AI for Climate Science

Philip Stier, Department of Physics, University of Oxford, UK

(philip.stier@physics.ox.ac.uk)

1. Introduction

While the introduction of AI has been a recent step change in many areas of science, the beginnings of AI in climate research are much less well defined. This is because climate research is by scale inherently big data intensive so that its research community had to develop automated tools for prediction and statistical data analysis from the onset, without knowledge that these approaches will be considered as AI or Machine Learning (ML, application of AI to learn from data without instruction) in the future (or even without being aware now). For example, regression and clustering have been applied for decades to understand associations between climate variables in observational and modelling data. Probabilistic methods and Bayes' theorem have been widely used to tackle inverse problems in the inference of climate parameter from Earth Observations, as e.g. in optimal estimation retrievals of atmospheric properties (e.g. temperature, clouds, composition) from satellite data [2]. Clustering techniques, such as k-means, have been widely used to identify and attribute weather patterns and cloud regimes [3]. Causal attribution techniques have been developed to isolate anthropogenic climate signals from natural variability, generally requiring significant dimensionality reduction e.g., through Principal Component Analysis [4].

Building on this early work, recent advances in AI are now transforming climate science, across all fields, which we will investigate in the following. The focus of this chapter is less on the specific AI/ML methods applied to each problem – there are plenty of choices, they are rapidly evolving, and there generally exists limited consensus on the optimal method for each task – but rather on the key application areas and opportunities for AI/ML to transform climate science.

2. Climate modelling

Climate models span a wide range of complexities, from simple zero-dimensional energy balance models with analytical solutions, to complex numerical Earth System Models predicting the transient evolution of the key Earth system components. In the following, we will be focusing on the latter type of models with an emphasis on the physical climate system. A climate model can be generalized as a map *C*, mapping the climate state vector $\mathbf{x}(t)$, from its initial state $\mathbf{x}(t_0)$ through time t, typically with prescribed spatiotemporally varying boundary conditions, such as natural (e.g. volcanic emissions, solar radiation) and anthropogenic (e.g. concentrations of CO_2, aerosols, land use) perturbations, represented here as vector $\mathbf{b}(t)$. *C* includes structural (due to the formulation of the underlying equations representing climate processes) and parametric (due to inexact knowledge of parameters $\boldsymbol{\theta}$ in these equations) uncertainties:

$$\mathbf{x}(\mathrm{t}) \;=\; C(\,\mathbf{x}(t_0)\,, \mathbf{b}(t), \boldsymbol{\theta})$$

C here represents a wide range of complex physical and, in Earth System Models, bio-chemical processes. A climate model generally discretizes Earth spatially (e.g. structured or unstructured grids, spherical harmonics) and physical conservation laws (of momentum, energy, mass), represented as systems of coupled partial differential equations, are numerically integrated forward in time. However, to maintain numerical stability [5], this time integration must be performed with short integration timesteps; for a typical climate model with 1 degree × 1 degree (approximately 100km) resolution, the timestep is typically O(10min), which has to be further reduced for higher spatial resolutions as will be discussed below. For a typical transient climate simulation from year 1700 to 2100 this requires $O(10^7)$ time-integration steps solving a complex set of equations for $O(10^6)$ grid-points. Note that climate models effectively integrate random manifestations of weather over time and are structurally often very similar to weather forecasting models so that many applications of AI discussed below apply to both domains.

However, key climate processes occur on small scales that cannot be explicitly resolved in such climate models. For example, clouds form droplets microscopically on a scale of O(10^{-6} m), which grow by diffusion and collisions to form precipitation of O(10^{-3} m). Macroscopically, clouds typically occur on scales of O(10^{2} m) to O(10^{5} m). As it is computationally not possible to globally resolve all of these scales, such processes are *parametrized*, i.e., approximated from the explicitly resolved large-scale climate variables. While tremendously successful in weather and climate prediction, such simplified parametrizations introduce structural deficiencies by construction, which contribute significantly to the remaining overall model uncertainty.

Another, often overlooked, problem is that climate models contain many such parameterizations to represent e.g., clouds, radiation, turbulence, gravity waves, oceanic mesoscale eddies. Due to the underlying complexity these are generally solved sequentially via operator splitting, separately computing the solution for each, and combining these separate solutions to calculate the overall tendency for each variable that is then integrated forward in time. In fact, typical climate models have three operator-split parameterizations for clouds alone: i) calculating the cloud amount, ii) the convective transport of momentum, water and tracers, and iii) the cloud microphysics, i.e., the evolution of droplets of crystals.

Physical climate modelling has pushed the envelopes of supercomputing from its inception; hence, it is not surprising that the climate modelling community now embraces the opportunities provided by the advancement of AI and machine learning as we will discuss below.

2.1 Emulating climate model parameterizations

Based on a long heritage of climate model development, most AI applications in climate modelling have focused on the emulation of climate model parameterizations *in existing climate modelling frameworks* with two main objectives:

i) Speedup, implementing emulators of existing parameterizations for reduced computational cost. Applications include:
   - Emulation of atmospheric radiative transfer using neural networks: it has been demonstrated that a neural network emulation of a longwave radiation code in a climate model achieved a speedup by 50-80 times faster than the original parameterization [6], an approach which has subsequently been extended to the shortwave radiation and improved in accuracy [7].
   - Emulation of aerosol microphysics: testing various ML approaches, such as neural networks, random forests and gradient boosting: it has been shown that a neural network can successfully emulate an aerosol microphysics module in a climate model, including physical constraints, with a speedup of over 60 times than the original model [8].

ii) Accuracy: the speedup provided by AI/ML allows for replacing (computationally affordable) simplified parameterizations with emulations of (previously computationally unaffordable) higher accuracy reference models. Applications include:
   - Replacement of uncertain atmospheric convective cloud parameterizations with neural networks, e.g., emulating 2D cloud resolving models for each column (superparameterization). [9] use a nine-layer deep, fully connected network with 256 nodes in each layer with around 0.5 million parameters to successfully replace the convection scheme in a climate model with multiyear simulations closely reproducing the mean climate of the cloud-resolving simulation as well as key aspects of variability. Interestingly, their neural network conserves energy approximately without incorporating this as a physical constraint during the training. Remaining problems with stability can be overcome by the explicit incorporation of physical constraints [10].
   - Emulation of gravity wave drag parameterizations: training on an increased complexity version of an existing scheme emulators have been built that produce more accurate weather

forecasts than the operational version while performing 10 times faster on GPUs than the existing scheme on a CPU [11].

- Representation of unresolved ocean mesoscale eddies: computationally efficient parameterizations of ocean mesoscale eddies have been developed based on relevance vector machines and convolutional neural networks from high-resolution simulations [12]. Physical constraints are explicitly embedded in the design of the network: the architecture forms the elements of a symmetric eddy stress tensor so that exact global momentum and vorticity conservation can be achieved, and the results are interpretable.

Based on the long heritage of specialized parameterization development, it is not entirely surprising that current approaches primarily focus on the speedup and improvement of parameterization in existing structural frameworks. However, the incorporation of AI/ML provides significant opportunities to overcome structural limitations of current climate models.

