A Multi-Scale Inference, Estimation, and Prediction Engine for Earth System Modeling

M. Anghel, B. Nadiga, N. Panda, A. Mohan, E. Hunke, C. Begeman
Los Alamos National Laboratory

Primary Focus Area: Predictive modeling via AI techniques. This paper suggests a framework for combining AI approaches with Earth system data and models. Here we attempt to unify AI approaches for a variety of water cycle applications presented in other white papers, to chart a multi-decadal investment plan for Earth system predictability.

Science challenge: We posit that AI methods can be leveraged to significantly enhance the predictive skill of forward Earth system modeling (ESM) activities. A hybrid framework incorporating traditional ESM modeling, inference methods, and AI techniques could make better use of both measured and computed information as well as computational resources by targeting inference tasks at program priorities, such as the predictability of precipitation extremes. This runtime pathway to closing the simulation/analysis-data/model improvement loop will streamline the traditional offline pathway to model improvement, which is based on domain science expertise, while suggesting guidance for further observations and measurements.

Rationale: ESMs are already hierarchical, multi-scale models, evolving the global Earth system through approximate numerical solutions to the laws of fluid dynamics and thermodynamics at ambitious scales ($O(10^7)$ cells in the atmosphere and $O(10^8)$ cells in the ocean). Explicitly resolving processes that operate on sub-grid scales remains too computationally demanding for the foreseeable future, and yet they are essential for decadal or longer simulations capturing water cycle extremes. For example, the major winter precipitation producers in the desert Southwest are often the result of fine-scale interactions within large-scale atmospheric patterns, necessitating representations of cloud microphysics and local interactions with mountain topography. Furthermore, the large-scale air masses and pressure systems are influenced by global teleconnections, bringing atmospheric moisture (from the south and west) and back-door cold fronts (from the east) into the region. Back-door cold fronts in the Southwest are extensions of the polar vortex, which is driven by the temperature gradient between the tropics and the Arctic. Thus, the predictability of Southwest precipitation is intertwined with changing conditions in the Arctic, also associated with small-scale processes such as sea ice fracturing.

To study these complex, multi-scale, water cycle problems, we propose a multi-scale ESM in which the large scale prognostic equations are coupled with subgrid-scale dynamics models (SGSM). Because of the strong dependence of subgrid dynamics on ambient conditions (ESM state variables), skillful representation of subgrid processes requires prognostic evolution of subgrid dynamics and physics in embedded small-scale models. The strategy presented here involves significantly reducing the computational cost of running this multi-scale ESM, enabling AI-mediated *in-situ* inference, uncertainty quantification, bias mitigation and data-model integration. We call this generic framework a *Multi-Scale Inference, Estimation, and Prediction Engine*. The approach is two-pronged. First, we use model reduction techniques to accelerate simulations of Earth system processes and to enable the efficient estimation of model parameters and their uncertainties. Second, we employ several AI methods for integrating observations and simulations. The goal is to leverage available data to enhance model capabilities, discover relationships between different scales, and reduce uncertainty and bias in predictions.

Components of the Overarching Framework

DOI: https://doi.org/10.2172/1769648
DOE BER, AI4ESP 2021 #AI4ESP1003 (https://www.ai4esp.org)

1) AI techniques present opportunities for learning **Reduced Order Models** (ROMs) from observed or simulated data. ROMs can make embedding small- and meso-scale processes in each column of a large-scale ESM computationally feasible[9,10,12]. Since an accurate, global, low-dimensional ESM does not exist, we propose to decompose the coarse-scale ESM model into a number of local interacting subsystems, each coupled locally to a SGSM. This framework is further discussed by Mohan et al.[8] Examples include cloud-resolving models, discrete element sea ice models, or vegetation models[12]. Elliott et al. offer an innovation for ROMs that represent chemical interactions, combining traditional ML approaches with chemistry knowledge from observations[4].

