Integrative data-driven approaches for characterization & prediction of aerosol-cloud processes

Authors: Lyndsay Shand[1], Ben Hillman[1], Lekha Patel[1], Gabriel Huerta[1], J. Derek Tucker[1], Andrea Staid[1], Don Lyons[1], Erin Schliep[2]
[1] Sandia National Laboratories, [2] University of Missouri

Focal Area: Advanced methods to glean insights from complex data (Focal Area 3).
Science Challenge: Fusing and interpreting the vast amount of data from disjoint sources for the purpose of elevating our understanding of aerosol-cloud interaction presents an enormous challenge and opportunity and is necessary to improve uncertainty quantification and predictions, especially for extreme events.

Rationale:

Aerosols perturb the atmosphere and climate both directly and indirectly. Indirectly, they cause changes in cloud properties, including cloud formation, cloud lifetime, and cloud radiative properties (Twomey, 1974 and Albrecht, 1989). Directly, aerosols can impact the frequency and intensity of water cycle-related extreme events, such as drought and flooding (Reed et al., 2019; Sillmann et al., 2019). Research linking observed aerosol-cloud-precipitation processes with numerical modeling has experienced much improvement in the treatment of ice nucleation, warm rain processes, secondary-organic aerosol formation, and convection (Fan et al., 2018; Zhao et al 2019, Mülmenstädt et al. 2020, Quaas et al. 2020). However, our knowledge on how aerosols mix with clouds, and thus on how aerosols impact the water-cycle, is still limited. **Large uncertainties in these processes remain due to both difficulties in measuring a mixed, non-idealized atmosphere, due to asynchronous model-observation linkages, and missing or incomplete parameterizations linking aerosol-cloud interactions.** This knowledge gap has led to large uncertainties in climate predictions, e.g. 20-year extreme precipitation predictions of the tropics and subtropics (Kharin et al., 2013). **"How are cloud processes affected by changes in anthropogenic aerosol production and how do these fluctuations translate to extreme precipitation events?" are nontrivial and transformational science questions that could greatly benefit from improved understanding and modeling of complex aerosol-cloud-precipitation interactions. Such improvements would necessarily reduce uncertainty in these predictions, leading to fewer model discrepancies, more confident predictions of extreme events (e.g. floods and droughts) and more informed decision making on climate policy.**

Aerosol effects on cloud processes are dynamic, multi-faceted, and interdependent, resulting in some of the largest uncertainties in climate models (IPCC, 2013). Expected effects of aerosol emissions on cloud systems include changes in cloud condensation nuclei, liquid water content, precipitation and albedo, to name a few (Durkee et al., 2018 and Possner et al., 2018). High frequency and high-resolution data collection and monitoring further challenges the study of these complex relationships. These rich data, which include in situ, ground-based, and space-based measurements, as well as data from multi-resolution numerical simulations, need to be integrated to improve our understanding of aerosol processes within clouds. While a wealth of high-resolution measurements has been collected with the objective of monitoring aerosol and cloud processes, especially by BER Atmospheric Radiation Measurement (ARM) facilities, NASA and NOAA field campaigns, and satellite observations, these data have been historically underutilized to improve our knowledge of the earth system. For example, in the weather forecasting community, only 3-5% of the vast amount of satellite data are used in numerical forecasts due in large part to the complexity of the science. This complexity presents significant "challenges to properly interpret and exploit the most interesting and potentially most valuable satellite data" (Boukabara, 2019). Recent and continuing advances in satellite observational capabilities have

DOI: https://doi.org/10.2172/1769729
DOE BER, AI4ESP 2021 #AI4ESP1120 (https://www.ai4esp.org)

yielded higher resolution data with more global coverage (e.g., *500* meters every 5-15 minutes from the geostationary GOES-17 satellite and *250* meters from near-polar orbiting MODIS). Linking these data sets with ground-based atmospheric radiation measurements such as those provided by the DOE's ARM program is of specific relevance to cloud-aerosol processes. Fusing the vast amount of data from disjoint sources for the purpose of elevating our understanding of aerosol-cloud interactions across space and time presents both an enormous challenge and great opportunity for improving climate predictions and uncertainty quantification, especially for extreme precipitation events.