Firstly, key climate processes are coupled on timescales much faster than the climate model integration timestep, for example the interaction between aerosols and clouds. However, under current operator splitting approaches these processes can only interact every host model timestep, typically O(10min). AI/ML offers the *opportunity to emulate such coupled processes jointly*, eliminating structural operator splitting errors – an opportunity that has not yet been widely capitalized on. Secondly, AI/ML also provides an opportunity to develop entirely new parameterizations, under incorporation of observational constraints, as we will now discuss.

2.2 Development of new climate model parameterizations

The development of climate model parameterizations generally combines theory with either detailed process or high-resolution modelling or observed relationships between small scale processes and resolved scale variables (e.g., cloud fraction parameterized as a function of grid-scale atmospheric humidity [13]). However, the *systematic* incorporation of observational constraints in the parameterization development has remained challenging. Consequently, observations are often primarily used for model evaluation and parameter tuning *after the structural model development* and often only based on aggregate statistics, such as gridded monthly means of satellite observations. It is surprising –and concerning – that current climate model development and evaluation make use of only a small fraction of the information content from an unprecedented amount of Earth observations available.

However, new approaches to systematically incorporate observations and high-resolution simulations in the development of parameterizations are being developed. A framework for machine-learning based parameterizations to learn from Earth observations and targeted high-resolution simulations has been outlined [14]. In this framework, parameters and parametric functions of parameterizations are learned by minimizing carefully chosen (yet still subjective) objective functions that penalize the mismatch between the simulations and observations or between the simulations and targeted high-resolution simulations. It should be noted that such methods still incorporate structural errors introduced by the specific formulation of the parametrization itself and its spatiotemporal coverage. For example, current convection parameterizations in climate models suffer from their restriction to a single model column, which makes it difficult to accurately represent larger organized cloud systems.

A particular challenge for observation-data driven approaches in climate modelling is that, by construction, we expect the climate system to change between the training period, for which observations are available, and under future climate change. Reliable satellite-based Earth Observations are available from about 1980 to present day, i.e., from a period that has already undergone significant climate change. Such dataset shifts arising from non-stationarity of the climate processes differ from most traditional machine learning applications, for which the underlying distributions are assumed to remain the same between training and test sets. In addition, there exists a risk of introducing selection biases due to the incompleteness of Earth Observations.

Machine learning strategies to limit the impact of dataset shifts from non-stationarity include active learning [15], querying for additional constraints from observations or process modelling when dataset shifts are detected [14]. Common approaches based on feature dropping, i.e., removing features underlying the dataset shifts, are undesirable in the presence of real physical shifts in the distribution of climate state variables under climate change. [16] develop "climate-invariant" mappings of thermodynamic variables in the emulation of sub-grid scale processes by physically rescaling the inputs and outputs of ML models to facilitate their generalization to unseen climates. However, it is not yet entirely clear how such methods generalize to state-dependent and potentially discontinuous feedback process, such as phase dependent cloud feedbacks [17] or potential bifurcation points [18].

2.3 Calibration of climate and climate process models

Climate models contain a significant number of uncertain model parameters $\boldsymbol{\theta}$. For example, the convective rate of mixing of clouds with their environment is highly uncertain, due to the large uncertainty associated with small-scale turbulent processes, but has a profound impact on upper tropospheric humidity, cloudiness and ultimately climate sensitivity [19]. During the model development phase, such climate model parameters are generally tuned ad-hoc to minimize biases against a sub-set of present-day observations. However, it is possible to create climate model variants with different parameter combinations that evaluate reasonably against present day observations and still simulate very different climate sensitivity [20]. Perturbed Parameter Ensembles (PPE), varying climate model parameters within their uncertainty bounds, have been proposed as a way to probe the full parametric uncertainty [21]. However, due to the large number of uncertain climate model parameters – e.g. 47 key parameters have been identified in the HadGEM3 climate model [22] – a very large number of simulations would be required to probe uncertainty in this 47-dimensional parameter space. This approach has been pioneered in the citizen-science project climateprediction.net [19, 21] based on the distributed computing power provided by a large number of volunteers. However, the computational demand limits application to lower resolution models of limited complexity or requires reducing the number of parameters through sensitivity analysis [e.g. 22].

The availability of AI/ML based emulation of climate model outputs provides an alternative approach for uncertainty estimation and calibration for climate models. A Bayesian framework for the calibration of numerical models with atmospheric applications has been introduced [23], based on Gaussian Processes [24]. In this framework, an ensemble of climate model simulations is performed that probe the parametric uncertainty in all parameter dimensions. Latin hypercubes [25] provide an efficient sampling of the parameter space, reducing the number of simulations to perform. Emulation, e.g., using Gaussian Processes, is performed on a set of model outputs both for observational constraint, as well as to assess the impact of this constraint on key climate metrics, such as radiative perturbations to the global energy balance. Calibration of the model corresponds to the inverse problem of finding the optimal combination of parameters $\boldsymbol{\theta}$ that best match a set of observations corresponding to the emulated model output, which can be densely sampled (as opposed to the sparsely performed climate model simulations). Due to structural model uncertainties and representation errors, a model will never exactly match a comprehensive set of observations. Hence, the goal of calibration is generally to minimize a suitably chosen objective function, commonly some form of distance metrics between the model output and corresponding observations taking into account observational errors, such as those used in history matching [26] using techniques, such as Approximate Bayesian Computation (ABC) or Markov-Chain Monte Carlo (MCMC). These approaches to climate model calibration have been pioneered in the context of constraining the highly uncertain effect of aerosols (air pollution particles) on clouds and climate [27].

Open-source tools providing a general workflow for emulating and calibrating climate models with a wide range of heterogeneous observations are now becoming available [28].

2.4 Digital Twin Earths

Through confluence of advances in high-performance computing and efficient computational solvers for non-hydrostatic equations of fluid flow at small scales it is now possible to numerically simulate the global Earth’s atmosphere at cloud-resolving kilometer scales [29]. Combination with

Earth Observations at similar scales would allow to exploit a much larger fraction of the information content from observations than previously possible – and to avoid key structural deficiencies associated with parameterizations, in particular for deep convective clouds or ocean mesoscale eddies, that can be removed as the associated flows are now explicitly resolved. The computational challenge is immense: e.g., the ICON model with 2.5km resolution has 84×10^6 grid columns with 90 levels and, for numerical stability, requires integration over climatological timescales in 4.5s timesteps. Storage of just 10 climate variables approaches 1 Tb per output timestep [29] and several Pb per day. Novel AI based approaches for data compression e.g., using autoencoders learning the underlying physics, encoding the data to a lower dimensional space for output and subsequent decoding during the analysis, could make data analysis tractable again [30].

The entrance of models with superior physical realism opens the doors for the development of *digital twins* of Earth [31]. Digital Twin Earths combine coupled physical and biogeochemical models of Earth system components, such as the atmosphere, oceans, the carbon cycle or even the biosphere, with Earth observations using data assimilation. Data assimilation aims for the optimal combination of observations with simulations in a physically consistent framework, which allows to compensate for observational gaps, while observations constrain model uncertainties due to remaining structural and parametric errors. Such Twin Earths will allow to timely assess the impacts of societal decisions under the impacts of climate change over timescales from days to decades.

While broad agreement currently exists that a solid physical basis is a precondition for the interpretability and trustworthiness of Digital Twin Earths [32], AI/ML will have a big part to play in their development. Key applications will include the emulation of remaining sub-grid scale parameterizations (e.g., turbulence, cloud microphysics, radiation), data-driven development of new parameterizations and model components as well as tools to analyze and query the vast datasets created by such models.

