1A.4 VALIDATION OF NWS HYDROLOGIC ENSEMBLE FORECAST SERVICE (HEFS) REAL-TIME PRODUCTS AT THE MIDDLE ATLANTIC RIVER FORECAST CENTER

Seann Reed*[1], Alaina MacFarlane[1]

[1]Middle Atlantic River Forecast Center, Eastern Region
National Weather Service, NOAA, State College, Pennsylvania

1. INTRODUCTION

National Weather Service (NWS) river forecast products provide a variety of users with information they need to make decisions. Users include the general public, emergency managers, water resources managers, hydroelectric power plant operators, and others. Many users like probabilistic information to help in decision making. Therefore, with support from the Office of Water Prediction (OWP), NWS River Forecast Centers (RFCs) continue working to improve our ensemble-based probabilistic forecasts.

RFCs began ensemble forecasting in the 1980s and 90s using Extended Streamflow Prediction (ESP) techniques (Day, 1985). ESP accounts for variations in watershed initial conditions (e.g. soil moisture and snow-pack) and the range of possible input forcings (precipitation and temperature) in the historical record (also referred to as climatology in this paper). ESP was designed primarily for 30 to 90 day water supply forecasts.

In 2012, the Meteorological Model Ensemble Forecast System (MMEFS) was implemented at four RFCs in the Eastern U.S. (Adams and Ostrowski, 2010). MMEFS provides short-term forecasts up to seven days. Unlike ESP, MMEFS uses ensemble output from meteorological model forecasts as input to hydrologic models. Despite often large uncertainties in precipitation forecasts and model precipitation conditional bias, MMEFS products provide a valuable heads-up for potential flooding in a region, up to seven days in advance. RFCs have received positive feedback from both emergency managers and water resources managers indicating that MMEFS hydrologic ensembles are useful for decision making.

More recently the Hydrologic Ensemble Forecast Service (HEFS) (Demargne et al., 2014) was developed to further improve river ensemble forecasting. HEFS offers several advantages compared to MMEFS: (1) the ability to bias-correct and downscale temperature and precipitation forecasts from meteorological models, (2) the ability to temporally merge these processed forcings from multiple sources (e.g. deterministic QPF, GEFS, CFSv2, and climatology) producing seamless short- to long-term forecasts, and (3) the ability to account for hydrologic model uncertainty in the final ensemble generation.

The HEFS module that performs tasks 1 and 2 is referred to as the Meteorological Ensemble Forecast Preprocessor (MEFP) and the module that performs task 3 is the Ensemble Post-Processor (EnsPost).

After several years of testing and running HEFS for selected watersheds in the Delaware River basin, MARFC began widespread production of HEFS products for public display on the NWS Advanced Hydrologic Prediction Service web site (https://water.weather.gov/ahps/) in 2017. Figures 1a and 1b are example HEFS graphics that are available for 117 of our forecast points. Figure 1a is the short-term forecast guidance product for 1 - 10 days. This is a new graphic style that was developed for HEFS (Carr et al., 2018). Figure 1b is a 30-day exceedance probability graph geared towards water supply managers that has been produced for many years using data from the ESP technique. Now, HEFS forecasts are used to make the exceedance probability graphs, integrating the intelligence of meteorological model information into this product.

---

[1]*Corresponding author address: Seann Reed, 328 Innovation Blvd., State College, PA 16803; e-mail: seann.reed@noaa.gov.

MMEFS and HEFS Implementation Details

The RFC MMEFS implementations include three separate sets of forecasts. MMEFS forecasts are created using the Global Ensemble Forecast System (GEFS), the North American Ensemble Forecast System (NAEFS), and the Short-Range Ensemble Forecast (SREF) system (See Operational Models Encyclopedia, 2019 for information on these systems). MMEFS simply runs the output forcings from these meteorological ensemble systems (GEFS, NAEFS, and SREF) through our hydrologic models.

The current MARFC HEFS-MEFP implementation uses the Global Ensemble Forecast System (GEFS) forcings for days 1 - 15 and resampled climatology forcings for days 16 - 30. We are not currently using CFSv2 because previous work has shown that CFSv2 adds little to no skill in our region above the MEFP resampled climatology (Brown, 2013).

As described by Demargne et al. 2014, HEFS-MEFP uses the mean of the input ensemble or input from a deterministic model (mean of the GEFS in our case), statistical parameters from historical analysis, and the Schaake Shuffle method to reconstruct plausible ensemble members (Schaake et al. 2007). A multi-year re-forecast dataset is required to estimate parameters for MEFP. For this reason, our current HEFS-MEFP implementation uses GEFS v.10 instead of the GEFS v.11 (the current operational version); the necessary re-forecast data are not available for GEFS v.11. We will begin using GEFS v.12 for HEFS in 2020, skipping over v. 11 for logistical reasons. MMEFS currently uses GEFS v. 11 because it does not require reforecasts.

We currently do not run the EnsPost component of HEFS for our short-term forecasts, i.e., we are not accounting for the hydrologic model uncertainty. We do not use EnsPost because EnsPost automatic adjustments can conflict with our manual operational model state corrections.

Goal

While HEFS offers potential advantages over MMEFS, there are enough differences in the algorithms and performance assessments to date that we have decided to keep both running in operations. This paper describes our first comprehensive evaluation of HEFS real-time products with an emphasis on comparison to the MMEFS methods. The goal is to summarize information about the accuracy of these products that can aid the decision making of our forecasters and external users. This verification information will also help us inform efforts at the NWS Office of Water Prediction (OWP) to improve HEFS, which is intended to be the standard tool for ensemble river forecasting moving forward.

Prior Work

Earlier MMEFS verification work by Philpott et al., 2012 focused on SREF output and showed that conditional biases in the precipitation forcing tended to produce over-forecasting of streamflow for events below flood stage and under-forecasting for events above flood stage.

Prior to HEFS implementation in each RFC, HEFS baseline validation is completed across each RFC region to ensure that HEFS consistently improves upon ESP (Lee et al., 2018). While valuable for identifying implementation or data problems and for understanding regional differences, improvement over ESP is a fairly low bar, as we know that GEFS should be able to predict the occurrence of rain events in the next 15 days better than climatology.

Baseline validation was done using hindcasting. In hindcasting, we first run observed historical forcings through the hydrologic models to generate model states for each day. Then forecasts are initiated each day from the appropriate hydrologic states and GEFS v.10 re-forecast data (Hamill et al., 2013) for that day.

The hindcasting methodology offers the benefit of being able to analyze many years of data with the same methodology; however, there is a key difference between these hindcasts and our operational forecasts — the soil moisture and snow states in our operational forecasts are updated to reflect current watershed conditions while the hindcasting algorithms we have available to us do not include state updating. We show an example of the impact of these initial state differences in the results section.

Instead of using hindcasting, this paper focuses on validation of archived MMEFS and HEFS data from real-time runs. Although this approach limits the length of the validation period, it provides information about the errors in our actual forecasts.

