A Hybrid Climate Modeling System Using AI-assisted Process Emulators

Jiali Wang, Rao Kotamarthi, Virendra Ghate, Bethany Lusch, Prasanna Balaprakash, Justin M. Wozniak, Xingqiu Yuan, William Pringle, Pengfei Xue, Julie Bessac, Won Chang

Focal Area

This white paper addresses Focus Area II. We advocate developing a hybrid modeling system to improve the understanding of decadal- and longer-scale predictability of high impact water cycle components. This hybrid model combines a partial differential equation (PDE)-based dynamic core with AI/ML based emulators to represent many of the computationally expensive processes in Earth's climate models. The hybrid modeling system has the potential to exploit emerging graphics processing unit (GPU)-accelerated architectures and allows for the generation of large ensemble (~1000's) simulations to better characterize the model uncertainty and understand predictability.

Science Challenge

The predictability of climate and weather models is generally assessed by determining the rate of error growth. The most commonly used method is to conduct ensemble simulations with small perturbations introduced in initial conditions; then the growth rate of the error from ensemble mean is obtained for a better assessment of the predictability. As several processes occur and change at much finer scales than the resolutions of current available climate models, they are often parameterized with multiple tunable parameters. Hence, very large member ensemble simulations are required for better understanding the model error, sensitivity to model initializations, parameter uncertainties, and the impact of data assimilation. This is particularly important for the water cycle component (e.g., precipitation processes) which is the most difficult to simulate in Earth system models. However, it is computationally infeasible to perform more than a few dozen simulations at horizontal spatial resolutions of 50 to 100 km, let alone convection-permitting scale that is needed for high impact events (e.g., extremes). As a result, we have been constrained to few-member ensembles to explore the vast array of possible predictions; and assume the limited ensemble members can represent the wide range of uncertainty due to perturbations introduced into the models. The hybrid modeling system can potentially address this critical scientific knowledge gap by enabling a large ensemble size and offer us better opportunities to understand model predictability for water cycle components. However, developing hybrid models that combine AI-assisted emulators with a PDE-based dynamic core is challenging largely due to the differences in software ecosystems for scientific computing and data science. This white paper discusses AI/ML and advanced computational technology that can potentially address these challenges and enable to answer the following scientific questions. While we focus on Great Lakes region because of our expertise and its novelty and complexity in terms of modeling, the methodology illustrated can be applied to assess the predictability of similar water cycle components over other regions.

- What are the sources of model errors and uncertainties in simulated clouds, precipitation and evapotranspiration over the Great Lakes region?
- Will large member ensembles help us identify improved regions of predictability when initialized with appropriate data (data assimilation)?
- Do hybrid models provide a faster model solution that can address the impacts of human activities on regional surface water cycle components?

Rationale

Gaps and needs: The barriers to Earth system predictability center on our inability to assess the predictability of this complex, coupled, non-linear system using methods of non-linear stability analysis because of the vast range of parameter spaces, such as errors in initializations, uncertainties in physics

DOI: https://doi.org/10.2172/1769645
DOE BER, AI4ESP 2021 #AI4ESP1142 (https://www.ai4esp.org)

schemes, and the accumulated model errors. This inability has led to the development of more empirical methods for assessing the model predictability, such as performing an ensemble of parameter perturbation in the physics parameterization; then the ensemble is assessed by traditional statistical techniques to explore the regions of predictability that are conditioned on selective parameters. However, a component of model error is inherently random (aleatory), and modeling them in a deterministic way limits predictability and contributes to structural uncertainties. Therefore, in addition to reducing the epistemic component of model error, we need to better understand the source of model errors and uncertainties through perturbation of model initial conditions, physical parameterizations and parameters, as well as external forcing. This requires a fast, accurate and stable modeling system that couples AI-based emulators (representing small scales) with PDE-based dynamic solvers (representing large scales) using advanced computational techniques (e.g., software integration, coupling, workflows) to achieve long time-period and/or large ensemble simulations.

Previous studies have developed emulators for complex climate models based on purely numerical simulations (e.g., Krasnopolsky 2005; Wang et al. 2019). This could be problematic especially for small-scale and fast-moving processes, which are often not or only partially resolved in these numerical models (e.g., the missing interaction between neighboring cells through cold pools; the three-dimensional nature of radiation). These processes may be directly trained by in-situ observations, remote sensing data, and super-resolution simulations. However, much remains unknown about how much data from the past and present could be used to train models for future prediction and how to make sure these models conserve important physical properties. In addition, even the emulators perform well offline it does not guarantee a stable online coupling with the dynamic solver because of many challenges as listed below.

While most of the emulators are developed and tested offline, there are studies that train the emulators while coupled with the dynamical solvers (online). However, these studies use very simple models and ignore important aspects (e.g., cloud water and ice) for simplicity (Brenowitz and Bretherton 2019; Maulik et al. 2019). An ongoing challenge, even for simple models, is to maintain the stability and forecast skill of the hybrid modeling system for a sufficiently long time period. Integrating emulators into existing PDE dynamic solvers is a substantial software architecture research challenge. First, the representation of data in the PDE models and in the emulators may differ greatly (e.g., processes, scales). Second, there is a concurrency management problem as emulators compete for resources with existing PDE models, and the relative computation times and resource requirements may change dynamically (e.g., time-stepping). Third, monitoring and debugging models that mix intermediate results from PDE models and emulators will be difficult (e.g., exponential growth of errors). Overall, the development of a hybrid modeling system will face similar challenges when compared to the development of conventional models such as the complexity of the Earth system with nonlinear interactions between model components, scale interactions, exponential growth of errors in initial conditions, numerical instabilities, and many more.

