Extending Predictions of Hazardous Weather Into the Medium-Range with Machine Learning

Aaron J. Hill, Russ S. Schumacher, and Jacob Escobedo
Dept. of Atmospheric Science, Colorado State University

Overview

- Excessive rainfall, tornadoes, large hail, and damaging winds occur on spatial and temporal scales that are not well represented in numerical weather prediction (NWP) model output
- Predictability limit for these hazards is short, so reliable probabilistic forecasts are needed rather than deterministic predictions that will inevitably have large errors
- Over the past several years, we have developed a suite of probabilistic forecast systems, referred to as Colorado State University-Machine Learning Probabilities (CSU-MLP; Schumacher et al. 2021) system
- The CSU-MLP uses the Global Ensemble Forecast System (GEFS) Reforecast dataset, historical observations of hazardous weather, and machine learning algorithms (Random Forests; RFs) to generate skillful, reliable guidance
- Operational forecasters can use these products as a "first guess" fields when generating outlooks
- **To extend the value of this prediction system, we have started building new ML models capable of probabilistically predicting hazards in the medium range (i.e., days 3 -- 8)**

Guiding Knowledge

- ML models developed for predictions of hazardous weather are typically trained using predictors in close spatio-temporal proximity to the weather event (e.g., Hill et al. 2020)
- The atmospheric evolution of prior days and local changes to the environment can often highlight an increased threat of severe weather
- We are exploring new predictor assembly methods that incorporate this knowledge of predictability and flow-dependencies into model training
- Objectively selecting meteorological predictors that signal increased threats for hazardous weather unique to different regions of the country

