AI predicted shifts in watershed hydrodynamics driven by extreme weather and fire

Erica Siirila-Woodburn[1], Cynthia Gerlein-Safdi[1], Juliane Mueller[1]
James Dennedy-Frank[1], Robinson Negron-Juarez[1];
[1]Lawrence Berkeley National Laboratory

Focal areas: *Primary* - (2) Predictive modeling through AI; *Secondary* - (3) Insight gleaned from complex data (both observed and simulated) using AI

Science challenge: Changes in climate are expected to exacerbate extreme weather conditions such as drought that will shape the landscape and contribute to the frequency and severity of wildfires. Following a wildfire, shifts in water interception, storage, and discharge alter the integrated hydrologic cycle in complex and sometimes counterintuitive ways [1] with potential for extreme rainfall and/or snowmelt to further modify a watershed's response. Precipitation frequency, magnitude, and phase changes will subsequently have unknown implications on water supply and flood hazard. Current models cannot provide robust predictions of these management-crucial hydrodynamic shifts.

Rationale*:* Model predictions of post-wildfire impacts on the hydrologic cycle are exceedingly limited for several reasons: 1) hydrologic processes are complex, and are impacted by bedrock through atmosphere interactions over a large range of spatio-temporal scales in practice limiting model predictions due to impractical computational expenses; 2) watershed hydrodynamics depend on local site attributes including fire severity, soil characteristics, vegetation type, and subsurface flow paths, which are difficult to measure; 3) the lack of *in-situ* data to characterize site characteristics and measure outcomes results in ill-parameterized and unconstrained models; and 4) as more observational data becomes available, the links between different model parameters are increasingly complex, and disparities between conceptualized and physical processes oftentimes becomes more apparent. This is especially true in the face of a changing climate, where a lack of historical analogues often acts as a "stress-test" on system physics. In some cases, models calibrated with historical datasets alone will inherently fail to simulate future conditions. Finally, 5) integrating big-data and models remains challenging, and classical approaches struggle to adapt. Thus, while advances in computing power and observational datasets offer new information at unprecedented resolutions, our ability to use these data to better understand and predict system behavior with improved models remains limited.

For a radical improvement in the predictive capability and fundamental understanding of wildfire and water to occur, a nexus of remote sensing, machine learning (ML), artificial intelligence (AI) and numerical modeling is required. Remote sensing can address points 2) and 3) above by downscaling large-scale data to the local scale to characterize sites and assess model performance [2]. ML and AI can address points 1) and 2) as they can provide surrogate models of the compute-intensive simulations, and thus reduce the computational burden. ML and AI are powerful tools capable of ingesting the large datasets generated from different sources *(in situ*, remote sensing, etc.) in order to untangle complex non-linear relationships between key site characteristics and hydrodynamic shifts, therefore aiding in issues related to points 1) and 4). Finally, AI and ML will help to assist difficulties in distilling big-data issues, addressing point 5).

By linking the impacts of climate change-enhanced fire risks to the water cycle, we can identify water "hotspots" where fire may have an outsized effect (societal, hydrologic, ecologic, and financial standpoints). While post-wildfire effects *directly* impact water flow (including increased runoff due to decreased soil permeability, decreased transpiration due to vegetation loss, and alteration of surficial and subsurface flow pathways given changes in hydrologic connectivity), several *indirect* impacts are noteworthy. These include changes to ecosystem health and habitats, such as the potential replacement of burnt native species with invasive ones. These changes have subsequent impacts on multi-sector water use (e.g. hydro-

DOI: https://doi.org/10.2172/1769660
DOE BER, AI4ESP 2021 #AI4ESP1123 (https://www.ai4esp.org)

electric power, agriculture, residential, etc.) which may limit the ability for demands to be met, as well as risk management decisions and infrastructure (including subsequent flooding, contamination, erosion, and landslides). These considerations are exceedingly important given that wildfires are more common in a changing climate, with significant financial and environmental costs. In 2020 wildfires caused 30 fatalities in the Western US and over $130 billion dollars in damage, not fully including potential damage to drinking water supplies.

Secondly, an associated benefit of this approach is the use of wildfires as emulators of the dangers of extreme conditions. Wildfires are both *caused by* and *amplified by* changing climate extremes. Specifically, droughts and the associated vegetation mortality and increased fuel loading contribute to wildfire ignition and severity; then, following the wildfire event, burn-scars are disproportionately impacted by extreme precipitation often resulting in extreme damage from flooding and landslides. In this regard, even though many future stressors to the environment are expected with climate change, post-wildfire effects are an especially attractive focal point to demonstrate ML and AI utility.

Narrative: The integration of ML with a) established and upcoming remote sensing observations and b) developments in the earth system and hydrologic models matched with exascale computing capabilities, provides promise that prediction and management of wildfire can be greatly improved in the next decade. Together, these technologies could potentially enable real-time management-relevant predictions of immediate- to-long-term water shifts in extreme weather following wildfires at a national scale, and the development of a post-wildfire assessment tool. Such an approach will supplement classical modeling approaches with surrogate models for reduced computational expense, novel scale-adaptive approaches for model selection, parameter translation, uncertainty quantification, and new developments in AI-assisted data assimilation. It will further use new, high-resolution satellite observations with advanced AI approaches to better understand the complex nonlinear relationships between key site characteristics and hydrodynamic shifts, and thus inform classical modeling approaches.

