AI-Constrained Bottom-Up Ecohydrology and Improved Prediction of Seasonal, Interannual, and Decadal Flood and Drought Risks

Forrest M. Hoffman[1], Jitendra Kumar[1], Zheng Shi*[1], Anthony P. Walker[1], Jiafu Mao[1], Yaoping Wang*[2], Abigail L. S. Swann[3], James T. Randerson[4], Umakant Mishra[5], Gabriel J. Kooperman[6], Hyungjun Kim[7], Chonggang Xu[8], Charles D. Koven[9], David Lawrence[10], Megan Fowler*[10], Martin De Kauwe[11], Belinda Medlyn[12], Lianhong Gu[1], Liz Agee*[1], Jeff Warren[1], Shawn Serbin[13], Alistair Rogers[13], Trevor F. Keenan[9], Nate McDowell[14], Nathan Collier[1], Sarat Sreepathi[1], Juan Restrepo[1], Rick Archibald[1], Feng Bao*[15], Richard Tran Mills[16]

*[1]Oak Ridge National Laboratory; [2]University of Tennessee; [3]University of Washington; [4]University of California Irvine; [5]Sandia National Laboratories; [6]University of Georgia; [7]University of Tokyo; [8]Los Alamos National Laboratory; [9]Lawrence Berkeley National Laboratory; [10]National Center for Atmospheric Research; [11]University of New South Wales; [12]Western Sydney University; [13]Brookhaven National Laboratory; [14]Pacific Northwest National Laboratory; [15]Florida State University; [16]Argonne National Laboratory; *Early Career*

Focal Areas

(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 composed of a hierarchy of models

(3) Insight gleaned from complex data (both observed & simulated) using AI, big data analytics, and other advanced methods, including explainable AI and physics- or knowledge-guided AI

Science Challenge

A framework that employs AI methods for hybrid process-/ML-based modeling in a hierarchy of models and data analytics can address uncertainties in ecohydrology by connecting plant community structure, hydraulics, and physiology, along with soil composition, biogeochemistry, and physics, to the water cycle. Doing so will (1) improve predictability, (2) fill in gaps where mechanistic insights are lacking (e.g., mesophyll conductance, rooting depth, groundwater access and recharge), and (3) stimulate hypothesis creation to drive theoretical development aimed at reconciling seasonal-to-interannual flood and drought risk for watersheds and interannual-to-decadal flood and drought risk from regions to continents.

Rationale

Water cycle extremes leading to floods or droughts pose risks to freshwater and food security, environmental sustainability, energy and transportation infrastructure, and human lives. Medium-range weather forecasts and seasonal-to-interannual climate predictions have significantly improved as model resolution has increased and atmospheric model tuning has advanced. However, prediction uncertainties associated with land–atmosphere interactions have only increased and represent a growing source of error in simulations. Soil moisture, which has inertial memory of weeks to years, influences the atmosphere through changes in evapotranspiration and the surface energy budget that affect air temperature and precipitation. Soil moisture is a function of inputs (precipitation) minus outputs (evapotranspiration, surface runoff, and groundwater access and recharge). The latter are largely controlled by seasonally variable vegetation cover and function, which also change over time because of rising atmospheric carbon dioxide levels and increasing global temperatures. Consistent with the four grand challenges in Earth and

DOI: https://doi.org/10.2172/1769668
DOE BER, AI4ESP 2021 #AI4ESP1060 (https://www.ai4esp.org)

environmental sciences identified as potentially benefiting from the application of AI methods (Chapter 2, Stevens et al., 2020), a framework that employs ML approaches could address uncertainties in ecohydrology by connecting plant physiology and hydraulics and soil composition and dynamics to soil moisture and water cycle extremes to improve predictability of flood and drought risk. Recent research has demonstrated strong but poorly quantified influences of rising CO_2 on soil moisture (Swann et al., 2016), global streamflow (Fowler et al., 2019), rainfall (Kooperman, Chen et al., 2018), and flood risk (Kooperman, Fowler et al., 2018) through plant physiological and hydrological effects, moderated by plant stress and drought (Ukkola et al., 2016). Addressing these uncertainties with ML will advance science in support of the five research questions for the Integrated Water Cycle Scientific Grand Challenge identified in the EESSD's Strategic Plan (US DOE, 2018). Moreover, this research will benefit from domain-specific explainable and knowledge-informed AI methods, will require high performance computing and neuromorphic resources and could offer a test bed for analog mixed stochastic/deterministic processes (e.g., light harvesting) suitable for quantum computing.

