



Are general circulation models obsolete?

V. Balaji^{a,b,1}, Fleur Couvreur^c, Julie Deshayes^d, Jacques Gautrais^e, Frédéric Hourdin^f, and Catherine Rio^g

Edited by Andrew Gettelman, National Center for Atmospheric Research; received April 1, 2022; accepted September 9, 2022 by Editorial Board Member Akkihebbal R. Ravishankara

Traditional general circulation models, or GCMs—that is, three-dimensional dynamical models with unresolved terms represented in equations with tunable parameters—have been a mainstay of climate research for several decades, and some of the pioneering studies have recently been recognized by a Nobel prize in Physics. Yet, there is considerable debate around their continuing role in the future. Frequently mentioned as limitations of GCMs are the structural error and uncertainty across models with different representations of unresolved scales and the fact that the models are tuned to reproduce certain aspects of the observed Earth. We consider these shortcomings in the context of a future generation of models that may address these issues through substantially higher resolution and detail, or through the use of machine learning techniques to match them better to observations, theory, and process models. It is our contention that calibration, far from being a weakness of models, is an essential element in the simulation of complex systems, and contributes to our understanding of their inner workings. Models can be calibrated to reveal both fine-scale detail and the global response to external perturbations. New methods enable us to articulate and improve the connections between the different levels of abstract representation of climate processes, and our understanding resides in an entire hierarchy of models where GCMs will continue to play a central role for the foreseeable future.

climate modeling | machine learning | model calibration | model hierarchy

The general circulation model, or GCM, is a mainstay of research into the evolving state of the Earth system over a range of timescales. The term dates back to the very origin of numerical simulation of the atmosphere (e.g., refs. 1 and 2). The equations governing the general circulation of fluids on a spinning sphere use the basic Navier–Stokes equations, whose form specialized for the planetary circulation was first formulated at the turn of the 20th century (e.g., refs. 3 and 4). However, closed-form solutions are not readily available, and their use as research and prediction tools had to await the advent of numerical solution in the 1950s (5).

The 2021 Nobel prizes in Physics honor some of the work done with GCMs. The first formal global warning of anthropogenic climate change, the Charney Report (6), was substantially based on the pioneering work of Syukuro Manabe, who confirmed 19th century speculations on the warming effect of adding CO₂. While normally attributed to Tyndall and Arrhenius, the earlier work of Eunice Foote has recently come to light (7). She, in fact, presciently remarked, “An atmosphere of that gas would give to our earth a high temperature” (8). While Foote and others were talking principally about the radiative effects of CO₂, it was Manabe

and others who included dynamical considerations, the transport of heat vertically through convection (9), as well as from the equator poleward through atmospheric and oceanic circulation (e.g., ref. 10). Besides, GCMs also play a central role in the work of another of the 2021 winners, Klaus Hasselmann, who laid the groundwork for the statistical methods behind the field of detection and attribution of climate change (e.g., ref. 11). The detection of climate change requires extracting the signal of forced response in simulations from natural variability, and the attribution of it to external climate forcing agents, such as CO₂ emissions, again requires counterfactual runs of a GCM where that particular forcing is absent.

It may seem an odd juncture, when a Nobel prize has just been awarded for GCM-based work, to speculate on the obsolescence of the GCM. However, there has been a considerable body of literature, for a while, arguing that the limitations of GCMs require a major overhaul for further progress in climate modeling. It has been noted (see, e.g., ref. 12) that the bounds of uncertainty on equilibrium climate sensitivity (ECS: the asymptotic response of a model climate to a doubling of CO₂ concentration) has not significantly diminished since the Charney Report (6). Furthermore, a systematic synthesis of multiple lines of evidence to constrain ECS in ref. 13 indicates, in several places, a diminishing role for GCMs relative to other sources of information. Some have taken a leap from here to assert that the entire project of parameterization—the discovery of parsimonious representation through insight or mathematical methods—may have no future (e.g., refs. 14 and 15), and that large-scale computation is the way forward.

It is perhaps no accident that this debate takes place at a particular inflection point in the history of computing (16), where it is now possible to marshal and extract information from data at an unprecedented scale, the era of big data and machine learning (ML). These methods have led to some spectacular successes in various fields: AlphaFold, for

Author affiliations: ^aCooperative Institute for Modeling the Earth System, Princeton University, NJ 08544; ^bLaboratoire des Sciences du Climat et de l'Environnement, Le Commissariat à l'Énergie Atomique et aux Énergies Alternatives, 91191 Gif-sur-Yvette, France; ^cCentre National de Recherches Météorologiques, University of Toulouse, Météo-France, CNRS, 31057 Toulouse Cedex, France; ^dSorbonne Universités-CNRS-Institut de recherche pour le développement (IRD) - Muséum National d'Histoire Naturelle (MNHN), Laboratory of Oceanography and Climate: Experiments and Numerical Approaches (LOCEAN), 75005 Paris, France; ^eCentre de Recherches sur la Cognition Animale, Centre de Biologie Intégrative, Université de Toulouse, CNRS, Université Paul Sabatier (UPS), 31062 Toulouse, France and; ^fLaboratoire de Météorologie Dynamique - Institut Pierre Simon Laplace (LMD-IPSL), Sorbonne University, CNRS, 75005 Paris, France

Author contributions: F.C. analyzed data; and V.B., F.C., J.D., J.G., F.H., and C.R. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. A.G. is a guest editor invited by the Editorial Board.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution License 4.0 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).

¹To whom correspondence may be addressed. Email: balaji@princeton.edu.

Published November 14, 2022.

example, can decipher the structure of complex molecules directly from data (17). This has led to speculation that we might have entered the era of “post-theory science” (18). This is a fierce debate in many fields: whether large-scale structure emerges directly from the addition of detail and data, and where the limits of models built from data might lie. We explore this debate here in the context of the modeling of climate, a complex system undergoing slow but inexorable global changes, but where the details matter as well.

The debate poses questions that resonate across all fields of science that use large-scale data and computation as a pillar of the scientific method, alongside theory and observations. In the discussion at the end, Section 6, we will point out certain parallels, particularly with the debate in neuroscience surrounding the Human Brain Project (HBP). Leading up that discussion, the paper will begin with an account of the structure of the GCM from the time of Manabe’s pioneering studies to the present day (Section 1), followed by an analysis of some aspects of climate modeling which have exposed GCMs to criticism (Section 2), interrogating the role of model resolution (Section 3), model calibration (Section 4), and the generation of counterfactuals (Section 5). It is our assertion that parameterized GCMs can expect to get a new lease of life at this moment through sophisticated approaches to model calibration based on methods borrowed from ML. We begin with our account of the GCM.

1. The Structure of the GCM, from Manabe to Present Day

The general circulation of the atmosphere and ocean can be described by equations of fluid flow and thermal energy transport and exchange. The numerical solution of a discretized form of these coupled partial differential equations can be written in the form

$$\frac{\partial \mathbf{x}}{\partial t} = R(\mathbf{x}) + U(\mathbf{x}) + P(\mathbf{x}) + F, \quad [1]$$

where \mathbf{x} is a state vector consisting of mass, momentum, and energy associated with fluid elements, as well as other quantities that can contribute to changes in state, which could include various phases of water in the atmosphere, salinity in the ocean, and trace elements, CO₂, methane, dust, and other species, of natural or human origin. $R(\mathbf{x}) + U(\mathbf{x})$ represents the dynamics, the Navier–Stokes equation. Because of the discretized nature of the numerical model, only part of the fluid motions are explicitly resolved (R). The unresolved (U) fluid motions (with scale smaller than one or several grid points) are represented in the form of a closure, that is, a representation of subgridscale dynamics in terms of resolved-scale state variables. P represents other processes that contribute to the thermodynamics; these can include diabatic processes associated with phase changes of water in the atmosphere, leading to the formation of clouds and rain, and the influence of solar radiation, and the equation of state of a complex fluid with many constituents. U and P are often collectively referred to as the physics. Finally, F represents the terms that are considered external (not simulated by the model) influences on the system, known as forcings; these can include solar radiation, volcanoes, anthropogenic emissions of CO₂, geothermal heating and other radiatively active quantities, or particulate

aerosols that can play a role in cloud formation. Both atmosphere and ocean are extremely shallow compared to Earth’s radius R ($H/R \approx 10^{-3}$, where H is the fluid height), so that the numerical treatment of vertical dynamics is generally different from the horizontal.

