2018

A Comparison of HREF and HRRRE Predictions for Ensemble Flash Flood Forecasting

Kyle K. Hugeback
Iowa State University

A Comparison of HREF and HRRRE Predictions for Ensemble Flash Flood Forecasting

Kyle K. Hugeback

Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa

Dr. Kristie Franz – Mentor

Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa

ABSTRACT

The objective of this study was to explore flash flood forecasting by looking at a comparison of streamflow discharge forecasts produced by the lumped Sacramento Soil Moisture Accounting Model (SAC-SMA) using quantitative precipitation forecasts (QPF) as input. The experimental High-Resolution Rapid Refresh Ensemble (HRRRE) and the operationally used High-Resolution Ensemble Forecast (HREF) were tested for three Iowa watersheds during the warm season. Past studies have found that warm season events pose the greatest uncertainty for rainfall prediction, which contributes to uncertainty in streamflow prediction. Precipitation ensembles help to cover the spread of uncertainty and provide value to hydrologic forecasts through the use of random perturbations to their input characteristics and boundary conditions. Datasets for the HRRRE and HREF were collected and evaluated, then processed to yield basin average QPF over the selected watersheds. The QPF ensembles were then fed into the SAC-SMA hydrologic model along with interpolated potential evapotranspiration (ET) and temperature, following a model spin up that ran from 2016 up to the time of the event. For the three watersheds and seven events stretching from the 20th of June to the 5th of September, the HREF outperformed its experimental counterpart. The forecasts produced using the operational HREF outperformed those produced using the experimental HRRRE for evaluations of peak discharge forecasts for both the full distribution of ensemble members and the ensemble mean. This is seen in lower biases between the mean discharge and the observed. Furthermore, the HREF had a lower ranked probability score than the HRRRE when evaluated for peak discharge. The HRRRE-based forecasts were more accurate for prediction of peak discharge timing than the HREF-based forecasts.

---

1. Introduction

According to the Water Resource Services Branch of the National Weather Service (NWS), total losses due to flooding in 2017 stand at $61.4 billion. In addition to the monetary loss, the human cost totaled 137 lives lost in the United States (NOAA, 2018). In the Midwest US, flooding poses a significant threat to agricultural production as well as homes and cities. Mitigating these risks is of the utmost importance.

Increasing warning times is essential to mitigating loss and allowing for a quicker mobilization from first responders. With larger river flood events, the rise in water can be seen more easily because response times—the time it takes for runoff to enter a stream and be reported by a river gauge—are on the order of many hours or days, depending upon the basin size and characteristics. Though there may be more warning associated with significant events, that does not make them easy to avoid. Advanced warning from hydrologic forecasts supports substantial community efforts to sandbag or erect waterproofing barriers. These forecasts are crucial in saving life and property.

The task of forecasting such events falls on River Forecast Centers and Weather Forecast Offices around the country. With so many basins, calibrations to make and parameters to keep track of, furthering the science of accurate and timely forecasts is key to helping the public and decreasing the burden on forecasters.

In this study, QPF from the High-Resolution Rapid Refresh Ensemble (HRRRE) and the operationally used High-Resolution Ensemble Forecast (HREF) was processed and run through the lumped SAC-SMA for the purpose of producing flood discharge forecasts to evaluate the use of these ensembles for streamflow prediction.

2. Literature Review

The warm season of the year is when the bulk of the precipitation in the Midwest and Central Plains is received. These events are associated with large thunderstorm complexes called Mesoscale Convective Systems or MCS's (Gallus, 2012). Convective storms are characterized by towering cumulus clouds on the meso- to micro-scale. Convective storms also produce the highest rainfall rates. Their propagation and duration depend heavily on the environment and large-scale forcing in the region. All those characteristics previously described contribute to the complexities of convective modeling. Sharp gradients, both spatially and temporally, of rainfall intensity give rise to errors when attempting to model MCS's and other related convective storms using coarse grids of even a few kilometers. Resolving phenomena, that lie within a grid box requires the use of parameterization. Parameterizations are a scheme of equations to simplify the computation of specific values, or parameters, that cannot be explicitly resolved. Oversimplifications, such as this, can be a contributing factor in the failure of models to produce accurate rainfall estimates (Gallus, 2012).

Ensembles are a group of model members that, through the use of random initial perturbations, provide a forecaster with a more probabilistic view of a weather event. By having multiple members, with a diversified set of physics schemes, it has been shown that ensemble forecasts reach a higher accuracy than any single member (Ebert, 2001; Du et al., 1997). Given the issues present with individual member forecasts, especially the shortfalls in warm season precipitation forecasting, it is imperative that ensemble forecasts be used to fill this informational gap (Gallus, 2012).

Though separate, generally an ensemble's members show similarities in their outputs. Gilmour et al. (2001) looked at the evolution of ensemble runs in terms of their linearity. Researchers found that due to the nature of perturbations, linearity between members of the ensemble may break down earlier than previously thought. The time frame discussed in this work was on the order of 24 hours.

