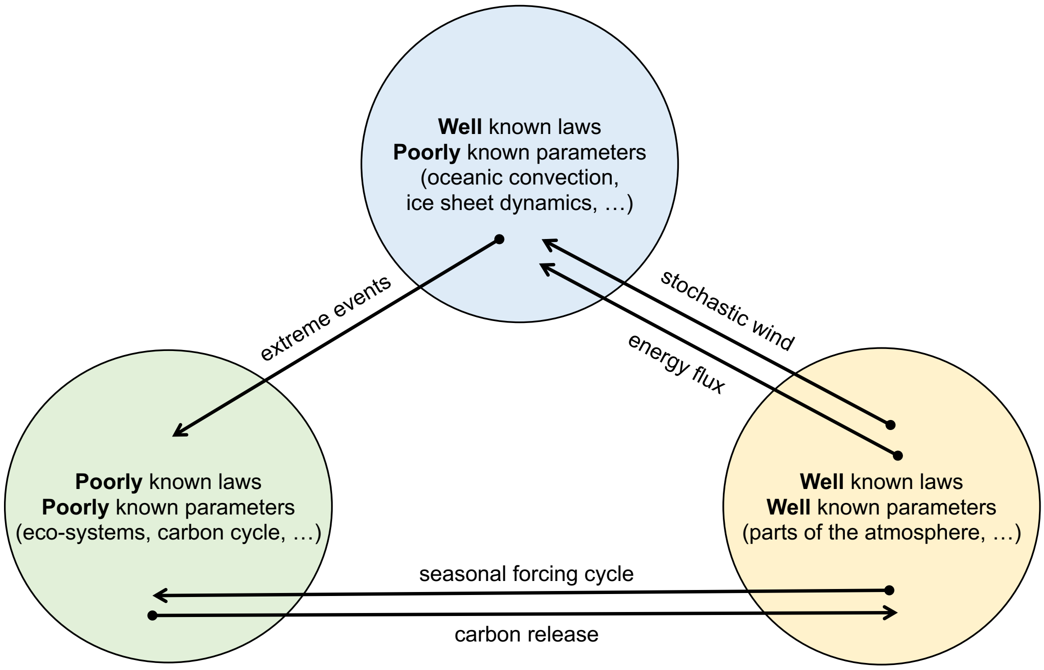

Figure 1: Symbolic representation of Earth system components in terms of knowledge clusters. Arrows indicate exemplary exchange of information between the clusters in terms of geophysical processes and coupling mechanisms. ML can take over different tasks depending on the cluster application, e.g., data exploration and analysis in case of poor process knowledge (green cluster), ESM enhancement by improving insufficient parameterizations and other simplifications in process-based models (blue cluster), or emulation and acceleration of well-understood process-based simulations (orange cluster). Similarly, ML can be applied on coupling mechanisms and interaction processes (arrows), utilizing adjacent clusters as training data pools.

Overview on Earth System Modelling and Earth System Observations

For some parts of the Earth system, the primitive physical equations of motion are known explicitly, such as the Navier-Stokes equations that describe the fluid dynamics of the atmosphere and oceans (Fig. 1). In practise, it is impossible to numerically resolve all relevant scales of the dynamics and approximations have to be made. For example, the fluid dynamical equations for the atmosphere and oceans are integrated on discrete spatial grids, and all processes that operate below the grid resolution have to be parameterised to assure a closed description of the system. Since the multi-scale nature of the dynamics of geophysical fluids implies that the subgrid-scale processes interact with the larger scales that are resolved by the model, (stochastic) parameterization of subgrid-scale processes is a highly non-trivial, yet unavoidable, part of climate modelling[5–7].

For other parts of the Earth system, primitive equations of motion, such as the Navier-Stokes equations for atmospheric motion, do not exist. Essentially, this is due to the complexity of the Earth system, where many phenomena that emerge at a macroscopic level are not easily deducible from microscopic scales that may or may not be well-understood. A typical example is given by ecosystems and the physiological processes governing the vegetation that covers vast parts of the land surface, as well as their interactions with the atmosphere, the carbon and other geochemical cycles. Also for these cases, approximations in terms of parameterizations of potentially crucial processes have to be made.

Regardless of the specific process, such parameterizations induce free parameters in ESMs, for which suitable values have to be found empirically. The size of state-of-the-art ESMs mostly prohibits systematic calibration methods such as, e.g., the ones based on Bayesian inference, and the models are therefore often tuned manually. The quality of the calibration as well as the overall accuracy of the model can only be assessed with respect to relatively sparse observations of the last 170 years, at most, and there is no way to assess the models' skill in predicting future climate conditions[8]. Although necessary, paramterizations can cause biases or structural model errors. The example of the discretized spatial grid suggests that the higher the spatial resolution of an ESM, the smaller the potential errors. Likewise, it is expected that the models' representation of the Earth system will become more accurate the more processes are resolved explicitly.

Despite the tremendous success of ESMs, persistent problems and uncertainties remain:

(1) A crucial quantity for the evaluation of ESMs is the equilibrium climate sensitivity (ECS), defined as the amount of equilibrium GMT increase that results from an instantaneous doubling of atmospheric carbon dioxide[9]. There remains a large ECS range in current ESM. From CMIP5 to CMIP6, the likely range of ECS has widened from 2.1–4.7°C to 1.8–5.6°C[10,11]. Reducing these uncertainties, and hence the uncertainties of future climate projections, is one of the key challenges in the development of ESMs.

(2) Both theoretical considerations and paleoclimate data suggest that several sub-systems of the Earth system can abruptly change their state in response to gradual changes in forcing[12,13]. There is concern that current ESMs will not be capable of predicting future abrupt climate changes, because the instrumental era of less than two centuries has not experienced comparable transitions, and model validation against paleoclimate data evidencing such events remains impossible due to the length of the relevant time scales[14]. In an extensive search, many relatively abrupt transitions have been identified in future projections of CMIP5 models[15], but due to the nature of these rare, high-risk events, the accuracy of ESM in predicting them remains untested.

(3) Current ESMs are not yet suitable for assessing the efficacy or the environmental impact of carbon dioxide removal techniques, which are considered key mitigation options in pathways realizing the Paris Agreement[16]. Further, ESMs still insufficiently represent key environmental processes such as the carbon cycle, water and nutrient availability, or interactions between land use and climate. This can impact the usefulness of land-based mitigation options that rely on actions such as biomass energy with carbon capture and storage or nature-based climate solutions[17,18].

(4) The distributions of time series encoding Earth system dynamics typically exhibit heavy tails. Extreme events such as heat waves and droughts, but also extreme precipitation events and associated floods, have always caused tremendous socio-economic damages. With ongoing anthropogenic climate change, such events are projected to become even more severe, and the attribution of extremes poses an-

other outstanding challenge of Earth system science[19]. While current ESMs are very skillful in predicting average values of climatic quantities, there remains room for improvement in representing extremes.

