This document serves as the final report for NOAA Award NA14NWS4680016 that covered the period May 1, 2014 – April 30, 2017.

This project had several goals focusing on the use of high-resolution ensemble WRF simulations to assist weather forecasters with improving forecasts of weather elements related to convective systems, with special emphasis on precipitation and ways that it could impact hydrological forecasting.

During the first year of the project, the newest available version of ARPS3DVAR code was implemented along with the latest version of WRF (at that time version 3.6.1). Scripts were set up to allow real time runs so that guidance could be supplied to the DMX NWS office to assist them with forecasts. Testing was initially done on some summer 2014 events for which data were archived. It was determined that initialization of the model must use a prior 3 or 6 hour forecast so that the WRF ensemble could be launched with radar data assimilation immediately at 00, 06, 12, or 18 UTC when convection was present in a domain centered on Iowa that included nine NWS radars. Simulations were integrated forward for 12 hours.

Also during the first year, we worked with Mike DeWeese at the NCRFC to obtain HL-RDHM parameters and forcing data from the operational system. We created a way to convert the WRF netCDF output into the format needed to supply QPF to the HL-RDHM model. Initial ensemble runs were done with the WRF model using two microphysical schemes (Thompson and Morrison) and two planetary boundary layer (PBL) schemes (YSU and MYNN2.5). Early results were discussed at the AMS 30th Conference on Hydrology in New Orleans in January, 2016.

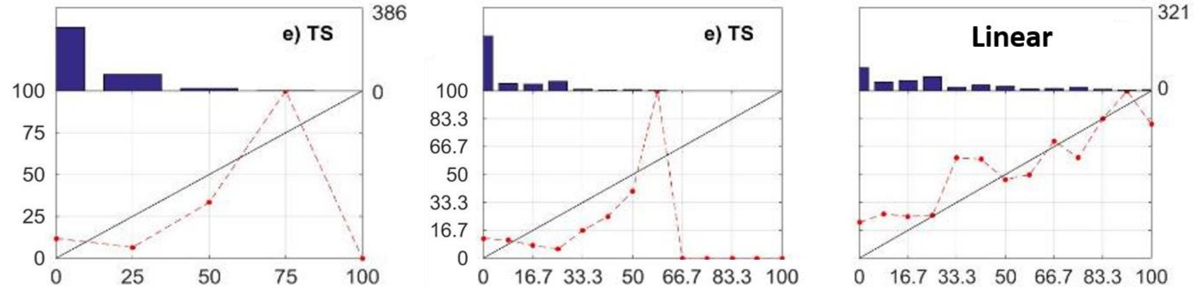
During year two of the project, 23 simulations were performed using the 4-member WRF ensemble for heavy rain events that occurred in Iowa during 2015. A presentation was given on the radar data assimilation portion of the project at the Unidata Users Workshop in June 2015 in Boulder, CO. The probability-matched mean in the simulations was used as input to the HL-RDHM hydrologic model, particularly the SAC-HTET model. The model was calibrated and evaluated for our test basin, the Squaw Creek Watershed near Ames, IA.

In addition, classification began on the simulated morphologies in the sample of cases in each ensemble member using the approach taken in previous studies (e.g. Duda and Gallus, Wea. Forecasting, 2010). A total of 35 morphologies were classified in the sample and scored using a metric adopted from Snively and Gallus (2014, Wea. Forecasting). The mean accuracy score in this sample was 0.49, a value that matched what was found in the larger sample of cases studied by Snively and Gallus. During the classification work, it became apparent that the spread in the 4-member ensemble was too small, and we thus began to explore the use of different initial and lateral boundary conditions to increase spread. Preliminary results of this work were presented at the AMS 28th Conf. on Weather Analysis and Forecasting/24th Conf. on Numerical Weather Prediction in Seattle during January 2017.

During the third year (a no cost extension year), we switched to a new ensemble that not only used mixed-physics but also mixed initial and lateral boundary conditions. We also increased from two to four the number of different microphysical and planetary boundary layer schemes used (Thompson, Morrison, Goddard, and WSM6 for microphysics; YSU, MYNN 2.5, QNSE, and MYJ for planetary boundary layer schemes). We simulated a total of 35 past events (both during 2015 and 2016) using the new ensemble. Our focus on this sample of cases became an investigation of how to best forecast convective morphology. In addition, we were alerted to problems in the SAC-HTET model by users at the NCRFC and we switched our hydrologic modeling to the SAC-HT model. In some initial testing early in year 3, we found an RMSE using the probability matched ensemble QPF in this hydrologic model of 222.1 cfs with a MAE of 134.2 cfs. These values were regarded as too high, and showed the serious impact of even relatively small displacement errors in warm season convective QPF.