Looking more specifically at the models featured in our study, Seo et al. (2018) investigated the uncertainty in High-Resolutions Rapid Refresh (HRRR, the same base model as the HRRRE) QPF forecasts during a September 2016 flooding event in Iowa. This study investigated uncertainty through mean areal precipitation (MAP) analysis and the hillslope-link hydrologic model (HLM). Their work concluded that the model QPF contributed to an overestimate in the peak of the flood wave. The peak also arrived earlier than seen in the observations. In their paper, they describe their work as continued evidence that the HRRR struggles with the prediction of short-term precipitation forecasts and uncorrected QPF. These forecasts introduce considerable uncertainty into the hydrologic model. With a moderate agreement between both discharge output from the HLM and MAP analysis showed that QPF error was the highest for short lead times. Only with 4 to 6 hours of lead time did uncertainty improve. In their discussion, they describe their work as continued evidence that the HRRR struggles with the prediction of short-term precipitation forecasts and uncorrected QPF from these forecasts brings in considerable uncertainty.

This same conclusion about uncertainty early in model runs was reached by Du et al. in their 1997 paper on short-term ensemble forecasting. They cited issues with ingesting initial conditions as an abundant source of error and added uncertainty. The HRRR model QPF overestimated the peak of the flood wave. The peak also arrived earlier than seen in the observations.

The second ensemble that this study focused on is the HREF. Though research done on the HREF is scarce, research has been completed on many of its components. A 2015 publication by Barthold et al. looked at the improvement of flash flood forecasting through the work of the Hydrometeorological Testbed at the Weather Prediction Center in partnership with the National Severe Storms Laboratory (NSSL). QPF from the Storm Scale Ensemble of Opportunity (SSEO) was incorporated into their comparisons to flash flood guidance. This is significant to this research because the HREF version two, which became operational in late 2017, is heavily based on the SSEO (UCAR, 2017). Barthold et al. (2015) concluded that improving datasets could contribute to better flash flood guidance. Storm-scale ensemble development is essential to the future of flash flood forecasting convective-allowing models.

With ensemble forecasts showing skill in some areas and uncertainty in others, adding the hydrology component to these forecast shows interesting information. In their study of water supply forecasts using the NWS's Ensemble Streamflow Prediction (ESP) system in the Colorado River Basin, Franz et al. (2002) brought forth good discussion on the usefulness of ensemble forecasts. They explained how single-member forecasts fail to carry the same depth of information that is contained in a full ensemble distribution. The strength of these ensembles lies in the probabilistic forecasts that can be derived from their members. Due to this, the authors of the study recommend others use ensemble forecasts in a probabilistic manner. In their evaluation of ESP techniques, they used a ranked probability score (RPS) and other distribution measures like reliability and discrimination (Franz, 2002).

The lumped Sacramento Soil Moisture Accounting Model (SAC-SMA), which has been chosen for this study, cannot operate strictly on QPFs. Another critical input is evapotranspiration (ET). A recent study from 2017 looked at making comparisons between several sources of ET data. Moreover, during their investigation, they used several of the same watersheds that this study will focus on (i.e., Squaw Creek and the South Skunk River), which allows their findings to more directly tie to this research. In the case where they looked specifically at the lumped SAC-SMA, the simulated potential evapotranspiration (PET) performances were worse than those provided by the North Central River Forecast Center (NCRFC), though the values used by the NWS are lower than the observed PET. When using the simulated PET from the MODIS satellite or ET demand curves, the lumped SAC-SMA model underestimated stream discharge and did not lead to any measurable improvements over the ET provided by the NCRFC (Bowman et al., 2017). Due to these findings, an ET dataset from the NCRFC will be used to conduct this study.

3. Data and Methods

a. Study Domain

The entire domain of this study lies within the state of Iowa. The three watersheds are shown in Figure 1. The first two watersheds chosen for this study are located in the central third of the state. The South Skunk River has two gauges in the Ames area. For the use in this study, the gauge upstream of the confluence with Squaw Creek was chosen. At this point in the river, it covers 315 square miles of drainage. The second watershed chosen was Squaw Creek in Ames. The creek is gauged roughly two and a half miles upstream from its confluence with the South Skunk. At the point of the gauge, it drains 204 square miles. The last of the three is the Turkey River. It is gauged in Spillville, where it drains an area of 177 square miles. This watershed sits on the edge of the Paleozoic Plateau of Northeast Iowa, offering up much hillier terrain than that of Squaw Creek or the South Skunk River.

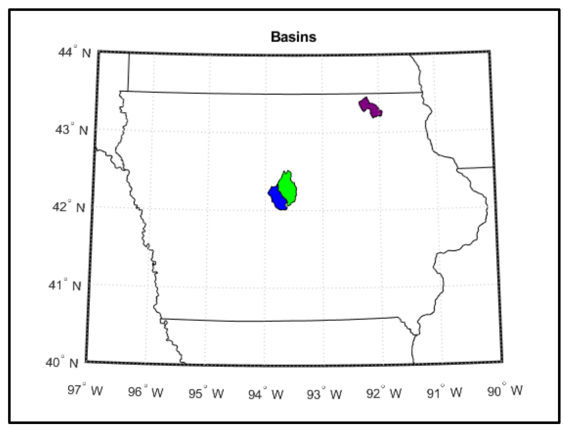


Figure 1: Watersheds used for this study include the Squaw Creek in Ames (blue), the South Skunk River near Ames (green), and the Turkey River at Spillville (purple).

b. Event Selection

For the selection of events to be used in this study, data was surveyed from the United State Geological Survey database. The HRRRE became the limiting factor. Its model runs could only be acquired up to 48 hours after they took place. Due to this, collecting a more significant number of near flood cases meant that the exceedance in the observed dataset had to be lowered. In the end, the lower threshold was moved to 70% of action stage.