In addition to the possible solutions to these fundamental challenges, improvements of the overall accuracy of ESMs can be expected from more extensive and more systematic integration of the process-based numerical models with observational data. Earth system observations (ESOs) are central to ESMs, serving a multitude of purposes. ESOs are used to evaluate and compare process-based model performance, to generate model parameters and initial model states, or as boundary forcing of ESMs[20,21]. ESOs are also used to directly influence the model output by either tuning or nudging parameters that describe unmodeled processes, or by the more sophisticated methods of data assimilation that alter the model's state variables to bring the model output in better agreement with the observations[22]. Gradient-based optimization, as in four-dimensional variational (4DVar) schemes, is the current state of the art for efficiency and accuracy, but currently requires time consuming design and implementation of adjoint calculation routines tailored to each model. Ensemble-based Kalman filter (EnKF) schemes are gradient-free but produce unphysical outputs and rely on strong statistical assumptions that are often unsatisfied, leading to biases and overconfident predictions[23]. The main problems of contemporary ESM data assimilation are 1) nonlinear dynamics and non-Gaussian error budgets in combination with the high dimensionality of many ESM components[24–26], and 2) constraining the governing processes over the different spatio-temporal scales found in coupled systems[27,28].

ESOs cover a wide range of spatio-temporal scales and types, ranging from a couple of centimeters to tens of thousands of kilometers, and from seconds and decades to millennia. The types of observations range from in-situ measurements of irregular times and spaces to global satellite-based data fields. Yet, the available observational data pool still contains large gaps in time and space that prevent building a holistic observation-driven picture of the coupled Earth system, which result from insufficient data resolution, too short observation time periods, and largely unobserved compartments of Earth systems like, for instance, abyssal oceans. The combination of these complex characteristics render Earth system observations both challenging and particularly interesting for AI applications.

From Machine Learning-based Data Exploration Towards Learning Physics

In contrast to other research branches[29–32], the usage of ML in Earth and climate sciences is still in its infancy. While current ML applications are mostly found in explorative studies and are still far away from operational usage, profound impact on research as well as on the supercomputing industry is expected[33]. A key observation is that ML concepts from computer vision and automated image analysis can be isomorphically transferred to ESO imagery and time series[34,35]. Pioneering studies demonstrated the feasibility of ML for remote sensing data analysis, classification tasks, and parameter inversion already in the 1990s[36–39], and climate-model emulation in the early 2000s[40]. The figurative Cambrian explosion of AI techniques in Earth and climate sciences, however, only began over the last five years and will rapidly continue throughout the coming decades.

ML has been applied across various spatial and temporal scales, ranging from short-term regional weather prediction to Earth-spanning climate phenomena. Significant progress has been made in developing purely data-driven weather prediction networks, aiming to explore alternative approaches to process-based model forecasts[41–43], or to emulate and accelerate computationally demanding components of weather forecasting systems like the parameterization of gravity wave drag[44] and the simulation of cloud processes[45]. However, current global data-driven ML weather forecasts operate on much lower resolution than state-of-the-art process-based models[46] and the lack of available training data likely prevents a closure of this gap in the near future[47]. Yet, ML for emulation and acceleration tasks could play an even more important role in this context (orange knowledge cluster in Fig. 1), particularly during the advent of exascale computing[48] and resolving related computational challenges and bottlenecks[49]. ML contributed to the pressing need to improve the predictability of natural hazards, for instance, by uncovering global extreme-rainfall teleconnections[50], or by improving long-term forecasts of the El Niño Southern Oscillation (ENSO)[51,52]. ML-based image filling techniques were utilized to reconstruct missing climate information, allowing to correct previous global temperature records[53]. Furthermore, ML was applied to analyze climate data sets, e.g., to extract specific forced signals from natural climate variability[54,55] or to predict clustered weather patterns[56]. In these applications, the ML tools function as highly specialized agents that help to uncover and categorize patterns in an automated way, which is particularly useful for observable processes that are only poorly described through physical laws or paramerizations (green knowledge cluster in Fig. 1). A key methodological advantage of ML in comparison to covariance-based spatial analysis lies in the possibility to map nonlinear processes[57,58]. At the same time, such trained neural networks lack actual physical process knowledge, as they solely function through identifying and generalizing statistical relations by minimizing pre-defined loss measures for a specific task[59]. Consequently, research on ML in Earth and climate science differs fundamentally from the previously described efforts of advancing ESMs in terms of methodological development and applicability.

Concepts of utilizing ML not only for physics-blind data analyses, but also as surrogates and methodological extensions for ESMs have only recently started to shape[60]. Scientists started pursuing the aim that ML methods learn aspects of Earth and climate physics, or at least plausibly relate cause and effect. The combination of ML with process-based modelling is the essential distinction from the previous ESO data exploration (blue knowledge cluster in Fig. 1). Lifting ML from purely diagnosis-driven usage towards the prediction of geophysical processes will also be crucial for aiding climate change research and the development of mitigation strategies[61].

Following this reasoning, ML methods can be trained with process-based model data to inherit a specific geophysical causation or even emulate and accelerate entire forward simulations. For instance, ML has been used in combination with ESMs and ESOs to invert space-borne oceanic magnetic field observations to determine the global ocean heat content[62]. Similarly, a neural network has been trained with a continental hydrology model to recover high-resolution terrestrial water storage from satellite gravimetry[63]. ML plays an important role for upscaling unevenly distributed carbon flux measurements to improve global carbon monitoring systems[64]. The eddy covariance technique was combined with ML to measure the net ecosystem exchange of CO_2 between ecosystems and the atmosphere, offering a unique opportunity to study ecosystem responses to climate change[65]. ML has shown success in representing subgrid-scale processes and other parameterizations of ESMs, given that sufficient training data were available. As such, neural networks were applied to approximate turbulent processes in ocean models[66] and atmospheric subgrid processes in climate models[67]. Here, substantial computational savings could be achieved[44,45], freeing resources that in turn could be used to improve the model simulations, e.g., by raising ensemble sizes or improve the numerical model's resolution. Several studies highlight the potential for ML-based parameterization schemes[68–72], helping step-by-step to gradually remove numerically and human-induced simplifications and other biases of ESMs[73]. Nevertheless, most ML parameterization schemes are still applied under idealized conditions, e.g., coarse model resolution, simplified physics, or reduced prognostic model variables. Transferring and testing these achievements on more complex ESM configurations remains an ongoing and open challenge[74].

While some well-trained ML tools and simple hybrids have shown higher predictive power than traditional process-based models, only the surface of new possibilities, but also of new scientific challenges, has been scratched. So far, ML, ESMs, and ESO have largely been independent tools. Yet, we have reached the understanding that applications of physics-aware ML and model-network hybrids pose huge benefits by filling up niches where purely process-based models persistently lack reliability[75].

The Fusion of Process-based Models and Artificial Intelligence

While the idea of hybrids of process-based and ML models is not new[76], the understanding of how ML can enhance process-based modelling has evolved based on the recent advances. The long-term goal will be to consistently integrate the recently discovered advantages of ML into the already decade-long source of process knowledge in Earth system science (Fig. 2). However, this evolution does not come without methodological caveats, which need to be investigated carefully. For the sake of comparability, we distinguish between weakly coupled NESYM hybrids, i.e., an ESM or AI technique benefits from information from the respective other, and strongly coupled NESYM hybrids, i.e., fully coupled model-network combinations that dynamically exchange information between each other.

