Deep learning to represent sub-grid processes in climate models

Stephan Rasp[a,b], **Michael S. Pritchard**[b], **and Pierre Gentine**[c]

[a]Meteorological Institute, Ludwig-Maximilian-University, 80333 Munich, Germany; [b]Department of Earth System Science, University of California, Irvine, CA 92697, USA; [c]Columbia University, Department of Earth and Environmental Engineering, Earth Institute, and Data Science Institute, New York, NY 10027, USA

This manuscript was compiled on August 13, 2018

The representation of nonlinear sub-grid processes, especially clouds, has been a major source of uncertainty in climate models for decades. Cloud-resolving models better represent many of these processes and can now be run globally but only for short-term simulations of at most a few years because of computational limitations. Here we demonstrate that deep learning can be used to capture many advantages of cloud-resolving modeling at a fraction of the computational cost. We train a deep neural network to represent all atmospheric sub-grid processes in a climate model by learning from a multi-scale model in which convection is treated explicitly. The trained neural network then replaces the traditional sub-grid parameterizations in a global general circulation model in which it freely interacts with the resolved dynamics and the surface-flux scheme. The prognostic multi-year simulations are stable and closely reproduce not only the mean climate of the cloud-resolving simulation but also key aspects of variability, including precipitation extremes and the equatorial wave spectrum. Furthermore, the neural network approximately conserves energy despite not being explicitly instructed to. Finally, we show that the neural network parameterization generalizes to new surface forcing patterns but struggles to cope with temperatures far outside its training manifold. Our results show the feasibility of using deep learning for climate model parameterization. In a broader context, we anticipate that data-driven Earth System Model development could play a key role in reducing climate prediction uncertainty in the coming decade.

Climate modeling | Sub-grid parameterization | Convection | Deep learning

Many of the atmosphere's most important processes occur on scales smaller than the grid resolution of current climate models, around 50–100 km horizontally. Clouds, for example, can be as small as a few hundred meters; yet they play a crucial role in determining the earth's climate by transporting heat and moisture, reflecting and absorbing radiation, and producing rain. Climate change simulations at such fine resolutions are still many decades away (1). To represent the effects of such sub-grid processes on the resolved scales, physical approximations—called *parameterizations*—have been heuristically developed and tuned to observations over the last decades (2). However, owing to the sheer complexity of the underlying physical system, significant inaccuracies persist in the parameterization of clouds and their interaction with other processes, such as boundary-layer turbulence and radiation (1, 3, 4). These inaccuracies manifest themselves in stubborn model biases (5–7) and large uncertainties about how much the earth will warm as a response to increased greenhouse gas concentrations (1, 8, 9). To improve climate predictions, therefore, novel, objective and computationally efficient approaches to sub-grid parameterization development are urgently needed.

Cloud-resolving models (CRMs) alleviate many of the issues related to parameterized convection. At horizontal resolutions of at least 4 km deep convection can be explicitly treated (10), which substantially improves the representation of land-atmosphere coupling (11, 12), convective organization (13) and weather extremes. Further increasing the resolution to a few hundred meters allows for the direct representation of the most important boundary-layer eddies, which form shallow cumuli and stratocumuli. These low clouds are crucial for the Earth's energy balance and the cloud-radiation feedback (14). CRMs come with their own set of tuning and parameterization decisions but the advantages over coarser models are substantial. Unfortunately, global CRMs will be too computationally expensive for climate change simulations for many decades (1). Short-range simulations covering periods of months or even a few years, however, are beginning to be feasible and are in development at modeling centers around the world (15–18).

In this study, we explore whether deep learning can provide an objective, data-driven approach to utilize high-resolution modeling data for climate model parameterization. The paradigm shift from heuristic reasoning to machine learning has transformed computer vision and natural language processing over the last few years (19) and is starting to impact more traditional fields of science. The basic building blocks of deep learning are deep neural networks which consist of several inter-connected layers of nonlinear nodes (20). They are capable of approximating arbitrary nonlinear functions (21) and can easily be adapted to novel problems. Furthermore, they can handle large datasets during training and provide fast

Significance Statement

Current climate models are too coarse to resolve many of the atmosphere's most important processes. Traditionally, these sub-grid processes are heuristically approximated in so-called parameterizations. However, imperfections in these parameterizations, especially for clouds, have impeded progress towards more accurate climate predictions for decades. Cloud-resolving models alleviate many of the gravest issues of their coarse counterparts but will remain too computationally demanding for climate change predictions for the foreseeable future. Here we use deep learning to leverage the power of short-term cloud-resolving simulations for climate modeling. Our data-driven model is fast and accurate thereby showing the potential of novel, machine learning-based approaches to climate model development.

SR conducted model simulations and data analysis with help from MSP. MSP and PG conceived the project. SR led the writing of the paper with input from MSP and PG.

The authors declare no conflict of interest.

