**NSF Convergence Accelerator Track D: Hidden water and hydrologic extremes: a groundwater data platform for machine learning and water management**

Groundwateris by far the largest unfrozen freshwater resource in the world. We know it is critically important to global water security and food production, especially during extreme events. Yet, this buried resource remains an unseen portion of the hydrologic cycle both literally and figuratively. Groundwater observations are sparse and grossly insufficient for forecasting and decision making. Furthermore, what observation data does exist, are heterogenous and difficult to access, often requiring significant interpretation. Physically-based groundwater models have greatly advanced in recent years and can begin to address observational gaps at the national level. However, the computational demands are currently too high for most applications outside of academia. As a result, the national and global tools that we use for decision making and planning rely on simplified groundwater representations, creating systematic biases in hydrologic forecasts and limiting our ability to plan for and manage extreme events.

This is a long standing problem that has proven a difficult challenge within the hydrologic community. Meaningful solutions will require interdisciplinary teams that can address data, modeling and human design challenges and that place the end user at the center of the process. The proposed groundwater convergence accelerator will do just that: bring together experts in groundwater, machine learning, data science and software engineering with water managers and forecasters. We will take a user-centered design approach to (1) make groundwater data and simulations FAIR (Findable, Accessible, Interoperable and Reusable), and (2) develop a user portal for building machine-learning groundwater models that can be incorporated into the existing workflows of applied users. Our resulting framework, *HydroFrame-ML* will consist of a national groundwater data platform and a user portal for building, evaluating and sharing groundwater models. While the data access and modelingplatform will be our most tangible deliverables, our ultimate goal is not to build tools *per se*, but rather to build community platforms that are useful and are used. Our overall success will be measured by the users of this platform and the improvements to their forecasting and decision making.

**Intellectual Merit:** For this project, we are bringing together the most physically rigorous national scale groundwater simulations developed through *HydroFrame* with national leaders in Earth Systems Modeling (GFDL) and water management (Reclamation). By providing end-to-end workflows combining state of groundwater science with operational management tools, *HydroFrame-ML* will advance both large-scale water management as well as our understanding of how human operations and groundwater interact in extreme events. Our products will provide innovative ways to improve forecasts and in the process will expand our knowledge about the (1) contributions of groundwater to extreme events in managed systems; (2) biases in our current risk-assessment approaches which do not consider groundwater; and (3) potential to improve long-term sustainability by more actively managing groundwater and accounting for groundwater surface water interactions in our projections.

**Broader Impacts:** HydroFrame-ML is motivated by, and structured around, applied solutions for water management planning and decision making. Extreme events like drought and floods have far-reaching societal impacts. They are common, costly and likely to get worse in the future. We have partnered with the Bureau of Reclamation, which is the largest wholesale water provider in the country, providing water to more than 31 million people and 10 million acres of farmland. They will be driving the design of our use cases and the metrics we use to evaluate our success in Phase 1, as well as helping us expand our team for Phase 2. Advances we make will translate directly to better water management outcomes that directly benefit society. Additionally, we believe that the clear societal relevance of water can help engage a diverse undergraduate population and broaden participation. We will develop hands-on activities and challenges designed to give undergraduates experience in machine learning and data science, in the context of pressing real-world challenges. We have a dedicated education coordinator for Phase 1 and plan to partner with successful STEM mentorship programs in phase 2 to build content that can help broaden participation of underrepresented students well beyond the timeframe of this project.