The emergent development of weak hybrids is predominantly driven by the aim to resolve the previously described ESM limitations, particularly unresolved and especially sub-grid scale processes (left branch of Fig. 2). Neural networks can emulate such processes after careful training with simulation data from a high-resolution model that resolves the processes of interest, or with relevant ESO data. The next methodological milestone will be the integration of such trained neural networks into ESMs for operational usage. First tests have indicated that the choice of the AI technique, e.g., neural networks versus random forests, seems to be crucial for the implementation of learning parameterization schemes, as they can significantly deteriorate the ESM's numerical stability[77]. Thus, it is not only important to identify how neural networks can be trained to resolve ESM limitations, but also how such ML-based schemes can be stabilized in the model physics context and how their effect on the process-based simulation can be evaluated and interpreted[78]. The limitations of ML-based parameterization approaches can vary widely for different problems or utilized models and, consequently, should be considered for each

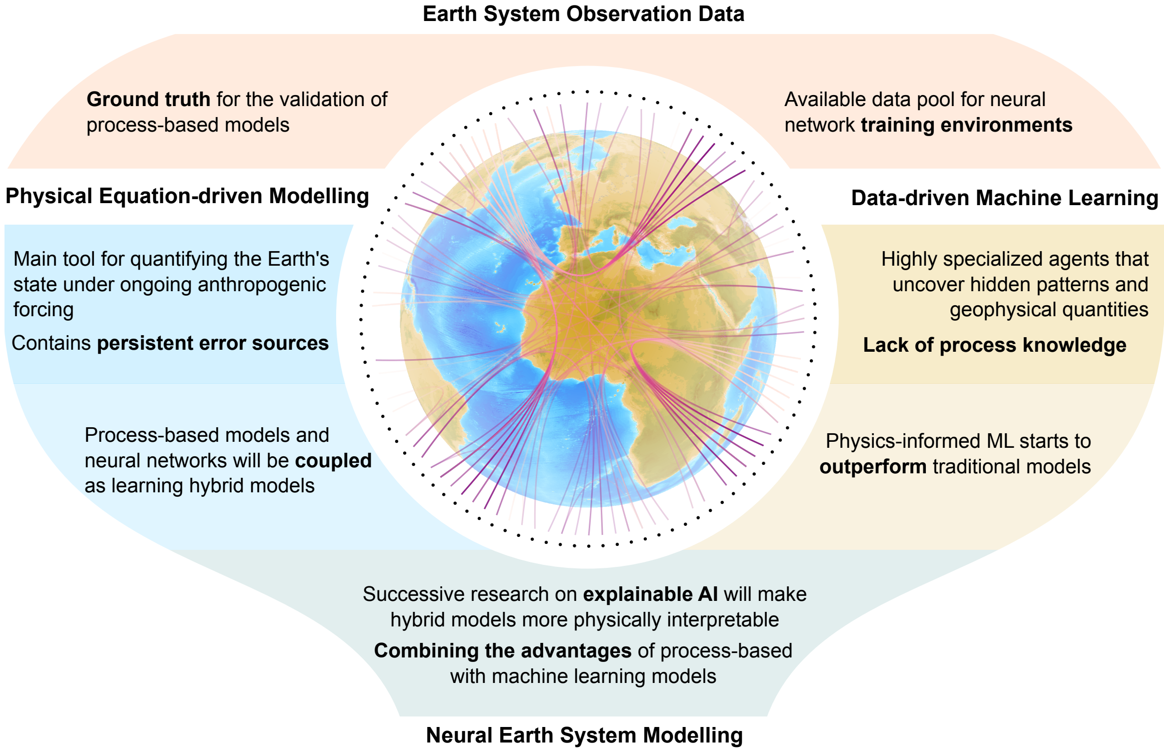

Figure 2: Successive stages of the fusion process of Earth system models and artificial intelligence towards Neural Earth System Modelling (NESYM). The left and right branches visualize the current efforts and goals for building weakly coupled hybrids (blue and yellow), which converge towards strongly coupled hybrids with the support from explainable AI (green). Details on weak and strong hybrids are provided in the main text.

learning task individually[79]. Nevertheless, several ideas have been proposed to stabilize ML parameterizations, e.g., by enforcing physical consistency through customized loss functions in neural networks and specific network architectures[71,80], or by optimizing the considered high-resolution model training data[72]. In addition, an ESM blueprint has been proposed, in which learning parameterizations can be targeted through searching an optimal fit of statistical measures between ESMs, observations, and high-resolution simulations[81]. While this is not strictly applying ML, the approach is well posed for exploring parameterizations suitable for smooth climate solutions, avoiding the problems of the EnKF techniques. In such a context, further efforts have been made to enhance an ESM not with ML directly, but in combination with a data assimilation system[22]. For instance, emulating a Kalman filter scheme with ML has been investigated[82,83], an ML-based estimation of atmospheric forcing uncertainties used as error covariance information in data assimilation has been proposed[84], ML for nudged hindcasts[74], as well as further types of Kalman-network hybrids[85,86]. Despite the shown potential for combining data assimilation and ML, it should be highlighted that many current challenges of data assimilation need to be solved for respective ML approaches as well, such as robust quantification of model and observation uncertainties and the optimal use of sparse observations[87].

In the second class of weak hybrids, the model and AI tasks are transposed, such that the information flow is directed from the model towards the AI tool (right branch of Fig. 2). Here, neural networks are trained directly with model state variables, their trajectories, or with more abstract information like seasonal signals, interannual cycles, or coupling mechanisms (knowledge cluster connections in Fig. 1). The goal of the ML application might not only be model emulation, but also inverting non-linear geophysical processes[62], learning geophysical causation[88], or predicting extreme events[89,90]. In addition to these inference and generalization tasks, a key question in this sub-discipline is whether a neural network can learn to outperform the utilized process-based trainer model in terms of physical consistency or predictive power. ESOs play a vital role in this context, as they can serve as additional training constraints for a neural network training, allowing it to build independent self-evaluation measures[63].

The given examples generally work well for validation and prediction scenarios within the given training data distribution. Out-of-distribution samples, in contrast, pose a huge challenge for supervised learning, which renders the "learning from the past" principle possibly ill-posed for prediction tasks in NESYM. Because of the both naturally and anthropogenically induced non-stationarity of the climate and Earth system, it will be very challenging – and in many cases impossible – for purely data-driven AI methods to perform accurate climate projections on their own. Nevertheless, some hope for purely data-driven AI approaches may remain for problems for which it can be convincingly argued that, for instance, the data distributions for colder-climate training conditions and warmer-climate projections overlap. But in practise it will be hard to guarantee that the unseen domains of the data distributions corresponding to a warmer climate are not relevant for a given process under study. Moreover, in specific cases the scales of the processes under study may to a first approximation be separable from the scales relevant in the context of anthropogenic climate change, but guaranteeing this in practice will again be very difficult.

Overcoming the overall limitations posed by the non-stationarity of the climate system requires a deeper holistic integration in terms of strongly coupled hybrids and the consideration of further, less constrained training techniques like unsupervised training[91] and generative AI methods[69,92,93]. For example, problems of pure AI methods with non-stationary training data can be attenuated by combining them with physical equations describing the changing energy-balance of the Earth system due to anthropogenic greenhouse-gas emissions[94]. A key distinction of strongly coupled hybrids is that the ML component can be further improved by continued training. As such, the dynamic exchange of information does not only mean that the ML part is repeatedly called after being trained for usage in a weak hybrid, but can further evolve based on the current model state, newly available observations, etc. In addition, first steps towards physics-informed AI have been made by ML-based and data-driven discovery of physical equations[95] and by the implementation of neural partial differential equations[96,97] into the context of climate modelling[98].

