Title: Framework for an adaptive integrated observation system using a hierarchy of machine learning approaches

Authors/Affiliations:
Jennifer Comstock, Joseph C. Hardin, and Zhe Feng – Pacific Northwest National Laboratory

Focal Area(s):

1. Data acquisition enabled by machine learning, AI, and advanced methods including experimental/network design/optimization, and hardware-related efforts involving AI
2. Insight gleaned from complex measurements using AI, big data analytics, and other advanced methods, including explainable AI and physics- or knowledge-guided AI

Science Challenge and Rationale:

Atmospheric processes are stochastic, occur at scales from the micrometer to many kilometers, and are constantly changing over time. Characterizing these interactions and associated environmental conditions using traditional measurement techniques is difficult and can take years to build statistics on atmospheric phenomena that occurs episodically. Developing new innovative approaches to modify sampling strategies in real-time would enable the routine collection of targeted measurements focused on a specific set of science questions.

Artificial intelligence provides an avenue to develop new algorithms and integrated instrument systems that work together to process, analyze, and interpret observations, make decisions based on current and changing conditions, and modify operational strategies on the fly. Atmospheric processes occur on a range of spatiotemporal scales and require measurements that provide spatial variability of surface measurements and profiles of atmospheric state, aerosol, and cloud properties. Scanning instruments, such as scanning cloud and precipitation radar and Doppler lidar, coupled with distributed instrument networks provide additional spatial context. Development of an integrated, adaptive observational system would transform atmospheric measurement strategies that learn from historical observations and apply that knowledge to adapt in real-time as the atmosphere evolves.

An example application of this framework is understanding the initiation and upscale growth of deep convection across geographic regimes, which can trigger extreme flooding events. Understanding and improving the predictability of water cycle extremes addresses a key DOE EESSD Grand Challenge. Processes controlling the initiation and intensification of deep convection is in part determined by the complex interactions between the land and atmosphere, and changes in environmental conditions over time. Understanding the linkages between these processes requires knowledge across various spatiotemporal scales.

Narrative:

Current observational systems that use fixed measurement strategies operate independently and provide snapshots of the evolution of atmospheric phenomena, such as the transition from shallow to deep convection. An adaptive system would analyze measurements from multiple

DOI: https://doi.org/10.2172/1769712
DOE BER, AI4ESP 2021 #AI4ESP1028 (https://www.ai4esp.org)

spatially distributed instruments and numerical weather prediction (NWP) forecasts and modify its measurement strategy based on the evolution of the atmospheric system as it passes through the observational spatial domain, informed by prior knowledge gained through historical observational analysis. An integrated framework for developing an adaptive observational system requires a hierarchy of machine learning applications each designed to cooperate and learn from observations. This multi-step process may include the following examples:

- **Representation Learning:** Apply ML classification techniques (e. g. unsupervised and semi-supervised learning) to observational data and model reanalysis to classify atmospheric conditions, such as cloud objects, environmental conditions, and meteorological regimes. There are a variety of data sources, such as DOE ARM, operational weather radar, land surface conditions, satellite, and surface-based networks (NASA AERONET, MPLNET, mesonets, radiation networks, etc.). Datasets would vary depending on the scientific focus (e. g. convection clouds, aerosol-cloud interactions, etc.). Labeled field campaign data can also be used to further train the ML model or provide evaluation data. By processing existing historical observations, we are able to characterize phenomenon that may not be fully represented in atmospheric models. Results will help constrain the parameter space for a specific science objective using prior archived observations.
- **Learning Linkages:** Apply ML systems (such as deep neural networks, or ensemble tree methods) within the identified parameter space to discover relationships between the evolution of measurable atmospheric variables and the life cycle of atmospheric systems. An example is how do certain approaching weather systems coupled with local surface and boundary layer conditions affect initiation of flood producing extreme deep convection at an ARM site. These relationships can inform scientists on new interpretations and reasoning of physical processes across scales. Use this knowledge to inform the basis of an adaptive instrument system.
- **Adapt to changing environment:** By combining the representations of phenomenon and their linkages to environment and evolution, build an integrated instrument system that adapts measurements to evolving atmospheric conditions using a learning multi-agent architecture and focus measurements on specific phenomena of interest. For example, radiosonde measurements and changes in atmospheric state measured by surface-based mesonets may indicate that convection is likely in a certain geographic region. Reinforcement Learning would help map linkages from the observations to changes in the instrument operations, such as changing sampling intervals or scanning strategies in real-time to focus instruments on atmospheric phenomena in a specific geographic domain. This science informed engineering approach will enable engineers to build intelligent AI systems that can identify atmospheric targets of interest and adapt and track their evolution. As edge computing brings online larger numbers of low-cost sensors, this multi-agent approach has the potential to scale to heterogeneous instrument networks. Real-time observations can be fed back into the system to improve performance and help the system improve over time.

While this system would benefit from and utilize NWP models to supplement incomplete observations of atmospheric environments such as large-scale flow regimes, a key strength is the ability to learn the best representations of systems for adaptive operations from data, even in the case where models may have biased representations of the phenomenon. Similarly, by combining historical observations (e.g. weather radar) and reanalysis (e.g. high-resolution numerical weather prediction models) with AI approaches, we can predict where particular cloud fields will occur and adjust our real-time measurement strategy accordingly.

<u>Outcomes</u>:
This white paper addresses EESSD Scientific Grand Challenges related to the Integrated Water Cycle and Data-Model Integration. This strategy would benefit data assimilation in high-resolution models; improve knowledge of physical processes by exploring predictive understanding using ML for focused science objectives; develop new model diagnostics and benchmarks based on improved statistics; and enhance execution of intensive observing periods or longer-term focused observation periods. Codes, workflows, and tools would be made available through open source repositories such as Git, and we expect that all data used is publicly available through government data portals.

<u>Suggested Partners/Experts (Optional):</u>
Skill sets needed for topics described in this white paper are machine learning expertise with domain knowledge, subject matter experts for specific scientific objectives including observational techniques and analysis (i.e., deep convection processes), and instrument hardware/software and data system expertise for instrument, sensor, and computational integration.

This white paper is synergistic with the paper "The Usage of Observing System Simulation Experiments (OSSE) and Reinforcement Learning to Optimize Experimental Design and Operation" led by Dr. Joseph Hardin also submitted to the AI4ESP call. Observational data mining tasks described here would help identify model biases, whereas the model and OSSE based approach focuses on strategic decisions and setting policy, which could inform the ML framework we have described.