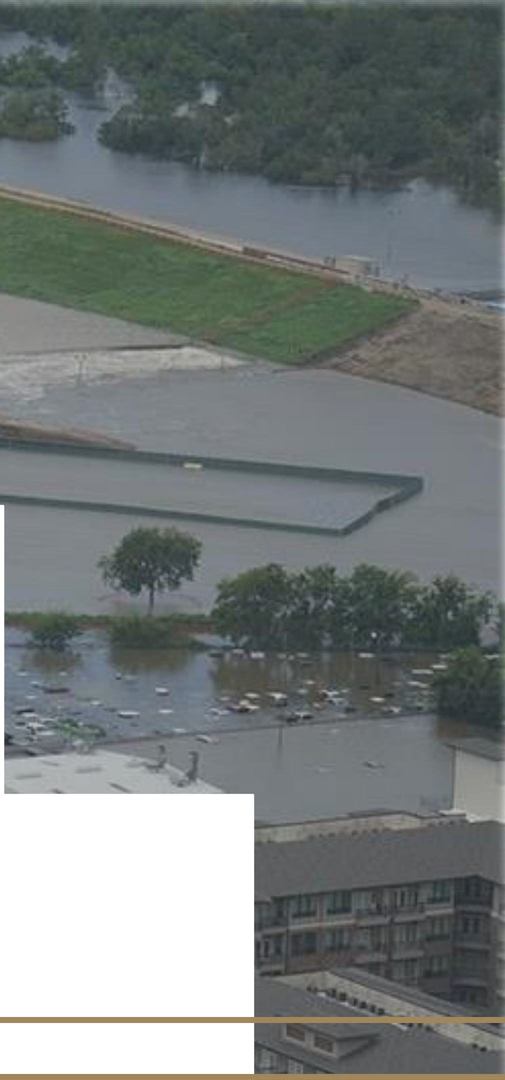
FLORIDA TECH

AGU100 ADVANCING EARTH AND SPACE SCIENCE

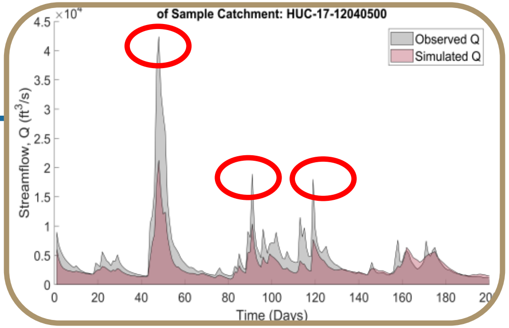
Fall 2020 Meeting
H188: Machine Learning in Hydrologic Forecasting I

Machine Learning for Flood Peak Prediction in Ungauged Basins

Zimeena Rasheed, Akshay Aravamudan, Ali Gorji Sefidmazgi, Georgios C. Anagnostopoulos, Efthymios I. Nikolopoulos



A need to PREDICT PEAK FLOWS



Objectives

Investigate the use of machine-learning-based algorithms to:

a) assess the relative importance of **dynamic and static variables** in flood response

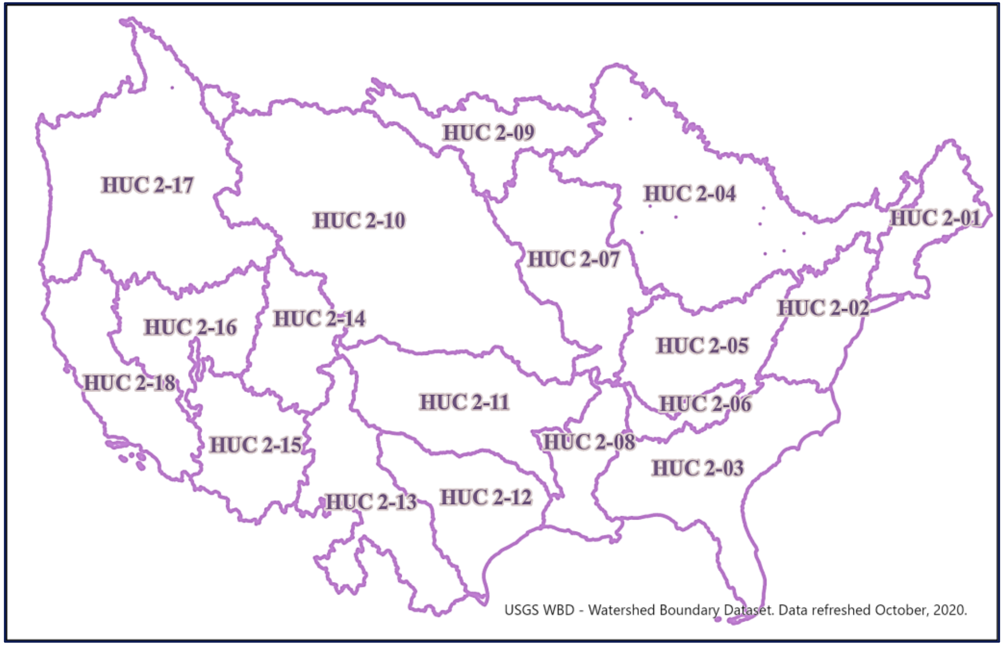
b) develop predictive **models for peak flow** response