Continuous maturing of the methodological fusion process will allow building hybrids of neural networks, ESMs and ESOs that dynamically exchange information. ESMs will soon utilize output from supervised and unsupervised neural networks to optimize their physical consistency and, in turn, feed back improved information content to the ML component. ESOs form another core element and function as constraining ground truth of the AI-infused process prediction. Similar to the adversarial game of generative networks[99], or coupling mechanisms in an ESM[100], also strongly coupled NESYM hybrids will require innovative interfaces that control the exchange of information that are, so far, not available. As the methodical range of weak and strong hybrids is too large to be summarized through a single overarching definition, we formulate key characteristics and define goals of Neural Earth System Modelling:

(1) Hybrids can reproduce and predict out-of-distribution samples and extreme events,

(2) hybrids perform constrained and consistent simulations that obey physical conservation laws despite potential shortcomings of the hybrids' individual components,

(3) hybrids include integrated adaptive measures for self-validation and self-correction, and

(4) NESYM allows replicability and interpretability.

While most studies implemented neural networks for ML in this context, the term Neural Earth System Modelling includes all AI techniques that help to achieve these goals. The ultimately goal of NESYM is to help scientists improving the current forecast limits of geophysical processes and to contribute towards understanding the Earth's susceptible state in a changing climate. Consequently, not only the fusion of ESM and AI will be in the research focus, but also AI interpretability and resolving the common notion of a black box.

Peering into the Black Box

ML has emerged as a set of methods based on the combination of statistics, applied mathematics and computer science, but it comes with a unique set of hurdles. Peering into the black box and explaining the decision making process of the ML method, termed explainable AI (XAI), is critical to the use of ML tools. Especially in the physical sciences, adaptation of ML suffers from a lack of interpretability, particularly supervised ML. In contrast and in addition to XAI stands the call for interpretable AI (IAI), i.e., building specifically interpretable ML models from the beginning on, instead of explaining ML predictions through post-process diagnostics[101].

Ensuring that what is 'learned' by the machine is physically tractable or causal, and not due to trivial coincidences[102,103], is important before ML tools are used, e.g., in an ESM setting targeted at decision making. Thus, interpretability and explainability provides the user with trust in the ML output, improving its transparency. This is critical for ML use in the policy-relevant area of climate science as society is making it increasingly clear that understanding the source of AI predictive skill is of crucial importance[104,105]. By analyzing the decision making process, climate scientists will be able to better incorporate their own physical knowledge into the ML method, ultimately leading to greater confidence in predictions. Perhaps least appreciated in geoscientific applications thus far is the use of IAI and XAI to discover new science[103,106] and assist in theoretic advances[107]. For example, when a ML model is capable of making skillful predictions, XAI allows us to ask "what did it learn?". In this way, ML models can act as investigative tools for discovery.

The power of XAI for climate, ocean and weather applications has very recently been demonstrated[106,108–110]. Tools for developing XAI models are referred to as Additive Feature Attribution (AFA)[111]. For example, neural networks coupled with the XAI attribution method known as layerwise relevance propagation (LRP)[112,113] have revealed modes of variability within the climate system, sources of predictability across a range of timescales, and indicator patterns of climate change[55,106]. There is also evidence that XAI methods can be used to evaluate climate models against observations, identifying the most important climate model biases for the specific prediction task[114]. However, these methods are in their infancy and there is vast room for advancements in their application, making it explicitly appropriate to employ them within the physical sciences[103,110]. In the context of the above, we emphasize, however, that IAI and XAI approaches should go side-by-side with well-posed physical research hypotheses. Also in this regard, we again highlight the importance of combining recent methods from AI with domain-specific physical understanding and the state of the art in process-based modelling.

Unsupervised ML, can be intuitively IAI through the design of experiments. For example, applying clustering on closed model budgets of momentum ensures all relevant physics are represented, and can be interpreted in terms of the statistically dominant balances between terms[115]. Similarly, 'equation driven' ML can be used to determine the salient terms given an array of mathematical operations, and suggest interpretable sub-gridscale parameterisation developments on this basis[66,95]. In this manner, dominant physical mechanisms or equation terms can be determined, generating new knowledge in physics and beyond[91,115,116]. Knowledge of dominant regimes can subsequently be used to engineer features for a well posed XAI application where the source of predictive ML skill is transparent[110]. Adversarial learning has been an effective tool for generating super-resolution fields of atmospheric variables in climate models[93]. Furthermore, unsupervised ML approaches have been proposed for discovering and quantifying causal interdependencies and dynamical links inside a system, such as the Earth's climate[88,117]. Another example of an ML application that can be termed IAI is equation discovery, for example using Relevance Vector Machines (RVM), which has been applied for ocean eddy parameterizations[95]. It is also of note that a revolution of analysis tools has been called for to evaluate climate models, and ML is poised to be part of this change[60,118,119].

Given the importance of both explainability and interpretability for improving ML generalization and scientific discovery, promoting collaborations between climate and AI scientists can help to develop methods that are tailored to the field's needs. This is not just an interesting exercise - it is essential for the proper use of AI for NESYM development and use. Earth and climate scientists can aid the development of consistent benchmarks that allow evaluating both stand-alone ML and hybrids in terms of geophysical consistency[120]. However, help of the AI community is needed to resolve other recently highlighted ML pitfalls. For example, identifying and avoiding shortcut learning[121] in hybrid models, developing ESM concepts of adversarial examples and deep learning artifacts[122], and developing AFA[123] tools appropriate for physical applications such as within XAI[110]. Only through combined efforts and continuous development of both ESM and AI can Neural Earth System Modelling emerge.

Concluding remarks

Our perspective should not only be seen as the outline of a promising scientific pathway to achieve a better understanding of the Earth's present and future state, but also as an answer to the recent support call from the AI community[124]. Based on the current state of applying AI to Earth system and climate sciences, further exploration of the full potential and, equally, the limits of AI in this field is important. Yet, this line of research is a high-risk venture with many potential pitfalls and dead ends. At this point, there is no guarantee that AI will be the missing key to overcome the grand challenges of Earth and climate sciences, some of which were described at the beginning of this perspective. In its current stage, it also seems unlikely that AI alone can solve the climate prediction problem. In the coming years, AI will necessarily need to rely on clear, physically meaningful research hypotheses, the geophysical determinism of process-based modelling, and on careful human evaluation against domain-specific knowledge. Along such lines, we believe that lasting progress beyond the current hype in applying AI to Earth system science will be possible. However, once we find solutions to the foreseeable limitations described above and can build interpretable and geophysically consistent AI tools, this next evolutionary step will seem much more likely.

