Comparing QPF Driven Streamflow Ensembles from Models of Varying Complexity

Kyle K. Hugeback and Dr. Kristie J. Franz

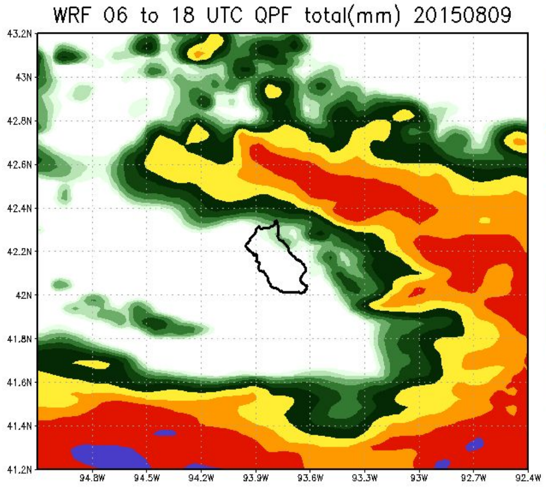
Iowa State University

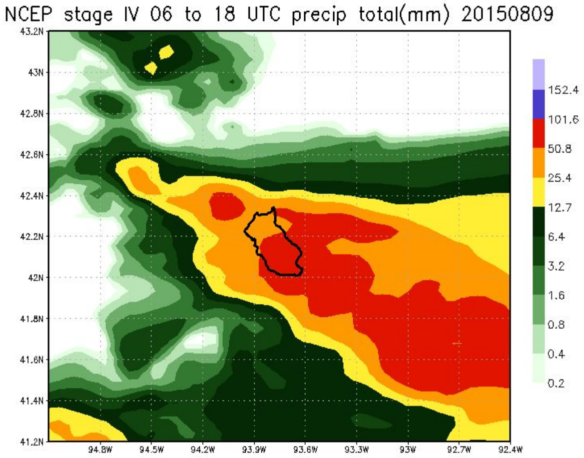
Department of Geological and Atmospheric Sciences

Presented by Dr. Kristie Franz

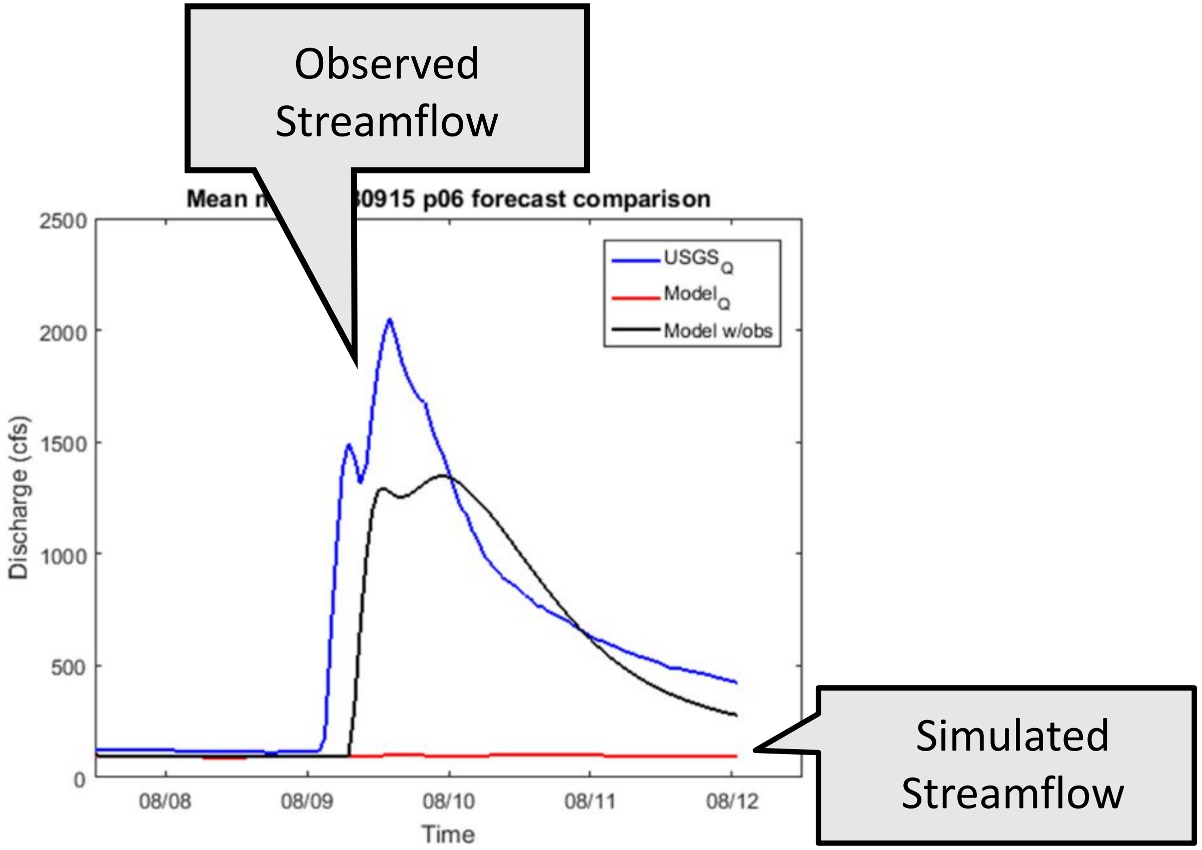
Motivation

Series of studies examining application of ensemble QPF and QPF adjustments for short-term streamflow forecasts (Carlberg et al., 2020, Goenner et al., 2020, Kiel et al., 2022, Hugeback et al., 2023)





Spatial displacement large enough to entirely miss watershed



Motivation

- We continue to explore QPF post-processing methods
  - Spatial displacement
  - Intensity
  - Timing
- Question arose, what hydrologic model should we use going forward?
- A second question was posed - does it matter?