In ref. 9, Manabe and Wetherald considered the time-asymptotic balance between the destabilization of an atmospheric column by radiation (warming at the surface, cooling aloft) and stabilization by convection, which transports heat vertically. Radiative–convective equilibrium (RCE) mediated by water vapor and CO₂ yields the basic behavior of global warming in a single atmospheric column. Later work (e.g., refs. 10 and 19) extended it to include horizontal transport of heat as well, from the equator to the poles, confirming the single-column result. To give a sense of the computational size, a typical GCM resolution today is about 50 km or less, compared to 500 km in ref. 10, and vertical resolutions have increased by a comparable factor; at this resolution, the required temporal resolution is measured in minutes. At these scales, the spatial grid may encompass 2×10^7 points, and a simulation of 100 y in length requires 3×10^6 time steps. In the decades since the pioneering work of Manabe, considerable ingenuity has gone into creating fast and accurate solutions to the dynamics and increasing its resolution.

An even more significant site of creativity has been in the physics ($U(\mathbf{x}) + P(\mathbf{x})$) which now encompasses myriad processes in the atmosphere and ocean contributing to the transport of heat and other quantities; these include processes associated with clouds and precipitation, mixing by unresolved motions in the ocean, and an increasingly sophisticated treatment of the terrestrial and marine biosphere, with the contribution of trace elements and aerosols, whose natural and anthropogenic emissions also play a role in understanding the response of the climate system to changes in F . The structure of the GCM (Fig. 1) now consists of a number of parameterizations representing individual processes and feedbacks,

$$U(\mathbf{x}) + P(\mathbf{x}) = \sum_p \mathcal{M}(\mathbf{x}, \lambda_p), \quad [2]$$

where λ_p represents a set of parameters that may be empirically set through comparison with observations or theory. Typically, these are bulk quantities, representative

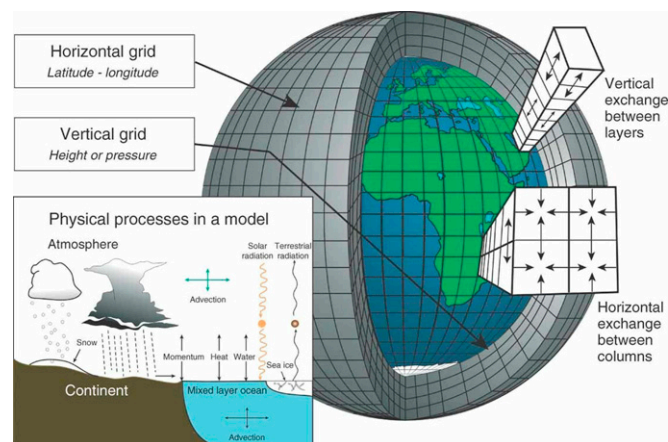


Fig. 1. Figure 2 from ref. 20: the classic structure of a GCM. A similar column structure is used in the ocean as well. Reprinted with permission from ref. 20.

values at the resolution of the model discretization. The climate is a multiscale system, encompassing processes from microscales (cloud droplet formation, stomatal exchange) to megascales (orbital changes, plate tectonics). Transport of heat and CO₂ into the ocean abyss occurs in the planetary scale on times measured in millennia.

We introduce a vocabulary for problems addressed with numerical models, where \mathbf{x} in Eq. 1 represents the state of the Earth system, and for identifying sources of error and uncertainty in predictions from these models.

- Note, first, that Eq. 1 represents the time trajectory of \mathbf{x} . The trajectory itself is the weather. The system is chaotic (21), which represents the first source of uncertainty. Numerical weather prediction (NWP) has improved over decades with better models, observations, and the techniques of data assimilation, which constrain trajectories to stay close to observations in a least-squares sense (22).
- The climate is the set of preferred states of the system, its attractors, discovered by running trajectories for a long time, and averaging over the weather. If F is held constant, the climate should be stationary, and the fluctuations around that state are its internal variability.
- The forcing F is time varying, with both natural and anthropogenic contributions, and the climate change problem consists in separating the forced response from internal variability. For purposes of policy related to the climate emergency, this constitutes studying the response to scenarios (trajectories of F ; see, e.g., ref. 23).
- Abrupt transitions, or tipping points, are changes in climate which are very fast relative to the rate of change of F . For instance, orbital changes are known to lead to long-term changes in climate, such as the glaciation-deglaciation cycles. Yet, the climate record contains transitions such as Heinrich events (24, 25), a rapid collapse and resumption of the planetary-scale meridional overturning circulation (MOC). We shall return several times to the MOC in this article as a canonical problem of concern for climate. The long intrinsic timescales of the forces shaping the MOC pose a unique set of challenges. Studies of individual processes (e.g., ref. 26) also show that the system may be capable of abrupt (relative to the rate of change of F) transitions.
- Parametric uncertainty arises from imperfect constraints, either from observations or theory, on λ_p . Structural uncertainty and structural error arise when no values of λ_p can fit the known constraints, leading to the conclusion that the form of the equations in \mathcal{M} could be improved. (It is, of course, possible to formulate structural uncertainty as parametric, by having a parameter that chooses one equation structure over another!) It has been noted that averaging over different GCMs (individual formulations of R , U , and P) can disguise structural error given statistical independence among GCMs (e.g., ref. 27). However, some biases remain common and persistent across many generations of models, suggesting gaps in understanding, or epistemic uncertainty (28).

For the purposes of this Perspective, we define the GCM as the tool used to study the response of the climate system to changes in F . The tool is broadly similar to (and, in many cases, is based on the same model code as) the

models used for NWP. In both cases, we follow trajectories of \mathbf{x} in time from a specified initial state. For weather, the trajectory itself is the solution, while, for climate, we are interested in the attractors of the landscape where the trajectories lie. Besides chaotic uncertainty, we must also contend with the structural and parametric uncertainty associated with the physics terms U and P . Finally, F and its time rate of change (usually ignored for weather, although not for climate) is also uncertain, as emissions trajectories are unpredictable (29). For timescales of interest to address the climate emergency, including the possibility of abrupt transitions, we generally need to run simulations of at least $\mathcal{O}(100)$ simulated years (SY) in length. The exploration of the different sources of uncertainty will require sampling at least $\mathcal{O}(100)$ model settings, as we shall show below in Section 5. A GCM is defined, for the purpose of this study, as an engine for simulating the Earth system, capable of running 100 instances of 100 SY in a reasonable amount of time and given adequate computing resources. The same tool may be configured for other purposes (e.g., process studies on a limited domain or in highly idealized settings, or higher resolution but shorter duration), but these constitute the minimum requirements for studying the response to changes in F .

2. What Ails the GCM?

There are several diagnoses of the weaknesses of GCMs. There is, first, the argument that the column abstraction breaks down in the presence of large-scale organization: the mesoscale organization of cloud systems in the atmosphere (30). A similar argument can be made for ocean dynamics as well, where mesoscale turbulence in the form of persistent eddies is able to deposit energy and momentum away from the source (e.g., ref. 31), in the form of Agulhas rings, for example (32). Such nonlocal effects of unresolved terms call into question the structure of the GCM that has been used since Manabe's pioneering calculations. Some aspects of the climate system have resisted efforts at representation in models, with stubborn biases and uncertainties. This has led some to question whether such processes are parameterizable at all (e.g., ref. 15). Following this line of thinking, it is now often contended that nothing short of resolving finer-scale motions, coupled with assimilation of present-day observations to control model biases, will, in fact, address these shortcomings, and that a future generation of models will address these issues through substantially higher resolution.

A second criticism of GCMs is around the consideration of the "tuning," or calibration, of climate models. As noted above, unresolved physics ($U + P$) is represented using equations with parameters constrained within some range by observations or theory. The coupled system is then further subjected to global constraints such as top-of-atmosphere energy balance (33). The fact that the models are tuned to reproduce some features of the observed planet is, in some quarters, viewed as rendering the results suspect.

