Discussion

Climbing down Charney's ladder: machine learning and the post-Dennard era of computational climate science

V. Balaji[1,2]

[1]Princeton University and NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA
[2]Institute Pierre-Simon Laplace, Paris, France

VB, 0000-0001-7561-5438

The advent of digital computing in the 1950s sparked a revolution in the science of weather and climate. Meteorology, long based on extrapolating patterns in space and time, gave way to computational methods in a decade of advances in numerical weather forecasting. Those same methods also gave rise to computational climate science, studying the behaviour of those same numerical equations over intervals much longer than weather events, and changes in external boundary conditions. Several subsequent decades of exponential growth in computational power have brought us to the present day, where models ever grow in resolution and complexity, capable of mastery of many small-scale phenomena with global repercussions, and ever more intricate feedbacks in the Earth system. The current juncture in computing, seven decades later, heralds an end to what is called Dennard scaling, the physics behind ever smaller computational units and ever faster arithmetic. This is prompting a fundamental change in our approach to the simulation of weather and climate, potentially as revolutionary as that wrought by John von Neumann in the 1950s. One approach could return us to an earlier era of pattern recognition and extrapolation, this time aided by computational power. Another approach could lead us to insights that continue to be expressed in mathematical equations. In either approach, or any synthesis of those, it is clearly no longer the steady march of the last few decades, continuing to add detail to ever more elaborate models. In this prospectus,

we attempt to show the outlines of how this may unfold in the coming decades, a new harnessing of physical knowledge, computation and data.

This article is part of the theme issue 'Machine learning for weather and climate modelling'.

1. Introduction

The history of numerical weather prediction and climate simulation is almost exactly coincident with the history of digital computing itself [1]. The story of these early days has been told many times (e.g. [2]), including by the pioneers themselves, who worked alongside John von Neumann starting in the late 1940s. We shall revisit this history below, as some of those early debates are being reprised today, the subject of this survey. In the seven decades since those pioneering days, numerical methods have become central to meteorology and oceanography. In meteorology and numerical weather prediction, a 'quiet revolution' [3] has given us decades of steady increase in the predictive skill of weather forecasting based on models that directly integrate the equations of motion to predict the future evolution of the atmosphere, taking into account thermodynamic and radiative effects and time-evolving boundary conditions, such as the ocean and land surface. This has been made possible by systematic advances in algorithms and numerical techniques, but most of all by an extraordinary, steady, decades-long exponential expansion of computing power. Figure 1 shows the expansion of computing power at one of the laboratories that in fact traces its history back to the von Neumann era. The very first numerical weather forecast was in fact issued in 1956 from the IBM 701 [4], which is the first computer shown in figure 1.

Numerical meteorology, based on a representation of real numbers in a finite number of bits, also serendipitously led to one of the most profound discoveries of the latter half of the twentieth century, namely that even completely deterministic systems have limits to the predictability of the future evolution of the system [5].[1] Simply knowing the underlying physics does not translate to an ability to predict beyond a point.

Using the same dynamical systems run for very long periods of time (what von Neumann called the 'infinite forecast' [6]), the field of climate simulation developed over the same decades. While simple radiative arguments for CO_2-induced warming were advanced in the nineteenth century,[2] numerical simulation led to a detailed understanding which goes beyond radiation and thermodynamics to include the dynamical response of the general circulation to an increase in atmospheric CO_2 (e.g. [8]).

The issue of *detail* in our understanding is worth spelling out. The exorbitant increase in computing power of the last decades have been absorbed in the adding of detail, principally in model spatial resolution, but also in the number of physical variables and processes simulated. That addition of detail has been shown to make significant inroads into understanding and the ability to predict the future evolution of the weather and climate system (henceforth Earth system). Not only does the addition of fine spatial and temporal scales—if one were to resolve clouds, for example—have repercussions at much larger (planetary) scales, but the details of interactions between different components of the Earth system—ocean and biosphere for example—have first-order effects on the general circulation as well.

It is the contention of this article that the continual addition of detail in our simulations is something that may have to be reconsidered, given certain physical limits on the evolution of computing hardware that we shall outline below. We may be compelled down some radically different paths.

The key feature of today's silicon-etched CMOS-based computing is that 'arithmetic will no longer run faster, and may even run slower, but we can perform more of it in parallel'. This has led to the revival of a computing approach which also can be traced back to the 1940s, using networks

[1]While this was formally known since Poincaré, Lorenz first showed it in a simple system of physical equations.

[2]This is generally attributed to Tyndall and Arrhenius, though earlier published research by Eunice Foote has recently come to light [7].

HISTORY OF GFDL COMPUTING
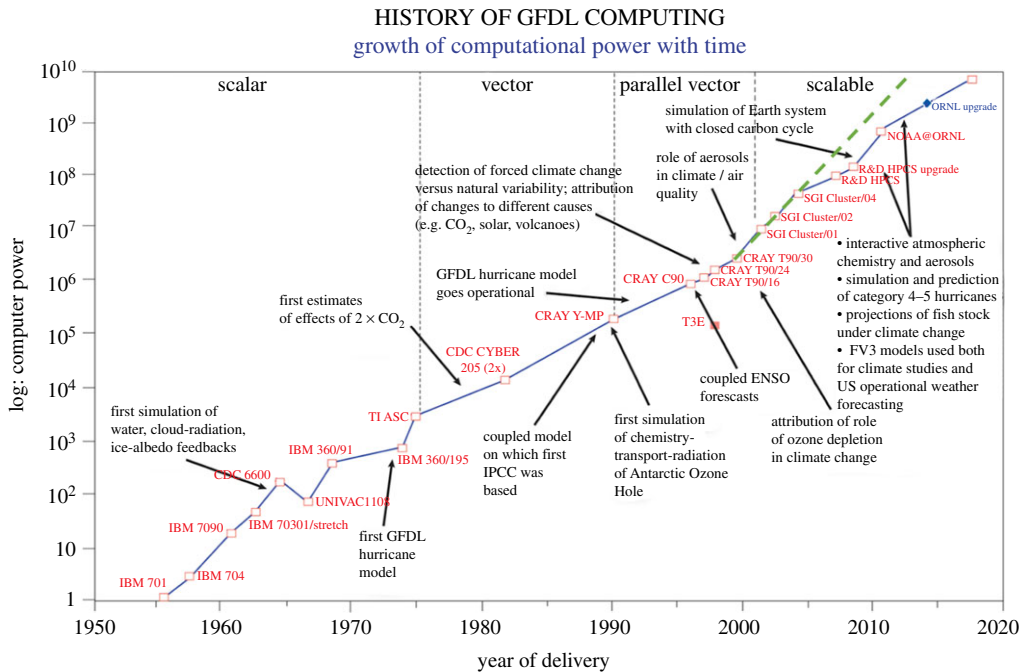### growth of computational power with time