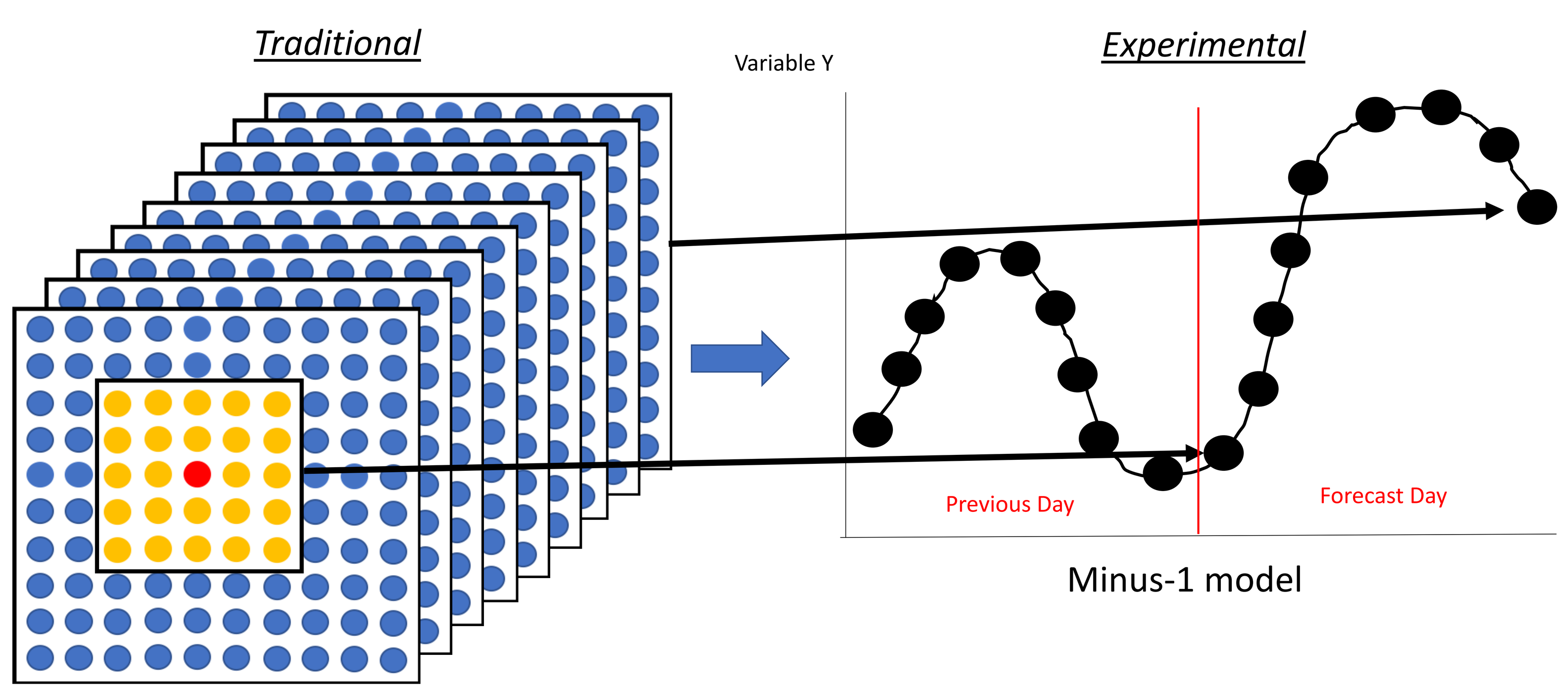


Figure 1: Schematic of predictor assembly for a meteorological variable in our (left) traditional method and (right) experimental method, in which spatiotemporal features are spatially averaged to create time series for training over the forecast period and one day prior (i.e., minus-1 model).

Experimental Setup

- Nine years of hindcasts from the GEFS/Reforecast dataset (2003-2012)
- Historical observations from Storm Data (i.e., filtered and quality controlled severe weather reports) defined over 24-hour periods
- Observations are paired to GEFS/R grid points on a 40-km grid across the CONUS; training is split up across three geographical regions
- Thermodynamic and kinematic features selected around training points (traditional method); those same features are spatially averaged in flow-dependent models to create time series over multiple days

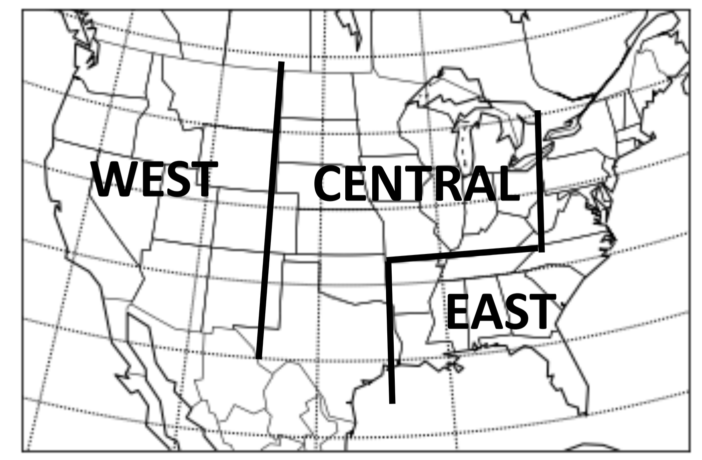


Figure 2: Geographical regions defining RF models trained and used to forecast across the CONUS.