In the longer-term future, progress in physics-constrained machine learning and explainable artificial intelligence (XAI) could lead to a paradigm shift away from computationally expensive numerical integration in short timesteps – but this will require a step-change in trustworthy AI and its broader perception.

3. Analysis of climate model data

3.1 Emulation

Current climate model intercomparisons create vast amounts of data, so big that they can no longer simply be queried to provide guidance to policy makers. The sixth phase of the Coupled Model Intercomparison Experiment (CMIP6) [33], the main international climate modelling exercise, includes several ten thousand simulation years from individual climate models; the size of its output is estimated to be around 18Pb [34]. Consequently, current climate model ensembles can only explore a very limited subset of the socio-economic scenarios [35] available to policy makers. Therefore, key policy decisions are generally based on simple global-mean climate parameters, most famously the goal to "limit global warming to well below 2, preferably to 1.5 degrees Celsius, compared to pre-industrial levels" in the Paris Agreement. Real-time information for policy makers, e.g., during United Nations Conference of the Parties (COP) negotiations, is generally provided by reduced complexity Integrated Assessment Models [e.g. 36]. While such models can be physically consistent and evaluated against complex climate models, they by construction neglect the complex regional variations of climate change and its impacts.

The availability of AI/ML techniques for the emulation of large-scale datasets provides opportunities to develop a new generation of reduced complexity climate model emulators for key climate parameters. These tools remain fully traceable to the output of large ensembles of complex climate models, e.g., CMIP6, and predict the global patterns of regional climate change.

From an AI/ML perspective, the emulation of climate model output is not dissimilar from climate modelling itself. Such an emulator can be generalized as a map E, mapping the climate state vector $\mathbf{x}$(t), through time t in dependence of transient boundary conditions $\mathbf{b}$(t).

$$\mathbf{x}(\mathrm{t}) = E(\mathbf{b}(t), \boldsymbol{\theta}_{\boldsymbol{ml}})$$

However, there exist a few conceptual differences: because E is learning the machine learning model parameters $\boldsymbol{\theta}_{\boldsymbol{ml}}$ from fitting of a precalculated training dataset $\mathbb{X}$ (the climate model output), no explicit integration in time is necessary. Further, because climate is defined as the longer-term average of individual weather patterns, it is generally possible to aggregate over larger temporal timescales, such as annual or even decadal means, although it may be desirable to also emulate the seasonal cycle as well as shorter term extremes or even full probability distributions. Finally, the primary application as a decision-making tool for choices of socio-economic scenarios means that boundary conditions, such as CO_2 emissions, can generally be aggregated to global means. However, it should be kept in mind that this may limit the applicability to e.g., regional precipitation response to short-lived climate forcers, such as aerosols, for which the emission location impacts on the response [37].

A benchmarking framework based on CMIP6 [33], ScenarioMIP [35] and DAMIP [38] simulations performed by a full complexity climate model has been developed [39], combined with a set of machine-learning based models. A range of emulators based on Gaussian Processes [24], Random Forests [40], Long short-term memory (LSTM) [41] neural networks as well as Convolutional Neural Networks (CNNs) [42] are available in a common framework suitable for emulation of Earth System components [28]. These emulators can predict annual-mean global distributions of key climate parameters, including extreme values, given a wide range of emissions pathways and allow to efficiently probe previously unexplored scenarios, a concept that could become invaluable in (computationally expensive) Digital Twin Earths.

3.2 Downscaling

The resolution of global climate models is generally too coarse to predict local climate impacts on scales of relevance e.g., for critical infrastructure, such as solar or wind power plants, or for river catchment areas, for the assessment of flood risk. Such higher resolution predictions have traditionally been made using regional climate models, which simulate regional climate in high resolution based on the boundary conditions from global climate models or using statistical downscaling, developing statistical relationships between resolved scale coarse climate variables and local or regional climate variables of interest [43, 44].

Treating a two-dimensional climate dataset as an image, the downscaling problem is closely related to the concept of super-resolution in machine learning, aiming to generate consistent high-resolution images from low-resolution input images. While this is in principle an ill-posed problem, machine learning based super-resolution methods take advantage of prior knowledge about the structure of the high-resolution images and have achieved remarkable accuracy. For climate downscaling, prior knowledge about high-resolution orography (in particular for precipitation, often triggered by flow over mountains) and surface characteristics (in particular for temperature as it affects absorption of sunlight and surface fluxes) are expected to improve the prediction. These and other physical constraints can be included explicitly or implicitly as part of high-resolution training datasets.

In supervised super resolution downscaling, a machine learning model is trained using high resolution climate datasets from observations, (e.g., precipitation data from radar networks or high-resolution models to predict high-resolution climate data from low resolution inputs, such as low-resolution climate models. Successful applications include: use of CNNs for continental-scale downscaling of temperature and precipitation [45, 46]; use of Generative Adversarial Networks (GANs) to downscale wind velocity and solar irradiance outputs from global climate models scales representative for renewable energy generation [47].

However, reliable high-resolution training datasets do not exist for all applications. For example, the assessment of climate impacts over Africa is of crucial societal importance but very limited observational networks exist that could be used as high-resolution training data. Hence, the development of unsupervised downscaling methods has a large potential. [48] treat downscaling as a domain alignment problem and develop an unsupervised model to learn a mapping between two random variables based on their shared structure with a predictive performance comparable to existing supervised methods.

Philip Stier, Department of Physics, University of Oxford, UK (philip.stier@physics.ox.ac.uk)

4. AI for Earth Observations in Climate Science

Observations have played a key role for the discovery and our understanding of key climate processes ever since the discovery of the importance of water vapor and carbon dioxide for atmospheric radiation by Eunice Foote [49]. Robust constraints from Earth Observations are of vital importance for trust in climate models and their predictions. Today, with an unprecedented amount of Earth Observation data available from spaceborne and ground-based observing systems, the role of observations in climate science is becoming increasingly limited by our ability to extract the relevant information content at scale.

For decades, the exploitation of Earth Observations for climate science has focused on pixel-by-pixel retrievals of relevant climate parameters, such as temperature profiles, atmospheric composition, and cloud properties, primarily from spectral radiance measurements. This approach has been tremendously successful but rejects a significant fraction of the information content available in Earth Observations arising from spatio-temporal structures beyond the single pixel. The advent of reliable and scalable AI/ML for feature detection and classification provides a unique opportunity to explore the full potential of Earth Observations to increase our understanding of climate process as well as for climate model constraint and evaluation.

4.1 Remote sensing retrievals

The retrieval of climate relevant parameters, represented as a state vector $\mathbf{x}$, from remote sensing observations is an inverse problem for which often no unique solutions exist. Generally, a forward radiative transfer model $\mathbf{F}(\mathbf{x})$ is used to simulate observations $\mathbf{y}$ of spectral radiances (passive instruments) or backscatter (active instruments) in the presence of measurement errors $\boldsymbol{\varepsilon}$ under assumption of prior knowledge $\mathbf{x_a}$ about the state vector. The retrieval aims to minimize the difference, generally expressed as cost function comprising of a data-fit term measuring the fit between the forward model $\mathbf{F}(\mathbf{x})$ and the observation $\mathbf{y}$, and a regularizer measuring the fit between the best estimate of the state vector $\hat{\mathbf{x}}$ and the prior $\mathbf{x_a}$. Such retrievals, often performed in the Bayesian framework of Optimal Estimation [2], have been successfully deployed in satellite remote sensing of climate parameters such as temperature and humidity profiles, cloud properties and atmospheric composition, and allow to systematically take into account multiple error sources [50]. However, significant uncertainties remain in the estimation of fundamental climate parameters; in fact, the uncertainty in our estimates of total ice water in the atmosphere is nearly as large as the total amount itself [51], as well as of associated cloud properties, such as the cloud droplet number concentration [52], severely limiting our ability to constrain current climate models. There remains significant potential to advance on current retrievals using AI, including multi-sensor fusion and the use of invertible neural networks. However, it should be noted that the retrieval problem is generally highly under-constrained. See further details in Chapter 17.