Three cases were found for the South Skunk River: June 20th, July 1st, and Sept. 5th. The Squaw Creek also saw peaks on July 1st and Sept. 5th. Lastly, the Turkey River saw local maxima in discharge around Sept. 2nd and the 5th.

c. SAC-SMA Model and Parameters

The hydrologic model being used in this study is the Sacramento Soil Moisture Accounting model. In its lumped configuration, it is being used operationally by the NWS River Forecast Centers. All the watersheds in this study are contained within their jurisdiction. The SAC-SMA conceptual rainfall-runoff model represents the soil with two layers (Burnash et al., 1973; Burnash, 1995). There is a thin surface layer, with a much more substantial sublayer contained below that. Each layer keeps track of several factors, such as free water, tension water storage, evapotranspiration and diffusion (Bowman et al., 2017). This model is a natural choice to use when comparing the operational feasibility of QPF forecasts because of its widespread use among RFC offices. For its operation, the SAC-SMA needs temperature (for the SNOW17 part of the model), ET, and precipitation.

Parameters that describe the water storage characteristics for each basin and a unit hydrograph for each watershed were obtained from the NCRFC. A full list of the parameters noted above is included in Table 1.

Table 1: SAC-SMA parameter descriptions have been adapted from Spies et al. 2015

UZTWM	Upper-zone tension water max storage
UZFWM	Upper-zone free water max storage
UZK	Upper-zone free water lateral depletion rate
ZPERC	Max percolation rate
REXP	Exponent of the percolation rate
LZTWM	Lower-zone tension water max storage
LZFSM	Lower-zone free water supplementary max storage
LZFPM	Lower-zone free water primary max storage
LZSK	Lower-zone water depletion rate
LZPK	Lower-zone primary free water depletion rate

d. QPF Data Sources

The first of the two sources of QPF data was from the HRRRE. This experimental model is attempting to improve 12-hour forecasts through real-time data assimilation and testing ensemble design for 36-hour forecasts among other things. The model is based on the Advanced Research, Weather Research and Forecasting Model (WRF-ARW) version 3 convective allowing model. The model runs a 15-km outer analysis grid with a nested 3-km grid. The ensemble consists of a nine-member spread. Each member is produced through random perturbations to the wind vectors U and V, temperature, water vapor mixing ratio, dry air in the column, and boundary conditions, as well as 15-minute radar data assimilation (Dowell et al., 2018). Data from the HRRRE was only available for 48 hours after each model run was released. HRRRE runs were acquired from the Earth System Research Laboratory database. In particular, the dataset used by this study was sourced from other researchers at Iowa State University.

The second source of data comes from the HREF, which is used operationally by the NCRFC for its streamflow forecasts. This is an ensemble system produced by the Storm Prediction Center that is the operational version of the Storm-Scale Ensemble of Opportunity (SSEO). The two most recent runs, of each of high-resolution WRF-ARW, Nonhydrostatic Multiscale Model on the B-grid (NMMB), “NSSL-like”-ARW, and a North American Mesoscale Forecast System (NAM) run that spans the full continental United States (CONUS), make up the group of eight members. The HREF is run using a 3.2 km grid for the CONUS. Because model runs are used to populate the full list of members, weighting must be applied to mitigate any influence that could be contributed by the time lag between model runs (UCAR, 2017).

e. MAP, MAT, and PET

Data for MAP and mean areal temperature (MAT) data from each watershed used in this study have been provided by the NCRFC in Chanhassen, MN. These data sets are used to produce warm starts for the model as to accurately represent the soil conditions before one of the chosen events. Similarly, PET comes from the NCRFC, but it is only reported on the 16th of each month.

f. Running the Hydrologic Model

Both ensemble datasets were processed to provide the necessary inputs for the hydrologic model. The HRRRE GRIB files were converted to NetCDF format using wgrib2 commands. Then both the HRRRE and HREF QPF had to be extracted, six-hour totals computed, and basin averages figured for each watershed. All of the extracting of QPF through basin averaging was done using Matlab scripts.

MAT and MAP were required to be delimited by date and field, then processed into a readable format for the model to interpret. A linear interpolation was done for the PET the produce data for each day of the year. Once again, Matlab scripts were made to aid in the data processing. With all the data in a usable format, MAP, MAT, and PET were combined to make three basin specific spin-up datasets.

Once the proper unit hydrograph and parameter file had been placed in the model’s working directory, a run could then be initiated with those spin-ups. It ran from the beginning of 2016 until the flood event. The spin-up was then cut off and replaced by the ensemble QPF. Due to the efficiency and speed of the lumped SAC-SMA, the model was run separately for each member of each ensemble for all events.

g. Key Analysis Methods

Before any runs of the SAC-SMA were made, basin average precipitation values were extracted and then compared to the MAP data. Equation 1 shows the simple subtraction used to find precipitation bias (P Bias). This can be considered a control variable for these forecasts.

$$QPF - MAP = PBias \quad (1)$$

Once model runs were completed, ensemble averages were computed. Those ensemble averages were compared to the observed peak discharge and its timing.