c) advance **flood warning systems**

Regional Models across CONUS

CAMeLs Data – for

670 catchments



Methodology

Data Preprocessing

Identify Flood Peaks and triggering storm features

Exploratory Data Analysis

Select predictors after noting correlations and trends

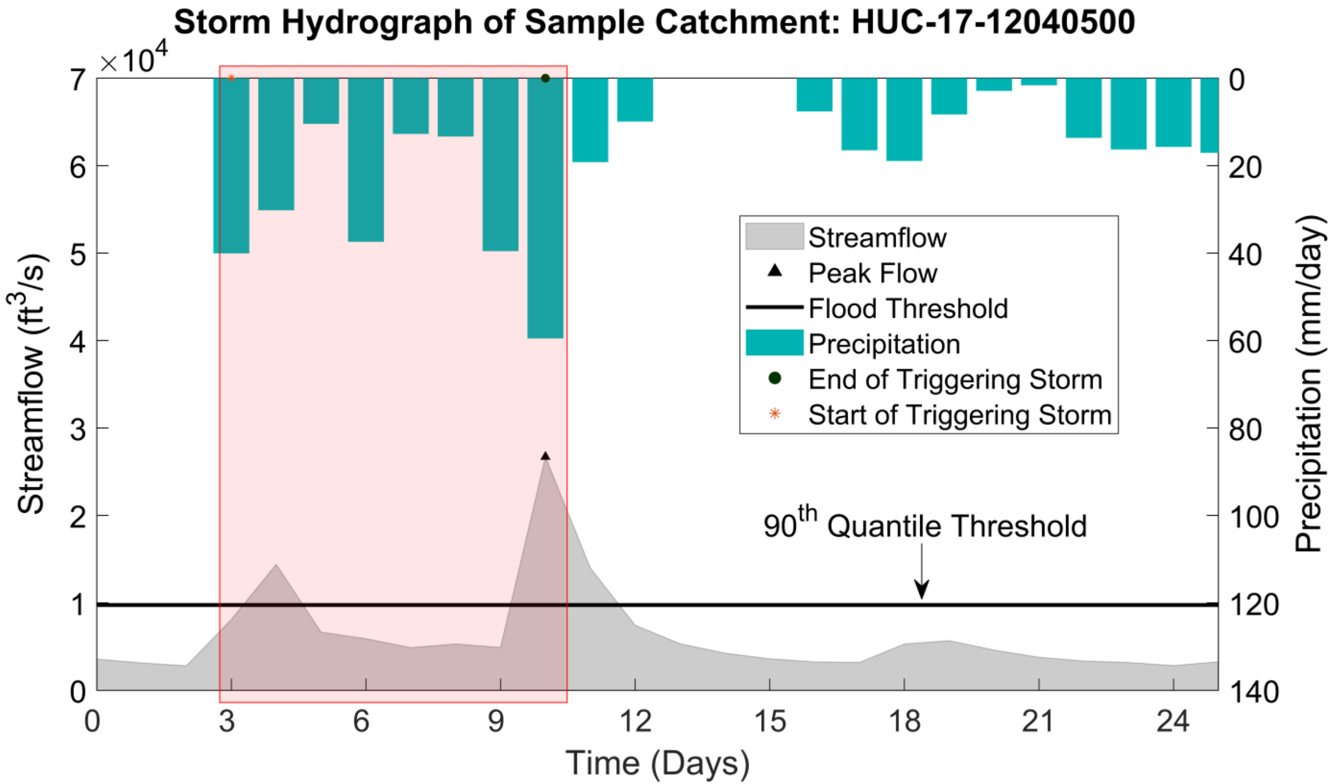
Model Validation

Develop predictive models; use LSTM (Kratzert et al. 2018) for comparison