(2) In general, ROMs for climate dynamics do not provide stable models that can predict its evolution over long time scales. **Distributed algorithms are needed for assessing the global long-term stability of the coupled low-dimensional approximations,** to detect where and when to correct the ROMs using observations or retrain them by improving the basis. Among the many AI approaches to the construction of ROMs, we suggest a POD-Galerkin projection because it can yield local ROMs in the form of polynomial ODEs, for which we have existing distributed strategies to estimate their stability. We can then learn structural corrections to the ROMs that provide dynamic stability. Tools to discriminate the intrinsic instability of the climate from the instability due to the imperfection of ROMs will be necessary, a fundamental question of identifying structural bias.

(3) **Data-guided parameter and state estimation** in global- and local-scale models can be posed as an optimization problem, to minimize the difference (cost) between the predicted model behavior and observations[8]. In their white paper, Bennet et al. discuss how AI-based data assimilation can be used for parameter estimation in the context of extreme hydroclimate events[2]. We propose that the embedded ROMs be constructed for both *forward* and *backward* (adjoint) models to achieve efficient parameter and state estimation. Derivatives of the cost function may be used to predict the response due to changes in the parameters, a Bayesian AI framework to estimate parameter uncertainties, and a Hamiltonian Monte Carlo approach to efficiently sample the posterior distribution in parameter space. We envision *local* inference strategies to estimate these parameters in parallel given the local (ESM, SGSM) pair models and data. Dubey et al. and Lin et al. explore how we might learn process representations from a combination of observations and model output[3,7]. Schwenk and Ren provide a tool for leveraging satellite data for a wide variety of data-driven tasks[11].

(4) **Forward propagation of uncertainties must be efficient** for long-term climate predictions using a range of emission scenarios and mitigation strategies. The construction of forward ROMs for the evolution of ESMs is not sufficient for the efficient propagation of uncertainties attached to the ESM trajectories. Moments propagation algorithms, like polynomial chaos expansion methods, or distributed methods to estimate approximations to the Peron-Frobenius operator that describe the evolution of densities over state and parameters, may be explored[9].

(5) Efficient implementation of ***in-situ* analysis methods** will probe the behavior of ESMs and efficiently use the massive amount of data they generate. Lawrence et al. demonstrate how statistical models may be fit *in situ* to characterize extreme event probability[6]. Begeman et al. propose a complementary deep-learning technique called Fractal Nets to analyze the underlying dynamics of both the observed and simulated Earth system and make predictions from parallel analysis of pairs of local data streams[1]. These Fractal Nets could allow us to address structural

model biases, understand hidden dynamical links (e.g., teleconnections that affect drought frequency and severity), and predict rare events. Rupe et al. suggest yet another approach through Transfer Operators to identify and predict coherent structures such as extreme events in the water cycle[10].

(6) **Adaptive Sampling (AS) strategies** to mitigate the prohibitive computational cost of directly embedding SGSMs. The AS strategy learns the map that provides (local) fine-scale feedback to the coarse ESM models. Since we expect the SGSM to explore a large number of dynamic regimes, we expect to adaptively update their reduced order approximations. AS can substantially reduce the total number of expensive fine-scale updates and produce a library of SGSMs, which is only updated when significant drift in the large-scale driving forces is detected. An AS library will operate at the coarse scale as well, since dynamic drift at this scale will require ROMs for the ESM to be updated as the system evolves.

(7) **Efficient experimental design** (ED) ascertains in real time the location, type, resolution and sampling rate of measurements necessary to improve predictive accuracy. Uncertainty quantification and sensitivity analyses using adjoint methods are critical for implementing ED strategies, which are dependent on the desired accuracy and the choice of climate variables. ED strategies must also weigh competing goals, for example, between prediction accuracy and prediction uncertainty. Fractal Nets can be used to determine the location and the observable with maximal causal influence on other climate observables, such as extreme precipitation, informing where to collect observations without an adjoint[1]. Hardin et al. discuss considerations for experimental design of atmospheric campaigns using Reinforcement Learning[5].

(8) **Real time detection of unanticipated events and anomalies, and their visualization,** are also *in-situ* analysis tasks. Many of these events are lost or require re-running a simulation with additional diagnostics, when an *a posteriori* analysis is not possible. Price et al. advocate for an AI-mediated identification of the precursors to anomalous ice-sheet events[9]. The implementation of such tools[2,5,10] will significantly enhance our understanding of climate and its impacts.