[2]To whom correspondence should be addressed. E-mail: s.rasp@lmu.de

predictions at inference time. All of these traits make deep learning an attractive approach for the problem of sub-grid parameterization.

Extending on previous offline or single-column neural network cumulus parameterization studies (22–24), here we take the essential step of implementing the trained neural network in a global climate model and running a stable, prognostic multi-year simulation. To show the potential of this approach we compare key climate statistics between the deep learning-powered model and its training simulation. Furthermore, we tackle two crucial questions for a climate model implementation: first, does the neural network parameterization conserve energy; and second, to what degree can the network generalize outside of its training climate? We conclude by highlighting crucial challenges for future data-driven parameterization development.

Model and neural network setup. Our base model is the super-parameterized Community Atmosphere Model v3.0 (SPCAM) (25) in an aquaplanet setup (see SI Appendix for details). The sea surface temperatures (SSTs) are fixed and zonally invariant with a realistic equator-to-pole gradient (26). The model has a full diurnal cycle but no seasonal variation. The horizontal grid spacing of the global circulation model (GCM) is approximately 2 degrees with 30 vertical levels. The GCM time step is 30 minutes. In super-parameterization, a two-dimensional CRM is embedded in each global circulation model grid column (27). This CRM explicitly resolves deep convective clouds and includes parameterizations for small-scale turbulence and cloud microphysics. In our setup, we use eight 4 km-wide columns with a CRM time step of 20 seconds, after Ref. (28). For comparison, we also run a simulation with the traditional parameterization suite (CTRLCAM) that is based on an undilute plume parameterization of moist convection. CTRLCAM exhibits many typical problems associated with traditional sub-grid cloud parameterizations: a double inter-tropical convergence zone (ITCZ) (5); too much drizzle and missing precipitation extremes; and an unrealistic equatorial wave spectrum with a missing Madden-Julian-Oscillation (MJO). In contrast, SPCAM captures the key benefits of full three-dimensional CRMs in improving the realism all of these issues with respect to observations (29–31). In this context, a key test for a neural network parameterization is whether it learns sufficiently from the explicitly resolved convection in SPCAM to remedy such problems while being computationally more affordable.

Analogous to a traditional parameterization, the task of the neural network is to predict the sub-grid tendencies as a function of the atmospheric state at every time step and grid column (Table S1). Specifically, we selected the following input variables: the temperature $T(z)$, specific humidity $Q(z)$ and wind profiles $V(z)$, surface pressure P_s, incoming solar radiation S_{in} and the sensible H and latent heat fluxes E. These variables mirror the information received by the CRM and radiation scheme with a few omissions (SI Appendix). The output variables are: the sum of the CRM and radiative heating rates ΔT_{phy}, the CRM moistening rate ΔQ_{phy}, the net radiative fluxes at the top of atmosphere and surface F_{rad} and precipitation P. The input and output variables are stacked to vectors $\mathbf{x} = [T(z), Q(z), V(z), P_s, S_{\text{in}}, H, E]^T$ with length 94 and $\mathbf{y} = [\Delta T_{\text{phy}}(z), \Delta Q_{\text{phy}}(z), F_{\text{rad}}, P]^T$ with length 65 and normalized to have similar orders of magnitude (SI Appendix). We omit condensed water to reduce the complexity of the problem (see Discussion). Furthermore, there is no momentum transport in our version of SPCAM. Informed by our previous sensitivity tests (24) we use one year of SPCAM simulation as training data for the neural network, amounting to around 140 million training samples.

The neural network itself $\hat{\mathbf{y}} = \mathcal{N}(\mathbf{x})$ is a nine layer deep, fully-connected network with 256 nodes in each layer. In total, the network has around half a million parameters that are optimized to minimize the mean squared error between the network's predictions $\hat{\mathbf{y}}$ and the training targets $\mathbf{y}$ (see SI Appendix). This neural network architecture is informed by our previous sensitivity tests (24). Using deep rather than shallow networks has two main advantages: first, deeper, larger networks achieve lower training losses; and second, deep networks proved more stable in the prognostic simulations (for details see SI Appendix and Fig. S1). Unstable modes and unrealistic artifacts have been the main issue in previous studies that used shallow architectures (22, 23).

Once trained, the neural network replaces the super-parameterization's CRM as well as the radiation scheme in CAM (NNCAM). In our prognostic global simulations, the neural network parameterization interacts freely with the resolved dynamics as well as with the surface flux scheme. The neural network parameterization speeds up the model significantly: NNCAM's physical parameterization is around 20 times faster than SPCAM's and even 8 times faster than NNCAM's, in which the radiation scheme is particularly expensive. The key fact to keep in mind is that the neural network does not become more expensive at prediction time even when trained with higher-resolution training data. The approach laid out here should, therefore, scale easily to neural networks trained with vastly more expensive three-dimensional global CRM simulations.