Acknowledgements

This study was funded by the Helmholtz Association and by the Initiative and Networking Fund of the Helmholtz Association through the project Advanced Earth System Modelling Capacity (ESM). NB acknowledges funding by the Volskwagen foundation and the European Union's Horizon 2020 research and innovation program under grant agreement No 820970 (TiPES). Support for EAB was provided, in part, by the U.S. National Science Foundation under AGS-1749261. MS acknowledges funding from the Cooperative Institute for Modeling the Earth System, Princeton University, under Award NA18OAR4320123 and from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce. The

statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of Princeton University, the National Oceanic and Atmospheric Administration, or the U.S. Department of Commerce.

Authors' contributions

CI conceived the paper and organized the collaboration. All authors contributed to writing and revising all chapters of this manuscript. In particular, NB and CI drafted the ESM overview, JSW and JS drafted the ESO and DA overview, CI and CK drafted the chapter 'From Machine Learning-based Data Exploration Towards Learning Physics', CI and JSW and NB drafted the chapter 'The Fusion of Process-based Models and Artificial Intelligence', MS and EAB and CI drafted the chapter 'Peering into the Black Box'.

Competing Interest

The authors declare no competing interest.

References

[1] Prinn, R. G. Development and application of earth system models. *Proceedings of the National Academy of Sciences* **110**, 3673–3680 (2013). URL `https://www.pnas.org/content/110/Supplement_1/3673`. `https://www.pnas.org/content/110/Supplement_1/3673.full.pdf`.

[2] Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of cmip5 and the experiment design. *Bulletin of the American Meteorological Society* **93**, 485 – 498 (2012). URL `https://journals.ametsoc.org/view/journals/bams/93/4/bams-d-11-00094.1.xml`.

[3] Stocker, T. *et al.* (eds.) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013). URL `www.climatechange2013.org`.

[4] Eyring, V. *et al.* Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development* **9**, 1937–1958 (2016). URL `https://gmd.copernicus.org/articles/9/1937/2016/`.

[5] Lin, J. W.-B. & Neelin, J. D. Considerations for stochastic convective parameterization. *Journal of the Atmospheric Sciences* **59**, 959 – 975 (2002). URL `https://journals.ametsoc.org/view/journals/atsc/59/5/1520-0469_2002_059_0959_cfscp_2.0.co_2.xml`.

[6] Klein, R. Scale-dependent models for atmospheric flows. *Annual Review of Fluid Mechanics* **42**, 249–274 (2010).

[7] Berner, J. *et al.* Stochastic parameterization: Toward a new view of weather and climate models. *Bulletin of the American Meteorological Society* **98**, 565 – 588 (2017). URL `https://journals.ametsoc.org/view/journals/bams/98/3/bams-d-15-00268.1.xml`.

[8] Knutti, R. Should we believe model predictions of future climate change? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **366**, 4647–4664 (2008). URL `https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2008.0169`. `https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2008.0169`.

[9] Knutti, R., Rugenstein, M. A. & Hegerl, G. C. Beyond equilibrium climate sensitivity. *Nature Geoscience* **10**, 727–736 (2017).

[10] Meehl, G. A. *et al.* Context for interpreting equilibrium climate sensitivity and transient climate response from the cmip6 earth system models. *Science Advances* **6** (2020). URL `https://advances.sciencemag.org/content/6/26/eaba1981`. `https://advances.sciencemag.org/content/6/26/eaba1981.full.pdf`.

[11] Zelinka, M. D. *et al.* Causes of higher climate sensitivity in cmip6 models. *Geophysical Research Letters* **47**, e2019GL085782 (2020). URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL085782`. E2019GL085782 10.1029/2019GL085782, `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019GL085782`.

[12] Lenton, T. M. *et al.* Tipping elements in the earth's climate system. *Proceedings of the National Academy of Sciences* **105**, 1786–1793 (2008). URL `https://www.pnas.org/content/105/6/1786`. `https://www.pnas.org/content/105/6/1786.full.pdf`.

[13] Boers, N., Ghil, M. & Rousseau, D.-D. Ocean circulation, ice shelf, and sea ice interactions explain dansgaard–oeschger cycles. *Proceedings of the National Academy of Sciences* **115**, E11005–E11014 (2018). URL `https://www.pnas.org/content/115/47/E11005`. `https://www.pnas.org/content/115/47/E11005.full.pdf`.

[14] Valdes, P. Built for stability. *Nature Geoscience* **4**, 414–416 (2011).

[15] Drijfhout, S. *et al.* Catalogue of abrupt shifts in intergovernmental panel on climate change climate models. *Proceedings of the National Academy of Sciences* **112**, E5777–E5786 (2015). URL `https://www.pnas.org/content/112/43/E5777`. `https://www.pnas.org/content/112/43/E5777.full.pdf`.

[16] Masson-Delmotte, V. *et al.* (eds.) *IPCC: Global Warming of 1.5 Degree Celsius: IPCC Special Report on the Impacts of Global Warming of 1.5 Degree Celsius* (Intergovernmental Panel on Climate Change (IPCC), 2018). URL `https://www.ipcc.ch/sr15`.

[17] Shukla, P. *et al.* (eds.) *Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable landmanagement, food security, and greenhouse gas*

fluxes in terrestrial ecosystems (Intergovernmental Panel on Climate Change (IPCC), 2019). URL https://www.ipcc.ch/srccl/.

[18] Pörtner, H. *et al.* (eds.) *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (Intergovernmental Panel on Climate Change (IPCC), 2019). URL https://www.ipcc.ch/srocc/.

[19] Otto, F. E. *et al.* Attribution of extreme weather events in Africa: a preliminary exploration of the science and policy implications. *Climatic Change* **132**, 531–543 (2015).

[20] Balsamo, G. *et al.* Satellite and in situ observations for advancing global earth surface modelling: A review. *Remote Sensing* **10** (2018). URL https://www.mdpi.com/2072-4292/10/12/2038.

[21] Hersbach, H. *et al.* The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146**, 1999–2049 (2020). URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803. https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3803.

[22] Evensen, G. *Data assimilation: the ensemble Kalman filter* (Springer Science & Business Media, 2009).

[23] Houtekamer, P. L. & Zhang, F. Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review* **144**, 4489–4532 (2016). URL http://journals.ametsoc.org/doi/10.1175/MWR-D-15-0440.1.

[24] van Leeuwen, P. J. Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society* **136**, 1991–1999 (2010). URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.699. https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.699.

[25] van Leeuwen, P. J., Künsch, H. R., Nerger, L., Potthast, R. & Reich, S. Particle filters for high-dimensional geoscience applications: A review. *Quarterly Journal of the Royal Meteorological Society* **145**, 2335–2365 (2019). URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3551. https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3551.

[26] Vetra-Carvalho, S. *et al.* State-of-the-art stochastic data assimilation methods for high-dimensional non-gaussian problems. *Tellus A: Dynamic Meteorology and Oceanography* **70**, 1–43 (2018). URL https://doi.org/10.1080/16000870.2018.1445364. https://doi.org/10.1080/16000870.2018.1445364.

[27] Penny, S. G. *et al.* Strongly coupled data assimilation in multiscale media: Experiments using a quasi-geostrophic coupled model. *Journal of Advances in Modeling Earth Systems* **11**, 1803–1829 (2019). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001652. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019MS001652.

[28] Browne, P. A., de Rosnay, P., Zuo, H., Bennett, A. & Dawson, A. Weakly coupled ocean–atmosphere data assimilation in the ecmwf nwp system. *Remote Sensing* **11** (2019). URL https://www.mdpi.com/2072-4292/11/3/234.