Motivation

Hydrologic Forecast System

- Temperature
- Precipitation
- Potential evapotranspiration

Inputs

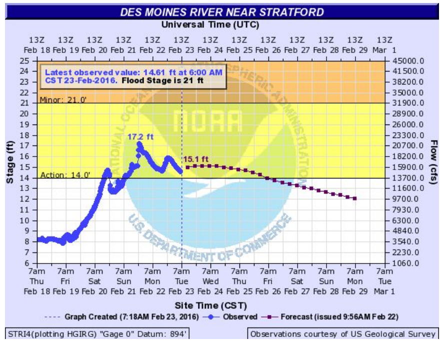
NWS models

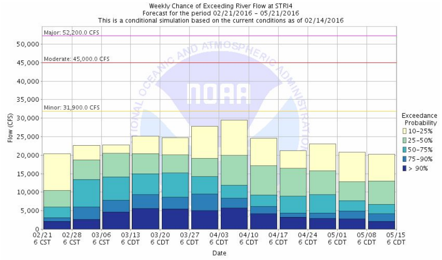
SNOW17
SAC-SMA
routing

Models

streamflow forecasts

Verification/ Observations





Uncertainty

Predicted precipitation adds its own sources of error and uncertainty

Research Question

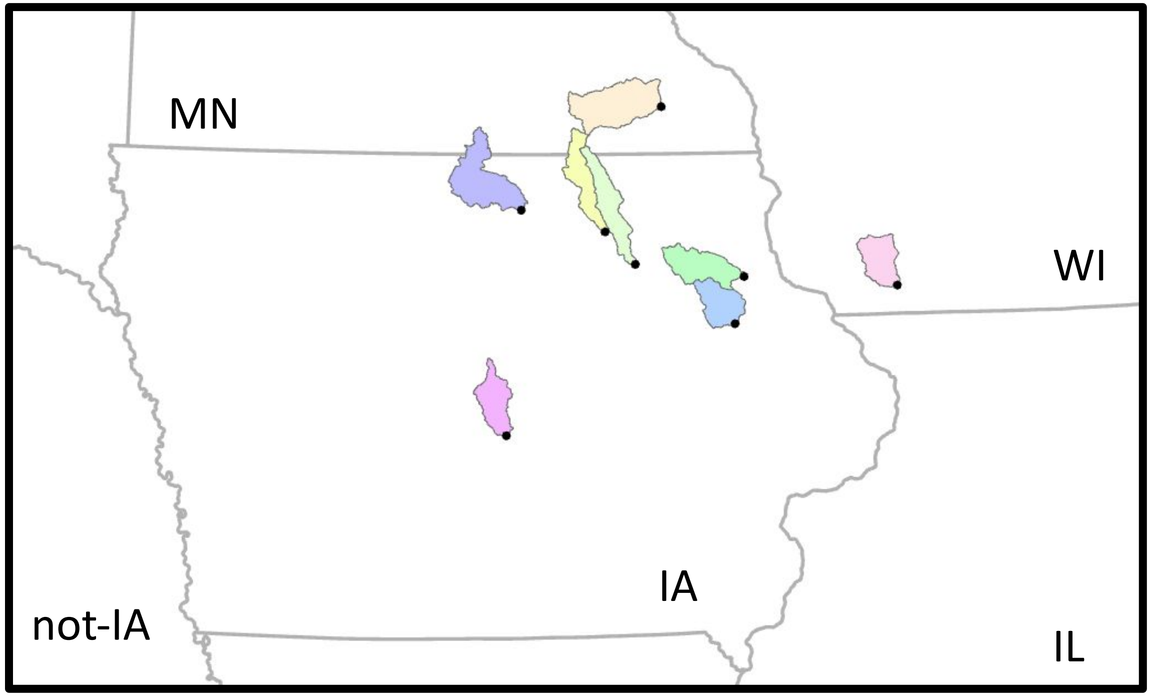
Does the uncertainty/error in the QPFs outweigh the uncertainty of the hydrologic modeling system, s.t. QPF-driven forecasts from different models are largely the same?

We hypothesize that the QPF error and uncertainty will dominate the streamflow prediction due to the challenges associated with accurately predicting precipitation over a specific watershed, similarly to the results found in Shu et al. 2022.

We verify and compare streamflow forecasts produced from three different hydrologic models given the same ensemble QPF.

Methods

- Study Sites
  - 8 headwaters
  - 700-1400 km^2
  - Basins chosen to have acceptable model NSE and Pbias from Hugeback et al. 2023 (WRF-hydro)
- Event selection
  - 2018 warm-season
  - Localized heavy rain across areas of the domain
  - Flash flooding watches and warnings issued for nearly all cases



Methods – Data Sources

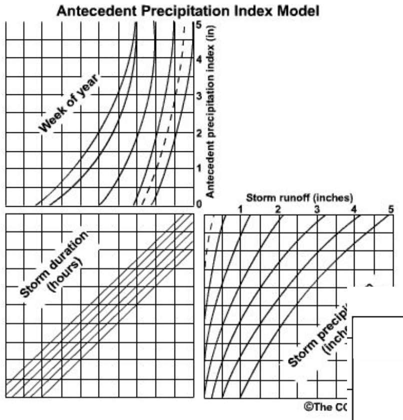
- High-Resolution Rapid Refresh Ensemble (HRRRE)
  - HRRRv3 dynamical core, 3km resolution
  - 9-member convection allowing ensemble
  - 36-hour simulation duration
  - Perturbations to water vapor mixing ratio, temperature, pressure, as well as zonal and meridional wind in the initial conditions and lateral boundary conditions

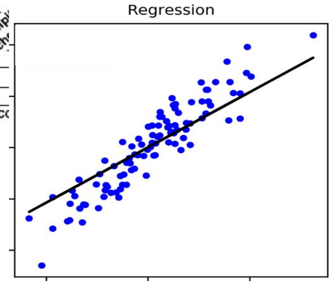
- North American Land Data Assimilation System, version 2 (NLDAS-2)
  - 1/8th degree resolution
  - 4 years worth of spin-up/training/calibration data
  - Oct 2013-spring 2018

