https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

Opinion: Optimizing climate models with process-knowledge, resolution, and AI

Tapio Schneider[1,2], L. Ruby Leung[3], and Robert C. J. Wills[4]

[1]California Institute of Technology, Pasadena, CA, USA
[2]Google Research, Mountain View, CA, USA
[3]Pacific Northwest National Laboratory, Richland, Washington, USA
[4]ETH Zurich, Zurich, Switzerland

Correspondence: Tapio Schneider (tapio@caltech.edu)

Abstract. Accelerating progress in climate modeling is urgent for proactive and effective climate change adaptation. The central challenge lies in accurately representing processes that are small in scale yet are climatically important, such as turbulence and cloud formation. These processes are not explicitly resolvable, necessitating the use of parameterizations. We propose a balanced approach that leverages the strengths of traditional process-based parameterizations and contemporary AI-based data-driven methods to model subgrid-scale processes. This strategy focuses on employing AI to derive data-driven closure functions from both observational and simulated data, integrated within parameterizations that encode system knowledge and conservation laws. Increasing resolution to resolve a larger fraction of small-scale processes can aid progress toward improved and interpretable climate predictions outside the observed climate distributions, but it must still allow the generation of large ensembles for model calibration and the broad exploration of possible climate outcomes—currently $O(10\,\mathrm{km})$ horizontal resolutions are feasible. By synergizing decades of scientific development with advanced AI techniques, this approach aims to significantly boost the accuracy, interpretability, and trustworthiness of climate predictions.

1 Introduction

Climate models serve two distinct purposes. First, they encode our collective knowledge about the climate system. They instantiate theories and provide a quantitative account of climate processes—the complex interplay of causes and effects that governs how the climate system operates. In this role, they belong to the realm of *episteme*, or explanatory science (Russo, 2000; Parry, 2021). Second, climate models function as practical tools that allow us to calculate how the climate system might behave under different circumstances that have not been directly observed. In this role, they fall under the realm of *techne*, or goal-oriented applied science (Russo, 2000; Parry, 2021). The requirements for climate models differ depending on their role as *episteme* or *techne*. As encodings of our understanding (*episteme*), climate models should strive for explainability and simplicity, even if it means sacrificing a certain level of accuracy. Understanding of the climate system at different levels of description emerges through a hierarchy of models, ranging from simpler ones such as one-dimensional radiative-convective equilibrium models to more complex ones such as atmospheric general circulation models with simplified parameterizations of

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

subgrid-scale processes (Held, 2005; Jeevanjee et al., 2017). On the other hand, as calculation tools (*techne*), climate models should aim to simulate the climate system as accurately as possible under unobserved circumstances.

Over the past six decades, climate modeling has operated under the tacit assumption that these two roles of climate models converge, implying that the most complex models reflecting our understanding of the system are also the most accurate tools for predicting its behavior in unobserved conditions. This is a desirable goal, but it may not always be attainable in systems as complex as the climate system.

In this essay, we specifically focus on climate models as *techne*, emphasizing their role as tools for accurately calculating the behavior of the climate system in unobserved circumstances. The goal of such calculations is to obtain statistics about the climate system, including average temperatures at specific locations and seasons, the probability that daily precipitation in a given region exceeds some threshold, or the covariance between temperature and humidity, which can lead to potentially dangerous humid heat extremes. These calculations correspond to what Lorenz (1975) defined as predictions of the second kind, where future climate statistics are estimated given evolving boundary conditions, such as human-induced greenhouse gas emissions. This contrasts with predictions of the first kind, which focus on forecasting the future state of a system given its initial conditions $\boldsymbol{\zeta}_0$, as seen in weather forecasting. Consequently, climate models as *techne* should aim to minimize a loss function of the form

$$\mathcal{L} = \|\langle \boldsymbol{y}(t) \rangle - \langle \boldsymbol{\mathcal{H}} \circ \boldsymbol{\mathcal{G}}(t; \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\nu}; \boldsymbol{\zeta}_0) \rangle\|_{\Gamma}^2. \tag{1}$$

Here, the angle brackets $\langle \cdot \rangle$ indicate an appropriate time averaging, such as a seasonal average over multiple years. The vector $\boldsymbol{y}(t)$ represents time-varying observables of the climate system, including those whose time average $\langle \boldsymbol{y}(t) \rangle$ gives rise to higher-order statistics such as the frequency of exceeding a daily precipitation threshold in a specific region. The climate model, denoted as $\boldsymbol{\mathcal{G}}(t; \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\nu}; \boldsymbol{\zeta}_0)$, is a mapping that takes parameter vectors $(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ and an initial condition vector $\boldsymbol{\zeta}_0$ (usually important only for slowly varying components of the climate system, such as oceans and ice sheets) to time-varying simulated climate states $\boldsymbol{\zeta}(t) = \boldsymbol{\mathcal{G}}(t; \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\nu}; \boldsymbol{\zeta}_0)$. The observation operator $\boldsymbol{\mathcal{H}}$ maps simulated climate states $\boldsymbol{\zeta}(t)$ to the desired observables $\boldsymbol{y}(t)$. Lastly, $\|\cdot\|_{\Gamma} = \|\boldsymbol{\Gamma}^{-1/2} \cdot\|_2$ represents a weighted Euclidean norm, or Mahalanobis distance. The weight is determined by the covariance matrix $\boldsymbol{\Gamma}$, which reflects model and observational errors and noise due to fluctuations from internal variability in the finite-time average $\langle \cdot \rangle$. The weighted Euclidean norm is chosen because the climate statistics are aggregated over time, meaning that, due to the central limit theorem, it is reasonable to assume that these statistics exhibit Gaussian fluctuations (Iglesias et al., 2013; Schneider et al., 2017a; Dunbar et al., 2021). However, the specific choice of norm in the loss function is not crucial for the following discussion. The essence is that the loss function penalizes mismatches between simulated and observed climate *statistics*, with less noisy statistics receiving greater weight.

To achieve accurate simulations of climate statistics, the objective is to minimize the loss function (1) with respect to the parameters $(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ for unobserved climate statistics $\langle \boldsymbol{y} \rangle$. Importantly, these climate statistics may not entirely fall within the distribution of observed climate statistics, particularly in the context of global warming projections. Therefore, the ability of a model to generalize beyond the distribution of the observed data becomes essential. Merely minimizing the loss over observed climate statistics or even driving the loss to zero in an attempt to imitate observations and pass a "climate Turing tes" (Palmer,

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

2016) is not sufficient. Instead, fundamental science and data science tools, such as cross-validation and Bayesian tools, need to be brought to bear to plausibly minimize the loss for unobserved statistics.

In the loss function, we distinguish three types of parameters:

1. The parameters $\boldsymbol{\theta}$ appear in process-based models of subgrid-scale processes, such as entrainment and detrainment rates in parameterizations of convection. These parameters are directly interpretable and theoretically measurable, although their practical measurement can be challenging.

2. The parameters $\boldsymbol{\lambda}$ represent the characteristics of the climate model's resolution, such as the horizontal and vertical resolution in atmosphere and ocean models.

3. The parameters $\boldsymbol{\nu}$ pertain to AI-based data-driven models that capture subgrid-scale processes or correct for structural model errors, either within process-based models of subgrid-scale processes or holistically for an entire climate model (Kennedy and O'Hagan, 2001; Levine and Stuart, 2022; Bretherton et al., 2022). These parameters are neither easily interpretable nor directly measurable but are learned from data.

This distinction among the parameters is useful as it reflects three different dimensions along which climate models can be optimized. First, optimization can be achieved by calibrating parameters and improving the structure of process-based models that represent subgrid-scale processes such as turbulence, convection, and clouds. These processes have long been identified as the dominant sources of biases and uncertainties in climate simulations (Cess et al., 1989; Bony and Dufresne, 2005; Stephens, 2005; Vial et al., 2013; Schneider et al., 2017b; Zelinka et al., 2020). Second, optimization can be accomplished by increasing the resolution of the models, which reduces the need for parameterization (Bauer et al., 2021; Slingo et al., 2022). Finally, optimization can be pursued by integrating AI-based data-driven models. These models have the potential to replace (Gentine et al., 2018; O'Gorman and Dwyer, 2018; Yuval and O'Gorman, 2020) or complement (Schneider et al., 2017a; Lopez-Gomez et al., 2022) process-based models for subgrid-scale processes. Additionally, they can serve as comprehensive error-corrections for climate models (Bretherton et al., 2022).

In the past two decades, efforts to optimize climate models have often focused on individual dimensions in isolation. For instance, Climate Process Teams, initiated under the U.S. Climate Variability and Predictability Program, have concentrated on enhancing process-based models by incorporating knowledge from observational and process-oriented studies into climate modeling (Subramanian et al., 2016). The resolution of atmosphere and ocean models has gradually increased, albeit at a pace slower than the advances in computer performance would have allowed (Schneider et al., 2017b). More recently, there have been calls to prioritize resolution increase, aiming to achieve kilometer-scale resolutions in the horizontal, with the expectation that this would alleviate the need for subgrid-scale process parameterizations (Bauer et al., 2021; Slingo et al., 2022). Moreover, there is a rapidly growing interest to advance climate modeling by utilizing AI tools, broadly understood to include tools such as Bayesian learning, deep learning, and generative AI (e.g., Schneider et al., 2017a; Reichstein et al., 2019; Chantry et al., 2021; Watson-Parris, 2021; Balaji et al., 2022; Irrgang et al., 2022; Schneider et al., 2023).

Beginning with a review of recent advances in the goodness-of-fit between climate simulations and observed records, here we will explore the potential benefits and challenges associated with optimizing each of the three dimensions mentioned earlier.

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

Our analysis will highlight the importance of adopting a balanced approach that encompasses progress along each dimension, as this is likely to yield the most robust and accurate climate models.

2 Evolution of climate models

The climate statistics $\langle \boldsymbol{y} \rangle$ used in the loss function (1) can vary depending on the specific application. For example, a national climate model may prioritize minimizing the loss within a particular country. However, there are several climate statistics that are generally considered important and should be included in any comprehensive loss function. Two such examples are the top-of-atmosphere (TOA) radiative energy fluxes and surface precipitation rates.

The inclusion of TOA radiative energy fluxes is crucial because accurately simulating these fluxes is a prerequisite for accurately simulating a wide range of climate statistics. After all, radiative energy is the primary energy that drives the climate system; changes in radiative energy fluxes caused by changes in greenhouse gas concentrations drive global climate change. Errors in radiative energy fluxes affect many aspects of a simulated climate, from wind to precipitation distributions. The balance of TOA radiative energy fluxes must also be closed to machine precision. This is necessary to achieve a steady climate, without drift, in centennial to millennial integrations with over 10^7 discrete timesteps—in what John von Neumann called the "infinite forecast" (Edwards, 2010), in contrast to the short-term integrations needed for weather forecasting, which have less stringent conservation requirements. Similarly, precipitation rates are of significant importance as they directly impact human activities. Achieving accurate simulations of precipitation rates relies on accurately simulating numerous subgrid-scale processes within the climate system. Therefore, precipitation is an emergent property that serves as a holistic metric to assess the goodness-of-fit of a climate model.

