Probabilistic machine learning for predicting atmospheric convection initiation

Greta Miller, Hannah Christensen, and Philip Stier
Atmospheric, Oceanic and Planetary Physics, University of Oxford
greta.miller@physics.ox.ac.uk

UNIVERSITY OF OXFORD

ENVIRONMENTAL RESEARCH DOCTORAL TRAINING PARTNERSHIP

Introduction

- Realistically representing atmospheric convection is important for accurate numerical weather and climate simulations.
- Parameterizing where and when convection occurs (convection trigger) is a well-known source of model uncertainty [1,2].
- While some global storm-resolving models omit the convection parameterization [3], it is unknown how well they simulate the initiation of convection without a parameterization scheme.

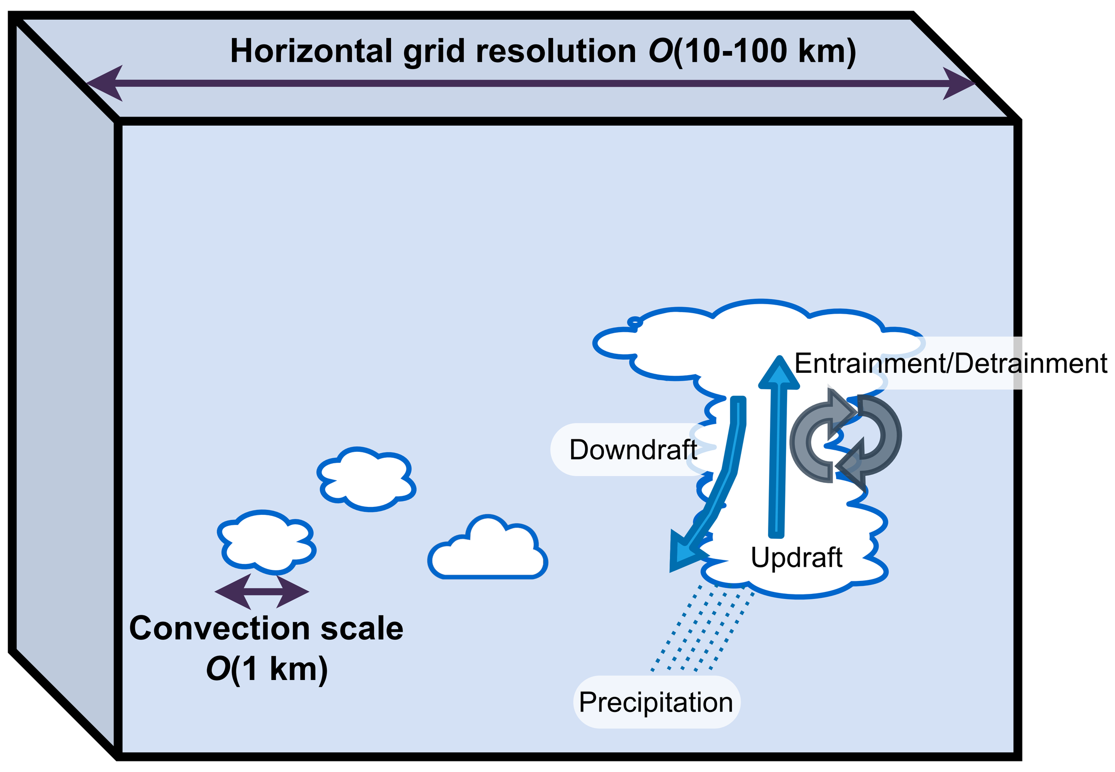

Fig 1. Convection parameterization schemes represent an ensemble of cumulus clouds with subgrid-scale processes that collectively modify the large-scale atmospheric state

How can we improve the representation of convection in climate and weather models?

→ Develop a probabilistic machine learning (ML) model that predicts the probability of the onset of deep convection

- Understand sources of uncertainty and the mechanisms driving the initiation of convection
- Characterize errors and assess the realism of convection triggering, especially in storm-resolving models

Methods

Data

- Observed large-scale atmospheric variables from Atmospheric Radiation Measurement (ARM) constrained variational analysis [4]
- Site: Southern Great Plains, USA, summer (June – August)
- Convective event: precipitation rate ≥ 0.5 mm/hr

Machine learning model: Random forest

- Randomly under-sample non-convective events to rebalance dataset
- Bayesian hyperparameter optimization to tune model settings
- Probability calibration (sigmoid method)

