Multisensor Agile Adaptive Sampling of Convective Storms Driven by Real-time Analytics

Pavlos Kollias[1,2], Edward Luke[1], Katia Lamer[1], Fan Yang[1], Mariko Oue[2], Nathan Urban[3] and Andrew Vogelmann[1]

1. Department of Environmental and Climate Sciences, Brookhaven National Laboratory
2. Division of Atmospheric Sciences, Stony Brook University
3. Computational Science Initiative, Brookhaven National Laboratory

Focal Area

Intelligent data acquisition in convective storms ("…*when and where matters*...") enabled by machine learning, AI using a dynamically optimized, data-driven observing network, and edge computing.

Science Challenge

Convective storms vertically transport water vapor and condensate from Earth's surface to the upper troposphere. Life on Earth is fundamentally linked to this transport which determines the hydrological cycle, and the intensity of severe weather responsible for the destruction of life and property. Despite advances in high-resolution modeling and better observational capabilities, the scientific community continues to be confronted with knowledge gaps about convective storms that limit our predictive capabilities. The ongoing developments in the high-resolution Energy Exascale Earth System Model (E3SM), large eddy simulations, and AI-based analytics to evaluate uncertainties are expected to provide a comprehensive framework for new scientific discovery. The model-experiment (MODEX) approach suggests that the aforementioned advancements in model development and AI-based inference techniques should be complemented by similar advancements in the experimental (observational) side so that the former does not outstrip the ability of the latter to provide meaningful constraints. *What are the recent advancements in observations that will provide the necessary leap forward in improving our predictive capabilities?* To address this question, we propose a new experimental paradigm called Multisensor Agile Adaptive Sampling (MAAS) that capitalizes on advancements in communications (5G), computational resources (edge/fog computing), sensor capabilities, and machine learning (ML) and AI techniques (Kollias et al., 2020). The MAAS framework allows for the collection of higher spatiotemporal resolution and quality observations of convective storms than is traditionally possible. The MAAS framework is scalable and applicable to atmospheric observatories such as those operated by the Department of Energy (DoE) Atmospheric Radiation Measurement (ARM) facility.

Rationale

The vast majority of our observing systems (i.e., DOE ARM observatories) interact with the atmosphere using a predetermined "stare" or "sit and spin" sampling strategy that is not adaptive to the atmospheric conditions. This approach is based on sampling all the sky equally, even clear skies without any atmospheric phenomena of interest. *Thus, our ability is precluded from revisiting more frequently in time and with higher spatial resolution the parts of the sky where a particular phenomenon exists*. In addition to the lack of an intelligent, automated observational framework, another limitation of the current observational paradigm is the current way in which multi-sensor observations are used. It is common to rely on multiple sensors to observe these parts of the atmosphere and gain insights regarding their interactions using multi-sensor retrieval techniques. Traditionally, the value of multi-sensor observations emerges long after their collection during a post-processing phase. *Unfortunately, any knowledge gained at that stage cannot be used to adjust and optimize the observing strategy*, often leaving an incomplete picture of the atmosphere.

DOI: https://doi.org/10.2172/1769753
DOE BER, AI4ESP 2021 #AI4ESP1070 (https://www.ai4esp.org)

A convective storm is a rapidly evolving system with strong spatial inhomogeneity. The aforementioned experimental bottleneck obstructs our knowledge advancement in convective processes that are ultimately associated with Earth's hydrological cycle and extreme weather phenomena. Our vision and efforts so far have focused on the development of the first-ever artificial intelligence (AI) interface for surface-based observatories (MAAS, Kollias et al., 2020). The AI system receives input from a complex ecosystem of cameras, satellite feeds, lidars, radars etc., to identify, target and sample convective storms. The rapid revisiting of these atmospheric features offers time-resolved observations (**delta-t**, the process-resolved measurements) accompanied by agile and adaptive sampling that substantially improves the spatial resolution and quality of the observations. Our intelligent observation can revisit in time more frequently and with higher spatial resolution. Such observations can for the first-time provide new insights into rapidly evolving atmospheric phenomena associated with convective storms.

Narrative

Recently, we demonstrated the value of MAAS at the SBU/BNL testbed located on Long Island, NY by adapting the sampling strategy of a phased-array radar and a polarimetric scanning cloud radar —two different yet uniquely complementary systems— using real-time observations from a geostationary satellite, a surface camera, and the radars themselves. The ARM facility currently supports the integration of the MAAS framework to provide adaptive sampling guidance for their scanning precipitation radar during the Tracking Aerosol Convection interactions ExpeRiment (TRACER) experiment in the Houston area for the 2021-2022 period. The future deployment of the third ARM Mobile Facility in the Southeastern United States (SEUS) is another candidate for implementation of the MAAS framework and the collection of a rich dataset of delta-t measurements in convective storms. In future expansions, MAAS could rely on observations from other operational sensors such as i) the NEXRAD which provides information about long range targets, ii) networks of surface cameras, and iii) lidars sensitive to cloud base height and clear air dynamics. The SBU/BNL testbed and the ARM TRACER and SEUS deployments will be our main source of observations and testbeds for implementing different versions of the MAAS framework and will provide invaluable data for improving and refining the framework. One important asset of the MAAS framework is the near-zero latency collection of data from space-based and surface-based remote sensors using high-speed communications and high-performance edge computing. These diverse datasets are quality controlled and co-registered to provide a multi-sensor view of the atmospheric state that is used to identify features of interest, nowcast their future location, and convert their trajectory to instrument sampling guidance. In the closed-loop process, ML and AI play a critical role. The following specific areas are where ML/AI techniques are necessary.