Finally, GCMs are now numerous [122 models from 44 institutions at the time of counting (34)]. Viewed as an ensemble of simulations, they embody considerable

structural or epistemic uncertainty, in the form of differing representations of the unresolved scales. The GCMs are not all statistically, or in terms of model code, independent of each other (28), and different evaluation metrics yield different, and contradictory, measures of model quality. The uncertainty bounds have, if anything, increased between the last two climate model assessment cycles, and the recently published Intergovernmental Panel on Climate Change *Sixth Assessment Report* (IPCC AR6) (35) notes that many models now produce ECS values outside the assessed “very likely” range, leading to an enhanced role for emulators (e.g., ref. 36), as discussed in Section 5. Models used in service of decision-making and policy, including those used in the recently concluded IPCC AR6, rely on emulators allowing for rapid exploration of multiple future scenarios or, through the use of statistical techniques, for the correction of biases. Furthermore, it is increasingly noted that models pegged to present-day climate do not do a good job of representing the climate fluctuations of the past, including past warm climates that may hold lessons for the climate emergency (e.g., ref. 37). We address these concerns in turn.

3. What Resolution Is Enough?

We begin with the question of resolution. As noted above, one suggested remedy for the weakness of GCMs is to increase the resolution until some of the nonlocal phenomena alluded to above are, in fact, resolved (ref. 38, for instance). The computational expense of such a model would require a substantial boost to computing capability (39).

Atmospheric and ocean dynamics fit within the broad contours of geophysical turbulence. At very large (planetary) scales, this looks like two-dimensional (2D) turbulence, known to have an energy spectrum with a k^{-3} spectrum. At smaller scales, baroclinicity starts to play a role, and 3D turbulence with a $k^{-5/3}$ spectrum. This is, in fact, observed in the atmosphere, seen, for example, in ref. 40. Similar spectra are observed in ocean turbulence, as well (41, 42). The key feature to underline for this discussion is that the 3D energy cascade continues all the way down to molecular scales. There is no fundamental scale separation in turbulence. Any truncation applied in order to create a discrete representation for numerical purposes is an arbitrary one, usually constrained by the available computing power.

We can look beyond turbulent energy spectra to the specificities of certain dynamical phenomena. In the atmosphere, one of the key phenomena of interest is moist convection. The pioneering work of Rayleigh and others shows that fluids heated from below will overturn, with the overturning motion in the form of “cells” that roughly scale with the height of the convecting fluid. In the case of atmospheric convection, this includes deep convection, which roughly scales with the height of the tropopause (~ 10 km; see, e.g., ref. 43), such as thunderstorms or tropical cyclones, and shallow convection within the planetary boundary layer (depth ~ 1 km), which can take many forms, including large stratocumulus decks thousands of kilometers in extent, playing a significant role in the planetary albedo and heat balance. GCMs have traditionally tried to represent these in parameterizations of the vertical transport of momentum,

heat, and moisture, as well as other tracers, by convection. Subgrid closures of deep convective processes can be based on an assumption of quasi-equilibrium between synoptic-scale destabilization of the column and the stabilization by convection (e.g., ref. 44, and its descendants). The slow rate of advances in these methods (45) has led to efforts where parameterizations are replaced with embedded cloud-resolving models (CRMs), known as “superparameterizations” (46), or, going even further, replacing shallow convection as well, with large-eddy simulation (LES) models (47). It is not clear whether, since their inception, this class of models have justified their extreme computational expense in terms of improved climate simulation.

Many aspects of clouds, such as the representation of moist convection, require us to invoke microphysical processes involved in the condensation of water vapor and the formation of falling hydrometeors. But the motions themselves can be captured in nonhydrostatic models. Limited area modeling of deep convection dates back to the 1970s, and advances in computing capacity in the intervening decades make global CRMs (GCRMs, also known as global cloud system resolving models, global convection-permitting models, global storm-resolving models, etc.; see ref. 48 for a defense of the term despite not quite resolving cloud dynamics) within reach. The DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains (DYAMOND) experiment (49) is a recent systematic comparison across multiple GCRMs. These models typically have horizontal resolutions in the range of 1 km to 5 km, considered sufficient to capture mesoscale convective organization and, at least marginally, resolve individual convective events. Aspects of convective organization, such as the formation of gust fronts with downstream surface density currents (50), remain below the resolution of GCRMs, and must be parameterized, as they are responsible for the initiation of new convective events. Boundary layer convection is responsible for shallow clouds, and requires at least $10\times$ higher resolution; this is, in fact, one of the largest sources of uncertainty in the current generation of models (51). This class of clouds will not be resolved by kilometer-scale models. And, of course, water as vapor or condensate is radiatively active, and particulates play a role in cloud formation, as well. Surface exchanges at very fine scale mediate aerosol emissions (e.g., ref. 52). Many of these aerosol-radiation-cloud processes take place at micrometer scale and will be forever outside any conceivable resolution of a numerical model on any known computational technology in the literature today.

CRMs (and their global incarnation, GCRMs) and LES models of the turbulent and cloudy boundary layer are widely used to study processes that cannot be resolved in GCMs, but perhaps we can learn from them, to inform the development of parameterizations. Such studies are usually mediated by single-column models (see, e.g., ref. 53). Such studies often show that the disparity between different CRM formulations remains comparable to those between GCMs. In a set of RCE (the same problem treated in a column by ref. 8) comparisons across both CRMs and GCMs (54), the uncertainty spreads across CRMs and GCMs with parameterized convection were quite comparable (55). In the first extensive comparison of GCRMs (necessarily

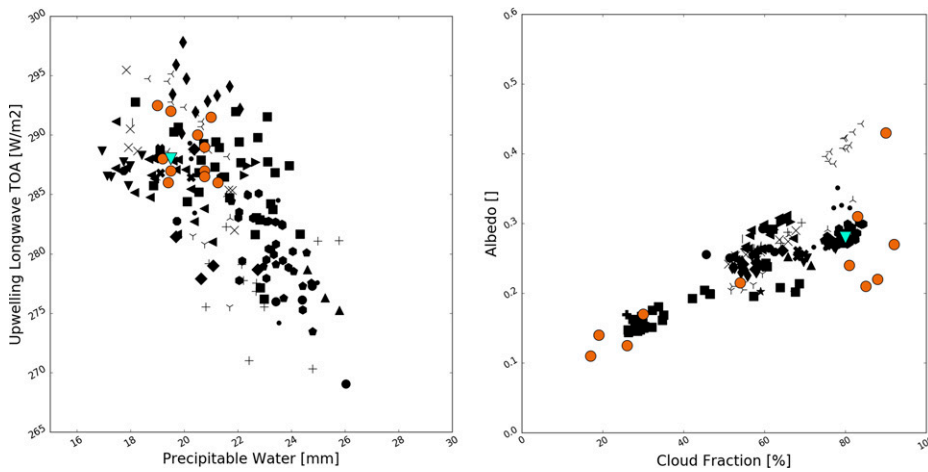


Fig. 2. Comparison of observable properties of cloud fields (August monthly mean outgoing longwave radiation at top of atmosphere (TOA), precipitable water, albedo, and cloud fraction) averaged over the Southern Atlantic (15 °W to 10 °E; 18 °S to 5 °S) in observations (green triangle), DYAMOND GCRMs (orange circles, from ref. 57) and in CMIP6 GCMs (black symbols). The added resolution in GCRMs reduces structural uncertainty in some respects but not others.

short runs, only 40 d compared to the GCM timescales outlined in Section 1) (49), there was considerable inter-model variability (56). In Fig. 2, we show a comparison of the relationship between precipitable water and outgoing top-of-atmosphere longwave radiation and albedo and low cloud cover from ref. 56 with the same quantities from the 6th Coupled Model Intercomparison Project (CMIP6) experiment. While the CRMs are closer to observations for the precipitable water and outgoing longwave radiation, the spread in low cloud cover and albedo is just as wide in GCRMs as in GCMs. The model spread in DYAMOND has been traced to differences in the treatment of boundary layer convection (57).

LES structural uncertainty tells a similar story: Comparisons across LES models of the boundary layer with or without clouds still show some reduction in spread in certain cases (58), but considerable spread in response to changes in configuration in specific aspects such as microphysics (59), turbulence closures (60, 61) and numerics (62).

Similarly, the representation of oceanic fronts, eddies, and currents is markedly different depending on the type of model that is used to represent them, whether turbulent mixing by eddies is resolved or parameterized. The ocean depth varies widely between pelagic zones and the abyss, and, consequently, the Rossby deformation radius—the length scale representing when mesoscale eddy mixing is significant—varies from $\mathcal{O}(100)$ km in the equatorial open ocean to $\mathcal{O}(1)$ km or less near coastlines and toward the poles (63). Attempts to build a “scale-aware” parameterization of eddy mixing (e.g., refs. 64 and 65) must contend with these variations.