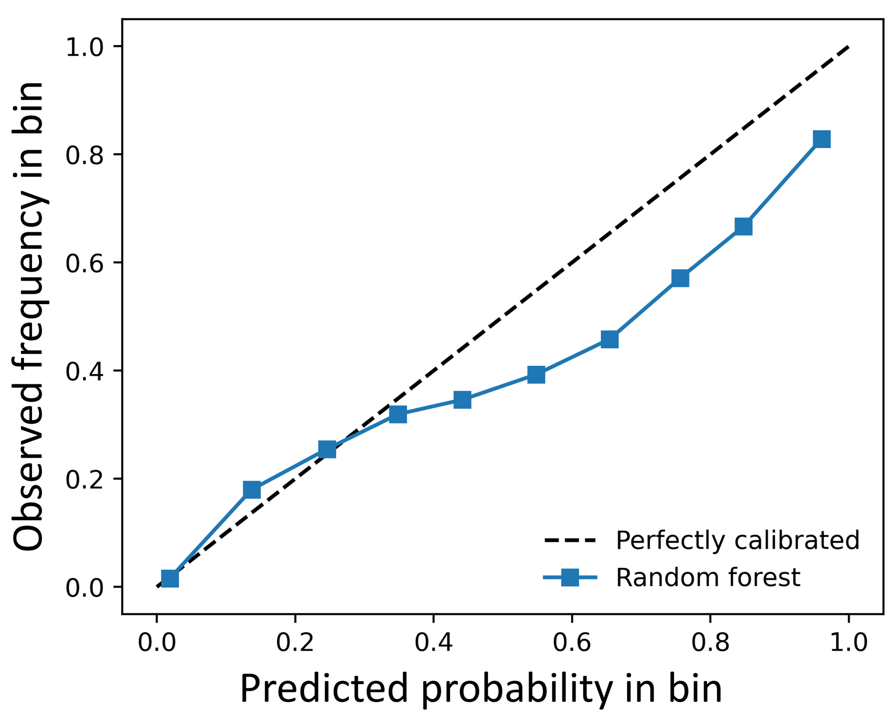

Fig 2. Reliability curve for random forest after probability calibration

Results

The generation of convective available potential energy from large-scale advection (dCAPE) is the most important predictor of convection, followed by specific humidity and the lifting condensation level (LCL).