4.2 Feature detection and tracking

Driven by a vast number of commercial applications, the availability of large-scale labeled training datasets [e.g. ImageNet: 53] as well as of open source platforms for machine learning, and the development of robust and reliable machine learning models, the detection and classification of features in images has rapidly advanced and now influences most aspects of modern life.

Maybe surprisingly, the use of observable features from Earth Observations, such as cloud patterns or ocean eddies, has so far remained a niche area for the systematic evaluation and constraint of climate models. One reason is that current low-resolution climate models have, by construction, limited skill in simulating such small-scale features and are therefore primarily evaluated based on aggregate statistics. But equally, the task may have simply felt too daunting as manual detection of features in vast Earth Observation datasets is simply not practical, except for low frequency events, such as tropical cyclones.

However, the availability of reliable and scalable AI/ML based feature detection (detecting objects) and semantic segmentation (classifying each pixel) techniques has started to rapidly transform the analysis of Earth Observations for climate science, as evident from the following examples centered on the role of clouds for climate, noting that there are similar applications across many areas of climate research:

Quasi-linear tracks from well-defined pollution sources, such as ships, volcanoes, cities or industrial areas, have been extensively used to study the effect of air pollution to brighten clouds [54] – but their manual detection has remained challenging. Deep convolutional neural networks with skip connection architecture (U-Net) have been trained based on hand-labeled ship-track data from the MODIS satellite instrument [55] and applied in a case study for the stratocumulus cloud deck off the coast of California. Applying a similar technique, it has recently been possible to catalogue all ship-tracks globally since 2002 and find marked reductions under emissions control regulations introduced in 2015 and 2020 [56]. Such AI based approaches allow the global assessment and continuous monitoring of anthropogenic climate perturbations.

Cloud feedbacks are the dominant contributor of inter-model spread in climate sensitivity among current climate models [57] and often involve subtle transitions between cloud regimes and morphologies that are difficult to simulate and need to be observationally constrained:

Low stratocumulus clouds scatter sunlight back to space, have large cloud fractions and therefore cool the Earth. It is therefore important to understand the transition from stratocumulus clouds to cumulus clouds with lower cloud fraction and therefore a weaker cooling effect and how this might be affected by climate change. Such transitions can occur as a pocket of open cells (POCs) in a closed stratocumulus cloud field and the effect of air pollution on this process had been hypothesized to have a significant effect on climate. However, such prior work was limited to case studies with a small number of POC occurrences. Training a modified ResNet-152 which has been pre-trained on ImageNet, it has been possible to analyze a 13-year satellite record to identify >8000 POCs globally [58]. This allowed for the first time to conclude that the overall radiative effect of POCs on climate is small.

To understand cloud feedbacks, it is key to understand which environmental factors control cloud regimes and morphology. A labeled dataset of mesoscale cloud organization crowd-sourced from the Zooniverse platform [59] has been used [60] to train deep learning algorithms for object detection [Resnet: 61] and subsequent image segmentation [Unet: 62, from fastai library]). The application to an 11-year satellite dataset then allowed to derive heat maps for individual cloud morphologies and to link them to the presence of their physical drivers, which may change under climate change.

However, the vast scale of climate datasets often makes hand-labelling impossible, so the development and deployment of unsupervised classification techniques has great potential. An unsupervised classification scheme for mesoscale cloud organization based on the ResNet-34 CNN residual network feeding the produced embeddings into a hierarchical clustering algorithm has been developed and trained on and applied to GOES-16 satellite images [63]. It was shown the derived cloud clusters have distinct cloud structures, radiative and morphological properties with unique physical characteristics.

4.3 Extreme event and anomaly detection

While much of the attention to climate change focuses on global mean temperature rise, the most severe impacts occur through the associated shift in the likelihood and strength of extreme events, such as heatwaves, flooding, synoptic storms, tropical cyclones, or droughts.

It would therefore seem expected that reliable objective techniques for the detection, quantification and prediction of extreme events were readily available. However, this is not the case, as illustrated for the case of observed changes in heavy precipitation shown in Figure 1. In fact, for large parts of the globe, including most of Africa and South America, there exists limited data or literature to even quantify observed changes, let alone robust methodologies. The selection of extreme weather events to be attributed to climate change [64, 65] has been generally performed heuristically, with the potential to introduce selection biases (and potentially ignoring less-developed and less-observed parts of the world). The secretary-general of the United Nations, António Guterres, has tasked the World Meteorological Organization in 2022 “to ensure every person on Earth is protected by early warning systems within five years” [66].

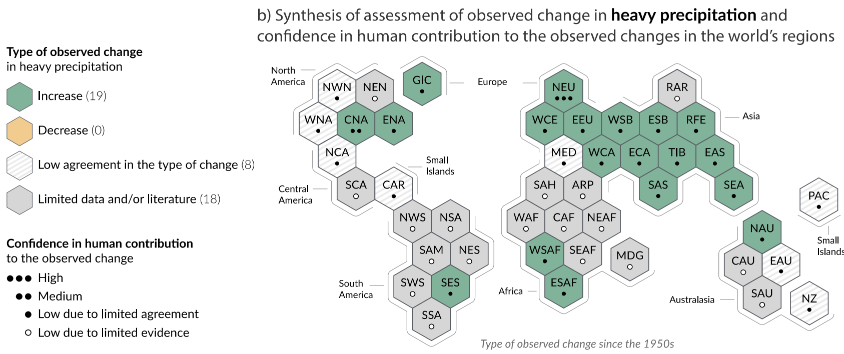

Figure 1: Intergovernmental Panel on Climate Change synthesis of current assessment of observed changes in heavy precipitation and confidence in human contribution to these changes [IPCC AR6 Summary for Policymakers: 1].

The capability of AI to detect changes in large-scale datasets provides an opportunity to transform our ability to detect, quantify and predict changes in extreme climate events.

As a first step, AI is now being used to objectively detect weather patterns associated with extreme events, such as detection of tropical cyclones, atmospheric rivers (as cause of heavy precipitation), or the detection of fronts, primarily using supervised or semi-supervised machine learning methods [67, 68]. Such methods rely on training datasets and toolkits are being developed to curate expert-labeled datasets for tropical cyclones and atmospheric rivers based on high-resolution model simulations [69].

However, such curated datasets remain limited in scope and only capture a limited subset of extreme event types. They currently do not capture key sources of extreme weather in the developing world, such as mesoscale convective systems and associated dust-storms.

The significant potential of AI to develop unsupervised techniques for the objective detection and classification of extreme events and in particular of their response to a changing climate remains underexplored.

