



ARL-TR-8835 • Oct 2019



**Application of the Scale Decomposition  
Technique for Assessing Precipitation  
Forecasts of a High-Resolution Weather  
Research and Forecasting Advanced Research  
WRF (WRF-ARW) Ensemble for a Complex  
Mixed Precipitation Event near the  
Washington, DC, Area**

**by John W Raby, Robert E Dumais, Huaqing Cai,  
Jeffrey A Smith, Leelinda P Dawson, and Brian P Reen**

## **NOTICES**

### **Disclaimers**

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.



# **Application of the Scale Decomposition Technique for Assessing Precipitation Forecasts of a High- Resolution Weather Research and Forecasting Advanced Research WRF (WRF-ARW) Ensemble for a Complex Mixed Precipitation Event near the Washington, DC, Area**

**John W Raby, Robert E Dumais, Huaqing Cai, Jeffrey A Smith,  
Leelinda P Dawson, and Brian P Reen**  
*Computational and Information Sciences Directorate, CCDC Army Research  
Laboratory*

**REPORT DOCUMENTATION PAGE**

*Form Approved*  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

<b>1. REPORT DATE (DD-MM-YYYY)</b> October 2019		<b>2. REPORT TYPE</b> Technical Report		<b>3. DATES COVERED (From - To)</b> 1 October 2018–12 September 2019	
<b>4. TITLE AND SUBTITLE</b> Application of the Scale Decomposition Technique for Assessing Precipitation Forecasts of a High-Resolution Weather Research and Forecasting (WRF), Advanced Research WRF (WRF-ARW) Ensemble for a Complex Mixed Precipitation Event near the Washington, DC, Area				<b>5a. CONTRACT NUMBER</b>	
				<b>5b. GRANT NUMBER</b>	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b> John W Raby, Robert E Dumais, Huaqing Cai, Jeffrey A Smith, Leelinda P Dawson, and Brian P Reen				<b>5d. PROJECT NUMBER</b>	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> CCDC Army Research Laboratory ATTN: FCDD-RLC-EM White Sands Missile Range, NM 88002-5501				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>  ARL-TR-8835	
<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b>				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>	
				<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION/AVAILABILITY STATEMENT</b> Approved for public release; distribution is unlimited.					
<b>13. SUPPLEMENTARY NOTES</b> ORCID IDs: Brian Reen, 0000-0002-2031-4731; Jeffrey Smith, 0000-0002-2920-749X					
<b>14. ABSTRACT</b> An ensemble of short-range weather forecasts was generated using the Advanced Research version of the Weather Research and Forecast model for a challenging winter precipitation forecast event near Washington, DC, on 9 February 2016. An assessment was conducted to quantify the uncertainty of the precipitation nowcasts produced by the 28-member ensemble. The assessment used a combination of tools to quantify uncertainty, which was the ensemble mean and observed accumulated precipitation, time series plots of forecast and observed accumulated precipitation, rank histograms, and 2-D observation ranks. These tools showed 1) uncertainty in the location of a maximum in accumulated precipitation, 2) the presence of inadequate spread in forecast precipitation relative to the spread in the observations, and 3) that there was spatial variation in the bias. The scale-decomposition technique was applied to better quantify the uncertainty by isolating the precipitation maximum and assessing the quality of the forecast structure in terms of the spatial scale of the error. The results showed that this technique provided an assessment of model skill as a function of precipitation threshold value and spatial scale. It enabled the separation of the larger errors, attributable to displacement, from the smaller errors attributable to smaller scale processes.					
<b>15. SUBJECT TERMS</b> scale decomposition, scale separation, thresholds, observations, model evaluation tools, wavelet transform					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>  UU	<b>18. NUMBER OF PAGES</b>  46	<b>19a. NAME OF RESPONSIBLE PERSON</b> John W Raby
<b>a. REPORT</b> Unclassified	<b>b. ABSTRACT</b> Unclassified	<b>c. THIS PAGE</b> Unclassified			<b>19b. TELEPHONE NUMBER (Include area code)</b> (575) 678-2004

## **Contents**

---

<b>List of Figures</b>	<b>iv</b>
<b>List of Tables</b>	<b>v</b>
<b>Summary</b>	<b>vi</b>
<b>1. Introduction</b>	<b>1</b>
<b>2. Preliminary Assessment Analysis and Results</b>	<b>3</b>
<b>3. Application of the Scale-Decomposition Technique</b>	<b>21</b>
<b>4. Conclusions and Summary</b>	<b>30</b>
<b>5. References</b>	<b>32</b>
<b>Appendix. Wavelet-Stat (W-S) Metrics</b>	<b>34</b>
<b>List of Symbols, Abbreviations, and Acronyms</b>	<b>36</b>
<b>Distribution List</b>	<b>38</b>