When looking more specifically at the behavior of individual members, a RPS was calculated to evaluate the usefulness of each ensemble in a probabilistic sense. Franz et al. (2003) describe the process of producing these scores. First, ensemble members must be broken up into streamflow non-exceedance categories, in this study the following are used: <50%, 50-70%, 70%-Action stage, Action-Flood stage, and >Flood stage. Then, forecast cumulative distributions (F_m) must be calculated. F_m looks at the frequency of members (f_j) occupying each of the categories (Equation 2):

$$F_m = \sum_{j=1}^{m} f_j, \quad m = 1, ..., J. \quad (2)$$

Similarly, observed streamflow is put through the same process, by assigning the category that was exceeded a value of 1, while all the lower categories were assigned a value of 0. In Equation 3, o_j will either be a one or a zero depending on whether the category was exceeded.

$$O_m = \sum_{j=1}^{m} o_j, \quad m = 1, ..., J. \quad (3)$$

RPS is then calculated as the summation of the square of the forecast cumulative distribution (F_m) minus the observed distribution (O_m) as displayed in Equation 4. An average can be taken of the RPS to be able to look at each ensemble across all events. This takes into account the spread of all members. The smaller the RPS, the closer the member spread was to mimicking the observed event.

$$RPS = \sum_{m=1}^{J} (F_m - O_m)^2. \quad (4)$$

Wet/dry biases were also computed. The unique perturbations, model configurations, and physics parameterizations used by the different members can lead to consistent biases. Looking at wet and dry biases can provide useful information about individual members affecting averages. Discharge bias (Q Bias) values were calculated for each member for each event. This was done by subtracting the observed peak discharge (Obs Q) from the member's peak discharge (Mem Q) (Equation 5). Furthermore, the average of all the events, and then that process was completed once more for all basins.

$$(Mem\ Q) - (Obs\ Q) = Q\ Bias \quad (5)$$

It is also important to note that a healthy spread conveys nearly as much critical information to a forecast as does a perfect forecast from one member of the ensemble when looking at a peak in the post-event stage. An ensemble distribution provides a forecaster with a "what if" scenario that can aid in the decision making regarding public response. An ensemble distribution can only aid a forecaster if the event was captured within the spread of the members. A good example of this is shown in Figure 2.

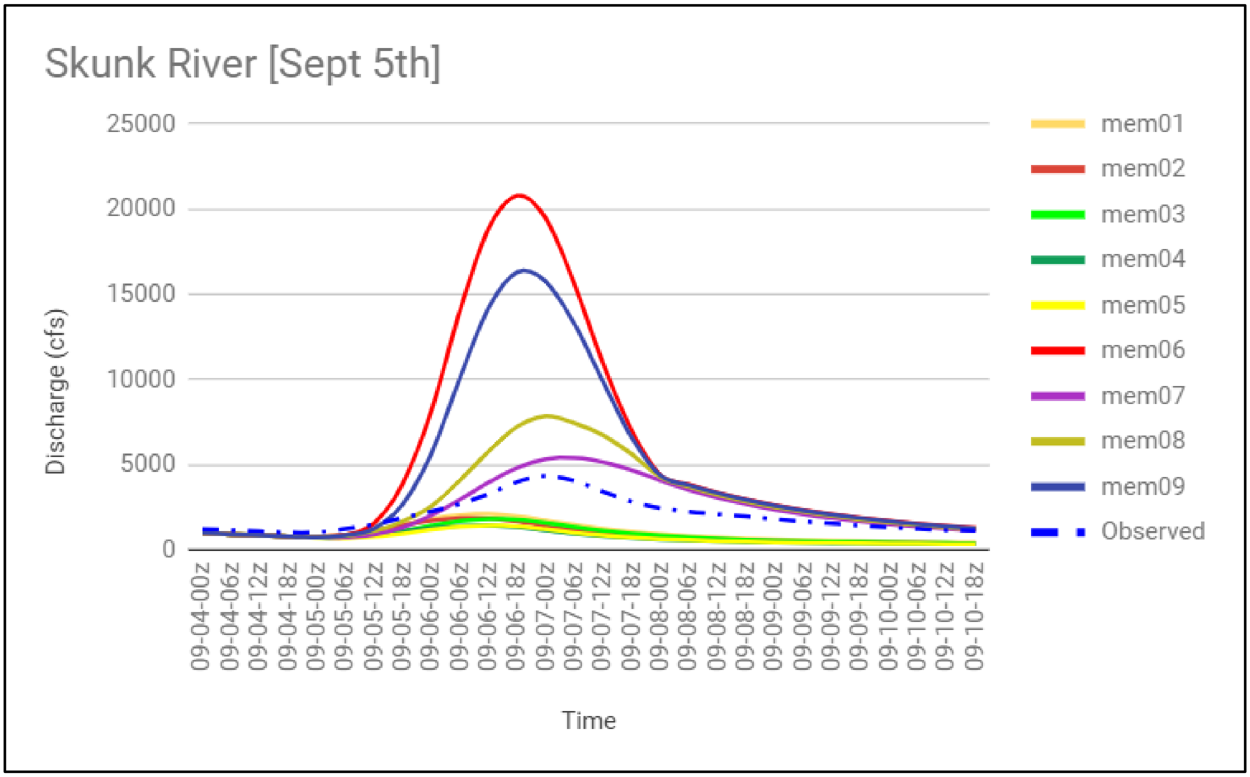


Figure 2: Individual member discharge from the HRRRE as compared to the observed discharge (blue dashed) for the September 5th event where the South Skunk River reach 70% of action stage.

