Can Machines Learn How Clouds Work? The Epistemic Implications of Machine Learning Methods in Climate Science

Suzanne Kawamleh*†

Scientists and decision makers rely on climate models for predictions concerning future climate change. Traditionally, physical processes that are key to predicting extreme events are either directly represented (resolved) or indirectly represented (parameterized). Scientists are now replacing physically based parameterizations with neural networks that do not represent physical processes directly or indirectly. I analyze the epistemic implications of this method and argue that it undermines the reliability of model predictions. I attribute the widespread failure in neural network generalizability to the lack of process representation. The representation of climate processes adds significant and irreducible value to the reliability of climate model predictions.

1. Introduction. The success of machine learning (ML) methods in facilitating scientific discovery in disciplines like biology (e.g., bioinformatics) has led some to declare data-intensive science as a fourth paradigm of science, a fundamental transformation of the scientific method (Gray 2007). The nascent field of *climate informatics* was established on the basis of the optimism that ML methods will have similarly profound effects in climate science (Monteleoni et al. 2013).

There is a growing literature exploring the potential of using ML techniques in connection with general circulation models (GCMs) to outperform

and possibly replace traditional models by improving predictions (Dueben and Bauer 2018), providing insight into key climate processes (Ganguly et al. 2014), facilitating causal discovery (Ebert-Uphoff and Deng 2015), and reducing biases and uncertainties in GCM simulations (Steinhaeuser, Chawla, and Ganguly 2010; Rasp, Pritchard, and Gentine 2018).[1] Some question the possibility of replacing physically based models with data-driven models: "can forecast models that are based on deep learning and training on atmospheric data compete with or even beat weather and climate models that are based on physical knowledge and the basic equations of motion?" (Dueben and Bauer 2018, 4000).

Whether ML methods can replace weather prediction and climate models is an open question. I focus my attention on a follow-up question: What would such a shift in climate science methodology imply for the reliability of climate model projections? I explore such implications by examining the increasingly common practice of replacing traditional model parameterizations with ML parameterizations in climate models. Whereas traditional parameterizations indirectly represent climate processes, ML parameterizations aim to secure the same predictive skill without directly or indirectly representing climate processes. I argue that the advent of ML methods, like neural network parameterization (NNP), fundamentally transforms the development and evaluation of climate models. I support my argument with a case study of how artificial neural networks (ANNs) are used to replace convective parameterizations in climate models. The climate models with ML parameterizations fail to be predictively accurate outside of the training data. I attribute this failure to the lack of process representation. The representation of climate processes adds significant and irreducible value to the reliability of climate model predictions.

2. Convective Parameterization

2.1. Convective Processes. Convection is heat and moisture transfer due to a current created by hot air rising and cold air sinking. Convection is a key process in the formation of clouds that (1) produces extreme precipitation, tornados, and so on, and (2) determines the magnitude of atmospheric warming in response to rising greenhouse gas concentrations. Scientists and decision makers rely on climate models for predictive insight concerning future climate change, particularly extreme events like droughts, tornadoes, and monsoons. However, many physical processes that are key to accurately predicting extreme events are not well understood and occur on scales much smaller than the model's resolution. But modeling at a convective-permitting

1. I use model predictions and projections interchangeably throughout the article. Model projections are model predictions that are conditional on certain emissions scenarios or initial conditions.

resolution at the global scale is computationally intensive. As a result, such processes are simplified and approximated in indirect representations called *parameterizations*.

2.2. Traditional Parameterization.   Parameterization is a method of indirectly representing such processes using simplified mathematical expressions. These expressions describe the effects of the process on the rest of the model as a function of variables that are directly represented in the model. The parameterization takes the state of the grid box as an input and calculates how large-scale variables evolve over time. A necessary condition for parameterization is the closure assumption, or the hypothesized existence of a physical relationship between the resolved large-scale variable and subgrid processes. This relationship facilitates the identification of appropriate large-scale variables with which scientists can indirectly represent the subgrid processes. Closure assumptions can be independently and empirically constrained by observational studies. For example, a common closure assumption is that cumulus convection occurs only if the environment is conditionally unstable (Betts and Miller 1984).