While our baseline HEFS forecasts go out 30 days, the analysis here focuses mostly on forecasts out 7 days to compare HEFS to MMEFS. We have both emergency management and water resource management partners who are interested in high, medium, and low flows in this seven day forecast window.

2. METHODOLOGY

In comparing HEFS to MMEFS, we choose to focus our comparisons on the 0 UTC MMEFS-NAEFS runs. We do not compare HEFS to MMEFS-SREF here because the SREF only provides forecasts out 3 days.

We choose the MMEFS-NAEFS over the MMEFS-GEFS because it tends to better represent possible spread. The NAEFS has 42 members and contains two underlying models, while the GEFS only has 21 members. In fact, the NAEFS 0 UTC and 12 UTC runs actually include 21 members of the GEFS and 21 members of the Canadian Ensemble Forecast. We choose the 0 UTC NAEFS because the HEFS is currently created once per day from the 0 UTC GEFS v.10 members.

One might argue that we give the MMEFS methodology an unfair advantage by choosing the NAEFS (which has more members than the GEFS and includes GEFS v. 11); however, the goal here is not to argue the scientific merits of one method over another. The goal is to develop useful guidance for users of currently available products. For this purpose we want to compare the 'best' MMEFS product available on a daily basis to the daily HEFS product.

For our statistical analysis, we use different methods to assess performance at high and low flows. Our analysis period is from January 1, 2017 to September 30, 2019. This period is selected because of the availability of concurrent archives for MMEFS and HEFS.

<u>High Flows:</u>

For high flows exceeding either a caution stage (also known as Action stage in NWS parlance) or a flood stage at a location, we look at contingency statistics such as Probability of Detection (POD), False Alarm Ratio (FAR), and Critical Success Index (CSI) (Jolliffe and Stephenson, 2012).

Contingency	Obs Y	Obs N
Fcst Y	A	B
Fcst N	C	D

POD = A/(A+C)
FAR = B/(A+B)
CSI = A/(A+B+C)

Observed instantaneous stage data for validation were downloaded from the USGS National Water Information System (https://waterdata.usgs.gov/nwis) for 103 stations. Data marked by the USGS as potentially impacted by river ice were eliminated from the analysis. Because we have a relatively short 2.75 year analysis period, we pool contingency tables for all stations in the analysis to maximize the sample size. The fact that 2018 into early 2019 was a wet period for our region also helps to increase our sample size.

We also pool forecasts with 2 to 6 day lead times for contingency table analysis. On each day of our analysis period we are assessing the ability of MMEFS-NAEFS and HEFS to predict whether or not caution stage or flood stage will be exceeded some time in the next 2 to 6 days.

In addition to contingency statistics, we qualitatively examine daily summary maps of HEFS and MMEFS forecasts across the region, which leads to conclusions consistent with the contingency table summaries.

To provide additional context on the performance of these probabilistic forecasts, we also compare the HEFS performance to our deterministic forecasts for days 1 - 3 in a few basins. For this comparison, we select Harper Tavern, PA and

Martinsburg, PA, where a remarkable number of distinct flood events occurred during the 2.75 year analysis period, 14 and 20 respectively. We look at the mean absolute error (MAE) of forecasts and observed pairs when the observations were above minor flooding.

Low and Medium Flows

For low flow assessment, we examine mean absolute error (MAE) as a function of lead time for two large basins in our region where we have interested partners concerned about water supply and hydroelectric power generation decisions — the Susquehanna River at Marietta, PA, and the Potomac River at Little Falls, MD.

3. RESULTS AND DISCUSSION

Initial Condition Impacts: Real-time vs. Hindcasts

As mentioned in the introduction, we focus our analysis on real-time HEFS products rather than HEFS hindcasts. The main difference in these two products is that our HEFS hindcasts do not currently use data assimilation to correct for initial conditions. Figure 2 shows an example of how much difference this makes for the Harper Tavern, PA, headwater basin. The plot shows relative mean error for daily real-time HEFS forecasts and HEFS hindcasts over the same analysis period. Statistics are analyzed for high flows (all flows exceeding the 90th percentile, Q(Pr=0.90)) and low flows (all flows exceeding 10th percentile, Q(Pr=0.1)). The real-time model state adjustments inserted by forecasters do improve the relative mean error statistics in this basin. For low flows, the improvement is about 11% at a 42-hour lead-time and drops to about 5% at a 30-day lead-time. For high flows, the improvement is about 8% at the 42-hour lead-time and drops to 1% for 30-day lead times. Overall, both sets of HEFS forecasts considerably underestimate flows, particularly for high flows. This is confirmed below for a broader set of basins.

High Flows

For the contingency statistics pooled over 103 locations, there were about 3600 'events' exceeding caution stage and 730 'events' exceeding minor flood stage (A+C in the tables). Because we tallied events daily using a sliding 4-day forecast window, this is roughly 4 times the number of caution and minor level exceedances in the observed record.

Figure 3 shows CSI, POD, and FAR results for caution stage. As shown in the bar charts, these statistics are tallied for different warning thresholds. A 0.1 warning threshold on the x-axis assumes that the ensemble forecasts predict action stage if only 10% of the ensemble members exceed action stage. Thus, more conservative forecasts are on the left of these charts and have higher PODs and higher FARs. For caution stage at warning thresholds of 0.1 and 0.3, HEFS and MMEFS CSI values are very similar, with HEFS having a slight edge. For thresholds of 0.5, 0.7, and 0.9, the MMEFS-NAEFS produces higher CSI.

Figure 4 shows very similar results for minor flooding. HEFS scores a higher CSI at the 0.1 threshold, but MMEFS-NAEFS scores higher for the 0.3, 0.5, 0.7, and 0.9 thresholds. Comparing Figures 3 and 4, both models show considerably higher CSI values for predicting caution stage compared to minor flood stage (max CSI values of 0.29 compared with ~ 0.1). This reflects the higher difficulty in predicting less frequent events. The highest CSI values of about 0.1 for minor flood prediction occur for HEFS at the 0.1 warning threshold and for MMEFS-NAEFS at the 0.3 and 0.5 warning thresholds.

Figures 3 and 4 also highlight the fact that the MMEFS-NAEFS is more bullish on flooding. MMEFS-NAEFS consistently produces a higher POD, but also a higher FAR than HEFS. At the lower warning thresholds, these tendencies balance out in yielding relatively similar CSI values.

Figures 5 and 6 show CSI results stratified by warm season (April - September) and cool season (October - March). In Figure 6 we can see that for minor flooding HEFS can achieve the highest CSI in the cold season at the 0.3 warning threshold (the lower of the two thresholds used in our map displays, e.g. Figure 7). In the warm season, MMEFS-NAEFS outperforms HEFS in terms of CSI at the 0.3 threshold but HEFS does better at the 0.1 threshold.