In the analysis, any event that fell within the ensemble spread was contained, while any observed peak that exceeded the ensemble spread was not contained. This allowed for the containing ratio (CR) to be calculated (Equation 6).

$$\frac{Events\ Contained}{Total\ Number\ of\ Events} = CR \quad (6)$$

4. Results

a. Precipitation Biases

Given that ensemble precipitation estimates are at the core of this study, it is essential to investigate those values first. As seen in Figure 3, the HRRRE and HREF showed a wet bias at the beginning of the period with both ensembles being dryer than the MAP dataset during the forecast period from 06z to 18z. The HRRRE had a higher bias in precipitation towards the end of the forecast period, with the highest value of 8.9mm. The average bias for the HRRRE was slightly wet at 2.29mm. The HREF had a similar overall wet bias at 2.21mm, though the bulk of that bias was accumulated in the first forecast period.

b. Peak Event Discharge vs. Ensemble Mean

After running the QPF through the SAC-SMA, it could be seen that the HRRRE ensemble average discharge, which from here on will be shortened to mean HRRRE-Q, overestimated the observed value in four

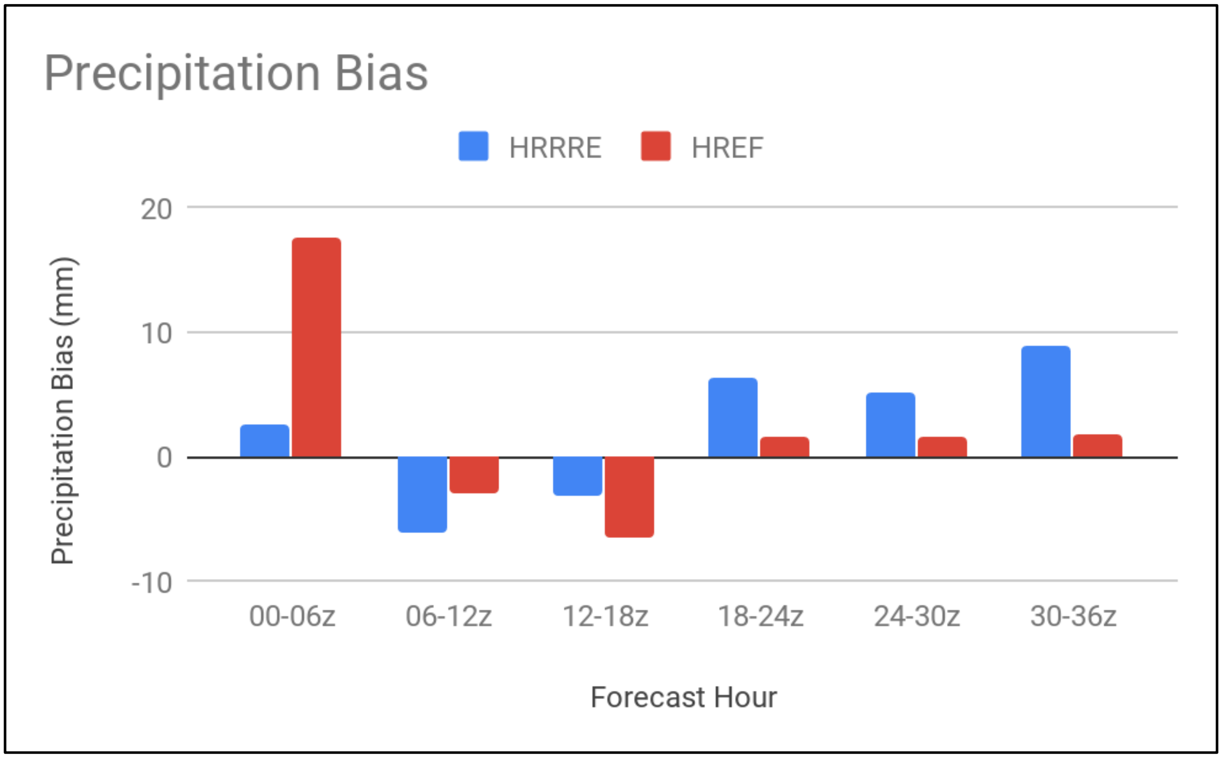


Figure 3: Ensemble precipitation bias (QPF - MAP) averaged for all members and all events at individual time steps.

of the events while underestimating the peak three times (Table 2). The mean HREF discharge (mean HREF-Q) had better performance with three overestimations, two underestimations, while one peak discharge fell within 200 cfs of the observed. This occurred on September 5th for Squaw Creek where it is shown that it reached 92% of the observed. On average the mean HRRRE-Q overestimated the peak by 2096 cfs, while the

Table 2: Ensemble mean peak discharge as compared to observed discharge for all events. This is expressed as the percent of the observed, i.e., 390 would equate to 390% of the observed value. A value greater than 100 represents an overestimation and vice versa for a value less than 100.