Other features such as vegetation, not discussed in this article, also exhibit heterogeneity at any conceivable resolution. While there is no doubt that, in general, increasing resolution reduces the number of processes needing to be parameterized (although placing increasing resolution demands on observations as well), and, in many cases, increases the fidelity of simulations, particularly in the short term, in the climate context, this must be weighed against the expense, which limits the simulation duration and consequently the ability of models to realize the aspects that regulate the climate on long timescales. The absence of any target resolution where one could argue that everything of interest to climate is resolved implies that these trade-offs are always with us.

4. Are There “Untuned” Models?

The question of the “tuning,” or calibration, of GCMs has been a fact of life since their inception, often not clearly described in the literature, until recent efforts to document the role of tuning in model development (33, 66). On the scientific level, these attempts highlight the central role of tuning in modeling, and open up new avenues in the use of automatic calibration techniques from ML.

In broad terms, we can define the process of tuning as one of finding suitable values of λ_p in Eq. 2 that best fit observations or theory, identifying it as an intrinsic and universal aspect of model development. We seek to achieve fidelity to each process \mathcal{M} in the model, as well as respecting global constraints across the coupled system: for example, conserving energy and mass, including of individual agents in the climate system, such as water. The global constraints must be applied when coupling models at any resolution. Tuning is thus a multistep process, where individual parameterizations \mathcal{M} are first tuned separately to within a desired tolerance, but which may then be refined in a second stage after the coupled model is built. Thus, tuning, which is often seen as an optimization of a loss function, may be redefined as identifying the subspace of parameters compatible with a number of constraints. While the procedures can be onerous, the process of calibration is central to model development and the way teams learn how parts of the coupled system respond to changes in others. The coupled system can yield surprises: In one example [the National Oceanic and Atmospheric Administration’s Geophysical Fluid Dynamics Laboratory (NOAA/GFDL) model (67, 68)], the coupled system had an ECS higher than was expected during the development of individual components. Thus, model calibration is not a weakness of models; it, in fact, holds the key to how model developers learn how their model behaves, and, consequently, how the Earth system regulates itself.

Historically, the tuning of models has been found to be expensive, if one uses the GCM itself as the forward model. In particular, key circulation features such as the MOC may be sensitive to tuning in ways that reveal themselves after simulations of $\mathcal{O}(100)$ SY (67). Tuning such models “by hand” can be an inefficient scattershot exploration of a small amount of the possible space of parametric uncertainty. This is, perhaps, one reason why some observers hold tuning in low regard. There is also the risk of tuning to a state

Example of iterative reduction of parameter space

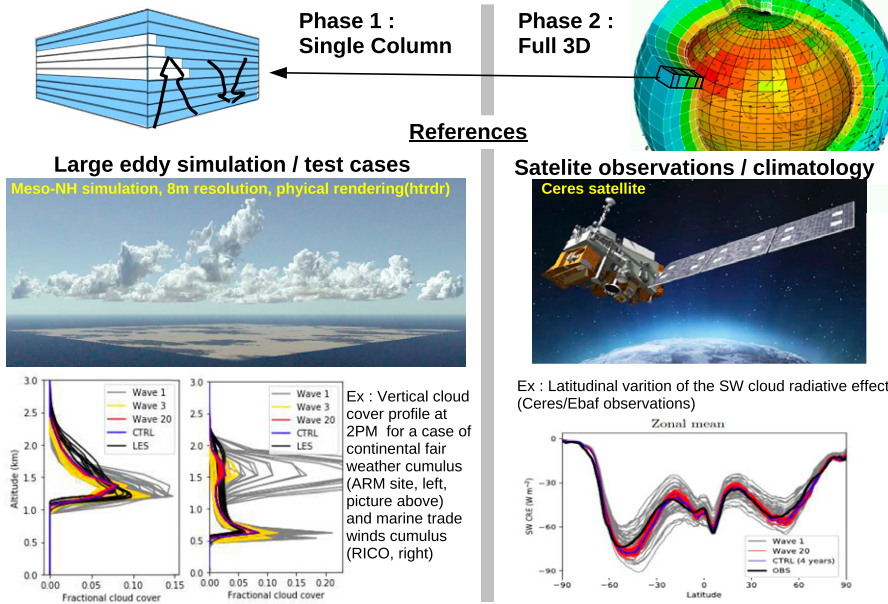


Fig. 3. A schematic of the use of LES models for parameter estimation, adapted from ref. 72. LES models of different observed cloud regimes are used to build emulators to estimate parameters governing the behavior of column-averaged quantities. Successive waves of HM deliver the minimal parameter space consistent with the LES data, the NROY space. The effect of clouds on absorbed shortwave radiation is reduced by a comparable amount to the previous laborious tuning “by hand” using the same target. ARM, atmospheric radiation measurement; RICO, Rain In Cumulus over Ocean experiment.

containing compensating errors that satisfy the constraints but for the wrong reasons.

As noted earlier, computing technology, at the present time, favors the algorithms of ML, the ability to add or extract patterns seen in large datasets. Tuning is a constant concern in the construction and optimization of methods, such as choosing the width, depth, and structure of a neural net, aspects often referred to as hyperparameters, as they are of the network, not of the process being emulated. They are chosen to meet the requirements of fidelity against the training dataset. Some data are often withheld to guard against overfitting to noisy data, and the withheld (“out of sample”) data can be used to validate the result as generalizable to novel situations not seen during training. “Physics-informed” ML (69), where the loss function can be made to penalize violations of global constraints such as conservation laws, also has a parallel to the tuning process described above.

In the HighTune project, an atmospheric boundary layer convection scheme based on the eddy-diffusivity mass-flux (EDMF) approach (70, 71) is calibrated to match results from an LES simulation (72, 73). The EDMF scheme, which has both upgradient (organizing) and downgradient (dissipative) components, is structured for problems such as cloudy convection and turbulence, but has parameters that must be empirically determined in a variety of boundary layer regimes. The process is illustrated in Fig. 3. A variety of cloud regimes is simulated in an LES constrained by observations; the LES serves as the “truth” for tuning. A single-column representation of EDMF is then compared against LES output using the “history-matching” (HM) method of ref. 74, exploring the range of uncertainty simultaneously across multiple parameters at a number of points. The full range of parametric uncertainty is then explored using emulators, as described in ref. 72. Rather than seeking a single optimum λ_p , HM seeks only to remove implausible regions of parametric space from consideration, leaving a “not ruled out yet” (NROY) region for consideration. Any parameter

values within NROY are possible valid values, and new constraints can be progressively added as needed. Finally, the HM method can also serve as a means of diagnosing structural error: A null NROY space (i.e., no possible values of λ_p permit \mathcal{M} to meet the desired constraints within the chosen tolerance) indicates that the representation in \mathcal{M} needs to be refined.

An independent effort from the Climate Modeling Alliance (75) project uses a somewhat different approach, but with the same ends, namely, to calibrate an EDMF scheme with a set of parameters and their uncertainty bounds (76, 77). There are also efforts to build a “library” of cloud regimes (78) for use in training. As noted above in Section 3, LES models carry their own structural uncertainty. The Development and Evaluation of PHYSical parameterizations of atmospheric models (79) project seeks to build systematic benchmarks and training datasets for similar efforts elsewhere.

This represents the first step toward a more systematic and objective exploration of parametric uncertainty than is possible when GCMs are directly used as the forward model. These methods are now being extended to consider the tuning of slower processes in the ocean, following ref. 80. Some of the subsequent problems of tuning coupled systems with multiple timescales are still being explored using highly simplified models (81). Preliminary results indicate that compensating errors and other defects inherent in the tuning of coupled systems may still be present, but it may be possible to diagnose those within the HM method.

5. Why Do We Need Counterfactual Earths, and Where Do They Come from?

In Sections 3 and 4, we have outlined a procedure for creating a tool for numerically simulating a very complex multiscale system at a maximal level of detail given computing limits, and calibrating the system within those limits to resemble the observed Earth as closely as possible.

