Title: Bridging Multiscale Processes in Earth System Models with Physics-Guided Hierarchical Machine Learning

Authors/Affiliations

Alexander Y. Sun[1], Bridget R. Scanlon[1], Clint Dawson[2], Paola Passalacqua[3], Dev Niyogi[3,4], Zong-Liang Yang[4], Susanne Pierce[5]
[1]Bureau of Economic Geology; [2]Department of Aerospace Engineering and Engineering Mechanics; [3]Department of Civil, Architectural and Environmental Engineering; [4]Department of Geological Sciences; [5]Texas Advanced Computing Center (all with The University of Texas at Austin)

Focal Area(s)

The focal area of this whitepaper is, "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)." The ideas and frameworks described herein are deemed site agnostic. As a use case, this group will initially focus on the coupling between land processes, surface/subsurface hydrological processes, coastal processes, and human activities in the U.S. Gulf states.

Science Challenge

The major science challenge addressed in this whitepaper is *how to improve the predictive capabilities of Earth system models (ESM) by combining a hierarchy of machine learning (ML) models to bridge scales and compensate for the missing and/or inaccurate process representation in different ESM components*. The theme of this whitepaper is directly related to the following grand challenge listed in the EESSD's Strategy Plan, Advance next-generation understanding of Earth system drivers and their effects on the integrated Earth-energy-human system.

Rationale

Research needs or gaps:

All events in the dynamically coupled Earth system ultimately result from numerous complex, interrelated processes acting across a wide range of spatial and temporal scales (Swain et al., 2020). The multiscale nature of these drivers/processes is further complicated by spatiotemporal feedbacks, climatic teleconnections, and by interactions between human activities and the Earth system's natural variability and trends. As a result, consideration of fine-scale processes is critical for accurate prediction of climate extremes, in the assessment of local- and regional-scale vulnerability to climate change, and for accurate assessment of the impact of climate extremes (Diffenbaugh et al., 2005). The predictive capability of many Earth system models (ESMs) and global hydrological models is presently hampered by (a) incomplete process understanding and representation, especially with regard to how changes propagate across interfaces within the Earth system components and across disciplinary boundaries (Blöschl et al., 2019); (b) the coarse grid resolution (10–100 km) used in many regional and global models; (c) suboptimal regionalization and inversion techniques used for estimating system parameters; and (d) lack of a wide spectrum of testbeds and benchmarks for validating model outcomes and for identifying the sources of uncertainty.

There is currently enormous interest in applying AI/ML to improve the predictive capability of ESMs and to enable rapid forecasting (Reichstein et al., 2019; Sun and Scanlon, 2019). Many Earth science ML models, however, have focused on the so-called single-domain, single-resolution uses, in which a model is trained to predict a single class of variables of interest (e.g., radiation, precipitation, streamflow, or soil moisture) by using a limited set of observations and/or model simulations that are

DOI: https://doi.org/10.2172/1769682
DOE BER, AI4ESP 2021 #AI4ESP1128 (https://www.ai4esp.org)

run at a fixed resolution. The resulting ML models typically do not allow for easy coupling across ESM domains. Most ML models are not tested over a wide range of dynamics, do not extrapolate well on out-of-class samples, and do not possess adaptive learning capabilities.

Barriers to progress:

We identify the following barriers to developing ML-enabled, next-generation integrated Earth-energy-human system models that result from the lack of:

(1) Explainable and interpretable AI/ML approaches for process identification and representation, especially relating to resolving subgrid processes and constitutive relations;

(2) Flexible schemes for combining process-based and ML models (i.e., hybrid modeling) across common ESM scales and subdomains;

(3) Domain-agnostic approaches for integrating ML models and submodels trained at different scales and on different components of an ESM, and

(4) ML-enabled large-scale rapid initial state forecasting and system optimization capability, which is required for ensemble-based forecasting and uncertainty quantification.

Expected benefits:

We envision a new hierarchical machine learning (HML) paradigm to address many of the aforementioned challenges/barriers, with a specific focus on extending the existing DOE model experiment (MODEX) frameworks, including E3SM, and testbeds to assess the impacts of climate extremes and human activities in coastal areas (**Figure 1**). Ultimately, we envision a "digital twin" of the Earth system to be created, which not only enables scenario/perturbation testing, but also communication between researchers and policy makers.