Basin	Skunk			Squaw		Turkey	
Event	06/20	07/01	09/05	07/01	09/05	09/02	09/05
HRRRE-Q (%)	390	28	147	27	277	51	179
HREF-Q (%)	228	194	73	138	92	56	46

Table 3: Ranked probability scores based on the peak discharge of each model member. The RPS for each event is shown according to the watershed where it occurred. The average for each ensemble is included in the right column.

Basin	Skunk			Squaw		Turkey		Average
Event	06/20	07/01	09/05	07/01	09/05	09/02	09/05	
HRRRE	0.642	1.605	0.617	1.605	0.667	0.111	0.049	0.757
HREF	0.813	0.422	0.641	0.422	0.656	0.250	0.625	0.547

mean HREF-Q saw a much smaller overestimation of only 330 cfs.

For the timing of the mean HRRRE-Q forecasts showed that it tended to occur just after the observed peak, i.e., less than a time step of difference. Across all events, the mean HREF-Q peak fell between one and two-time steps earlier than observed.

c. *Peak event Flooding vs. Ensemble Members*

i. *Discharge Ranked Probability Score*

When all the events are averaged together, we can see in Table 3 that the HREF discharge (HREF-Q) RPS came out to be 0.21 lower than that of the HRRRE discharge (HRRRE-Q.

ii. *Timing Ranked Probability Score*

When looking at how the timing of peak discharge for the ensemble members matched up with the observed values, we see different results than were seen with peak discharge alone. The HRRRE-Q had a lower score than the HREF-Q by 0.28 as shown in Table 4.

Table 4: Ranked probability scores based on the timing of peak discharge of each model member as compared to the observed. Each event value is shown for the watershed where it occurred. The average for each ensemble is included in the bottom row.

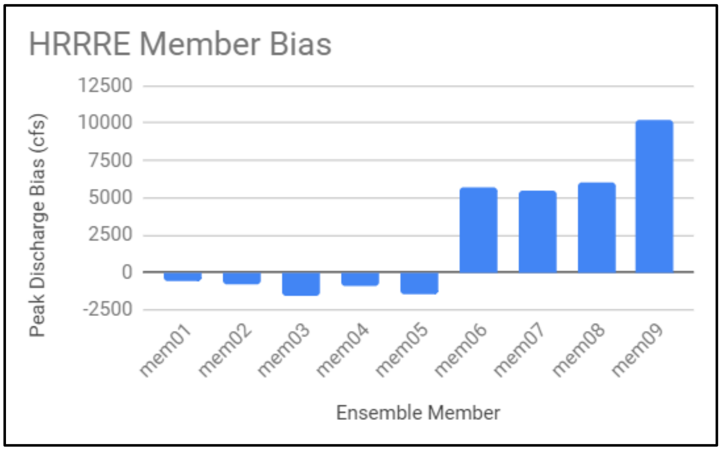
Basin	Skunk			Squaw		Turkey		Average
Event	06/20	07/01	09/05	07/01	09/05	09/02	09/05	
HRRRE	0.148	1.963	0.978	1.667	0.432	1.037	1.000	1.032
HREF	0.297	0.891	2.141	0.656	1.625	0.578	3.000	1.313

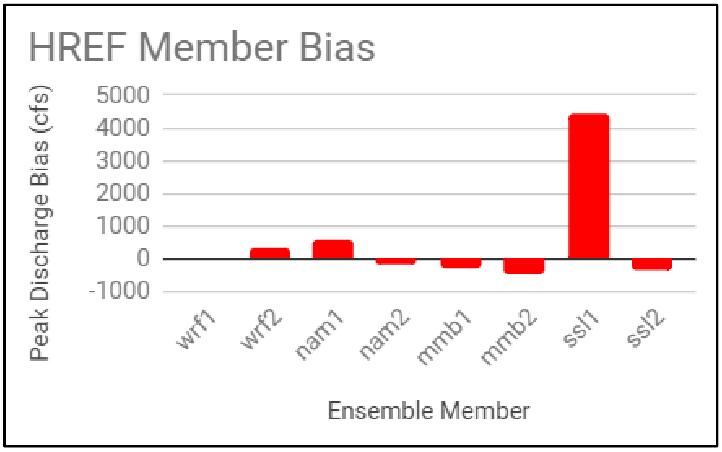
d. Individual Member Performance

It can be seen in Figures 4 and 5 that several members showed consistent dry biases, those members include one through five of the HRRRE. The HREF saw half of its eight members having a dry bias, though to a lesser degree than the HRRRE. However, neither ensemble could overcome an overall wet bias with the HRRRE-Q coming out to be almost 2000 cfs wetter than the HREF-Q. The HRRRE-Q average was heavily skewed by the extreme wet bias of member nine (Table 5).

e. Member Spread

Of the seven events that this study focused on, the HRRRE-Q failed to capture four of them, while the HREF-Q managed to only miss out of two of the peaks. Both of the HREF's shortfalls occurred in the Turkey River basin (Table 6).





Figures 4 and 5: Ensemble wet/dry bias (Member Peak – Observed Peak) computed for individual members across all events and watersheds. Figure 4 is shown for the HRRRE as a blue histogram. Figure 5 is shown for the HREF as a red histogram.