Figure 1 assesses the evolution of climate models over the past two decades in simulating the observed climatology of TOA radiative energy fluxes and precipitation rates, setting aside temporarily that the loss minimization should occur for unobserved records. The figure displays the median root mean square (rms) error between model seasonal climatologies and observations, with all data interpolated to a common 2.5° latitude-longitude grid.[1] The plot includes three generations of climate models from the Coupled Model Intercomparison Project (CMIP) as well as recent higher-resolution simulations. It is evident that, over time, there has been a gradual improvement in the fidelity of models in simulating TOA radiative energy fluxes and precipitation. For example, in CMIP6 (late-2010s), the median rms error relative to CMIP3 (mid-2000s) is reduced by 15% for precipitation, 31% for TOA outgoing longwave flux, and 30% for TOA reflected shortwave flux (all values indicate seasonal-mean improvements). Individual modeling centers, such as the National Center for Atmospheric Research (NCAR), have surpassed this median rate of improvement, with rms error reductions of 30% for precipitation, 49% for TOA outgoing longwave flux, and 36% for TOA reflected shortwave flux in the progression from CCSM3 to CESM2. These improvements primarily stem from advances in process-based parameterizations and model tuning (e.g., Danabasoglu et al., 2020). The average resolution has also increased

[1] That is, what is displayed in Fig. 1 are unweighted errors, in contrast to the loss function (1), which downweights mismatches between simulations and observations for variables that have high error variance, e.g., because of internal variability in finite-time averages.

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

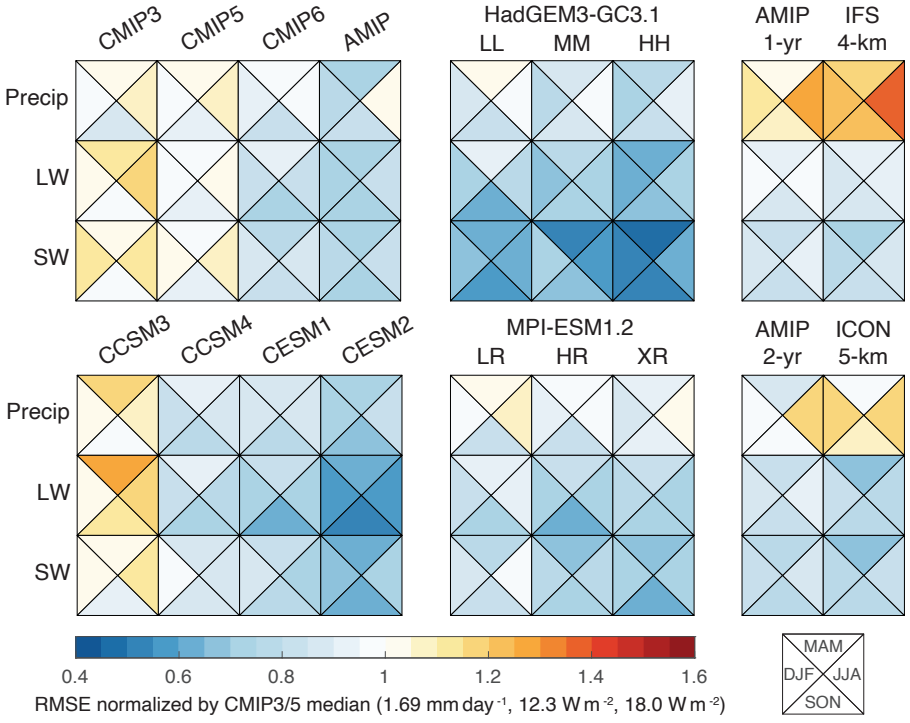

Figure 1. Normalized rms error (RMSE) in the seasonal climatology of precipitation, top-of-atmosphere (TOA) longwave (LW) radiation, and TOA shortwave (SW) radiation for different models and model intercomparison projects. The rms errors are relative to climatologies from GCPC (Adler et al., 2018) and CERES-EBAF (Loeb et al., 2009) datasets over the period 2001–2020. CMIP and AMIP rms errors represent median values over the included models. Climatologies are computed as follow: for CMIP3 over 2001–2020 for the B1 scenario; for CMIP5/6 over 2001–2020 for a combination of the historical and RCP4.5/SSP2.45 scenarios; for AMIP over 1995–2014 for models contributing to CMIP6; for HadGEM3-GC31 and MPI-ESM1-2 over 1995–2014 for the historical simulations; and for the kilometer-scale resolution simulations over 1 simulated year for IFS and over 2 simulated years for ICON (nextGEMS cycle 2), shown together with AMIP averaged over the same simulation length. The rms error is normalized by the median across CMIP3 and CMIP5 models for each field and across all seasons, with normalization constants shown below the colorbar. HadGEM3-GC3.1 and MPI-ESM1.2 from HighResMIP (Haarsma et al., 2016) are sorted in order of increasing horizontal resolution (see Table 1).

across the model generations, shifting from around 200–400 km horizontally in the atmosphere in CMIP3 to around 100–200 km in CMIP6 (Schneider et al., 2017b; Intergovernmental Panel on Climate Change, 2021).

To specifically examine the impact of resolution, we consider two models from the High Resolution Model Intercomparison Project (HighResMIP; Haarsma et al., 2016): HadGEM3-GC3.1 (Roberts et al., 2019) and MPI-ESM1.2 (Gutjahr et al., 2019). These models have conducted simulations at three different resolutions, with horizontal resolutions in the atmosphere between 25 and 200 km (Table 1). Both models exhibit a modest but consistent reduction in error metrics as resolution increases. However, there is one exception: the doubling in atmospheric horizontal resolution, without an increase in ocean resolution or

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

Table 1. Atmosphere and ocean model resolutions of HighResMIP simulations included in Fig. 1.

	Atmos. Res.	Ocean Res.	Vertical Levels
HadGEM3-GC3.1-LL	N96 (135 km)	100 km	85
HadGEM3-GC3.1-MM	N216 (60 km)	25 km	85
HadGEM3-GC3.1-HH	N512 (25 km)	8 km	85
MPI-ESM1.2-LR	T63 (200 km)	150 km	47
MPI-ESM1.2-HR	T127 (100 km)	40 km	95
MPI-ESM1.2-XR	T255 (50 km)	40 km	95

atmospheric vertical resolution, from MPI-ESM1.2-HR to MPI-ESM1.2-XR did not result in an improvement in error metrics. This finding suggests that ocean resolution and atmospheric vertical resolution are also important factors contributing to the improvements with resolution.

Recently, there has been a push to increase the resolution of climate models even further to kilometer scales, allowing for partial resolution of deep convection and potential improvements in simulating precipitation and its extremes (Bauer et al., 2021; Slingo et al., 2022). Figure 1 displays the rms errors of two such models (IFS and ICON) in simulating the seasonal climatology of TOA radiative energy fluxes and precipitation. These simulations use a prescribed seasonally varying climatology of sea surface temperatures (SSTs) and can be compared to the ensemble of coarser-resolution models in the Atmospheric Model Intercomparison Project (AMIP). Figure 1 also shows the rms errors in AMIP for 1- and 2-year averaging periods equivalent to the length of the IFS and ICON simulations, respectively. Compared to the coarser-resolution AMIP simulations, the kilometer-scale simulations show modest improvements in TOA shortwave fluxes and longwave fluxes, but increased errors in precipitation. These simulations highlight that higher resolution alone does not guarantee an improved fit in climate simulations. Many crucial climate-regulating processes, such as shallow clouds and cloud microphysics, remain unresolved at kilometer-scales, requiring appropriate parameterization. Extensive calibration or even re-design of subgrid-scale parameterizations is necessary to reduce large-scale biases that can otherwise exceed those seen in coarser-resolution models (Hohenegger et al., 2023).

Figure 2 provides a more detailed illustration of how kilometer-scale models can inherit longstanding biases from coarse-resolution models. The figure compares August precipitation between observations and simulations. The simulations include coarse-resolution AMIP models and a set of kilometer-scale simulations conducted under the DYAMOND project (Stevens et al., 2019). The figure reveals that the kilometer-scale simulations capture more intricate details in the precipitation patterns, such as the strong orographic precipitation in the Himalayas, New Guinea, and the Sierra Madre Occidental. However, they still exhibit similar large-scale biases as the coarse-resolution simulations, such as excessive precipitation over the tropical regions of the south Pacific and Indian Oceans, commonly referred to as the double-ITCZ bias (Tian and Dong, 2020). The double-ITCZ bias has important implications for regional precipitation projections over land (Dong et al., 2021).

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

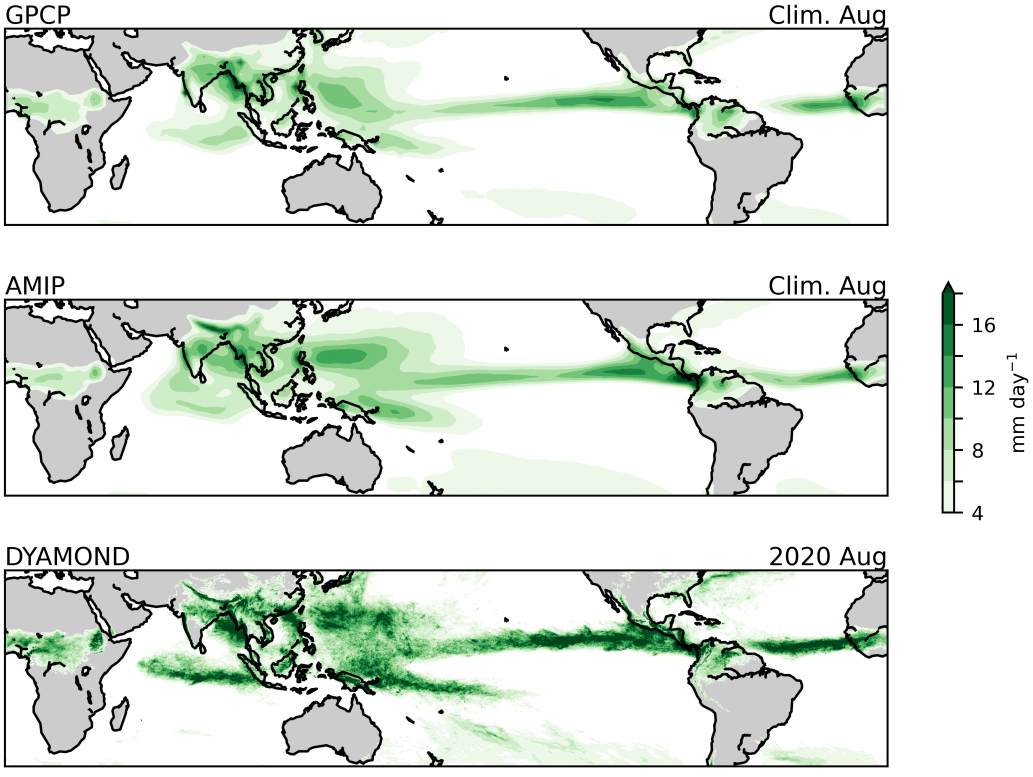

Figure 2. August precipitation in satellite observations (top), coarse-resolution AMIP simulations (middle), and kilometer-scale DYAMOND simulations (bottom). Observations are based on the Global Precipitation Climatology Project (GPCP) (Adler et al., 2018). AMIP simulations are from 14 models that participated in CMIP6. Both GPCP data and AMIP simulations are interpolated to a common $1^\circ \times 1^\circ$ grid, and the August climatology is derived from 1979–2014. The DYAMOND results, shown at the model native resolutions, are based on the average of 5 models with horizontal resolutions ranging from 3.3 km to 7.8 km for August 2020. Figure adapted from Zhou et al. (2022).