Since 2012 we have served maps on the web similar to those in Figure 7 to assess the MMEFS output in a region. Now that we have HEFS running at most of the same locations, we will soon provide similar maps for HEFS output.

Figure 7 shows January 22, 2019, forecasts for a typical cool season event. Both MMEFS-NAEFS (upper left) and HEFS (upper right) indicate flood potential along the East Coast in the Richmond to Washington to Philadelphia to New York corridor. The third panel in Figure 7 shows the forecast points where flooding actually occurred: Walton, NY; Neversink Reservoir, NY; Boonton, NJ; Pine Brook, NJ; Norristown, PA; Camp Hill, PA; and Harper Tavern, PA. Consistent with our statistical results, NAEFS tends to over-predict the amount of flooding and HEFS tends to under-predict. For the January 22 forecast, NAEFS had 5 correct and 22 false alarms while the HEFS had 1 correct and 3 false alarms. The flooding occurred between late afternoon on Jan 24 and noon on Jan 25. Thus, the ensemble forecasts were providing this level of skill with about a 2.5 day lead time.

In the ensemble map displays like Figure 7, circles are used to represent a 30% exceedance probability and squares are used to represent a 70% chance. Our statistical results suggest that paying attention to the 30% threshold is a good choice to achieve relatively high CSI values from both MMEFS-NAEFS and HEFS. We also see that for predicting minor flooding, MMEFS-NAEFS would be the model of choice in the warm season and HEFS in the cool season at the 30% threshold (Figure 6). Because the underlying models in both MMEFS-NAEFS and HEFS are not convective allowing, it makes sense that the more bullish NAEFS performs relatively well in the warm season.

Most of our development and implementation focus in the near-term will be on improving HEFS. We will be upgrading HEFS to use GEFS v.12 in 2020 when GEFS v.12 hindcasts become available. We also plan to enhance our HEFS implementation so that it runs four times per day instead of just once. Use of additional forcing sources to improve short-term prediction is also a possibility within the flexible HEFS software framework. Lastly, the Office of Water Prediction is planning some HEFS algorithm improvements, including a method to improve bias correction for high flows.

A challenge that remains in providing the best possible HEFS forecasts is to make these forecasts fully consistent with our deterministic forecasts. This is currently not the case because our deterministic forecasts are updated more frequently, meaning they can leverage more recent precipitation observations, more recently updated model forecasts, higher resolution model forecasts, and the latest forecaster modifications, including in-between synoptic times. Because of these factors, our deterministic forecasts are likely to be more accurate than the mean of the HEFS forecasts in the short-term, particularly for high flow events. Figure 8 shows that this was indeed the case for days 1-3 at Harper Tavern, PA, and Martinsburg, WV, during our 2.75 year analysis period. Figure 8 shows mean absolute error for only the forecast-observed pairs when the observed stage exceeded flood stage. Results for these two basins are shown because they experienced an unusually high number of floods, 20 for Martinsburg and 14 for Harper Tavern during this period.

Low and Medium Flows

Figures 9 and 10 show MAE for the Susquehanna River at Marietta, PA (26,000 mi^2), and the Potomac River at Little Falls, MD (11,560 mi^2). The left panels show statistics for higher flows, considering only flows exceeding the 90th and 95th percentiles, while the right panels show statistics for a larger range of flows, considering flows exceeding the 50th (medium) and 10th (low) percentiles. At Marietta, the MMEFS-NAEFS has lower errors than HEFS for high flows. At low and medium flows, MMEFS-NAEFS and HEFS have very similar MAE for days 1-3 but HEFS has lower errors on days 3-7. Results are similar at Little Falls, except HEFS performance is similar to MMEFS-NAEFS for flows above the 90th percentile.

4. SUMMARY AND CONCLUSIONS

The NWS Middle Atlantic River Forecast Center has been running the Hydrologic Ensemble Forecast Service (HEFS) for 117 of our forecast points since 2017. HEFS has features, e.g.

forcings bias correction, which can improve upon the Meteorological Model Ensemble Forecast System (MMEFS), which has been running since 2012. However, the HEFS implementation does not always outperform MMEFS and is still being improved. We currently run both systems in parallel to leverage the benefits of each. The analyses reported here were initiated to improve our understanding of HEFS and MMEFS performance and provide some quantitative guidance on how and when the forecasts from each system should be used.

We first analyzed ensemble forecasts and observed stage data at 103 river gages for the period January 1, 2017 to September 30, 2019. We looked at the ability of these models to predict the occurrence of either caution stage or flood stage exceedance in a 2-6 day forecast window by computing POD, FAR, and CSI.

This analysis shows that MMEFS-NAEFS has higher PODs but also higher FARs compared to HEFS, regardless of the warning probability threshold selected. MMEFS-NAEFS and HEFS were not too far apart in the level of CSI that could be achieved when an optimal warning threshold is selected; a lower warning threshold needs to be selected for HEFS. If we look at a minor flood warning threshold of 0.3, meaning 30% of the ensemble members are indicating flooding, HEFS has a small CSI edge for cool season events while MMEFS-NAEFS maintains a small edge in the warm season.

Both models attained substantially higher CSI values when predicting caution stage compared with predicting minor flooding. The best CSI values are close to 0.3 for caution stage but only slightly exceeding 0.1 for minor flood stage. This highlights the challenge of predicting flooding in small basins with lead times of 2 days or more.

While it is challenging to predict flood levels in specific, small basins two days in advance, these probabilistic forecasts provide useful information in developing our 5-Day Flood Outlook Products and giving our partners actionable information at a sub-regional level for time horizons beyond our 3-day deterministic forecasts.

For two locations with frequent flooding in the past three years, we compared our deterministic flood forecasts to the ensemble mean from HEFS in the first three days. The deterministic flood forecasts have lower MAE. This is likely because we are able to update these forecasts more often, with more recent precipitation observations, more recent precipitation forecasts and more recent streamflow observations. Thus, we still recommend using our deterministic forecasts for days 1-3 and ensemble forecasts beyond day 3. With support from the Office of Water Prediction, we are planning improvements to our HEFS implementations over the next several years that will begin to make our deterministic and ensemble flood forecasts more consistent.

We verified in two large basins of interest for water supply applications that HEFS products do outperform MMEFS-NAEFS products for low and medium flows in a 7-day forecast window. Although not reported here, previous work also showed that 30-day water supply products based on HEFS are more skillful than our traditional ESP products.

ACKNOWLEDGEMENTS

The Office of Water Prediction Hydrologic Ensemble Forecast Development team has provided substantial support during the last several years to implement and verify HEFS.

REFERENCES

Adams, T.E., and J. Ostrowski, May 2010: Short Lead-Time Hydrologic Ensemble Forecasts from Numerical Weather Prediction Model Ensembles, World Environmental and Water Resources Congress, https://doi.org/10.1061/41114(371)237.