Model Analysis/ Interpretation

Evaluation based on region and flood severity

Data Preprocessing



Rule-Based Models are “explainable”

Rule-Based

- Decision Tree
- Histogram-based Gradient Boost Regressor
- Random Forest

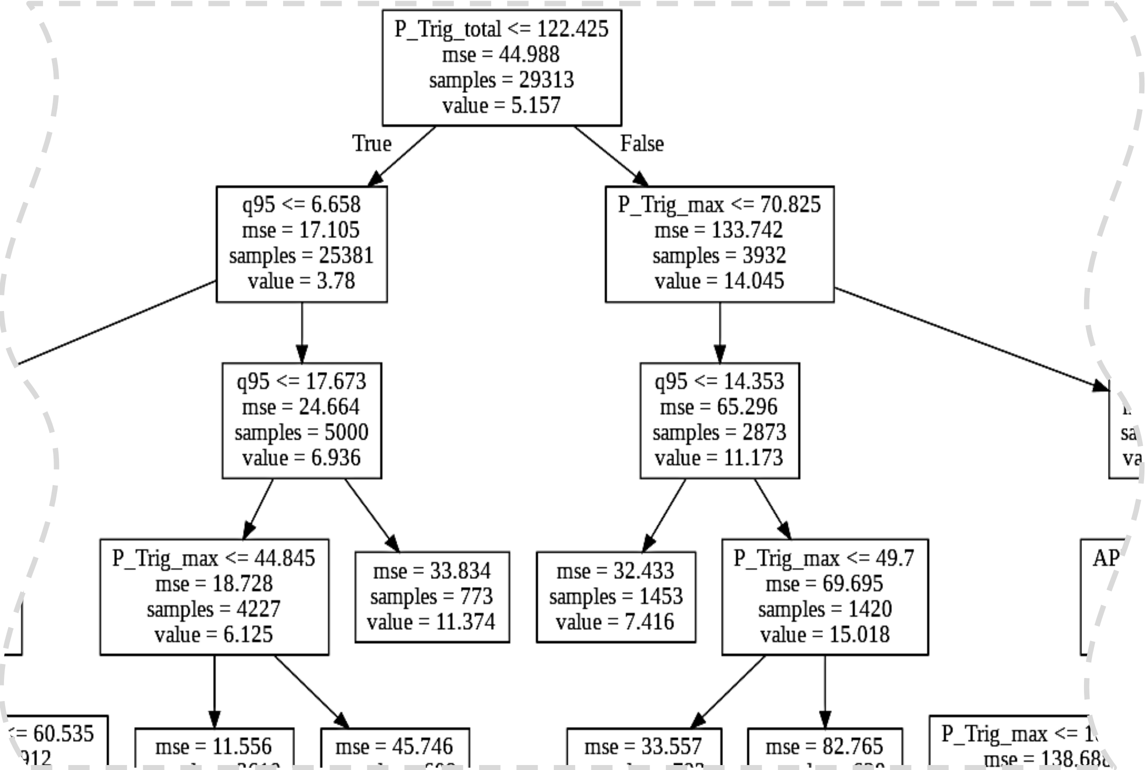
Non-Rule-Based

- Linear/LASSO Regression
- Multi-Perceptron (“Neural Networks”)
- Kernel (Ridge) Regression