However, observational studies like Thompson et al. (1979) demonstrated that instability was inversely related to convective activity. Thus, observational studies challenged the soundness of this assumption and decreased confidence in the projections of models that incorporated this assumption. This is one example of how traditional parameterizations incorporate physical and empirical knowledge to effectively constrain model projections.

2.3. Parametric Uncertainty.   Nevertheless, the indirect representation of climate processes in traditional parameterization does not provide a strong enough constraint on model predictions. The parameterization of cloud-related processes, particularly convection, is the primary driver of intermodel disagreement and the source of the largest uncertainty for model predictions of climate change (Bony et al. 2015). This is due to parametric uncertainty concerning the best parameter value for a model.

An important difference between various parameterization schemes of a climate process is the estimation of the parameter values associated with the large-scale variables used to indirectly represent subgrid processes. Such parameter values are poorly constrained by first principles or observational data. A common practice is to *tune* a climate model, whereby a scientist estimates or makes ad hoc changes to the values of uncertain parameters to improve the statistical fit between model results and observational data. Given the large set of parameters required to start up and run a complex model like a GCM, this parametric uncertainty is compounded. This is significant given that parametric uncertainty concerning ideal parameter values gives rise to uncertainty about model predictions and behavior.

Scientific practices, like tuning, can limit the value of using model fit assessments to build confidence in model predictions. Model fit assessment, assessing the statistical fit of model results to observational data, is a poor indication of the model's ability to adequately represent key processes and predict the observed values of variables related to those processes since the model was deliberately tuned to achieve fit to observational data. Extensively tuning a model's parameters to fit past data undercuts the model's skill in making reliable predictions about a future climate that is different from the past and present climate to which the model parameters were tuned.

Traditional parameterization methods face challenges from parametric uncertainty and widespread practices like parameter tuning. Both challenges are rooted in the indirect representation of key processes that provides a weaker-than-needed constraint on model projections. These limitations give rise to the development of cloud-resolving models (CRMs).

3. Cloud-Resolving Models. A CRM is a high-resolution model that directly represents subgrid processes like deep convective clouds. As a result, CRMs lead to significant improvements in model predictions and reduce model uncertainty (Prein et al. 2015, 340).

Example: Zelinka et al. (2017).—The following example highlights how high resolution process models have advanced scientific understanding and reduced uncertainty. In the Third Assessment Report of the UN Intergovernmental Panel on Climate Change (IPCC), Stocker et al. state that the accurate and reliable simulation of climate change is contingent on climate models' adequate representation of climate processes, especially feedback processes (2001, 419–31). Cloud feedback processes are the source of "the greatest uncertainty in future projections of climate" and, in fact, "even the sign of this feedback remains unknown" (419), with some models producing a positive net feedback and others producing a negative net feedback. However, by the IPCC Fifth Assessment Report, scientists were able to confidently (90% confidence, "very likely") state that "the sign of the net radiative feedback due to all cloud types is . . . likely positive" (Zelinka et al. 2017, 676) with all global climate models simulating a positive net cloud feedback.

Zelinka et al. attribute the improvement in model estimates of cloud feedback across five IPCC assessment reports to "high resolution process models" that "illuminated the competing processes that govern changes in low cloud coverage and thickness" (2017, 677). This is because process models directly represented low cloud processes, like cloud microphysics and phases, which were found to produce positive feedbacks, thus improving model estimates of net radiative feedback. The direct representation of such processes introduces additional physical and empirical constraints that improve climate models' predictive performance.

However, although a CRM has substantial advantages over traditional parameterizations, it is too computationally intensive to run at a global scale for climate prediction. As such, scientists are developing NNPs in an attempt to capture the predictive advantages of the CRM without directly representing clouds and convective processes.