Figure 1. History of computational power at the NOAA Geophysical Fluid Dynamics Laboratory. Computational power is measured in aggregate floating-point operations per second, scaled so that the IBM 701 machine equals 1. Epochs (scalar, vector, parallel, scalable) show the dominant technology of the time. Landmark advances in climate science are shown. The green dashed line shows the logarithmic slope of increase in the early 2000s. Courtesy Youngrak Cho and Whit Anderson, NOAA/GFDL. (Online version in colour.)

of simulated 'neurons' to mimic processing in the human brain. The brain does in fact make use of massive parallelism but with slow 'clock speeds'. While the initial excitement around neural networks in the 1960s (e.g. the 'perceptron' model of [9]) subsided as progress stalled due to the computational limitations of the time, these methods have undergone a remarkable resurgence in many scientific fields in recent years, as the algorithms underlying learning models are ideally suited to today's hardware for arithmetic. While the meteorological community may have initially been somewhat reticent (for reasons outlined in [10]), the last 2 or 3 years have witnessed a great efflorescence of the literature applying *machine learning* (ML)—as it is now called—in Earth system science. This special issue itself is evidence.

We argue in this article that this represents a sea change in computational Earth system science that rivals the von Neumann revolution. Indeed, some of the debates around ML today—pitting 'model-free' methods against 'interpretable AI'[3] for example—recapitulate those that took place in the 1940s and 1950s, when numerical meteorology was in its infancy, as we shall show.

In subsequent sections, we revisit some of the early history of numerical meteorology, to foreshadow some of the key debates around ML today (§2). In §3, we explore the issues around *Dennard scaling* leading to the current impasse in the speed of arithmetic *in silico*. In §4, we look at the state of the art of computational climate science, to see the limits of conventional approaches, and ways forward using learning algorithms. Finally, in §5, we look at the prospects and pitfalls in our current approaches, and outline a programme of research that may—or may not, as the jury is still out—revolutionize the field by judiciously combining physical insight, statistical learning and the harnessing of randomness.

---

[3] AI, or artificial intelligence, is a term we shall generally avoid here in favour of terms like ML, which emphasize the statistical aspect, without implying insight.

2. From patterns to physics: the von Neumann revolution

The development of dynamical meteorology as a science in the twentieth century has been told by both practitioners (e.g. [11,12]) and historians (e.g. [13,14]). The pioneering papers of Vilhelm Bjerknes (e.g. [15]) are often used as a signpost marking the beginning of dynamical meteorology, and we choose to use Vilhelm Bjerknes to highlight a fundamental dialectic that has enlivened the field since the beginning, and to this day, as we shall see below. Bjerknes pioneered the use of partial differential equations (the first use of the 'primitive equations') to represent the state of the circulation and its time evolution, but closed-form solutions were hard to come by. Numerical methods were also immature, and basic facts about the computational stability of the methods were yet unknown. The failed attempts by Richardson [16] involving thousands of human 'computers' are also well documented. One could consider it the first attempt at parallel processing, and at 64 000 'cores', it would be considered still quite cutting-edge today!. Bjerknes attempted a 'graphical calculus' using drawing tools but they were imprecise at 2–3 decimal digits of precision, as stated in [13, p. 87]. We note this as we shall return to the issue of numerical precision later in §3. Finally, abandoning the equations-based approach (global data becoming unavailable during war and its aftermath also played a role), Bjerknes reverted to making maps of air masses and their boundaries [15]. Forecasting was often based on a vast library of paper maps to find a map that resembled the present, and looking for the following sequence, what we would today recognize as Lorenz's analogue method [17]. Nebeker has commented on the irony that Bjerknes, who laid the foundations of theoretical meteorology, was also the one who developed practical forecasting tools 'that were neither algorithmic nor based on the laws of physics' [14].

We see here, in the sole person of Bjerknes, several voices in a conversation that continues to this day. One conceives of meteorology as a science, where everything can be derived from the first principles of classical fluid mechanics. A second approach is oriented specifically towards the goal of predicting the future evolution of the system (weather forecasts) and success is measured by forecast skill, by any means necessary. This could for instance be by creating approximate analogues to the current state of the circulation and relying on similar past trajectories to make an educated guess of future weather. One can have understanding of the system without the ability to predict; one can have skilful predictions innocent of any understanding. One can have a library of training data, and learn the trajectory of the system from that, at least in some approximate or probabilistic sense. If no analogue exists in the training data, no prediction is possible.

The physics-based approach came again to the forefront after the development of digital computing starting in the late 1940s, mainly centred around John von Neumann, Jule Charney, Joseph Smagorinsky and others at the Institute for Advanced Study (IAS) in Princeton. Once again, this story has been vividly told both by historians ([2,13]) and the participants ([6,18]), and we would not dare to try to tell it better here. Some of the pioneers are shown in this photograph taken by Smagorinsky himself (figure 2).

After Bjerknes' turn from physics to practical forecasting, which had become the coin of the realm, and weather forecasting was entirely based on what we would today called pattern recognition, performed by human meteorologists. What were later called 'subjective' forecasts depended a lot on the experience and recall of the meteorologist, who was generally not well-versed in theoretical meteorology, as Phillips remarks in [19]. Rapidly evolving events without obvious precursors in the data were often missed.

Charney's introduction of a numerical solution to the barotropic vorticity equation [20], and its execution on ENIAC, the first operational digital computer, essentially led to a complete reversal of fortune in the race between physics and pattern recognition. Programmable computers (where instructions were loaded in as well as data) came soon after, and the next landmark calculation of Phillips [21] came soon thereafter. Charney had spoken of 'climbing the ladder' of a hierarchy of models of increasing complexity, and the concept of the model hierarchy is something we shall revisit later as well. The Phillips 2-layer model was the next rung on the ladder.

