Understanding Training Data Components for Excessive Rainfall Machine-Learning Models: A look inside the Unified Flooding Verification System

Mitchell R. Green (Central Michigan University), Aaron J. Hill (Colorado State University), Russ S. Schumacher (Colorado State University)

Problem Statement

The challenge of forecasting excessive rainfall arises from needing to incorporate not only the resolution of magnitude and spatial distribution of an event but also the current surface hydrology and terrain.

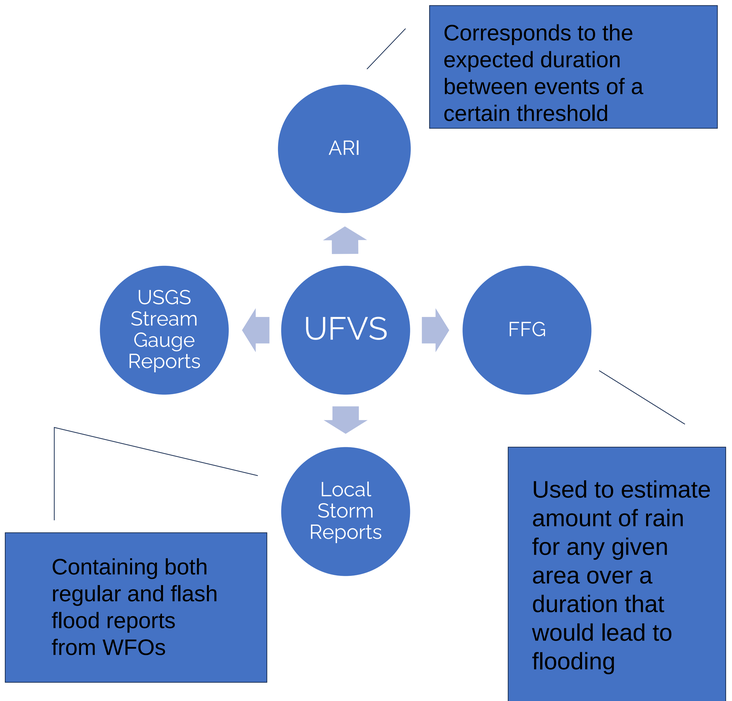
Relying on a single source of verification such as flash flood guidance (FFG) or local flood reports is insufficient as each source are subject to specific errors.

Understanding the internal components of training datasets is important in better understanding how precipitation machine-learning models trained on this dataset output probabilistic forecasts.

This project will analyze the stage IV precipitation analysis of 5-year average recurrence interval (ARI) and FFG exceedances and look for biases between different versions of the CSU-MLP precipitation model.

UFVS

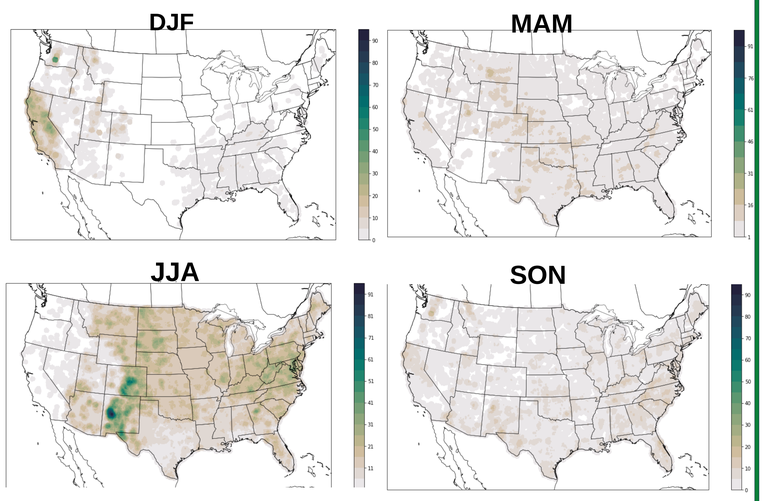
The Unified Flooding Verification System (UFVS) was developed as a comprehensive dataset containing multiple observations and flood proxies to be used for excessive rainfall verification.