The subsequent analyses are computed from five-year prognostic simulations after a one-year spin-up. All neural network, model and analysis code is available online (SI Appendix).

Results.

Mean climate. To assess NNCAM's ability to reproduce SPCAM's climate we start by comparing the mean sub-grid tendencies and the resulting mean state. The mean sub-grid heating (Fig. 1*A*) and moistening rates (Fig. S2) of SPCAM and NNCAM are in close agreement with a single latent heating tower at the ITCZ and secondary free-tropospheric heating maxima at the mid-latitude storm tracks. The ITCZ peak, which is co-located with the maximum SSTs at 5°N, is slightly sharper in NNCAM compared to SPCAM. In contrast, CTRLCAM exhibits a double ITCZ signal, a common issue of traditional convection parameterizations (5). The resulting mean state in temperature (Fig. 1*B*), humidity and wind (Fig. S2*B* and *C*) of NNCAM also closely resembles SPCAM throughout the troposphere. The only larger deviations are temperature biases in the stratosphere. Since the mean heating rate bias there is small, the temperature anomalies most likely have a secondary cause—for instance differences in circulation or internal variability. In any case, these deviations are not of obvious concern because the upper atmosphere is poorly resolved in our setup and highly sensitive to changes in the model setup (Fig. S5C and D). In fact, CTRLCAM has even larger differences compared to SPCAM in the stratosphere but also throughout the troposphere for all variables.

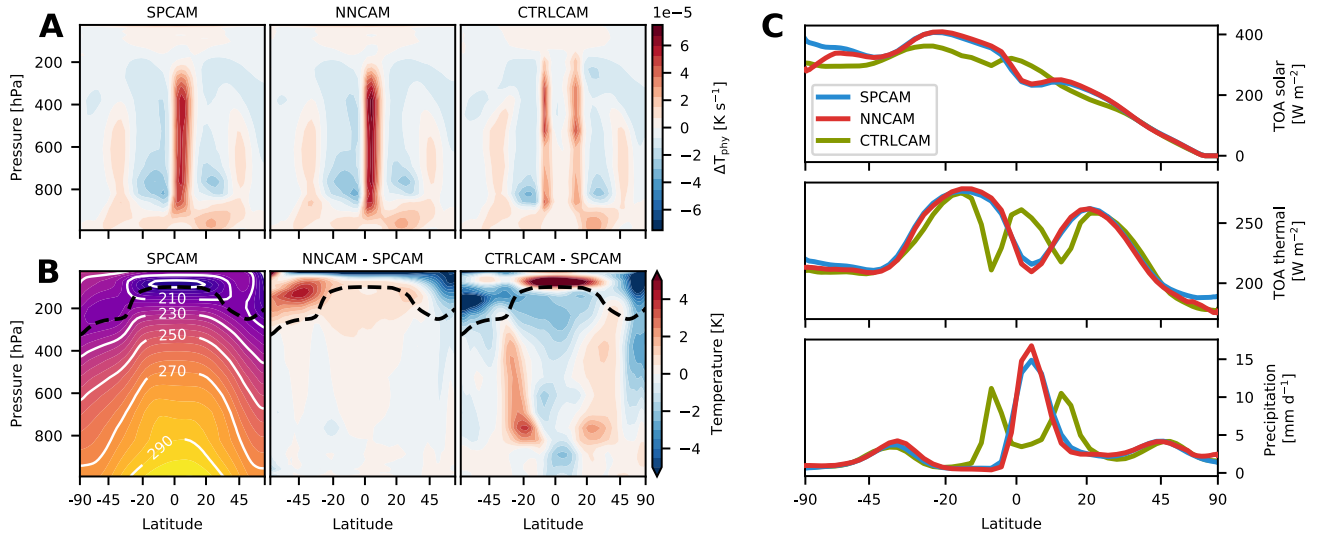

Fig. 1. All figures show longitudinal and five year-temporal averages. (A) Mean convective and radiative sub-grid heating rates ΔT_{phy}. (B) Mean temperature T of SPCAM and biases of NNCAM and CTRLCAM relative to SPCAM. The dashed black line denotes the approximate position of the tropopause, determined by a $\partial p\theta$ contour. (C) Mean shortwave (solar) and longwave (thermal) net fluxes at the top of the atmosphere and precipitation. Note that in all figures the latitude axis is area-weighted.

The radiative fluxes predicted by the neural network parameterization also closely match those of SPCAM for most of the globe, whereas CTRLCAM has large differences in the tropics and subtropics caused by its double ITCZ bias (Figs. 1*C* and S2*D*). Towards the poles NNCAM's fluxes diverge slightly, the reasons for which are yet unclear. The mean precipitation of NNCAM and SPCAM follows the latent heating maxima with a peak at the ITCZ, which again is slightly sharper for NNCAM.

In general, the neural network parameterization, freely interacting with the resolved dynamics, reproduces the most important aspects of its training model's mean climate to a remarkable degree, especially compared to the standard parameterization.

