1. Title: Elucidating and predicting the dynamic evolution of water and land systems due to natural and energy-related forcings

2. Authors/Affiliations: Grant Bromhal (Lead), NETL; Chung Yan Shih and Gavin Liu, NETL Support Contractors; David Dzombak, Carnegie Mellon; Srikanta Mishra and Jared Schuetter, Battelle Memorial Institute; Alexander Sun, University of Texas; George Guthrie, LANL

3. Focal Area(s):
3. Insight gleaned from complex data (both observed and simulated) using AI, big data analytics, and other advanced methods, including explainable AI and physics- or knowledge-guided AI; &
1. Data acquisition and assimilation enabled by machine learning, AI, and advanced methods including experimental/network design/optimization, unsupervised learning (including deep learning), and hardware-related efforts involving AI (e.g., edge computing)

4. Science Challenge:
Interactions between water, land, and energy systems are complex and occur on a variety of scales, ranging from local to basinal to regional. Accurately predicting the behavior of ground water and surface water systems for 5-10 years and beyond requires an understanding of the current system and the ability to model both the natural system at scale and human-induced forcings related to energy and other activities. Artificial intelligence and machine learning (AI/ML) combined with modern compilation and integration efforts for U.S. groundwater and surface water systems present potential solutions to bolstering detailed physics-based models of these systems[1]. Big data tied with ML and physics-based modeling can drive breakthroughs in understanding the earth system, but research is often impeded by data access (e.g., privacy issues), quality, formats, gaps, multi-source, multi-scale, integration, and spatiotemporal challenges[2]. Effective integration of real data and simulated (synthetic) data that fill gaps is critical. Overcoming these complex data and model integration challenges will enable a transformational approach to acquiring enhanced understanding of environmental systems.

5. Rationale:
Human systems—particularly those associated with energy generation and use—impact water resources across many scales and through a variety of processes. Attempts to understand natural-water systems dynamics (surface water to ground water to deep brines) must account for current and future impacts due to energy systems. Earth-system predictions must capture a variety of scales: local (e.g., single site), basinal (e.g., interaction coupled to basin dynamics), and regional (e.g., watersheds).Hydrologic systems are increasingly affected by large-volume withdrawal associated with energy operations (e.g., power plants, shale oil/gas wells) and human and agricultural consumption, as well as large-volume injection, such as geologic CO_2 storage. ML and data science approaches can be used to augment traditional physics-based models in a variety of related fields, such as applications to groundwater modeling[3], for predicting land subsidence[4], and simulating groundwater dynamics in river basins[5]. The National Academies[6] identified some key missing elements: (a) incorporation of time-lapse high spatial resolution satellite imagery; (b) utilization of data to inform the human relationship with groundwater; (c) a consistent framework to process information, conduct model simulation, and utilize ML; and (d) information systems to handle large datasets and multiple models.

DOI: https://doi.org/10.2172/1769701
DOE BER, AI4ESP 2021 #AI4ESP1018 (https://www.ai4esp.org)