Over the past two decades, then, climate models have shown gradual improvements in key metrics, with error reductions of 10–20% per decade, as seen in Figure 1 and in other studies (Bock et al., 2020). However, there are still errors that are large compared to the climate change signals we aim to predict. For instance, the radiative forcing due to doubling CO_2 concentrations is about $4\ \mathrm{W\ m^{-2}}$, while rms errors in TOA radiative energy fluxes are $O(10\ \mathrm{W\ m^{-2}})$. The response of climate models to increasing greenhouse gas concentrations also varies widely across models. For example, the time when the 2°C warming threshold of the Paris agreement is exceeded varies by several decades among models (Schneider et al., 2017b; Intergovernmental Panel on Climate Change, 2021). This indicates that there is significant room for further improvement.

Given the significant errors in simulating the current climate and the uncertainties in future projections, there's a large gap between the demands placed on climate models for adaptation decisions—such as designing stormwater management systems or sea walls to handle a 100-year flood in the decades ahead—and the capabilities of models today (Fiedler et al., 2021;

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

President's Council of Advisors on Science and Technology, 2023). Yet, the need for such decision-making is immediate. Therefore, it is urgent to accelerate the improvement of climate models, aiming for a step change enhancement in both accuracy and usability for decision-making, beyond the gradual advances of recent decades. The question is how to achieve such a step change.

3 Process-based parameterizations

The uncertainties and biases in climate simulations, as shown in Figs. 1 and 2, have their roots in the parameterization of unresolved small-scale processes. So far, these processes have been primarily parameterized based on process knowledge, in a reductionist approach. For example, the influential work of Arakawa and Schubert (1974) laid the foundation for widely used parameterizations of moist convection, employing a reductionist process model of convective plumes that are, at all times, in statistical equilibrium with their environment and incorporate environmental air by entrainment. Research over the past two decades has focused on refining the formulation of the entrainment rate, a key control on climate model sensitivity to greenhouse gas concentrations (Stainforth et al., 2005; Knight et al., 2007). Typically, this rate is represented by a constant parameter $\epsilon = \theta$ or a parametric function $\epsilon = \epsilon(z, \boldsymbol{\zeta}; \boldsymbol{\theta})$ of height z and (usually local) plume and environmental properties encoded in the model state $\boldsymbol{\zeta}$ (e.g., de Rooy et al., 2013; Yeo and Romps, 2013; Anber et al., 2019; Savre and Herzog, 2019; Cohen et al., 2020). Similarly, diffusive closures of various types have been commonly employed for boundary layer turbulence in the atmosphere and oceans. These closures employ diffusivities that may depend on height, other flow variables, or a turbulence kinetic energy determined by separate equations, and they are sometimes augmented by correction terms to represent upgradient fluxes (e.g., Mellor and Yamada, 1982; Large et al., 1994; Lock et al., 2000).

The process-based approach offers the advantage that the parameters or parametric functions that require closure are interpretable and theoretically measurable. For example, Monin-Obukhov similarity theory reduced the problem of parameterizing turbulence in a thin ($\sim$100 m) layer near the surface to finding universal functions that characterize the vertical structure of turbulent fluxes (Foken, 2006). Later, these functions were empirically derived based on measurements over a field of wheat stubble during the summer of 1968 in Kansas (Businger et al., 1971); they have since been widely incorporated into climate models. This represents a success story for the process-based approach. It led to a parametrically sparse and interpretable representation of near-surface turbulent fluxes. It applies not only to summer conditions over Kansas wheat fields but also demonstrates invariance properties that make it applicable across most of the globe, with a few limitations, particularly in convective situations.

However, despite this progress, the dominant source of uncertainties and biases in climate simulations, even 50 years after the introduction of the convection parameterization by Arakawa and Schubert (1974), lies in the representation of turbulence, convection, and clouds above the near-surface layer. This indicates that the reductionist approach to developing process-based models for these components has encountered significant challenges. For example, measuring entrainment rates directly, despite being theoretically possible, remains challenging in practice, both in observational data and high-resolution simulations (e.g., Romps, 2010). The search for universal functions to accurately represent entrainment has been unsuccessful thus far. Con-

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

sequently, the process-based approach to modeling convection and clouds is widely perceived as being deadlocked (Randall et al., 2003; Randall, 2013).

However, prematurely dismissing process-based modeling as obsolete would ignore its advantages and its potential for further development. Contrasting the achievements of Monin-Obukhov similarity theory and moist convection parameterizations is illuminating. Monin-Obukhov similarity theory systematically coarse-grained the equations of motion, employing controlled approximations and identifying the nondimensional groups of variables that govern near-surface turbulent fluxes. The approach reduced the closure problem to finding universal functions of the identified nondimensional groups, with well-defined limits in different scenarios. This led to its near-universal applicability. In contrast, typical moist convection parameterizations in current use emerged phenomenologically, without a systematic coarse-graining of the known equations of motion through controlled approximations. Even when starting from a rigorous basis like the Arakawa and Schubert (1974) parameterization, operational parameterizations often introduced artificial scale breaks between boundary layer turbulence, shallow convection, and deep convection, or even between convection over land and oceans, leading to separate parameterizations with discontinuous differences in parameters and structure. Such discontinuities do not exist in nature. As a result, these parameterizations lack well-defined limits. For example, they do not converge to a well-defined dry limit when the latent heats of fusion and vaporization of water approach zero, and they do not converge in resolution. This approach hindered the systematic removal of unnecessary approximations, particularly as model resolution increased and common assumptions, such as small plume area fractions relative to the host model's grid scale, or statistical equilibrium between moist convection and the environment, became inadequate (Arakawa et al., 2011; Arakawa and Wu, 2013; Randall, 2013). Therefore, rather than declaring process-based modeling for moist convection and other complex processes at a dead end, a more nuanced perspective recognizes the need for further development with greater mathematical and physical rigor, particularly in light of the abundant data and enhanced computational capabilities available today that surpass what the early pioneers of these approaches had at their disposal. The invariance properties, such as conservation laws and symmetries, inherited by this approach from the underlying equations of motion, may well hold the key to developing universal parameterizations that enable us to minimize the loss (1) for unobserved climate statistics and generalize beyond the observed distributions.[2]

These considerations suggest that successful process-informed parameterizations satisfy three clear requirements:

1. Parameterizations should be based on the governing equations of subgrid-scale processes whenever possible. These equations can be systematically coarse-grained through methods like conditional averaging or moment equations derived from distribution assumptions on subgrid-scale fluctuations. Examples of such approaches include Lappen and Randall (2001), Golaz et al. (2002), Guo et al. (2015), Firl and Randall (2015), Tan et al. (2018), Thuburn et al. (2018), Cohen et al. (2020), and Lopez-Gomez et al. (2020).

2. Artificial scale breaks, such as those between shallow and deep convection, should be avoided. These breaks have no correspondence in nature but introduce correlated parameters that are difficult to calibrate. For example, parameters

[2]That is, progress on macroscopic *techne* here hinges on microscopic *episteme*.

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

in shallow and deep convection schemes are necessarily correlated. Moreover, bridging these scale breaks becomes challenging as resolution increases.

3. When scale separation is absent between the parameterized subgrid-scale processes and the resolved grid-scale, parameterizations must incorporate subgrid-scale memory and stochastic terms. This means that convection and cloud parameterizations, for example, need to be explicitly time-dependent (i.e., have memory) and cannot be assumed to be in instantaneous equilibrium with the environment. Homogenization theories such as those of Mori and Zwanzig, which employs averaging but also shows how fluctuations about averages arise on the macroscale, supports the inclusion of these features (Majda et al., 2003; Wouters and Lucarini, 2013; Lucarini et al., 2014; Wouters et al., 2016; Lucarini and Chekroun, 2023).

Developing process-based parameterizations by systematically coarse-graining equations of motion will lead to unclosed terms, similar to the universal functions in Monin-Obukhov similarity theory. These terms, whether they contain parameters, parametric functions, or non-parametric functions, present excellent targets for AI-enabled learning from data.

There is accumulating evidence that a program focused on process-based parameterizations can achieve success. For example, Lopez-Gomez et al. (2020) and Cohen et al. (2020) have demonstrated the effectiveness of a unified parameterization approach for boundary layer turbulence and moist convection. This parametrically sparse approach, based on conditionally averaged equations of motion, accurately represents a wide range of cloud dynamics observed on Earth, from stable boundary layers to stratocumulus-topped boundary layers and deep convection. Furthermore, Lopez-Gomez et al. (2022) have shown that machine learning can be employed to identify closure functions in these parameterizations, such as entrainment rates that depend on nondimensional groups.

As climate models reach resolutions where deep convection becomes marginally resolved, using an inadequate deep-convection parameterization based on instantaneous statistical equilibrium may well be less effective than not using any parameterization at all. In the kilometer-scale simulations shown in Figures 1 and 2, deep convection parameterizations are entirely turned off for this reason (Stevens et al., 2019). However, parameterizations for boundary layer turbulence and low-cloud cover are usually kept, and sometimes also those for shallow convection, even though they were originally developed for resolutions in the hundred kilometer range, where, for example, assumptions of instantaneous statistical equilibrium of subgrid-scales with resolved scales are more justifiable. As seen in the above figures, this approach has not achieved the anticipated success; in particular, it has not significantly improved large-scale precipitation simulations at kilometer-scale resolutions. Consequently, it is essential to advance the development of parameterizations that effectively bridge the scales between marginally resolved convection and the dynamics that remain unresolved in this resolution range, in addition to parameterizations of yet smaller scales, such as the microphysics of cloud droplet and ice crystal formation.