Using the updated ensemble members outlined above, a total of 35 morphologies were reclassified from 32 12h WRF simulations. Multiple methods of created a forecast of morphology from the ensemble were investigated. The “majority rules” average ensemble mean accuracy skill score was 0.50, roughly the same accuracy score found by Snively and Gallus study (0.49). Because our results of underforecasting bow echoes (BE) and squall lines with trailing stratiform precipitation (TS) were similar to Snively and Gallus, two ensemble mean bias corrections were also tested. The first bias correction reclassified a convective mode to a BE (TS) if at least 2 members simulated a BE (TS) at a given hour, which resulted in the same average ensemble mean accuracy skill score (0.50). The second correction reclassified the ensemble average to BE (TS) when at least one BE (TS) was simulated. This correction resulted in a lower average ensemble mean accuracy skill score of 0.42, where scores of 7 events improved and scores of 13 events declined. We also looked at hourly probabilistic forecasts of simulated convective mode. First, we calculated forecast probabilities using the percentage of members simulating a convective mode for each hour that convection was observed (hereafter, referred to as direct method). Second, we incorporated a 3-hour neighborhood approach, which considered the three hours surrounding an observation. For instance, an observed convective mode at 12 UTC used the simulated output from 11, 12, and 13 UTC. This allowed for the possibility of catching small timing errors. To evaluate these probabilistic forecasts, reliability diagrams were used. In general, the reliability of probabilistic forecasts for each of the nine individual convective modes performed poorly, particularly with forecast probabilities of greater than 50%. However, when the convective modes are grouped into more general categories of linear and cellular, the forecast reliability markedly improved. A third forecasting method incorporated the forecast probability of a convective mode occurring during a 6 or 12 hour period. The probability of a convective mode occurring over the forecast time period was calculated as the frequency of simulated convective modes among all four members. The divergence skill score (DSS) developed by Weijs et al. (*Mon. Wea. Rev.*, 2010) was used to verify this method. A DSS determines whether information was gained (reduced uncertainty) or lost (increased uncertainty) from a given forecast when compared to a climatological forecast. Of the 32 simulated events evaluated, 68.8%, 65.6%, and 66.7% of the forecasts gained information for the full 12 hour period, first 6 hour period and second 6 hour period, respectively. A paper titled “WRF ensemble prediction of convective morphology: A preliminary examination” is in preparation for submission to *Wea. Forecasting* in May 2017.

Examples of the reliability diagrams for squall lines with trailing stratiform rain (TS) and broader linear group. The first diagram is reliability for the direct forecast probabilistic method, while the 2nd and 3rd diagrams are reliabilities using the neighborhood forecast probabilistic method.



We also tested the probability of precipitation (PoP) – quantitative precipitation forecasting technique (Schaffer et al., *Wea. Forecasting,* 2011), the last component we had proposed to explore in our original proposal. Brier Scores were calculated individually for the 2015 events and 2016 events using training data from the 2007-2008 and 2010 years from Kochasic and Gallus (*In press*) and we focused on the 0.01, 0.05 and 0.10 in. thresholds. Since the 45x45 neighborhood size performed the best in the Kochasic and Gallus study, we initially evaluated that neighborhood size. The Brier Scores were worse for both years compared to the Kochasic and Gallus study using either training data sets. We proceeded to then evaluate the two most similar neighborhood sizes, 35x35 and 55x55, for additional comparisons. The Brier Scores were again worse for the current study than the Kochasic and Gallus study. As with the Kochasic and Gallus study, however, the 45x45 neighborhood size performed the best when producing PoP forecasts. In most cases, the first six hour PoP forecasts performed better than the second six hour forecasts. Similarly, as the precipitation thresholds increased, the PoP forecasts performed better. Although there were far fewer events in 2016, the Brier Scores were fairly similar to the 2015 Brier Scores.