[29] Voulodimos, A., Doulamis, N., Doulamis, A. & Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience* **2018** (2018).

[30] Brown, T. B. *et al.* Language models are few-shot learners. *arXiv* (2020). `2005.14165`.

[31] Loh, E. Medicine and the rise of the robots: A qualitative review of recent advances of artificial intelligence in health. *BMJ Leader* **2**, 59–63 (2018).

[32] Girasa, R. Ai as a disruptive technology. In *Artificial Intelligence as a Disruptive Technology*, 3–21 (Springer, 2020).

[33] Bauer, P. *et al.* The digital revolution of earth-system science. *Nature Computational Science* **1**, 104–113 (2021).

[34] Lary, D. J., Alavi, A. H., Gandomi, A. H. & Walker, A. L. Machine learning in geosciences and remote sensing. *Geoscience Frontiers* **7**, 3–10 (2016). URL `http://dx.doi.org/10.1016/j.gsf.2015.07.003http://linkinghub.elsevier.com/retrieve/pii/S1674987115000821`.

[35] Salcedo-Sanz, S. *et al.* Machine learning information fusion in Earth observation: A comprehensive review of methods, applications and data sources. *Information Fusion* **63**, 256–272 (2020).

[36] Dawson, M., Olvera, J., Fung, A. & Manry, M. Inversion of Surface Parameters Using Fast Learning Neural Networks. In *[Proceedings] IGARSS '92 International Geoscience and Remote Sensing Symposium*, vol. 2, 910–912 (IEEE, 1992). URL `http://ieeexplore.ieee.org/document/578294/`. `arXiv:1011.1669v3`.

[37] Miller, D. M., Kaminsky, E. J. & Rana, S. Neural network classification of remote-sensing data. *Computers and Geosciences* **21**, 377–386 (1995).

[38] Serpico, S. B., Bruzzone, L. & Roli, F. An experimental comparison of neural and statistical non-parametric algorithms for supervised classification of remote-sensing images. *Pattern Recognition Letters* **17**, 1331–1341 (1996). URL `https://doi.org/10.1016/S0167-8655(96)00090-6`.

[39] Hsieh, W. W. & Tang, B. Applying Neural Network Models to Prediction and Data Analysis in Meteorology and Oceanography. *Bulletin of the American Meteorological Society* **79**, 1855–1870 (1998). URL `http://journals.ametsoc.org/doi/abs/10.1175/1520-0477{%}281998{%}29079{%}3C1855{%}3AANNMTP{%}3E2.0.CO{%}3B2`. `z0022`.

[40] Knutti, R., Stocker, T. F., Joos, F. & Plattner, G. K. Probabilistic climate change projections using neural networks. *Climate Dynamics* **21**, 257–272 (2003).

[41] Arcomano, T. *et al.* A Machine Learning-Based Global Atmospheric Forecast Model. *Geophysical Research Letters* **47**, 1–9 (2020).

[42] Weyn, J. A., Durran, D. R. & Caruana, R. Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data. *Journal of Advances in Modeling Earth Systems* (2019).

[43] Weyn, J. A., Durran, D. R. & Caruana, R. Improving Data-Driven Global Weather Prediction Using Deep Convolutional Neural Networks on a Cubed Sphere. *Journal of Advances in Modeling Earth Systems* **12** (2020).

[44] Chantry, M., Hatfield, S., Duben, P., Polichtchouk, I. & Palmer, T. Machine learning emulation of gravity wave drag in numerical weather forecasting (2021). `2101.08195`.

[45] Gettelman, A. *et al.* Machine learning the warm rain process. *Journal of Advances in Modeling Earth Systems* **13**, e2020MS002268 (2021). URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002268`. E2020MS002268 2020MS002268, `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020MS002268`.

[46] Rasp, S. & Thuerey, N. Data-driven medium-range weather prediction with a resnet pretrained on climate simulations: A new model for weatherbench. *Journal of Advances in Modeling Earth Systems* **13**, e2020MS002405 (2021). URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002405`. E2020MS002405 2020MS002405, `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020MS002405`.

[47] Palmer, T. A vision for numerical weather prediction in 2030 (2020). `2007.04830`.

[48] Neumann, P. *et al.* Assessing the scales in numerical weather and climate predictions: will exascale be the rescue? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **377**, 20180148 (2019). URL `https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2018.0148`. `https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2018.0148`.

[49] Kurth, T. *et al.* Exascale deep learning for climate analytics. In *SC18: International Conference for High Performance Computing, Networking, Storage and Analysis*, 649–660 (2018).

[50] Boers, N. *et al.* Complex networks reveal global pattern of extreme-rainfall teleconnections. *Nature* **566**, 373–377 (2019). URL `http://dx.doi.org/10.1038/s41586-018-0872-x`.

[51] Ham, Y.-g., Kim, J.-h. & Luo, J.-j. Deep learning for multi-year ENSO forecasts. *Nature* **573**, 568–572 (2019). URL `http://dx.doi.org/10.1038/s41586-019-1559-7http://www.nature.com/articles/s41586-019-1559-7`.

[52] Yan, J., Mu, L., Wang, L., Ranjan, R. & Zomaya, A. Y. Temporal Convolutional Networks for the Advance Prediction of ENSO. *Scientific Reports* **10**, 1–15 (2020).

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

[54] Barnes, E. A., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C. & Anderson, D. Viewing Forced Climate Patterns Through an AI Lens. *Geophysical Research Letters* **46**, 13389–13398

(2019). URL https://doi.org/10.1029/2019GL084944https://onlinelibrary.wiley.com/doi/10.1029/2019GL084944.

[55] Barnes, E. A. *et al.* Indicator Patterns of Forced Change Learned by an Artificial Neural Network. *Journal of Advances in Modeling Earth Systems* **12** (2020). 2005.12322.

[56] Chattopadhyay, A., Hassanzadeh, P. & Pasha, S. Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data. *Scientific Reports* **10**, 1317 (2020). URL http://dx.doi.org/10.1038/s41598-020-57897-9http://arxiv.org/abs/1811.04817http://www.nature.com/articles/s41598-020-57897-9. 1811.04817.

[57] Ramachandran, P., Zoph, B. & Le, Q. V. Searching for activation functions. *arXiv* 1–13 (2017). 1710.05941.

[58] Lu, Z., Hunt, B. R. & Ott, E. Attractor reconstruction by machine learning. *Chaos* **28** (2018). URL http://dx.doi.org/10.1063/1.5039508. 1805.03362.

[59] Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986). arXiv:1011.1669v3.

[60] Reichstein, M. *et al.* Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019). URL http://dx.doi.org/10.1038/s41586-019-0912-1.

[61] Huntingford, C. *et al.* Machine learning and artificial intelligence to aid climate change research and preparedness. *Environmental Research Letters* **14** (2019).

[62] Irrgang, C., Saynisch, J. & Thomas, M. Estimating global ocean heat content from tidal magnetic satellite observations. *Scientific Reports* **9**, 7893 (2019). URL http://dx.doi.org/10.1038/s41598-019-44397-8.