Forecast Example

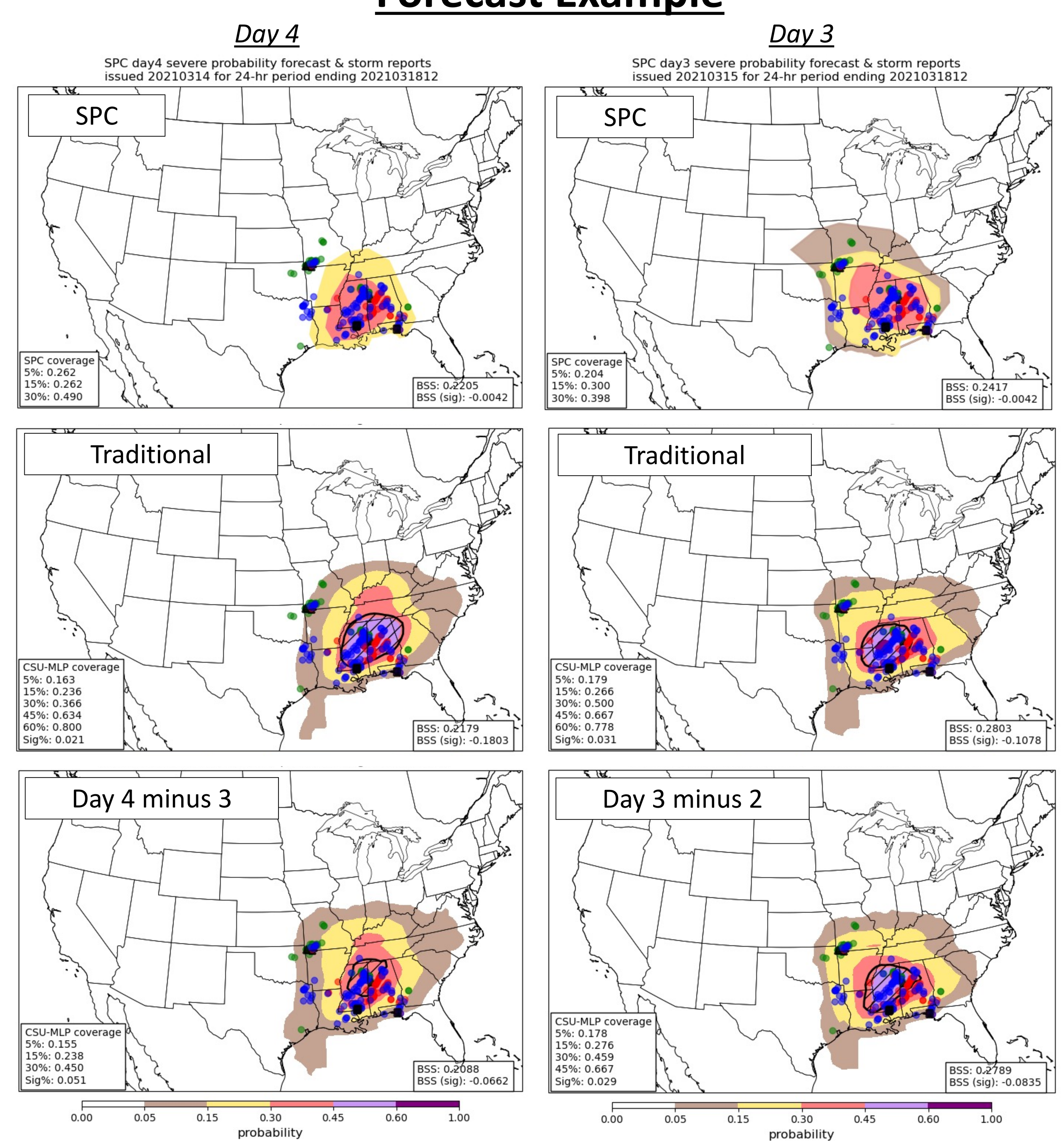


Figure 3: Example (left) day-4 and (right) day-3 probabilistic forecasts of any severe weather report (colored contours) for the period 1200 UTC 17 March 2021 – 1200 UTC 18 March 2021. (top) Outlooks from the Storm Prediction Center (SPC), (middle) traditional model forecasts, and (bottom) experimental, uncalibrated forecasts for day 4 minus-3 (i.e., includes three prior days) and day 3 minus-2 (i.e., includes two prior days) models. Skill scores are depicted in the bottom right and colored dots represent severe weather reports from the SPC. Spatial coverage of reports in respective contours labeled in bottom left.