Variability. Next, we investigate NNCAM's ability to capture SPCAM's higher-order statistics—a crucial test since climate modeling is as much concerned about variability as it is about the mean. One of the key statistics for end users is the precipitation distribution (Fig. 2*A*). CTRLCAM shows the typical deficiencies of traditional convection parameterizations—too much drizzle and a lack of extremes. SPCAM remedies these biases and has been shown to better fit to observations (31). The precipitation distribution in NNCAM closely matches that of SPCAM, including the tail. The rarest events are slightly more common in NNCAM than in SPCAM, which is consistent with the narrower and stronger ITCZ (Fig. 1*A* and *C*).

We now focus on the variability of the heating and moistening rates (Figs. 2*B* and S3*A*). Here, NNCAM shows reduced variance compared to SPCAM and even CTRLCAM, mostly located at the shallow cloud level around 900 hPa and in the boundary-layer. Snapshots of instantaneous heating and moistening rates (Fig. S3*B* and *C*) confirm that the neural network's predictions are much smoother, i.e. they lack the vertical and horizontal variability of SPCAM and CTRLCAM. We hypothesize that this has two separate causes: first, low training skill in the boundary-layer (24) suggests that much of SPCAM's variability in this region is chaotic and, therefore, has limited inherent predictability. Faced with such seemingly random targets during training, the deterministic neural network will opt to make predictions that are close to the mean in order to lower its cost function across samples. Second, the omission of condensed water in our network inputs and

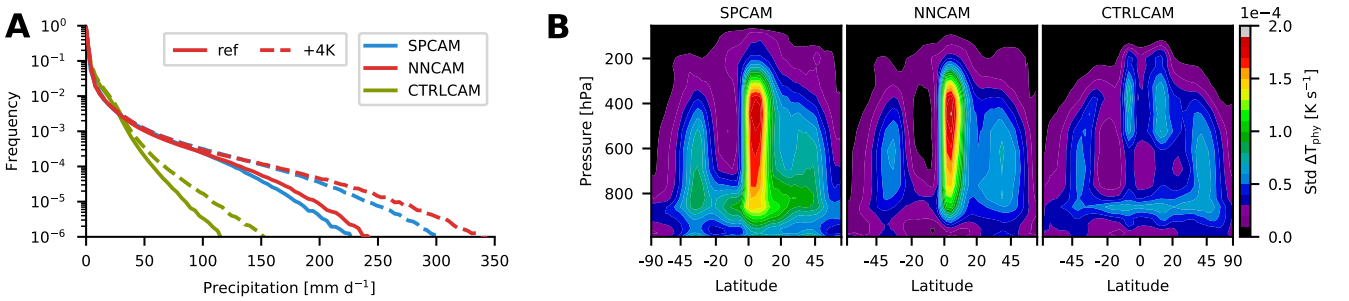

Fig. 2. (A) Precipitation histogram of time-step (30 minutes) accumulation. The bin width is 3.9 mm d^{-1}. Solid lines denote simulations for reference SSTs. Dashed lines denote simulations for +4K SSTs (explanation in Generalization section). The neural network in the +4K case is NNCAM-ref+4K. (B) Zonally averaged temporal standard deviation of convective and radiative sub-grid heating rates ΔT_{phy}.

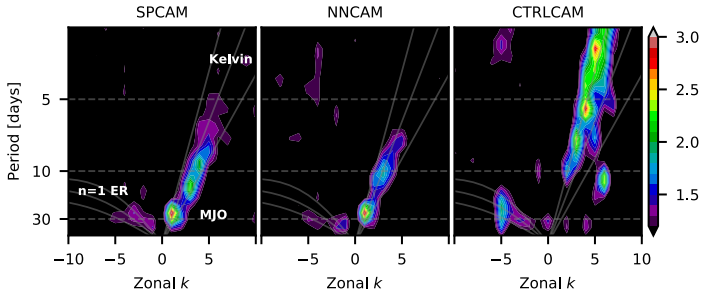

Fig. 3. Space-time spectrum of the equatorially symmetric component of 15S-15N daily precipitation anomalies divided by background spectrum after Fig. 3b in Ref. (32). Negative (positive) values denote westward (eastward) traveling waves.

outputs limits NNCAM's ability to produce sharp radiative heating gradients at the shallow cloud tops. Because the circulation is mostly driven by mid-tropospheric heating in tropical deep convection and mid-latitude storms, however, the lack of low-tropospheric variability does not seem to negatively impact the mean state and precipitation predictions. This result is also of interest for climate prediction in general.