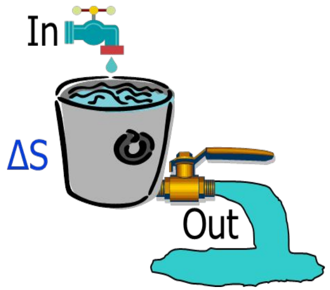
- Models run at 1 hour time step
- Inputs averaged to watershed scale depending on model requirements

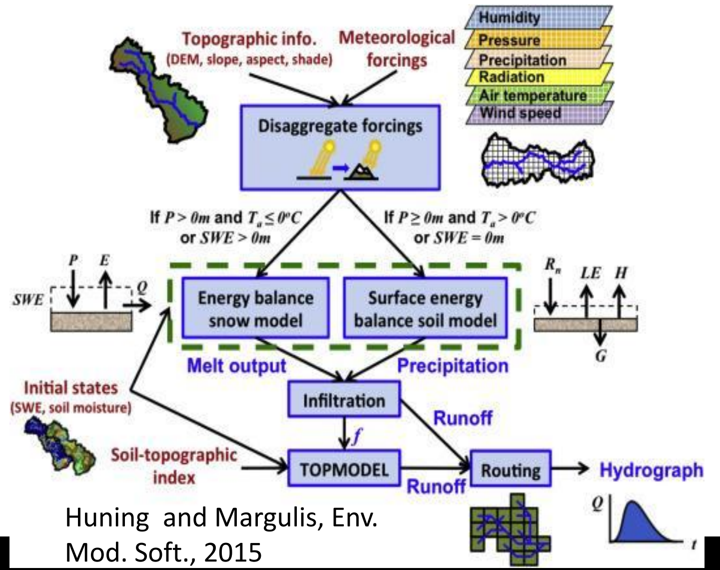
Background

Process representation





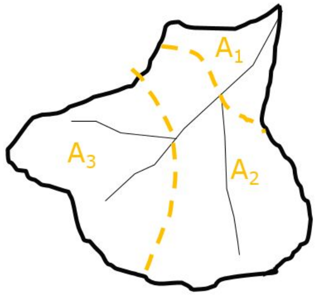




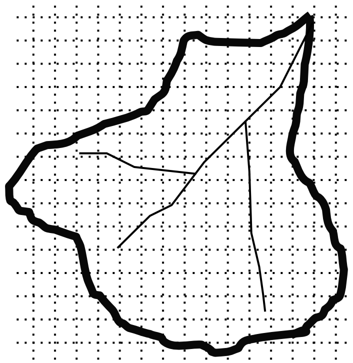
Huning and Margulis, Env. Mod. Soft., 2015

- A "plethora" of hydrologic models with different degrees of complexity is available?

Spatial representation

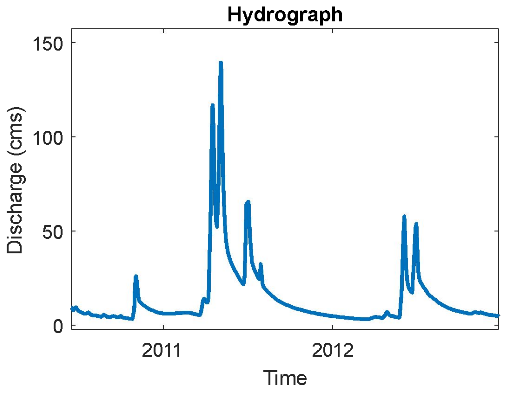


Lumped



Distributed

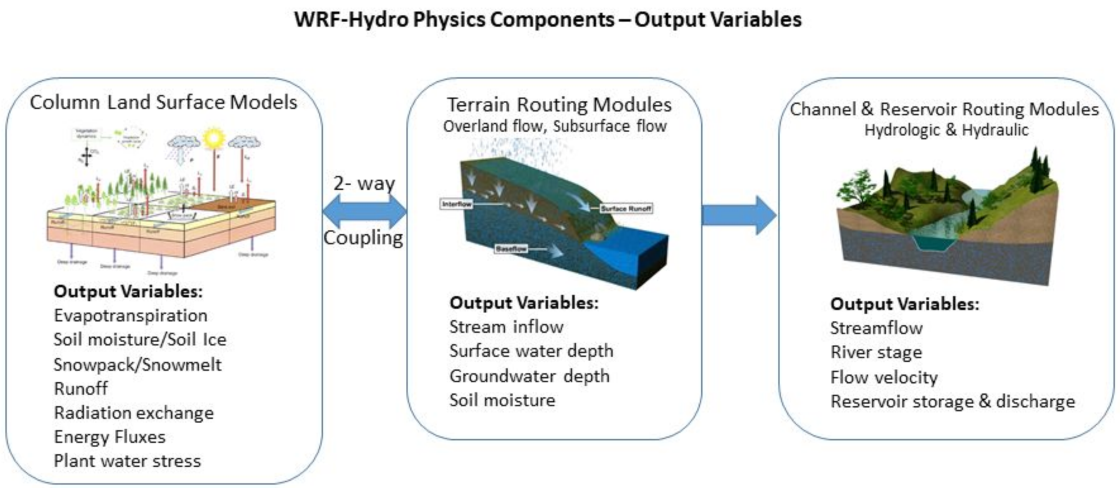
Temporal representation



Continuous vs event-based

Methods – Models

- **WRF-Hydro v 5.1.1 in National Water Model 2.0 configuration** (Gochis et al. 2020)



  - Fully distributed
  - Most physics-based of three
  - NoahMP LSM – 1km grid
    - 2m soil depth, 4 layers
    - Bucket model for ground water
    - 250m surface and subsurface routing, steepest decent
  - Muskingum-Cunge reach-based routing

