Title: AI-Assisted Parameter Tuning Will Speed Development and Clarify Uncertainty in E3SM

Authors: Peter Caldwell (LLNL), Chris Golaz (LLNL), Peter Bogenschutz (LLNL), Marcus van Lier-Walqui (Columbia University), Aaron Donahue (LLNL), Chris Vogl (LLNL), Barry Rountree (LLNL), Aniruddha Marathe(LLNL), and Tapasya Patki (LLNL)

Focal Area: This idea is best aligned with **predictive modeling**. It targets Earth System Model improvement and uncertainty quantification

Science Goal: Use AI to efficiently tune uncertain physical parameters in the high-resolution E3SM model in order to optimize agreement with observations

Rationale: Most aspects of climate models are grounded in sophisticated and rigorous physical and mathematical formulae. Unfortunately, flaws in these formulations exist and are amplified by imperfect interactions between processes, yielding simulations which disagree with observed climate. This conflict between bottom-up physics-based model predictions and top-down observational constraints gives rise to a decidedly non-rigorous component of model development: parameter tuning. Reframing tuning as optimization of a mathematically-defined objective function would go a long way towards improving climate model credibility. Automating this process would save time, improve simulation quality, and ensure that tuning remains possible as models become more complex and move to higher resolutions.

Currently, the final tuning stage of an Earth system model (ESM) requires several person-years of effort running simulations with different parameter settings in an educated guess-and-check mode, using expert judgement to choose which configuration is best based on a broad and sometimes poorly defined set of climate metrics. **This manual tuning places a huge burden on scientists' time and opens climate modeling to accusations of lacking rigor (Hourdin et al., 2017, Schmidt, et al., 2017).**

Model formulation uncertainties are currently either ignored or approximated by the spread of a relatively small set of models which have each been tuned to provide a single 'best-guess' set of parameters. In the few studies where multiple realistic parameter settings were explored (Stainforth et al., 2005, Mauritsen et al., 2012, Golaz et al., 2013), parameter choice was generally found to have an important effect on climate change predictions. **Producing several sets of "best" parameters would go a long way towards quantifying parametric uncertainty in climate models but is currently impractical given the cost of manual tuning.**

Efficient tuning would also allow modeling centers to easily tell whether a new candidate parameterization will actually improve the integrated model or not. Currently, new parameterizations are evaluated without retuning the model because doing so would be prohibitively expensive and time consuming. Because errors in existing schemes are usually masked by compensatory tuning, new schemes often perform worse because they upset balances between compensating errors. Conversely, untuned new schemes often seem to solve long-standing model problems… but offer little benefit once retuned. A great example of this is

DOI: https://doi.org/10.2172/1769663
DOE BER, AI4ESP 2021 #AI4ESP1021 (https://www.ai4esp.org)

the balance between reflective stratocumulus clouds and dim trade cumulus clouds (Caldwell et al., 2019, Guo et al., 2021). Models tend to produce clouds with similar shortwave reflection everywhere, so tuning for energy balance requires stratocumulus which are too dim and shallow convection which is too bright. Thus new stratocumulus-focused parameterizations invariably make stratocumulus look better in initial tests but look about the same after retuning. **If tuning was automated and cheap enough that it could be performed for each new parameterization, modeling centers could be much more objective in their decision to accept or reject a new parameterization.**

<u>**Narrative:**</u> Figure 1 illustrates the 4 steps needed to automate the E3SM tuning process. Step 1: Uncertain physics parameters in E3SM are explored using a **perturbed physics ensemble** (PPE).