The tropical wave spectrum (32) depends vitally on the interplay between convective heating and large-scale dynamics. This makes it a demanding, indirect test of the neural network parameterization's ability to interact with the dynamical core. Current-generation climate models are still plagued by issues in representing tropical variability: in CTRLCAM, for instance, moist Kelvin waves are too active and propagate too fast while the MJO is largely missing (Fig. 3). SPCAM drastically improves the realism of the wave spectrum (29), including in our aquaplanet setup (26). NNCAM captures the key improvements of SPCAM relative to CTRLCAM: a damped Kelvin wave spectrum, albeit slightly weaker and faster in NNCAM, and an MJO-like intra-seasonal, eastward traveling disturbance. The background spectra also agree well with these results (Fig. S6*A*)

Overall, NNCAM's ability to capture key advantages of the cloud-resolving training model—representing precipitation extremes and producing realistic tropical waves—is to some extent unexpected and represents a major advantage compared to traditional parameterizations.

Energy conservation. A necessary property of any climate model parameterization is that it conserves energy. In our setup, energy conservation is not prescribed during network training. Despite this, NNCAM conserves column moist static energy to a remarkable degree (Fig. 4*A*). Note that because of our omission of condensed water, the balance shown is only approximately true and exhibits some scatter even for SPCAM. The spread is slightly larger for NNCAM, but all points lie within a reasonable range, which shows that NNCAM never severely violates energy conservation. These results suggest that the neural network has approximately learned the physical relation between the input and output variables without being instructed to. This permits a simple post-processing of the neural network's raw predictions to enforce exact energy conservation. We tested this correction without noticeable changes to the main results. Conservation of total moisture is equally as important but the lack of condensed water makes even an approximate version impossible.

The globally integrated total energy and moisture are also stable without noticeable drift or unreasonable scatter for multi-year simulations (Fig. 4*B*). This is still true for a 50-year NNCAM simulation that we ran as a test. The energy conservation properties of the neural network parameterization are promising and show that, to a certain degree, neural networks can learn higher-level concepts and physical laws from the underlying dataset.

Generalization. A key question for the prediction of future climates is whether such a neural network parameterization can generalize outside of its training manifold. To investigate this we run a set of sensitivity tests with perturbed SSTs. We begin by breaking the zonal symmetry of our reference state by adding a wavenumber one SST perturbation with 3K amplitude (Fig. 5*A*; SI Appendix). Under such a perturbation SPCAM develops a thermally direct Walker circulation within the tropics with convective activity concentrated at the downwind sector of the warm pool. The neural network trained with the zonally invariant reference SSTs only (NNCAM) is

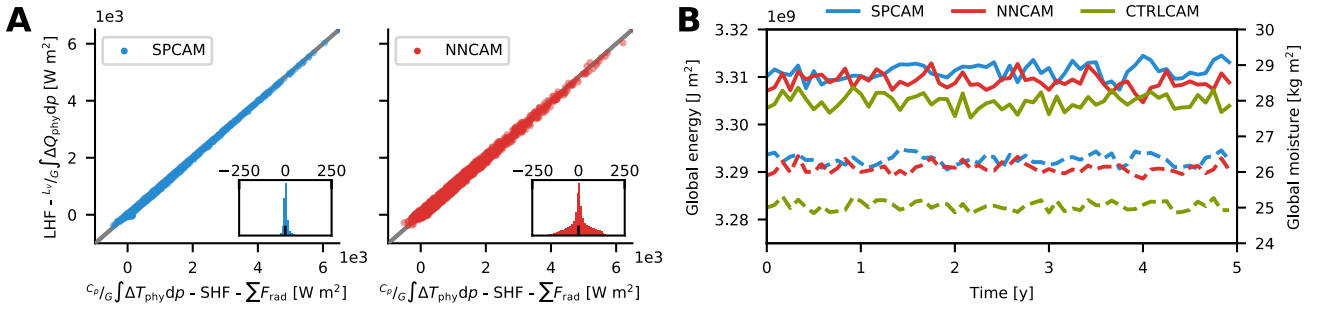

Fig. 4. (A) Scatter plots of vertically integrated column heating $^{C_p}/_G\int\Delta T_{\rm phy}{\rm d}p$ minus the sensible heat flux H and the sum of the radiative fluxes at the boundaries $\sum F_{\rm rad}$ against the vertically integrated column moistening $^{L_v}/_G\int\Delta T_{\rm phy}{\rm d}p$ minus the latent heat flux H. Each dot represent a single prediction at a single column. A total of ten time steps are shown. Inset show distribution of differences. (B) Globally integrated total energy (static, potential and kinetic; solid) and moisture (dashed) for the five-year simulations after one year of spin-up.