Brown, J.D. 2013: Verification of long-range temperature, precipitation and streamflow forecasts from the Hydrologic Ensemble Forecast Service (HEFS) of the U.S. National Weather Service. Technical Report prepared by Hydrologic Solutions Limited for the U.S. National Weather Service, Office of Hydrologic Development, 128pp.

Carr, R.H., B. Montz, K. Semmens, K. Maxfield, S. Connolly, P. Ahnert, R. Shedd, and J. Elliott, 2018: Major Risks, Uncertain Outcomes: Making

Ensemble Forecasts Work for Multiple Audiences. *Wea. Forecasting,* **33**, 1359–1373, https://doi.org/10.1175/WAF-D-18-0018.1

Day, G. N., 1985: Extended streamflow forecasting using NWSRFS, J. of *Water Resources Planning and Management*, 111, 157-170.

Demargne, J., L. Wu, S.K. Regonda, J.D. Brown, H. Lee, M. He, D.J. Seo, R. Hartman, H.D. Her, M. Fresch, J. Schaake, and Y. Zhu, Jan 2014: The Science of NOAA's Operational Hydrologic Ensemble Forecast Service, , *Bull. Amer. Meteor. Soc.,* **95**, 79–98, https://doi.org/10.1175/BAMS-D-12-00081.1

Global Ensemble Forecast System, https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-ensemble-forecast-system-gefs

Hamill, T.M., Bates, G.T., Whitaker, J. S., Murray, D.R., Fiorino, M., Galarneau Jr., T., Zhu, Y., and Lapenta, W. October 2013: NOAA's second-generation global medium-range ensemble reforecast data set. . *Bull. Amer. Meteor. Soc.,* **94**, 1553–1565, https://doi.org/10.1175/BAMS-D-12-00014.1

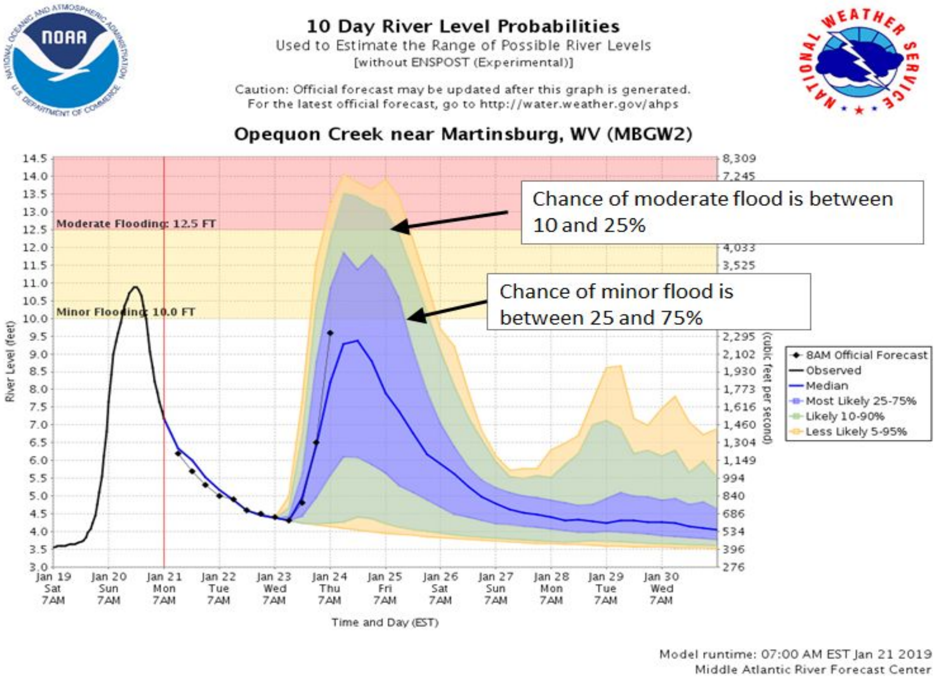
Jolliffe, I.T., and Stephenson, D.B, Editors, 2012: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, John Wiley & Sons Ltd.

Lee, H. S., Liu, Y., Brown, J., Ward, J., Maestre, A., Fresch, M. A., Herr, H., and Wells, E., 2018: Validation of Ensemble Streamflow Forecasts from the Hydrologic Ensemble Forecast Service (HEFS), AGU Fall Meeting, https://ui.adsabs.harvard.edu/abs/2018AGUFM.H31A..05L.

Operational Models Encyclopedia - Probabilistic Models, 11/8/2019: https://sites.google.com/a/ucar.edu/model-encyclo-prob/probabilistic

Philpott, A. W., P. Wnek, and J. D. Brown, 2012: Verification of ensemble river forecasts at the Middle Atlantic River Forecast Center. [Available online at https://ams.confex.com/ams/92Annual/webprogram/Paper199532.html.]

Schaake, J, J. Demargne, R. Hartman, M. Mullusky, E. Welles, L. Wu, H. Herr, X. Fan, and D.J. Seo, 2007, Precipitation and temperature ensemble forecasts from single-value forecasts, Hydrology and Earth System Sciences, https://doi.org/10.5194/hessd-4-655-2007.



10 Day River Level Probabilities
Used to Estimate the Range of Possible River Levels
[without ENSPOST (Experimental)]

Caution: Official forecast may be updated after this graph is generated.
For the latest official forecast, go to http://water.weather.gov/ahps

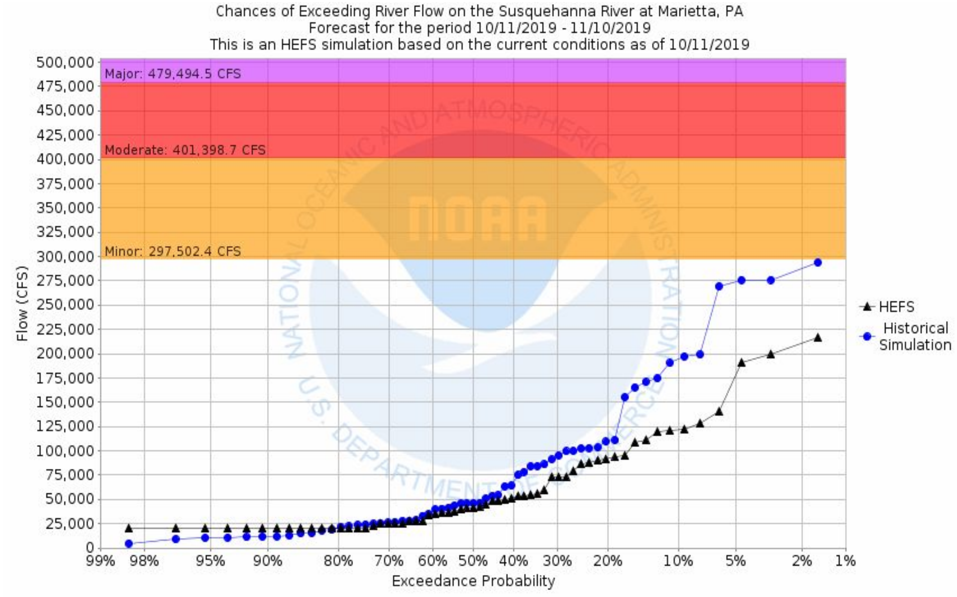
Opequon Creek near Martinsburg, WV (MBGW2)

Chance of moderate flood is between 10 and 25%

Chance of minor flood is between 25 and 75%

Model runtime: 07:00 AM EST Jan 21 2019
Middle Atlantic River Forecast Center

Figure 1a. HEFS Short-term Probabilistic Guidance Product.