4.4 Learning relationships between climate variables and climate processes

Traditionally, the vast majority of use cases of Earth Observations in climate science have focused on specific climate variables. However, key climate processes involve multiple variables so methods to discover and exploit the relationship between such variables could provide additional, and potentially more direct, constraints on the underlying physical processes.

For example, the linear ML method of ridge regression has been used to quantify how clouds respond to changes in the environment from Earth Observations and climate model simulations [70]. This allows the identification of key processes underlying cloud feedbacks and to provide a tighter constraint on the amplifying effect of clouds on global warming. Using a wider range of cloud controlling factors and a non-linear approach based on gradient boosting decision trees [LGBM: 71], regimes of cloud controlling environmental variables from Earth Observations have been identified [72] that provide a new constraint on the representation of clouds in global climate models.

However, it should be noted that the climate system is a coupled dynamical system so that care must be taken when evaluating statistical relationships between climate variables as correlation does not necessarily imply causation.

4.5 Causal discovery and attribution

In a large-scale dynamical system such as the Earth, inevitably many climate variables are coupled. Consequently, when observing relationships between climate variables, e.g., through regression or correlation, it is generally not clear which relationships are causal and which are driven by a common driver (confounder) [73].

In climate modelling, mechanism denial, perturbed parameter ensembles or adjoint methods are widely used to identify and quantify causal relationships related to the importance of specific processes.

Optimal fingerprinting methods have been developed for the detection of climate impacts and their attribution to specific anthropogenic forcers [4]. This typically involves comparison of model-simulated spatio-temporal patterns of the climate response to external forcings to the observational record in a reduced dimensional space, e.g., from Principal Component Analysis. Both simulated and observed response are normalized by internal variability to improve the signal to noise ratio. The observed climate change is regarded as a linear combination of signals from climate forcers, such as greenhouse gases or air pollution, and the internal climate variability and the magnitude of the response to each forcing is estimated using linear regression [cf. 74]. However, such fingerprinting methods rely on climate models accurately representing the key climate processes.

Hence, methods that allow for causal discovery and attribution solely from observations are required. As it is generally not desirable to conduct large-scale control experiments with the climate system, there exists significant potential for advanced statistical and AI causal methodologies. Recent advances in causal inference based on graph-based structural causal models, where climate variables are the nodes, edges indicate causal connections and arrows include causal directions, are now being increasingly applied in climate science [cf. 75]. New methods combining conditional independence tests (to identify potential causal links) with a causal discovery algorithm (to remove false positive links) have been developed to estimate causal networks from large-scale time-series datasets [76]. However, it is worth pointing out that many climate observations are discontinuous. For example, most satellite-based Earth observations stem from sun-synchronous polar orbiting satellites with a fixed overpass time, generally providing only one measurement per day (for retrievals relying on solar wavelengths) so with a temporal resolution that is lower than e.g., the lifetime of most individual clouds. New causal inference methods suitable for discontinuous observations based on causal forests and neural networks are being developed in the context of the effect of air pollution on clouds and climate [77].

5. How AI/ML will transform climate science

Climate change is one of the greatest challenges facing Planet Earth. Achieving the goals of the Paris Agreement requires the fastest transformation of the world's economy that has ever been attempted. It is therefore vital that the underpinning scientific evidence is robust, interpretable, and trustworthy. Physical understanding will always remain at its core.

At the same time, advances in climate research are held back by our ability to simulate climate at sufficiently small scales that resolve key processes explicitly, as well as by our ability to interpret the vast amount of data from climate models and Earth observations. It is becoming increasingly clear that AI and Machine Learning will make transformational contributions to both areas – with the challenge to deliver physically constrained, trustworthy, and explainable results.

For climate modelling, it seems inconceivable that anything else than hybrid models, combining numerical solutions of fundamental physical equations with faster and/or more accurate AI components will dominate the short- to medium-term future. And AI will dominate for heuristic model components for which no closed set of physical questions exist, such as models of biogeochemical cycles and ecosystems. The question remains primarily what fraction of the physical climate models will ultimately be replaced by AI, which in turn depends on its physical consistency and interpretability, underpinning trust. Challenges include the optimal incorporation of physical constraints; the development of climate invariant model components to deal with non-stationarity as the climate system will change between the training period and future climate; and, importantly, interpretability: as the climate system does not allow for control experiments, it is fundamental that climate predictions remain interpretable as basis for trust. In addition to speedup by fast emulation, specific opportunities to improve climate models with AI include: the emulation of parameterizations from more accurate reference models and observations; the possibility to avoid operator splitting between different climate processes by emulating multiple climate model components together; the potential to overcome locality of current parameterizations through consideration of non-local inputs to the emulation [78]; the potential to introduce memory through approaches such as LSTM, e.g., for convection parameterizations; as well as exploiting the opportunities provided by the coevolution of AI based and reduced-precision climate model components [79].

Philip Stier, Department of Physics, University of Oxford, UK (philip.stier@physics.ox.ac.uk)

Already, AI is transforming our ability to analyze the vast output of climate model simulations and this impact is only going to increase in the future. AI based emulation of climate model scenarios will provide easily accessible spatio-temporal guidance to policy makers that remains fully traceable to the underpinning complex climate model simulations. Future opportunities include the emulation of the full probability distribution of climate variables to flexibly assess extreme events as well as the inclusion of regional emission variations in the emulators to provide guidance on their regional and global impacts. Combination of emulation of climate model output with AI based downscaling techniques, as well as further advances in unsupervised downscaling techniques could ultimately provide accessible high-resolution data of climate change and impacts for local and regional decision making.

While much of the current attention on AI in climate science focuses on modelling, AI will entirely transform our ability to interpret Earth Observations to exploit the full information content available. This is not without challenges: climate observations are heterogeneous, discontinuous, on non-Euclidian spaces, and big – with each snapshot being vastly bigger than typical image resolutions used in standard machine learning applications and total data volumes increasing by many Pb each year. The general presence of a multitude of confounding factors makes causal attribution from observations alone challenging. However, future opportunities are plentiful: AI will make it possible to detect, track and label climate phenomena all the way to individual clouds or ocean eddies. Combination with causal discovery tools will allow identification and quantification of key drivers of climate change. It will facilitate the objective detection of climate impacts and changing extreme events, e.g., flooding, wildfires, droughts, and heatwaves, across the full observational record and its attribution to anthropogenic activities, for example replacing prevailing linear methods with non-linear neural networks in detection and attribution applications.

Ultimately, the challenge will be to optimally combine physical models and Earth observations with AI to accelerate climate science.

For the foreseeable future, the O(1km) resolution of even storm-resolving Digital Twin Earths currently in planning will not be sufficient to resolve key processes, such as low-cloud feedbacks that dominate current climate model uncertainty so they will need to be informed, possibly by AI, by even higher-resolution Large Eddy Simulation (LES) or Direct Numerical Simulation (DNS) models. Likewise, it will not be possible to routinely run storm-resolving Digital Twin Earths for a large number of scenarios over the wide parameter space explored in current climate model intercomparisons so they may need to be complemented with low-resolution twins for long-term climate scenario simulations. Biases introduced by the low-resolution twin parameterizations could be systematically corrected for using ML learned correction terms derived from the high-resolution twins, as has been recently successfully demonstrated [80]. And data from the high-resolution twins could be used as training data for supervised super-resolution downscaling of low-resolution twin to predict climate impacts and in particular extreme events on scales relevant for decision making. Such a vision of twin Digital Twin Earths (let's call it DTE^2) is presented in Figure 2 – and would not be possible without AI for climate science.