This Multi-Scale Inference, Estimation, and Prediction Engine can advance the study of extreme water cycle events. One starting place is coupling a cloud-resolving, forward SGSM with the ESM, and using its adjoint for uncertainty propagation and parameter tuning. AS strategies would be used to augment training data to improve SGSM fidelity. Fractal Nets or Transfer Operators in combination with learning from large-scale (e.g., satellite) data can be used to detect the teleconnections important for a local area, and to detect changes in those large-scale dynamics on climatic timescales. ROMs could also be used to simulate orographic effects, leveraging local data (e.g., stream gauges and the ARM SAIL campaign in Colorado) to learn local precipitation dynamics. In situ analysis methods would be embedded in the simulations for early detection of anomalous precipitation events, and to provide additional data such as high temporal-resolution output for the precipitation event and watershed effects. Implementing these AI methods and representations within an ESM will require significant software engineering advances to efficiently balance these tasks and their memory requirements. The framework that we propose is modular, employing generic AI, modeling, and inference methods (model order reduction, adaptive sampling, Markov Chain Monte Carlo) that transcend traditional boundaries between scientific disciplines.

References

1. Begeman, C., M. Anghel, I. Chattopadhyay. Interpretable Deep Learning for the Earth System with Fractal Nets. White paper submitted to DOE/BER/EESSD AI4ESP call, February 2021.
2. Bennett, K., M. Vesselinov, S. Karra. Improved characterization and parameterization of extreme events in E3SM. White paper submitted to DOE/BER/EESSD AI4ESP call, February 2021.
3. Dubey, M., J. Benedict, B. Nadiga, P. Johnson, H. Viswanathan, P. Chylek & S. Cairns. AI for Extreme Volcanic Climate Forcing and Feedback Forecasting in the 21^{st} century. White paper submitted to DOE/BER/EESSD AI4ESP call, February 2021.
4. Elliott, S., A. Piliouras, B. Nadiga, F. Hoffman, J. Kumar, J. Tang, U. Mishra, et al. Jaynesian Analysis of Environmental Chemistry: Systems Model Component Integration via the Arctic Aquatic Carbon Cycle. White paper submitted to DOE/BER/EESSD AI4ESP call, February 2021.
5. Hardin, J. C. et al. The Usage of Observing System Simulation Experiments and Reinforcement Learning to Optimize Experimental Design and Operation. White paper submitted to DOE/BER/EESSD AI4ESP call, February 2021.
6. Lawrence, E., A. Biswas, L. Van Roekel, N. Urban. In situ inference for Earth system predictability. White paper submitted to DOE/BER/EESSD AI4ESP call, February 2021.
7. Lin, Y., C. Begeman, and J. Schwenk. Physics-Consistent Deep Learning Method with Adaptive Data Augmentation. White paper submitted to DOE/BER/EESSD AI4ESP call, February 2021.
8. Mohan, A., Z. Cao, M. Anghel, M. Maltrud, C. Sweeney, M. McKerns, X. Asay-Davis. Multiscale Reduced Order Modeling and Parameter Estimation for Climate Sciences. White paper submitted to DOE/BER/EESSD AI4ESP call, February 2021.
9. Price, S., M. Hoffman, M. Perego. Rapid sea level rise from Antarctica: early detection and uncertainty quantification. White paper submitted to DOE/BER/EESSD AI4ESP call, February 2021.
10. Rupe, A., V. Vesselinov, B. Nadiga, D. DeSantis, M. Anghel. Transfer Operator Framework for Earth System Predictability and Water Cycle Extremes. White paper submitted to DOE/BER/EESSD AI4ESP call, February 2021.
11. Schwenk, J., C. Ren. A new era of observationally-informed E3SM: cycleGAN for unifying imagery archives. White paper submitted to DOE/BER/EESSD AI4ESP call, February 2021.
12. Xu, C. Learning the process structure of dynamic vegetation models. White paper submitted to DOE/BER/EESSD AI4ESP call, February 2021.