Narrative: <u>Leveraging Artificial Intelligence (AI)</u>

We wish to leverage recent advances in computing and analytical techniques to develop tools that can fully fuse and interpret data of varying sources, resolutions, and fidelities. The aforementioned abundance of data provides an excellent opportunity for researchers, but much work needs to be done to assimilate data into digestible forms for statistical and machine learning (ML) algorithms (Maskey et al., 2020). Cutting-edge analytical methods, including both statistical and ML techniques, have been recently developed to study complex environmental processes but have yet to be applied in practice (e.g. McDermott and Wikle, 2020 and references therein). Such methods are capable of spanning the continuum from observation to predictive modeling and are ideal for fusing and interpreting convoluted data collected globally, and thus for filling knowledge gaps related to aerosol-cloud-precipitation processes. ML techniques can integrate new knowledge with existing modeling efforts to reduce uncertainties and radically improve the predictive capabilities of earth system models (ESMs). However, noteworthy methodological and computational challenges remain and limit significant improvements in the prediction and uncertainty quantification for ESMs. These challenges include fusing large datasets and sources with complex dependencies, and accounting for spatial and temporal misalignment and varying measurement error.

Statistical and ML methods are generally unconstrained by physical understanding (e.g., Reichstein et al (2019) and Maskey et al (2020)), however, these methods can be extended to include physical constraints and properly account for the uncertainty of using such information (e.g., McDermott and Wikle, 2020 and Berliner et al., 2000). The risky and paradigm-shifting work that we believe should be done with AI and aerosol-cloud interactions is as follows. We need to set aside the overly constrained parameterizations of aerosol properties, such as size and composition to enable the integration of other aerosol-related retreivals, such as aerosol optical depth and all phases of water. This will require building statistical models that can appropriately characterize the complex dynamics and a short-term forecasting capability using machine learning techniques for a non-precipitating cloud parcel in a select region before then extending the techniques to a larger domain. This builds upon observations and numerical simulations of Amiri-Farahani et al. (2017), Durkee et al. (2018), Possner et al. (2018), Zheng et al. (2020), and Patel and Shand (2020).

<u>Statistics for Interpretability and Uncertainty Quantification</u>

We propose developing tailored statistical methods that can account for the dynamic evolution of aerosols over space and time **by handling** key simultaneous effects of atmospheric processes, multiple complementary data sources, and multiple sources of uncertainty, especially measurement error. Advanced statistical approaches have proven to make significant progress to address all of these complexities. For example, Shi and Cressie (2007) estimate global and regional trends of aerosol- and weather-related measurements and multivariate space-time statistical approaches have been proposed for characterizing the dynamic relationship across space and time between multiple spatial processes (e.g., Gelfand and Banerjee, 2010 and Datta et al., 2016). Hooten and Wikle (2012) proposed an agent-based model in a hierarchical Bayesian framework to intuitively parameterize complex systems of dynamic space-time processes. Schliep et al., (2015) proposed a model to successfully fuse aerosol data from ground-based and remote sensing sources and overcome challenges due to misaligned data sources, correlation across space and time, and extensive missingness of ground-based monitoring data. Such approaches are promising but have yet to be developed for aerosols and clouds. Patel and Shand (2020), for example, demonstrated a statistical model for the behavior of aerosols emitted from ships in low-lying clouds in a marine environment and account for key atmospheric drivers and dynamics through stochastic differential equations.

Deep Learning & Machine Learning for Improved Prediction

We propose adapting and developing ML approaches to build surrogates of complex parameterizations of earth system processes when large amounts of data are available. These methods enable the integration of insights gleaned from statistical methods with ESMs. Existing methods that we can leverage are the following: Weber et al. (2020) adapts a convolutional neural network as a surrogate to efficiently predict short-term precipitation forecasts, while Hughes et al. (2018) and Zheng et al. (2020) demonstrate the use of gradient-boosting algorithms to represent and predict aerosol mixing-states. Representations of small-scale atmospheric processes via General Adversarial Networks (GANs), have been demonstrated by Berner et al., 2017 and Gagne et al., 2020. Historically, modelers have relied upon heuristics to identify extreme events which cannot adequately capture the complexity of the system and are known to be unreliable (Barbuzano, 2020). In recent weather modeling advances, Chattopadhyay et al. (2020) adapts a capsule neural network to accurately predict extreme weather events. Deep Learning methods have been shown to be particular useful for building surrogate models of ESMs and have the potential for efficient prediction of extreme climate events. Qi and Majda (2020) leverage deep convolutional neural networks to predict extreme events in turbulent dynamical systems. McDermott and Wikle (2020) rely on reservoir computing to employ cutting-edge deep echo state networks along with a hierarchical Bayesian framework to advance the deterministic multi-scale Lorenz-96 model (Lorenz, 1996). This work shows promise of reliable and efficient long-lead forecasting in complex non-linear systems. Although these approaches are promising, significant advances are still needed to even begin to represent the climate dynamics, provide uncertainty quantification and accurately predict extreme events and within ESMs.