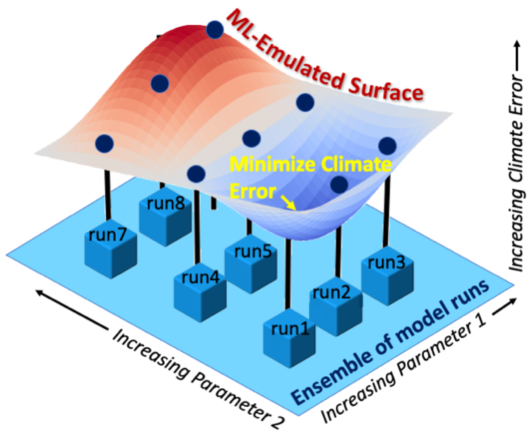

Figure 1: Graphical depiction of optimization strategy

Step 2: A **metric of climate skill** is created and applied to each simulation. Step 3: A machine-learning-based **emulator** is created which predicts this climate skill metric as a function of the PPE's uncertain parameters in a computationally efficient way. Step 4: the emulator is probed for combinations of parameters which **optimize climate skill**. Although this plan is straight-forward, operational use in E3SM will require intensive research into each of the steps above. Emulation is the primary use of AI in this project, but ensemble creation and optimization also require AI-related technologies.

Because each major change to E3SM requires a new PPE, PPE efficiency is essential. To take advantage of DOE Leadership Computing Facilities queueing policies and to provide the parallelism required for exascale architectures, many runs need to be bundled into each job submission. AI-based operating-system level optimization to interleave calculations between ensemble members will be necessary for performance improvement. Though Latin hypercube sampling is efficient, the shear number of uncertain climate parameters poses challenges. Previous AI-assisted tuning efforts have adjusted as many as 46 parameters at once (Elsaesser et al., 2020), but needed PPE size increases quickly as number of parameters increases. For tractability, manual tuning is typically done by adjusting just a few parameters at a time. What is the optimal tradeoff between tuning parameters and computational efficiency? Down-selecting parameters via a preliminary PPE before performing the actual emulation study may be required. If so, teasing out the most important yet orthogonal set of parameters to target will be an interesting AI problem.

Creation of the climate skill metric is probably the most important and most challenging step of this process. As noted above, it is hard to define a complete list of important climate aspects (though many partial lists like www.cesm.ucar.edu/working_groups/Atmosphere/metrics.html exist). We expect definition of an appropriate metric to be an iterative process where an optimum set of parameters is found, it's corresponding model simulation is evaluated by experts and deficiencies are identified, and these deficiencies are codified in a new climate skill metric for the next round of evaluation. As a result, a byproduct of this effort will be a better codification of what aspects of a simulation developers actually care about. Regularizing variables with different units and magnitudes is also necessary and will require thought.

Emulating climate metrics isn't cutting-edge science, but using cheaper model versions to emulate a model too expensive to sample frequently is. We know of one climate study which accomplishes this (Anderson and Lucas, 2018) by simply including model resolution as a tuning parameter in their random-forest emulator. Using this approach, they were able to get skillful precipitation and top-of-atmosphere radiation prediction from just 2 ensemble members per uncertain parameter rather than the 10 simulations per parameter typically required for Latin-hypercube sampling (Loeppsky et al., 2009). Further research in this area is needed. In particular, machine learning algorithms that are able to leverage the fundamental difference between model resolution and tuning parameters may yield better skill. For example, transfer learning applied to a low-resolution baseline to develop a high-resolution neural net could reduce overall cost. Similarly, Gaussian Process Estimation (GPE) could first be applied to low-resolution data to develop the prior used for GPE emulation of high-resolution data. These and other techniques should be compared to identify the most accurate and efficient approach to using cheap model runs to build faithful emulators for more expensive climate simulations. Note that emulation for parameter optimization doesn't require perfect accuracy since predicted optimal parameter settings will be validated via actual E3SM simulations and eventual science applications will run the physical model rather than its emulator.

Best-performing parameters can be identified either in a probilistic way by seeking the most likely set of parameters for the climate skill metric given uncertain observational data or by ignoring uncertainty and simply applying optimization methods to the skill metric. Markov Chain Monte Carlo (MCMC) in particular has proven to be successful and popular for climate applications (e.g. Lee et al., 2011, Elsaesser et al., 2020, Cleary et al., 2020). Because GPE provides a measure of its own uncertainty in emulating actual model output, it is particularly attractive to use in probilistic approaches, which can then incorporate information about emulator error in their predictions (e.g. Johnson et al., 2015). Optimization techniques which ignore uncertainty have also been applied successfully (e.g. Zhang et al, 2015), but to our knowledge only using smaller sets of uncertain parameters. Because having multiple "best" sets of possible parameters was identified in the Rationale Section as being important, optimization strategies will need to be local in nature rather than seeking a single global optimum.