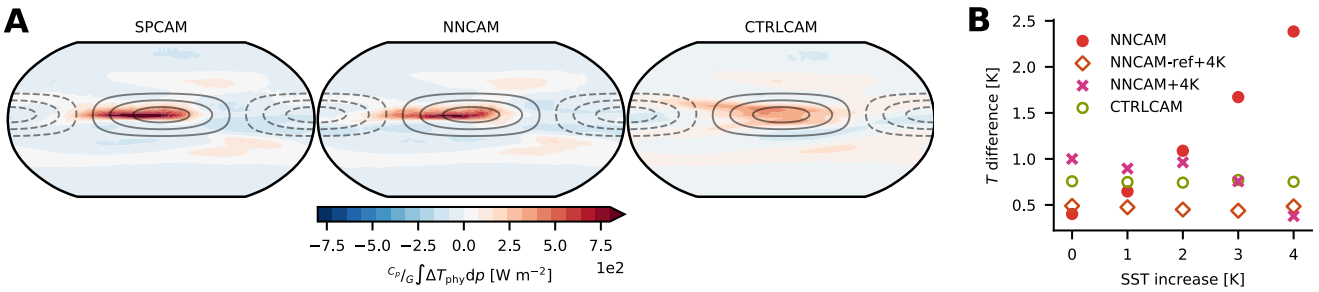

Fig. 5. (A) Vertically integrated mean heating rate ${}^{C_p}/_G \int \Delta T_{\mathrm{phy}} \mathrm{d}p$ for zonally perturbed SSTs. Contour lines show SST perturbation in 1 K intervals starting at 0.5 K. Dashed contours represent negative values. (B) Global mean mass-weighted absolute temperature difference relative to SPCAM reference at each SST increment. The different NNCAM experiments are explained in the corresponding text.

able to generate a similar heating pattern even though the heating maximum is slightly weaker and more spread out. The resulting mean temperature state in the troposphere is also in close agreement, with biases of less than 1 K (Fig. S4). Moreover, NNCAM runs stably despite the fact that the introduced SST perturbations exceed the training climate by as much as 3 K. CTRLCAM, for comparison, has a drastically damped heating maximum and a double ITCZ to the west of the warm pool.

Our next out-of-sample test is a global SST warming of up to 4 K in 1 K increments. We use the mass-weighted absolute temperature differences relative to the SPCAM reference solution at each SST increment as a proxy for the mean climate state difference (Fig. 5*B*). The neural network trained with the reference climate only (NNCAM) is unable to generalize to much warmer climates. A look at the mean heating rates for the +4K SST simulation reveals that the ITCZ signal is washed out and unrealistic patterns develop in and above the boundary-layer (Fig. S5*B*). As a result the temperature bias is significant, particularly in the stratosphere (Fig. S5*D*). This suggests that the neural network cannot handle temperatures that exceed the ones seen during training. To test the opposite case, we also trained a neural network with data from the +4K SST SPCAM simulation only (NNCAM+4K). The respective prognostic simulation for the reference climate has a realistic heating rate and temperature structure at the equator but fails at the poles, where temperatures are lower than in the +4K training dataset (Fig. S5*A* and *C*).

Finally, we train a neural network using half a year of data from the reference and the +4K simulations each, but not the intermediate increments (NNCAM-ref+4k). This version performs well for the extreme climates and also in between (Figs. 5*B* and S5). Reassuringly, NNCAM-ref+4K is also able to capture important aspects of global warming: an increase in the precipitation extremes (Fig. 2*A*) and an amplification and acceleration of the MJO and Kelvin waves (Fig. S6*B*). These sensitivity tests suggest that the neural network is unable to extrapolate much beyond its training climate but can interpolate in between extremes.

Discussion. In this study we have demonstrated that a deep neural network can learn to represent sub-grid processes in climate models from cloud-resolving model data at a fraction of the computational cost. Freely interacting with the resolved dynamics globally, our deep learning-powered model produces a stable mean climate that is close to its training climate, including precipitation extremes and tropical waves. Moreover, the neural network learned to approximately conserve energy without being told so explicitly. It manages to adapt to new surface forcing patterns but struggles with out-of-sample climates. The ability to interpolate between extremes suggests that short-term, high-resolution simulations which target the edges of the climate space can be used to build a comprehensive training dataset. Our study shows a potential way for data-driven development of climate and weather models. Opportunities but also challenges abound.

An immediate follow-on task is to extend this methodology to a less idealized model setup and incorporate more complexity in the neural network parameterization. This requires ensuring positivity of water concentrations and stability which we found challenging in first tests. Predicting the condensation rate, which is not readily available in SPCAM, could provide a convenient way to ensure conservation properties. Another intriguing approach would be to predict sub-grid fluxes instead of absolute tendencies. However, computing the flux divergence to obtain the tendencies amplifies any noise produced by the neural network. Future efforts using machine learning parameterizations should systematically address these issues. Additional complexities like topography, aerosols and chemistry will present further challenges but none of those seem insurmountable from our current vantage point.

Limitations of our method when confronted with out-of-sample temperatures are related to the traditional problem of overfitting in machine learning—the inability to make accurate predictions for data unseen during training. Convolutional neural networks and regularization techniques are commonly used to fight overfitting. It may well be possible that a combination of these and novel techniques improves the out-of-sample predictions of a neural network parameterization. Note also that our idealized training climate is much more homogeneous than the real world climate, for instance a lack of the El Niño-Southern Oscillation, which probably exacerbated the generalization issues.

