AI4ESP White Paper:
Making Atmospheric Convective Parameterizations Obsolete with Machine Learning Emulation

Walter M. Hannah (LLNL)

Michael S. Pritchard (UC Irvine), John M. Peters (NPS), Peter M. Caldwell (LLNL), Yang Tian (LLNL), Youngsoo Choi (LLNL), Aaron Donahue (LLNL), Ben Hillman (SNL)

Focal Area

Parameterizations of moist convection in atmospheric models are notoriously problematic, and while global cloud resolving models (GCRM) are often touted as the ultimate solution, the computational cost is a considerable hurdle to overcome. Machine learning emulation of GCRMs for predictive modelling can leverage the DOE's computational resource investments and allow widespread use of GCRMs such that traditional parameterizations become obsolete for most applications.

A World Without Parameterized Convection

Many atmospheric scientists direct their entire career towards improving the representation of convective processes in global models, which are widely understood to be the largest sources of uncertainty (Sherwood et al. 2014; Sanderson e al. 2008). However, there is only so much progress we can expect to make using parameterizations based on "low-order" models of cloud processes. Many of us hope that incremental improvements to parameterizations of clouds and microphysics will somehow bring us to a model that we consider to be faithful resemblance to the real atmosphere. Unfortunately, clouds are comically more complicated than these low-order approaches could ever be, so we shouldn't trick ourselves into thinking that we can engineer away our model biases with better low-order models of convection.

Over the last decade there has been a steady increase in the number of models run at convective resolving scales over a global domain. There has also been work to develop hybrid approaches, such as regionally refined grids and multiscale modelling frameworks, for conducting experiments that attempt to strike a balance between fidelity and computational burden. Despite this progress, the technical challenges that accompany these feats of experimentation will not be completely solved by clever coding or larger arrays of compute nodes. A big reason for this problem is that as the increased workload from finer grid spacing limits the model throughput due to the corresponding decrease in time step and associated increase in the frequency of interprocess communication.

Thus, it may seem that a future without traditionally parameterized global models is an unattainable fantasy, but advances in machine learning may provide a path towards widespread use of finely resolved models via emulation. In other words, a model could exist as a combination of the full physics model and a suite of trained emulators that can reproduce the behavior of the full model at a fraction of the cost. This notion of replacing poorly understood physics with an even more poorly understood "black box" surrogate model will undoubtably make many scientists uncomfortable. However, this does not need to be a scary proposition, and it doesn't necessarily mean that there's no place for developing simplified, physically-based models to bolster our understanding of the various phenomena. Simple models will always have a role in distilling a problem down to its simplest possible explanation, but we shouldn't have to tolerate these simplifications when societal needs mandate accurate projections that

DOI: https://doi.org/10.2172/1769746
DOE BER, AI4ESP 2021 #AI4ESP1056 (https://www.ai4esp.org)

encompass a larger manifold of complexity.

How Do We Get There?

Machine learning emulation of atmospheric models is still in its infancy. Recent studies have begun exploring how to emulate the physics calculations with success (Gentine et al. 2018; O'Gorman and Dwyer 2018; Rasp et al. 2018; Brenowitz and Bretherton 2019; Benowitz et al. 2020), but there are still many areas to investigate. All of these previous studies have made sensible decisions to simplify the problem (ex. aquaplanet) in order to more clearly demonstrate how well a surrogate model can emulate the full model. These surrogate models can successfully reproduce things that they were not constrained to do, like energy conservation, but it remains to be seen whether these results hold up when fully coupled to ocean, ice, and land models. Emulating more complex models might require more supervised training methods to constrain the surrogate models. Answering the many open questions on this topic requires incentives to optimize the methods and infrastructure so that these tools can be widely adopted for scientific research.

Optimizing the Training Methods and Data

Building surrogate models involves subjective choices, such as the structure of nodes in a neural network or the number of trees in a random forest. In some ways these choices are just engineering exercises, but they could have scientific consequences if the resulting surrogate model suffers from a lack of stochasticity or overfitting to the present climate. Identifying and minimizing these problems will require detailed comparison across this large array of possibilities.