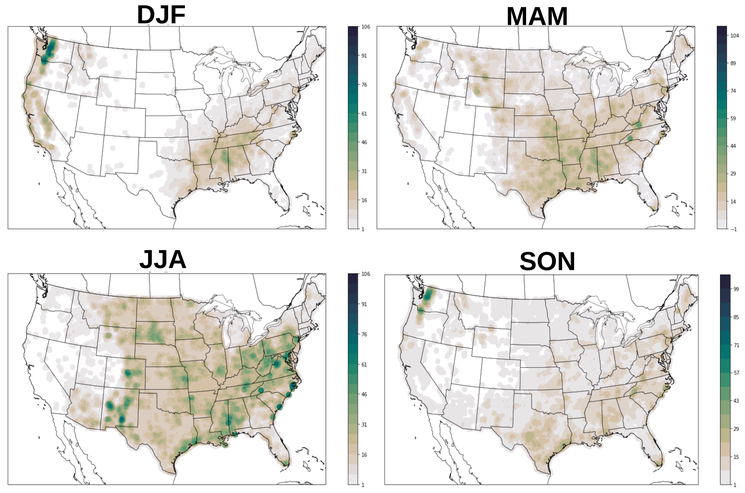
Erickson et al. 2019

Analysis

5-year ARI Exceedances

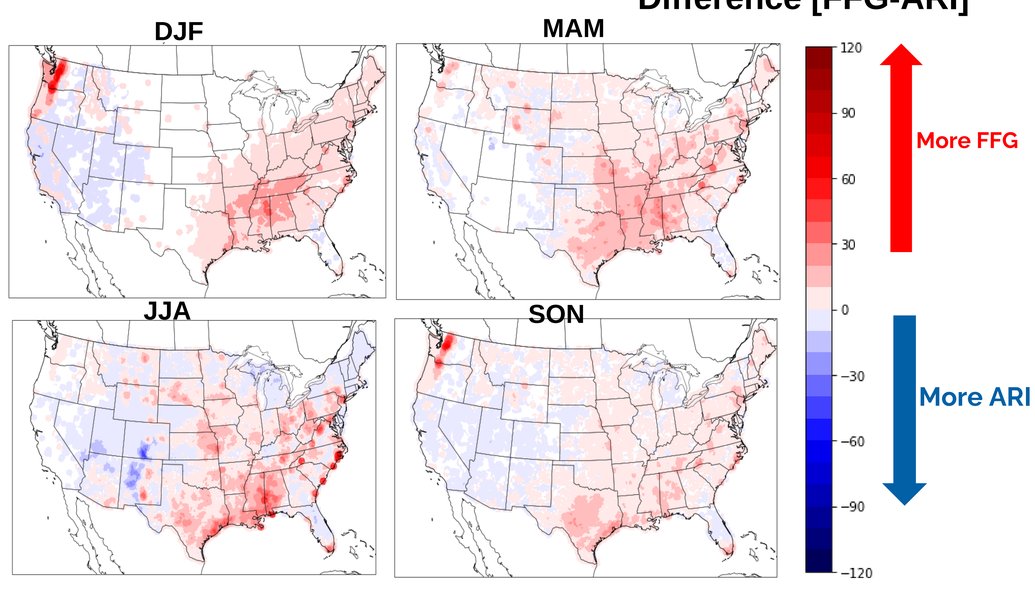


FFG Exceedances



Findings

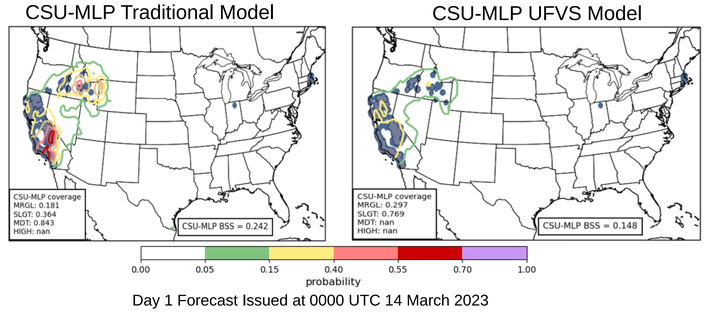
Difference [FFG-ARI]



FFG is exceeded more frequently in the southeastern CONUS than ARI year-round. Alternatively, ARI thresholds are exceeded more frequently in the western half of the CONUS.

This raises questions on how probability forecasts from machine learning versions trained on this dataset are created in these subregions

Event Studies



Day 1 Forecast Issued at 0000 UTC 14 March 2023

A study of the 14 March 2023 atmospheric river event in California highlights the difference in maximum probabilities issued by two versions of the CSU-MLP model

Future Work

Further analysis of other internal components such as USGS reports to determine their influence, if any, on probability forecasts.

Assess accuracy of differences among version outputs,

Investigating the reasoning behind differences with FFG and ARI exceedances that occur in specific geographic regions.

Acknowledgements

I thank my mentors and program directors for their endless support. *This work has been supported by the National Science Foundation Research for Undergraduates Site in Earth System Science at Colorado State University under the agreement No. AGS-1950172.*

Contact: green1mr@cmich.edu