Convolutional and recurrent neural networks could be used to capture spatial and temporal dependencies, such as propagating mesoscale convective systems or convective memory across time steps. Furthermore, generative adversarial networks (20) could be one promising avenue towards creating a stochastic machine learning parameterization that captures the variability of the training data. Random forests (33) have also recently been applied to learn and model sub-grid convection in a global climate model (34). Compared to neural

networks, they have the advantage that conservation properties are automatically obeyed but suffer from computational limitations.

Recently, it has been argued (35) that machine learning should be used to learn the parameters or parametric functions within a traditional parameterization framework rather than the full parameterization as we have done. Because the known physics are hard-coded this could lead to better generalization capabilities, a reduction of the required data amount and the ability to isolate individual components of the climate system for process studies. On the flip side, it still leaves the burden of heuristically finding the framework equations, which requires splitting a coherent physical system into sub-processes. In this regard, our method of using a single network naturally unifies all sub-grid processes without the need to prescribe interactions.

Regardless of the exact type of learned algorithm, once implemented in the prognostic model some biases will be unavoidable. In our current methodology there is no way of tuning after the training stage. We argue, therefore, that an online learning approach, where the machine learning algorithm runs and learns in parallel with a CRM is required for further development. Super-parameterization presents a natural fit for such a technique. For full global CRMs this likely is more technically challenging.

A grand challenge is how to learn directly from observations—our closest knowledge of the truth—rather than high-resolution simulations which come with their own baggage of tuning and parameterization (turbulence and microphysics) (35). Complications arise because observations are sparse in time and space and often only of indirect quantities, for example satellite observations. Until data assimilation algorithms for parameter estimation advance, learning from high-resolution simulations seems the more promising route towards tangible progress in sub-grid parameterization.

Our study presents a paradigm shift from the manual design of sub-grid parameterizations to a data-driven approach that leverages the advantages of high-resolution modeling. This general methodology is not limited to the atmosphere but can equally as well be applied to other components of the Earth system and beyond. Challenges must still be overcome, but advances in computing capabilities and deep learning in recent years present novel opportunities that are just beginning to be investigated. We believe that machine learning approaches offer great potential that should be explored in concert with traditional model development.

Materials and Methods

Detailed explanations of the model and neural network setup can be found in SI Appendix. This also contains links to the online code repositories. The raw model output data is available from the authors upon request.

ACKNOWLEDGMENTS. SR acknowledges funding from the German Research Foundation (DFG) project SFB/TRR 165 "Waves to Weather". MP acknowledges funding from the DOE SciDac and Early Career programs DE-SC0012152 and DE-SC0012548 and the NSF programs AGS-1419518 and AGS-1734164. PG acknowledges funding from the NSF programs AGS-1734156 and AGS-1649770, the NASA program NNX14AI36G and the DOE Early Career program DE-SC0014203. Computational resources were provided through the NSF XSEDE allocations TG-ATM120034 and TG-ATM170029. We thank Gaël Reinaudi, David Randall, Galen Yacalis, Jeremy McGibbon, Chris Bretherton, Phil Rasch, Tapio Schneider, Padhraic Smyth and Eric Nalisnick for helpul conversations during this work.