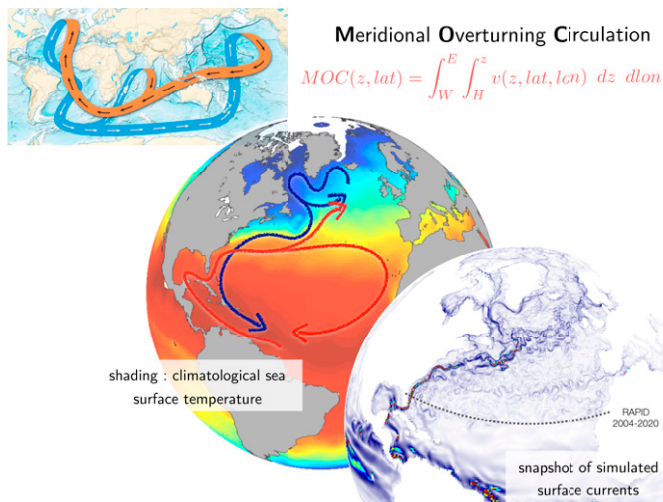


Fig. 4. Different views of the MOC, from models at various levels of Charney's ladder. They all exhibit fidelity to aspects of our knowledge of the MOC, and our understanding of the MOC lies in the composite.

We use this system to study Earth's climate and its history; the appearance of life and the maintenance of an atmosphere, ocean, and land surface suitable for life; the fluctuations of the past and the recent period of relative stability that allowed for the possibility of settlements and agriculture; and the consequences of the Industrial Revolution leading to the current climate emergency. As an object of science, Earth poses a key problem, in that there is only one instance of it, and one temporal trajectory of its state, that we can observe. Simulated Earths remain our only means of exploring different hypotheses about how the system works, and other trajectories it might have taken and might take in the future. It also remains our only means of exploring responses to the climate emergency, and understanding and predicting the impacts of different policy choices of mitigation, adaptation, and resilience building.

Using techniques pioneered by Hasselmann, another of 2021's Nobels in Physics (e.g., ref. 11), the techniques of detection and attribution have helped us understand the role of different forcing agents in F , and their contribution to a changing climate. First, a comparison of present-day climate with a counterfactual simulation known as the "pre-Industrial control" (where F is held constant at its value in 1850 CE), allows us to unequivocally detect that the climate has changed, beyond the bounds of simulated random internal variability. Attribution of climate change to particular forcing agents is performed using either single-forcing runs, where all contributors to F are held constant save one, or agents grouped into greenhouse gases only, natural only, etc. (82), to tease apart the contribution of different forcings to climate change, as well as their nonlinearities (as the contributions may not be simply additive). The number of individual forcings considered now number over 10 (see, e.g., figure 2c in *IPCC AR6 Summary for Policymakers* (35)). In addition, ref. 23 considers at least eight pathways of future evolution of forcing. Together, such simulations represent a requirement for GCMs to explore hundreds of counterfactual pathways of climate evolution.

We draw attention to two major implications of this for the use of GCMs. First, GCMs are used to explore a counterfactual space, not directly constrained by observations. This

includes chaotic uncertainty (different trajectories of \mathbf{x} under changes in initial conditions, which provide a measure of the intrinsic internal variability), and to counterfactual settings of F , where various actual observed forcings are turned on or off. An individual GCM (a particular formulation of the terms R , U , and P) is first calibrated against theory and observations as described in Section 4 and, once this is satisfactory, is used to explore an even vaster counterfactual space (see, e.g., figure 4 from ref. 29). The requirement to be able to simulate counterfactuals must be taken into account in the context of ML as well, as we shall discuss below in Section 6.

Second, in view of the computational expense of GCMs outlined in Section 1 (see also ref. 83), it has proved prohibitively expensive to explore all the forcings, and their potential future pathways, using GCMs. The IPCC has turned, instead, to an extended use of emulators. Note that, unlike the emulators in Section 4, which attempt to mimic individual climate processes, these are emulators of the whole climate system, which attempt to predict the response of the entire system, usually in the form of an integrated measure such as the global mean surface temperature, to changes in F . Regional patterns of climate change can be inferred by coupling with techniques such as pattern scaling (84). The emulators are all reduced-complexity models of various flavors, ranging from relatively simple regressions trained on recent historical data, to dynamical systems models such as impulse-response models, to highly simplified, and usually 1D, GCMs (85). Their advantage is that they are typically millions of times faster than GCMs (86), although their lack of internal physical consistency poses epistemic risk.

One of the key measures of the climate response to CO_2 , the ECS, is itself a counterfactual, as it is based on an asymptotic equilibrium that is never observed in nature. It is nonetheless useful as a measure in order to project a range of responses to scenarios (possible future trajectories of F). Furthermore, while the recent (and most precise) observational record of the satellite era is too short to constrain GCMs adequately, there are other indirect means of placing limits on ECS, including paleoclimate data and constraints on individual processes contributing to the ECS (13). The recent IPCC concluded that many GCMs were providing ECS outside the "very likely" range, and used emulators where ECS is a tunable parameter, to refine the consensus projections and their uncertainty bounds (86).

Reduced complexity models have also been used extensively to study the potential for abrupt transitions in the climate system, for which there is some evidence in the paleoclimate record (87). The MOC (see Fig. 4) is a canonical feature of the climate system with the potential for metastability: Under sufficient fresh water input in the North Atlantic, from retreating continental ice sheets, for instance, the MOC can "collapse." Yet, GCMs exhibit metastability of the MOC less readily than reduced-complexity models, and eddy-resolving models even less so (88). A concern often expressed is that GCMs are too stable to perturbations (37).

The ability to run a wide variety of century- or millennial-scale simulations is essential for an evaluation of many aspects of climate risk, such as tipping points. High-resolution model trajectories constrained by assimilation of recent

observations will do little to mitigate this concern. The discussion above in Section 4 points to ways forward using objective methods of emulation of GCMs, just as GCMs themselves learn to emulate CRMs and LES models. This would alleviate the epistemic risk associated with emulators.

6. What Might a Future Modeling Landscape Look Like?

In a celebrated 1972 essay “More is different” (89), Philip Anderson, another Nobel laureate in Physics, argued that many of our sciences abstract reality at different levels of complexity, and struck a cautionary note on the limits of assuming that one level of explanation is “nothing more” than an expression of aggregate behavior of the elements at a deeper level of abstraction. The understanding of such emergent behavior of complex networked systems (CNS) happens at multiple levels, and the reduction of the properties at one level to the one below may not be computationally tractable or of practical use, even if it is true in principle. The Earth system is an exemplar of a CNS, as discussed here: bringing together domains as different as fluid dynamics, radiative transfer, chemistry, biology, and ecology, over a range of time and space scales. Yet the assembly as a whole appears to persist in stable states for millennia. The scientific puzzles related to the climate emergency center on understanding the emergent balance of terms perpetuating stable states, and what thresholds of these balances we are exceeding with anthropogenic perturbations.

As simulation is now a pillar of the scientific method, one temptation at this point is to turn over the problem of emergence to the computer itself. Many diverse fields of science have attempted to simulate a CNS at a high level of network fidelity, explicitly incorporating all the interactions and feedbacks. The question is then, Can such a numerical system spontaneously exhibit higher levels of organization? In neuroscience for example, we can pose this question in the context of the emergence of higher-level brain function from the details expressed or simulated at the level of individual neurons. The HBP (90) attempts to do just this, and others have questioned the limits of this approach (e.g., ref. 91, which examines limits to this approach in the context of vision, or ref. 92, which looks at “top-down” effects where the large-scale structure regulates the behavior at the neuronal level).

The equivalent question to pose in climate science is whether the stable states of the climate, where \mathbf{x} from Eq. 1 remains near an attractor despite fluctuations around it, emerge directly from the assembly of the system with all its details. We also would like to see whether such a detailed simulation can accurately capture the responses of the system to changes in F . Despite the “fast” physics of the atmosphere and planetary surface, we have seen that the system contains natural and forced variability at “slow” timescales, regulated by emergent features of the general circulation such as the MOC. The timescales of changes in forcing, such as anthropogenic emissions since the Industrial Revolution, and of responses in features such as MOC demand modeling tools capable of century-long simulations under many possible forcing pathways, 100 simulations of 100 SY each, as argued in Section 1. We

have reviewed the arguments for kilometer-scale models in Section 2, which are quite limited in capability compared to the GCMs [optimistically, perhaps capable of $\mathcal{O}(10)$ simulations each of $\mathcal{O}(10)$ SY in length on the largest available computing platform, an order of magnitude below that (39, 93)]. Simultaneously, those exploring policy responses require very fast models for exploring many policy options under many forcing scenarios.