HGBR and RF for predicting peaks

Ensemble tree-based learning algorithms

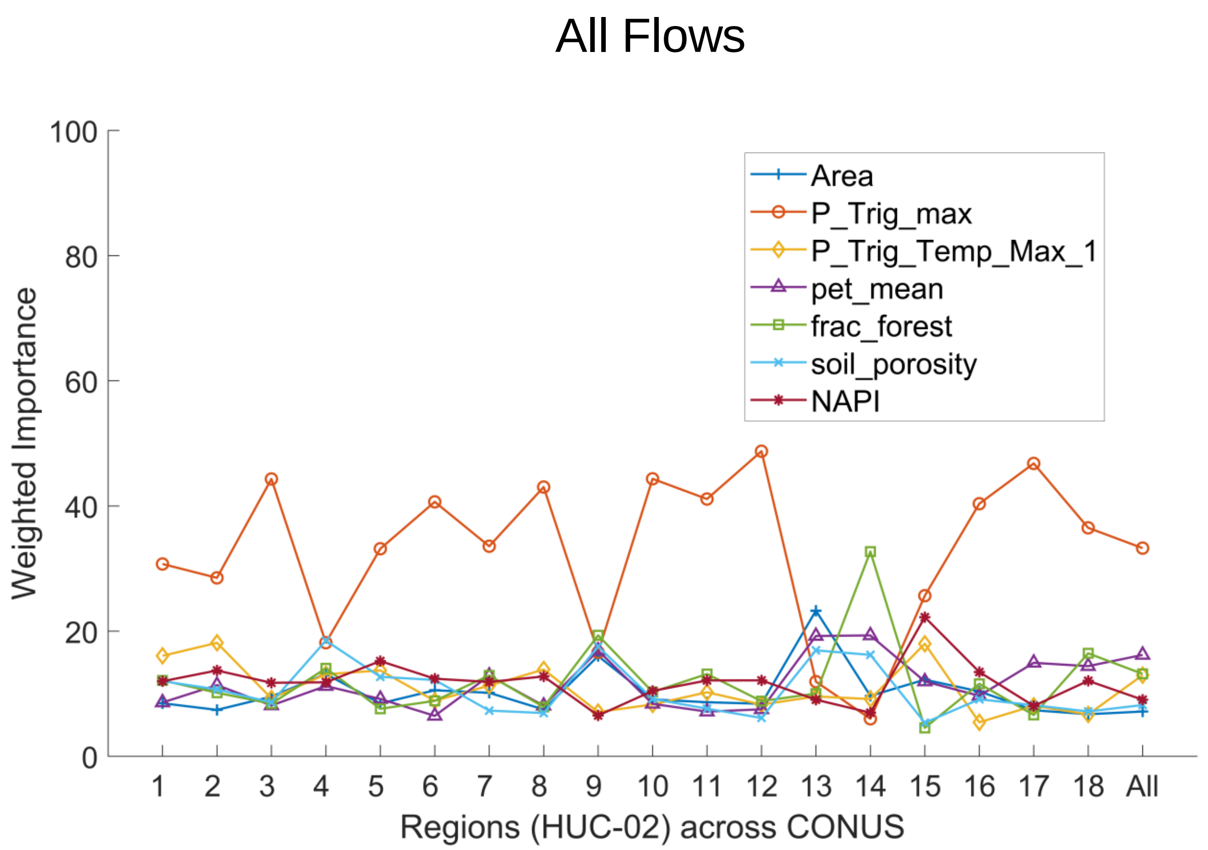
- HGBR – iterative
- RF - average



Summary of models and performance metrics

Summary of performance metrics for all models

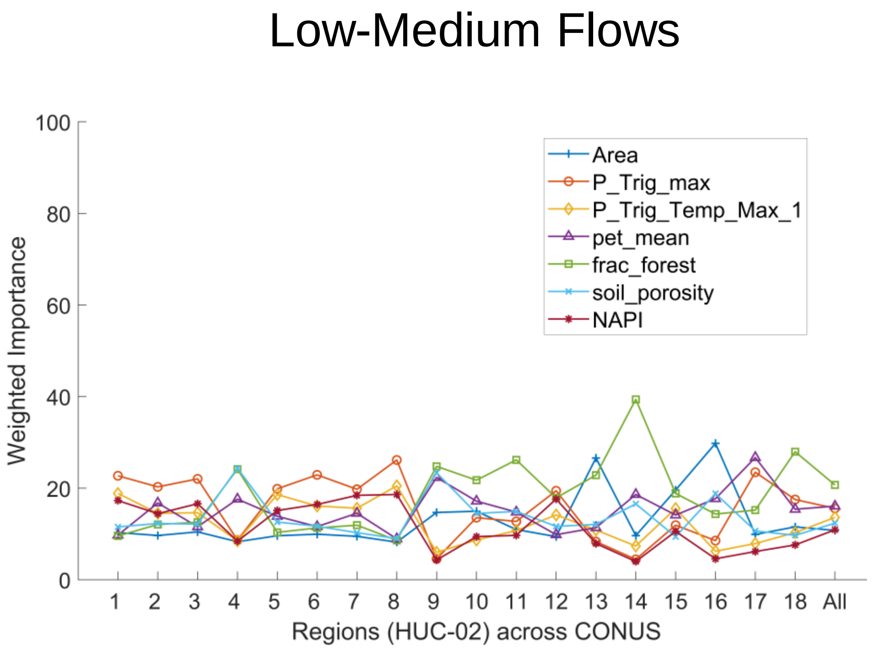
Model	Best Validation RMSE	Validation R^2
Histogram Based Gradient Boost Regressor	**4.08**	0.66