It was not long before forecasts based on simplified numerical models outperformed subjective forecasts. Forecasting skill, measured by errors in the 500 hPa geopotential height, was clearly
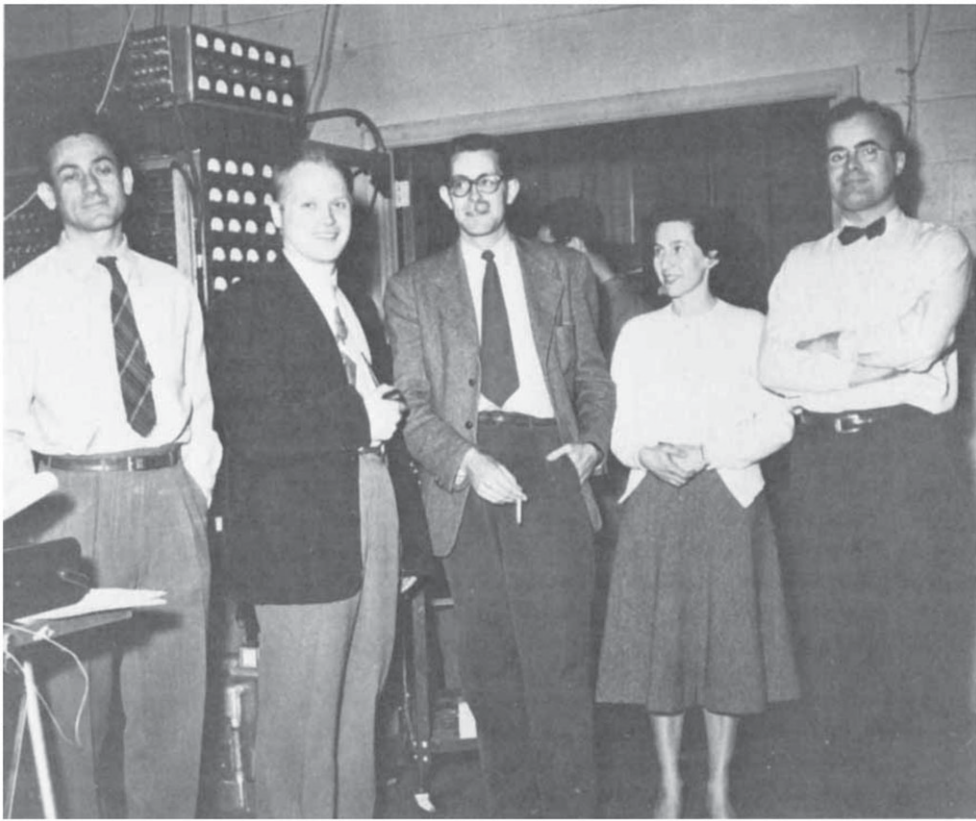
Some of the members of the IAS Meteorology group in 1952. Left to right: J. G Charney,
N. A. Phillips, G. Lewis, N. Gilbarg, G. W. Platzman (behind the camera: J Snagorinsky).
The IAS computer is in the background.

Figure 2. Some of the pioneers of computational Earth system science, photographed by Joe Smagorinsky. From [6].

better in the numerical forecasts after Phillips' breakthrough (see fig. 1 in [22]). Forecasts based
on integrating forward prognostic equations are now so taken for granted that Edwards remarks
in [13] that he found it hard to convince some of the scientists he met in the 1990s that it was
decades since the founding of theoretical meteorology before physics outdid simple heuristics
and theory-free pattern recognition in forecast skill.

The same tools, numerical models on a digital computer, were quickly also put to use to see
what long-term fluctuations of weather might look like, von Neumann's 'infinite forecast'. Models
of the ocean circulation had begun to appear (e.g. [23]) showing low-frequency (by atmospheric
weather standards) variations, and the significance of atmosphere-ocean coupling had been
observed [24]. The first coupled model of Manabe & Bryan [25] appeared in Smagorinsky's
laboratory. Smagorinsky had remarked that while the 'deterministic non-periodic' behaviour [5]
of geophysical fluids, which we now know as chaos, placed limits on predictability, the statistics of
weather fluctuations in the asymptotic limit could still be usefully studied [6], and his laboratory
soon became a centrepiece of this body of research. Within a decade, such models were the basic
tools of the trade for studying the asymptotic equilibrium response of Earth system to changes in
external forcing, the new field of computational climate science.

The practical outcomes of these studies, namely the response of the climate to anthropogenic
CO_2 emissions, raised public alarm with the publication of the Charney Report in 1979 [26].
Around the same time, a revolution in technology made computing cheap and ubiquitous, a mass
market commodity. This is seen in figure 1 as a transition from specialized 'vector' processors

to parallel clusters built of 'commodity' components. Computational climate science, now with planetary-scale societal ramifications, became a rapidly growing field expanding across many countries and laboratories, which could all now aspire to the scale of computing required to work out the implications of anthropogenic climate change. There was never enough computing: it was clear, for example, that clouds were a major unknown in the system (as noted already in the Charney Report) and were (and still are, see [27]) well below the resolution of models able to exploit the largest computers available. The models were resource-intensive, ready to soak up every last cycle of computation and byte of memory. A more sophisticated understanding of the Earth system also began to bring more processes into the simulations, now an integrated whole with physics, chemistry and biology components. We were still climbing Charney's hierarchy and adding complexity, but quite often the new components, were empirical, sometimes well-founded in observations in nature or in the laboratory, but often curve-fits to observations that were usually inadequate in time and space. Ecosystems and the turbulent planetary boundary layer are prime examples.

While the hierarchy ladder in the early days could be described theoretically in terms of the non-dimensional numbers that governed which terms to include or neglect in the equations, a standard approach in fluid dynamics, newer rungs in the ladder were governed by dimensional numbers in equations whose structure itself was empirically determined, with parameters poorly constrained by data, or indeed with no observable counterpart in nature. When these components are assembled into a coupled system, we are left with many errors and biases that are minimized during a further 'tuning' or calibration phase [28] varying some of these free parameters. It has been shown for example that a model's skill at reproducing the twentieth-century temperature record can be tuned up or down without violating fidelity to the process-level empirical constraints [29]. Some obvious 'fudge factors' such as flux adjustments [30] have been eliminated over the years. However, recalcitrant errors remain. The so-called double-ITCZ bias for example, which, despite glimmers of progress (see varied explanations in [31–33]) has remained stubbornly resistant to any amount of reformulation or tuning across many generations of climate models [34–36]. It is contended by many that no amount of fiddling with parameterizations can correct some of these long-standing biases, and only direct simulation resolving key features is likely to lead to progress (e.g. [37, Box 2]).