4. Neural Network Parameterizations

4.1. Artificial Neural Networks. Neural networks have been successfully used in a wide array of applications such as image recognition, self-driving cars, and so on. Given a large body of data, neural networks are trained to describe the evolution of nonlinear processes. Many climate processes are nonlinear, and different components of the climate system interact in a nonlinear way. Because of the immense complexity of the climate system and the technical limitations of computing power, it is necessary to approximate the nonlinear functions that describe climate processes. ANNs have a property, the universal approximation function, that enables the ANN to approximate any nonlinear deterministic function. This property holds irrespective of the character of the application (Earth's climate system) or limited knowledge of underlying processes (cloud-related processes) that may pose a challenge for the development of a clear physically based algorithm (Schmidhuber 2015). Thus, ANNs are a promising method by which scientists can better incorporate cloud-related processes, like convection, which cannot be adequately represented by physical equations in GCM.

4.2. Neural Network Parameterization. Scientists are training an ANN to *learn* the improved representation of convective clouds from a CRM. This involves fitting a statistical model to the output data of a CRM. The ML model predictions are then repeatedly compared to the CRM predictions with the aim of minimizing error between the ML model output and CRM output. Thus, the ML model learns the mapping between the input and output variables from the CRM, without being explicitly programmed. The trained neural network then replaces the traditional parameterization in a GCM (now a hybrid GCM) and interacts with other parts of the model.

Climate scientists check for three things when testing the NNP: whether the hybrid GCM with the NNP can accurately simulate (1) basic climate statistics, (2) patterns of climate variability and extreme events, and (3) climate change. The ability to simulate climate change is a test of the generalizability of the NNP beyond the training data. Does the hybrid GCM exhibit model fit for climate statistics and variability because the NNP memorized the CRM data or because it learned some basic physical relations underlying the CRM output data and driving climate change from the CRM? This is important because

one of the aims of developing ML parameterizations is to generate insight into physical processes from large climate data sets (Monteleoni et al. 2013; Ganguly et al. 2014).

Example: Rasp et al. (2018).—Rasp et al. (2018) train a nine-layer-deep neural network to learn atmospheric subgrid processes from a multiscale model that explicitly resolves convection. They use Community Atmosphere Model v3.0 (SPCAM) as the CRM. This CRM explicitly represents deep convective clouds and uses parameterizations for small-scale turbulence and cloud microphysics. The CRM is embedded in each grid column of the GCM. The embedded CRM outputs predictions of the subgrid tendencies of climate variables like humidity as a function of the atmospheric state at each grid column and for every time step. A neural network is trained on a year's worth of CRM output data for these variables and generates its own predictions of the output variables. It is tested to see whether it can learn from the explicitly represented convection in SPCAM.

Rasp et al. then replace a traditional parameterization in a GCM with the trained neural network. The neural network version of CAM is called Neural Network Community Atmosphere Model (NNCAM), and they run simulations for 5 and 50 years. A key test for Rasp et al. is whether the neural network can learn from SPCAM to avoid the shortcomings characteristic of traditional parameterizations, such as a double ITCZ (a precipitation bias). They then run a stable multiyear predictive simulation using the hybrid GCM with the trained neural network.

5. What Machines Fail to Learn

5.1. Failure of Generalizability.   When scientists speak of ML models competing or beating physically based climate models, they generally mean in terms of predictive performance. It is widely accepted that ML models, neural networks in particular, are highly predictive and generalize very well but at the expense of being "black boxes" or opaque to scientific understanding and physical interpretation (McGovern et al. 2019).

Generalizability is a question of how well the ML model learns by testing its predictive skill in new situations. With respect to Rasp et al.'s (2018) study, can NNCAM accurately simulate climate and learn the effects of unprecedented forcing levels, like high sea surface temperatures of 4 K, which are not seen in the training data? Can neural nets learn the effects of extremely high CO_2 levels or predict climate system behavior in areas not included in the training data set?

Unfortunately, ML parameterizations fail to generalize. Despite Rasp et al.'s claims that "NNCAM's ability . . . represents a major advantage compared with traditional parameterizations" (2018, 9687), NNCAM fails to generalize

outside the training data set. The failure in generalizability is a significant blow to the utility of NNPs in GCMs for climate model prediction.

