Hybridizing Machine Learning and Physically-based Earth System Models to Improve Prediction of Multivariate Extreme Events (AI Exploration of Wildland Fire Prediction)

Yufei Zou[1], Philip J. Rasch[1], Hailong Wang[1]

[1]Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA.

Focal Areas: This project responds to two focal areas identified in the DOE Call for AI4ESP White Papers: 1) Predictive modeling through the use of artificial intelligence (AI) techniques, and 2) insights gleaned from complex data using explainable AI and big data analytics.

Science Challenge: Large wildland fires (hereafter wildfires) appearing as high-impact compound climate extreme events are closely related to hydroclimate and water cycle extremes that modulate surface fuel supply and combustibility [1]. These compound events have multivariate climatic features (e.g., temperature, precipitation, relative humidity, wind, lightning) and societal drivers (e.g., forest management, land use change, human caused ignitions) [2-5]. Meanwhile, they induce strong feedbacks to the coupled atmosphere, biosphere, and hydrosphere by perturbing regional and global radiation budget as well as ecological, biogeochemical, and water cycles across multiple spatiotemporal scales [5-7]. The nonlinear interactions between these natural and anthropogenic components of the Earth system are too complex to be completely and adequately represented in today's Earth system models (ESMs) [8]. The inherent stochastic nature of fire activity at all scales further increases the difficulty of its prediction using ESMs that are usually developed from deterministic equations and parameterizations [9]. Besides, concurrence of long-term (decadal to interdecadal) global climate change and fire regime shifts overlapping with short-term (intraseasonal to interannual) variations of regional fire weather and burning activity confound predictability of these compound extreme events.

We propose to address the above scientific challenges by using machine learning (ML)-based data-driven modeling techniques to integrate observations and physically-based ESMs' simulations in a computationally efficient hybrid prediction system. This prediction system is supposed to characterize the wildfire's sensitivity to climate and exogenous drivers at high resolution (~ 0.25°) on subseasonal to seasonal (S2S) timescales providing improved predictability and explainability. We will use the system to help identify: (1) What are the computational elements of a hybrid system needed to predict compound climate extreme events such as global wildfires? (2) What are the key drivers (either natural or anthropogenic) that modulate short-term variations of multivariate fire weather and burning activity over different regions? How can one take advantage of those driver-response relationships to improve the predictability of large wildfires on S2S time scales? (3) What are the underlying physical mechanisms and sources of improved predictability? Which ML techniques are optimal in revealing and adapting these mechanisms?

Rationale: Robust driver-response relationships existing on different spatiotemporal scales lay the foundation for predictability of climate extreme events for either physically-based ESMs or statistical/empirical models. Successful S2S prediction of global wildfires relies on correctly capturing both climate teleconnections (which modulates local fire weather and surface fuel) and human influence (both ignition and suppression effects) on top of seasonal variations in climate. Unfortunately, current fire models (using either empirical or process-based approaches) [10-12] have shown limited skills in simulating and predicting large wildfires at shorter and finer scales, especially over extratropical regions, due to incomplete understanding of the complex

DOI: https://doi.org/10.2172/1769718
DOE BER, AI4ESP 2021 #AI4ESP1155 (https://www.ai4esp.org)

interactions between large-scale circulation-driven hydroclimate changes (e.g., drought), synoptic fire weather fluctuations (e.g., lightning/gusty wind), and vegetation dynamics (e.g., fuel supply/aridity) as well as large uncertainties associated with human influence (e.g., fire ignition/prevention/suppression). The coarse resolution grids (~ 1°) and deficient structure of current generation ESMs also limit their modeling capability to simulate fire-related climatic and ecological feedback processes (e.g., fire plumes and their interactions with cloud systems; fire-induced vegetation dynamics) at sub-grid fine scales. Therefore, global fire prediction and evaluation of fire impacts do not improve dramatically as model resolution increases owing to the above structural deficiencies in present ESMs, even given a very substantial increase in computational cost for high-resolution ESM simulations.

To improve understanding of the factors that influence fire predictability, it is necessary to continuously improve almost every aspect of the major components (i.e., atmosphere, land, ocean, sea ice) of ESMs by better simulating fire-related dynamic, physical, and ecological processes associated with both natural and human dimensions of Earth system variability [8]. However, this pathway is computationally expensive and demanding as discussed above. An alternative approach is to build data-driven statistical/empirical models that bypass incomplete physical understanding of fire processes and computational limitations in ESMs to meet the realistic needs of fire risk assessment and management [9]. Although this data-driven approach is attractive due to lower development and implementation cost, it has its own challenges and disadvantages like scarce data and data structure/quality issues for model training and evaluation, difficulties in determining complete model predictors and describing nonlinear and nonstationary driver-response coupled relations, and a lack of explainability of prediction results. Traditional statistical models (e.g., multiple linear regression models; autoregressive integrated moving average models) used by previous studies [10, 11] are more susceptible to these problems than rapidly developing ML and deep learning (DL) models [13] that usually have weaker model assumptions and less susceptibility to poor data quality but also require larger data quantity and more sophisticated model training processes.

