EGU24-11708, updated on 23 Jul 2024
https://doi.org/10.5194/egusphere-egu24-11708
EGU General Assembly 2024

Medium-Range Excessive Rainfall Prediction with Machine Learning

Aaron Hill[1,3] and Russ Schumacher[2]
[1]University of Oklahoma, School of Meteorology, Norman, United States of America (ahill@ou.edu)
[2]Colorado State University, Department of Atmospheric Science, Fort Collins, United States of America (russ.schumacher@colostate.edu)
[3]NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography, Norman, United States of America

The prediction of excessive rainfall using numerical weather prediction (NWP) models is unequivocally difficult owing to the myriad of complexities that must be resolved (e.g., parent storm dynamics, microphysics) in order to forecast the placement and intensity of rainfall correctly. However, machine learning (ML) has provided a new avenue by which we can generate predictions of excessive rainfall with sufficient lead time to inform decision makers and planners to the threat of inclement weather. ML techniques are able to decode known long-standing relationships between environmental predictors and convective hazards from long historical records, and they have demonstrated tremendous value in predicting weather hazards at longer lead times (e.g., Hill et al. 2023). Further, continued effort by the meteorological community to explain ML models and their forecasts is building trust between developers and end users. As a result, their use in meteorological hazard forecasting is expanding, particularly into the medium range (e.g., 4-8 days) when forecasters are reliant on relatively coarse NWP models to create forecasts.

In this work, we are using Random Forests (RFs) to generate daily probabilistic forecasts of excessive rainfall at 1-8 day lead times. The RFs are trained using output from the Global Ensemble Forecast System and historical observations of excessive rainfall. Environmental parameters like precipitable water and CAPE, as well as modeled precipitation, are spatiotemporally arranged so the RFs can learn spatial and diurnal patterns that associate with excessive rainfall. The RF models are evaluated against a spatio-temporally varying climatology and show skill out to 7 days, and routinely outperform human-based forecasts past a 1-day lead time. In this presentation, we will highlight performance characteristics of the RFs into the medium-range (e.g., out to 8 days) and discuss the implications of excessive rainfall definitions in RF model training. Additionally, we will present an ensemble prediction framework that provides estimates of uncertainty and ranges of forecast solutions that operational forecasters desire at extended lead times.