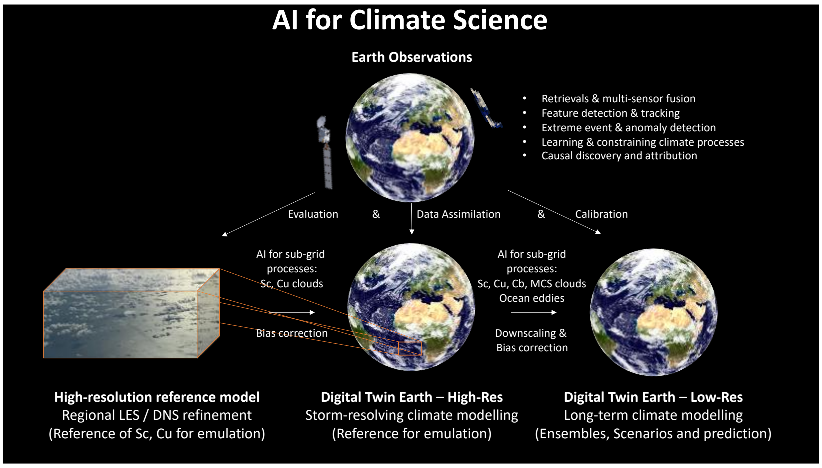

Figure 2: Applications of AI for Climate Science in the context of Digital Twin Earths.

Philip Stier, Department of Physics, University of Oxford, UK (philip.stier@physics.ox.ac.uk)

References

[1] V. Masson-Delmotte, P. Zhai, A. Pirani *et al.*, "IPCC, 2021: Summary for Policymakers," in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.*, vol. In Press. Cambridge, UK and New York, USA: Cambridge University Press, 2021, pp. 3-32.

[2] C. D. Rodgers, "Retrieval of Atmospheric-Temperature and Composition from Remote Measurements of Thermal-Radiation," *Rev Geophys,* vol. 14, no. 4, pp. 609-624, 1976, doi: 10.1029/RG014i004p00609.

[3] C. Jakob and G. Tselioudis, "Objective identification of cloud regimes in the Tropical Western Pacific," *Geophysical Research Letters,* vol. 30, no. 21, Nov 1 2003, doi: 10.1029/2003GL018367.

[4] K. Hasselmann, "Optimal Fingerprints for the Detection of Time-Dependent Climate-Change," *J Climate,* vol. 6, no. 10, pp. 1957-1971, Oct 1993, doi: 10.1175/1520-0442(1993)006<1957:OFFTDO>2.0.CO;2.

[5] R. Courant, K. Friedrichs, and H. Lewy, "Über die partiellen Differenzengleichungen der mathematischen Physik," *Mathematische Annalen,* vol. 100, pp. 32-74, 1928, doi: 10.1007/BF01448839.

[6] V. M. Krasnopolsky, M. S. Fox-Rabinovitz, and D. V. Chalikov, "New approach to calculation of atmospheric model physics: Accurate and fast neural network emulation of longwave radiation in a climate model," *Mon Weather Rev,* vol. 133, no. 5, pp. 1370-1383, May 2005, doi: Doi 10.1175/Mwr2923.1.

[7] V. M. Krasnopolsky, M. S. Fox-Rabinovitz, Y. T. Hou *et al.*, "Accurate and Fast Neural Network Emulations of Model Radiation for the NCEP Coupled Climate Forecast System: Climate Simulations and Seasonal Predictions," *Mon Weather Rev,* vol. 138, no. 5, pp. 1822-1842, May 2010, doi: 10.1175/2009mwr3149.1.

[8] P. Harder, D. Watson-Parris, P. Stier *et al.*, "Physics-Informed Learning of Aerosol Microphysics," *Environmental Data Science,* vol. 1, E20, 2022, doi: 10.1017/eds.2022.22.

[9] S. Rasp, M. S. Pritchard, and P. Gentine, "Deep learning to represent subgrid processes in climate models," *P Natl Acad Sci*

USA, vol. 115, no. 39, pp. 9684-9689, Sep 25 2018, doi: 10.1073/pnas.1810286115.

[10] J. Yuval, P. A. O'Gorman, and C. N. Hill, "Use of Neural Networks for Stable, Accurate and Physically Consistent Parameterization of Subgrid Atmospheric Processes With Good Performance at Reduced Precision," *Geophysical Research Letters,* vol. 48, no. 6, Mar 28 2021, doi: 10.1029/2020GL091363.

[11] M. Chantry, S. Hatfield, P. Dueben *et al.*, "Machine Learning Emulation of Gravity Wave Drag in Numerical Weather Forecasting," *J Adv Model Earth Sy,* vol. 13, no. 7, Jul 2021. [Online]. Available: <Go to ISI>://WOS:000682994000010.

[12] L. Zanna and T. Bolton, "Data-Driven Equation Discovery of Ocean Mesoscale Closures," *Geophysical Research Letters,* vol. 47, no. 17, Sep 16 2020. [Online]. Available: <Go to ISI>://WOS:000572406100081.

[13] H. Sundqvist, E. Berge, and J. E. Kristjansson, "Condensation and Cloud Parameterization Studies with a Mesoscale Numerical Weather Prediction Model," *Mon Weather Rev,* vol. 117, no. 8, pp. 1641-1657, Aug 1989, doi: Doi 10.1175/1520-0493(1989)117<1641:Cacpsw>2.0.Co;2.

[14] T. Schneider, S. W. Lan, A. Stuart *et al.*, "Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations," *Geophysical Research Letters,* vol. 44, no. 24, pp. 12396-12417, Dec 28 2017, doi: 10.1002/2017gl076101.

[15] B. Settles, "Active Learning Literature Survey," in "Computer Sciences Technical Report," University of Wisconsin–Madison, Wisconsin–Madison, 2009, vol. 1648.

[16] T. Beucler, M. Pritchard, J. Yuval *et al.*, "Climate-Invariant Machine Learning," 2021, doi: 10.48550/arXiv.2112.08440.

[17] I. Tan, T. Storelvmo, and M. D. Zelinka, "Observational constraints on mixed-phase clouds imply higher climate sensitivity," *Science,* vol. 352, no. 6282, pp. 224-227, Apr 8 2016. [Online]. Available: <Go to ISI>://WOS:000373681600046.

[18] T. M. Lenton, H. Held, E. Kriegler *et al.*, "Tipping elements in the Earth's climate system," *P Natl Acad Sci USA,* vol. 105, no. 6, pp. 1786-1793, Feb 12 2008. [Online]. Available: <Go to ISI>://WOS:000253261900005.

[19] D. A. Stainforth, T. Aina, C. Christensen *et al.*, "Uncertainty in predictions of the climate response to rising levels of greenhouse gases," *Nature,* vol. 433, no. 7024, pp. 403-406, Jan 27 2005, doi: 10.1038/nature03301.

[20] T. Mauritsen, B. Stevens, E. Roeckner *et al.*, "Tuning the climate of a global model," *J Adv Model Earth Sy*, vol. 4, Aug 7 2012, doi: 10.1029/2012ms000154.

[21] M. Allen, "Do-it-yourself climate prediction," *Nature*, vol. 401, no. 6754, pp. 642-642, Oct 14 1999, doi: Doi 10.1038/44266.