Knowledge Gained

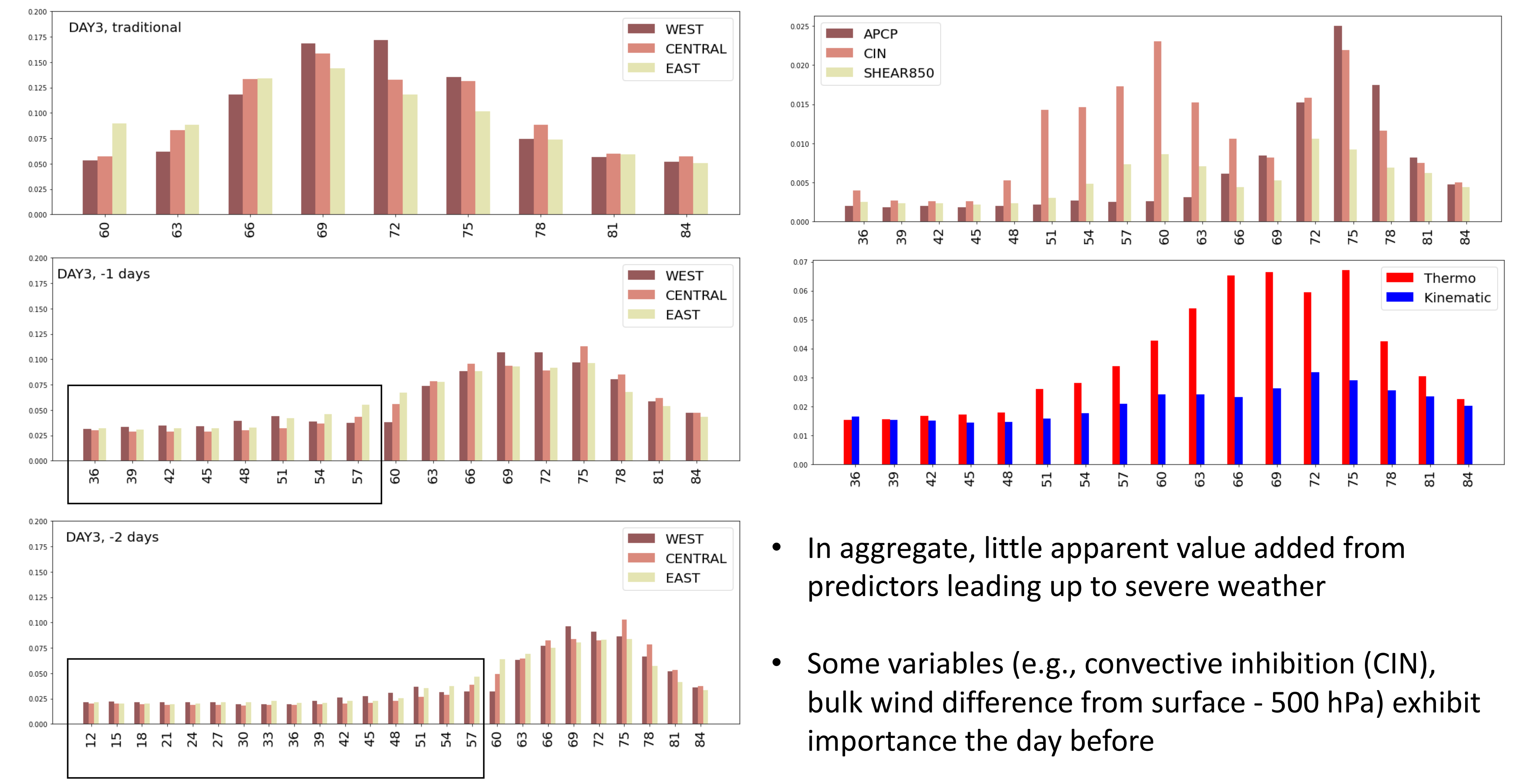
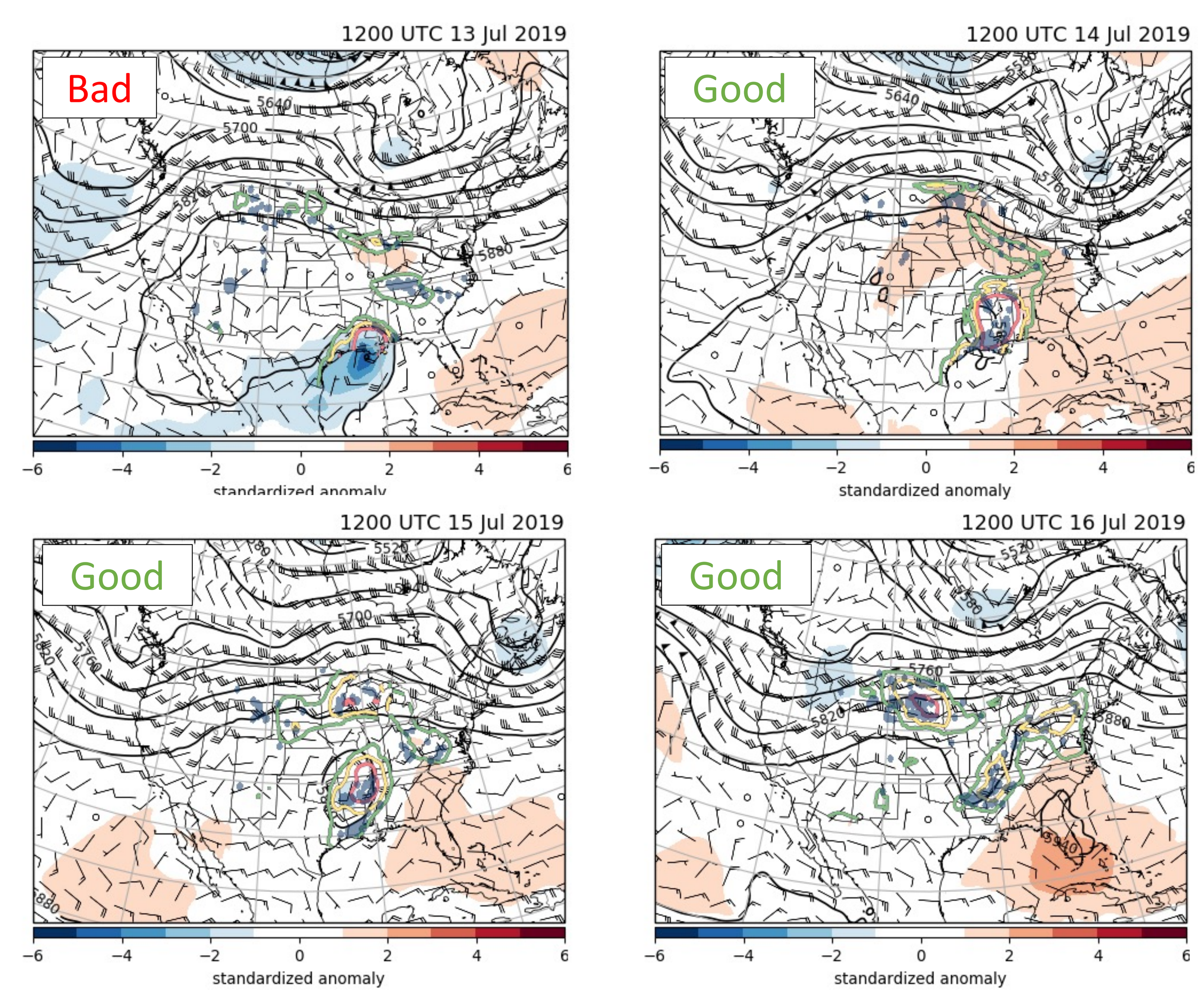