The revolution begun by von Neumann and Charney at the IAS, and the subsequent decades of exponential growth in computing shown in figure 1, have led to tremendous leaps forward as well as more ambiguous indicators of progress. What initially looked like a clear triumph of physics allied with computing and algorithmic advances now shows signs of stalling, as the accumulation of *detail* in the models—both in resolution and complexity—leads to some difficulty in the interpretation and mastery of model behaviour. Exponential growth curves in the real-world eventually turn sigmoid, and this may be true of Earth system modelling as well. More concretely, the exponential growth of figure 1 is also levelling off, with no immediate recourse, as we show below in §3. This is leading to a turn in computational climate science that may be no less far-reaching than the one wrought in Princeton in 1950.

3. The end of Dennard scaling: slower arithmetic, but more of it

The state of play in climate computing at the cusp of the ML explosion was reviewed in [38]. Earth system modelling faces certain intrinsic problems which keeps it from realizing even the potential of current computing. Considerable creative energy is being devoted to this problem, as outlined in [38], but we shall not revisit those efforts here. Suffice it to say that these issues are not simply a matter of better software engineering, but fundamental aspects of the algorithms of fluid dynamics. But for now, we restrict ourselves to the physical limits of current computing technology, and how we may adapt to life on the cusp of this technological transition.

Over almost four decades of the commodity microprocessor revolution alluded to earlier, there had been an extraordinary and relentless expansion in computing capacity, as seen in figure 1. This is quite often referred to as 'Moore's Law', based on Gordon Moore's observation that the

Table 1. The basis for Dennard scaling. With every shrink cycle in transistor fabrication, we gain the linear shrink factor κ in circuit switching speed, while maintaining power density constant. Adapted from table 1 in [39].

| device parameter	scaling factor
doping concentration	κ
transistor dimension	$1/\kappa$
voltage V	$1/\kappa$
current I	$1/\kappa$
capacitance C	$1/\kappa$
delay time per circuit VC/I	$1/\kappa$
power dissipation per circuit VI	$1/\kappa^2$
power density VI/A	1

number of transistors, the basic building blocks of digital computers, in a given area of silicon substrate doubles every 18 months, as we make miniaturization gains in complementary metal-oxide-silicon (CMOS) fabrication. Underlying Moore's Law is the physics of *Dennard scaling* [39]. The microprocessor revolution is fuelled by our ability to etch circuits at finer scale with each cycle in fabrication, resulting in faster switching (and thus the speed of an arithmetic operation) without any increase in power requirements, as shown in the last row of table 1.

As transistor dimensions continue to shrink, Dennard scaling is approaching its physical limits, as noted by [40]: for example at the current 5 nm fabrication dimension, transistors are about 30 atoms across. At the physical limit of CMOS, Dennard scaling breaks down [41]. First of all, clock speeds no longer increase. More critically, power density no longer stays constant, decreasing performance per watt. The increase in power dissipation per unit area also implies increased heat dissipation, leading to the phenomenon of 'dark silicon' [42], where large sections, often more than 50% of the chip surface have to be turned off (no current) at any given moment, in order to stay within safe thermal limits. This means that the number of operations per cycle is actually well below what is theoretically possible.

The sigmoid taper of exponential growth in microprocessor speed is clearly seen in figure 3. While the number of transistors still remains on a doubling trajectory, the actual arithmetic speed (chip frequency in MHz) and wattage have levelled off. The transistors now go to increase the number of logical cores on the chip. As a result, arithmetic speed is now static or may even become slower to maintain the power and cooling envelope, but more arithmetic can be concurrently executed. The difficulties of continuing to increase resolution under these constraints have been described in [43].

So, certain physical limits of current computing have been reached. While there are efforts to imagine radically novel computing methods, including 'non-von Neumann' methods such as quantum or neuromorphic computing [44], those are still on the drawing board. Interesting forays edging towards 'inexact computing' [45,46] remain in exploratory stages as well.

It is clear, however, that absent some unforeseen advance, current approaches to Earth system modelling will not be advancing in the same fashion as the last several decades. An analysis of the potential performance of a global cloud-resolving model on an exascale machine in [47] showed that it would potentially take an entire such machine to run that model at a speed of 1 simulated year per day (SYPD). Climate science requires not just computing *capability* (SYPD) but also *capacity*, the ability to run multiple copies of the model to sample various kinds of uncertainty. A suite of typical climate experiments aimed at studying the response of the model to various perturbations is measured in $\mathcal{O}(10\,000)$ years. Simply developing and testing a model, including the calibration described in §2, may cost five times as much [48]. Similar costs obtain in establishing a model's predictive skill by running a suite of retrospective hindcasts. Furthermore, the model studied in [47] has vastly less *complexity* (defined in [49] as the number

42 years of microprocessor trend data
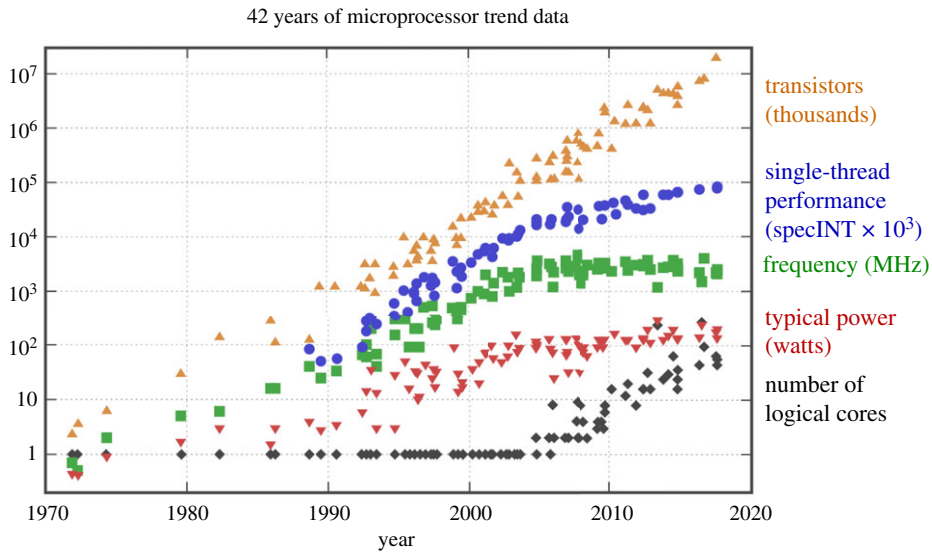
original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham,
K. Olukotun, L. Hammond, and C. Batten New plot and data collected for 2010–2017 by K. Rupp

Figure 3. Dennard scaling tails off at the end of four decades of microprocessor miniaturization. From 42 years of microprocessor trend data (https://www.karlrupp.net/2018/02/42-years-of-microprocessor-trend-data/), courtesy Karl Rupp. (Online version in colour.)

of distinct physical variables simulated by a model) than a typical workhorse model. It is clear that the additional of detail, in resolution and complexity, to models cannot continue as before. A fundamental rethinking of the decades-long climb up the Charney ladder is long overdue.