In conclusion, we propose developing advanced analytical methods tailored to the challenges discussed in this paper to transform our understanding of complex aerosol-cloud-precipitation processes. These methods, along with substantial cross-disciplinary collaborations, will enable the discovery and integration of newfound insights with existing climate models such as DOE's E3SM. This work will not only improve the interpretability and predictability of ESMs but also enhance the science and ensure reusable and reproducible findings. Developing broadly applicable tools with statistics and ML will ultimately enable AI that will substantially improve DOE's confidence in predicting extreme events such as sudden drought or flooding and set DOE up for success to contribute paradigm-changing improvements to earth system predictability. This work follows the Earth and Environmental Systems Sciences Division Strategic Plan to "develop an improved capability for Earth system prediction on seasonal to multidecadal time scales to inform the development of resilient U.S. energy strategies" through the Data-Model integration Scientific Grand Challenge to "support the integration and management of models, experiments, and observations across a hierarchy of scales and complexity" (DOE CESD Strategic Plan 2018). We look forward to the collaboration with colleagues interested in these methodologies to create new solutions for the predictability of the Earth System.

Suggested Partners/Experts:

Chris Wikle (WikleC@Missouri.edu), Professor of Statistics and Department Chair @ University of Missouri, Expert in statistical methods for complex processes

Dan Cooley (cooleys@stat.colostate.edu), Professor of Statistics @ Colorado State University– extreme value theory applied to atmospheric and environmental sciences, NCAR collaborator

Doug Nychka (douglasnychka@gmail.com), Professor of Applied Math Statistics @ Colorado School of Mines, Emeritus Senior Scientist @ NCAR– spatial statistics for large climate data

Dorit Hammerling (hammerling@mins.edu), Professor of Applied Math and Statistics @ Colorado School of Mines – Lead the Statistics and Data Science Section in the Institute for Mathematics Applied to Geosciences at the National Center for Atmospheric Research (NCAR)

Murali Haran (mharan@stat.psu.edu), Professor and Department Chair @ Penn State University, spatial statistics for climate science applications, complex computer models, statistical emulation and calibration

References

Amiri-Farahani, A., R. J. Allen, D. Neubauer, and U. Lohmann. (2017). Impact of Saharan dust on North Atlantic marine stratocumulus clouds: importance of the semidirect effect, *Atmos. Chem. Phys*., 17, 6305–6322, https://doi.org/10.5194/acp-17-6305-2017.

Barbuzano, J. (2020). Teaching machines to detect climate extremes, *Eos, 101,* https://doi.org/10.1029/2020EO145702.

Berliner, L. M., Wikle, C. K., and Cressie, N. (2000). Long-Lead Prediction of Pacific SSTs via Bayesian Dynamic Modeling. *Journal of Climate* 13, 22, 3953-3968.

Berner, J., U. Achatz, L. Batte, A. De La Cámara, H. Christensen, M. Colangeli, and Y. I. Yano. (2017). Stochastic parameterization: Towards a new view of weather and climate models. Bulletin of the *American Meteorological Society*, 98, 565–588.

Boukabara, Sid-Ahmed (2019). AI for Earth observation and numerical weather prediction, *SpaceNews.* April 2019. Available at: https://spacenews.com/ai-for-earth-observation-and-numerical-weather-prediction/. (Accessed: January 2021)

Chattopadhyay, A., Nabizadeh, E., and Hassanzadeh, P. (2020). Analog forecasting of extreme-causing weather patterns using deep learning**.** *Journal of Advances in Modeling Earth Systems*.

Datta, A., Banerjee, S., Finley, A. O., & Gelfand, A. E. (2016). Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets. *Journal of the American Statistical Association*, 111(514), 800–812. https://doi.org/10.1080/01621459.2015.1044091

DOE EESSD Strategic Plan. (2018). Available at: https://tes.science.energy.gov/strategicplan/index.shtml.

Durkee, P.A., K. J. Noone, and R. T. Bluth. (2000). The Monterey area ship track experiment. *Journal of the Atmospheric Sciences*, 57:2523-2541.

Fan, J., Rosenfeld, D., Zhang, Y. et al. (2018), "Substantial convection and precipitation enhancements by ultrafine aerosol particles" *Scienc*e. 26 Jan 2018: Vol. 359, Issue 6374, pp. 411-418.