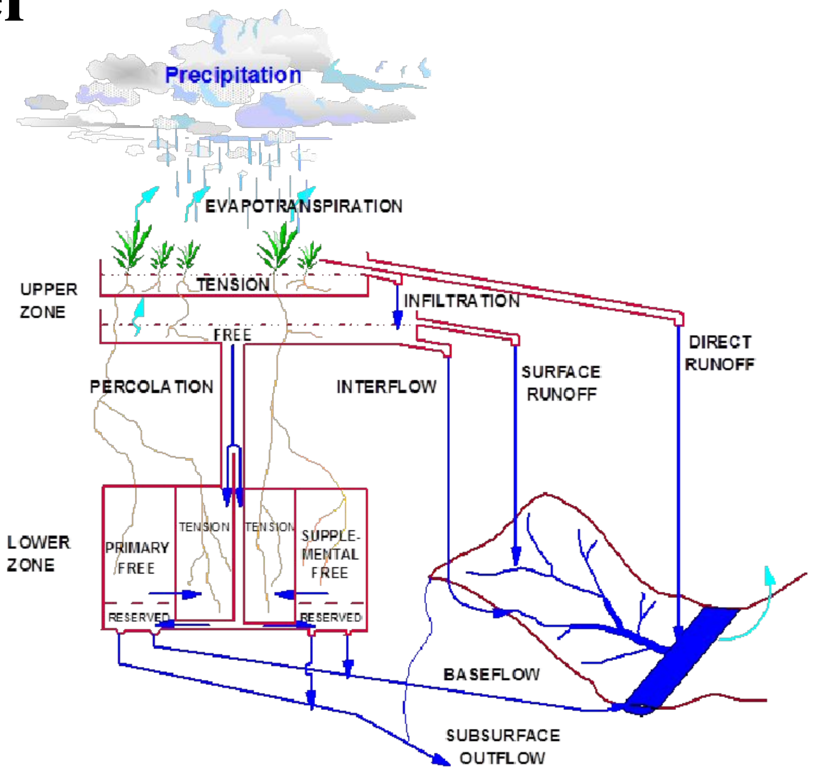
  - **Subdomain pulled from the NWM**
    - **Parameters and watershed configurations**

Methods – Models

- **Sacramento Soil Moisture Accounting Model** (SAC-SMA ; Burnash et al. 1973)

  - Lumped model
  - Conceptual processes
  - Two soil layers
    - Tension and free water storage
  - Five components of runoff

  - **Converted from 6-hourly timestep to 1-hourly**
    - **Ran simple SCE parameter calibration due to this conversion**
  - **1-hr unit hydrograph acquired from NCRFC**



Methods – Models

- **Long-Short Term Memory Recurrent Neural Network (LSTM)**
  - Machine learning model
  - Same input fields as WRF-Hydro but as hourly basin averages
    - Time fields added – Year, Month, Day, Hour, Minute
    - Normalized between 0 and 1
  - Model structure testing and hyperparameter tuning
    - Optimization of the model loss function (MSE)

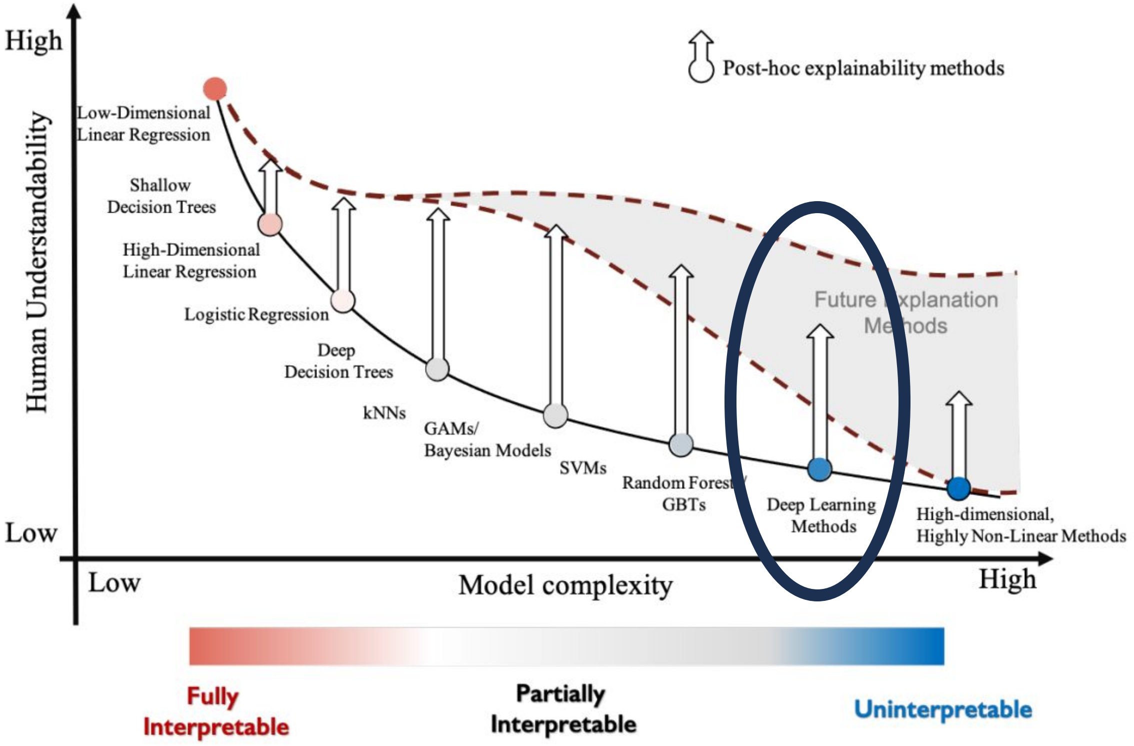


Figure from Flora et al. 2022

Methods – Models

Final test -

- **Multimodel Ensemble**
  - Constructed from all three hydrologic model outputs, 27 members