5.2. NNCAM Failure and Beyond. Despite initially positive results, the hybrid GCM (NNCAM) generalizes poorly to out-of-sample temperatures. NNCAM undergoes three sensitivity tests with perturbed sea surface temperatures to assess the parameterization's generalizability outside the range of its training data. NNCAM is able to run stably with sea surface temperature perturbations up to 3 K. However, when the perturbations have an amplitude of 4 K, the NNCAM is unable to generalize: "the ITCZ signal is washed out and unrealistic patterns develop in and above the boundary layer. . . . As a result the temperature bias is significant, particularly in the stratosphere" (Rasp et al. 2018, 9688). For these reasons, Rasp et al. conclude that "the neural network cannot handle temperatures that exceed the ones seen during training" (9687). Rasp et al. attribute this failure of generalizability to the traditional problem of overfitting in ML (9688).

This failure in generalizability is a recurring problem for ML parameterization models (see table 1).[2] The consistent failure to generalize is problematic on two fronts. First, the goal of adopting ML methods in climate modeling was to leverage the advantages of CRMs for global climate prediction. Inadequate predictive skill is a failure to fulfill this primary aim. Second, neural networks are often presumed to be highly predictive but not explanatory. Much attention has been paid to opening the "black box" of neural networks and the challenges neural networks pose for understanding and explanation (McGovern et al. 2019). However, surprisingly, prediction is also a challenge for neural networks in climate science applications. In the next section, I offer an explanation that relates NNPs' lack of process representation to their predictive inadequacy.

6. What Went Wrong? An Explanation. ML parameterizations fail to generalize because, as Rasp et al. note, they overfit the CRM output data. In the process of training the neural network on the output CRM data, the neural network remembers and reproduces the mapping between the input and output variables of that particular CRM data set instead of learning the general physical or causal relations that underlie the CRM output data. As such, the neural network fails to accurately predict outside the training data set even though it fits the CRM data very well—too well in fact.

There are generally two ways to improve a model's ability to predict out of sample: improving and evaluating process representations. Both strategies

2. Scher and Messori (2019) investigate the generalization properties of neural networks for climate science. Their findings are directly relevant for, but not limited to, the generalizability of convective parameterizations.

TABLE 1. Studies That Investigate Machine Learning Convective
Parameterizations and Their Generalizability

Study	Does It Generalize?
Rasp, Pritchard, and Gentine (2018)	No
O'Gorman and Dwyer (2018)	No
Scher and Messori (2019)	No
Yuval and O'Gorman (2020)	No

impose physical and empirical constraints that are more robust against overfitting than model fit assessment. This is important because NNPs are particularly vulnerable to overfitting due to the training of NNPs on previously tuned CRMs. However, ML parameterizations like NNPs do not represent processes directly or indirectly, ruling out both strategies for improving model predictions, leaving scientists with model fit assessment as the primary form of model evaluation. This makes for an unfortunate dilemma.

6.1. From Tuning to Training and Back Again. The original concern with using model fit assessment to support a model's adequacy for purpose was that model performance may be due to ad hoc tuning rather than a model's accurate representation or prediction of climate processes. This concern is exacerbated with neural network models that are trained on a CRM whose parameters have been individually tuned to observational data and that may still incorporate parameterized components.

For example, the CRM in the Rasp et al. (2018) study—SPCAM—still parameterizes cloud microphysics and small-scale turbulence. The CRM includes processes that are directly represented and others that are parameterized because they are not well understood. The values of those parameters are estimated to maximize the statistical fit of the CRM's output with the observational data. The ANN is then trained on the output data of the previously tuned CRM. NNP predictions are then repeatedly compared to the CRM output with the aim of minimizing error between the NNP predictions and CRM output (see fig. 1). This can be summarized, as follows.