MLP (Multi - layer Perception)	4.26	0.63
Random Forest	**4.49**	0.55
Lasso Regression	4.54	0.58
Linear Regression	4.55	0.58
Kernel (ridge) Regression	4.64	0.56
Decision Tree	4.96	0.5
LSTM	**6.08**	

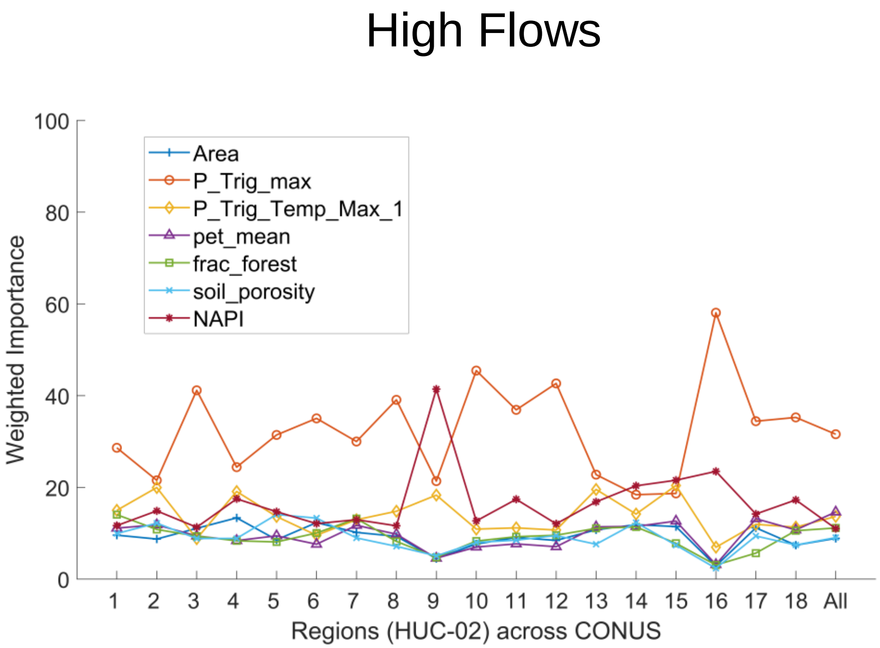


Importance of Variables

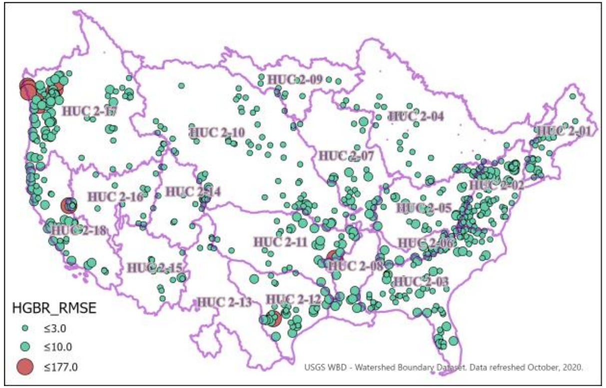
Regional dependence for predicting peak flows across CONUS