[63] Irrgang, C., Saynisch-Wagner, J., Dill, R., Boergens, E. & Thomas, M. Self-Validating Deep Learning for Recovering Terrestrial Water Storage From Gravity and Altimetry Measurements. *Geophysical Research Letters* **47** (2020). URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2020GL089258https://onlinelibrary.wiley.com/doi/10.1029/2020GL089258.

[64] Jung, M. *et al.* Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the fluxcom approach. *Biogeosciences* **17**, 1343–1365 (2020). URL https://bg.copernicus.org/articles/17/1343/2020/.

[65] Tramontana, G. *et al.* Partitioning net carbon dioxide fluxes into photosynthesis and respiration using neural networks. *Global Change Biology* **26**, 5235–5253 (2020). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.15203. https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.15203.

[66] Bolton, T. & Zanna, L. Applications of Deep Learning to Ocean Data Inference and Subgrid Parameterization. *Journal of Advances in Modeling Earth Systems* **11**, 376–399 (2019). URL `http://doi.wiley.com/10.1029/2018MS001472`.

[67] Rasp, S., Pritchard, M. S. & Gentine, P. Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences* **115**, 9684–9689 (2018). URL `http://arxiv.org/abs/1806.04731http://www.pnas.org/lookup/doi/10.1073/pnas.1810286115`. `1806.04731`.

[68] O'Gorman, P. A. & Dwyer, J. G. Using Machine Learning to Parameterize Moist Convection: Potential for Modeling of Climate, Climate Change, and Extreme Events. *Journal of Advances in Modeling Earth Systems* (2018). `1806.11037`.

[69] Gagne, D. J., Christensen, H. M., Subramanian, A. C. & Monahan, A. H. Machine Learning for Stochastic Parameterization: Generative Adversarial Networks in the Lorenz '96 Model. *Journal of Advances in Modeling Earth Systems* **12** (2020). `1909.04711`.

[70] Han, Y., Zhang, G. J., Huang, X. & Wang, Y. A Moist Physics Parameterization Based on Deep Learning. *Journal of Advances in Modeling Earth Systems* 0–2 (2020).

[71] Beucler, T., Pritchard, M., Gentine, P. & Rasp, S. Towards Physically-consistent, Data-driven Models of Convection. *arXiv* 2–6 (2020). URL `http://arxiv.org/abs/2002.08525`. `2002.08525`.

[72] Yuval, J. & O'Gorman, P. A. Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature Communications* **11**, 1–10 (2020). URL `http://dx.doi.org/10.1038/s41467-020-17142-3`. `2001.03151`.

[73] Brenowitz, N. D. & Bretherton, C. S. Prognostic Validation of a Neural Network Unified Physics Parameterization. *Geophysical Research Letters* 1–10 (2018). URL `https://agupubs-onlinelibrary-wiley-com.emedien.ub.uni-muenchen.de/doi/pdf/10.1029/2018GL078510http://doi.wiley.com/10.1029/2018GL078510`.

[74] Watt-Meyer, O. *et al.* Correcting weather and climate models by machine learning nudged historical simulations. *Earth and Space Science Open Archive* 13 (2021). URL `https://doi.org/10.1002/essoar.10505959.1`.

[75] Pathak, J. *et al.* Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **28**, 041101 (2018). URL `https://doi.org/10.1063/1.5028373`. `https://doi.org/10.1063/1.5028373`.

[76] Krasnopolsky, V. M. & Fox-Rabinovitz, M. S. Complex hybrid models combining deterministic and machine learning components for numerical climate modeling and weather prediction. *Neural Networks* **19**, 122–134 (2006).

[77] Brenowitz, N. D. *et al.* Machine Learning Climate Model Dynamics: Offline versus Online Performance. *ArXiv* 1–6 (2020). URL `http://arxiv.org/abs/2011.03081`. `2011.03081`.

[78] Brenowitz, N. D., Beucler, T., Pritchard, M. & Bretherton, C. S. Interpreting and Stabilizing Machine-Learning Parametrizations of Convection. *Journal of the Atmospheric Sciences* **77**, 4357–4375 (2020). URL `https://journals.ametsoc.org/view/journals/atsc/77/12/jas-d-20-0082.1.xml`. `2003.06549`.

[79] Seifert, A. & Rasp, S. Potential and limitations of machine learning for modeling warm-rain cloud microphysical processes. *Journal of Advances in Modeling Earth Systems* (2020). URL `https://onlinelibrary.wiley.com/doi/10.1029/2020MS002301`.

[80] Beucler, T., Rasp, S., Pritchard, M. & Gentine, P. Achieving conservation of energy in neural network emulators for climate modeling. *arXiv* 2–6 (2019). `1906.06622`.

[81] Schneider, T., Lan, S., Stuart, A. & Teixeira, J. Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations. *Geophysical Research Letters* **44**, 12,396–12,417 (2017). `1709.00037`.

[82] Cintra, R. S. & Velho, H. F. d. C. Data Assimilation by Artificial Neural Networks for an Atmospheric General Circulation Model: Conventional Observation. *Bulletin of the American meteorological Society* **77**, 437–471 (2014). URL `http://arxiv.org/abs/1407.4360http://www.intechopen.com/books/advanced-applications-for-artificial-neural-networks/data-assimilation-by-artificial-neural-networks-for-an-atmospheric-general-circulation-model`. `1407.4360`.

[83] Wahle, K., Staneva, J. & Guenther, H. Data assimilation of ocean wind waves using Neural Networks. A case study for the German Bight. *Ocean Modelling* **96**, 117–125 (2015). URL `http://linkinghub.elsevier.com/retrieve/pii/S146350031500116Xhttp://dx.doi.org/10.1016/j.ocemod.2015.07.007`.

[84] Irrgang, C., Saynisch-Wagner, J. & Thomas, M. Machine Learning-Based Prediction of Spatiotemporal Uncertainties in Global Wind Velocity Reanalyses. *Journal of Advances in Modeling Earth Systems* **12** (2020).

[85] Brajard, J., Carrassi, A., Bocquet, M. & Bertino, L. Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the lorenz 96 model. *Journal of Computational Science* **44**, 101171 (2020). URL `http://www.sciencedirect.com/science/article/pii/S1877750320304725`.

[86] Ruckstuhl, Y., Janjić, T. & Rasp, S. Training a convolutional neural network to conserve mass in data assimilation. *Nonlinear Processes in Geophysics Discussions* **2020**, 1–15 (2020). URL `https://npg.copernicus.org/preprints/npg-2020-38/`.

[87] Geer, A. J. Learning earth system models from observations: machine learning or data assimilation? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **379**, 20200089 (2021). URL https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2020.0089. https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2020.0089.

[88] Runge, J. *et al.* Inferring causation from time series in Earth system sciences. *Nature Communications* **10**, 1–13 (2019). URL http://dx.doi.org/10.1038/s41467-019-10105-3.

[89] Boers, N. *et al.* Prediction of extreme floods in the eastern Central Andes based on a complex networks approach. *Nature Communications* **5**, 1–7 (2014).

[90] Qi, D. & Majda, A. J. Using machine learning to predict extreme events in complex systems. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 52–59 (2020).

[91] Sonnewald, M., Dutkiewicz, S., Hill, C. & Forget, G. Elucidating ecological complexity: Unsupervised learning determines global marine eco-provinces. *Science Advances* **6**, 1–12 (2020).