Just around the time the state of play in climate computing was reviewed in [38], the contours of the revival of ML using artificial neural networks (ANNs) were beginning to take shape. Deep learning (DL) using multiple neuronal layers were showing significant skill in many domains. As noted in §1, ANNs existed alongside the physics-based models of von Neumann and Charney for decades, but may have languished as the computing power and parallelism were not available. The new processors emerging at the right of figure 3 in the twilight of Dennard scaling, are ideally suited to ML: the typical DL computation consists of dense linear algebra, scalable almost at will, able to reduce memory bandwidth at reduced precision without loss of performance. Processors such as the graphical processing unit (GPU) and most especially the TPU (tensor processing unit) showed themselves capable of running a typical DL workload at close to the theoretical maximum performance of the chip [50]. There are many challenges to executing conventional equation-based arithmetic on these chips, not least of which is their low precision, often as low as 3 digits, the same as for the manual arithmetic that limited Bjerknes, see §2. While continuing to explore low-precision arithmetic (e.g. [51]), we have begun to explore ML itself in the arsenal of Earth system modelling. We turn now in §4 to an assess the potential of ML to show us a way out of the current computing impasse.

4. Learning physics from data

The articles in this special issue form a wide spectrum representing the state of the art in the use of ML in Earth system science, and we do not propose to offer a broad or comprehensive review. Instead, we aim to demonstrate using a judicious selection of a few threads of research from the current literature how Earth system modelling's turn towards ML reprises some of the fundamental issues that arose in the pioneering era, outlined above in §2.

Recall Bjerknes's turn away from theoretical meteorology upon finding that the tools at his disposal were not adequate to the task of making predictions from theory. It is possible that the computational predicament we now find ourselves in, as outlined above in §3, is a historical parallel, and we too shall turn towards practical, theory-free predictions. An example would be the prediction of precipitation from a sequence of radar images [52], where 'optical flow' (essentially, extrapolating various optical transformations such as translation, rotation, stretching, intensification) is compared and found competitive with persistence and short term model-based forecasts. Similarly, ML methods have shown exceptional forecast skill at longer timescales, including breaking through the 'spring barrier' (the name given to a dramatic reduction in forecast skill in models initialized prior to boreal spring) in ENSO predictability [53]. Interestingly, the Lorenz analogue method [17] also shows longer term ENSO skill (with no spring barrier) compared to dynamical models [54], a throwback to the early days of forecasting as described in §2. These and other successes in purely 'data-driven' forecasting (though the ENSO papers use model output as training data) have led to speculation in the media that ML might indeed make physics-based forecasting obsolete (see for example, Could Machine Learning Replace the Entire Weather Forecast System?[4] in *HPCWire*). ML methods (in this case, recurrent neural networks) have also shown themselves capable of reproducing a time series from canonical chaotic systems with predictability beyond what dynamical systems theory would suggest, e.g. [55] (which indeed explicitly makes a claim to be 'model-free'), [56]. Does this mean we have come full circle on the von Neumann revolution, and return to forecasting from pattern recognition rather than physics? The answer of course is contingent on the presumption that the training data in fact is comprehensive and samples all possible states of the system. For the Earth system, this is a dubious proposition, as there is variability on all time scales, including those longer than the observational record itself. A key issue for all data-driven approaches is that of *generalizability* beyond the confines of the training data.

Turning to climate from weather, we look at the aspects of Earth system models that while broadly based on theory, are structured around an empirical formulation rather than from first principles. These are areas obviously ripe for a more directly data-driven approach. These often are based around the parameterized components of the model that deal with 'sub-gridscale' processes operating below the truncation imposed by discretization. A key feature of geophysical fluid flow is the three-dimensional turbulence cascade continuous from planetary scale down to the Kolmogorov length scale (e.g. [57, fig. 1]), which must be truncated somewhere for a discrete numerical representation. ML-based representation of sub-grid turbulence is one area receiving considerable attention now [58]. Other sub-gridscale aspects particular to Earth system modelling where ML could play a role include radiative transfer and the representation of clouds, which now has a rich literature, including in this special issue. As noted just above, 'data-driven' methods often rely on the output of models. This is particularly so for the case of sub-gridscale physics, where learning often involves *emulating* the behaviour of models in which the phenomenon is resolved.

ANNs have the immediate advantage of often being considerably faster than the component they replace [59,60], thus directly responsive to the computational challenge laid down in §3. They additionally have the advantage of being differentiable, which sub-grid physics often is not: this has a key advantage in the use of data assimilation (DA) techniques to constrain model trajectory. The formal equivalences between DA and ML have been demonstrated in [61], for instance between adjoint models used in DA and the back-propagation technique in ML. Furthermore, the calibration procedures of [28] are very inefficient with full ESMs, and may be considerably accelerated using *emulators* derived by learning [62].

ML nonetheless poses a number of challenging questions that we are now actively addressing. The usual problems of whether the data is representative and comprehensive, and on generalizability of the learning, continue to apply. There is a conundrum in deciding where

---

[4]https://www.hpcwire.com/2020/04/27/could-machine-learning-replace-the-entire-weather-forecast-system/, retrieved 27 July 2020.

the boundary between being physical knowledge driven and data-driven lies. We outline some key questions being addressed in the current literature, and not least in this special issue. We take as a starting point, a particular model component (such as atmospheric convection or ocean turbulence) that is now being augmented by learning.

Do we take the structure of the underlying system as it is now, and use learning as a method of reducing parametric uncertainty? Emerging methods potentially do allow us to treat structural and parametric error on a common footing (e.g. [63]), but we still may choose to go either structure-free, or attempt to discover the structure itself.

If we divest ourselves of the underlying structure of equations, we have a number of issues to address. As the learning is only as good as the training data, we may find that the resulting ANN violates some basic physics (such as conservation laws), or does not generalize well [64]. This can be addressed by a suitable choice of basis in which to do the learning, as subsequent work [65] shows. A similar consideration was seen in [66], which when trained on current climate failed to generalize to a warmer climate, but the loss of generalizability could be addressed by a suitable choice of input basis, using lapse rates rather than temperatures, for example (P O'Gorman 2020, personal communication).