Benefits: The hybrid modeling system will be a crucial element of the Data-Model Integration Scientific Grand Challenge in the EESSD strategic vision. We will leverage Argonne's expertise developed from coupling large computing frameworks with general-purpose data science tools to deploy a hybrid modeling system that will be able to achieve long time-period and large-member ensemble simulations. The system will be able to (1) represent all relevant scales and features over the globe, (2) allow scale interactions and (3) represent chains of complex interactions between weather features. The system can offer immediate advantages over currently achievable small ensembles or short simulations of a climate model. It will allow us to conduct high-resolution simulations in both space and time scales (e.g., call radiative forcing every time step) and identify regions that require more expensive simulations for reducing model errors or more observations for better understanding certain mechanisms. The rich data will also provide an opportunity to re-evaluate the commonly used statistical methods (e.g., annual block maxima for generalized extreme value theory; and assumption of linear trend in time). Last but not least, the use of observational data in developing these emulators will provide the ability to identify critical observations and develop numerical/observational and computational strategies to improve both their representation and the predictability of the model.

Narrative

This whitepaper proposes to build individual emulators for each physical process, which allows us to identify each physical process's contribution to model error and to tune the parameters of each parameterization while coupled with the PDE-based numerical model. However, individual emulators will not be able to consider the coupling effect between different physical processes. We could use AI techniques to test their coupling effect, for example, for a given physical process and the data, we can build a diverse set of AI/ML model ensembles using scalable neural architecture search. These model ensembles can be combined and calibrated against observations using model fusion and meta learning techniques. We will include the parameters of each parameterization as a dependence in the emulators. When we train the emulators we will optimize the parameters of each parameterization by leveraging scalable reinforcement learning, Bayesian optimization, and derivative-based/free optimization methods which can optimize the parameters more efficiently than conventional random sampling or grid search.

We propose that developing emulators based on both observation data and very high-resolution simulations (e.g., large eddy simulation scale) is necessary to develop the appropriate hybrid models that will allow us to explore the predictability questions. Data examples include the Atmospheric Radiation Measurement's observatory and routine high resolution modeling at the Southern Great Plains. For certain locations, if we have both observations and model simulations, we can use observations to improve the simulations. This improvement most likely can also help the neighboring locations, for example, using transfer learning and continual learning methods. Similar learning methods can be also helpful when the emulator trained on historical data is used for a different climate state. In this case re-training is required with a small amount of data from the new climate state.

To couple AI/ML-based emulators and complex PDE-based dynamic solvers and respond to the software architecture research challenges we listed in **Rationale**, we will leverage Argonne's strong expertise in software development and data science, and develop (1) data adapters that allow in-memory, distributed data to be presented to the emulators; (2) lightweight workflow-like components to distribute work on demand; and (3) decision tracking components to capture the data that goes in and out of the emulators. At the Argonne Leadership Computing Facility (ALCF), as we prepare for central processing unit (CPU) and GPU hybrid supercomputers such as the upcoming exascale Aurora machine, we are researching ways to couple AI and numerical simulations. One big advantage of replacing process models with ML-based emulators is that ML models tend to run efficiently on GPUs. We could first distribute the training of emulator ensembles across GPUs; once the training is nearly complete, we could further refine the emulators with online training, using the loss across the whole hybrid model, with the dynamic core on CPUs. Finally, creating training data will require massive computational resources for direct numerical simulations, especially at very high spatial and temporal resolutions, but this could be feasible at the ALCF. The auto-ML algorithms can facilitate the process of feature reduction and searching for an optimal network topology and greatly improve the productivity.

A Hybrid Climate Modeling System Using AI-assisted Process Emulators

Suggested Partners/Experts (Optional)

References (Optional)

1. Wang, J., Balaprakash, P., and Kotamarthi, V.R. (2019): Fast domain-aware neural network emulation of a planetary boundary layer parameterization in a numerical weather forecast model, Geosci. Model Dev., 12 (10), 4261-4274. https://doi.org/10.5194/gmd-12-4261-2019
2. Krasnopolsky, V. M., Fox-Rabinovitz, M. S., and Chalikov, D. V.: New approach to calculation of atmospheric model physics: Ac- curate and fast neural network emulation of long wave radiation in a climate model, Mon. Weather Rev., 133, 1370–1383, 2005.
3. Brenowitz, N. D., & Bretherton, C. S. (2019). Spatially extended tests of a neural network parametrization trained by coarse-graining. *Journal of Advances in Modeling Earth Systems*, *11*, 2728–2744. https://doi. org/10.1029/2019MS001711
4. Maulik, R., San, O., Rasheed, A., & Vedula, P. (2019). Subgrid modelling for two-dimensional turbulence using neural networks. Journal of Fluid Mechanics, 858, 122-144. doi:10.1017/jfm.2018.770