[22] D. M. H. Sexton, C. F. McSweeney, J. W. Rostron *et al.*, "A perturbed parameter ensemble of HadGEM3-GC3.05 coupled model projections: part 1: selecting the parameter combinations," *Clim Dynam*, vol. 56, no. 11-12, pp. 3395-3436, Jun 2021, doi: 10.1007/s00382-021-05709-9.

[23] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *J R Stat Soc B*, vol. 63, pp. 425-450, 2001, doi: Doi 10.1111/1467-9868.00294.

[24] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning* (Adaptive computation and machine learning). Cambridge, Mass.: MIT Press, 2006, pp. xviii, 248 p.

[25] M. D. Mckay, R. J. Beckman, and W. J. Conover, "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, vol. 21, no. 2, pp. 239-245, 1979, doi: Doi 10.2307/1268522.

[26] I. Vernon, M. Goldstein, and R. G. Bower, "Galaxy Formation: a Bayesian Uncertainty Analysis Rejoinder," *Bayesian Anal*, vol. 5, no. 4, pp. 697-708, 2010, doi: 10.1214/10-Ba524rej.

[27] L. A. Lee, K. J. Pringle, C. L. Reddington *et al.*, "The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei," *Atmos Chem Phys*, vol. 13, no. 18, pp. 9375-9377, 2013, doi: 10.5194/acp-13-9375-2013.

[28] D. Watson-Parris, A. Williams, L. Deaconu *et al.*, "Model calibration using ESEm v1.1.0-an open, scalable Earth system emulator," *Geosci Model Dev*, vol. 14, no. 12, pp. 7659-7672, Dec 20 2021. [Online]. Available: <Go to ISI>://WOS:000731831100001.

[29] B. Stevens, M. Satoh, L. Auger *et al.*, "DYAMOND: the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains," *Prog Earth Planet Sc*, vol. 6, no. 1, Dec 2019, doi: 10.1186/s40645-019-0304-z.

[30] B. Stevens. "Faster Faster — machine learning our way to a new generation of storm-resolving Earth system models." United Nations ITU. https://aiforgood.itu.int/event/ai-and-climate-science-bjorn-stevens/ (accessed 1 December 2022, 2021).

[31] P. Bauer, P. D. Dueben, T. Hoefler *et al.*, "The digital revolution of Earth-system science. ," *Nature Computational Science*, vol. 1, pp. 104-113, 2021, doi: 10.1038/s43588-021-00023-0.
[32] M. Reichstein, G. Camps-Valls, B. Stevens *et al.*, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195-204, Feb 14 2019, doi: 10.1038/s41586-019-0912-1.
[33] V. Eyring, S. Bony, G. A. Meehl *et al.*, "Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization," *Geosci Model Dev*, vol. 9, no. 5, pp. 1937-1958, 2016, doi: 10.5194/gmd-9-1937-2016.
[34] V. Balaji, K. E. Taylor, M. Juckes *et al.*, "Requirements for a global data infrastructure in support of CMIP6," *Geosci Model Dev*, vol. 11, no. 9, pp. 3659-3680, Sep 11 2018, doi: 10.5194/gmd-11-3659-2018.
[35] B. C. O'Neill, C. Tebaldi, D. P. van Vuuren *et al.*, "The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6," *Geosci Model Dev*, vol. 9, no. 9, pp. 3461-3482, Sep 28 2016, doi: 10.5194/gmd-9-3461-2016.
[36] R. J. Millar, Z. R. Nicholls, P. Friedlingstein *et al.*, "A modified impulse-response representation of the global near-surface air temperature and atmospheric concentration response to carbon dioxide emissions," *Atmos Chem Phys*, vol. 17, no. 11, pp. 7213-7228, Jun 16 2017. [Online]. Available: <Go to ISI>://WOS:000403779200004.
[37] P. Stier, S. C. v. d. Heever, M. Christensen *et al.*, "Multifaceted aerosol effects on precipitation " *Nat Geosci*, vol. in review, 2022.
[38] N. P. Gillett, H. Shiogama, B. Funke *et al.*, "The Detection and Attribution Model Intercomparison Project (DAMIP v1.0) contribution to CMIP6," *Geosci Model Dev*, vol. 9, no. 10, pp. 3685-3697, Oct 18 2016, doi: 10.5194/gmd-9-3685-2016.
[39] D. Watson-Parris, Y. Rao, D. Olivié *et al.*, "ClimateBench: A benchmark dataset for data-driven climate projections," *J Adv Model Earth Sy*, vol. 14, 2022, doi: 10.1029/2021MS002954.
[40] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition*, 1995: IEEE, pp. 278-282.
[41] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
[42] Y. L. Cun, B. Boser, J. S. Denker *et al.*, "Handwritten digit recognition with a back-propagation network," in *Advances in*

neural information processing systems 2, vol. 2, D. Touretzky Ed. Denver, USA: Morgan Kaufmann Publishers Inc., 1990, pp. 396–404.

[43] T. M. L. Wigley, P. D. Jones, K. R. Briffa *et al.*, "Obtaining Sub-Grid-Scale Information from Coarse-Resolution General-Circulation Model Output," *J Geophys Res-Atmos,* vol. 95, no. D2, pp. 1943-1953, Feb 20 1990, doi: DOI 10.1029/JD095iD02p01943.

[44] G. Burger, T. Q. Murdock, A. T. Werner *et al.*, "Downscaling Extremes-An Intercomparison of Multiple Statistical Methods for Present Climate," *J Climate,* vol. 25, no. 12, pp. 4366-4388, Jun 15 2012, doi: 10.1175/Jcli-D-11-00408.1.

[45] J. Bano-Medina, R. Manzanas, and J. M. Gutierrez, "Configuration and intercomparison of deep learning neural models for statistical downscaling," *Geosci Model Dev,* vol. 13, no. 4, pp. 2109-2124, Apr 28 2020, doi: 10.5194/gmd-13-2109-2020.

[46] T. Vandal, E. Kodra, S. Ganguly *et al.*, "DeepSD: Generating High Resolution Climate Change Projections through Single Image Super-Resolution," p. arXiv:1703.03126, 2017, doi: 10.48550/arXiv.1703.03126.

[47] K. Stengel, A. Glaws, D. Hettinger *et al.*, "Adversarial super-resolution of climatological wind and solar data," *P Natl Acad Sci USA,* vol. 117, no. 29, pp. 16805-16815, Jul 21 2020, doi: 10.1073/pnas.1918964117.

[48] B. Groenke, L. Madaus, and C. Monteleoni, "ClimAlign: Unsupervised statistical downscaling of climate variables via normalizing flows," 2020, doi: 10.48550/arXiv.2008.04679.

[49] E. Foote, "Circumstances Affecting the Heat of the Sun's Rays," *The American Journal of Science and Arts,* vol. 22, no. 66, pp. 383-384, 1856.

[50] C. D. Rodgers, *Inverse Methods for Atmospheric Sounding* (Series on Atmospheric, Oceanic and Planetary Physics, no. 2). Singapore: World Scientific, 2000.

[51] O. Boucher, D. Randall, P. Artaxo *et al.*, "Clouds and Aerosols," in *Working Group I Contribution to the Intergovernmental Panel on Climate Change Firth Assessment Report*, S. Fuzzi, J. Penner, V. Ramaswamy, and C. Stubenrauch Eds., (Climate Change 2013: The Physical Science Basis. Geneva: Cambridge University Press, 2013.