Variable behavior changes in response to flood severity

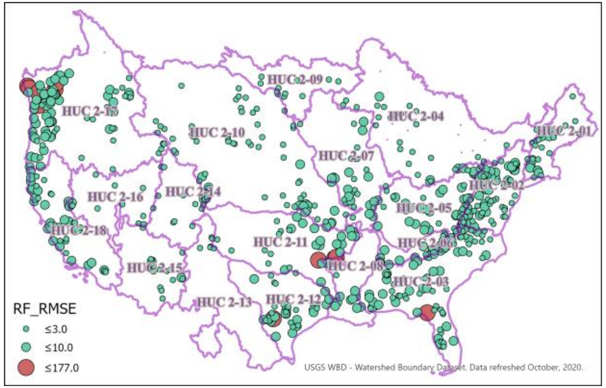




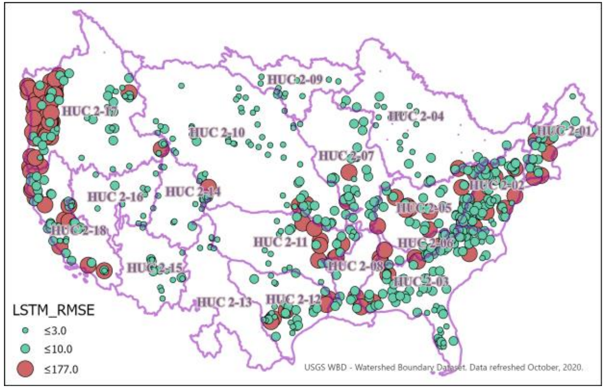
Mapping performance to Catchments



HGBR

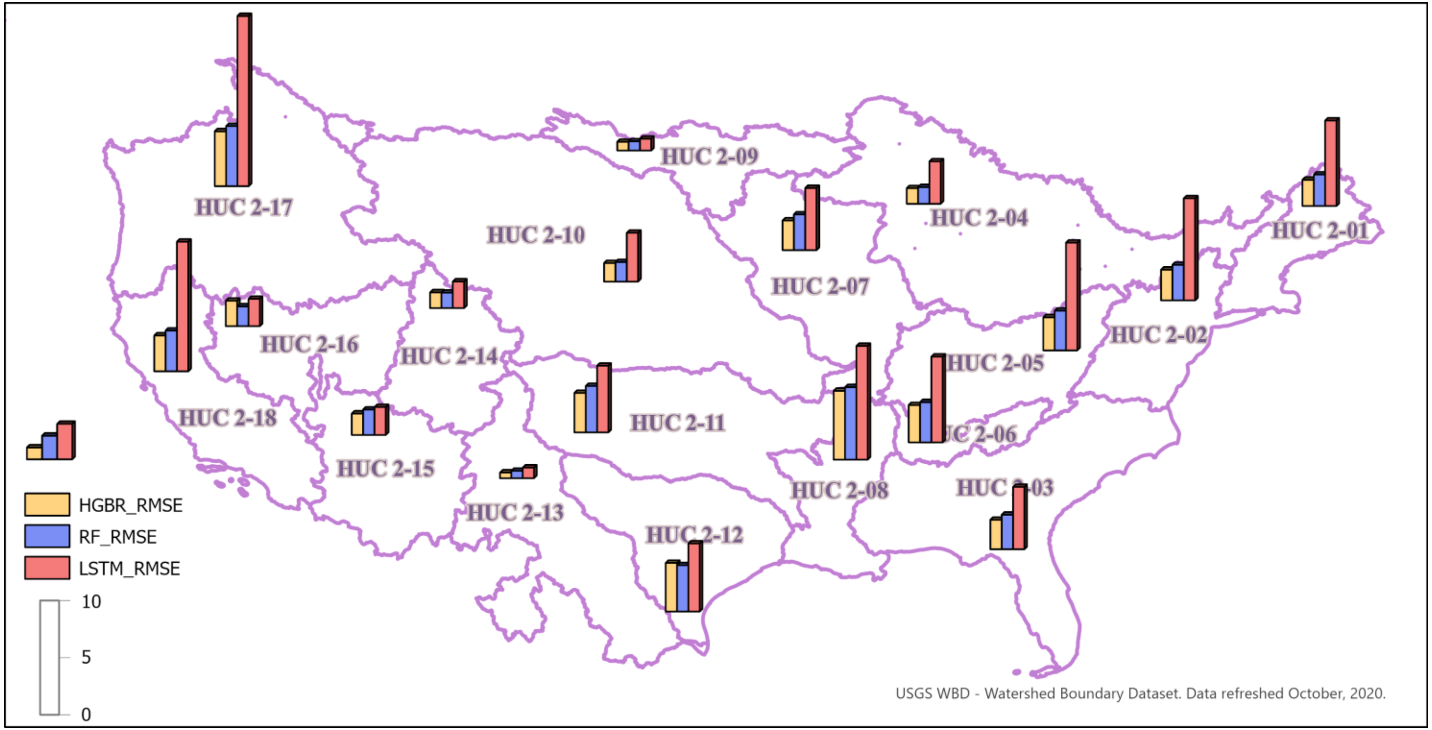


RF



LSTM

Mapping performance to Regions



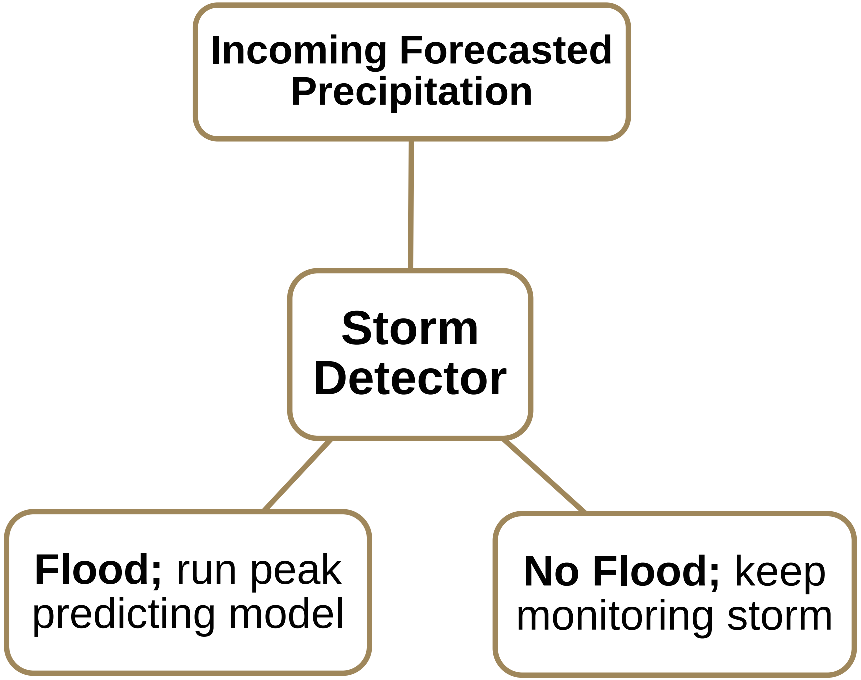
Compares the magnitude of RMSE per Region for three models

Conclusion

- HGBR and RF: demonstrated promising results for flood peak predictions
- Regional dependence of peak-flow predictions noted both per model and among models
- Precipitation controls flood response in high-flow events but its importance reduces for low-moderate flow events

Future Direction

- The second phase – addressing “**ungauged**” catchments
- From “Understanding” → “Application” via an **operational flood-prediction** framework



References/Acknowledgements

Addor, N., Newman, A. J., Mizukami, N. and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21, 5293–5313, doi:10.5194/hess-21-5293-2017, 2017

A. J. Newman, M. P. Clark, K. Sampson, A. Wood, L. E. Hay, A. Bock, R. J. Viger, D. Blodgett, L. Brekke, J. R. Arnold, T. Hopson, and Q. Duan: Development of a large-sample watershed-scale hydrometeorological dataset for the contiguous USA: dataset characteristics and assessment of regional variability in hydrologic model performance. Hydrol. Earth Syst. Sci., 19, 209-223, doi:10.5194/hess-19-209-2015, 2015

F. Kratzert, D. Klotz, C. Brenner, K. Schulz, & M. Herrnegger. (2018). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, *22*, 6005–6022. https://doi.org/10.5194/hess-22-6005-2018

We would like to acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation.

Nikolopoulos and Rasheed would like to acknowledge support from NSF grant # 1934712 (Collaborative Research: Near Term Forecasts of Global Plant Distribution, Community Structure, and Ecosystem Function)