Chances of Exceeding River Flow on the Susquehanna River at Marietta, PA
Forecast for the period 10/11/2019 - 11/10/2019
This is an HEFS simulation based on the current conditions as of 10/11/2019

Figure 1b. HEFS 30-day Flow Exceedance Probabilities.

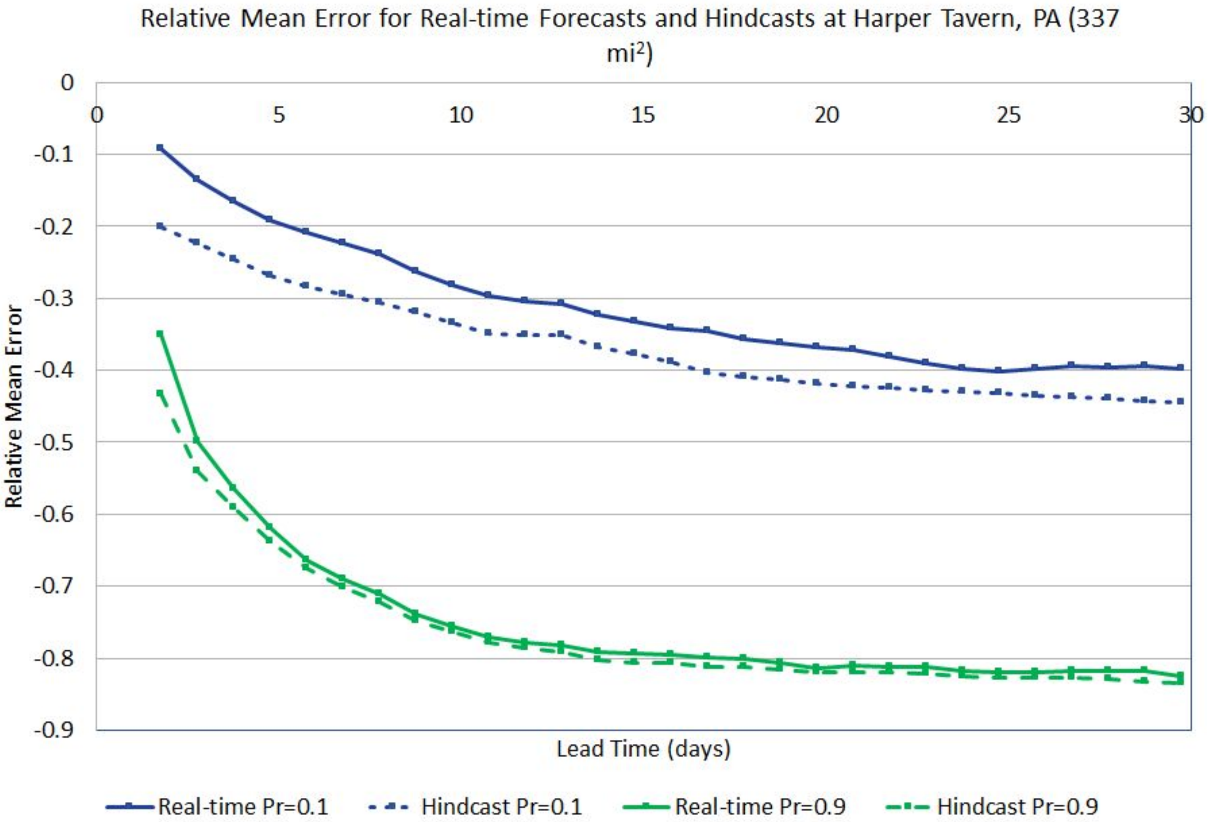


Figure 2. Relative Mean Error of HEFS daily real-time forecasts compared to daily hindcasts for Harper Tavern, PA (337 mi2), January 1, 2017 - September 30, 2019. Blue lines consider all flows greater than Q(Pr=0.1) and green lines consider all flows greater than Q(Pr=0.9) ('high flows').