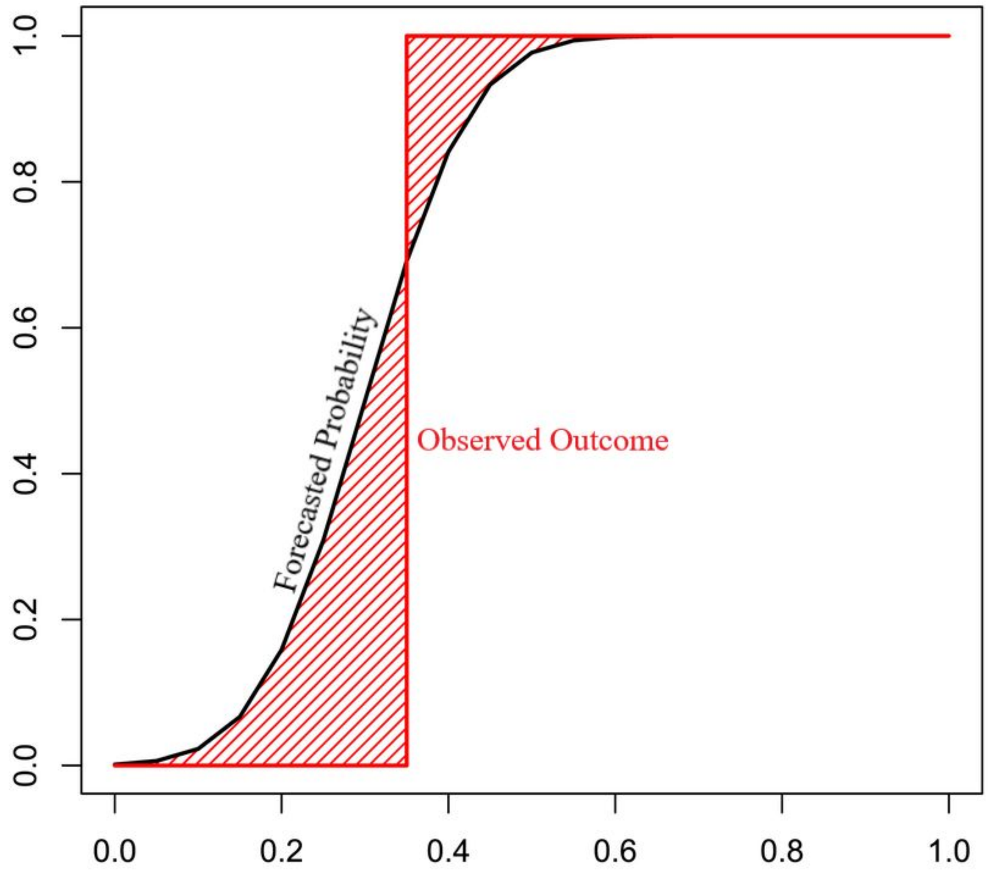
  - Ensemble weighting method
    - MSE in peak discharge for events used to weight the QPF+model members
    - Conducted validation by bootstrapping to hold out one event, and recalculate weights
    - Each member was ranked for all events, the average weight based on the aggregate of all the bootstraps was used for the final ranking and weighting for each gauge location

Methods – Ensemble Forecast Verification

- Evaluation based on probability of peak discharge falling into NWS flood categories.

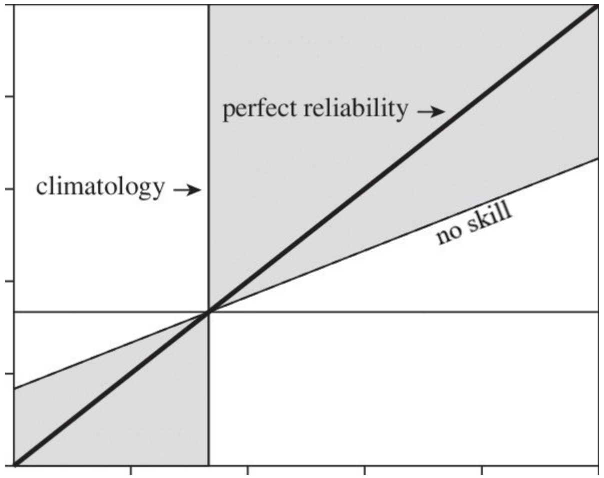
- Brier Score
  - Dichotomous
  - Based on action stage threshold

- Ranked Probability Score
  - Continuous
  - Based on all stages (e.g., no-action, action, minor, etc.)