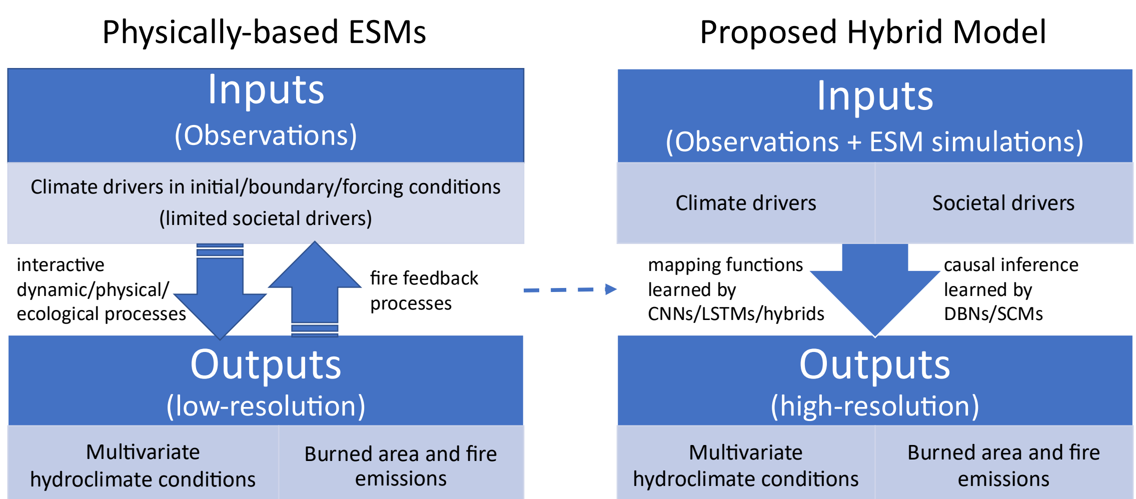

Figure 1 Comparison of different predictive modeling approaches

The distinct characteristics of different modeling approaches along with increasingly available data from both global remote sensing and improved ESMs (e.g., a new process-based RESFire model with region-specific climate-fire-ecosystem interactions [14]) and rapidly advancing hardware/software development environment (e.g., CPU-GPU heterogeneous computing) motivate us to consider a hybrid way of integrating physically-based ESMs and data-driven ML/DL models for a high-resolution global fire prediction system with improved predictability and explainability at reduced computational costs (Fig. 1). We believe this system will provide insight into the physical mechanisms explaining multivariate hydroclimate changes and sheds light on the prediction of other compound climate extreme events. It would also meet the realistic needs for reliable early warning of severe fire hazards and serve as a benchmark for guiding and evaluating fire modeling development in the next-generation exascale ESMs such as E3SM.

Narrative: The proposed hybrid prediction system will be built using predictors identified with appropriate ML algorithms such as the Random Forest method that take into account comprehensive climatic and anthropogenic driving factors, and robust input-output mapping information learned from both observations and current coarse resolution ESMs. We suggest some candidate predictors for the fire prediction system would include both natural (e.g., global gridded SST, SIC, soil moisture, etc.) and human (e.g., population density, traffic and power networks, land use change, etc.) factors, which provide more comprehensive and abundant driving force information than what were used by previous empirical fire models such as selective regional averaged SST or synthetic climate indices [10, 11].

Many ML/DL algorithms are candidates for fire predictive modeling. We are particularly interested in using DL neural networks (e.g., convolutional neural networks (CNNs) [13], long short-term memory networks (LSTMs) [15], or their hybrids) that are designed to automatically learn complex (linear and/or nonlinear) mapping functions from multiple inputs to outputs with several appealing features including robustness to noise and nonlinearity as well as spatiotemporal feature and dependence learning [16], which make them suitable for global fire prediction with strong seasonality and nonlinearity. We also plan to use causal inference methods (e.g., dynamic Bayesian networks (DBNs) [17], structural causal models (SCMs) [18]) when focusing on system explainability (i.e., causal interdependencies of the underlying system) rather than predictability, which are complementary to neural networks that tend to be overparameterized at the expense of interpretability [16].

We suggest that these methods could be applied to analyzing both global observational fire data (e.g., GFEDv4 [19, 20]) and ensembles of fire model output data from ESMs (e.g., FireMIP results under different climate scenarios [21, 22]) for training and evaluation of the new hybrid system. Fidelity measures could use those employed in the International Land Model Benchmarking (ILAMB) system [23]. A combination of ML/DL algorithms and multi-source (i.e., observations and ESM simulations) big data analytics can simultaneously improve S2S global fire prediction skills and reveal underlying climate-fire teleconnection mechanisms at a lower cost. Moreover, the application of causal inference methods in analysis and evaluation of physically-based ESM simulations based on their underlying causal interaction structures provides a new perspective other than traditional ways using climate variable statistics to understand the similarity and differences between models and observations [18].