Narrative

Scientific and technical description of the opportunities and approach:

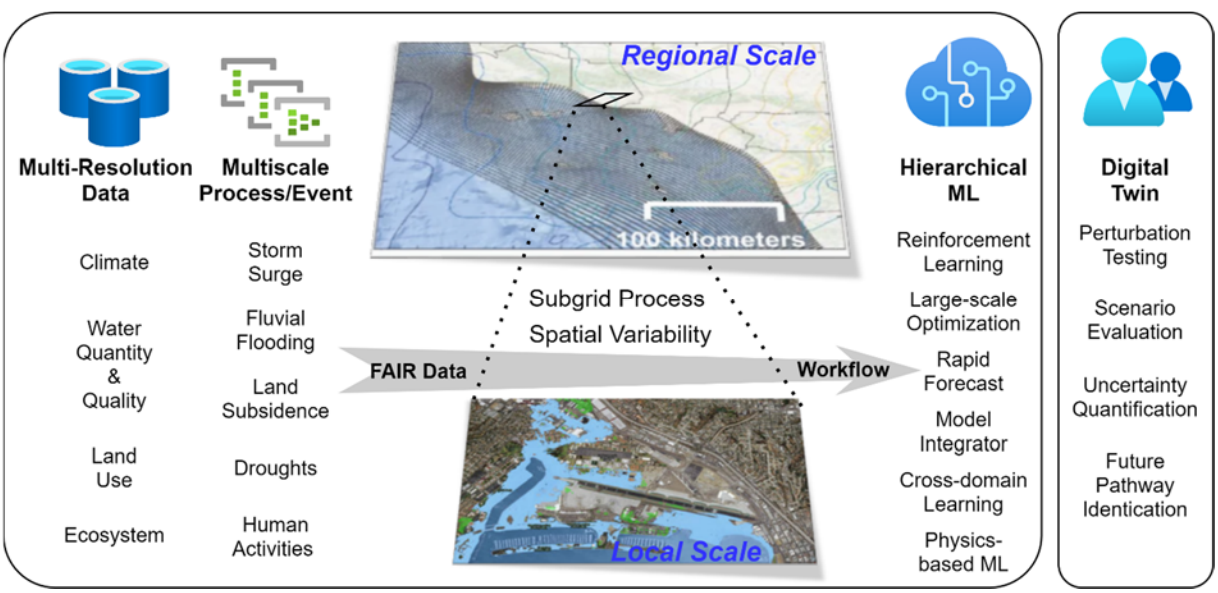

Figure 1. A framework for enabling a Digital Twin of the Earth system by integrating multiresolution Earth observations, multiscale process models, and hierarchical ML.

Both dry and wet events over the contiguous United States are projected to increase by the end of the century, in a spatially heterogeneous manner (Hsiang et al., 2017). Globally, climate change is intensifying tropical cyclones, accelerating sea-level rise, and increasing coastal flooding. Coastal areas and river deltas are especially vulnerable to flooding because of their low elevations and densely populated cities (Edmonds et al., 2020). ML-enabled, high-fidelity ESMs are needed to test the response of the Earth system to external forcing in a manner meaningful to both modelers and decision makers.

Instead of developing and using many disparate ML models, as is usually done in current practices, we envision a hierarchical machine learning (HML) framework to unify multi-purpose and multiscale ML models for ESM. Our HML framework incorporates elements of recent developments in (a) cross-domain ML mapping (Kim et al., 2017), (b) semantic-based automated workflow based on the findability, accessibility, interoperability, and reusability (FAIR) data principles (Goble et al., 2020); (c) federated learning for large-scale systems (Bonawitz et al., 2019), and (d) hybrid process-based and ML modeling. The outcomes of the HML framework will be used to enable a digital twin of the Earth system.

We envision such an HML framework to consist of three layers. The base layer of the HML will consist of physics-informed ML and deep generative learning models to tackle model initial state, subgrid process representation, model fidelity, and process closures (Long et al., 2018; Meng and Karniadakis, 2020). The barriers to running the ESM models at higher resolution are (a) computational power; (b) limits to model physics and parameterizations used (e.g. some cloud schemes only work beyond certain grid spacing), and (iii) lack of input data at high resolution. The ML models incorporate the underlying physics to regularize learning processes, thus mitigating the issue of limited training samples. Cross-domain mapping, representing one of the most interesting achievements in the modern deep learning era (Zhu et al., 2017), provides a new data-driven approach for coupling different Earth science processes across scales and in both forward and inverse directions (Sun and Tang, 2020; Sun, 2018), mitigating the impact of incomplete process understanding and inaccurate parameterization.