When training a surrogate model of a model physics package that is meant to run on the same grid as the full model, the obvious choice of training data is the raw inputs to the physics calculations, as in Rasp et al. (2018). However, there may be just as much utility in emulating the physics over a coarser grid, as in Brenowitz and Bretherton (2019), which will require further choices on how to coarse-grain the finely resolved model data. Coarse-graining GCRM data also opens interesting possibilities on whether a surrogate model would benefit from including a representation of the "sub-grid" convective variance or organization that is lost when moving to a coarser grid. Similarly, the effects of fine-scale topography could potentially be encoded into the training data set to improve the performance over mountainous terrain where the dynamical core is given a smoother surface.

The results of Brenowitz and Bretherton (2019) suggest that expert human intervention may be required to cultivate the optimal training data set that yields a stable emulator. Ideally these lessons would be codified into rules based on physical reasoning so that work on similar problems would not have to learn the same lessons repeatedly. But even if such rules can be formulated, the community would benefit from a rich history of published work exploring the various limitations of different training approaches.

Frequent Surrogate Model Training with Massively Parallel Processing

In our vision of a world where atmospheric simulations are either done with an expensive GCRM or a cheap emulation of the fine-scale physics, there will still be a need for active development and improvement to the full physics model. Much of this development and testing can be done on reduced domains with prescribed or parameterized large-scale forcing, but there will be a recurring need to produce new surrogate models along the way. It is not clear how this constant iteration of re-training with such massive datasets can be accommodated.

Part of the issue of optimizing the training process comes from not knowing how much training data is sufficient. Instead of running a GCRM continuously for one or multiple years, it may be sufficient to use a targeted set of short hindcasts that cover a range of conditions, including various phases of interannual and interdecadal modes of variability. If this approach is feasible, then hindcast ensembles could be run in parallel to dramatically speed up the production of data for surrogate model training and validation. Determining the validity of these strategies also invokes the question of whether short term validation of a surrogate model will translate to simulation fidelity on longer time scales.

Another hurdle to supporting frequent surrogate model training is the fact that training often requires that all training data is accessible from the memory on a single compute node, but this creates a huge problem when the training data set becomes much larger. This problem is commonly avoided in engineering applications of machine learning emulation by creating a network of emulators, each of which can be trained to emulate a specific physical region. This approach is not likely to be feasible for climate projections as we want the physics to be consistent every location to accommodate drastic changes at the surface caused by melting ice or shifting coastlines.

Summary

The potential for machine learning based emulation to unleash a new era in earth system modeling is a tantalizing prospect, but the community lacks the experience that is critical for scaling this technology up to emulate global cloud resolving models. Funding should be directed towards optimizing the methods and infrastructure needed to make surrogate model training a standard component of Earth system models in the exascale era.

References

1. Brenowitz, N. D., and C. S. Bretherton, 2019: Spatially Extended Tests of a Neural Network Parametrization Trained by Coarse-Graining, *J. Adv. Model. Earth Syst.*, **11**, 2728– 2744.
2. Brenowitz, N. D., T. Beucler, M. S. Pritchard, and C. S. Bretherton, 2020: Interpreting and Stabilizing Machine-Learning Parametrizations of Convection, *J. Atmos, Sci.*, *77*(12), 4357-4375.
3. Gentine, P., M. S. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis, 2018: Could machine learning break the convection parameterization deadlock? *Geophys. Res. Lett.*, **45**, 5742–5751.
4. O'Gorman, P. A., and J. G. Dwyer, 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.
5. Rasp, S., M. S. Pritchard, and P. Gentine, (2018): Deep learning to represent subgrid processes in climate models. *Proc. Nat. Acad. Sci.*, **115**, 9684–9689.
6. Sanderson, B.M., Piani, C., Ingram, W.J. *et al.* Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed-physics GCM simulations. *Clim Dyn* **30,** 175–190 (2008).
7. Sherwood, S., Bony, S. & Dufresne, JL. Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature* **505,** 37–42 (2014).