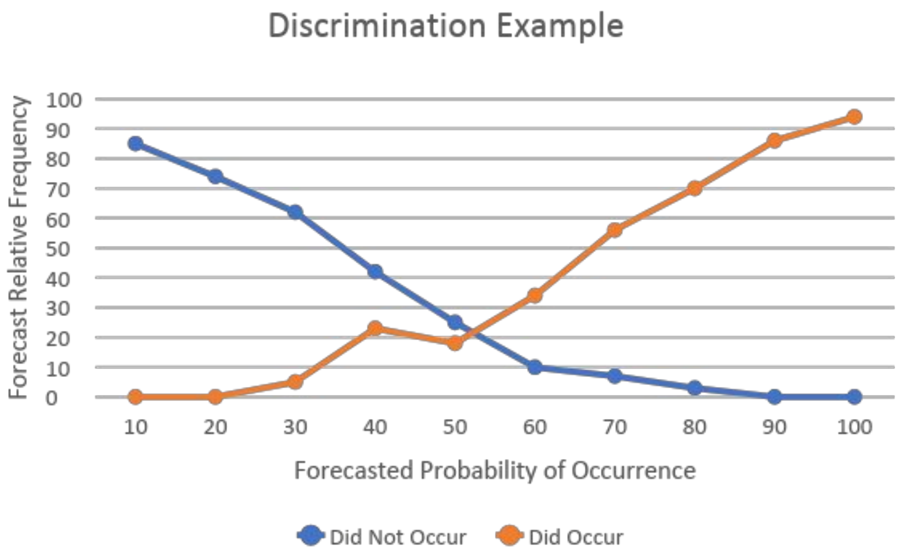


Methods – Ensemble Forecast Verification

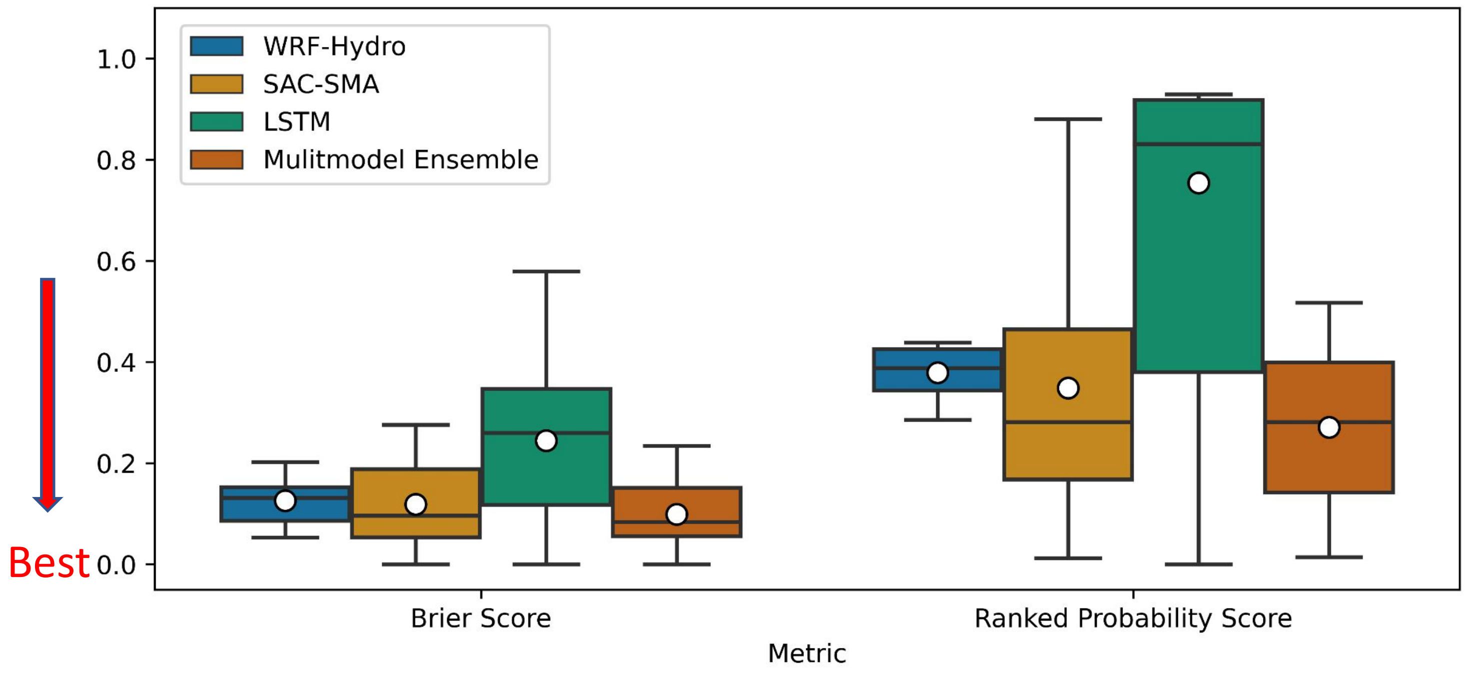
- Reliability
  - Relative frequency of observations given the forecasted probability
  - Action stage



- Discrimination
  - Relative frequency of forecasts given the observation
  - Tells us if action stage was forecasted for cases where action stage was or was not observed