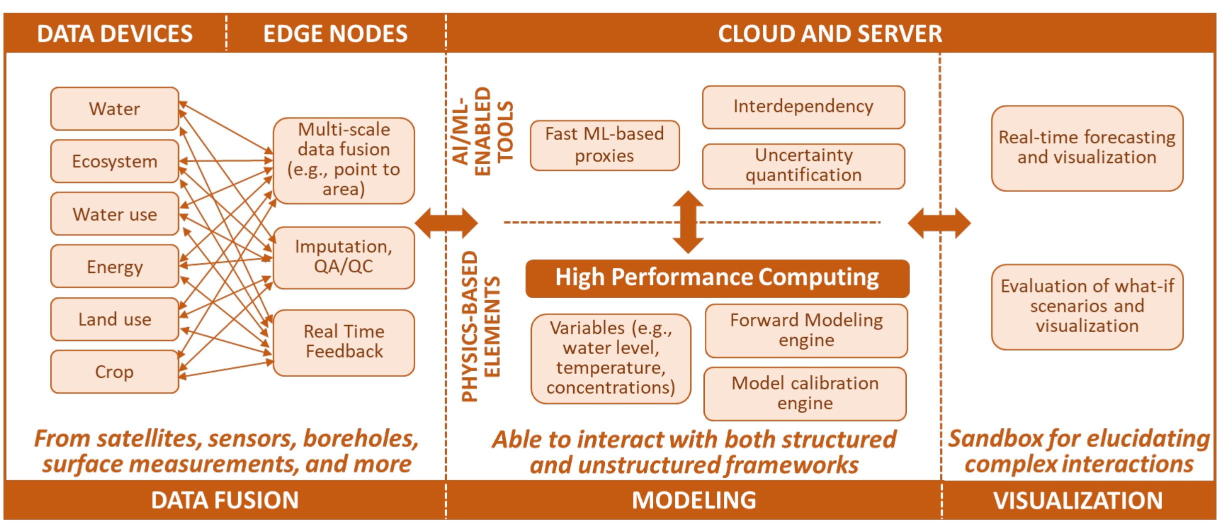
Our desire is to bring significant computational power in the form of advanced physics-based models, augmented by the application of AI/ML and an integrated visualization system, to elucidate how surface and subsurface water systems interact on a variety of scales. These capabilities can help show how human-induced forcings impact the natural water system and predict changes over the next decade and beyond.

6. Narrative:

Understanding the links between water and energy, as related to their availability and demand across spatial and temporal scales, is critical for better representing and predicting the evolution of climate-water-land processes within earth systems models (ESMs). While ESMs typically expect gridded data format, multisector data comes in heterogeneous formats and resolution, exhibiting various imaging geometries, physical meanings, and statistics/semantics. For modeling, the integration of multi-sensor, multisector data under FAIR principles (i.e., findable, accessible, interoperable, and reusable) is a major barrier to developing the next-generation ESM. For application, a major gap is around the discrepancy between current ESM output resolution (10-100 km) local management scales (~10 m).

This group is in a unique position to address many of the data imputation, scaling, and homogenization (DISH) challenges arising from multi-sector dynamics (MSD) modeling, which has recently been identified as a high research priority[7]. Specifically, we will leverage recently advanced ML-assisted automated workflows and federated learning approaches, as well as the collective expertise in water/energy modeling, to transform and ingest big water and energy use data from the public and private sectors. The results will improve the current capability of the DOE E^{3}SM[8] to capture infrastructure evolution (e.g. transmission lines, gas pipelines, water management systems) under climate changes, allowing better representation of the co-evolving human and natural systems within the E^{3}SM.

The group will develop and populate a framework, building from advances in DOE's SMART Initiative[9] regarding integration of highly variable data sources for use with combined physics-based models and AI/ML algorithms to help transform understanding of how surface-subsurface water systems interact by enabling rapid exploration of "what-if" scenarios and calibrated forecasts for local, basinal, and regional water-land systems.



As schematically depicted above, our proposed approach involves the following elements:

Data Fusion: Pertinent data will be obtained from a variety of sources (e.g., satellite-based InSAR, water level gauges, subsurface pressure and temperature sensors), including soft data that informs on human activity related to hydrologic impacts. We will leverage DOE-generated local, basinal, regional, and global measurements from numerous sources/time scales, as well as partner with other governmental and public organizations, such as the USGS's water resources center, and the Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI)[10], a non-profit organization that works closely with the NOAA National Water Center[11] to strengthen the nation's water forecast capabilities for floods and droughts. The group has strong existing relationships with both organizations.

NETL has an innovative data platform, the Energy Data eXchange (EDX), recently named DOE's first geospatial data hub, to manage and share data, and which can be made to link with BER's ESS-DIVE repository, and serve as the gateway and platform to manage data for the group. This data platform can be used to facilitate data integration for groundwater and surface water systems. For example, it can incorporate the power of edge nodes used in processing data from a wide range of sensing devices, enabling real-time feedback and federated learning.