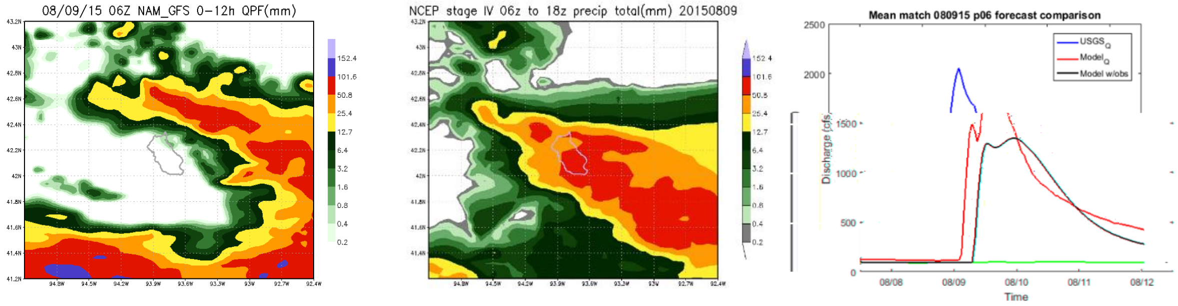
The tables below show the Brier Scores from the 2015 and 2016 events for the 45x45 neighborhood size with the 2007-2008 training data. Kochasic and Gallus Brier scores are also included as a comparison.

0-6 hour forecast	2015 Brier Score	2016 Brier Score	Kochasic Brier Score
0.01 in. threshold	0.1469	0.1808	0.0868
0.05 in. threshold	0.1204	0.1192	0.0474
0.10 in. threshold	0.0931	0.0683	0.0279

7-12 hour forecast	2015 Brier Score	2016 Brier Score	Kochasic Brier Score
0.01 in. threshold	0.1848	0.1766	0.0786
0.05 in. threshold	0.1510	0.1576	0.0471
0.10 in. threshold	0.1039	0.0961	0.0303

Finally, we investigated further the use of an ensemble mean QPF as forcing for the SAC-HT hydrological model to generate streamflow predictions. This was done for the Squaw Creek USGS outlet in Ames, IA and for all 32 simulations. The probability matched mean method was used to create the ensemble mean QPF as outlined by Ebert (*Mon. Wea. Rev.*, 2001). After calibration, the distributed SAC-HT was initialized by with a roughly 4 months spin up period using NCRFC precipitation observations. At the time of the WRF initialization, the probability matched mean QPF was used as an input into the SAC-HT model to create a 12 hour forecast. The SAC-HT run was continued for 7 days to capture the hydrograph recession. We compared the simulated streamflow using WRF ensemble QPF input, simulated streamflow with only NCRFC observation input, and USGS observed streamflow. Overall, the simulated streamflow using WRF ensemble QPF input performed rather poorly, underforecasting streamflow in most cases. Error in storm placement in the QPF was found to be a considerable factor in the poor performance as determined by visually comparing QPF to Stage 4 QPE for 8 events. Figures below show one event and an example QPF; the Squaw Creek watershed is shown for reference. These comparison figures gave insight into the location differences of the heaviest precipitation. Despite typically underforecasting rainfall over the Squaw Creek watershed, the timing of the most intense rainfall over these periods performed well. In most cases, the location of the heaviest QPF was displaced by a relatively short distance, typically within tens of km. However, the distance displacement was usually enough to locate the most intense QPF output outside of the Squaw Creek watershed.

Below are the 12h QPF from 06 – 18 UTC on 09 August, 2015 for one ensemble member, the corresponding Stage 4 QPE, and resulting streamflow comparison for Squaw Creek in central Iowa (blue – USGS obs, black – model with NCRFC obs, red – model using QPF from WRF ensemble).



Key Findings:

- Convection-allowing WRF simulations used to create a four member ensemble with radar data assimilation do a reasonably good job with the timing of warm season heavy rainfall events, but displacement errors of a few tens of kilometers result in large errors when the predicted short-term precipitation is used in the SAC-HT model. Work is needed to identify methods that account for the displacement errors, perhaps through the use of larger ensembles or neighborhood techniques that shift the predicted precipitation. Based on our findings, probabilistic guidance for streamflow is recommended for small basins during the warm season because the small displacement errors for heavy convective rainfall are unavoidable.

- A small WRF ensemble may be able to provide skillful forecasts of convective morphology (and hence most likely severe storm impacts) and would be most effect for forecasts of broader system types such as linear, nonlinear, and cellular. Underdispersion in the ensemble causes unreliable forecasts in some situations for specific prediction of subtypes of systems such as squall lines with trailing stratiform rain. Future work should explore the use of larger ensembles.
- A short-range small convection-allowing ensemble has increased spread when the members use mixed initial and lateral boundary condition information in addition to mixed physics.