***Multispectral Satellite Feature Identification*:** The current generation of Geostationary Operational Environmental Satellite (GOES-16/17) constitutes a significant technological leap forward from its predecessor in spatial, temporal, and spectral resolution (consisting of 16 wavelength bands). This exceptional increase in resolution is opening new windows into atmospheric processes of interest. It is reasonable to expect that the signatures of most of these atmospheric processes are optimally described by not one, but by some non-linear, potentially time-evolving, combination of these bands. A thunderstorm is a good example of a process that constantly evolves over the course of its lifecycle. Many begin their existence as a simple cumulus cloud which is easy to pinpoint in location. But predicting in advance where such a cumulus cloud will form and, of all those that form, which are most likely to develop into thunderstorms are both important questions to which machine learning can be brought to bear. As a thunderstorm develops and assumes an increasingly complex structure, pinpointing the location of its constituent components (for example its active core) becomes more difficult from a spaceborne perspective, as ice lofted to high altitudes spreads out over a large area, obscuring details below. However, there are still signals observable related to the core's position. The core tends to overshoot the ice, producing a temporally dynamic area of variable cloud reflectance. In addition, lightning is strongly correlated with the core and detectable by yet another sensor on the payload of the GOES. A machine learning system can synthesize these observational components into a more holistic view, creating a best-estimate position for real-time tracking by surface-based sensors, and developing a better understanding of the fundamental physics. Such a system would require a number of AI components. Data-driven spatiotemporal forecasting, possibly integrated with real-time forecast models or their ML surrogates, is needed to predict the formation of thunderstorms and their cumulus precursors. Feature identification, tracking, and filtering, in tandem with dynamical forecasting and multi-modal data fusion, are needed to infer the locations of storm constituents. Reinforcement learning and optimization under uncertainty will be needed to prioritize sensor targeting and learn closed-loop sensor control policies that update in real-time as data comes in.

<u>Real-time Model-Observation Integration:</u> A paradigm used in the atmospheric sciences consists of a feedback loop where observations (e.g., radar-centric variables) are converted into the physical quantities used by models to drive their improvement. Likewise, the output from these models is often converted through instrument simulators back into observable parameters for comparison with the "truth". Although the model-observation integration has been used in current nowcasting or variational simulation techniques, predicting sub-grid scale clouds (storms) is still challenging due to several limitations both in observation (limited observation coverage, limited observables, uncertainties in the physical quantity retrievals) and model (spatiotemporal resolutions). Edge computing and ML techniques will complement these limitations and allow for better prediction even of sub-grid atmospheric phenomena beyond just the model-observation integration. AI may be employed to develop fast ML surrogates within which approximate data assimilation and parameter estimation may be performed in real-time, in combination with Bayesian uncertainty quantification and objective-driven decision theory applied to attention-based models to select atmospheric conditions of relevance to reducing model predictive uncertainties. Data assimilation may be accomplished entirely within the ML surrogates using transfer learning concepts, pre-training to model dynamics, and updating the surrogate with real-time data.

<u>Supervised Learning by Expert Crowd Sourcing:</u> An agile remote sensing system that includes a user interface with sufficient sophistication to allow members of the broader scientific community to interact with and influence it in real time opens the door to generating a dataset of highly relevant labels that can then be used to supervise the training of a machine learning system. For example, if a team of scientists with access to such an interface are studying the conditions most conducive to triggering a thunderstorm and the storm's subsequent lifecycle, their collective interaction with it will generate a knowledgebase of aspects of the atmospheric state most relevant to the problem. This knowledgebase can then train a system to add value to a field campaign as it progresses, or for subsequent campaigns. If a researcher points a cursor at a blob in a radar image and clicks their mouse to direct a sensor toward it, information has been gained that something special may distinguish that blob as particularly relevant, according to the domain knowledge of the expert. This draws upon a long history of human-in-loop label annotation for ML training data.

<u>Reference:</u> Kollias, P., Luke, E., Oue, M., & Lamer, K. (2020). Agile adaptive radar sampling of fast-evolving atmospheric phenomena guided by satellite imagery and surface cameras. *Geophysical Research Letters*, 45, e2020GL088440. https://doi.org/10.1029/2020GL088440