The sense that GCMs may be obsolete comes from these conflicting demands: very high resolution for some key processes, which restrict the ability to explore and quantify uncertainties and study low-frequency variability; the need to explore many counterfactual scenarios for constructing climate policy, which cannot be guided simply by present-day observations; and the long simulation times needed to understand prior episodes of abrupt climate change.

As noted in the comparison with neuroscience, this is a debate, across many fields, on how much reliance to place on the most complex models capturing detail using the elements of the finest representation of all the elements in a complex system. Such simulations are sometimes now called “digital twins,” borrowing a term from manufacturing (e.g., ref. 94), where it originally meant an exact digital copy of an engineered system and its specifications. It has been increasingly used to describe complex systems, including living systems (95). But, as noted in ref. 95, for sufficiently complex systems, we need simulations at multiple levels of abstraction and complexity.

The MOC is a case in point. At the broadest level, Earth is warmed by radiation from the sun around the tropics, and loses heat near the poles, implying an equator-to-pole transport of heat in the ocean-atmosphere system. In the ocean, for the current planetary topography, this results in what is referred to as the ocean’s “conveyor belt” for heat and energy (96). Looking a little closer, we see important features such as the horizontal gyres that also contribute to the meridional transport of heat northward in the North Atlantic (97), an important feature regulating the climate on the continents around (98). Zooming farther in, we can see that turbulent eddies continue to transport heat across the meridional flow in the North Atlantic (99). All levels of explanation are broadly consistent with data, and, while each rung of the ladder of complexity can be described in terms of residuals of a fine-scale balance from a level below, the balance is regulated by global constraints such as the equator-to-pole radiative imbalance. In short, it is the entire hierarchy that constitutes our understanding of the MOC.

In the modeling of complex multiscale systems, calling one of these layers a “twin” at the expense of others appears to be a rhetorical overreach (100). Indeed, in speaking of the HBP in ref. 101, the ambition is sharply circumscribed:

In constructing a ‘digital twin’ of a living organ, one is confronted by important challenges over and above those encountered when constructing the digital twin of an inanimate object. Therefore, the concept of the ‘digital twin’ in this context needs to be carefully defined to provide clarity on its limitations and to avoid creating unrealistic expectations of exact fidelity ... The digital twin is thus a copy in the practical sense, usually associated with a model of a function or process, and its power lies

in its usefulness in dealing with relevant problems faced by its physical counterpart without the need (and certainly not the claim) of capturing every single detail thereof.

In the climate context, this could describe any numerical model dating back to ref. 5!

We have demonstrated (Section 4) that there are methods for making the model hierarchy traceable, by showing how to derive a parameterization on the basis of a model farther up what Charney called the model “ladder” (16). It might be worth including such high-resolution benchmarks for the training of parameterizations in future model intercomparison projects for a new generation of GCMs calibrated using the methods outlined here. Similar methods can be imagined to link the reduced-complexity models of ref. 85 to GCMs, from which we can carefully explore how to remove the biases in the GCMs themselves, such as those outlined in Section 2.

We contend, in this article, that the GCM remains the indispensable meeting point of these divergent directions. Its structure encapsulates our fundamental understanding of how the climate works, and represents an astute assembly of choices and trade-offs that are versatile enough to meet the challenges outlined here. The GCM's column structure reflects the importance of separating the vertical in the model topology, and the importance of convection and its timescales in atmosphere and ocean. The structural independence of columns has been seen as a limitation, but new methods can use nonlocal predictors (102). The stochastic parameterizations mentioned in Section 3 also impose nonlocal (in space and time) coherence to the stochasticity (103).

Just as data assimilation techniques for constraining trajectories toward a time series of observations led to major improvements in NWP (22), these new methods, based on data from process models such as CRMs, or observations, as well as imposing physical constraints, hold out the possibility of efficiently approaching the attractors of the system, the key feature distinguishing the climate problem from weather, as outlined in Section 1. The recommendation is

to adopt a rigorous, transparent, and reproducible tuning process, rather than assuming tuning will simply disappear when we simulate at a high enough level of detail.

Finally, we do not wish to forget the use of models as pedagogical tools, for students to explore climate response, to have “fun”—a point repeatedly made by Manabe after receiving the Nobel Prize, for example, in his first Nobel press conference (104). Models must, in addition, be easy to use to explore their sensitivity to counterfactual changes, and explore novel and risky ideas on how it responds to perturbations.

The future modeling landscape must rest on the principle of a traceable model hierarchy (16). Models at every level of the hierarchy have their own forms of structural uncertainty, as noted above; this uncertainty does not vanish at any conceivable model resolution possible on any known computational technology. Each model can be put to multiple uses and subject to diverse physical and computational constraints. The traditional GCM, with its ability to combine resolved dynamics with unresolved physics for a nonstationary Earth system, remains the crossroads between models built for other purposes: models that can resolve some of the physical uncertainties, but in limited settings; models that can be used to study transitions in the climate system that are abrupt events between millennia-long stable states; and emulators that produce corrected data for downstream users. Each of these involve trade-offs sacrificing accuracy in one part of the climate system against another. The GCM will remain the essential element ensuring that these trade-offs remain within reasonable limits for the entire Earth system.

6.1. Data, Materials, and Software Availability. There are no data underlying this work.

ACKNOWLEDGMENTS. V.B. is supported by the Cooperative Institute for Modeling the Earth System, Princeton University (Award NA18OAR4320123 from the National Oceanic and Atmospheric Administration, US Department of Commerce), and by the French Agence National de Recherche “Investissements d’avenir” Make Our Planet Great Again Program (ANR-17-MPGA-0010). We thank Naser Mahfouz, Zhihong Tan, and two anonymous reviewers for substantive and detailed reviews that have immeasurably improved the manuscript.

1. N. A. Phillips, The general circulation of the atmosphere: A numerical experiment. *Q. J. R. Meteorol. Soc.* **82**, 123–164 (1956).
2. E. N. Lorenz, *The Nature and Theory of the General Circulation of the Atmosphere* (World Meteorological Organization, Geneva, 1967).
3. V. Bjerknes, Das problem der wettverhers-age, betrachtet vom standpunkte der mechanik und der physik [in German]. *Meteor. Z.* **21**, 1–7 (1904).
4. J. Bjerknes, On the structure of moving cyclones. *Mon. Weather Rev.* **47**, 95–99 (1919).
5. J. Charney, R. Fjortoft, J. von Neumann, Numerical integration of the barotropic vorticity equation. *Tellus* **2**, 237–254 (1950).
6. J. G. Charney *et al.*, *Carbon Dioxide and Climate: A Scientific Assessment* (National Academies, 1979).
7. R. P. Sorenson, “Eunice Foote’s pioneering research on CO₂ and climate warming” (Search and Discovery Article 70092, American Association of Petroleum Geologists, 2011).
8. A. Huddleston, Happy 200th birthday to Eunice Foote, hidden climate science pioneer. Climate.gov. (2019). <https://www.climate.gov/news-features/features/happy-200th-birthday-eunice-foote-hidden-climate-science-pioneer>. Accessed 12 October 2022.
9. S. Manabe, R. T. Wetherald, Thermal equilibrium of the atmosphere with a given distribution of relative humidity. *J. Atmos. Sci.* **24**, 241–259 (1967).
10. S. Manabe, R. T. Wetherald, The effects of doubling the CO₂ concentration on the climate of a general circulation model. *J. Atmos. Sci.* **32**, 3–15 (1975).
11. K. Hasselmann, Optimal fingerprints for the detection of time-dependent climate change. *J. Clim.* **6**, 1957–1971 (1993).
12. S. Bony *et al.*, “Carbon dioxide and climate: Perspectives on a scientific assessment” in *Climate Science for Serving Society*, G. R. Asrar, J. W. Hurrell, Eds. (Springer, 2013), pp. 391–413.
13. S. C. Sherwood *et al.*, An assessment of Earth’s climate sensitivity using multiple lines of evidence. *Rev. Geophys.* **58**, e2019RG000678 (2020).
14. B. Stevens *et al.*, The added value of large-eddy and storm-resolving models for simulating clouds and precipitation. *Kisho Shushi. Dai2shu* **98**, 395–435 (2020).
15. J. Slingo *et al.*, Ambitious partnership needed for reliable climate prediction. *Nat. Clim. Chang.* **12**, 499–503 (2022).
16. V. Balaji, Climbing down Charney’s ladder: Machine learning and the post-Dennard era of computational climate science. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **379**, 20200085 (2021).
17. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
18. L. Spinney, Are we witnessing the dawn of post-theory science? *The Guardian*, 9 January 2022. <https://www.theguardian.com/technology/2022/jan/09/are-we-witnessing-the-dawn-of-post-theory-science>. Accessed 12 October 2022.
19. S. Manabe, J. Smagorinsky, R. F. Strickler, Simulated climatology of a general circulation model with a hydrologic cycle. *Mon. Weather Rev.* **93**, 769–798 (1965).
20. P. N. Edwards, History of climate modeling. *WIREs Clim. Chang.* **2**, 128–139 (2011).
21. E. N. Lorenz, On the predictability of hydrodynamic flow. *Trans. N. Y. Acad. Sci.* **25**, 409–432 (1963).
22. P. Bauer, A. Thorpe, G. Brunet, The quiet revolution of numerical weather prediction. *Nature* **525**, 47–55 (2015).
23. C. Tebaldi *et al.*, Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6. *Earth Syst. Dyn.* **12**, 253–293 (2021).
24. W. Broecker, G. Bond, M. Klas, E. Clark, J. McManus, Origin of the northern Atlantic’s Heinrich events. *Clim. Dyn.* **6**, 265–273 (1992).
25. J. F. McManus, R. Francois, J. M. Gherardi, L. D. Keigwin, S. Brown-Leger, Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. *Nature* **428**, 834–837 (2004).