1. P is a key parameter in a CRM. Scientists may tune the parameter, or estimate the value of P, to maximize the statistical fit between the observational data and the CRM output.
2. A neural network is trained on the output data of the CRM, which are calculated from the CRM as a whole including P.
3. NNP predictions are repeatedly compared to the CRM outputs with the aim of minimizing the statistical distance between the CRM and NNP predictions.
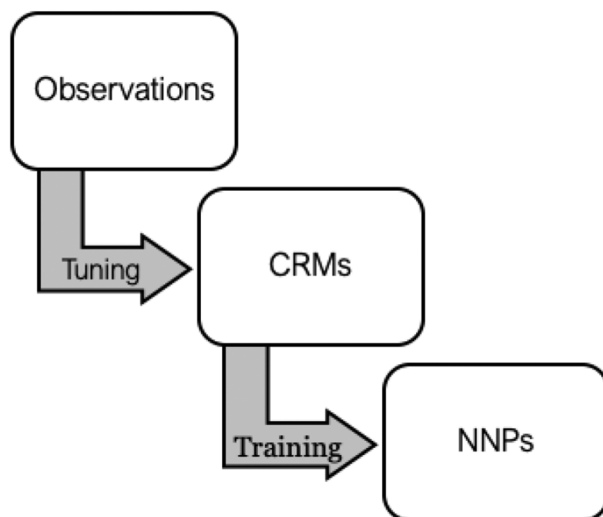
Figure 1. Visual representation of the cascade of uncertainty in neural network parameterizations due to model tuning and training practices.

This means neural network predictions may fit with CRM predictions for any number of reasons: (1) chance, (2) adequacy, (3) tuning/overfitting, or (4) some combination of these.

Training ML parameterizations on previously tuned CRMs has two consequences. First, it is more challenging to identify how model performance depends on tuning and training procedures. Second, it diminishes the epistemic import of model fit assessment, the only form of model evaluation available. It is more likely that there would be good model fit whether the model is or is not adequately predictive out of sample because the purpose of tuning and training is to optimize model fit to observational and CRM output data, respectively. The iterative training of an NNP on a previously tuned CRM makes 3 the most likely explanation for model fit.

6.2. Potential Objection. Some may object that this concern is overblown. As long as the variables assessed for model fit are distinct from the parameters subject to tuning methods, then the epistemic value of good model fit can be upheld. In fact, the CRM parameters most likely to be tuned in an ad hoc manner are, for example, cloud microphysics parameters, not variables associated with directly represented convective clouds or large-scale variables like temperature. When the NNP is trained on the CRM, it is being evaluated for model fit with respect to mean climate statistics like mean temperature. This independence of the variables tuned and variables

evaluated for fit ought to dispel worries about the dependence of fit or performance on tuning practices.

However, in the development of the NNP, we are concerned not just with tuning but with tuning coupled with training practices. The independence is no longer that of just the variable assessed for fit and the parameter tuned. Scientists now face the threefold challenge of distinguishing whether performance depends on tuning, training, or the compounded effects of training a model on a previously tuned CRM. Complex dependencies arise out of practices and interactions that span three related contexts: observational data, the CRM, and the neural network model. Put simply, epistemic opacity is deepened in more than one way (Alvarado and Humphreys 2017).

Furthermore, we know that the neural network model is trained on CRM data concerning the same climate variables for which it is subsequently evaluated using model fit assessment. For example, in Rasp et al. (2018), NNCAM is trained on 140 million training samples from SPCAM, about a year's worth of training data, for temperature and wind profiles. Rasp et al. then evaluate NNCAM's ability to reproduce SPCAM's climate with respect to temperature and wind profiles (9685). Neural networks improve performance via training methods whose efficacy relies on a dependence between the variables used in training and those evaluated for fit with the CRM output data. Training improves performance because it exploits exactly the type of dependence that scientists need to rule out in order for model fit evaluations to have epistemic value. Otherwise, scientists risk overtuning or overfitting to past or recent conditions, and "the model's predictive accuracy might well deteriorate as GMST projections are made for farther and farther into the future" (Parker 2009, 239)—a failure in generalizability.

6.3. Significance. This is a significant concern since the evaluation of the NNP centers on model fit assessment. Rasp et al. assess the hybrid GCMs simulation of (1) key climate statistics like mean climate and climate variability, (2) properties like energy conservation, and (3) the degree to which the hybrid GCM can generalize outside the training data. They do so by comparing NNCAM's ability to reproduce SPCAM's climate, a form of model fit assessment whose value relies on the improved statistical fit of SPCAM to observational data relative to the traditional parameterization. Rasp et al. find that the hybrid GCM, NNCAM, successfully reproduces important aspects of the SPCAM training model's mean climate, key patterns of climate variability, and effectively conserves energy (2018, 9687). This is taken to support NNCAM's adequacy for accurately representing and predicting the outcome of processes associated with mean climate statistics, variability, and thermodynamic principles. Rasp et al. claim that the neural network learned two things from the data set: (1) the higher-level concept of energy conservation

and (2) the physical relation between input and output variables. Thus, they conclude that NNCAM's performance "is to some extent unexpected and represents a major advantage compared with traditional parameterizations" (9687) and take their study to present a "paradigm shift" in the design of sub-grid parameterizations (9688).