[92] Leinonen, J., Guillaume, A. & Yuan, T. Reconstruction of Cloud Vertical Structure With a Generative Adversarial Network. *Geophysical Research Letters* **46**, 7035–7044 (2019). URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2019GL082532.

[93] Stengel, K., Glaws, A., Hettinger, D. & King, R. N. Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 16805–16815 (2020).

[94] Huber, M. & Knutti, R. Anthropogenic and natural warming inferred from changes in Earth's energy balance. *Nature Geoscience* **5**, 31–36 (2012).

[95] Zanna, L. & Bolton, T. Data-driven Equation Discovery of Ocean Mesoscale Closures. *Geophysical Research Letters* 1–13 (2020).

[96] Lagaris, I. E., Likas, A. & Fotiadis, D. I. Artificial neural networks for solving ordinary and partial differential equations. *IEEE Transactions on Neural Networks* **9**, 987–1000 (1998). 9705023.

[97] Raissi, M., Perdikaris, P. & Karniadakis, G. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics* **378**, 686–707 (2019). URL https://linkinghub.elsevier.com/retrieve/pii/S0021999118307125.

[98] Ramadhan, A. *et al.* Capturing missing physics in climate model parameterizations using neural differential equations. *arXiv* (2020). URL http://arxiv.org/abs/2010.12559. 2010.12559.

[99] Goodfellow, I. J. *et al.* Generative adversarial networks. *arXiv* (2014). 1406.2661.

[100] Hurrell, J. W. *et al.* The community earth system model: A framework for collaborative research. *Bulletin of the American Meteorological Society* **94**, 1339 – 1360 (01 Sep. 2013). URL https://journals.ametsoc.org/view/journals/bams/94/9/bams-d-12-00121.1.xml.

[101] Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* **1**, 206–215 (2019). URL http://dx.doi.org/10.1038/s42256-019-0048-x. 1811.10154.

[102] Balaji, V. Climbing down Charney's ladder: Machine Learning and the post-Dennard era of computational climate science. *arXiv* (2020). URL http://arxiv.org/abs/2005.11862. 2005.11862.

[103] Sonnewald, M. *et al.* Bridging observation, theory and numerical simulation of the ocean using machine learning. *arXiv* (2021). 2104.12506.

[104] Ethics guidelines for trustworthy ai. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (2019). Accessed: 2021-01-06.

[105] Executive order on promoting the use of trustworthy artificial intelligence in the federal government. https://www.whitehouse.gov/presidential-actions/executive-order-promoting-use-trustworthy-artificial-intelligence-federal-government/ (2020). Accessed: 2021-01-06.

[106] Toms, B. A., Barnes, E. A. & Ebert-Uphoff, I. Physically Interpretable Neural Networks for the Geosciences: Applications to Earth System Variability. *Journal of Advances in Modeling Earth Systems* **12**, 1–20 (2020). URL http://arxiv.org/abs/1912.01752https://onlinelibrary.wiley.com/doi/abs/10.1029/2019MS002002. 1912.01752.

[107] Kaiser, B. E., Saenz, J. A., Sonnewald, M. & Livescu, D. Objective discovery of dominant dynamical processes with intelligible machine learning. *arXiv* (2021). 2106.12963.

[108] McGovern, A. *et al.* Making the black box more transparent: Understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society* **100**, 2175 – 2199 (01 Nov. 2019). URL https://journals.ametsoc.org/view/journals/bams/100/11/bams-d-18-0195.1.xml.

[109] Ebert-Uphoff, I. & Hilburn, K. Evaluation, tuning and interpretation of neural networks for working with images in meteorological applications. *Bulletin of the American Meteorological Society* 1 – 47 (31 Aug. 2020). URL https://journals.ametsoc.org/view/journals/bams/aop/BAMS-D-20-0097.1/BAMS-D-20-0097.1.xml.

[110] Sonnewald, M. & Lguensat, R. Revealing the impact of global heating on north atlantic circulation using transparent machine learning. *Earth and Space Science Open Archive* 27 (2021). URL https://doi.org/10.1002/essoar.10506146.1.

[111] Beucler, T., Ebert-Uphoff, I., Rasp, S., Pritchard, M. & Gentine, P. Machine learning for clouds and climate (invited chapter for the agu geophysical monograph series "clouds and climate"). *Earth and Space Science Open Archive* 27 (2021). URL `https://doi.org/10.1002/essoar.10506925.1`.

[112] Olden, J. D., Joy, M. K. & Death, R. G. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling* **178**, 389 – 397 (2004). URL `http://www.sciencedirect.com/science/article/pii/S0304380004001565`.

[113] Bach, S. *et al.* On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLOS ONE* **10**, 1–46 (2015). URL `https://doi.org/10.1371/journal.pone.0130140`.

[114] Barnes, E. A., Mayer, K., Toms, B., Martin, Z. & Gordon, E. Identifying opportunities for skillful weather prediction with interpretable neural networks. *arXiv* (2020). `2012.07830`.

[115] Sonnewald, M., Wunsch, C. & Heimbach, P. Unsupervised Learning Reveals Geography of Global Ocean Dynamical Regions. *Earth and Space Science* **6**, 784–794 (2019).

[116] Callaham, J. L., Koch, J. V., Brunton, B. W., Kutz, J. N. & Brunton, S. L. Learning dominant physical processes with data-driven balance models. *Nature Communications* **12** (2021). URL `http://dx.doi.org/10.1038/s41467-021-21331-z`.

[117] Runge, J., Nowack, P., Kretschmer, M., Flaxman, S. & Sejdinovic, D. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances* **5** (2019). URL `https://advances.sciencemag.org/content/5/11/eaau4996`. `https://advances.sciencemag.org/content/5/11/eaau4996.full.pdf`.

[118] Eyring, V. *et al.* Taking climate model evaluation to the next level. *Nature Climate Change* **9**, 102–110 (2019). URL `http://dx.doi.org/10.1038/s41558-018-0355-y`.

[119] Schlund, M. *et al.* Constraining Uncertainty in Projected Gross Primary Production With Machine Learning. *Journal of Geophysical Research: Biogeosciences* (2020).

[120] Rasp, S. *et al.* WeatherBench: A benchmark dataset for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems* (2020). URL `http://arxiv.org/abs/2002.00469`. `2002.00469`.

[121] Geirhos, R. *et al.* Shortcut learning in deep neural networks. *Nature Machine Intelligence* **2**, 665–673 (2020). URL `http://dx.doi.org/10.1038/s42256-020-00257-zhttp://www.nature.com/articles/s42256-020-00257-z`. `2004.07780`.

[122] Buckner, C. Understanding adversarial examples requires a theory of artefacts for deep learning. *Nature Machine Intelligence* (2020). URL `http://dx.doi.org/10.1038/s42256-020-00266-y`.

[123] Alvarez-Melis, D. & Jaakkola, T. S. On the robustness of interpretability methods. *arXiv* (2018). `1806.08049`.

[124] Rolnick, D. *et al.* Tackling Climate Change with Machine Learning. *arXiv* (2019). URL http://arxiv.org/abs/1906.05433. 1906.05433.