Figure 4: Gini feature importances for day-3 (top left) traditional, (middle left) minus-1, and (middle bottom) minus-2 RF models aggregated across all variables by forecast hour. (top right) Same as in other panels, but for accumulated precipitation (APCP), convective inhibition (CIN), and 0 – 850 hPa wind shear (SHEAR850) in the East region. (bottom right) Same as in other panels, but thermodynamic and kinematic variables are aggregated separately from East region.

- In aggregate, little apparent value added from predictors leading up to severe weather
- Some variables (e.g., convective inhibition (CIN), bulk wind difference from surface - 500 hPa) exhibit importance the day before
- Local changes to thermodynamics "build up" prior to severe weather event more so than kinematics

Moving Forward



- Jacob Escobedo is developing techniques to understand under what flow conditions are our ML predictions (in this example, for excessive rainfall prediction) failing and succeeding, e.g., weakly- or strongly-forced environments, tropical cyclones, winter vs. summer extratropical weather systems?
- E.g., what happened in day-1 forecasts for a tropical cyclone beginning landfall?
- Alternative ML methods that capture non-local features important to the forecast (e.g., CNNs)

Figure 5: 0 – 500 hPa geopotential height (black contours) and height anomaly (filled contours) at 1200 UTC on each day across the period 13 – 16 July 2019 with observations of excessive rainfall (filled dots) and day-1 forecast probabilities (unfilled color contours) from the CSU-MLP suite of excessive rainfall-predicting models. The top left panel represents one of the worst forecasts (BAD) over a year and a half verification period, while the other three GOOD forecasts.

References

Schumacher et al. 2021: From random forests to flood forecasts: A research to operations success story. Bulletin of the American Meteorological Society, in press, doi:10.1175/BAMS-D-20-0186.1

Hill et al. 2020: Forecasting severe weather with random forests. Monthly Weather Review, 148, 2135--2161, doi:10.1175/MWR-D-19-0344.1

Acknowledgements

This work is supported by NOAA Joint Technology Transfer Initiative grants NA18OAR4590378 and NA20OAR4590350.

Contact

aaron.hill@colostate.edu