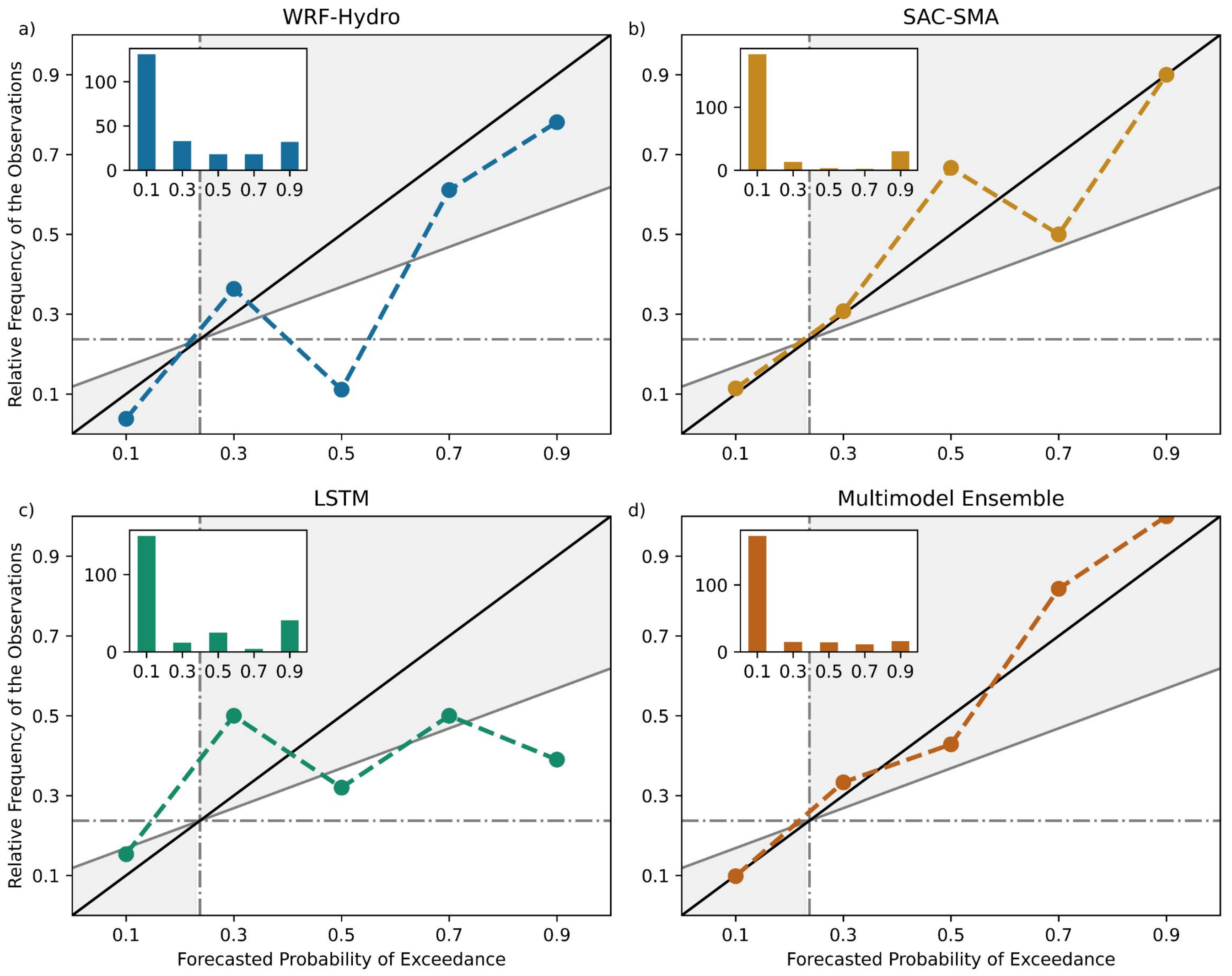
Results



- Brier Score
  - WRF-hydro and SAC-SMA forecasts had similar mean BS
  - WRF-Hydro scores most consistent across all gauge points and events
- Ranked Probability Score
  - SAC-SMA median score is best, but range larger than WRF-Hydro
- LSTM forecasts had largest range and poorer verification scores overall.
- Multimodel improved BS and RPS

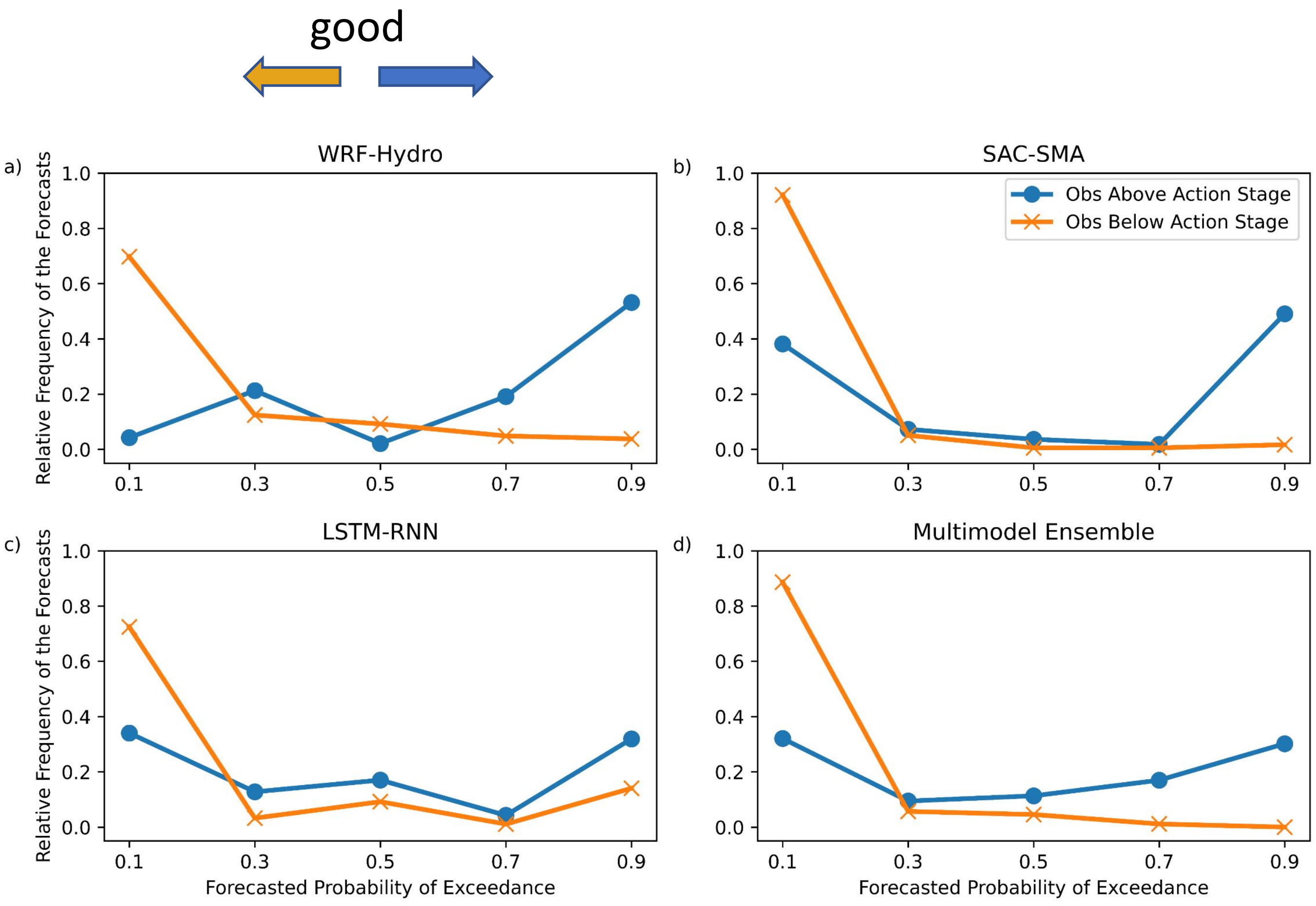
Results

- Reliability
  - WRF-Hydro and SAC-SMA forecasts have best reliability, with a few exceptions
  - SAC-SMA forecasts probabilities are poorly distributed
    - Low sample sizes in mid-probabilities
  - LSTM poorest reliability, esp. at high probabilities
  - Multimodel ensemble best reliability, slight underforecasting at higher forecast probability