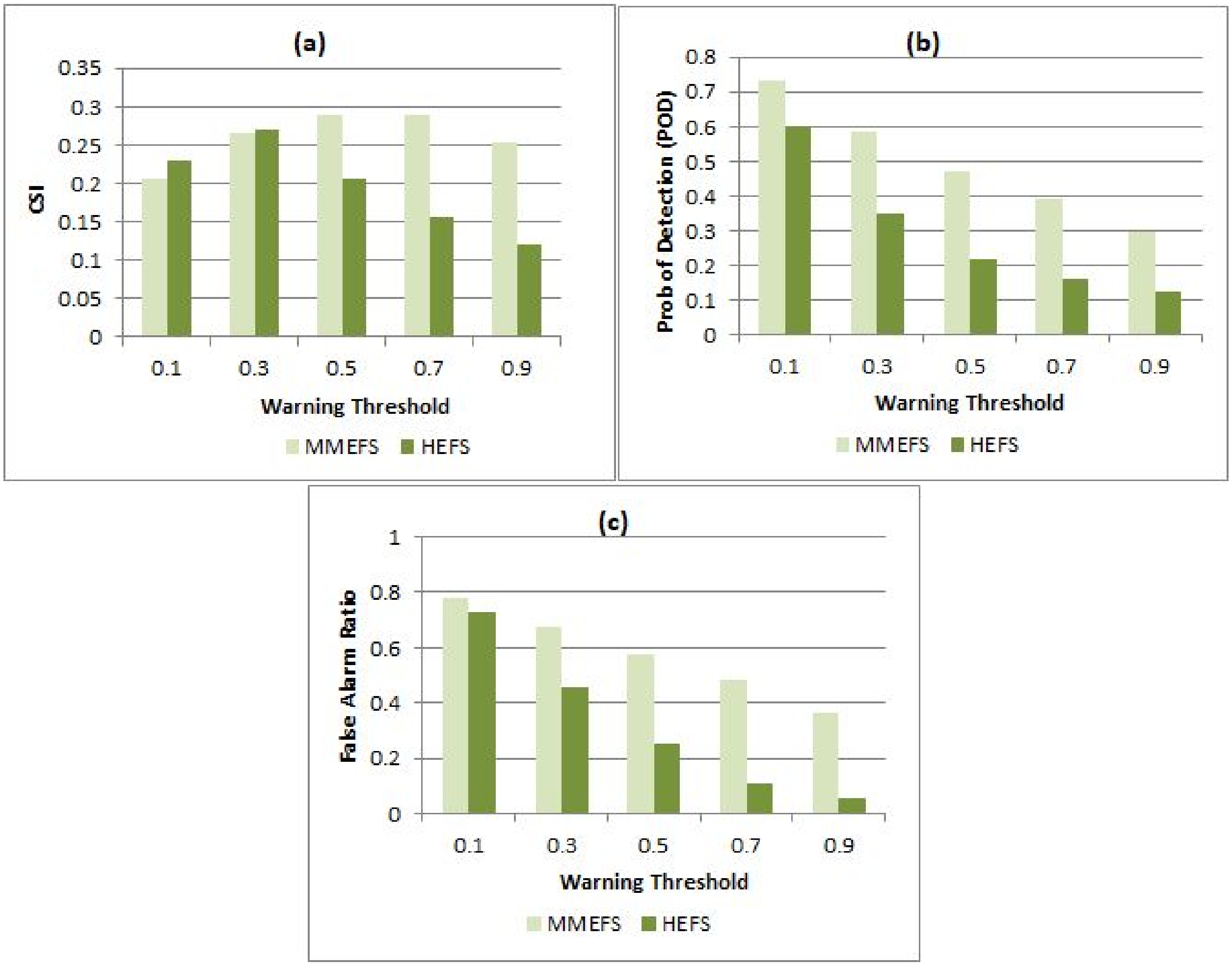


Figure 3. Caution stage statistics for the 2 - 6 day lead time. Pooled results for 103 validation locations over 2.75 years. a) is CSI, b) is POD, c) is FAR.

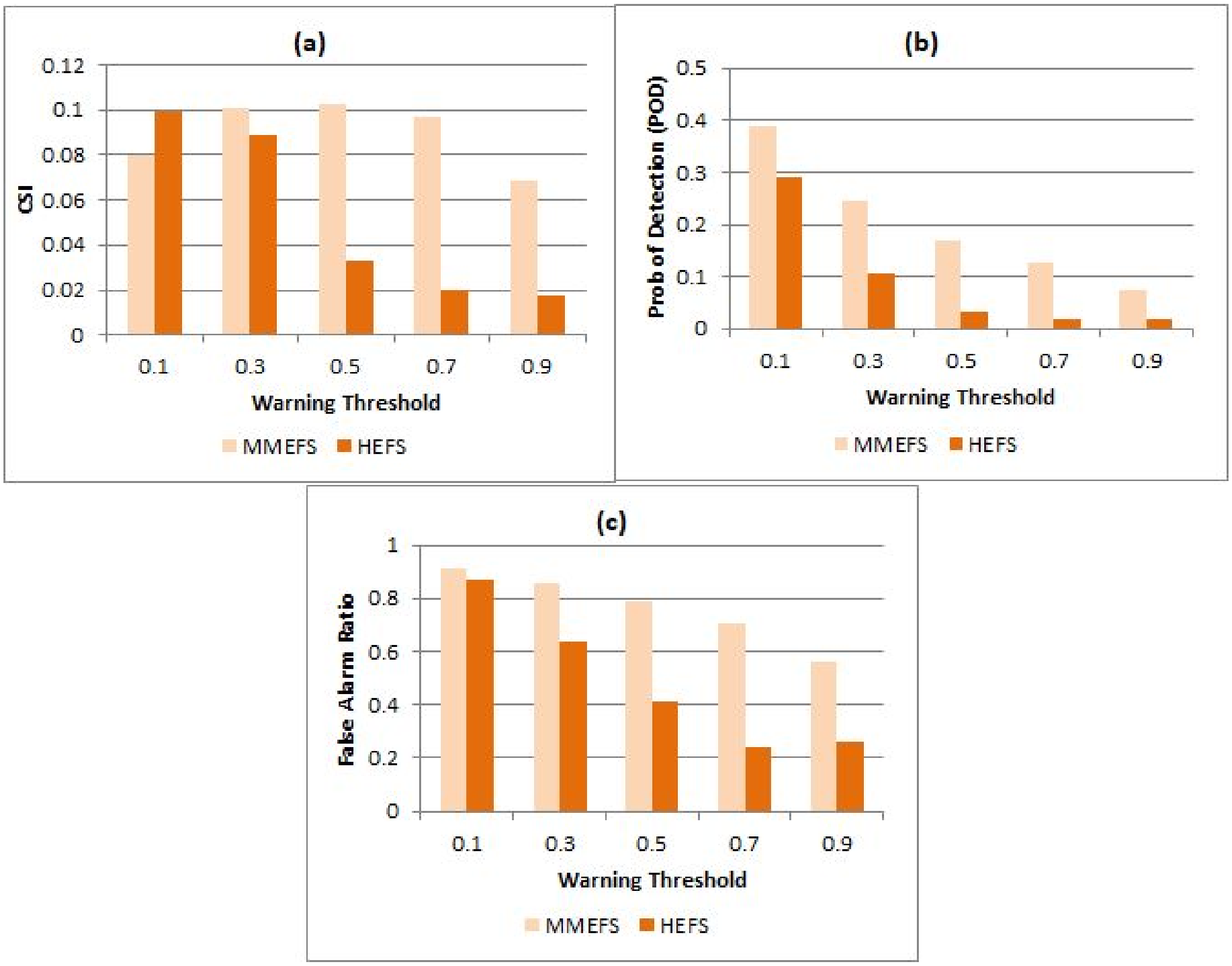


Figure 4. Minor flood stage and 2 - 6 day lead time. Pooled results for 103 validation locations over 2.75 years. a) is CSI, b) is POD, c) is FAR.

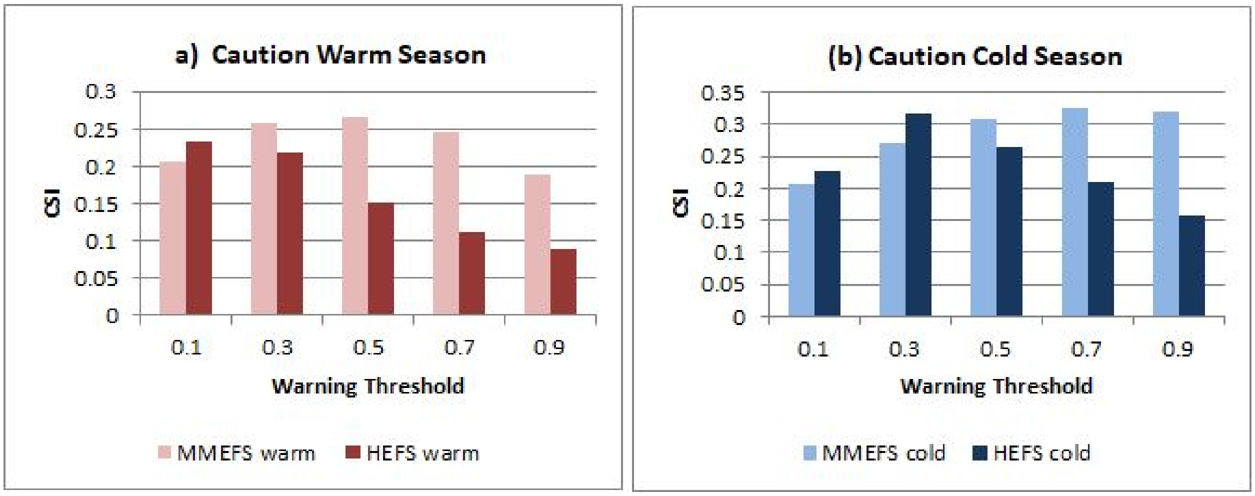


Figure 5. Caution stage statistics (CSI only) by season. Pooled results for 103 validation locations over 2.75 years.