5. Discussion

Both ensembles ended up with a wet bias. This fact supports another one of findings of Seo et al. (2018). The HRRRE QPF struggled the most of all, with a higher average precipitation bias than the HREF QPF, as compared to the MAP dataset provided by the NCRFC. Both models experienced an

Table 5: Average wet/dry bias as compared between individual ensemble member and observed peak discharge (Member Peak - Observed Peak) for all events in a given watershed is shown. The average for both ensembles is shown to the right in the light gray column.

	Skunk Bias (cfs)	Squaw Bias (cfs)	Turkey Bias (cfs)	Model Average
HRRRE	3255.84	1213.70	2486.93	2318.82
HREF	1805.63	624.29	-1527.76	300.72

Table 6: Indication of when the observed peak fell within the ensemble spread of each event. Also included here is the containing ratio (CR) for both ensembles in the rightmost column.								
Basin	Skunk			Squaw		Turkey		CR
Event	06/20	07/01	09/05	07/01	09/05	09/02	09/05	
HRRRE	Yes	No	Yes	No	Yes	No	Yes	57%
HREF	Yes	Yes	Yes	Yes	Yes	No	No	71%

interesting dry bias between hours 6 and 18. Seo et al. (2018) found that errors in QPF only seemed to recede after around 6 hours. Considering this the large wet bias seen for the HRRRE somewhat agree with their findings, though a larger bias is seen in the HREF. This bias in basin average precipitation became apparent when looking at peak discharge magnitudes after data had been run through the SAC-SMA. When the average HRRRE-Q bias was calculated for its members across all events, it came out to be nearly 2000 cfs greater than the HREF-Q. The worst contributor to this was member nine with an average bias of over 10000 cfs.

When looking at the ensemble average peak discharge, the HREF QPF outperformed the HRRRE QPF once again. The mean HREF-Q peak tended to come later than was observed, while the mean HRRRE-Q came closer by being a tad early than the observed peak. In previous research, peak discharge was seen to arrive earlier in the models than observed, leading us to see that the HRRRE operated similarly to those results, though the HREF did not (Seo et al., 2018).

It is essential to look at the bigger picture as the ensemble average only displays a small portion of the value conveyed by the full distribution of members. It is crucial that an ensemble can capture an event within its range of "what if" forecasts. The HREF was superior at containing the observed peak in its member spread by a margin of 71% events to the HRRRE's 57%. This may be due to a greater variety of physics packages being used in the formulation of the HREF (UCAR, 2017). Both ensembles are relatively small with eight and nine members respectively. Therefore, having these relatively low containing ratios is not shocking. As seen by Ebert (2001), no matter the specific inputs, the larger the ensemble, the higher the likelihood of capturing an event.

The RPS of peak discharge across all basins and events would seem to support the HREF's case as the better of the two ensembles for hydrologic forecasting by coming in with a lower score, indicating less deviation from the observed. Only when taking a critical look at the RPS computed for the timing of peak discharge did the HRRRE come out ahead. This time the HRRRE beat out its competitor by 0.28, suggesting that the HRRRE-Q for individual members, much like the mean HRRRE-Q, had greater ease in predicting the timing of the flood peak.

One of the limiting factors in this study came from the low number of cases. Attempts to use flood stage and action stage as a benchmark for event selection yielded a disappointingly small number of usable cases. At first, 70% of action stage method produced plenty of usable cases for each watershed. Only later was it realized that an issue in the HRRRE runs dealing with bad data being stored in the QPF grids further reduce the number of cases. Any events, before June 20th, 2018, were unusable due to the model not running in the proper configuration. It is apparent that a greater sample size would improve a study like this and open the door to more rigorous statistical testing.

As a caveat, the SAC-SMA in the lumped configuration used here is limited to using six-hour time steps. Both ensembles, however, produce hourly data. Thus, some detail may have been lost in averaging and may have affected the results. This would have likely smoothed out the data because differences were collected between each six-hour time step rather than going hour by hour.

6. Conclusions

The objective of this study was to explore flash flood forecasting by looking at a comparison of the experimental HRRRE and the operationally used HREF through streamflow modeling of the ensemble QPF employing the lumped SAC-SMA over three Iowa watersheds. This was achieved after isolating six-hour QPF from the available model runs, taking basin averages for each time step and, running the SAC-SMA with a lengthy spin up followed by the previously mentioned QPF values.

The data produced by this study would suggest that the HREF QPF provides more valuable forecasts than the HRRRE QPF when looking at flood peak discharge. The HRRRE underperformed in every metric besides the prediction of the timing of peak discharge. This would lead us to believe that improvements may need to be made to increase the viability of the HRRRE as an operationally used ensemble for hydrologic forecasting.

For the HRRRE, issues with documentation of model configuration and complications within the model corrupting fields of QPF data contributed to a small sample size of useful cases and caused headaches during data processing. These are problems to be expected for an experimental model. With time, these complications should fade, and thus this same topic may be revisited, hopefully with a more substantial amount of usable cases. Furthermore, as the HRRRE is developed, more members may be added.