Narrative

Given the availability of large volumes of observational data and *in situ* measurements, and rapidly expanding computational resources, the Earth system modeling community is adopting data-driven approaches for high resolution weather and climate simulations (Schneider et al., 2017), and DOE is poised to respond to such a community blueprint to improve Earth system predictability. An AI framework could be used to integrate the wealth of leaf-level fluorescence and gas exchange measurements (e.g., Leafweb), AmeriFlux and FLUXNET ecosystem fluxes, and Free Air Carbon Dioxide Enrichment (FACE) and Spruce and Peatland Responses Under Changing Environments (SPRUCE) data to develop a unified treatment of stomatal responses, assimilation, and acclimation to changes in hydrology and soil moisture. Specifically, we suggest ML-based models of stomatal conductance and plant hydraulics be employed to produce a hybrid process-based/ML-based land model for DOE's E3SM with the aim of reducing the uncertainty of soil moisture and carbon assimilation. Such hybrid ecohydrology models could also inform watershed models to deliver dynamic ecological process representations often absent in such models. In addition, ML models can be created to improve the characterization of soil organic carbon and soil bulk properties to further reduce soil moisture uncertainties. Moreover, ML methods should be explored to scale leaf-level and ecosystem processes to the watershed scale for seasonal-to-interannual predictions, through a hierarchy of ML and process-based models, and further to regional and continental scales for interannual-to-decadal predictions. Since plant and soil processes respond to climate change, FACE and SPRUCE data should be used to develop climate-adaptive ML models for the processes described above.

Part 1. Hybrid AI Models of Plant Physiology and Hydraulics

This approach could enable significant steps forward in developing and integrating new and alternative parameterizations within E3SM to produce a hybrid process-based/ML modeling framework (Reichstein et al., 2019). The requirement for reducing uncertainties in ecohydrological processes dictates prioritizing process representations of land–atmosphere interactions (energy, water, and carbon) that (1) are highly uncertain but for which observational constraint data are available and (2) are computationally expensive. A specific example is the *CanopyFluxesMod* module within the E3SM Land Model (ELM) that iteratively solves coupled equations for temperature, moisture, and carbon that underpin metabolism, growth, respiration, and transpiration processes. It is the most computationally expensive science component of the ELM, and a large volume of leaf-level and ecosystem measurements, now meeting FAIR data standards (Ely et al., 2021), are available to develop an array of ML-based models of stomatal behavior and plant hydraulics for different optimal growth strategies and different degrees of model complexity (Fisher and Koven, 2020). This effort could pioneer deployment of stochastic network methods on mixed CPU-GPU architectures. Emerging ML methods combined with data assimilation to handle nonlinear and non-Gaussian processes could be applied within large, complex coupled systems, like E3SM, to perform uncertainty quantification for process-based model components. It may also be possible to reduce uncertainties from some parameters or variables, enabling exploration of uncertainties in the other mechanistic process representations. Because many of the mechanisms involved may be represented as stochastic processes, known memory/time quantum advantages for stochastic simulations might be leveraged in their study on future quantum computers.

Part 2. AI/ML Methods for Spatial prediction of Fluxes and Soil Characteristics

Since soil bulk properties, and soil organic carbon in particular, control soil moisture and evapotranspiration, water cycle predictions are strongly dependent on correctly characterizing those properties in models. Recent studies show that AI/ML approaches show promise in improving spatial prediction of soil and environmental properties, including soil organic carbon stocks and land surface fluxes, with field data (Mishra et al., 2020). ML approaches have also been used to derive scaling functions of soil and ecosystem properties (Adhikari et al., 2020). AI-based scaling functions can be used to upscale soil properties and processes for model initialization and downscale them from coarse Earth system/land surface model predictions for evaluation. New ML methods can decompose signals from time series flux data to improve causal inference from environmental drivers to improve process understanding. ML approaches can be used to derive relationships among environmental factors, plant water uptake, water extremes, and soil properties that can be used as potential process-based benchmarks for baseline connections between fluxes and soil conditions in ESMs.

Part 3. Scaling Leaf- and Organism-Level Responses to Ecosystems and Watersheds

Measurements of leaf-level responses to environmental variations are used to develop process understanding for models of photosynthesis and stomatal conductance, and the models are then used in an integrated framework (e.g., *CanopyFluxesMod* module in ELM) to simulate canopy-level fluxes. Using a ML approach, such measurements could be directly related to the measurements made at the canopy scale reducing uncertainties in the canopy integration schemes. Similarly, ML methods can be applied to scale up plant responses—informed by ecosystem- and watershed-scale measurements, upscaled soil properties, and remote sensing data—to bound water budgets for watersheds and quantify risks of flooding and drought, particularly under water cycle extremes. Cluster analysis, Kriging, random forests, and various forms of artificial neural networks can be applied now; but newer, more sophisticated mathematical methods, including backward stochastic differential equations combined with ML and ensembles of stochastic emulator simulations, could better capture dynamics across scales. While the fundamental motivation is to improve mechanistic understanding of these processes across scales, by connecting this chain of hierarchical ML-empowered models to weather forecasting systems, the results can be used to inform probabilistic risk analysis to quantify flood risks for urban areas and other built infrastructure and to better quantify drought impacts on streamflow for energy and water utilities.