[52] D. P. Grosvenor, O. Sourdeval, P. Zuidema *et al.*, "Remote Sensing of Droplet Number Concentration in Warm Clouds: A Review of the Current State of Knowledge and Perspectives," *Rev Geophys*, vol. 56, no. 2, pp. 409-453, Jun 2018, doi: 10.1029/2017rg000593.

[53] J. Deng, W. Dong, R. Socher *et al.*, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 20-25 June 2009 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.

[54] M. Christensen, A. Gettelman, J. Cermak *et al.*, "Opportunistic Experiments to Constrain Aerosol Effective Radiative Forcing," *Atmos. Chem. Phys. Discuss.*, vol. 2021, pp. 1-60, 2021, doi: 10.5194/acp-2021-559.

[55] T. L. Yuan, C. Wang, H. Song *et al.*, "Automatically Finding Ship Tracks to Enable Large-Scale Analysis of Aerosol - Cloud Interactions," *Geophysical Research Letters*, vol. 46, no. 13, pp. 7726-7733, Jul 16 2019. [Online]. Available: <Go to ISI>://WOS:000476960100074.

[56] D. Watson-Parris, M. W. Christensen, A. Laurenson *et al.*, "Shipping regulations lead to large reduction in cloud perturbations," *Proceedings of the National Academy of Sciences*, vol. 119, no. 41, p. e2206885119, 2022, doi: doi:10.1073/pnas.2206885119.

[57] M. D. Zelinka, T. A. Myers, D. T. Mccoy *et al.*, "Causes of Higher Climate Sensitivity in CMIP6 Models," *Geophysical Research Letters*, vol. 47, no. 1, Jan 16 2020, doi: 10.1029/2019GL085782.

[58] D. Watson-Parris, S. A. Sutherland, M. C. Christensen *et al.*, "A Large-Scale Analysis of Pockets of Open Cells and Their Radiative Impact," *Geophysical Research Letters*, vol. 48, no. 6, Mar 28 2021, doi: 10.1029/2020GL092213.

[59] S. Rasp, H. Schulz, and B. Stevens. "Sugar, flower, fish or gravel." https://www.zooniverse.org/projects/raspstephan/sugar-flower-fish-or-gravel (accessed 1 December, 2022).

[60] S. Rasp, H. Schulz, S. Bony *et al.*, "Combining Crowdsourcing and Deep Learning to Explore the Mesoscale Organization of Shallow Convection," *B Am Meteorol Soc*, vol. 101, no. 11, pp. E1980-E1995, Nov 2020. [Online]. Available: <Go to ISI>://WOS:000598066100007.

[61] T.-Y. Lin, P. Goyal, R. Girshick *et al.*, "Focal Loss for Dense Object Detection," 2017, doi: 10.48550/arXiv.1708.02002.

[62] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015, doi: 10.48550/arXiv.1505.04597.

[63] L. Denby, "Discovering the Importance of Mesoscale Cloud Organization Through Unsupervised Classification," *Geophysical Research Letters*, vol. 47, no. 1, Jan 16 2020. [Online]. Available: <Go to ISI>://WOS:000513983400007.

[64] M. Allen, "Liability for climate change," *Nature*, vol. 421, no. 6926, pp. 891-892, Feb 27 2003. [Online]. Available: <Go to ISI>://WOS:000181186900017.

[65] "World Weather Attribution." https://www.worldweatherattribution.org (accessed 7 April, 2022).

[66] World Meteorological Organiszation "Early Warning systems must protect everyone within five years," ed. Geneva: World Meteorological Organiszation, 2022.

[67] E. Racah, C. Beckham, T. Maharaj *et al.*, "ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events," 2016, doi: 10.48550/arXiv.1612.02095.

[68] Y. Liu, E. Racah, Prabhat *et al.*, "Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets," 2016, doi: 10.48550/arXiv.1605.01156.

[69] Prabhat, K. Kashinath, M. Mudigonda *et al.*, "ClimateNet: an expert-labeled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather," *Geosci Model Dev*, vol. 14, no. 1, pp. 107-124, 2021, doi: 10.5194/gmd-14-107-2021.

[70] P. Ceppi and P. Nowack, "Observational evidence that cloud feedback amplifies global warming," *P Natl Acad Sci USA*, vol. 118, no. 30, Jul 27 2021, doi: 10.1073/pnas.2026290118.

[71] G. Ke, Q. Meng, T. Finley *et al.*, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

[72] A. Douglas and P. Stier, "Establishing Regimes of Cloud Controlling Factors using Machine Learning," *Journal of Geophysical Research*, vol. submitted, 2022.

[73] H. Reichenbach, *The direction of time*. Berkeley: University of California Press (in eng), 1956.

[74] G. Hegerl and F. Zwiers, "Use of models in detection and attribution of climate change," *Wires Clim Change,* vol. 2, no. 4, pp. 570-591, Jul-Aug 2011, doi: 10.1002/wcc.121.
[75] J. Runge, S. Bathiany, E. Bollt *et al.*, "Inferring causation from time series in Earth system sciences," *Nat Commun,* vol. 10, Jun 14 2019.
[76] J. Runge, P. Nowack, M. Kretschmer *et al.*, "Detecting and quantifying causal associations in large nonlinear time series datasets," *Science Advances,* vol. 5, no. 11, Nov 2019.
[77] A. Jesson, P. Manshausen, A. Douglas *et al.*, "Using Non-Linear Causal Models to Study Aerosol-Cloud Interactions in the Southeast Pacific," 2021, doi: 10.48550/arXiv.2110.15084.
[78] P. Wang, J. Yuval, and P. A. O'Gorman, "Non-Local Parameterization of Atmospheric Subgrid Processes With Neural Networks," *J Adv Model Earth Sy,* vol. 14, no. 10, 2022, doi: 10.1029/2022MS002984.
[79] P. D. Dueben, J. Joven, A. Lingamneni *et al.*, "On the use of inexact, pruned hardware in atmospheric modelling," *Philos T R Soc A,* vol. 372, no. 2018, Jun 28 2014, doi: 10.1098/rsta.2013.0276.
[80] C. S. Bretherton, B. Henn, A. Kwa *et al.*, "Correcting Coarse-Grid Weather and Climate Models by Machine Learning From Global Storm-Resolving Simulations," *J Adv Model Earth Sy,* vol. 14, no. 2, Feb 2022, doi: 10.1029/2021MS002794.

Philip Stier, Department of Physics, University of Oxford, UK (philip.stier@physics.ox.ac.uk) 33

Index keywords

- climate
- machine learning
- Earth Observations
- clustering
- causal attribution
- climate models
- Earth System Models
- clouds
- atmosphere
- ocean
- weather
- supercomputing
- emulation
- parameterization
- convection
- stationarity / non-stationarity
- feedbacks
- calibration
- perturbed parameter ensemble
- Bayes
- Gaussian Processes
- Latin hypercubes
- Twin Earths
- Digital Twin Earths
- explainable artificial intelligence / XAI
- Coupled Model Intercomparison Project / CMIP
- downscaling
- remote sensing
- radiative transfer
- feature detection
- tracking
- ship tracks
- extreme event
- tropical cyclones

- causal discovery
- air pollution
-