

Making better Earth system models with machine learning

Kelly Kochanski, University of Colorado Boulder

What is machine learning?

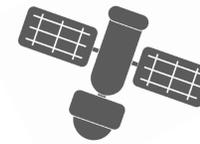
Machine learning (ML) is a set of statistical and computational techniques that use patterns in data to complete tasks without explicit instructions.

What should I know now?

The Earth is a complex, nonlinear system, and machine learning is the state of the art of complex systems modeling and prediction.

In the last few years, new streams of satellite and simulation data have made more ML techniques feasible for Earth science, and new advances in interpretable ML and model-data integration have made ML more compatible with Earth science models.

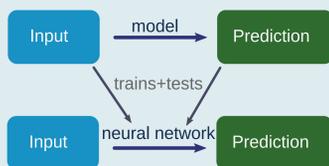
Right now, new applications need new Earth/ML expertise. That's you.



Emulating expensive physics

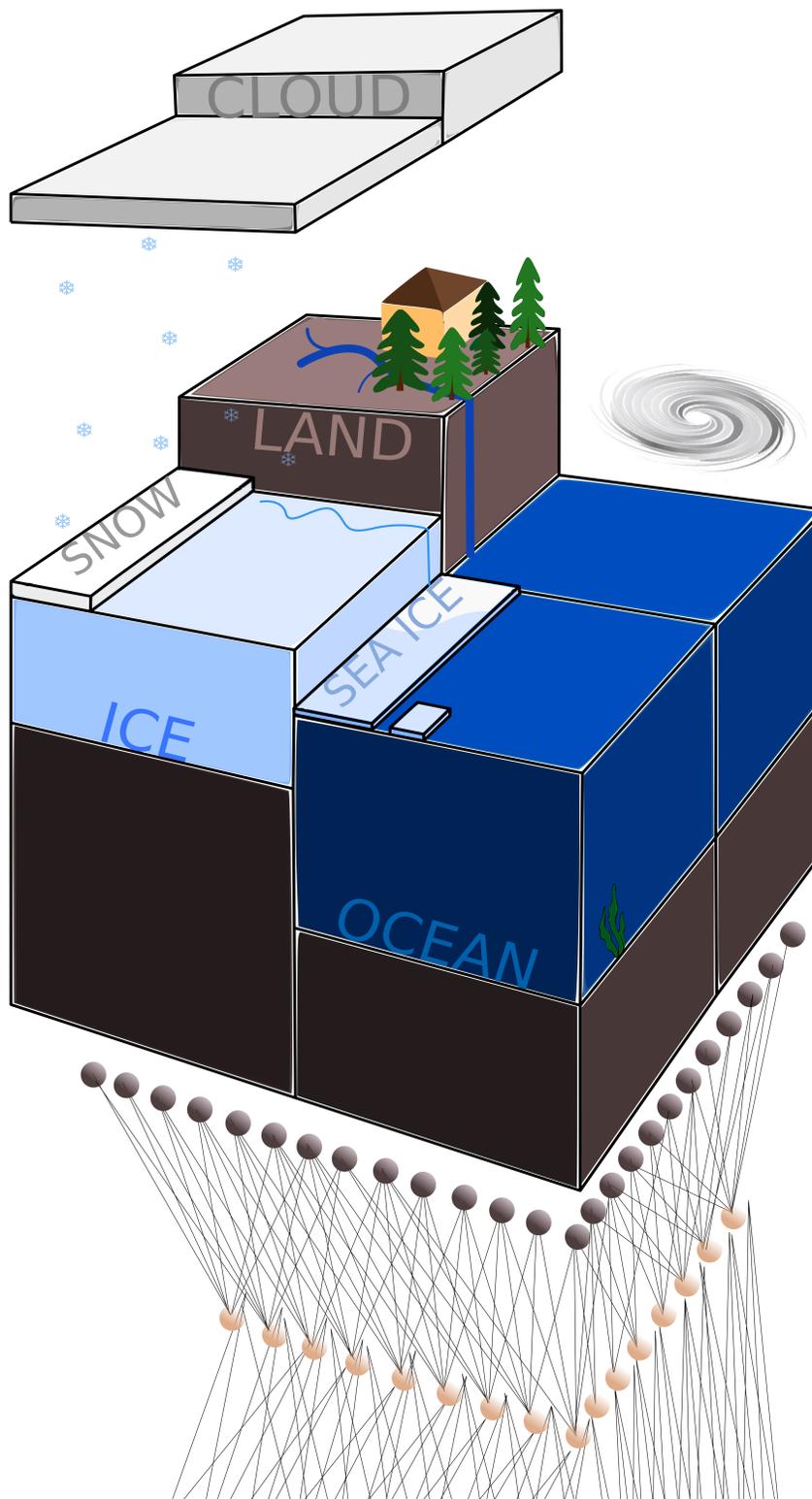
Many successful Earth science models, such as cloud-resolving atmospheric models, are too expensive to run within large, coupled models, or large multi-model experiments.

- Recent work has shown that deep and convolutional neural networks can emulate many physics models (e.g. [1,2]).



- The resulting neural networks have different cost-accuracy trade-offs than the original models, and are often easier to parallelize or run on GPUs.

