Structurally flexible cloud microphysics, observationally constrained at all scales via ML-accelerated Bayesian inference

Authors: Marcus van Lier-Walqui, Hugh Morrison, Greg Elsaesser,

Peter Caldwell, Pierre Gentine, Derek Posselt, Sean Santos, Po-Lun Ma

February 2021

Focal Areas: The proposed research addresses all three focus areas, but primarily (2): *Predictive modeling through the use of AI techniques and AI-derived model components; the use of AI and other tools to design a prediction system comprising of a hierarchy of models (e.g., AI driven model/component/parameterization selection)*

Science Challenge: We discuss the challenge of developing observationally informed parameterizations of microphysics for use at a hierarchy of modeling scales. Our proposed approach is applicable to any domain that suffers from a two-fold parameterization problem, where physical processes are not resolved at the model scale (the first problem), and those processes are uncertain at any scale (the second problem). For such problems, a physical approach facilitates modeling across scales, as well as systematic observational inference accelerated by machine learning (ML) surrogate models, and quantification of physical uncertainties.

Rationale Microphysics schemes in weather and climate models account for the interactions of cloud and precipitation particles, the processes that control their growth, and their effects on the dynamic and thermodynamic state of the atmosphere. It is well known that microphysics schemes are prime targets for improvement, as they suffer from physical deficiencies that are difficult to resolve even with detailed process-level studies [13]. A number of previous Machine Learning (ML) studies have focused on macroscale cloud parameterizations (including convection) using high-resolution simulations as training data [2, 17, 8]. These efforts depend critically on the capacity to design trustworthy high-resolution simulations, which is limited for microphysics. There is also a strong and warranted demand for *physical* parameterizations, i.e. parameterizations that are physically interpretable with recognizable physical components. Integrating observational constraints across scales (e.g. from in situ to global satellite), and implementing these insights in a hierarchy of models (e.g. from sub-kilometer resolution large eddy simulations (LES) to global climate models) is challenging without basing those in physical relationships that are valid at all scales. While the appeal of ML lies in a structural flexibility that allows (with adequate training data) performance to move beyond that of current physically-based schemes, another approach (largely unexplored) is to build structural flexibility into physically-based parameterizations to yield similar benefits. Whatever approach is taken, it is highly likely that uncertainties will remain in *any* future microphysics scheme, in part owing to the difficulty in directly observing microphysical

DOI: https://doi.org/10.2172/1769779
DOE BER, AI4ESP 2021 #AI4ESP1079 (https://www.ai4esp.org)

processes, and in part owing to observational uncertainties. That is not to say that uncertainty cannot be reduced, but that at each step it should be *quantified*; schemes must therefore embrace the probabilistic nature of the current state of knowledge.

Narrative We contend that progress in microphysics parameterization development for Earth system models depends critically on three factors:

1. A physically interpretable model
2. A methodology for addressing both parametric and structural errors
3. Quantification of physical uncertainties

Factor (1) reflects the benefits that physically-based models afford for integrating observational insights from all their available scales, thus facilitating constraint by laboratory studies in microphysics, multi-platform field campaign data, as well as satellite cloud observations. Additionally, physically-based insights can be naturally applied to a hierarchy of models, from high-resolution LES to global Earth System Models. Fully data-driven ML approaches are typically trained on vast quantities of data that are usually provided by some "reference" model. While they are well suited to reproducing their training data, they often have difficulty extrapolating outside these ranges of behavior — a critical requirement of any Earth system model. Methods for physical regularization of ML codes can help in this regard [1], but still do not solve the problem of how to combine, e.g. in situ and satellite-scale observational constraints within a single model framework. However, a more fundamental problem is that for microphysics, no true reference model exists — even at the particle scale processes such as rain droplet collisional breakup and snow particle aggregation are highly uncertain.

To their favor, ML approaches directly address factor (2) by providing a structurally flexible framework for simulation of physical processes, in this way offering improvement beyond the bounds of what is possible within existing physical parameterizations. The structural errors that hamper traditional physical schemes are expected to be important for cloud microphysics. In [4] it is noted that a ML parameterization for cloud-to-rain autoconversion is improved when it is allowed to depend on both rain number concentration and rain mass concentration; this is an example of where adding a new structural term improves model fidelity. Existing physical schemes have limited ability to adjust these fixed structural elements — typically they have some small number of parameters that can be adjusted, but they also make assumptions that are "hard-coded" into their structure and cannot be easily (or systematically) altered. For example, a process such as autoconversion is itself somewhat unphysical, as it is based on an ad hoc separation of liquid drops into two size-based categories (cloud and precipitation) — this constitutes a particularly deep-rooted structural uncertainty in most modern bulk microphysics schemes. Progress in microphysics requires the ability to systematically adjust and resolve structural components with the ultimate goals of: theoretical consistency, predictive fidelity, and model parsimony. Neither the first nor the third goals are easily attainable by purely data-driven ML approaches.