1. Schneider T, et al. (2017) Climate goals and computing the future of clouds. *Nature Climate Change* 7(1):3–5.
2. Hourdin F, et al. (2017) The Art and Science of Climate Model Tuning. *Bulletin of the American Meteorological Society* 98(3):589–602.
3. Stevens B, Bony S, Ginoux P, Ming Y, Horowitz LW (2013) What Are Climate Models Missing? *Science* 340(6136):1053–1054.
4. Bony S, et al. (2015) Clouds, circulation and climate sensitivity. *Nature Geoscience* 8(4):261–268.
5. Oueslati B, Bellon G (2015) The double ITCZ bias in CMIP5 models: interaction between SST, large-scale circulation and precipitation. *Climate Dynamics* 44(3-4):585–607.
6. Arnold NP, Randall DA (2015) Global-scale convective aggregation: Implications for the Madden-Julian Oscillation. *Journal of Advances in Modeling Earth Systems* 7(4):1499–1518.
7. Gentine P, et al. (2013) A Probabilistic Bulk Model of Coupled Mixed Layer and Convection. Part I: Clear-Sky Case. *Journal of the Atmospheric Sciences* 70(6):1543–1556.
8. Bony S, Dufresne J (2005) Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters* 32(20):L20806.
9. Sherwood SC, Bony S, Dufresne JL (2014) Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature* 505(7481):37–42.
10. Weisman ML, Skamarock WC, Klemp JB (1997) The Resolution Dependence of Explicitly Modeled Convective Systems. *Monthly Weather Review* 125(4):527–548.
11. Sun J, Pritchard MS (2016) Effects of explicit convection on global land-atmosphere coupling in the superparameterized CAM. *Journal of Advances in Modeling Earth Systems* 8(3):1248–1269.
12. Leutwyler D, Lüthi D, Ban N, Fuhrer O, Schär C (2017) Evaluation of the convection-resolving climate modeling approach on continental scales. *Journal of Geophysical Research: Atmospheres* 122(10):5237–5258.
13. Muller C, Bony S (2015) What favors convective aggregation and why? *Geophysical Research Letters* 42(13):5626–5634.
14. Soden BJ, Vecchi GA (2011) The vertical distribution of cloud feedback in coupled ocean-atmosphere models. *Geophysical Research Letters* 38(12).
15. Miyamoto Y, et al. (2013) Deep moist atmospheric convection in a subkilometer global simulation. *Geophysical Research Letters* 40(18):4922–4926.
16. Bretherton CS, Khairoutdinov MF (2015) Convective self-aggregation feedbacks in near-global cloud-resolving simulations of an aquaplanet. *Journal of Advances in Modeling Earth Systems* 7(4):1765–1787.
17. Yashiro H, et al. (2016) Resolution Dependence of the Diurnal Cycle of Precipitation Simulated by a Global Cloud-System Resolving Model. *SOLA* 12(0):272–276.
18. Klocke D, Brueck M, Hohenegger C, Stevens B (2017) Rediscovery of the doldrums in storm-resolving simulations over the tropical Atlantic. *Nat. Geosci.* 10(12):891–896.
19. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444.
20. Goodfellow I, Bengio Y, Courville A (2016) *Deep Learning*. (MIT Press).
21. Nielsen MA (2015) *Neural Networks and Deep Learning*. (Determination Press).
22. Krasnopolsky VM, Fox-Rabinovitz MS, Belochitski AA (2013) Using Ensemble of Neural Networks to Learn Stochastic Convection Parameterizations for Climate and Numerical Weather Prediction Models from Data Simulated by a Cloud Resolving Model. *Advances in Artificial Neural Systems* 2013:1–13.
23. Brenowitz ND, Bretherton CS (2018) Prognostic Validation of a Neural Network Unified Physics Parameterization. *Geophys. Res. Lett.* 45(12):6289–6298.
24. Gentine P, Pritchard M, Rasp S, Reinaudi G, Yacalis G (2018) Could Machine Learning Break the Convection Parameterization Deadlock? *Geophys. Res. Lett.* 45(11):5742–5751.
25. Collins WD, et al. (2006) The Formulation and Atmospheric Simulation of the Community Atmosphere Model Version 3 (CAM3). *Journal of Climate* 19(11):2144–2161.
26. Andersen JA, Kuang Z (2012) Moist Static Energy Budget of MJO-like Disturbances in the Atmosphere of a Zonally Symmetric Aquaplanet. *Journal of Climate* 25(8):2782–2804.
27. Khairoutdinov MF, Randall DA (2001) A cloud resolving model as a cloud parameterization in the NCAR Community Climate System Model: Preliminary results. *Geophysical Research Letters* 28(18):3617–3620.
28. Pritchard MS, Bretherton CS, DeMott CA (2014) Restricting 32-128 km horizontal scales hardly affects the MJO in the Superparameterized Community Atmosphere Model v.3.0 but the number of cloud-resolving grid columns constrains vertical mixing. *Journal of Advances in Modeling Earth Systems* 6(3):723–739.
29. Benedict JJ, Randall DA (2009) Structure of the Madden–Julian Oscillation in the Superparameterized CAM. *Journal of the Atmospheric Sciences* 66(11):3277–3296.
30. Arnold NP, Randall DA (2015) Global-scale convective aggregation: Implications for the Madden-Julian Oscillation. *Journal of Advances in Modeling Earth Systems* 7(4):1499–1518.
31. Kooperman GJ, Pritchard MS, O'Brien TA, Timmermans BW (2018) Rainfall From Resolved Rather Than Parameterized Processes Better Represents the Present-Day and Climate Change Response of Moderate Rates in the Community Atmosphere Model. *Journal of Advances in Modeling Earth Systems* 10(4):971–988.
32. Wheeler M, Kiladis GN (1999) Convectively Coupled Equatorial Waves: Analysis of Clouds and Temperature in the Wavenumber–Frequency Domain. *Journal of the Atmospheric Sciences* 56(3):374–399.
33. Breiman L (2001) Random Forests. *Machine Learning* 45:5–32.
34. O'Gorman PA, Dwyer JG (2018) Using machine learning to parameterize moist convection: potential for modeling of climate, climate change and extreme events. arXiv: 1806.11037.
35. Schneider T, Lan S, Stuart A, Teixeira J (2017) Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations. *Geophysical Research Letters* 44(24):396–12.