However, if the good model fit of NNCAM to the CRM model—whether for mean statistics, variability, or thermodynamics—is due to overfitting that particular CRM data set, then model fit is a poor test of NNCAM's ability to learn either (1) the higher-level concept of energy conservation or (2) the physical relation between input and output variables. The failure of NNCAM's generalizability outside the training data supports the attribution of NNCAM's performance to the extensive tuning and training of the NNP.

Contrary to the claims of Rasp et al., the NNP has failed to learn any meaningful underlying principle or physical relations among relata. Rather, the NNP cannot make accurate or reliable predictions for temperatures that lie outside of the range to which it had been repeatedly and iteratively made to fit with by the training procedure. The limited success of the NNP with respect to the training data is due to the development and evaluation methods used in connection with the NNP, not the NNP's having learned or captured underlying relations that give rise to those mean climate statistics, variability, and so on.

6.4. Summary.   Improving and evaluating process representations are primary means of improving the reliability of model predictions and provide safeguards against overfitting through physical/empirical constraints on model predictions. This is especially valuable in the context of NNPs and other ML applications for which overfitting is a particularly trenchant challenge due to the unique coupling of tuning and training practices in the development process. But, such process-based forms of model improvement and evaluation are unavailable because NNPs do not directly or indirectly represent climate processes. This leaves scientists with model fit assessment as the only and primary form of model evaluation for NNPs. However, model fit assessment is not robust against overfitting and is of limited epistemic value due to the tuning and training methods involved in ML model development. Thus, the replacement of traditional and cloud-resolving parameterizations with NNPs undermines the reliability of model projections. Traditional and cloud-resolving parameterizations represented processes directly or indirectly, and this process representation has added an irreducible value for the reliability of model predictions because it (1) provides physical/empirical constraints and (2) facilitates forms of model development and evaluation that guard against overfitting. This is particularly important for making reliable out-of-sample predictions. The advent of ML methods, like neural networks, transforms the development and evaluation of climate model parameterizations—not always for the better.

7. Conclusion. One of the central aims of Rasp et al.'s development of an NNP was to capture all the advantages of a CRM's improved process representation at a fraction of the computational cost (2018, 9684). However, the NNP did not represent those processes; it attempts to bypass the direct and indirect representation of physical processes that give rise to the output data by identifying and reproducing quantitative relations that hold among variables in the CRM output data. In short, the trained NNP fails to learn convection and generalize beyond its training data because it fails to represent the causal convective processes that relate the climate variables of interest. One cannot, as Rasp et al. had hoped, reap the predictive advantages of high-resolution cloud models while leaving out the key reason for their improved performance—the direct and improved representation of subgrid processes that govern model predictions. The very representation of processes adds significant and irreplaceable value for the reliability of climate model predictions.

REFERENCES

Alvarado, Rafael, and Paul Humphreys. 2017. "Big Data, Thick Mediation, and Representational Opacity." *New Literary History* 48 (4): 729–49.

Betts, Alan K., and M. J. Miller. 1984. *A New Convective Adjustment Scheme*. Technical report no. 43. Reading: European Centre for Medium Range Weather Forecasts.

Bony, Sandrine, et al. 2015. "Clouds, Circulation and Climate Sensitivity." *Nature Geoscience* 8, no. 4 (April): 261–68.

Dueben, Peter D., and Peter Bauer. 2018. "Challenges and Design Choices for Global Weather and Climate Models Based on Machine Learning." *Geoscientific Model Development* 11 (10): 3999–4009.

Ebert-Uphoff, Imme, and Yi Deng. 2015. "Identifying Physical Interactions from Climate Data: Challenges and Opportunities." *Computing in Science and Engineering* 17 (6): 27–34.