Ideally, we would like to go much further and actually learn the underlying physics. There have been attempts to learn the underlying equations for well-known systems ([67,68]) and efforts underway in climate modelling as well, to learn the underlying structure of parameterizations from data. In this context, 'learning the physics' means, for instance, starting from data and writing down closed-form equations for the effect of unresolved physics on resolved-scale tendencies. This is by its nature seeking a *sparse* representation, as a closed-form equation is one can write down, and must have a limited number of terms on the right-hand side. For tests of the canonical systems in the articles cited above, this is guaranteed, as the systems we are learning are known to be sparse in this sense. The challenge is to attempt to learn it for systems where a sparse representation is not known or guaranteed to exist [65] is an interesting step forward in this direction. However, the current methods still face some limitations. In particular the methods only seek representations within a library of possible eigenfunctions that can be used. Ingenuity and physical insight still play a role in finding the basis that yields a sparse, or parsimonious, representation.

Finally, we pose the problem of coupling. We have noted earlier the problem of calibration of models, which is done first at the component level, to bring each individual process within observational constraints, and then in a second stage of calibration against systemwide constraints, such as top of atmosphere radiative balance [28]. The issue of the stability of ANNs when integrated into a coupled system is also under active study at the moment (e.g. [69]). The question of whether tightly coupled subsystems should be jointly or separately learned still remains an open question.

5. Climbing down the ladder: prospects for ML-based Earth system models

We have highlighted in this article a historical progression in Earth system modelling, which we describe here as the addition of *detail* to our models, in the form of resolution and complexity. While this has resulted in tremendous leaps in understanding, there have been some aspects where we have failed to advance for decades. In most accounts, this can be traced to failures in representation of the form of sub-gridscale phenomena such as clouds. The solution may come when we in fact no longer need to seek representations, as we shall be resolving such phenomena directly. We have noted possible setbacks to this approach given physical limits on computing technology.

We have gazed into the computational abyss before us, and seen how ML may offer ways to adapt to what today's machines do best, which is statistical ML. It is a transition in our approach to predicting the Earth system that is potentially as far-reaching as that of von Neumann and Charney, algorithms that learn rather than do what they are told. How this will unfold is yet to be seen.

At first glance, it may appear as though we are turning our backs on the von Neumann revolution, going back from physics to simply seeking and following patterns in data. While such black-box approaches may indeed be used for certain activities, we are seeing many attempts to go beyond those, and ensure that the learning algorithms do indeed respect physical constraints even if not present in the data.

Even 'data-driven' methods quite often rely on the output of models, as noted at various points in the text: indeed reconstructed historical observations (reanalysis datasets) are reliant on models to produce a physically consistent multivariate dataset out of diverse observations. The observed climate record is also too short for unambiguously separating natural from forced variations: we rely on models to supplement that record. It is unlikely that raw data alone will suffice to train ML algorithms. Models that operate at the limit of resolution on the largest computers available will indeed be among the tools of the trade, but they will likely not be the workhorses of modelling, which require many runs to see how they respond to perturbations. Building reduced order models, lower in resolution and complexity, may well become one of the principal uses of the extreme-scale models.

It was said of Charney's first computation that he knew what to leave out for a feasible calculation. Since then, computing power has grown, and the models have climbed the ladder of a hierarchy of complexity. But as Held has remarked [70], it is necessary to descend the hierarchy as well, to pass from simulation to understanding. If ML-based modelling needs a manifesto, it may be this: to *learn from data not just patterns, but simpler models*, climbing down Charney's ladder. The path down the ladder will not be an exact reprise, but will include what we learned on the rungs on the way up. The vision is that these models will leave out the details not needed in an understanding of the underlying system, and learning algorithms will find for us underlying 'slow manifolds' [71], and the basis variables that yield a succinct, or parsimonious, mapping. That is the challenge before us.

References

1. Balaji V. 2013 Scientific computing in the age of complexity. *XRDS* **19**, 12–17. (doi:10.1145/2425676.2425684)
2. Dahan-Dalmedico A. 2001 History and epistemology of models: meteorology (1946–1963) as a case study. *Arch. Hist. Exact Sci.* **55**, 395–422. (doi:10.1007/s004070000032)

[5]https://github.com/karlrupp/microprocessor-trend-data, retrieved 27 July 2020.

3. Bauer P, Thorpe A, Brunet G. 2015 The quiet revolution of numerical weather prediction. *Nature* **525**, 47–55. (doi:10.1038/nature14956)

4. Cressman GP. 1996 The origin and rise of numerical weather prediction. In *Historical essays on meteorology 1919–1995: the diamond anniversary history volume of the American meteorological society* (ed JR Fleming), pp. 21–39. Boston, MA: American Meteorological Society.

5. Lorenz EN. 1963 On the predictability of hydrodynamic flow. *Trans. N.Y. Acad. Sci.* **25**, 409–432. (doi:10.1111/j.2164-0947.1963.tb01464.x)

6. Smagorinsky J. 1983 The beginnings of numerical weather prediction and general circulation modeling: early recollections. In *Advances in Geophysics* (ed. B Saltzman), vol. 25 of *Theory of climate*, pp. 3–37. New York, NY: Elsevier.

7. Sorenson RP. 2011 Eunice Foote's pioneering research on CO_2 and climate warming. *Search Discovery*. Article #70092. (http://www.searchanddiscovery.com/documents/2011/70092sorenson/ndx_sorenson.pdf)

8. Manabe S, Wetherald RT. 1975 The effects of doubling the CO_2 concentration on the climate of a general circulation model. *J. Atmos. Sci.* **32**, 3–15. (doi:10.1175/1520-0469(1975)032<0003:TEODTC>2.0.CO;2)

9. Block HD, Knight Jr B, Rosenblatt F. 1962 Analysis of a four-layer series-coupled perceptron. II. *Rev. Mod. Phys.* **34**, 135–142. (doi:10.1103/RevModPhys.34.135)

10. Hsieh WW, Tang B. 1998 Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bull. Am. Meteorol. Soc.* **79**, 1855–1870. (doi:10.1175/1520-0477(1998)079<1855:ANNMTP>2.0.CO;2)

11. Lorenz EN. 1967 *The nature and theory of the general circulation of the atmosphere*, vol. 218. Geneva, Switzerland: World Meteorological Organization.