Gelfand AE and Banerjee S. (2010). Multivariate Spatial Process Models. In: Gelfand AE, Diggle PJ, Fuentes M, Guttorp P, editors. *Handbook of Spatial Statistics*. Boca Raton, FL: Chapman & Hall/CRC. pp. 495–516.

Hooten, M. B. and C. K. Wikle. (2010). Statistical Agent-Based Models for Discrete Spatio-Temporal Systems, *Journal of the American Statistical Association*, 105:489, 236-248.

Hughes, M., Kodros, J. K., Pierce, J. R., West, M., and Riemer, N. (2018). Machine Learning to Predict the Global Distribution of Aerosol Mixing State Metrics. Atmosphere, 9(1), 15.

IPCC, 2013: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Kharin, V.V., F. W. Zwiers, X. Zhang, et al. (2013). Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change* **119,** 345–357. https://doi.org/10.1007/s10584-013-0705-8.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences,* 20(2), 130–141.

Maskey, M., H. Alemohammad, K. J. Murphy, and R. Ramachandran (2020), Advancing AI for Earth science: A data systems perspective, *Eos*, 101, https://doi.org/10.1029/2020EO151245. Published on 06 November 2020.

Mülmenstädt, J., C. Nam, M. Salzmann, et al. (2020). Reducing the aerosol forcing uncertainty using observational constraints on warm rain processes. *Science Advances* 29, Vol. 6, no. 22.

McDermott, P. L. and Wikle, C. K. (2018) Deep echo state networks with uncertainty quantification for spatio-temporal forecasting, *Environmetrics*, 30.

Patel, L., and L. Shand (2020). Simulating cloud-aerosol interactions made by ship emissions. In JSM Proceedings, Section on Statistics and the Environment. Alexandria, VA: American Statistical Association. 2282-2295.

Possner, A., H. H. Wang, R. Wood, K. Caldeira, and T. P. Ackerman (2018). The efficacy of aerosol–cloud radiative perturbations from near-surface emissions in deep open-cell stratocumuli. *Atmospheric Chemistry and Physics* 18(23), 17475–17488.

Qi, D. and Majda, J. (2020). Using Machine Learning to predict extreme events in complex systems. *Proceedings of the National Academy of Sciences* Jan 2020, 117 (1) 52-59.

Quaas, J., A. Arola, B. Cairns,et al. (2020). Constraining the Twomey effect from satellite observations: issues and perspectives. *Atmospheric Chemistry and Physics.*, 20, 15079–15099.

Reed, K. A., J. T. Bacmeister, J. J. A.Huff, X. Wu, S. C. Bates, and N. A. Rosenbloom. (2019). Exploring the impact of dust on North Atlantic hurricanes in a high-resolution climate model. *Geophysical Research Letters*, 46, 1105– 1112.

Reichstein, M., Camps-Valls, G., Stevens, B. et al. (2019) Deep learning and process understanding for data-driven Earth system science. *Nature* 566, 195–204.

Schliep, E. M., Gelfand, A. E., and Holland, D. M. (2015). Autoregressive spatially varying coefficients model for predicting daily $PM_{2.5}$ using VIIRS satellite AOT, *Adv. Stat. Clim. Meteorol. Oceanogr.*, 1, 59-74. https://doi.org/10.5194/ascmo-1-59-2015, 2015.

Sillmann, J., C. W. Stjern, G. Myhre, *et al.* (2019). Extreme wet and dry conditions affected differently by greenhouse gases and aerosols. *npj Clim Atmos Sci* 2, 24.

U.S. DOE. (2018). *Climate and Environmental Sciences Division Strategic Plan 2018–2023*, DOE/SC–0192, U.S. Department of Energy Office of Science. Accessed at: science.osti.gov/-/media/ber/pdf/workshop-reports/2018_CESD_Strategic_Plan.pdf.

Zhao, B., Wang, Y., Gu, Y. *et al.* (2019). Ice nucleation by aerosols from anthropogenic pollution. *Nat. Geosci.* **12,** 602–607.

Zheng, Z., Curtis, J. H., Yao, Y., Gasparik, J. T., Anantharaj, V. G., Zhao, L., West, M., & Riemer, N. (2020). Estimating Submicron Aerosol Mixing State at the Global Scale with Machine Learning and Earth System Modeling. *Earth and Space Science*. https://doi.org/10.1029/2020EA001500

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND2021-1584 A