4 Resolution

Climate is regulated by turbulent motions in the atmosphere and oceans. Horizontal motions transport energy, momentum, and, in the atmosphere, water vapor, shaping surface temperatures, winds, and precipitation patterns. Vertical motions couple the

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

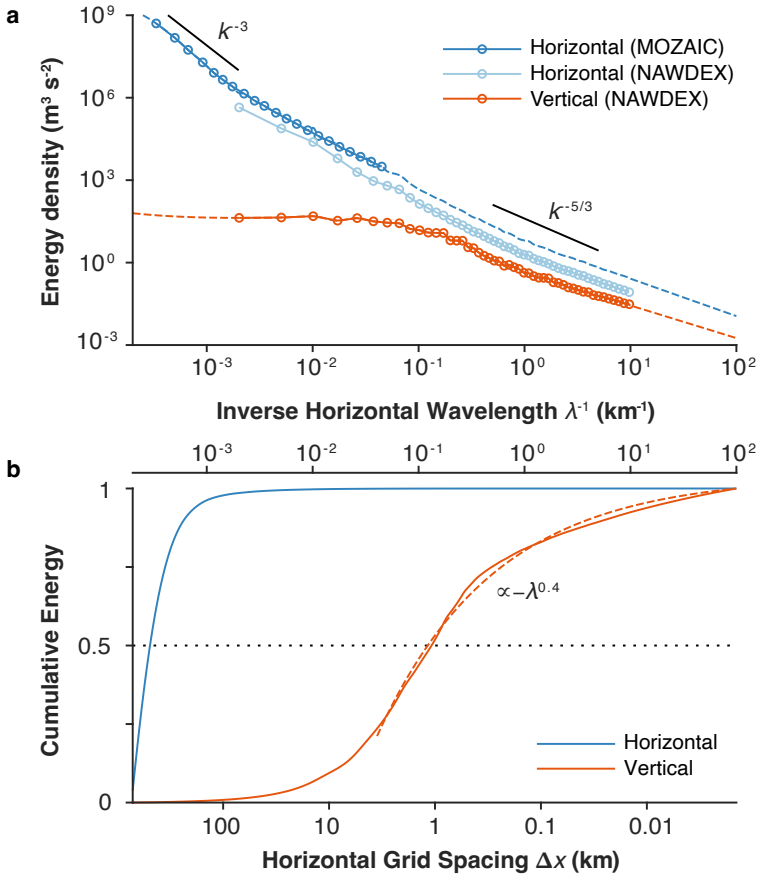

Figure 3. Kinetic energy spectra and cumulative energy in the atmosphere. (a) Spectral kinetic energy density based on aircraft measurements, shown as a function of inverse wavelength $\lambda^{-1} = k/(2\pi)$. (b) Cumulative kinetic energy from $\lambda_{\max} = 5000$ km to λ on the upper horizontal axis, normalized by the energy for $\lambda_{\min} = 10$ m; the lower horizontal axis expresses λ through the required horizontal model grid spacing Δx, using $\lambda \approx 7\Delta x$. Blue for horizontal motion kinetic energy, red for vertical motion kinetic energy. Data from Callies et al. (2014) (MOZAIC) and Schumann (2019) (NAWDEX). Dashed lines in (a) indicate linear extrapolations in log-log space, except for the dashed blue line where NAWDEX and MOZAIC data overlap: there the dashed line represents the NAWDEX spectrum multiplied by a fitting constant to match the MOZAIC spectrum. Cumulative energies are obtained by numerical integration over the spectra, including the extrapolations. The dashed line in (b), for wavelengths $\lambda \leq 25$ km, represents a power law fit $1 - a(\lambda^{\beta} - \lambda_{\min}^{\beta})$ to the cumulative vertical kinetic energy, with $\beta \approx 0.4$ and $a \approx 0.2$. Note that the extrapolations of vertical kinetic energy to large scales may not be very accurate due to possible deviations from a completely flat spectrum, which can slightly shift the position of the inflection point in the cumulative energy (Skamarock et al., 2014; Schumann, 2019).

atmosphere and surface, creating clouds, driving precipitation, and mixing heat and tracers like carbon dioxide in the oceans. Representing these turbulent motions accurately is crucial for climate models, but challenging due to their vast range of length scales, from planetary to millimeter scales.

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

Figure 3a shows the kinetic energy spectrum of horizontal and vertical motions in the atmosphere, measured by aircraft. The spectra are displayed as functions of the inverse of the horizontal wavelength λ, which is proportional to the horizontal wavenumber $k = 2\pi/\lambda$. At large scales (small wavenumbers), the spectrum of horizontal kinetic energy follows a k^{-3} power law, as predicted by geostrophic turbulence theory (Vallis, 2006, chapter 9). At mesoscales below approximately 500 km, the spectrum becomes shallower, resembling a $k^{-5/3}$ power law. The reason for this change has been debated. The shallower spectrum seems to be caused by linear inertia-gravity waves, which are internal waves modified by planetary rotation that coexist with the nonlinear, primarily geostrophic, atmospheric turbulence (Dewan, 1979; VanZandt, 1982; Callies et al., 2014).

At scales greater than 10–20 km, the kinetic energy of vertical motions is much weaker than that of horizontal motions, with a relatively flat spectrum. This difference is mainly due to two factors: (1) The scale of vertical motions is limited by the depth of the troposphere (about 10–20 km), which contains the most important vertical motions; and (2) the vertical velocity depends on the divergence of the horizontal velocity, which is weaker (by a factor of order Rossby number) than the dominant rotational velocity at large scales, though it becomes comparable to it on mesoscales. The divergence involves horizontal derivatives, leading to a multiplication by k^2 of the kinetic energy spectra in wavenumber space at horizontal scales above 10–20 km, where the vertical depth scale is constrained by the depth of the troposphere (see Schumann (2019) for a detailed model, from which these insights are drawn). This results in the relatively flat spectrum with low vertical kinetic energy at larger horizontal scales. At horizontal scales smaller than about 10–20 km, where the horizontal scale is comparable to the vertical scale and not constrained by the depth of the troposphere, the vertical kinetic energy spectrum starts to decay following a rate of roughly $k^{-5/3}$, like the horizontal kinetic energy spectrum. At yet smaller scales in the meter range, the turbulence becomes increasingly isotropic, which results in a $k^{-5/3}$ power law due to three-dimensional turbulence following a Kolmogorov spectrum. The figure shows an extrapolation of both the horizontal and vertical kinetic energy spectra from the smallest measured scale near 100 m down to 10 m for illustrative purposes. However, in reality, the spectra continue without a break to the Kolmogorov scale at millimeters, where kinetic energy is dissipated.

As climate model resolution increases, the continuity of the atmospheric energy spectrum implies a gradual improvement as resolved motions replace imperfectly parameterized smaller scales. To quantitatively assess the benefits of higher resolution in climate models, we integrate the energy spectra $\hat{E}(k)$ over a wavenumber interval from $k_{\min} = 2\pi/\lambda_{\max}$ to $k = 2\pi/\lambda$:

$$E(k_{\min}, k) \propto \int_{k_{\min}}^{k} \hat{E}(k')\, dk'. \tag{2}$$

Figure 3b illustrates the cumulative energy contained between $\lambda_{\max} = 5000$ km and a given λ on the upper horizontal axis, normalized by the cumulative energy extrapolated to $\lambda_{\min} = 10$ m. Because of the steepness of the horizontal kinetic energy spectrum at large scales, the benefits of increased resolution for horizontal kinetic energy level off at wavelengths just under 1000 km, corresponding to a grid spacing around $\Delta x \approx 150$ km (lower horizontal axis in Fig. 3) because the minimum wavelength λ a model can resolve is approximately $7\Delta x$ (Skamarock, 2004; Wedi, 2014; Klaver et al., 2020). Climate models reached this "geostrophic turbulence plateau" in resolution in the past decade (Schneider et al., 2017b). However, the vertical kinetic energy spectrum remains relatively flat at larger scales, leading to continued benefits in resolving vertical kinetic energy

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

as λ decreases.[3] Concretely, the data in Fig. 3 indicate that resolving wavelengths of 1000 km, 100 km, and 10 km (grid spacings Δx of about 150 km, 15 km, and 1.5 km, respectively) increases the fraction of resolved vertical kinetic energy between 5000 km and 10 m from 0.6% to 7% and 43%, respectively. The returns on increasing resolution only begin to diminish for wavelengths below 1 km, that is, grid spacings $\Delta x \lesssim 150$ m. The specific results depend on the data and extrapolations used (Schumann, 2019), but the main finding is clear: Even at kilometer-scale resolution, most vertical motions require parameterization, especially those generating the low clouds that impact Earth's energy balance (Bony and Dufresne, 2005; Stephens, 2005; Vial et al., 2013; Schneider et al., 2017b, 2019). As we push the resolution frontier, it is crucial to concurrently improve parameterization of smaller-scale turbulent motions.

Increasing horizontal model resolution comes with a substantial computational cost, which grows as $(\Delta x)^{-3}$ for fixed vertical resolution. This cost arises from the increasing number of horizontal grid points, $\propto (\Delta x)^{-2}$, and the need for smaller timesteps, $\propto (\Delta x)^{-1}$, to maintain numerical stability. To illustrate using the previous example from Fig. 3, reducing the grid spacing from $\Delta x \approx 15$ km to $\Delta x \approx 1.5$ km increases the computational cost by a factor 1000, while increasing the cumulative vertical kinetic energy resolved by a factor 6, from 7% to 43%. That is, the rate at which the resolved vertical kinetic energy improves (a low power of Δx) is very small relative to the additional computational expenditure. From this perspective, increasing horizontal resolution is an inefficient means of improving climate models. Moreover, the vertical grid spacing Δz must also be considered. Increasing vertical resolution incurs a more modest computational cost, typically scaling as $(\Delta z)^{-1}$. This is because fast vertical dynamics are generally treated implicitly in climate models to circumvent timestep limitations, ideally using implicit solvers with computational costs linear in $(\Delta z)^{-1}$. As hinted at in Fig. 1, increasing vertical resolution at coarser horizontal resolution can be advantageous because it can improve the representation of parameterized subgrid-scale dynamics (Harlaß et al., 2015; Kawai et al., 2019; Smalley et al., 2023). At higher horizontal resolutions where resolved dynamics become more isotropic, proportionately increasing both vertical and horizontal resolution becomes necessary, leading to a computational cost that scales even less favorably, like $(\Delta x)^{-4}$.

Therefore, making trade-offs in optimizing the parameters $\boldsymbol{\lambda}$ characterizing model resolution is essential because, even at foreseeable future resolutions, unresolved scales of atmosphere and ocean turbulence, plus even finer scales controlling cloud microphysics and other processes, will still require parameterization. In particular, it is crucial to invest some of the available computational resources in enabling parameterizations to be calibrated with and learn from data, which requires hundreds to thousands of climate simulations. This practical limitation, in addition to the need to run large ensembles of climate simulations to broadly explore possible climate outcomes, will determine the resolution attainable in climate models. In the atmosphere and oceans, this routinely achievable resolution currently lies in the $O(10\ \mathrm{km})$ range (Schneider et al., 2023). By finding the right balance between resolution and parameterization learning and calibration, we can make significant strides in improving climate simulations within realistic computational constraints.

[3] For a $k^{-\alpha}$ spectrum, the cumulative vertical kinetic energy scales as $-\lambda^{\beta}$, where $\beta = \alpha - 1$. This gives $\beta = 2/3$ for $\alpha = 5/3$; the curves in Fig. 3, for wavelengths $\lambda < 25$ km, are fit well with $\beta \approx 0.4$, showing that there is no qualitative change in behavior over those scales.