Modeling: Existing numerical models (e.g., E^3SM, PFLOTRAN, NWM) among the proposal partners will be used to help develop fast ML-based proxy models for the physics-based modeling platforms such as coupled Hydrologic-Thermal-Chemical-Biological (HTCB) models developed by DOE. ML algorithms from the SMART Initiative will be modified to capture the output of 3-D models that are correlated in space and time. The adoption of physics-informed ML approaches (e.g., through the use of novel loss functions) that impose physics-based constraints on the output of purely data-driven modeling strategies will be key to addressing situations where the data are voluminous but still sparse in space and/or time. By developing near real-time mechanisms for evaluating system behavior, as manifested through observations of the state variables, a calibrated, rapid forecasting model can be created. The outcome would be a modeling framework that would enable the rapid evaluation of "what-if" scenarios to transform how we interact with the natural system data to improve our understanding of the groundwater-surface water systems.

Visualization: To fully enable the transformative natural of AI/ML to predicting natural system behavior, rapid visualization capabilities must be enabled. The SMART Visualization Prize Challenge is producing a novel experiential visualization platform that can be adapted to the water-land-energy systems being addressed here. The visualization tools that have been developed for the NWM for a variety of scales can also be utilized. Resulting dashboards and other visualization tools can display the results of modeling activities in an intuitive and interactive manner. These will enable scientists to study data and system behavior as though playing in a "data sandbox" that can help transform the types of questions that they can ask and answer to probe the system. Such capabilities can also help communicate important information to decision makers, policy analysts and other stakeholders.

7. Suggested Partners/Experts (Optional)

Grant Bromhal, Srikanta Mishra, or George Guthrie from this white paper team can present on the SMART Initiative.

Kelly Rose, NETL, can speak on the Energy Data Exchange and related data acquisition, assimilation, and management activities at NETL.

Binayak Mohanty, Regents Professor, Texas A&M University. An expert on multi-scale hydrologic data acquisition and assimilation for the unsaturated zone (involved in both sensor-based field projects and physics- and data-driven modeling projects).

Jared Bales, Executive Director, CUASHI. A water resources and hydrological modeling expert, he develops and implements the overall strategy for CUAHSI, maintains relationships with its partners, and creates new partnerships.

David Rounce, Professor, Carnegie Mellon University. A water resources and hydrological modeling expert, his work involves computational modeling and use of large data sets to model the surface water hydrology of glaciers and related systems.

8. References (Optional)

1. Shen, C. et al. (2018), "Incubating deep-learning-powered hydrologic science advances as a community," *Hydrol. Earth Syst. Sci.*, 22:5639-5656.
2. Department of Energy (2019). "Ai for Science", in "Report on the Department of Energy Town Halls on Artificial Intelligence (AI) for Science," Argonne National Laboratory, Lemont, IL, available: https://www.anl.gov/ai-for-science-report.
3. Sahoo, S., TA Russo, J. Elliott and I. Foster (2017). "Machine learning algorithms for modeling groundwater level changes in agricultural regions of the US," *Water Resources Research*, https://doi.org/10.1002/2016WR019933.
4. Smith, RG and S. Majumdar (2020). "Groundwater storage loss association with land subsidence in Western United States mapped using machine learning," *Water Resources Research*, https://doi.org/10.1029/2019WR026621.
5. Chen, C., W. He, H. Zhou et al. (2020). "A comparative study among machine learning and numerical models for simulating groundwater dynamics in the Heihe River Basin, northwestern China," *Nature Science Reports*, https://doi.org/10.1038/s41598-020-60698-9
6. National Academies Press (2019). Groundwater Recharge and Flow: Approaches and Challenges for Monitoring and Modeling using Remotely Sensed Data, Chapter 4, Mitigating Groundwater Model Uncertainties, accessed at: https://www.nap.edu/read/25615/chapter/5.
7. Cronin, J., Anandarajah, G., & Dessens, O. (2018). Climate change impacts on the energy system: a review of trends and gaps. *Climatic change*, *151*(2), 79-93.
8. Energy Exascale Earth System model (E3SM) web site: https://e3sm.org/

9. Science-informed Machine Learning for Accelerating Real Time Subsurface Decisions (SMART) Initiative web site: https://edx.netl.doe.gov/smart/

10. Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI) web site: https://www.cuahsi.org/

11. NOAA National Weather Center (NWC) web site: https://water.noaa.gov/about/nwc