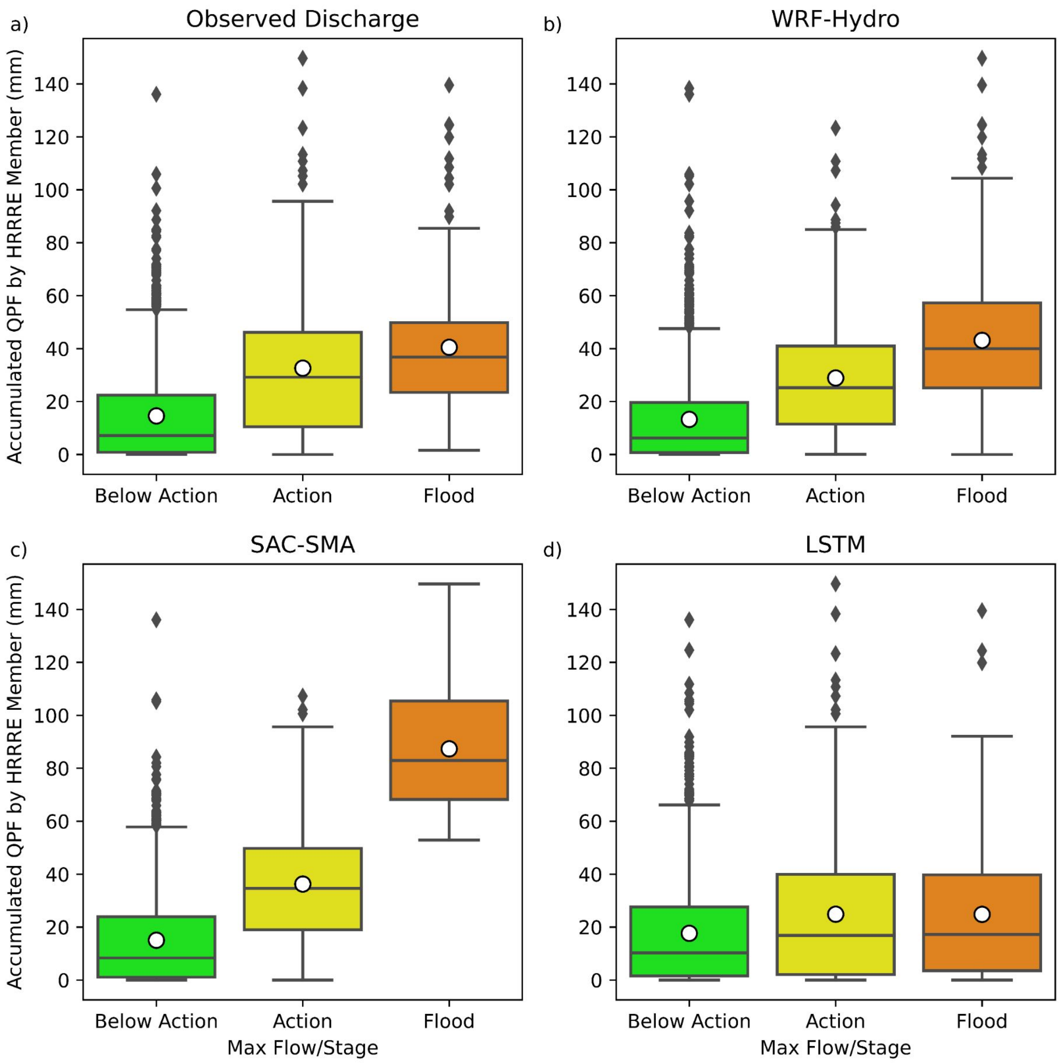
Results

- Discrimination
  - WRF-Hydro and SAC-SMA forecasts best discrimination for flow above and below-action stage
  - SAC-SMA tends to give low probability for action stage when it occurred
  - LSTM forecasts ok for events below action stage
    - Poor performance when event exceeded action stage
  - Multimodel: did not improve upon WRF-Hydro or SAC-SMA



Results

- QPF vs Discharge (observed/simulated)
  - Larger discharge associated with higher predicted precipitation
  - WRF-Hydro relationship similar to obs Q
  - SAC-SMA sensitive to higher amounts of QPF, results in significant streamflow response
  - LSTM seems to be indifferent to amount of QPF input



Conclusions

- SAC-SMA and WRF-Hydro forecasts performed well when evaluated for prediction of action stage

- LSTM performed the worst overall
    - Potentially data limited.
    - We used equal amount of input/training for all models, which may have left LSTM lacking the number of high-flow events needed in it's training sets to have a chance of success without the addition of physical process equations.

- Does the hydrologic model matter?
    - TAKE AWAY #1 ➡ Yes!
    - The model is not a simple transfer function.
    - QPF-driven ensembles from different models have different characteristics.

Conclusions

Summary of rankings from multimodel ensemble

Site	Best 9	Worst 9
DARW3	SAC-SMA	WRF-Hydro
RPMM5	LSTM	WRF-Hydro
IONI4	WRF-Hydro*	LSTM
MCWI4	LSTM*	WRF-Hydro
MCHI4	SAC-SMA	LSTM
TPLI4	WRF-Hydro*	LSTM
VLPI4	SAC-SMA	WRF-Hydro
MGOI4	WRF-Hydro	LSTM

* - Mixed with SAC-SMA members

- Which model is the best choice?
  - SAC-SMA was most often in the top 9 models (highest weight)
  - LSTM received the highest weights for two basins
    - Very poor performance at other sites may be reason for low average forecast skill
  - TAKE AWAY #2 ➡ Different models performed differently depending on the basin – there is no "one-size-fits-all" approach. Multimodel ensemble approach continues to show promise.

Thank you for listening!