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

5 AI for learning parameterizations

Even at the highest achievable climate model resolutions, parameterizing small-scale processes remains essential. Process-based parameterizations, encoding conservation laws and invariance properties, are promising for generalizing beyond observed climate data, but they include unclosed terms and functions. AI-based data-driven methods, broadly understood to include a spectrum of methods from Bayesian to deep learning, can aid in learning about these terms and functions or about entire parameterizations, thus reducing inaccuracies in climate models and potentially also allowing the quantification of the remaining uncertainties.

AI approaches require the specification of a loss function. The most suitable is a loss function (1) that penalizes differences between simulated and observed climate statistics. This function should include variables such as TOA radiative energy fluxes and global precipitation fields, as shown in Fig. 1. It might also include higher-order statistics, such as the covariance between surface temperature and cloud cover (Schneider et al., 2017a). This represents an emergent constraint: a statistic that, across climate models, correlates with the response of cloud cover to greenhouse gas concentration increases (e.g., Klein and Hall, 2015; Brient and Schneider, 2016; Caldwell et al., 2018; Hall et al., 2019). These emergent constraint statistics, previously used only for retrospective model assessments, can be proactively minimized in the loss function to improve model accuracy in simulating greenhouse gas responses.

However, using climate statistics in a loss function challenges traditional machine learning (ML) methods. Supervised learning (SL), the dominant ML approach, depends on labeled input-output pairs for process modeling and learns regressions of outputs onto inputs. For example, a convection parameterization requires at least temperature and humidity inputs, which must be paired with the output—convective time tendencies of temperature and humidity—for training. Since such data are unavailable from Earth observations, SL has been limited to simulated data (e.g., O'Gorman and Dwyer, 2018; Rasp et al., 2018; Gentine et al., 2018; Yuval and O'Gorman, 2020; Yu et al., 2023). Conversely, the climate statistics in the loss function (1) provide only indirect information about processes such as convection. For example, the loss function may include fields such as precipitation and cloud cover—noisy fields with missing data that are influenced by multiple processes, including but not limited to convection (Schneider et al., 2017a).

To illustrate, consider determining closures in a conservation equation:

$$\frac{Dq}{Dt} = \mathcal{F} + \mathcal{S}. \tag{3}$$

Here, $q(\boldsymbol{x},t)$ is a tracer, such as total specific humidity, dependent on space $\boldsymbol{x}$ and time t, and $D/Dt = \partial/\partial t + \boldsymbol{u}\cdot\nabla$ is the material derivative with fluid velocity $\boldsymbol{u}(\boldsymbol{x},t)$. The quantities on the left-hand side are taken to be resolved on the model's grid. The right-hand side consists of two components: $\mathcal{F}(\boldsymbol{x},t)$ represents unresolved subgrid-scale flux divergences in need of parameterization; $\mathcal{S}(\boldsymbol{x},t)$ denotes all other, separately modeled sources and sinks.

In SL approaches, the objective is to map the model state $\boldsymbol{\zeta}$ to approximate SGS flux divergences $\hat{\mathcal{F}}(\boldsymbol{\zeta};\boldsymbol{\nu})$ holistically, with parameters $\boldsymbol{\nu}$ (e.g., neural network weights and biases). To make the problem tractable, the mapping is usually considered locally in the horizontal, mapping column states $\boldsymbol{\zeta}(x,y,z_i,t)$ at discrete levels z_i (for $i=1,\ldots,N_z$) to parameterized flux divergences $\hat{\mathcal{F}}_j\big(\boldsymbol{\zeta}(x,y,z_i,t);\boldsymbol{\nu}\big)$ at levels z_j. This is achieved by using the column state $\boldsymbol{\zeta}(x,y,z_i,t)$ as input, and the remainder

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

$Dq/Dt - \mathcal{S}$ as output, to learn a regression $\hat{\mathcal{F}} \approx Dq/Dt - \mathcal{S} + \epsilon$. The material derivative Dq/Dt is rolled out over time intervals typically spanning hours to days. The aim is to minimize the residual ϵ over parameters $\boldsymbol{\nu}$, typically using methods such as backpropagation that compute gradients of the loss function with respect to the parameters $\boldsymbol{\nu}$.

This approach leverages the expressive capabilities of deep learning and has shown some promise, as evidenced in studies demonstrating that moist convection or ocean turbulence parameterizations that are accurate over short time intervals can be effectively learned in this manner (e.g., O'Gorman and Dwyer, 2018; Rasp et al., 2018; Gentine et al., 2018; Bolton and Zanna, 2019; Yuval and O'Gorman, 2020; Zanna and Bolton, 2020). Error corrections to existing parameterizations have been learned in a similar manner (Bretherton et al., 2022). However, focusing solely on minimizing the short-term residual ϵ presents several limitations:

1. The learned parameterization $\hat{\mathcal{F}}$ may not necessarily minimize the climate-relevant loss function (1), which is concerned with longer-term statistics.

2. Supervised learning of the parameterization $\hat{\mathcal{F}}$ is typically restricted to data generated computationally in higher-resolution simulations, restricting it to the emulation of imperfect models, because labeled parameterization output $Dq/Dt - \mathcal{S}$ is generally not available from Earth observations.

3. The parameterization $\hat{\mathcal{F}}$, learned holistically for the multitude of processes it comprises, usually does not generalize well out of the training distribution, necessitating training with a broad range of simulated climates (e.g., O'Gorman and Dwyer, 2018).

4. Climate models incorporating the learned parameterization $\hat{\mathcal{F}}$ often struggle with conserving essential quantities such as energy and exhibit instabilities during extended integrations (e.g., Brenowitz and Bretherton, 2018), because minimizing the short-term residual ϵ does not inherently ensure conservation or stability.

5. The learned parameterization $\hat{\mathcal{F}}$ is resolution dependent because it is based on discretized model states at a particular resolution, necessitating re-training whenever the resolution is changed.

Some of the challenges associated with SL approaches in climate modeling can be addressed or alleviated. For example, longer roll-outs of the material derivative Dq/Dt have been shown to reduce instabilities when integrating the learned parameterization $\hat{\mathcal{F}}$ into a climate model (Brenowitz et al., 2020), and constraints on the loss function can be used to enforce conservation laws (Beucler et al., 2021). Additionally, the issue of resolution dependence in the learned parameterization $\hat{\mathcal{F}}$ can be tackled by shifting from learning a finite-dimensional discrete mapping between model grid points to learning an infinite-dimensional operator. Such operators map between function spaces; they would effectively represent the atmospheric or oceanic column state as a continuous function, rather than as a set of discrete points (e.g., Nelsen and Stuart, 2021; Kovachki et al., 2023). This approach allows for a more flexible representation of the underlying physical processes, potentially adaptable to different resolutions without the need for retraining.

An alternative approach that avoids the restrictions of SL views learning parameterizations $\hat{\mathcal{F}}$ and ML parameters $\boldsymbol{\nu}$ within them as an inverse problem, minimizing a climate-relevant loss function (1) (Kovachki and Stuart, 2019). This loss function is

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

based on data that are only indirectly informative about the process being modeled; that is, the parameterization $\hat{\mathcal{F}}$ influences the climate model output $\mathcal{G}(t;\boldsymbol{\theta},\boldsymbol{\lambda},\boldsymbol{\nu};\boldsymbol{\zeta}_0)$ in the loss function only indirectly, through the complex and nonlinear interactions of other components in the climate model (Schneider et al., 2017a). Gradients of the loss function with respect to the parameters $\boldsymbol{\nu}$ in this case would involve differentiation through the model $\mathcal{G}$, which may not be differentiable or difficult to differentiate.

In this context, learning about the parameterization $\hat{\mathcal{F}}$ is no longer a straightforward regression of outputs onto inputs. However, this does not preclude the inclusion of parameter-rich and expressive deep learning models within the parameterization. The parameters $\boldsymbol{\nu}$ can be estimated by minimizing a climate-relevant loss function (1) using derivative-free ensemble Kalman inversion techniques, which can be used with models for which gradients are unavailable (Kovachki and Stuart, 2019). Furthermore, this inverse problem approach not only makes it possible to learn from heterogeneous and noisy Earth observations but also allows for the quantification of uncertainties (e.g., Cleary et al., 2021; Huang et al., 2022). Stochastic elements can also be incorporated in the parameterizations (e.g., Schneider et al., 2021b), which, as discussed in section 3, is particularly relevant in the absence of clear scale separation, offering a more principled and realistic representation of climate processes.

This expanded view of applying AI methods in climate models broadens the scope of where these methods can be effectively integrated. Instead of focusing on areas where SL is feasible, the emphasis shifts to where AI can have the most significant impact. The key challenge in climate modeling and prediction is minimizing the loss function (1) for unobserved climate statistics, especially in global warming scenarios where these statistics may fall outside the range of observed data. While traditional methods such as withholding part of the data for cross-validation are essential, they fall short in ensuring model generalization beyond the training dataset. This limitation is evident when predicting, for example, the direct effects of CO_2 on cloud cover (Bretherton, 2015) or on the photosynthetic productivity of the biosphere (Luo, 2007), which are challenging to infer from the limited variations in recent CO_2 concentrations.

A valuable insight emerges from the success of similarity theories, such as the Monin-Obukhov similarity theory discussed in section 3, which generalized effectively from a few specific measurements to a wide range of global conditions. Similarly, AI methods may be most effective when used to learn universal functions of relevant nondimensional variable groups: functions that likely remain invariant across different climates and are well-sampled in current climate conditions, including the seasonal cycle whose amplitude in many quantities exceeds the climate change signals we expect for the coming decades (Schneider et al., 2021a). If such learning minimizes a loss (1) in an online setting, it leads to long-term stable models, because existence of the statistics in the loss function implies that the model is stable over the timescales over which the statistics are aggregated. For example, rather than learning the convective flux divergence $\hat{\mathcal{F}}$ for water or energy holistically, it is likely beneficial to focus on learning key unknown functions such as entrainment and detrainment rates within the coarse-grained conservation laws for water and energy, embedded online in a larger forward model. Lopez-Gomez et al. (2022) demonstrated that this approach can successfully learn parameterizations that generalize well to warmer climates not encountered during training. An ancillary benefit is that quantification of uncertainties becomes more straightforward, and the parameterizations remain interpretable, facilitating the investigation of mechanisms, for example, of cloud feedbacks and the differential effects of changing greenhouse gas concentrations and warming in them. Models for structural errors can similarly be incorporated where the errors are actually made—within the parameterizations of unresolvable small-scale processes (Levine and Stuart, 2022; Wu et al., 2024).

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

Therefore, we advocate for an approach that leverages our extensive knowledge of conservation laws, expressed as partial differential equations, and enhances it with AI methods to learn about closure functions in parameterizations where reductionist first-principle approaches fall short. The central challenge is to find a balance: using first principles to encode system knowledge and conservation laws for generalization and interpretability, while avoiding overly rigid constraints that limit the model's adaptability to diverse data sets. This balance will vary across different components of the climate system. For instance, first principle modeling has proven less effective than data-driven approaches for river flows and snowpack thickness, where systematic coarse-graining is challenging and observed spatial and temporal variations plausibly capture future scenarios (Kratzert et al., 2018, 2019; Nearing et al., 2021; Kratzert et al., 2023; Moshe et al., 2020; Charbonneau et al., 2023). In contrast, modeling phenomena such as turbulence, convection, and clouds may benefit more from reductionist process-informed modeling, among other reasons because direct radiative effects outside observed climates impact clouds and cannot be learned solely from data. Striking the right balance is crucial for developing climate models that are physically grounded and trustworthy for predictions beyond observed climates, yet flexible enough to integrate a wide range of observational data, leading to more accurate and reliable predictions.