Suggested Partners/Experts

Roles of White Paper Contributors

Domain Model-Data Analytics Experts: Forrest M. Hoffman[1], Jitendra Kumar[1], Zheng Shi*[1], Anthony P. Walker[1], Jiafu Mao[1], Yaoping Wang*[2], Abigail L. S. Swann[3], James T. Randerson[4], Umakant Mishra[5], Gabriel J. Kooperman[6], Hyungjun Kim[7]

Modelers and Model Developers: Chonggang Xu[8], Charles D. Koven[9], David Lawrence[10], Megan Fowler*[10], Martin De Kauwe[11], Belinda Medlyn[12]

Laboratory and Field Data Experts: Lianhong Gu[1], Liz Agee*[1], Jeff Warren[1], Shawn Serbin[13], Alistair Rogers[13], Trevor F. Keenan[9], Nate McDowell[14]

Mathematicians and Computer Scientists: Nathan Collier[1], Sarat Sreepathi[1], Juan Restrepo[1], Rick Archibald[1], Feng Bao*[15], Richard Tran Mills[16]

References

Adhikari, K., Mishra, U., Owens, P. R., Libohova, Z., Wills, S. A., Riley, W. J., Hoffman, F. M., and Smith, D. R. 2020. Importance and strength of environmental controllers of soil organic carbon changes with scale, *Geoderma*, 375, 114472, doi:10.1016/j.geoderma.2020.114472.

Ely, K. S., et al. 2021. A reporting format for leaf-level gas exchange data and metadata, *Ecol. Inform.*, 61, 101232, doi:10.1016/j.ecoinf.2021.101232.

Fisher, R. A., and Koven, C. D. 2020. Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems, *J. Adv. Model. Earth Syst.*, 12(4), e2018MS001453, doi:10.1029/2018MS001453.

Fowler, M. D., Kooperman, G. J., Randerson, J. T., and Pritchard, M. S. 2019. The effect of plant physiological responses to rising CO_2 on global streamflow, *Nature Clim. Change*, 9:873–879, doi:10.1038/s41558-019-0602-x.

Kooperman, G. J., Chen, Y., Hoffman, F. M., Koven, C. D., Lindsay, K., Pritchard, M. S., Swann, A. L. S., and Randerson, J. T. 2018. Forest response to rising CO_2 drives zonally asymmetric rainfall change over tropical land, *Nature Clim. Change*, 8(5):434–440, doi:10.1038/s41558-018-0144-7.

Kooperman, G. J., Fowler, M. D., Hoffman, F. M., Koven, C. D., Lindsay, K., Pritchard, M. S., Swann, A. L. S., and Randerson, J. T. 2018. Plant-physiological responses to rising CO_2 modify simulated daily runoff intensity with implications for global-scale flood risk assessment, *Geophys. Res. Lett.*, 45(22):12,457–12,466, doi:10.1029/2018GL079901.

Mishra, U., Gautam, S., Riley, W. J., and Hoffman, F. M. 2020. Ensemble machine learning approach better predicts the spatial heterogeneity of surface soil organic carbon stocks in data-limited northern circumpolar region, *Frontiers in Big Data*, 3, 40, doi:10.3389/fdata.2020.528441.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat. 2019. Deep learning and process understanding for data-driven Earth system science, *Nature*, 566(7743): 195–204, doi:10.1038/s41586-019-0912-1.

Schneider, T., Lan, S., Stuart, A., and Teixeira, J. 2017. Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations, *Geophys. Res. Lett.*, 44, 12,396–12,417, doi:10.1002/2017GL076101.

Stevens, R., Taylor, V., Nichols, J., Maccabe, A. B., Yelick, K., and Brown, D. 2020. *AI for Science*, United States, doi:10.2172/1604756.

Swann, A. L. S., Hoffman, F. M., Koven, C. D., and Randerson, J. T. 2016. Plant responses to increasing CO_2 reduce estimates of climate impacts on drought severity, *Proc. Nat. Acad. Sci.*, 113(36):10019–10024, doi:10.1073/pnas.1604581113.

Ukkola, A., Prentice, I., Keenan, T. F., van Dijk, A. I. J. M., Viney, N. R., Myneni, R. B., and Bi, J. 2016. Reduced streamflow in water-stressed climates consistent with CO_2 effects on vegetation, *Nature Clim. Change*, 6:75–78, doi:10.1038/nclimate2831.

US DOE. 2018. *Earth and Environmental Systems Sciences Division Strategic Plan 2018–2023*, DOE/SC-0192, U.S. Department of Energy, Office of Science, https://science.osti.gov/-/media/ber/pdf/workshop-reports/2018_CESD_Strategic_Plan.pdf