Factor (3) is related to the previous two factors, and is an acknowledgement of the fact that model error will remain in virtually any future representation of atmospheric processes. Recent work by Morrison et al. [13] argues that uncertainties are likely to remain in microphysics for the foreseeable future owing to the difficulty in observing the microscale nature of precipitation and cloud particle interactions and growth mechanisms. Even if single-particle interaction and growth processes become better understood, there will remain the "first" parameterization problem: models cannot resolve the spatial scales upon which these processes occur. Thus, this implies that, to the extent possible, schemes should be capable of characterizing their own uncertainties, in turn requiring the ability to quantify uncertainty in scheme parameters as well as uncertainty in scheme structure. In this way we can quantify observationally driven advances in terms of reduction in scheme uncertainties.

We argue that in light of the above issues, it is critically important that *physical* models remain a focus of development for Earth System Models in general, and cloud microphysics parameterization schemes in specific. By employing a physical basis, theoretical knowledge, process-level insights from laboratory, in situ and remote sensing observations (such as those from DOE-supported field campaigns) can be unified with global satellite-based observational constraints within a Bayesian framework. For example, a detailed process-level study might generate a PDF over some parameters of interest (e.g. [20]); this PDF may then be used as the prior estimate to be successively refined into a *posterior* estimate by application of *new* observational constraint for, e.g., a global GCM simulation [6]. This process can be repeated with new observations, and at each step uncertainty is quantified via Bayes theorem, which natively treats quantities probabilistically.

The aforementioned approach does not directly address structural uncertainties in microphysical parameterization schemes, and doing so requires formulation of parameterization schemes wherein both parameters and structural choices are systematically adjustable. A microphysics parameterization currently in active development, the Bayesian Observationally-constrained Statistical-physical Scheme (BOSS, [14, 21]) is one such method — in BOSS microphysical processes are represented as generic power series expressions, and other structural adjustment is accomplished via the number and choice of drop-size distribution predictive variables. Similar approaches are possible for ice microphysics, aerosol processes, and cloud parameterization, but require community buy-in and support from sponsoring programs. Furthermore, we foresee a role for ML in investigation of an optimal basis for model structure, e.g. following recent work in equation discovery [3], we suggest research in *hybrid* physical-ML models built on physical foundations with structural options discovered via ML.

While the construction of suitably flexible microphysics schemes fulfills an essential prerequisite, there is work to be done to develop and refine the formalism by which observations will constrain these new models. Specifically, investment is needed on optimal methods for performing observationally-driven inference and quantifying uncertainty in physically-based parameterizations. In part, this relies on Bayesian methods like Markov Chain Monte Carlo — continued development of these methods is occurring , and new algorithms must be tested in ESM applications. Additionally, related methods such as non-Gaussian Kalman methods [16], importance sampling [15], and filtering algorithms such as Particle Filters [12] should be explored and tested. Specifically to address structural deficiencies, approaches for systematically inferring model structure should be tested, such as Bayesian model selection [9], transdimensional inference [18], and model expansion [7]. Thus far, the authors know of no such work in Earth System Modeling or microphysics.

Finally, Bayesian methods are often prohibitively expensive, requiring many thousands or even millions of model simulations. Here ML plays a critical role by providing computationally cheap *surrogate models* (aka model *emulators*) — machine learning codes such as neural networks or Gaussian process models [11] that emulate the sensitivities of the physical models to parameter variation. These methods, trained on perturbed parameter ensembles, have been applied to global model components successfully [11, 10], and have been proposed as a general strategy for ESM tuning [19, 5] — a proposal borne out by preliminary work applying emulator-based MCMC parameter estimation for development of the upcoming version of the NASA GISS ModelE global climate model [6]. Much work remains to investigate the optimal use of ML surrogates, assess their limitations, and develop new techniques tailored to Earth System models.

References

[1] T. Beucler, M. Pritchard, S. Rasp, J. Ott, P. Baldi, and P. Gentine. Enforcing analytic constraints in neural-networks emulating physical systems. *arXiv [physics.comp-ph]*, (1909.00912v5), 2021.

[2] N. D. Brenowitz and C. S. Bretherton. Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, 45, 2018.

[3] S. L. Brunton, J. L. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, mar 2016.

[4] J. C. Chiu, C. K. Yang, P. J. van Leeuwen, G. Feingold, R. Wood, Y. Blanchard, F. Mei, and J. Wang. Observational constraints on warm cloud microphysical processes using machine learning and optimization techniques. *Geophysical Research Letters*, 48, 2021.