The middle layer of the HML will consist of the domain agnostic model integrators. For this purpose, the HML will use a FAIR-based computational workflow to describe and automate a complex multi-step end-to-end process, involving data collection, data preparation, model simulation, and ML training and predictive analytics. Model integration will be built on existing standards [e.g., OpenMI (Harpham et al., 2019)] but will be extended to be compatible with ML models. The workflow will benefit from recent advances in general workflow management systems (e.g., KNIME) and cloud-based container orchestration (Kubernetes) to scale the computations required for training and coupling the myriad of ML models.

The top layer of the HML, which is the user-facing side of the envisioned digital twin, will provide functionalities to fine tune the ESM outputs and to enable applications, including metalearning, ensemble generation, and scenario generation and optimization. Reinforcement learning will be implemented using a hierarchy of ML surrogate models to enable the rapid assessment of pathways, as well as impact assessment of system perturbations induced by extreme events.

The HML-based workflow will be demonstrated for the U.S. Gulf Coast, for which the group has collectively amassed significant modeling expertise and data (e.g., high-resolution DEM, GPS, soil moisture, land use, and surface and groundwater quantity and quality). By working with partners, we will also identify a current DOE-BER SFA as an additional use case.

Activities that will advance the science:

The HML framework lays out a number of novel components for enabling AI/ML applications in the EESSD domain, filling the gap in data integration, multiscale ML-model coupling, and system integration and optimization. Implementing such a system will maximize the existing DOE-BER investment on E3SM development, significantly improving capabilities for coping with a changing climate.

Suggested Partners/Experts (Optional)

- PNNL ESS-SFA
- ASGC and JGCRI earth system modeling teams
- Texas GLO

References

Blöschl, G. et al., 2019. Twenty-three unsolved problems in hydrology (UPH)–a community perspective. Hydrological sciences journal, 64(10): 1141-1158.

Bonawitz, K. et al., 2019. Towards federated learning at scale: System design. arXiv preprint arXiv:1902.01046.

Diffenbaugh, N.S., Pal, J.S., Trapp, R.J., Giorgi, F., 2005. Fine-scale processes regulate the response of extreme events to global climate change. Proceedings of the National Academy of Sciences, 102(44): 15774-15778.

Edmonds, D.A., Caldwell, R.L., Brondizio, E.S., Siani, S.M., 2020. Coastal flooding will disproportionately impact people on river deltas. Nature communications, 11(1): 1-8.

Goble, C. et al., 2020. FAIR computational workflows. Data Intelligence, 2(1-2): 108-121.

Harpham, Q., Hughes, A., Moore, R., 2019. Introductory overview: The OpenMI 2.0 standard for integrating numerical models. Environmental Modelling & Software, 122: 104549.

Hsiang, S. et al., 2017. Estimating economic damage from climate change in the United States. Science, 356(6345): 1362-1369.

Kim, T., Cha, M., Kim, H., Lee, J.K., Kim, J., 2017. Learning to discover cross-domain relations with generative adversarial networks, International Conference on Machine Learning. PMLR, pp. 1857-1865.

Long, Z., Lu, Y., Ma, X., Dong, B., 2018. Pde-net: Learning pdes from data, International Conference on Machine Learning. PMLR, pp. 3208-3216.

Meng, X., Karniadakis, G.E., 2020. A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse PDE problems. Journal of Computational Physics, 401: 109020.

Reichstein, M. et al., 2019. Deep learning and process understanding for data-driven Earth system science. Nature, 566(7743): 195-204.

Sun, A., Tang, G., 2020. Downscaling Satellite and Reanalysis Precipitation Products Using Attention-Based Deep Convolutional Neural Nets. Front. Water 2: 536743. doi: 10.3389/frwa.

Sun, A.Y., 2018. Discovering state ‐ parameter mappings in subsurface models using generative adversarial networks. Geophysical Research Letters, 45(20): 11,137-11,146.

Sun, A.Y., Scanlon, B.R., 2019. How can big data and machine learning benefit environment and water management: A survey of methods, applications, and future directions. Environmental Research Letters.

Swain, D.L., Singh, D., Touma, D., Diffenbaugh, N.S., 2020. Attributing extreme events to climate change: a new frontier in a warming world. One Earth, 2(6): 522-527.

Zhu, J.-Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks, Proceedings of the IEEE international conference on computer vision, pp. 2223-2232.