12. Held IM. 2019 100 years of progress in understanding the general circulation of the atmosphere. *Meteor. Monogr.* **59**, 6.1–6.23. (doi:10.1175/AMSMONOGRAPHS-D-18-0017.1)

13. Edwards P. 2010 *A vast machine: computer models, climate data, and the politics of global warming*. Cambridge, MA: The MIT Press.

14. Nebeker F. 1995 *Calculating the weather: meteorology in the 20th century*. Amsterdam, The Netherlands: Elsevier.

15. Bjerknes V. 1921 The meteorology of the temperate zone and the general atmospheric circulation. *Mon. Weather Rev.* **49**, 1–3. (doi:10.1175/1520-0493(1921)49<1:TMOTTZ>2.0.CO;2)

16. Richardson LF. 2007 *Weather prediction by numerical process*. Cambridge, UK: Cambridge University Press.

17. Lorenz EN. 1969 Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.* **26**, 636–646. (doi:10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2)

18. Platzman GW. 1979 The ENIAC computations of 1950–gateway to numerical weather prediction. *Bull. Am. Meteorol. Soc.* **60**, 302–312. (doi:10.1175/1520-0477(1979)060<0302:TECOTN>2.0.CO;2)

19. Phillips NA. 1990 The emergence of quasi-geostrophic theory. In *The atmosphere – A challenge*, pp. 177–206. Berlin, Germany: Springer.

20. Charney J, Fjortoft R, von Neumann J. 1950 Numerical integration of the barotropic vorticity equation. *Tellus* **2**, 237–254. (doi:10.3402/tellusa.v2i4.8607)

21. Phillips NA. 1956 The general circulation of the atmosphere: a numerical experiment. *Q. J. R. Meteorol. Soc.* **82**, 123–164. (doi:10.1002/qj.49708235202)

22. Shuman FG. 1989 History of numerical weather prediction at the national meteorological center. *Weather Forecast* **4**, 286–296. (doi:10.1175/1520-0434(1989)004<0286:HONWPA>2.0.CO;2)

23. Munk WH. 1950 On the wind-driven ocean circulation. *J. Meteorol.* **7**, 80–93. (doi:10.1175/1520-0469(1950)007<0080:OTWDOC>2.0.CO;2)

24. Namias J. 1959 Recent seasonal interactions between north pacific waters and the overlying atmospheric circulation. *J. Geophys. Res.* **64**, 631–646. (doi:10.1029/JZ064i006p00631)

25. Manabe S, Bryan K. 1969 Climate calculations with a combined ocean-atmosphere model. *J. Atmos. Sci.* **26**, 786–789. (doi:10.1175/1520-0469(1969)026<0786:CCWACO>2.0.CO;2)

26. Charney JG, Arakawa A, Baker DJ, Bolin B, Dickinson RE, Goody RM, Leith CE, Stommel HM, Wunsch CI. 1979 Carbon dioxide and climate: a scientific assessment.

27. Schneider T, Teixeira J, Bretherton CS, Brient F, Pressel KG, Schär C, Siebesma AP. 2017 Climate goals and computing the future of clouds. *Nat. Clim. Change* **7**, 3–5. (doi:10.1038/nclimate3190)

28. Hourdin F *et al.* 2017 The art and science of climate model tuning. *Bull. Am. Meteorol. Soc.* **98**, 589–602. (doi:10.1175/BAMS-D-15-00135.1)

29. Golaz J-C, Horowitz LW, Levy H. 2013 Cloud tuning in a coupled climate model: Impact on 20th century warming. *Geophys. Res. Lett.* **40**, 2246–2251. (doi:10.1002/grl.50232)

30. Shackley S, Risbey J, Stone P, Wynne B. 1999 Adjusting to policy expectations in climate change modeling. *Clim. Change* **43**, 413–454. (doi:10.1023/A:1005474102591)

31. Xiang B, Zhao M, Held IM, Golaz J-C. 2017 Predicting the severity of spurious 'double ITCZ' problem in CMIP5 coupled models from AMIP simulations. *Geophys. Res. Lett.* **44**, 1520–1527. (doi:10.1002/2016GL071992)

32. Wittenberg AT *et al.* 2018 Improved simulations of tropical pacific annual-mean climate in the GFDL FLOR and HiFLOR coupled GCMs. *J. Adv. Model. Earth Syst.* **10**, 3176–3220. (doi:10.1029/2018MS001372)

33. Woelfle MD, Bretherton CS, Hannay C, Neale R. 2019 Evolution of the Double-ITCZ bias through CESM2 development. *J. Adv. Model. Earth Syst.* **11**, 1873–1893. (doi:10.1029/2019MS001647)

34. Lin J-L. 2007 The double-ITCZ problem in IPCC AR4 coupled GCMs: Ocean–atmosphere feedback analysis. *J. Clim.* **20**, 4497–4525. (doi:10.1175/JCLI4272.1)

35. Li G, Xie S-P. 2014 Tropical biases in CMIP5 multimodel ensemble: the excessive equatorial Pacific cold tongue and double ITCZ problems. *J. Clim.* **27**, 1765–1780. (doi:10.1175/JCLI-D-13-00337.1)

36. Tian B, Dong X. 2020 The Double-ITCZ Bias in CMIP3, CMIP5, and CMIP6 models based on annual mean precipitation. *Geophys. Res. Lett.* **47**, e2020GL087232. (doi:10.1029/2020GL087232)

37. Palmer T, Stevens B. 2019 The scientific challenge of understanding and estimating climate change. *Proc. Natl Acad. Sci. USA* **116**, 24 390–24 395. (doi:10.1073/pnas.1906691116)

38. Balaji V. 2015 Climate computing: the state of play. *Comput. Sci. Eng.* **17**, 9–13. (doi:10.1109/MCSE.2015.109)

39. Dennard RH, Gaensslen FH, Rideout VL, Bassous E, LeBlanc AR. 1974 Design of ion-implanted MOSFETs with very small physical dimensions. *IEEE J. Solid-State Circuits* **9**, 256–268. (doi:10.1109/JSSC.1974.1050511)

40. Markov I. 2014 Limits on fundamental limits to computation. *Nature* **512**, 147–154. (doi:10.1038/nature13570)

41. Chien AA, Karamcheti V. 2013 Moore's Law: the first ending and a new beginning. *Computer* **12**, 48–53. (doi:10.1109/MC.2013.431)