Data Integration - To observe wildfires and the hydrological response to post-fire precipitation, modern remote sensing platforms could be leveraged to understand various aspects of the water cycle. Soil moisture measurements from satellites such as CYGNSS and SMAP provide high-resolution soil moisture data key for understanding fire probability and severity, as well as understanding those conditions post-wildfire. These data can be complemented with soil moisture measurements from networks of micrometeorological towers flux towers (e.g. Ameriflux). Sensors with improved spatial, temporal and physical capabilities have potential to launch in the next decade. GRACE and its follow-on mission are able to capture shifts in groundwater storage, while the new ECOSTRESS and GEDI missions improve our understanding of vegetation response to stress and pre- and post-wildfire vegetation cover. A key satellite to quantify fire behaviour is GOES that provides images every 5 minutes. These high spatio-temporal resolutions datasets will give higher predictive power in the context of ML and AI frameworks. For example, in the presence of extreme conditions such as high-intensity precipitation events, the occurrence of sudden changes in surface water flows and sharp subsurface wetting fronts can be used to inform if these hotspots are more or less susceptible to sudden changes. Finally, more long-standing satellites such as Landsat, Sentinel, or MODIS can provide many key surface parameters, including vegetation indices, land surface type, or active/burning fires locations pre and post fire occurrence.

Data optimization - ML developments will be essential to assimilate data of different types and use such data for calibration and uncertainty quantification of process-based simulations. ML models are highly versatile and can approximate complex relationships between different types of data (numerical, categorical, different scales). Traditional approaches such as Gaussian process models are not well-suited to deal with the large variability of data associated with studying post-extreme event dynamics. The computational burden of making predictions with ML models is significantly lower than running the simulation, and thus using ML models during optimization and uncertainty quantification as surrogates of the expensive simulation will lead to substantial speedups.

Learning process interaction - The use of ML techniques (e.g transfer entropy) based on information theory can help uncover the interactive patterns of underlying processes that control hydrodynamics shifts. The strength of the relationships between target variables and predictor variables implies the relative importance of different driving factors in determining these shifts. The lead-lag times associated with each directional relationship could also inform how long such directional coupling takes place due to system inertia and nonlinearity. Integration of ML techniques and land-atmosphere models in a spatially explicit manner is key to determine hydrodynamics shifts. We envision that ML models can be constructed such that they are able to detect and take into account the different length scales over which different processes are interconnected. Secondly, because wildfires are now more frequent, more severe, and longer in duration, growing data around wildfire behavior can be leveraged in ML-enabled technologies. For example, ML can be used to understand triggers for the onset of the event and if factors contribute to its duration and spatial extent. ML-aided pattern recognition and classification will provide utility in this regard.

Predictive/surrogate modeling - After learning the spatio-temporal patterns of importance (i.e. the predictors), ML models can be used to determine the probabilities of hydrodynamic shifts and their relevant quantities. ML models will be developed such that they adhere to watershed physics constraints (such as topography) in order to make accurate predictions. New approaches for incorporating these constraints may be required. ML models can also be used to improve physical representations in process-based models, which would also help guide potential directions of future DOE model developments such as E3SM and FATES. These ML models can also be used to surrogate data into process-based models for physical processes that are computationally expensive or for physical schemes that do not currently exist. All of these possibilities will improve prediction of hydrological shifts and changes in the water budget as well as inform risk assessment and management. Proper tools could then be created to communicate these results in a timely manner to the public, first responders, stakeholders, and other interested partners.

Data-model integration and regions of potential interest - The approach outlined here will incorporate in-situ and remotely sensed data both for model development and validation. Near-global coverage of most remote sensing platforms relevant to this work means that study sites can be chosen anywhere in the world. Locations of particular relevance might include the Western United States, which have experienced both an increase in wildfires frequency, size, and severity over the last decades. This approach could include data integration from several other federal partners (e.g. USGS and NASA). Because the methods are site-agnostic, opportunities exist to leverage several existing DOE investments. Example test-beds include the site of the Watershed-Function SFA (Upper Colorado River Basin) or one of the field pilot sites of the NGEE-Tropics project (Puerto Rico, Brazil, or Panama), where wildfire occurrences have varying degrees of historical precedents. In particular, the Pantanal in Brazil, is the largest wetland in the world but has also experienced unprecedented wildfires in recent years, with devastating consequences for this unique ecosystem.

Relevance to EESSD and availability of tools and codes- Understanding hydrological shifts associated with wildfires and extreme weather events is highly relevant to the Integrated Water Cycle Grand Challenge (GC), the Biogeochemistry Scientific GC, and the Drivers and Responses in Earth System Scientific GC, and the Data-Model Integration Scientific GC. Tool developments through the above framework would use open source codes which would be publicly available online, as well as data through ESS-DIVE. These practices would lead to greater stakeholder engagement, aid in reproducibility and transparency, and promote greater scientific community engagement.

Suggested partners/experts:
General: E3SM-FATES team, CASCADE team, Rubisco team, UCs and CSUs;

Specific: Vincent Ambrossia (NASA), Ruby Leung (PNNL), Chonggang Xu (LANL), DWR, Laura Condon (UAZ), Brandon Collins (USFS), Chris Ruf (University of Michigan), Clara Chew (UCAR), Dave Sapsis (Cal-FIRE).

References:
[1] Maina and Siirila-Woodburn (2020): https://doi.org/10.1002/hyp.13568
[2] Negron-Juarez et al. (2020): https://doi.org/10.5194/bg-17-6185-2020