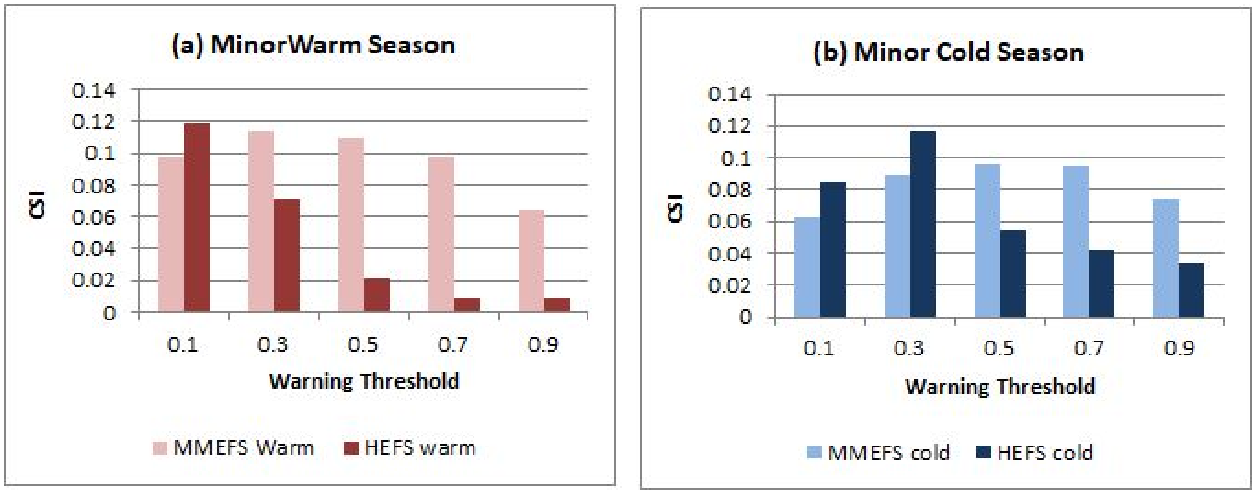


Figure 6. Minor flood stage statistics (CSI only) by season. Pooled results for 103 validation locations over 2.75 years.

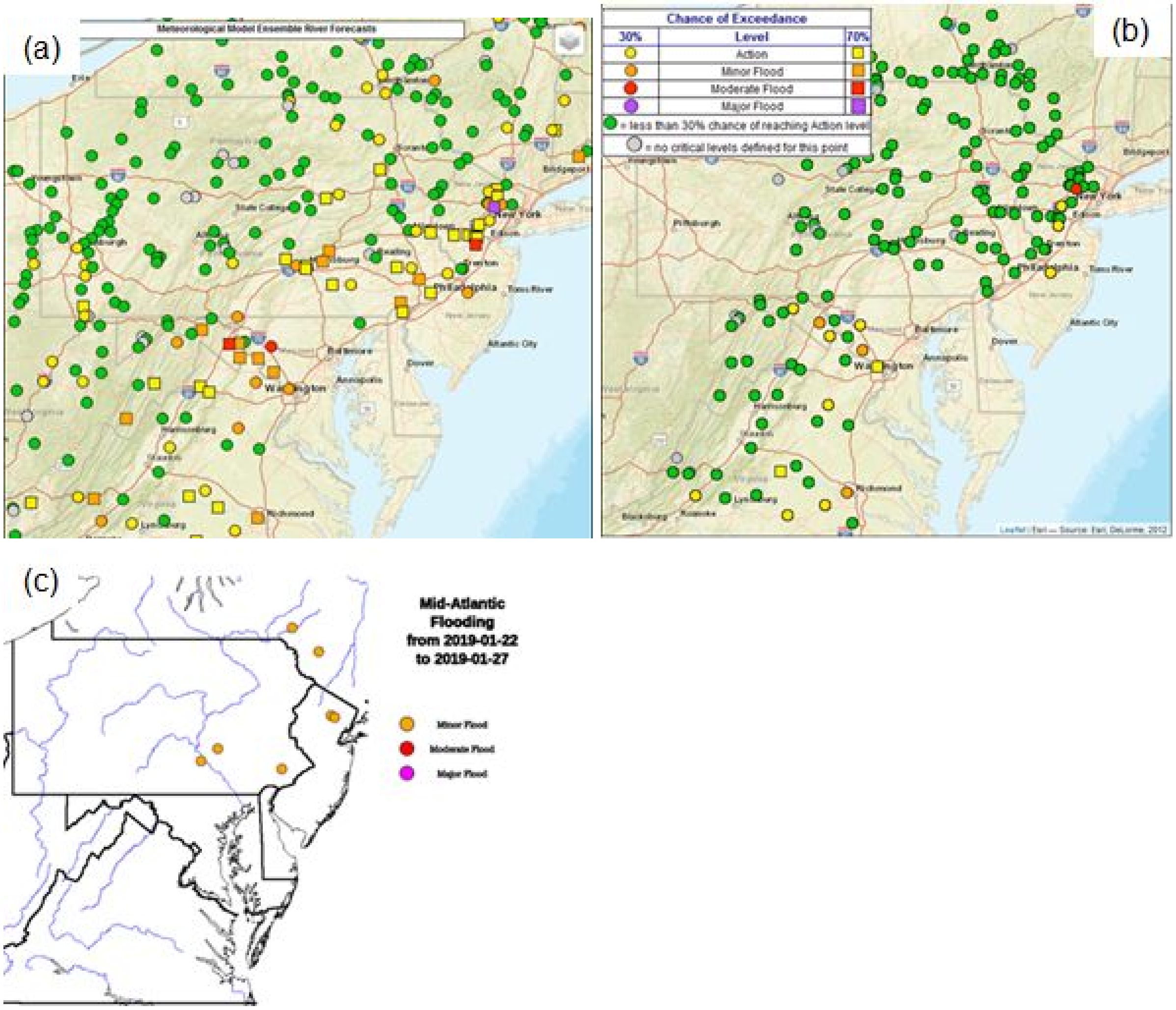


Figure 7. (a) MMEFS-NAEFS forecast based on Jan 22 0z, 2019 NAEFS run (upper left) and (b) MMEFS HEFS forecast based on Jan 22 0z GEFS run (upper right), and (c) seven forecast points reached minor flooding between late afternoon on 1/24 and noon on 1/25 (bottom).