As highlighted in the Rationale Section, this plan would transform climate science by making it more objective, by enabling better understanding of parametric uncertainty, and by freeing its lead developers to pursue science instead of tuning. The steps to this plan are straightforward and an initial implementation could be accomplished quickly, but an efficient/effective solution will require concerted effort. A major benefit of this plan is that it uses AI effectively without replacing physical-relationship-based governing equations with black-box relationships trained on current-climate data and known to not extrapolate well.

Suggested Partners/Experts:

- Ken Carslaw is brilliant, a good speaker, and an expert in this area
- This project is synergistic with white papers by:
    - Benj Wagman
    - Marcus van Lier-Walqui

Bibliography:

Anderson, G.J. and Lucas, D.D. (2018). Machine Learning Predictions of a Multi-Resolution Climate Model Ensemble. *Geophysical Research Letters*, 45, 4273-4280. https://doi.org/10.1029/2018GL077049

Caldwell, P. M., Mametjanov, A., Tang, Q., Van Roekel, L. P., Golaz, J.-C., Lin, W. et al. (2019). The DOE E3SM coupled model version 1: Description and results at high resolution. *Journal of Advances in Modeling Earth Systems*, 11, 4095– 4146. https://doi.org/10.1029/2019MS001870

Cleary, E. , Garbuno-Inigo, A., Lan, S., Schneider T., and Stuart, A. M. (2021). Calibrate, Emulate, Sample. *J. Computational Phys.* 424. https://doi.org/10.1016/j.jcp.2020.109716

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

Golaz, J.-C., Horowitz, L. W., and Levy, H. (2013), Cloud tuning in a coupled climate model: Impact on 20th century warming, *Geophys. Res. Lett.*, 40, 2246– 2251, doi:10.1002/grl.50232.

Guo, Z., Griffin, B. M., Domke, S., & Larson, V. E. (2021). A parameterization of turbulent dissipation and pressure damping time scales in stably stratified inversions, and its effects on low clouds in global simulations. *Journal of Advances in Modeling Earth Systems*, 13. https://doi.org/10.1029/2020MS002278

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., Rauser, F., Rio, C., Tomassini, L., Watanabe, M., & Williamson, D. (2017). The Art and Science of Climate Model Tuning, *Bulletin of the American Meteorological Society*, *98*, 589-602. https://doi.org/10.1175/BAMS-D-15-00135.1

Johnson, J. S. , Cui, Z., Lee, L. A. , Gosling, J. P., Blyth, A. M., and Carslaw, K. S. (2015). Evaluating uncertainty in convective cloud microphysics using statistical emulation. *Journal of Advances in Modeling Earth Systems*, 7:162–187. https://doi.org/10.1002/2014MS000383

Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W., and Spracklen, D. V. (2011). Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters. *Atmospheric Chemistry and Physics*, 11, 12253–12273. https://doi.org/10.5194/acp-11-12253-2011

Loeppky, J.L., Sacks, J., and Welch, W.J. (2009). Choosing the Sample Size of a Computer Experiment: A Practical Guide, *Technometrics*, 51:4, 366-376. https://doi.org/10.1198/TECH.2009.08040

Mauritsen, T., et al. (2012), Tuning the climate of a global model, *J. Adv. Model. Earth Syst.*, 4, doi:10.1029/2012MS000154.

Schmidt, Gavin A., et al. (2017) "Practice and Philosophy of Climate Model Tuning across Six US Modeling Centers." *Geoscientific Model Development* 10, 3207–23. https://doi.org/10.5194/gmd-10-3207-2017.

Stainforth, D., et al. (2005). Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* 433, 403–406. https://doi.org/10.1038/nature03301

Zhang, T., Li, L., Lin, Y., Xue, W., Xie, F., Xu, H., and Huang, X. (2015): An automatic and effective parameter optimization method for model tuning, *Geosci. Model Dev*., 8, 3579–3591. https://doi.org/10.5194/gmd-8-3579-2015