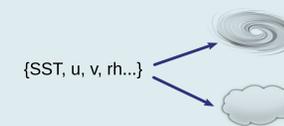
- Neural networks provide a framework that works for both data-generated and physics-based model components, and may make it possible to build physics-based models that learn select parameters continuously from data; see [3].



Identifying extreme events

Identifying extreme events within weather, climate and other data sets is a classification problem, with a twist: extreme events are, by definition, rare.

- Deep neural networks [9] and other machine learning techniques [10] have been used to track cyclones, weather fronts, tropical storms, and tornadoes.



Classifying land surfaces

With the growth of satellite imagery and other satellite datasets, tracking changes and processes on the surface of the Earth is increasingly an image recognition problem.

- Many machine learning techniques are great for classifying land use, including vegetation, crop cover, and dust sources, from satellite imagery; see [11,12,13].

- [14] proposes to create a new dataset of environmental imagery, called EnviroNet, to accelerate ML development for environmental science. If you have images you need to classify, the author is looking for test problems.

Using ensembles

Most models are uncertain. This uncertainty can often be reduced with multi-model ensembles, and ensemble predictions are a good problem for ML.

- At any given point in time, one of the ensemble members makes a better prediction than the others. Identifying this member is a tracking and classification problem that can be solved with online algorithms; see [4].

- Ensembles are expensive, but random forests may make it possible to get great results from an ensemble of mostly cheap, low resolution models; see [5].

Down-scaling

High-resolution results are great, but high-resolution models are expensive.

- Many machine learning techniques have been extensively used for down-scaling weather and climate data for local use; see for example [6].

- New developments in 'image super-resolution', e.g. [7], are likely to enable new down-scaling methods based on image recognition techniques.

Tracking spatial patterns

Many geoscientific trends can be visualized as videos. These include topographic evolution, sea ice, surface temperatures, and vegetation patterns.

- New video prediction techniques, such as [8], may make it possible to predict the evolution of image-like patterns directly from data.



What's coming next?

ML may change the way that scientific modeling is done. From a ML standpoint, learning from an existing model has many advantages: modelers can generate new training and test data on-demand, and the new ML model inherits some community trust from the old one. New research is working to develop the use of ML in models.

- New maps from neural networks to ODEs and vice versa [15, 16]
- Neural networks are able to pre-emptively detect instabilities [17] and sources of uncertainty [18,19] in physics models, and to help with model optimization [20].

New ML models for climate will be most successful if they are closely integrated into existing scientific models. This has been emphasized, again and again, by authors who have laid future paths for artificial intelligence within Earth science. New models need to leverage existing knowledge to make good predictions with limited data. In ten years, we will have more satellite data and more interpretable ML. For now, however, ML models must be creatively designed to work within existing models. The best of these models are likely to be built by close-knit teams including both climate and computational scientists.

References

- [1] P. Gentine et al. Could Machine Learning Break the Convection Parameterization Deadlock? GRL, 2018.
- [2] S. Rasp et al. Deep learning to represent subgrid processes in climate models. PNAS, 2018.
- [3] M. Reichstein et al. Deep learning and process understanding for data-driven Earth system science. Nature, 2019.
- [4] C. Monteleoni et al. Tracking climate models. Statistical Analysis and Data Mining, 2011.
- [5] G. Anderson and D. Lucas. Machine Learning Predictions of a Multiresolution Climate Model Ensemble. GRL, 2018.
- [6] W. Li et al. Evaluation of ML algorithms in spatial downscaling of MODIS land surface temperature, 2019.
- [7] Ledig et al. Photo-realistic single image super-resolution using a generative adversarial network. arXiv:1609.04802, 2016.
- [8] M. Mathieu et al. Deep multi-scale video prediction beyond mean square error. arXiv, 2016.
- [9] Y. Liu et al. Application of deep convolutional neural networks for detecting extreme weather in climate datasets. arXiv, 2016.
- [10] V. Lakshmanan and Travis Smith. An objective method of evaluating and devising storm tracking algorithms. Weather and Forecasting, 2010.
- [11] D. J. Lary, et al. Machine learning in geosciences and remote sensing. Geoscience Frontiers, 2015.
- [12] M. Kussul et al. Deep learning classification of land cover and crop types using remote sensing data, 2017.
- [13] D. J. Lary. Artificial Intelligence in Geoscience and Remote Sensing, 2009.
- [14] S. K. Mukkavilli. EnviroNet: ImageNet for environment, 2019.
- [15] M. Raissi and G. E. Karniadakis. Hidden physics models: machine learning of nonlinear partial differential equations. J. Computational Physics, 2018.
- [16] R. T. O. Chen et al. Neural ordinary differential equations. NIPS, 2018.
- [17] M. Jiang et al. A supervised learning framework for arbitrary Lagrangian-Eulerian simulations. ICMLA, 2016.
- [18] J. Ling and J. Templeton. Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier-Stokes uncertainty. Physics of Fluids, 2015.
- [19] B. Lakshminarayanan et al. Simple and scalable predictive uncertainty estimation using deep ensembles. NIPS, 2017.
- [20] J. Thangarajan et al. Bootstrapping parameter space exploration for fast tuning, 2018.

Thanks

David Rolnick, Priya Donti, S. Karthik Mukkavilli, Ben Kravitz, Andrew Ross, John David Gagne II, Ghaleb Abdulla, Goodwin Gibbons



See more at

www.climatechange.ai

(Tackling climate change with ML' paper release planned for early June)