Ganguly, A. R., E. A. Kodra, Ankit Agrawal, A. Banerjee, S. Boriah, S. N. Chatterjee, S. O. Chatterjee, A. Choudhary, D. Das, and J. Faghmous. 2014. "Toward Enhanced Understanding and Projections of Climate Extremes Using Physics-Guided Data Mining Techniques." *Nonlinear Processes in Geophysics* 21 (4): 777–95.

Gray, Jim. 2007. "Data Management: Past, Present, and Future." ArXiv, Cornell University. https://arxiv.org/abs/cs/0701156.

McGovern, Amy, Ryan Lagerquist, David John Gagne, G. Eli Jergensen, Kimberly L. Elmore, Cameron R. Homeyer, and Travis Smith. 2019. "Making the Black Box More Transparent: Understanding the Physical Implications of Machine Learning." *Bulletin of the American Meteorological Society* 100 (11): 2175–99.

Monteleoni, Claire, Gavin A. Schmidt, Francis J. Alexander, Alexandru Niculescu-Mizil, Karsten Steinhaeuser, Michael Tippett, Arindam Banerjee, M. Benno Blumenthal, Auroop R. Ganguly, Jason E. Smerdon, and M. Tedesco. 2013. "Climate Informatics." In *Computational Intelligent Data Analysis for Sustainable Development*, ed. T. Yu, N. Chawla, and S. Simoff, 81–126. Data Mining and Knowledge Discovery Series. London: Chapman & Hall.

O'Gorman, Paul A., and John G. Dwyer. 2018. "Using Machine Learning to Parameterize Moist Convection: Potential for Modeling of Climate, Climate Change, and Extreme Events." *Journal of Advances in Modeling Earth Systems* 10 (10): 2548–63.

Parker, Wendy S. 2009. "Confirmation and Adequacy-for-Purpose in Climate Modelling." In *Proceedings of the Aristotelian Society, Supplementary Volumes* 83:233–49.

Prein, Andreas F., et al. 2015. "A Review on Regional Convection-Permitting Climate Modeling: Demonstrations, Prospects, and Challenges." *Reviews of Geophysics* 53 (2): 323–61.

Rasp, Stephan, Michael S. Pritchard, and Pierre Gentine. 2018. "Deep Learning to Represent Subgrid Processes in Climate Models." *Proceedings of the National Academy of Sciences* 115 (39): 9684–89.

Scher, Sebastian, and Gabriele Messori. 2019. "Generalization Properties of Feed-Forward Neural Networks Trained on Lorenz Systems." *Nonlinear Processes in Geophysics* 26 (4): 381–99.

Schmidhuber, Jürgen. 2015. "Deep Learning in Neural Networks: An Overview." *Neural Networks* 61:85–117.

Steinhaeuser, Karsten, Nitesh V. Chawla, and Auroop R. Ganguly. 2010. "Complex Networks in Climate Science: Progress, Opportunities and Challenges." In *Proceedings of the 2010 Conference on Intelligent Data Understanding, CIDU 2010, October 5–6, 2010, Mountain View, California, USA*, ed. Ashok N. Srivastava, Nitesh V. Chawla, Philip S. Yu, and Paul Melby, 16–26. Moffett Field, CA: NASA Ames Research Center.

Stocker, T. F., et al. 2001. "Physical Climate Processes and Feedbacks." In *Climate Change 2001: The Scientific Basis*, ed. J. T. Houghton et al., 417–70. Cambridge: Cambridge University Press.

Thompson, Robert M., Jr., Steven W. Payne, Ernest E. Recker, and Richard J. Reed. 1979. "Structure and Properties of Synoptic-Scale Wave Disturbances in the Intertropical Convergence Zone of the Eastern Atlantic." *Journal of the Atmospheric Sciences* 36 (1): 53–72.

Yuval, Janni, and Paul A. O'Gorman. 2020. "Stable Machine-Learning Parameterization of Subgrid Processes for Climate Modeling at a Range of Resolutions." *Nature Communications* 11 (1): 1–10.

Zelinka, Mark D., David A. Randall, Mark J. Webb, and Stephen A. Klein. 2017. "Clearing Clouds of Uncertainty." *Nature Climate Change* 7 (10): 674–78.