6 A balanced path forward

Climate models, as a form of *techne*, aim to provide the most accurate and reliable predictions of how the climate system's statistics will change under unobserved conditions, such as increased greenhouse gas concentrations. Improving climate models is urgent for proactive and effective adaptation to the coming climate changes. However, current models fall short in accuracy and reliability (Fiedler et al., 2021; President's Council of Advisors on Science and Technology, 2023), as evidenced by their still significant errors in simulating observed climate statistics (Fig. 1).

Progress in climate modeling has been gradual, achieved primarily through increasing resolution and refining process-based parameterizations for small, unresolvable scales. Yet, neither approach alone, nor in combination, seems likely to produce a significant leap in model accuracy and reliability. The complexity of the climate system limits the effectiveness of reductionist approaches in developing process-based parameterizations. Additionally, while higher resolution is beneficial, it is no panacea. At any resolution reachable in the foreseeable future, many aspects, such as large portions of the atmosphere's vertical motion and finer scales such as those controlling cloud microphysics, will remain unresolvable.

AI tools, broadly defined, hold promise for breakthroughs due to their capacity to learn from high-resolution simulations and from the extensive array of Earth observations available. However, they cannot operate in isolation. Climate change prediction is an archetypal out-of-distribution prediction challenge. It is hard to envision how an unsupervised AI system could learn the effects of unseen greenhouse gas concentrations on aspects such as cloud cover or the biosphere's photosynthetic productivity using only higher-resolution simulations and current and recent past observations. Unlike weather forecasting, where short-term predictions can be validated daily and long-term stability and conservation properties of simulations are less critical, climate prediction lacks the luxury of immediate validation. Conservation, long-term stability in an "infinite forecast," and reliable generalization beyond observed climate states are essential.

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

Therefore, a balanced approach that capitalizes on the strengths of all three dimensions—advancing process-based parameterizations, maximizing resolution while allowing large ensembles of simulations, and harnessing AI tools to incorporate data-driven models where reductionism reaches its limits—appears to be the most promising path forward (Schneider et al., 2021a). In situations where we have well-defined equations of motion and can systematically coarse-grain them, AI may be optimally employed to learn data-driven yet climate-invariant closure functions of nondimensional variable groups arising within coarse-grained equations. This approach is akin to how data have been used to close Monin-Obukhov similarity theory for the atmospheric surface layer. Conversely, in situations where first-principle modeling and systematic coarse-graining are less effective, but where spatial and temporal climate variations—particularly the seasonal cycle—may plausibly represent future climate states, more direct data-driven models could prove more fruitful. This may be particularly relevant for various aspects of land surface modeling, such as snow, vegetation, and river models.

Well-defined competitive challenge problems, employing open benchmark data, shared code, and clear quantitative success metrics, may catalyze advances in climate modeling, as they have in other areas, such as machine vision, natural language processing, and protein folding (Donoho, 2023). For example, benchmark challenges for cloud parameterizations could leverage libraries of high-resolution simulations, employing current-climate simulations for training and altered climate conditions for evaluation (e.g., Hourdin et al., 2021; Shen et al., 2022; Lopez-Gomez et al., 2022; Yu et al., 2023). Other benchmark challenges may focus on the seasonal cycle of land carbon uptake, evapotranspiration, snow cover, or river discharge, utilizing a subset of the available data for model training while reserving other regional datasets for evaluation. Such structured challenges can foster innovation in climate process modeling and can help determine what balance between process-based and data-driven methods is most successful.

Ultimately, the utility of climate predictions for decision-making hinges on their trustworthiness and their ability to explore a broad range of possible climate outcomes through large ensembles (Deser et al., 2020; Bevacqua et al., 2023). A balanced approach, grounded in decades of accumulated intellectual capital and methodical, rigorous approximations, is likely to foster such trust. This approach can enable a clear tracing of the causal chain leading to possible climate changes, allowing for interpretation and scrutiny in line with centuries-old scientific traditions. If successful, this strategy may eventually also narrow the gap between *episteme* and *techne* in climate modeling and deepen our understanding of the climate system's complexities through the investigation of models that integrate data-driven components. Such a convergence would mark a significant advance in both the science and practical application of climate modeling.

Code and data availability. The data and code needed to produce Fig. 1 is available at https://doi.org/10.22002/9dzs6-tmg69.

Author contributions. Tapio Schneider: Conceptualization, formal analysis, visualization, writing (original draft); Ruby Leung: Conceptualization, visualization, writing (editing and review); Robert Wills: formal analysis, visualization, writing (editing and review).

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

Competing interests. The authors declare no competing interests.

Acknowledgements. We thank Joern Callies and Ulrich Schumann for providing the data for Figure 3 and for their valuable discussions; Yi-Fan Chen, Raffaele Ferrari, Nadir Jeevanjee, and Thomas Müller for insightful comments on drafts; and Duan-Heng Chang for identifying a critical typo in a previous version of the analysis script for Figure 1.

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

References

Adler, R. F., Sapiano, M. R. P., Huffman, G. J., Wang, J.-J., Gu, G., Bolvin, D., Chiu, L., Schneider, U., Becker, A., Nelkin, E., Xie, P., Ferraro, R., and Shin, D.-B.: The Global Precipitation Climatology Project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation, Atmosphere, 9, https://doi.org/10.3390/atmos9040138, 2018.

Anber, U. M., Giangrande, S. E., Donner, L. J., and Jensen, M. P.: Updraft constraints on entrainment: insights from Amazonian deep convection, J. Atmos. Sci., 76, 2429–2442, https://doi.org/10.1175/JAS-D-18-0234.1, 2019.

Arakawa, A. and Schubert, W. H.: Interaction of a cumulus cloud ensemble with the large-scale environment. Part I, J. Atmos. Sci., 31, 674–701, 1974.

Arakawa, A. and Wu, C.-M.: A unified representation of deep moist convection in numerical modeling of the atmosphere: Part I, J. Atmos. Sci., 70, 1977–1992, https://doi.org/10.1175/JAS-D-12-0330.1, 2013.

Arakawa, A., Jung, J.-H., and Wu, C.-M.: Toward unification of the multiscale modeling of the atmosphere, Atmos. Chem. Phys., 11, 3731–3742, https://doi.org/10.5194/acp-11-3731-2011, 2011.

Balaji, V., Couvreux, F., Deshayes, J., Gautrais, J., Hourdin, F., and Rio, C.: Are general circulation models obsolete?, Proc. Natl. Acad. Sci., 119, e2202075 119, https://doi.org/10.1073/pnas.2202075119, 2022.

Bauer, P., Stevens, B., and Hazeleger, W.: A digital twin of Earth for the green transition, Nature Climate Change, 11, 80–83, https://doi.org/10.1038/s41558-021-00986-y, 2021.

Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., and Gentine, P.: Enforcing analytic constraints in neural networks emulating physical systems, Phys. Rev. Lett., 126, 098 302, https://doi.org/10.1103/PhysRevLett.126.098302, 2021.

Bevacqua, E., Suarez-Gutierrez, L., Jézéquel, A., Lehner, F., Vrac, M., Yiou, P., and Zscheischler, J.: Advancing research on compound weather and climate events via large ensemble model simulations, Nature Comm., 14, 2145, https://doi.org/10.1038/s41467-023-37847-5, 2023.

Bock, L., Lauer, A., Schlund, M., Barreiro, M., Bellouin, N., Jones, C., Meehl, G., Predoi, V., Roberts, M., and Eyring, V.: Quantifying progress across different CMIP phases with the ESMValTool, J. Geophys. Res. Atmos., 125, e2019JD032 321, https://doi.org/10.1029/2019JD032321, 2020.

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

Bony, S. and Dufresne, J. L.: Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models, Geophys. Res. Lett., 32, L20 806, https://doi.org/10.1029/2005GL023851, 2005.

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

Brenowitz, N. D., Beucler, T., Pritchard, M., and Bretherton, C. S.: Interpreting and stabilizing machine-learning parametrizations of convection, J. Atmos. Sci., 77, 4357–4375, https://doi.org/10.1175/JAS-D-20-0082.1, 2020.

Bretherton, C. S.: Insights into low-latitude cloud feedbacks from high-resolution models, Phil. Trans. R. Soc. Lond. A, 373, 20140 415, https://doi.org/10.1098/rsta.2014.0415, 2015.

Bretherton, C. S., Henn, B., Kwa, A., Brenowitz, N. D., Watt-Meyer, O., McGibbon, J., Perkins, W. A., Clark, S. K., and Harris, L.: Correcting coarse-grid weather and climate models by machine learning from global storm-resolving simulations, J. Adv. Model. Earth Sys., 14, e2021MS002 794, https://doi.org/10.1029/2021MS002794, 2022.

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

Brient, F. and Schneider, T.: Constraints on climate sensitivity from space-based measurements of low-cloud reflection, J. Climate, 29, 5821–5835, https://doi.org/10.1175/JCLI-D-15-0897.1, 2016.

Businger, J. A., Wyngaard, J. C., Izumi, Y., and Bradley, E. F.: Flux-profile relationships in the atmospheric surface layer, J. Atmos. Sci., 28, 181–189, 1971.

Caldwell, P. M., Zelinka, M. D., and Klein, S. A.: Evaluating emergent constraints on equilibrium climate sensitivity, J. Climate, 31, 3921–3942, https://doi.org/10.1175/JCLI-D-17-0631.1, 2018.

Callies, J., Ferrari, R., and Bühler, O.: Transition from geostrophic turbulence to inertia–gravity waves in the atmospheric energy spectrum, Proc. Natl. Acad. Sci., 111, 17 033–17 038, https://doi.org/10.1073/pnas.141077211, 2014.

Cess, R. D., Potter, G., Blanchet, J., Boer, G., Ghan, S., Kiehl, J., Le Treut, H., Li, Z.-X., Liang, X.-Z., Mitchell, J., et al.: Interpretation of cloud-climate feedback as produced by 14 atmospheric general circulation models, Science, 245, 513–516, 1989.

Chantry, M., Christensen, H., Dueben, P., and Palmer, T.: Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft AI, Phil. Trans. Roy. Soc. A, p. 20200083, https://doi.org/10.1098/rsta.2020.0083, 2021.

Charbonneau, A., Deck, K., and Schneider, T.: A physics-constrained neural differential equation for data-driven seasonal snowpack forecasting, Artificial Intelligence for the Earth Systems, 2023.

Cleary, E., Garbuno-Inigo, A., Lan, S., Schneider, T., and Stuart, A. M.: Calibrate, emulate, sample, J. Comp. Phys., 424, 109 716, 10.1016/j.jcp.2020.109716, 2021.