## List of Figures

---

Fig. 1	WRE-N ensemble domains.....	2
Fig. 2	WRE-N ensemble mean 6-h accumulated precipitation (millimeters) at 1800 UTC from the 12-h forecast.....	6
Fig. 3	Observed 6-h accumulated precipitation (millimeters) at 1800 UTC...	7
Fig. 4	GFS deterministic 6-h accumulated precipitation (millimeters) at 1800 UTC from 12-h forecast.....	8
Fig. 5	WRE-N ensemble mean 6-h accumulated precipitation (millimeters) at 1800 UTC from the 12-h forecast with location where time series data shown in the following plot were extracted.....	9
Fig. 6	Observed 6-h accumulated precipitation (millimeters) at 1800 UTC. Location marked shows where Fig. 7 time series data were extracted. ....	10
Fig. 7	Time series of accumulated precipitation (millimeters) from the 28 WRE-N ensemble members and the observations for the period 1200 to 1800 UTC for the location shown in Figs. 5 and 6.....	11
Fig. 8	Observed 6-h accumulated precipitation (millimeters) at 1800 UTC from the 12-h forecast with location marked where time series data in the following plot were extracted .....	12
Fig. 9	WRE-N ensemble mean 6-h accumulated precipitation (millimeters) at 1800 UTC with location marked where time series data in Fig. 10 were extracted.....	13
Fig. 10	Time series of accumulated precipitation (millimeters) from the 28 WRE-N ensemble members and the observations for the period 1200 to 1800 UTC for the location shown in Figs. 8 and 9.....	14
Fig. 11	WRE-N ensemble mean 6-h accumulated precipitation (millimeters) at 1800 UTC from the 12-h forecast with location marked where WRE-N time series data in the following plot were extracted .....	15
Fig. 12	Observed 6-h accumulated precipitation (millimeters) at 1800 UTC with location marked where time series data in Fig. 13 were extracted .....	16
Fig. 13	Time series of accumulated precipitation (millimeters) from the 28 WRE-N ensemble members for the location shown in Fig. 11 and the observations for the location shown in Fig. 12 both for the period 1200 to 1800 UTC .....	17
Fig. 14	Ensemble member 06 and observed 6-h accumulated precipitation (millimeters) at 1800 UTC.....	18
Fig. 15	Uniform rank histogram example .....	19
Fig. 16	Rank histogram for the WRE-N ensemble 6-hr accumulated precipitation forecast at 1800 UTC.....	19

Fig. 17	WRE-N 6-h accumulated precipitation observation ranks at 1800 UTC .....	20
Fig. 18	Graphic produced by W-S showing the 6-h precipitation (millimeters) from (left) member 06 and (right) observations, along with a $32 \times 32$ grid square tile where a 2-D Haar wavelet filter was applied.....	22
Fig. 19	W-S output graphic of the binary difference field for threshold value $\geq 3.0$ mm .....	23
Fig. 20	W-S output graphic of the first spatial scale (4 km) for threshold value $\geq 3.0$ mm .....	24
Fig. 21	W-S output graphic of the second spatial scale (8 km) for threshold value $\geq 3.0$ mm .....	24
Fig. 22	W-S output graphic of the third spatial scale (16 km) for threshold value $\geq 3.0$ mm .....	25
Fig. 23	W-S output graphic of the fourth spatial scale (32 km) for threshold value $\geq 3.0$ mm .....	25
Fig. 24	W-S output graphic of the fifth spatial scale (64 km) for threshold value $\geq 3.0$ mm .....	26
Fig. 25	W-S output graphic of the sixth spatial scale (128 km) for threshold value $\geq 3.0$ mm .....	26
Fig. 26	W-S output graphic of the fifth spatial scale (64 km) for threshold value $\geq 6.0$ mm .....	28
Fig. 27	W-S output graphic of the fifth spatial scale (64 km) for threshold value $\geq 9.0$ mm .....	29

## List of Tables

---

Table 1	Intensity Skill Scores for spatial scales 4 to 128 km and threshold values of $\geq 3.0$ , $\geq 6.0$ , and $\geq 9.0$ mm .....	29
---------	--	----