[5] E. Cleary, A. Garbuno-Inigo, S. Lan, T. Schneider, and A. M. Stuart. Calibrate, emulate, sample. *arXiv [stat.CO]*, (2001.03689v1), January 2020.

[6] G. S. Elsaesser. NASA GISS ModelE3: New tuning framework and evaluation against global satellite datasets. *AGU Fall Meeting*, (A047-04), 2020.

[7] A. Gelman and C. R. Shalizi. Philosophy and the practice of bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, 66(1):8–38, Feb 2012.

[8] P. Gentine, M. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis. Could machine learning break the convection parameterization deadlock. *Geophysical Research Letters*, 45, 2018.

[9] M. Höge, A. Guthke, and W. Nowak. The hydrologist's guide to Bayesian model selection, averaging and combination. *Journal of Hydrology*, 572:96–107, 2019.

[10] J. S. Johnson, Z. Cui, L. A. Lee, J. P. Gosling, A. M. Blyth, and K. S. Carslaw. Evaluating uncertainty in convective cloud microphysics using statistical emulation. *Journal of Advances in Modeling Earth Systems*, 7:162–187, 2015.

[11] L. A. Lee, K. S. Carslaw, K. J. Pringle, G. W. Mann, and D. V. Spracklen. Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters. *Atmospheric Chemistry and Physics*, 11(23):12253–12273, Dec 2011.

[12] P. J. Leeuwen, H. R. Künsch, L. Nerger, R. Potthast, and S. Reich. Particle filters for high-dimensional geoscience applications: A review. *Quarterly Journal of the Royal Meteorological Society*, 145(723):2335–2365, may 2019.

[13] H. Morrison, M. Lier-Walqui, A. M. Fridlind, W. W. Grabowski, J. Y. Harrington, C. Hoose, A. Korolev, M. R. Kumjian, J. A. Milbrandt, H. Pawlowska, and et al. Confronting the challenge of modeling cloud and precipitation microphysics. *Journal of Advances in Modeling Earth Systems*, May 2020.

[14] H. C. Morrison, M. van Lier-Walqui, M. R. Kumjian, and O. P. Prat. A Bayesian approach for statistical-physical bulk parameterization of rain microphysics, Part I: Scheme description. *Journal of the Atmospheric Sciences*, 77(3):1019–1041, March 2020.

[15] M. Morzfeld, M. S. Day, R. W. Grout, G. S. H. Pau, S. A. Finsterle, and J. B. Bell. Iterative importance sampling algorithms for parameter estimation. *SIAM Journal on Scientific Computation*, 40(2):B329–B352, 2018.

[16] D. J. Posselt and C. H. Bishop. Nonlinear data assimilation for clouds and precipitation using a gamma inverse-gamma ensemble filter. *Quarterly Journal of the Royal Meteorological Society*, 144(716):2331–2349, Oct 2018.

[17] S. Rasp, M. S. Pritchard, and P. Gentine. Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, 115(39):9684–9689, Sep 2018.

[18] M. Sambridge, T. Bodin, K. Gallagher, and H. Tkalcic. Transdimensional inference in the geosciences. *Philosophical Transactions of the Royal Society A*, 371, 2013.

[19] T. Schneider, S. Lan, A. Stuart, and J. Teixeira. Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations. *Geophysical Research Letters*, pages 12396–12417, 2017.

[20] R. S. Schrom, M. van Lier-Walqui, M. R. Kumjian, J. Y. Harrington, A. A. Jensen, and Y.-S. Chen. Radar-based bayesian estimation of ice crystal growth parameters within a microphysical model. *Journal of the Atmospheric Sciences*, 78(2):549–569, feb 2021.

[21] M. van Lier-Walqui, H. Morrison, M. R. Kumjian, K. J. Reimel, O. P. Prat, S. Lunderman, and M. Morzfeld. A Bayesian approach for statistical-physical bulk parameterization of rain microphysics, Part II: Idealized Markov chain Monte Carlo experiments. *Journal of the Atmospheric Sciences*, 77(3):1043–1064, March 2020.

Suggested Partners/Experts

- **Yun Qian** (PNNL) Has performed PPE-based tuning on E3SM, has ample experience with parameter estimation for this model.
- **Kara Lamb** (Columbia University) Is exploring advanced methods for ML in Earth sciences, including for addressing structural errors.

Related White Papers

- **Ann Fridlind** (NASA/GISS) Dr. Fridlind's recommendation for a probabilistic MIP matches our own emulator-accelerated parameter estimation performed at the global scale. Additionally, Dr. Fridlind's white paper emphasizes the importance of laboratory process studies, something facilitated by structurally flexible physical parameterizations.
- **Peter Caldwell** (LLNL) Dr. Caldwell's proposal to expand the use of perturbed physics ensembles (PPEs) and ML emulators is an important step to constraining physically based microphysics in global ESMs.