Cohen, Y., Lopez-Gomez, I., Jaruga, A., He, J., Kaul, C. M., and Schneider, T.: Unified entrainment and detrainment closures for extended eddy-diffusivity mass-flux schemes, J. Adv. Model. Earth Sys., 12, e2020MS002 162, https://doi.org/10.1029/2020MS002162, 2020.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2 (CESM2), J. Adv. Model. Earth Sys., 12, e2019MS001 916, https://doi.org/https://doi.org/10.1029/2019MS001916, e2019MS001916 2019MS001916, 2020.

de Rooy, W. C., Bechtold, P., Fröhlich, K., Hohenegger, C., Jonker, H., Mironov, D., Siebesma, A. P., Teixeira, J., and Yano, J.-I.: Entrainment and detrainment in cumulus convection: an overview, Quart. J. Roy. Meteor. Soc., 139, 1–19, https://doi.org/10.1002/qj.1959, 2013.

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., Fiore, A., Frankignoul, C., Fyfe, J. C., Horton, D. E., et al.: Insights from Earth system model initial-condition large ensembles and future prospects, Nature Climate Change, 10, 277–286, 2020.

Dewan, E. M.: Stratospheric wave spectra resembling turbulence, Science, 204, 832–835, 1979.

Dong, L., Leung, L. R., Lu, J., and Song, F.: Double-ITCZ as an emergent constraint for future precipitation over Mediterranean climate regions in the North Hemisphere, Geophysical Research Letters, 48, e2020GL091 569, 2021.

Donoho, D.: Data science at the singularity, arXiv:2310.00865v1, https://arxiv.org/pdf/2310.00865.pdf, 2023.

Dunbar, O. R. A., Garbuno-Inigo, A., Schneider, T., and Stuart, A. M.: Calibration and uncertainty quantification of convective parameters in an idealized GCM, J. Adv. Model. Earth Sys., 13, e2020MS002 454, https://doi.org/10.1029/2020MS002454, 2021.

Edwards, P. N.: A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming, MIT Press, 2010.

Fiedler, T., Pitman, A. J., Mackenzie, K., Wood, N., Jakob, C., and Perkins-Kirkpatrick, S. E.: Business risk and the emergence of climate analytics, Nature Climate Change, 11, 87–94, https://doi.org/10.1038/s41558-020-00984-6, 2021.

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

Firl, G. J. and Randall, D. A.: Fitting and analyzing LES using multiple trivariate Gaussians, J. Atmos. Sci., 72, 1094–1116, https://doi.org/10.1175/JAS-D-14-0192.1, 2015.

Foken, T.: 50 Years of the Monin-Obukhov similarity theory, Bound.-Layer Meteor., 119, 431–447, https://doi.org/10.1007/s10546-006-9048-6, 2006.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., and Yacalis, G.: Could machine learning break the convection parameterization deadlock?, Geophys. Res. Lett., 45, 5742–5751, https://doi.org/10.1029/2018GL078202, 2018.

Golaz, J.-C., Larson, V. E., and Cotton, W. R.: A PDF-based model for boundary layer clouds. Part I: Method and model description, J. Atmos. Sci., 59, 3540–3551, 2002.

Guo, H., Golaz, J.-C., Donner, L. J., Wyman, B., Zhao, M., and Ginoux, P.: CLUBB as a unified cloud parameterization: Opportunities and challenges, Geophys. Res. Lett., 42, 4540–4547, https://doi.org/10.1002/2015GL063672, 2015.

Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J. H., von Storch, J.-S., Brüggemann, N., Haak, H., and Stössel, A.: Max Planck Institute Earth System Model (MPI-ESM1.2) for the High-Resolution Model Intercomparison Project (HighResMIP), Geosci. Model Dev., 12, 3241–3281, https://doi.org/10.5194/gmd-12-3241-2019, 2019.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geosci. Model Dev., 9, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016, 2016.

Hall, A., Cox, P., Huntingford, C., and Klein, S.: Progressing emergent constraints on future climate change, Nature Climate Change, 4, 269–278, 2019.

Harlaß, J., Latif, M., and Park, W.: Improving climate model simulation of tropical Atlantic sea surface temperature: The importance of enhanced vertical atmosphere model resolution, Geophys. Res. Lett., 42, 2401–2408, https://doi.org/10.1002/2015GL063310, 2015.

Held, I. M.: The gap between simulation and understanding in climate modeling, Bull. Amer. Meteor. Soc., pp. 1609–1614, https://doi.org/10.1175/BAMS-86-11-1609, 2005.

Hohenegger, C., Korn, P., Linardakis, L., Redler, R., Schnur, R., Adamidis, P., Bao, J., Bastin, S., Behravesh, M., Bergemann, M., et al.: ICON-Sapphire: simulating the components of the Earth system and their interactions at kilometer and subkilometer scales, Geosci. Model Dev., 16, 779–811, 2023.

Hourdin, F., Williamson, D., Rio, C., Couvreux, F., Roehrig, R., Villefranque, N., Musat, I., Diallo, F. B., Fairhead, L., and Volodina, V.: Process-based climate model development harnessing machine learning: II. Model calibration from single column to global, J. Adv. Model. Earth Sys., 13, e2020MS002 225, https://doi.org/10.1029/2020MS002225, 2021.

Huang, D. Z., Schneider, T., and Stuart, A. M.: Iterated Kalman methodology for inverse problems, J. Comp. Phys., 463, 111 262, https://doi.org/10.1016/j.jcp.2022.111262, 2022.

Iglesias, M. A., Law, K. J. H., and Stuart, A. M.: Ensemble Kalman methods for inverse problems, Inverse Problems, 29, 045 001 (20pp), https://doi.org/10.1088/0266-5611/29/4/045001, 2013.

Intergovernmental Panel on Climate Change: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, New York, NY, 2021.

Irrgang, C., Boers, N., Sonnewald, M., Barnes, E. A., Kadow, C., Staneva, J., and Saynisch-Wagner, J.: Towards neural Earth system modelling by integrating artificial intelligence in Earth system science, Nat. Mach. Intell., pp. 667–674, https://doi.org/10.1038/s42256-021-00374-3, 2022.

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

Jeevanjee, N., Hassanzadeh, P., Hill, S., and Sheshadri, A.: A perspective on climate model hierarchies, J. Adv. Model. Earth Sys., 9, 1760–1771, https://doi.org/10.1002/2017MS001038, 2017.

Kawai, H., Yukimoto, S., Koshiro, T., Oshima, N., Tanaka, T., Yoshimura, H., and Nagasawa, R.: Significant improvement of cloud representation in the global climate model MRI-ESM2, Geosci. Model Dev., 12, 2875–2897, 2019.

Kennedy, M. C. and O'Hagan, A.: Bayesian calibration of computer models, J. Roy. Statist. Soc. B, 63, 425–464, https://doi.org/10.1111/1467-9868.00294, 2001.

Klaver, R., Haarsma, R., Vidale, P. L., and Hazeleger, W.: Effective resolution in high resolution global atmospheric models for climate studies, Atmos. Sci. Lett., 21, e952, 2020.

Klein, S. A. and Hall, A.: Emergent Constraints for Cloud Feedbacks, Curr. Clim. Change Rep., 1, 276–287, https://doi.org/10.1007/s40641-015-0027-1, 2015.

Knight, C. G., Knight, S. H. E., Massey, N., and Allen, M. R.: Association of parameter, software, and hardware variation with large-scale behavior across 57,000 climate models, Proc. Natl. Acad. Sci., 104, 12 259–12 264, https://doi.org/10.1073/pnas.0608144104, 2007.

Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., and Anandkumar, A.: Neural operator: Learning maps between function spaces, J. Mach. Learn. Res., 24, 1–97, 2023.

Kovachki, N. B. and Stuart, A. M.: Ensemble Kalman inversion: A derivative-free technique for machine learning tasks, Inverse Problems, 35, 095 005, 2019.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrol. Earth Syst. Sci., 22, 6005–6022, https://doi.org/10.5194/hess-22-6005-2018, 2018.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., and Nearing, G.: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, Hydrol. Earth Syst. Sci., 23, 5089–5110, https://doi.org/10.5194/hess-23-5089-2019, 2019.

Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., Shalev, G., and Matias, Y.: Caravan—A global community dataset for large-sample hydrology, Nature Scientific Data, 10, 61, 2023.

Lappen, C.-L. and Randall, D. A.: Toward a unified parameterization of the boundary layer and moist convection. Part I: A new type of mass-flux model, J. Atmos. Sci., 58, 2021–2036, 2001.

Large, W. G., McWilliams, J. C., and Doney, S. C.: Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization, Rev. Geophys., 32, 363–403, https://doi.org/10.1029/94RG01872, 1994.

Levine, M. E. and Stuart, A. M.: A framework for machine learning of model error in dynamical systems, Comm. Amer. Math. Soc., 2, 283–344, https://doi.org/10.1090/cams/10, 2022.

Lock, A. P., Brown, A. R., Bush, M. R., Martin, G. M., and Smith, R. N. B.: A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests, Mon. Wea. Rev., 128, 3187–3199, 2000.

Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., Manalo-Smith, N., and Wong, T.: Toward optimal closure of the Earth's top-of-atmosphere radiation budget, J. Climate, 22, 748–766, https://doi.org/10.1175/2008JCLI2637.1, 2009.

Lopez-Gomez, I., Cohen, Y., He, J., Jaruga, A., and Schneider, T.: A generalized mixing length closure for eddy-diffusivity mass-flux schemes of turbulence and convection, J. Adv. Model. Earth Sys., 12, e2020MS002 161, https://doi.org/10.1029/2020MS002161, 2020.

Lopez-Gomez, I., Christopoulos, C., Ervik, H. L., Dunbar, O. R. A., Cohen, Y., and Schneider, T.: Training physics-based machine-learning parameterizations with gradient-free ensemble Kalman methods, J. Adv. Model. Earth Sys., 14, e2022MS003 105, https://doi.org/10.1029/2022MS003105, 2022.

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

Lorenz, E. N.: Climatic predictability, in: The Physical Basis of Climate and Climate Modelling, edited by Bolin, B. and Coauthors, vol. 16 of *GARP Publication Series*, pp. 132–136, World Meteorological Organization, 1975.

Lucarini, V. and Chekroun, M. D.: Theoretical tools for understanding the climate crisis from Hasselmann's programme and beyond, Nature Rev. Phys., 2023.

Lucarini, V., Blender, R., Herbert, C., Ragone, F., Pascale, S., and Wouters, J.: Mathematical and physical ideas for climate science, Rev. Geophys., 52, 809–859, https://doi.org/10.1002/2013RG000446, 2014.

Luo, Y.: Terrestrial carbon-cycle feedback to climate warming, Annu. Rev. Ecol. Evol. Syst., 38, 683–712, 2007.

Majda, A. J., Timofeyev, I., and Vanden-Eijnden, E.: Systematic strategies for stochastic mode reduction in climate, J. Atmos. Sci., 60, 1705–1722, 2003.

Mellor, G. L. and Yamada, T.: Development of a turbulence closure model for geophysical fluid problems, Rev. Geophys., 20, 851–875, 1982.