In the future, it would serve to add more watersheds to this study. Along those same lines, using the same dataset as investigated here, but using a different hydrologic model could serve as a healthy comparison to these results. For instance, using a distributed model may allow the HRRRE to show more skill in its forecasts due to its slightly finer 3km nested grid rather than just looking at basin averages like that seen in the lumped SAC-SMA.

7. Acknowledgements

The author thanks Dr. Franz for her advising and guidance on this project. Special thanks to Brian Connelly from the NCRFC for his help in securing the HREF, MAP, MAT, and PET datasets. Moreover, this work would not have been completed on time if it wasn't for the continued support of Andrew Goenner. He had several Matlab scripts that acted as templates for the preprocessing. Andrew also contributed heavily to the code trouble shooting. Lastly, thanks are in order for Benjamin Kiel for his help in acquiring the HRRRE dataset.

8. References

Barthold, F.E., T.E. Workoff, B.A. Cosgrove, J.J. Gourley, D.R. Novak, and K.M. Mahoney, 2015: Improving Flash Flood Forecasts: The HMT-WPC Flash Flood and Intense Rainfall Experiment. *Bull. Amer. Meteor. Soc.*, **96**, 1859–1866, https://doi.org/10.1175/BAMS-D-14-00201.1

Bowman, A. L., K. J. Franz, and T. S. Hogue, 2017: Case Studies of a MODIS-Based Potential Evapotranspiration Input to the Sacramento Soil Moisture Accounting Model. J. Hydrometeor., 18, 151–158, https://doi.org/10.1175/JHM-D-16-0214.1

Burnash, R. J. C., 1995: The National Weather Service River Forecast System—Catchment modeling. Computer Models of Watershed Hydrology, V. P. Singh, Ed., Water Resources Publications, 311–366.

Burnash, R. J. C., R. L. Ferral, and R. A. McGuire, 1973: A generalized streamflow simulation system: Conceptual models for digital computers. Joint Federal and State River Forecast Center, U.S. National Weather Service, and California Department of Water Resources Tech. Rep., 204.

Dowell, D., C. Alexander, T. Alcott, T. Ludwig, 2018: HRRR Ensemble (HRRRE) Guidance 2018 HWT Spring Experiment. ESRL/GSD. Accessed 19 September 2018, https://rapidrefresh.noaa.gov/internal/pdfs/2018_Spring_Experiment_HRRRE_Documentation.pdf

Du, J., S.L. Mullen, and F. Sanders, 1997: Short-Range Ensemble Forecasting of Quantitative Precipitation. *Mon. Wea. Rev.*, **125**, 2427–2459, https://doi.org/10.1175/1520-0493(1997)125<2427:SREFOQ>2.0.CO;2

Ebert, E.E., 2001: Ability of a Poor Man's Ensemble to Predict the Probability and Distribution of Precipitation. *Mon. Wea. Rev.*, **129**, 2461–2480, https://doi.org/10.1175/1520-0493(2001)129<2461:AOAPMS>2.0.CO;2

Franz, K. J., and T. S. Hogue, 2011: Evaluating uncertainty estimates in hydrologic models: borrowing measures from the forecast verification community. Hydrology and Earth System Sciences, 15, 3367–3382, https://doi.org/10.5194/hess-15-3367-2011

Franz, K. J., H. C. Hartmann, S. Sorooshian, and R. Bales, 2003: Verification of National Weather Service Ensemble Streamflow Predictions for Water Supply Forecasting in the Colorado

River Basin. J. Hydrometeor., 4, 1105–1118, https://doi.org/10.1175/1525-7541(2003)004<1105:VONWSE>2.0.CO;2

Gallus, W. A., Jr., 2012: The Challenge of Warm-Season Convective Precipitation Forecasting. Rainfall Forecasting, Nova Science Publishers, 129-160, ISBN 978-61942-134-9

Gilmour, I., L.A. Smith, and R. Buizza, 2001: Linear Regime Duration: Is 24 Hours a Long Time in Synoptic Weather Forecasting?. *J. Atmos. Sci.*, **58**, 3525–3539, https://doi.org/10.1175/1520-0469(2001)058<3525:LRDIHA>2.0.CO;2

NOAA, 2018: Flood Deaths and Direct Damages by State Water Year 2017 (October 1, 2016 - September 30, 2017). Accessed 9 September 2018, http://www.nws.noaa.gov/os/water/Flood%20Loss%20Reports/WY17%20Flood%20Deaths%20and%20Direct%20Damagesv2.pdf

Seo, B., F. Quintero, and W.F. Krajewski, 2018: High-Resolution QPF Uncertainty and Its Implications for Flood Prediction: A Case Study for the Eastern Iowa Flood of 2016.*J. Hydrometeor.*, **19**, 1289–1304, https://doi.org/10.1175/JHM-D-18-0046.1

Spies, R.R., K.J. Franz, T.S. Hogue, and A.L. Bowman, 2015: Distributed Hydrologic Modeling Using Satellite-Derived Potential Evapotranspiration. *J. Hydrometeor.*, **16**,129–146, https://doi.org/10.1175/JHM-D-14-0047.1

UCAR, 2017: Operational Models Encyclopedia, Accessed 19 September 2018, https://sites.google.com/a/ucar.edu/model-encyclo-prob/probabilistic/href/initial-conditions