## Summary

---

An ensemble of short-range weather forecasts were generated using the Advanced Research version of the Weather Research and Forecast model for a challenging winter precipitation forecast event near Washington, DC, on 9 February 2016. The ensemble members differed primarily in the source of initial and boundary conditions, the physics parameterizations, and other configuration options employed. A preliminary assessment was conducted to quantify the uncertainty of the accumulated precipitation short-lead-time nowcasts produced by the ensemble, which consisted of 28 members. The assessment used a combination of tools to quantify uncertainty in ensembles. These tools consisted of the ensemble mean and observed accumulated precipitation, time series plots (often referred to as spaghetti plots of forecast and observed accumulated precipitation), rank histograms, and 2-D observation ranks. These revealed 1) uncertainty in the location of a maximum in accumulated precipitation, 2) the presence of inadequate spread in forecast precipitation relative to the spread in the observations, 3) positive and negative biases, and 4) spatial variation in the bias. Analysis of these results revealed the need for additional verification to further quantify the uncertainty arising from the spatial errors generated due to the displacement of the precipitation maximum. The scale-decomposition technique was applied to better quantify the uncertainty by isolating the precipitation maximum and assessing the quality of the forecast structure in terms of the spatial scale of the error. The results showed that this technique provided an assessment of model skill as a function of precipitation threshold value and spatial scale, and enabled the separation of the larger errors, attributable to displacement, from the smaller errors attributable to smaller-scale processes.

## 1. Introduction

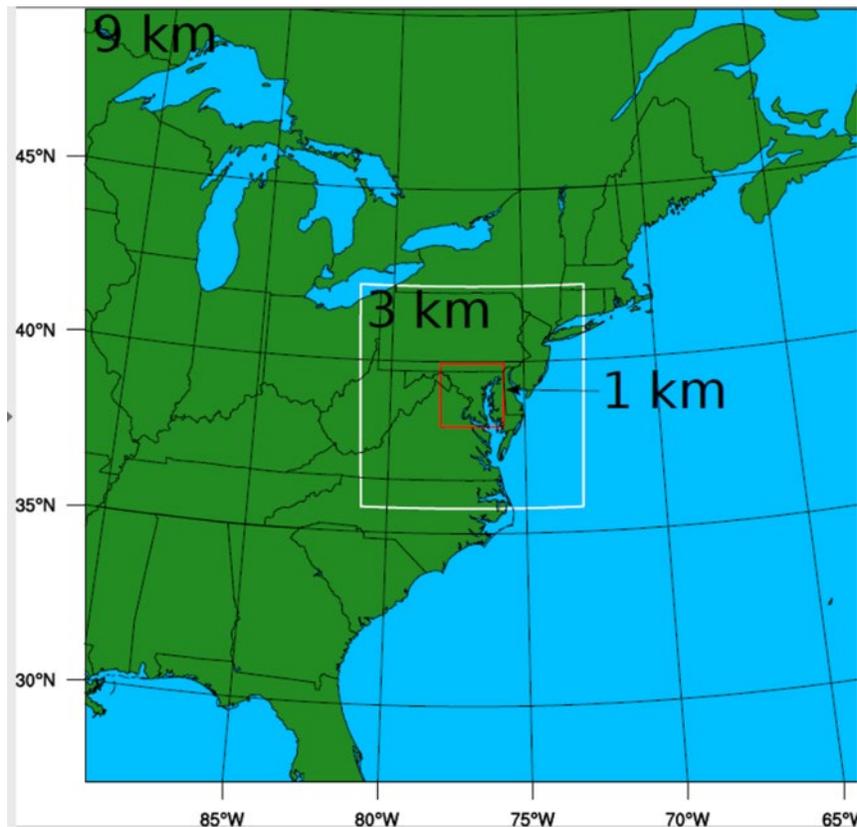
---

A 28-member Weather Running Estimate–Nowcast (WRE-N) ensemble was specifically configured to evaluate the value of high spatial and temporal resolutions when nowcasting weak-to-moderate winter systems. These systems were associated with challenging precipitation events in terms of amount, type, location, and onset for a limited area domain. Evaluation of precipitation forecasts entailed the assessment of several model performance aspects, such as the precipitation amount and type, surface temperature, ground temperature, boundary layer thickness, cloud amount, and ceiling height 1–2 days prior to the event. The chosen case study date had a winter system that presented these challenges over the Maryland–Virginia–Delaware (Delmarva) area on 9 February 2016. This study focused on one of the stated aspects, which was accumulated precipitation forecasting over very-short-range time periods.

Dumais et al. (2018) describe in detail the mixed winter precipitation event and the challenges it presented for forecasters and planners in the days leading up to 9 February 2016. The output from several Numerical Weather Prediction (NWP) models revealed a great variation of solutions indicative of the complexity of the mesoscale structure associated with this system. In general, there was a tendency to forecast snow north and northwest of the Washington, DC (DC), metro area, and rain to the south and over the Delmarva Peninsula. The snow forecasts ranged from heavier snow bands supporting accumulation in the north, to only slushy light accumulations in other areas. However, there was considerable disagreement in the model guidance depending on which model output was used for decision making. This high level of uncertainty in the days and hours preceding the event challenged emergency managers responsible for decisions regarding whether or not to mobilize resources for snow removal on and around the DC metro area. Challenges involved in predicting the impact for civilian concerns are not unlike those faced by military decision makers who are faced with the responsibility to plan and execute military operations using guidance from decision aids based on NWP model output.

