EGU2020-9820
https://doi.org/10.5194/egusphere-egu2020-9820
EGU General Assembly 2020

Machine learning for detection of climate extremes: New approaches to uncertainty quantification

William Collins[1,2], Travis O'Brien[3], Mr Prabhat[4], and Karthik Kashinath[4]
[1]Climate & Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, United States of America (wdcollins@lbl.gov)
[2]Department of Earth and Planetary Science, University of California, Berkeley, California, United States of America (wdcollins@berkeley.edu)
[3]Department of Earth and Atmospheric Sciences, Indiana University, Bloomington, Indiana, United States of America (obrienta@iu.edu)
[4]National Energy Research Scientific Computing Center, Lawrence Berkeley National Laboratory, Berkeley, California, United States of America (prabhat@lbl.gov and kkashinath@lbl.gov)

Machine learning (ML) has proven to be a very powerful body of techniques for identifying rare but highly impactful weather events in huge volumes of climate model output and satellite data. When these events and the changes in them are studied in the context of global warming, these phenomena are known as climate extremes. This talk concerns the challenges in applying ML to identify climate extremes, which often center on how to provide suitable training data to these algorithms. The challenges are:

- In many cases, the official definitions for the weather events in the current climate are either ad hoc and/or subjective, leading to considerable variance in the statistics of these events even in literature concerning the historical record;
- Operational methods for identifying these events are also typically quite ad hoc with very limited quantification of their structural and parametric uncertainties; and
- Both the generative mechanisms and physical properties of these events are both predicted to evolve due to well-understood physics, and hence the training data set should but typically does not reflect these secular trends in the formation and statistical properties of climate extremes.

We describe several approaches to addressing these issues, including:

- The recent creation of the first labeled data set specifically designed for algorithm training on atmospheric extremes, known as ClimateNet;
- Probabilistic ML algorithms that identify events based on the level of agreement across an ensemble of operational methods;

- Bayesian methods for that identify events based on the level of agreement across an ensemble of human expert-generated labels; and
- The prospects for physics-based detection using fundamental properties of the fluid dynamics (i.e., conserved variables and Lyapunov exponents) and/or information-theoretic concepts.