The source code and generated datasets of the proposed hybrid global fire prediction system will be deposited to publicly accessible code hosting and data repository platforms such as GitHub for evaluation, reproduction, and continuous development. The high-resolution global fire prediction datasets are expected to benefit practical fire risk management and responses in federal and state agencies such as Federal Emergency Management Agency (FEMA) and California Department of Forestry and Fire Protection (CALFIRE) as well as several DOE-funded scientific focus areas and research programs such as NGEE-Arctic, NGEE-Tropics, RUBISCO, and HiLAT-RASM for improved Earth system predictability.

References

1. Krawchuk, M.A. and M.A. Moritz, *Constraints on global fire activity vary across a resource gradient.* Ecology, 2011. **92**(1): p. 121-132.
2. Aldersley, A., S.J. Murray, and S.E. Cornell, *Global and regional analysis of climate and human drivers of wildfire.* Science of the Total Environment, 2011. **409**(18): p. 3472-3481.
3. Zscheischler, J., et al., *Future climate risk from compound events.* Nature Climate Change, 2018. **8**(6): p. 469-477.
4. Raymond, C., et al., *Understanding and managing connected extreme events.* Nature Climate Change, 2020. **10**(7): p. 611-621.
5. Marlon, J.R., et al., *Long-term perspective on wildfires in the western USA.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(9): p. E535-E543.
6. Zou, Y.F., et al., *Using CESM-RESFire to understand climate-fire-ecosystem interactions and the implications for decadal climate variability.* Atmospheric Chemistry and Physics, 2020. **20**(2): p. 995-1020.
7. Li, F. and D.M. Lawrence, *Role of Fire in the Global Land Water Budget during the Twentieth Century due to Changing Ecosystems.* Journal of Climate, 2017. **30**(6): p. 1893-1908.
8. Hantson, S., et al., *The status and challenge of global fire modelling.* Biogeosciences, 2016. **13**(11): p. 3359-3375.
9. Taylor, S.W., et al., *Wildfire Prediction to Inform Fire Management: Statistical Science Challenges.* Statistical Science, 2013. **28**(4): p. 586-615.
10. Shen, H.Z., et al., *Global Fire Forecasts Using Both Large-Scale Climate Indices and Local Meteorological Parameters.* Global Biogeochemical Cycles, 2019. **33**(8): p. 1129-1145.
11. Chen, Y., et al., *Forecasting Global Fire Emissions on Subseasonal to Seasonal (S2S) Time Scales.* J Adv Model Earth Syst, 2020. **12**(9): p. e2019MS001955.
12. Hantson, S., et al., *Quantitative assessment of fire and vegetation properties in simulations with fire-enabled vegetation models from the Fire Model Intercomparison Project.* Geoscientific Model Development, 2020. **13**(7): p. 3299-3318.
13. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* Nature, 2015. **521**(7553): p. 436-444.
14. Zou, Y.F., et al., *Development of a REgion-Specific Ecosystem Feedback Fire (RESFire) Model in the Community Earth System Model.* Journal of Advances in Modeling Earth Systems, 2019. **11**(2): p. 417-445.
15. Hochreiter, S. and J. Schmidhuber, *Long short-term memory.* Neural Computation, 1997. **9**(8): p. 1735-1780.
16. Reichstein, M., et al., *Deep learning and process understanding for data-driven Earth system science.* Nature, 2019. **566**(7743): p. 195-204.
17. Murphy, K.P., *Dynamic Bayesian networks: representation, inference and learning*, in *Computer Science*. 2002, University of California, Berkeley.

18. Runge, J., et al., *Inferring causation from time series in Earth system sciences.* Nature Communications, 2019. **10**.
19. Giglio, L., J.T. Randerson, and G.R. van der Werf, *Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4).* Journal of Geophysical Research-Biogeosciences, 2013. **118**(1): p. 317-328.
20. van der Werf, G.R., et al., *Global fire emissions estimates during 1997-2016.* Earth System Science Data, 2017. **9**(2): p. 697-720.
21. Li, F., et al., *Historical (1700-2012) global multi-model estimates of the fire emissions from the Fire Modeling Intercomparison Project (FireMIP).* Atmospheric Chemistry and Physics, 2019. **19**(19): p. 12545-12567.
22. Rabin, S.S., et al., *The Fire Modeling Intercomparison Project (FireMIP), phase 1: experimental and analytical protocols with detailed model descriptions.* Geoscientific Model Development, 2017. **10**(3): p. 1175-1197.
23. Collier, N., et al., *The International Land Model Benchmarking (ILAMB) System: Design, Theory, and Implementation.* Journal of Advances in Modeling Earth Systems, 2018. **10**(11): p. 2731-2754.