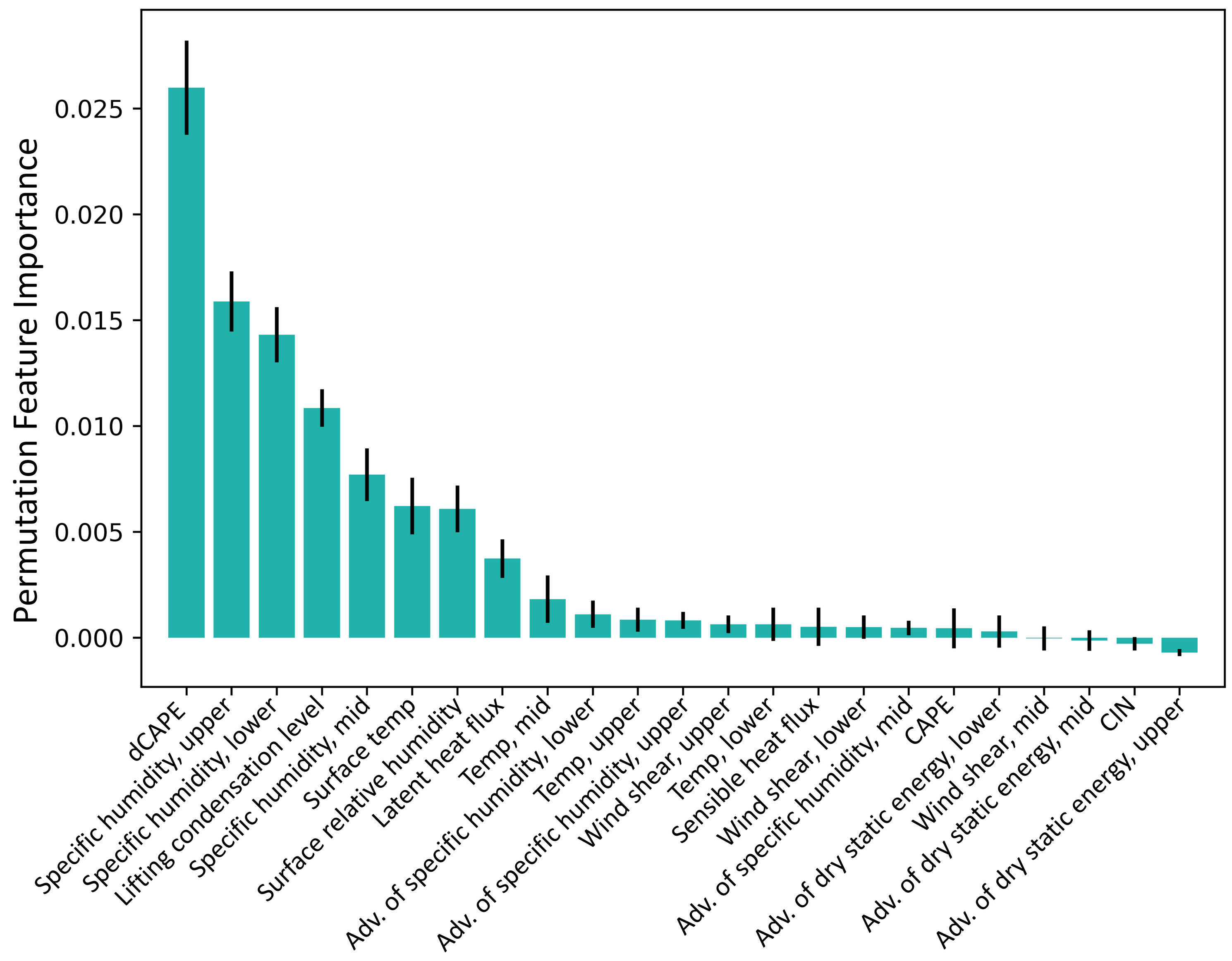

Fig 3. Feature importance ranking shows which inputs to the random forest are the most important predictors of convection for the Southern Great Plains site

The ML model captures the diurnal cycle and time series of precipitating convection reasonably well.

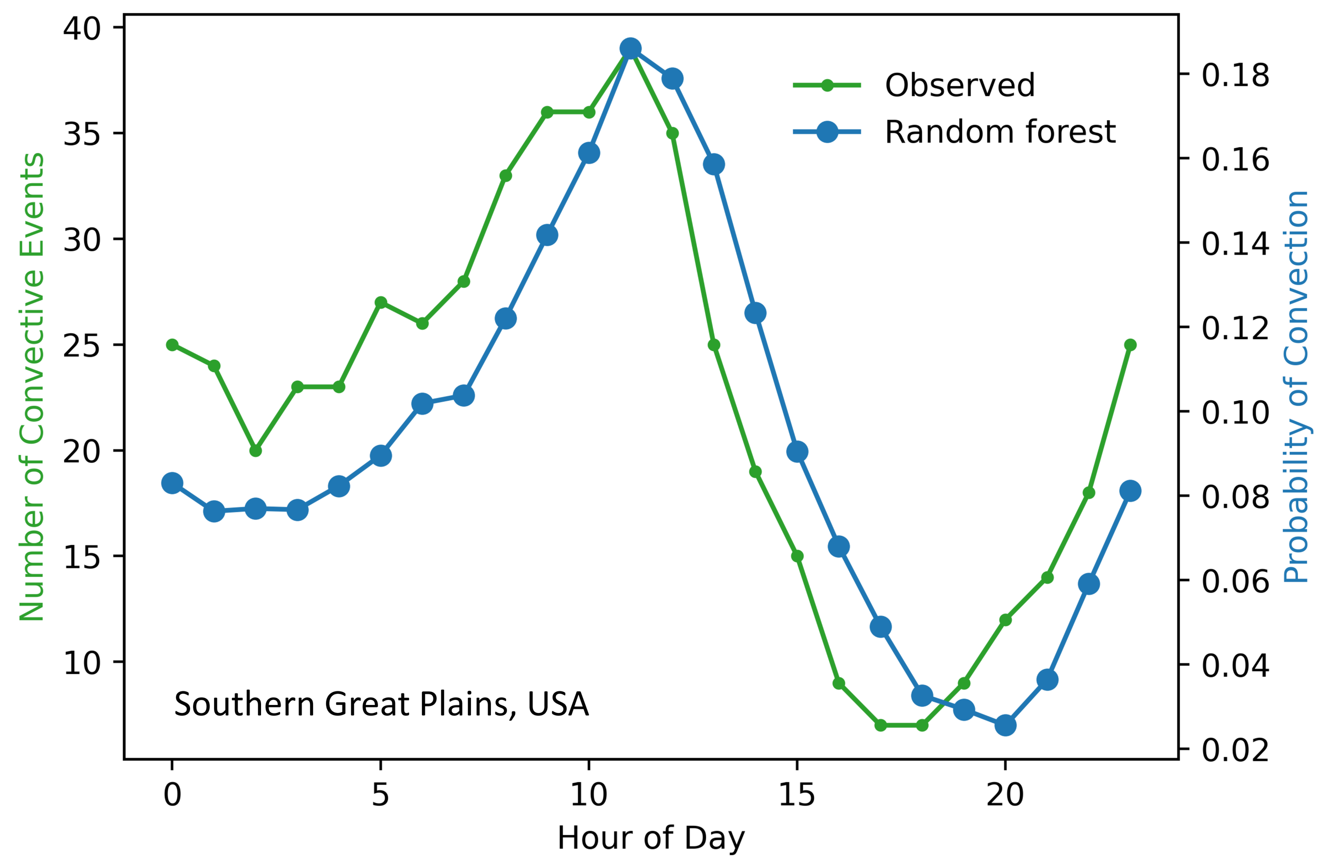

Fig 4. Diurnal cycle of convection for observations (green) vs model (blue)