26. T. Schneider, C. M. Kaul, K. G. Pressel, Possible climate transitions from breakup of stratocumulus decks under greenhouse warming. *Nat. Geosci.* **12**, 163–167 (2019).
27. T. Reichler, J. Kim, How well do coupled models simulate today's climate? *Bull. Am. Meteorol. Soc.* **89**, 303–312 (2008).
28. G. Abramowitz *et al.*, ESD reviews: Model dependence in multi-model climate ensembles: Weighting, sub-selection and out-of-sample testing. *Earth Syst. Dyn.* **10**, 91–105 (2019).
29. E. Hawkins, R. Sutton, The potential to narrow uncertainty in regional climate predictions. *Bull. Am. Meteorol. Soc.* **90**, 1095–1108 (2009).
30. P. S. Schumacher, K. L. Rasmussen, The formation, character and changing nature of mesoscale convective systems. *Nat. Rev. Earth Environ.* **1**, 300–314 (2020).
31. P. B. Rhines, Geostrophic turbulence. *Annu. Rev. Fluid Mech.* **11**, 401–441 (1979).
32. L. C. Jackson *et al.*, The mean state and variability of the north Atlantic circulation: A perspective from ocean reanalyses. *J. Geophys. Res. Ocean.* **124**, 9141–9170 (2019).
33. F. Hourdin *et al.*, The art and science of climate model tuning. *Bull. Am. Meteorol. Soc.* **98**, 589–602 (2017).
34. ESGF Data Statistics, CMIP6 models and modelling institutes (updated daily). esgf-ii.cmc.csi.cmc.ca/esgf-dashboards/ui/data-archive/CMIP6.html. Accessed 12 October 2022.
35. V. Masson-Delmotte *et al.*, *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Intergovernmental Panel on Climate Change, Geneva, Switzerland, 2021).
36. Z. Hausfather, K. Marvel, G. A. Schmidt, J. W. Nielsen-Gammon, M. Zelinka, Climate simulations: Recognize the 'hot model' problem. *Nature* **605**, 26–29 (2022).
37. P. Valdes, Built for stability. *Nat. Geosci.* **4**, 414–416 (2011).
38. C. Schär *et al.*, Kilometer-scale climate models: Prospects and challenges. *Bull. Am. Meteorol. Soc.* **101**, E567–E587 (2020).
39. P. Neumann *et al.*, Assessing the scales in numerical weather and climate predictions: Will exascale be the rescue? *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **377**, 20180148 (2019).
40. G. Nastrom, K. S. Gage, A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft. *J. Atmos. Sci.* **42**, 950–960 (1985).
41. D. Stammer, C. W. Böning, Mesoscale variability in the Atlantic Ocean from Geosat altimetry and WOCE high-resolution numerical modeling. *J. Phys. Oceanogr.* **22**, 732–752 (1992).
42. X. Capet, J. C. McWilliams, M. J. Molemaker, A. F. Schepetkin, Mesoscale to submesoscale transition in the California current system, Part III: Energy balance and flux. *J. Phys. Oceanogr.* **38**, 2256–2269 (2008).
43. V. Balaji, J. L. Redelsperger, Sub-gridscale effects in mesoscale deep convection: Initiation, organization and turbulence. *Atmos. Res.* **40**, 339–381 (1996).
44. A. Arakawa, W. H. Schubert, Interaction of a cumulus cloud ensemble with the large-scale environment, part I. *J. Atmos. Sci.* **31**, 674–701 (1974).
45. D. Randall, M. Khairoutdinov, A. Arakawa, W. Grabowski, Breaking the cloud parameterization deadlock. *Bull. Am. Meteorol. Soc.* **84**, 1547–1564 (2003).
46. M. F. Khairoutdinov, D. A. Randall, A cloud resolving model as a cloud parameterization in the NCAR community climate system model: Preliminary results. *Geophys. Res. Lett.* **28**, 3617–3620 (2001).
47. H. Parishani, M. S. Pritchard, C. S. Bretherton, M. C. Wyant, M. Khairoutdinov, Toward low-cloud-permitting cloud superparameterization with explicit boundary layer turbulence. *J. Adv. Model. Earth Syst.* **9**, 1542–1571 (2017).
48. M. Satoh *et al.*, Global cloud-resolving models. *Curr. Clim. Change Rep.* **5**, 172–184 (2019).
49. B. Stevens *et al.*, DYAMOND: The Dynamics of the atmospheric general circulation modeled on non-hydrostatic domains. *Prog. Earth Planet. Sci.* **6**, 61 (2019).
50. J. Y. Grandpeix, J. P. Lafore, A density current parameterization coupled with Emanuel's convection scheme. Part I: The models. *J. Atmos. Sci.* **67**, 881–897 (2010).
51. M. D. Zelinka *et al.*, Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.* **47**, e2019GL085782 (2020).
52. G. de Leeuw *et al.*, Production flux of sea spray aerosol. *Rev. Geophys.* **49**, RG2001 (2011).
53. D. A. Randall, Km. Xu, R. J. C. Somerville, S. Iacobellis, Single-column models and cloud ensemble models as links between observations and climate models. *J. Clim.* **9**, 1683–1697 (1996).
54. A. A. Wing *et al.*, Clouds and convective self-aggregation in a multimodel ensemble of radiative-convective equilibrium simulations. *J. Adv. Model. Earth Syst.* **12**, e2020MS002138 (2020).
55. T. Becker, A. A. Wing, Understanding the extreme spread in climate sensitivity within the radiative-convective equilibrium model intercomparison project. *J. Adv. Model. Earth Syst.* **12**, e2020MS002165 (2020).
56. C. Heim, L. Hentgen, N. Ban, C. Schär, Inter-model variability in convection-resolving simulations of subtropical marine low clouds. *Kisho Shushi. Daijshu* **99**, 1271–1295 (2021).
57. H. M. Christensen, O. G. A. Driver, The fractal nature of clouds in global storm-resolving models. *Geophys. Res. Lett.* **48**, e2021GL095746 (2021).
58. A. P. Siebesma *et al.*, A large eddy simulation intercomparison study of shallow cumulus convection. *J. Atmos. Sci.* **60**, 1201–1219 (2003).
59. M. C. vanZanten *et al.*, Controls on precipitation and cloudiness in simulations of trade-wind cumulus as observed during RICO. *J. Adv. Model. Earth Syst.* **3**, M06001 (2011).
60. R. J. Beare *et al.*, An intercomparison of large-Eddy simulations of the stable boundary layer. *Boundary Layer Meteorol.* **118**, 247–272 (2006).
61. F. Couvreux *et al.*, Intercomparison of large-Eddy simulations of the Antarctic boundary layer for very stable stratification. *Boundary Layer Meteorol.* **176**, 369–400 (2020).
62. K. G. Pressel, S. Mishra, T. Schneider, C. M. Kaul, Z. Tan, Numerics and subgrid-scale modeling in large eddy simulations of stratocumulus clouds. *J. Adv. Model. Earth Syst.* **9**, 1342–1365 (2017).
63. R. Hallberg, Using a resolution function to regulate parameterizations of oceanic mesoscale eddy effects. *Ocean Model.* **72**, 92–103 (2013).
64. M. F. Jansen, A. J. Adcroft, R. Hallberg, I. M. Held, Parameterization of eddy fluxes based on a mesoscale energy budget. *Ocean Model.* **92**, 28–41 (2015).
65. S. D. Bachman, The GM+E closure: A framework for coupling backscatter with the Gent and McWilliams parameterization. *Ocean Model.* **136**, 85–106 (2019).