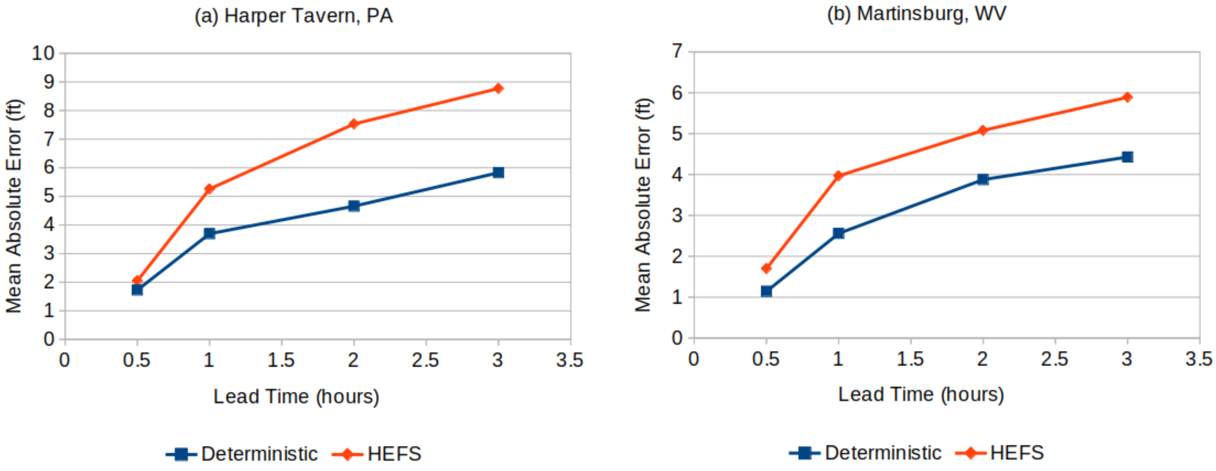


Figure 8. Mean absolute error for forecast-observed pairs with observed stages above flood stage. This is during a 2.75 year period in which 14 distinct flood events occurred at Harper Tavern and 20 occurred at Martinsburg.

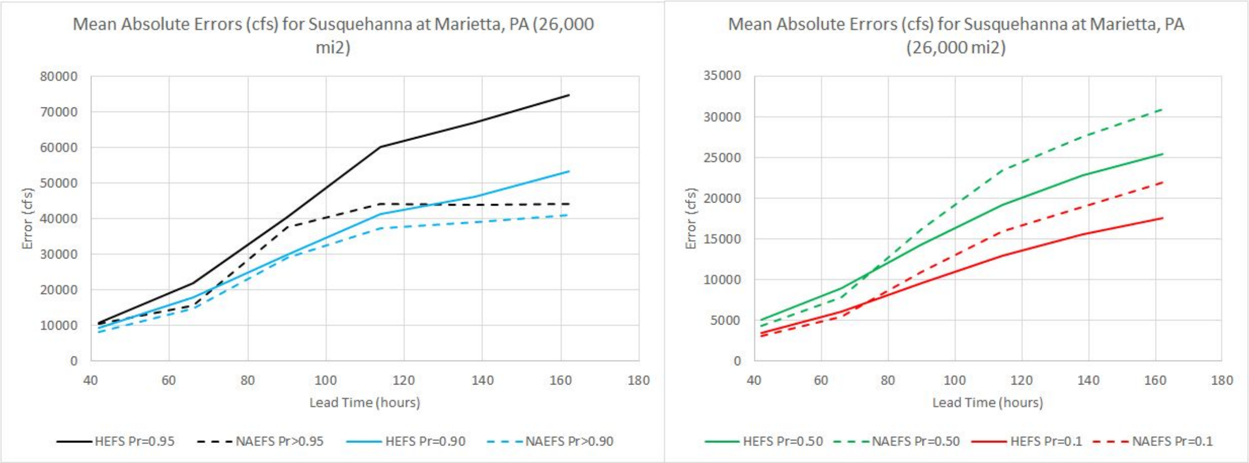


Figure 9. Differences in MAE for higher flows (left panel) and lower flows (right panel) at Marietta, PA. MMEFS-NAEFS has lower MAE for high flows and HEFS has lower MAE for medium and low flows.

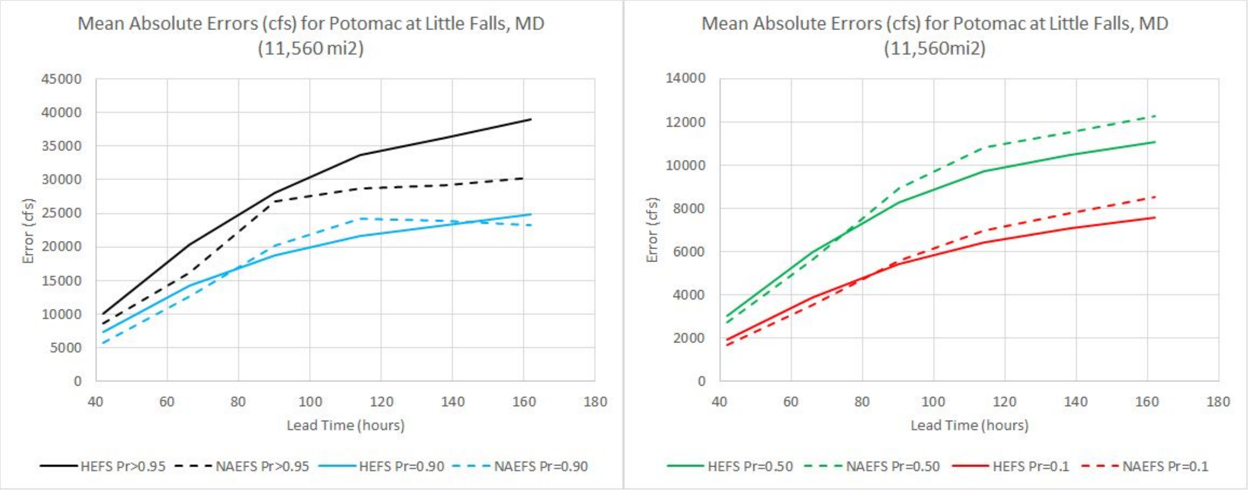


Figure 10. Differences in MAE for higher flows (left panel) and lower flows (right panel) at Little Falls, MD. MMEFS-NAEFS has lower MAE for high flows and HEFS has lower MAE for medium and low flows.