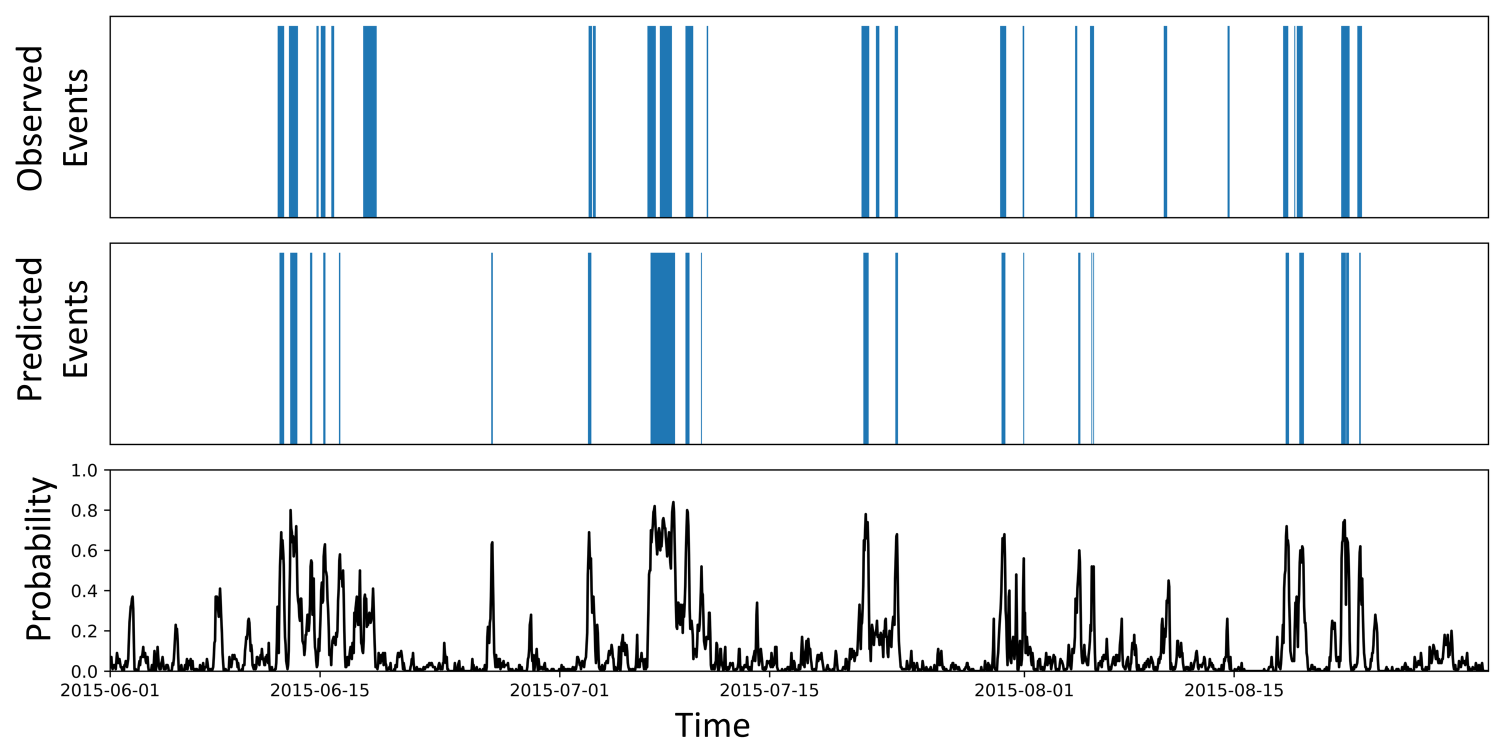

Fig 5. Time series of observations of convection vs model prediction and model probability for June – August 2015

Future Work

- Further model development
- Use global precipitation and reanalysis datasets to expand the ML model globally (not site-specific)
- Evaluate storm-resolving simulations of convection triggering using the probabilistic ML model
- Implement the probabilistic ML model as a stochastic parameterization for the convection trigger in a standard weather model

References

[1] Zhang, T. et al. (2021). *Journal of Advances in Modeling Earth Systems*, 13(5), 1–19.
[2] Ukkonen, P., & Mäkelä, A. (2019). *Journal of Advances in Modeling Earth Systems*, 11(6), 1784–1802
[3] Stevens, B. et al. (2019). *Progress in Earth and Planetary Science*, 6(1).
[4] Tao, C., & Xie, S. Constrained Variational Analysis (60VARANARAP). Atmospheric Radiation Measurement (ARM) User Facility.