42. Esmaeilzadeh H, Blem E, Amant RS, Sankaralingam K, Burger D. 2013 Power challenges may end the multicore era. *Commun. ACM* **56**, 93–102. (doi:10.1145/2408776.2408797)

43. Schär C *et al.* 2020 Kilometer-Scale climate models: prospects and challenges. *Bull. Am. Meteorol. Soc.* **101**, E567–E587. (doi:10.1175/BAMS-D-18-0167.1)

44. Conte TM, DeBenedictis EP, Gargini PA, Track E. 2017 Rebooting computing: the road ahead. *Computer* **50**, 20–29. (doi:10.1109/MC.2017.8)

45. Palem KV. 2014 Inexactness and a future of computing. *Phil. Trans. R. Soc. A* **372**, 20130281. (doi:10.1098/rsta.2013.0281)

46. Palmer T. 2015 Modelling: build imprecise supercomputers. *Nature News* **526**, 32–33. (doi:10.1038/526032a)

47. Neumann P *et al.* 2019 Assessing the scales in numerical weather and climate predictions: will exascale be the rescue? *Phil. Trans. R. Soc. A* **377**, 20180148. (doi:10.1098/rsta.2018.0148)

48. Adcroft A *et al.* 2019 The GFDL global ocean and sea ice model OM4.0: model description and simulation features. *J. Adv. Model. Earth Syst.* **11**, 3167–3211. (doi:10.1029/2019MS001726)

49. Balaji V *et al.* 2017 CPMIP: measurements of real computational performance of Earth system models in CMIP6. *Geoscientific Model Dev.* **10**, 19–34. (doi:10.5194/gmd-10-19-2017)

50. Jouppi NP *et al.* 2017 In-datacenter performance analysis of a tensor processing unit. In *Proc. of the 44th Annual Int. Symp. on Computer Architecture*, pp. 1–12.

51. Chantry M, Thornes T, Palmer T, Düben P. 2018 Scale-selective precision for weather and climate forecasting. *Mon. Weather Rev.* **147**, 645–655. (doi:10.1175/MWR-D-18-0308.1)

52. Agrawal S, Barrington L, Bromberg C, Burge J, Gazen C, Hickey J. 2019 Machine learning for precipitation nowcasting from radar images. http://arxiv.org/abs/1912.12132.

53. Ham Y-G, Kim J-H, Luo J-J. 2019 Deep learning for multi-year ENSO forecasts. *Nature* **573**, 568–572. (doi:10.1038/s41586-019-1559-7)

54. Ding H, Newman M, Alexander MA, Wittenberg AT. 2019 Diagnosing secular variations in retrospective ENSO seasonal forecast skill using CMIP5 model-analogs. *Geophys. Res. Lett.* **46**, 1721–1730. (doi:10.1029/2018GL080598)

55. Pathak J, Hunt B, Girvan M, Lu Z, Ott E. 2018 Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir computing approach. *Phys. Rev. Lett.* **120**, 024102. (doi:10.1103/PhysRevLett.120.024102)

56. Chattopadhyay A, Hassanzadeh P, Subramanian D. 2020 Data-driven predictions of a multiscale Lorenz 96 chaotic system using machine-learning methods: reservoir computing, artificial neural network, and long short-term memory network. *Nonlinear Process. Geophys.* **27**, 373–389. (doi:10.5194/npg-27-373-2020)

57. Nastrom G, Gage KS. 1985 A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft. *J. Atmos. Sci.* **42**, 950–960. (doi:10.1175/1520-0469(1985)042<0950:ACOAWS>2.0.CO;2)

58. Duraisamy K, Iaccarino G, Xiao H. 2019 Turbulence modeling in the age of data. *Annu. Rev. Fluid Mech.* **51**, 357–377. (doi:10.1146/annurev-fluid-010518-040547)

59. Krasnopolsky V, Fox-Rabinovitz M, Chalikov D. 2005 New approach to calculation of atmospheric model physics: accurate and fast neural network emulation of longwave radiation in a climate model. *Mon. Weather Rev.* **133**, 1370–1383. (doi:10.1175/MWR2923.1)

60. Rasp S, Pritchard MS, Gentine P. 2018 Deep learning to represent subgrid processes in climate models. *Proc. Natl Acad. Sci. USA* **115**, 9684–9689. (doi:10.1073/pnas.1810286115)

61. Kovachki NB, Stuart AM. 2019 Ensemble Kalman inversion: a derivative-free technique for machine learning tasks. *Inverse Prob.* **35**, 095005. (doi:10.1088/1361-6420/ab1c3a)

62. Williamson D, Goldstein M, Allison L, Blaker A, Challenor P, Jackson L, Yamazaki K. 2013 History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Clim. Dyn.* **41**, 1703–1729. (doi:10.1007/s00382-013-1896-4)

63. Williamson D, Blaker AT, Hampton C, Salter J. 2015 Identifying and removing structural biases in climate models with history matching. *Clim. Dyn.* **45**, 1299–1324. (doi:10.1007/s00382-014-2378-z)

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

65. Zanna L, Bolton T. 2020 Data-driven equation discovery of ocean mesoscale closures. *Geophys. Res. Lett.* **47.17**, e2020GL088376.

66. O'Gorman PA, Dwyer JG. 2018 Using machine learning to parameterize moist convection: potential for modeling of climate, climate change, and extreme events. *J. Adv. Model. Earth Syst.* **10**, 2548–2563. (doi:10.1029/2018MS001351)

67. Schmidt M, Lipson H. 2009 Distilling free-form natural laws from experimental data. *Science* **324**, 81–85. (doi:10.1126/science.1165893)

68. Brunton SL, Proctor JL, Kutz JN. 2016 Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **113**, 3932–3937. (doi:10.1073/pnas.1517384113)

69. Brenowitz ND, Beucler T, Pritchard M, Bretherton CS. 2020 Interpreting and stabilizing machine-learning parametrizations of convection. http://arxiv.org/abs/2003.06549 [physics].

70. Held I. 2005 The gap between simulation and understanding in climate modeling. *Bull. Am. Met. Soc.* **86**, 1609–1614. (doi:10.1175/BAMS-86-11-1609)

71. Lorenz EN. 1992 The slow manifold 'What Is It? *J. Atmos. Sci.* **49**, 2449–2451. (doi:10.1175/1520-0469(1992)049<2449:TSMII>2.0.CO;2)

72. Rupp K. 2020 48 years of microprocessor trend data. Zenodo repository. (doi:10.5281/zenodo.3947824)