Moshe, Z., Metzger, A., Elidan, G., Kratzert, F., Nevo, S., and El-Yaniv, R.: HydroNets: Leveraging river structure for hydrologic modeling, in: ICLR Workshop on AI for Earth Sciences, 2020.

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., and Gupta, H. V.: What Role Does Hydrological Science Play in the Age of Machine Learning?, Water Resources Res., 57, e2020WR028 091, https://doi.org/10.1029/2020WR028091, 2021.

Nelsen, N. H. and Stuart, A. M.: The random feature model for input-output maps between Banach spaces, SIAM J. Sci. Comput., 43, A3212–A3243, https://doi.org/10.1137/20M133957X, 2021.

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

Palmer, T. N.: A personal perspective on modelling the climate system, Proc. R. Soc. Lond. A, 472, 20150 772, https://doi.org/10.1098/rspa.2015.0772, 2016.

Parry, R.: *Episteme* and *Techne*, in: The Stanford Encyclopedia of Philosophy, edited by Zalta, E. N., Metaphysics Research Lab, Stanford University, 2021.

President's Council of Advisors on Science and Technology: Extreme weather risk in a changing climate: Enhancing prediction and protecting communities, Tech. rep., Executive Office of the President of the United States, Washington, D.C., https://www.whitehouse.gov/wp-content/uploads/2023/04/PCAST_Extreme-Weather-Report_April2023.pdf, 2023.

Randall, D. A.: Beyond deadlock, Geophys. Res. Lett., 40, 5970–5976, https://doi.org/10.1002/2013GL057998, 2013.

Randall, D. A., Khairoutdinov, M., Arakawa, A., and Grabowski, W.: Breaking the cloud parameterization deadlock, Bull. Amer. Meteor. Soc., 84, 1547–1564, https://doi.org/10.1175/BAMS-84-11-1547, 2003.

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

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, https://doi.org/10.1038/s41586-019-0912-1, 2019.

Roberts, M. J., Baker, A., Blockley, E. W., Calvert, D., Coward, A., Hewitt, H. T., Jackson, L. C., Kuhlbrodt, T., Mathiot, P., Roberts, C. D., Schiemann, R., Seddon, J., Vannière, B., and Vidale, P. L.: Description of the resolution hierarchy of the global coupled HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments, Geosci. Model Dev., 12, 4999–5028, https://doi.org/10.5194/gmd-12-4999-2019, 2019.

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

Romps, D. M.: A direct measure of entrainment, J. Atmos. Sci., 67, 1908–1927, https://doi.org/10.1175/2010JAS3371.1, 2010.

Russo, F.: Techno-Scientific Practices: An Informational Approach, Rowman & Littlefield Publishers, 2000.

Savre, J. and Herzog, M.: A general description of entrainment in buoyant cloudy plumes including the effects of mixing induced evaporation, J. Atmos. Sci., 76, 479–496, https://doi.org/10.1175/JAS-D-17-0326.1, 2019.

Schneider, T., Lan, S., Stuart, A., and Teixeira, J.: Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations, Geophys. Res. Lett., 44, 12 396–12 417, https://doi.org/10.1002/2017GL076101, 2017a.

Schneider, T., Teixeira, J., Bretherton, C. S., Brient, F., Pressel, K. G., Schär, C., and Siebesma, A. P.: Climate goals and computing the future of clouds, Nature Climate Change, 7, 3–5, https://doi.org/10.1038/nclimate3190, 2017b.

Schneider, T., Kaul, C. M., and Pressel, K. G.: Possible climate transitions from breakup of stratocumulus decks under greenhouse warming, Nature Geosci., 12, 163–167, 2019.

Schneider, T., Jeevanjee, N., and Socolow, R.: Accelerating progress in climate science, Physics Today, 74, 44–51, https://doi.org/10.1063/PT.3.4772, 2021a.

Schneider, T., Stuart, A. M., and Wu, J.: Learning stochastic closures using ensemble Kalman inversion, Trans. Math. Appl., 5, 1–31, https://doi.org/10.1093/imatrm/tnab003, 2021b.

Schneider, T., Behera, S., Boccaletti, G., Deser, C., Emanuel, K., Ferrari, R., Leung, L. R., Lin, N., Müller, T., Navarra, A., Ndiaye, O., Stuart, A., Tribbia, J., and Yamagata, T.: Harnessing AI and computing to advance climate modeling and prediction, Nature Climate Change, 13, 887–889, https://doi.org/10.1038/s41558-023-01769-3, 2023.

Schumann, U.: The horizontal spectrum of vertical velocities near the tropopause from global to gravity wave scales, J. Atmos. Sci., 76, 3847–3862, https://doi.org/10.1175/JAS-D-19-0160.1, 2019.

Shen, Z., Sridhar, A., Tan, Z., Jaruga, A., and Schneider, T.: A Library of Large-Eddy Simulations Forced by Global Climate Models, J. Adv. Model. Earth Sys., 14, e2021MS002 631, https://doi.org/10.1029/2021MS002631, 2022.

Skamarock, W. C.: Evaluating mesoscale NWP models using kinetic energy spectra, Mon. Wea. Rev., 132, 3019–3032, 2004.

Skamarock, W. C., Park, S.-H., Klemp, J. B., and Snyder, C.: Atmospheric kinetic energy spectra from global high-resolution nonhydrostatic simulations, J. Atmos. Sci., 71, 4369–4381, https://doi.org/10.1175/JAS-D-14-0114.1, 2014.

Slingo, J., Bates, P., Bauer, P., Belcher, S., Palmer, T., Stephens, G., Stevens, B., Stocker, T., and Teutsch, G.: Ambitious partnership needed for reliable climate prediction, Nature Climate Change, 12, 499–503, https://doi.org/10.1038/s41558-022-01384-8, 2022.

Smalley, M. A., Lebsock, M. D., and Teixeira, J.: Quantifying the impact of vertical resolution on the representation of marine boundary layer physics for global-scale models, Mon. Wea. Rev., 2023.

Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D. J., Kettleborough, J. A., Knight, S., Martin, A., Murphy, J. M., Piani, C., Sexton, D., Smith, L. A., Spicer, R. A., Thorpe, A. J., and Allen, M. R.: Uncertainty in predictions of the climate response to rising levels of greenhouse gases, Nature, 433, 403–406, 2005.

Stephens, G. L.: Cloud feedbacks in the climate system: A critical review, J. Climate, 18, 237–273, https://doi.org/10.1175/JCLI-3243.1, 2005.

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., Düben, P., Judt, F., Khairoutdinov, M., Klocke, D., Kodama, C., Kornblueh, L., Lin, S.-J., Neumann, P., Putman, W. M., Röber, N., Shibuya, R., Vanniere, B., Vidale, P. L., Wedi, N., and Zhou, L.: DYAMOND: The DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains, Prog. Earth Planet. Sci., 6, 61, https://doi.org/10.1186/s40645-019-0304-z, 2019.

https://doi.org/10.5194/egusphere-2024-20
Preprint. Discussion started: 24 January 2024

Subramanian, A., Ummenhofer, C., Giannini, A., Holland, M., Legg, S., Mahadevan, A., Perovich, D., Small, J., Teixeira, J., and Thompson, L.: Translating process understanding to improve climate models. A US CLIVAR White Paper., Report 2016-3, US CLIVAR, https://doi.org/10.5065/D63X851Q, 2016.

Tan, Z., Kaul, C. M., Pressel, K. G., Cohen, Y., Schneider, T., and Teixeira, J.: An extended eddy-diffusivity mass-flux scheme for unified representation of subgrid-scale turbulence and convection, J. Adv. Model. Earth Sys., 10, 770–800, https://doi.org/10.1002/2017MS001162, 2018.

Thuburn, J., Weller, H., Vallis, G. K., Beare, R. J., and Whitall, M.: A framework for convection and boundary layer parameterization derived from conditional filtering, J. Atmos. Sci., 75, 965–981, https://doi.org/10.1175/JAS-D-17-0130.1, 2018.

Tian, B. and Dong, X.: The double-ITCZ bias in CMIP3, CMIP5, and CMIP6 models based on annual mean precipitation, Geophys. Res. Lett., 47, e2020GL087 232, https://doi.org/10.1029/2020GL087232, 2020.

Vallis, G. K.: Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation, Cambridge University Press, Cambridge, UK, 2006.

VanZandt, T. E.: A universal spectrum of buoyancy waves in the atmosphere, Geophys. Res. Lett., 9, 575–578, 1982.

Vial, J., Dufresne, J.-L., and Bony, S.: On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates, Clim. Dyn., 41, 3339–3362, https://doi.org/10.1007/s00382-013-1725-9, 2013.

Watson-Parris, D.: Machine learning for weather and climate are worlds apart, Phil. Trans. R. Soc. A, 379, 20200 098, https://doi.org/0.1098/rsta.2020.0098, 2021.

Wedi, N. P.: Increasing horizontal resolution in numerical weather prediction and climate simulations: illusion or panacea?, Phil. Trans. R. Soc. A, 372, 20130 289, https://doi.org/10.1098/rsta.2013.0289, 2014.

Wouters, J. and Lucarini, V.: Multi-level dynamical systems: Connecting the Ruelle response theory and the Mori-Zwanzig approach, J. Stat. Phys., 151, 850–860, https://doi.org/10.1007/s10955-013-0726-8, 2013.

Wouters, J., Dolaptchiev, S. I., and Lucarini, V.: Parameterization of stochastic multiscale triads, Nonlin. Processes Geophys., 23, 435–445, https://doi.org/10.5194/npg-23-435-2016, 2016.

Wu, J., Levine, M., Shen, Z., Schneider, T., and Stuart, A. M.: Learning about structural errors in models of complex dynamical systems, J. Comp. Phys., in preparation., 2024.

Yeo, K. and Romps, D. M.: Measurement of convective entrainment using Lagrangian particles, J. Atmos. Sci., 70, 266–277, https://doi.org/10.1175/JAS-D-12-0144.1, 2013.

Yu, S., Hannah, W. M., Peng, L., Bhouri, M. A., Gupta, R., Lin, J., Lütjens, B., Will, J. C., Beucler, T., Harrop, B. E., et al.: ClimSim: An open large-scale dataset for training high-resolution physics emulators in hybrid multi-scale climate simulators, arXiv preprint arXiv:2306.08754, 2023.

Yuval, J. and O'Gorman, P. A.: Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions, Nature Comm., 11, 1–10, https://doi.org/10.1038/s41467-020-17142-3, 2020.

Zanna, L. and Bolton, T.: Data-driven equation discovery of ocean mesoscale closures, Geophys. Res. Lett., 47, e2020GL088 376, https://doi.org/10.1029/2020GL088376, 2020.

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S. A., and Taylor, K. E.: Causes of higher climate sensitivity in CMIP6 models, Geophys. Res. Lett., 47, e2019GL085 782, https://doi.org/10.1029/2019GL085782, 2020.

Zhou, W., Leung, L. R., and Lu, J.: Linking large-scale double-ITCZ bias to local-scale drizzling bias in climate models, Journal of Climate, 35, 7965–7979, 2022.