66. G. A. Schmidt *et al.*, Practice and philosophy of climate model tuning across six U.S. modeling centers. *Geosci. Model Dev.* **10**, 3207–3223 (2017).
67. I. M. Held *et al.*, Structure and performance of GFDL's CM4.0 climate model. *J. Adv. Model. Earth Syst.* **11**, 3691–3727 (2019).
68. M. Zhao *et al.*, The GFDL global atmosphere and land model AM4.0/LM4.0: 2. Model description, sensitivity studies, and tuning strategies. *J. Adv. Model. Earth Syst.* **10**, 735–769 (2018).
69. G. E. Karniadakis *et al.*, Physics-informed machine learning. *Nat. Rev. Phys.* **3**, 422–440 (2021).
70. F. Hourdin, F. Couvreux, L. Menut, Parameterization of the dry convective boundary layer based on a mass flux representation of thermals. *J. Atmos. Sci.* **59**, 1105–1123 (2002).
71. C. Rio, F. Hourdin, F. Couvreux, A. Jam, Resolved versus parameterized boundary-layer plumes. Part II: Continuous formulations of mixing rates for Mass-Flux schemes. *Boundary Layer Meteorol.* **135**, 469–483 (2010).
72. F. Couvreux *et al.*, Process-based climate model development harnessing machine learning: I. A calibration tool for parameterization improvement. *J. Adv. Model. Earth Syst.* **13**, e2020MS002217 (2021).
73. F. Hourdin *et al.*, Process-based climate model development harnessing machine learning: II. Model calibration from single column to global. *J. Adv. Model. Earth Syst.* **13**, e2020MS002225 (2021).
74. D. Williamson *et al.*, History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Clim. Dyn.* **41**, 1703–1729 (2013).
75. Climate Modeling Alliance, *Climate Modeling Alliance*. <https://clima.caltech.edu/>. Accessed 12 October 2022.
76. O. R. A. Dunbar, A. Garbuno-Inigo, T. Schneider, A. M. Stuart, Calibration and Uncertainty Quantification of Convective Parameters in an Idealized GCM. *J. Adv. in Modeling Earth Sys.* **13**, e2020MS002454. <https://doi.org/10.1029/2020MS002454>. Accessed 24 October 2022.
77. E. Cleary, A. Garbuno-Inigo, S. Lan, T. Schneider, A. M. Stuart, Calibrate, emulate, sample. *J. Comput. Phys.* **424**, 109716 (2021).
78. Z. Shen *et al.*, A Library of Large-Eddy Simulations Forced by Global Climate Models. *J. Adv. in Modeling Earth Sys.* **14**, no. 3 (2022): e2021MS002631. <https://doi.org/10.1029/2021MS002631>. Accessed 24 October 2022.
79. National Centre for Meteorological Research, DEPHY. <https://www.umr-cnrm.fr/spip.php?article930&lang=en>. Accessed 12 October 2022.
80. D. B. Williamson, A. T. Blaker, B. Sinha, Tuning without over-tuning: Parametric uncertainty quantification for the NEMO ocean model. *Geosci. Model Dev.* **10**, 1789–1816 (2017).
81. L. Jackson, P. Forster, The role of Climate Model Emulators in the IPCC 6th Assessment Report: An Online Workshop for the Scientific Community 30 September 2021. Zenodo. <https://doi.org/10.5281/zenodo.5615538>. Accessed 24 October 2022.
82. N. P. Gillett *et al.*, The detection and attribution model intercomparison project, (DAMIP v1.0) contribution to CMIP6. *Geosci. Model Dev.* **9**, 3685–3697 (2016).
83. V. Balaji *et al.*, CPMP: Measurements of real computational performance of Earth system models in CMIP6. *Geosci. Model Dev.* **10**, 19–34 (2017).
84. C. Tebaldi, J. M. Arblaster, Pattern scaling: Its strengths and limitations, and an update on the latest model simulations. *Clim. Change* **122**, 459–471 (2014).
85. Z. R. J. Nicholls *et al.*, Reduced complexity model intercomparison project phase 1: Introduction and evaluation of global-mean temperature response. *Geosci. Model Dev.* **13**, 5175–5190 (2020).
86. L. Jackson, P. Forster, The role of climate model emulators in the IPCC 6th assessment report: An online workshop for the scientific community (2021).
87. V. Brovkin *et al.*, Past abrupt changes, tipping points and cascading impacts in the Earth system. *Nat. Geosci.* **14**, 550–558 (2021).
88. P. R. Gent, A commentary on the Atlantic meridional overturning circulation stability in climate models. *Ocean Model.* **122**, 57–66 (2018).
89. P. W. Anderson, More is different. *Science* **177**, 393–396 (1972).
90. K. Amunts *et al.*, Linking brain structure, activity, and cognitive function through computation. *eNeuro* **9**, ENEURO.0316-21.2022 (2022).
91. Y. Frégnac, Big data and the industrialization of neuroscience: A safe roadmap for understanding the brain? *Science* **358**, 470–477 (2017).
92. G. Ellis, Top-down effects in the brain. *Phys. Life Rev.* **31**, 11–27 (2019).
93. N. P. Wedi *et al.*, A baseline for global weather and climate simulations at 1 km resolution. *J. Adv. Model. Earth Syst.* **12**, e2020MS002192 (2020).
94. W. Kitzinger, M. Karner, G. Traar, J. Henjess, W. Sihn, Digital Twin in manufacturing: A categorical literature review and classification. *IFAC PapersOnLine* **51**, 1016–1022 (2018).
95. J. Wen, B. Gabrys, K. Musial, Towards digital twin oriented modelling of complex networked systems and their dynamics: A comprehensive survey. *arXiv [Preprint]* (2022). <https://doi.org/10.48550/arXiv:2202.09363>. Accessed 24 October 2022.
96. W. S. Broecker, The great ocean conveyor. *Oceanography (Wash. D.C.)* **4**, 79–89 (1991).
97. R. Lumpkin, K. Speer, Large-scale vertical and horizontal circulation in the North Atlantic ocean. *J. Phys. Oceanogr.* **33**, 1902–1920 (2003).
98. J. B. Palter, The role of the Gulf stream in European climate. *Annu. Rev. Mar. Sci.* **7**, 113–137 (2015).
99. A. M. Treguier, J. Deshayes, C. Lique, R. Dussin, J. M. Molines, Eddy contributions to the meridional transport of salt in the North Atlantic. *J. Geophys. Res. Ocean.* **117**, C05010 (2012).
100. K. Evers, A. Salles, Epistemic challenges of digital twins & virtual brains perspectives from fundamental neuroethics. *Scio* **21**, 27–53 (2021).
101. K. Amunts *et al.*, The coming decade of digital brain research - A vision for neuroscience at the intersection of technology and computing. Zenodo. <https://zenodo.org/record/6630232?Y1AT0C-B3vw>. Accessed 24 October 2022.
102. P. Wang, J. Yuval, P. A. O'Gorman, Non-local parameterization of atmospheric subgrid processes with neural networks. *arXiv [Preprint]* (2022). <https://doi.org/10.48550/arXiv:2201.00417>. Accessed 24 October 2022.
103. M. Leutbecher *et al.*, Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Q. J. Royal Meteorol. Soc.* **143**, 2315–2339 (2017).
104. Princeton University, "News conference for 2021 Nobel Prize in Physics: Syukuro Manabe" (video recording 2021). <https://www.youtube.com/watch?v=ButzK41Qpww>. Accessed 12 October 2022.