Dumais et al. (2018) designed an NWP modeling experiment that involved running several simulations for this winter situation case study as an ensemble. The NWP model used was a triple-nest implementation of the Advanced Research version of the Weather Research and Forecast model (WRF-ARW). This model was configured similarly to that tested as a prototype US Army Combat Capabilities Development Command Army Research Laboratory (ARL) WRE-N system (Dumais et al. 2013). The WRF is described in detail by Skamarock et al. (2008). WRF and WRE-N are used interchangeably throughout this report. The WRE-N was allowed to “spin-up” during a 6-h preforecast period, and then run in free

forecast mode for 12 h. The output from the initial 6 h of this free forecast period is the focus of this study. All model simulations were run from 2016 09 Feb, 0600 Coordinated Universal Time (UTC) until 10 Feb, 0000 UTC, with model output provided in hourly forecast intervals. The center of the triple-nested domains was located near the DC metro area (Fig. 1). The outer domain encompassed much of the eastern United States and the Atlantic Ocean from Canada to the Gulf of Mexico, using a 9-km grid spacing. The next domain was focused on the mid-Atlantic region from Ohio extending offshore east of Maryland, and from North Carolina to southern New York. This domain used a 3-km grid. The inner domain, centered over the Delmarva Peninsula, using a 1-km grid.



**Fig. 1 WRE-N ensemble domains**

The WRE-N can use the WRF 4-D data assimilation (FDDA) observation nudging option to incorporate regional and local direct weather observations when available (Dumais et al. 2013). In this ensemble, the use of FDDA varied among the simulations. Dumais et al. (2018) describe, in detail, the methodology used to produce the various members of the model ensemble. The concept was to combine the physics diversity, initial conditions/lateral boundary condition (IC/LBC) diversity, and the diversity in other model details. Diversity in IC/LBC was employed by using various external operational models for initializing and

providing time-dependent lateral boundary conditions for the outer nest. The external models used were either from the National Center for Environmental Prediction (NCEP) deterministic Global Forecast System (GFS) 0.5° cycle based at 0600 UTC 09 Feb, the NCEP Global Ensemble Forecast System (GEFS) 0.5° ensemble model cycle based at 0600 UTC 09 Feb, or the NCEP mesoscale North American Model deterministic model cycle based at 0600 UTC 09 Feb.

The approach used for conducting the preliminary assessment to quantify the uncertainty of the forecast accumulated precipitation involved comparing forecast and observed accumulated precipitation. A tool was used to compute various statistics that quantify the overall character and structure of forecast accumulated precipitation for the ensemble (as a whole). Time series analysis was used to compare forecast accumulated precipitation from the individual ensemble members with each other and with the observations. Observation rank histograms were used to compare the ensemble and observed probability distribution of accumulated precipitation to see how well the ensemble represented the observations, and if there is bias. The 2-D observation rank graphics provided insight on the spatial variation of bias. The preliminary results gave a general sense of the uncertainty, but lacked specifics on the amount of uncertainty as a function of precipitation threshold value and spatial scale. To address this deficiency, a scale-decomposition technique was applied to one ensemble member, enabling the evaluation of its effectiveness in providing more-specific spatial verification information.

## **2. Preliminary Assessment Analysis and Results**

---

---

Traditional grid-to-point methods can verify the skill of NWP in predicting continuous meteorological variables through the computation of such statistics as mean error and root-mean-square error, which characterize model accuracy over the entire domain. When these techniques are applied to high-resolution models, such as the WRE-N, the results can give misleading error estimates when compared with lower-resolution models, which often score better when using these techniques. The issue is the inability of the verification technique to evaluate the true skill of higher-resolution forecasts. High-resolution models replicate mesoscale atmospheric features in a way that is more representative of the actual phenomenon owing to their use of a reduced grid spacing over smaller domains, higher-resolution land-surface models, and better parameterization of subgrid physical processes.

For this preliminary assessment, the initial approach was to characterize the uncertainty for the ensemble in terms of forecasting a particular feature of the accumulated precipitation that might capture errors resulting from the spatial

displacement of that feature. To conduct this type of assessment, an Ensemble-Stat (E-S) tool developed by the National Center for Atmospheric Research (NCAR) as part of a suite of tools used for model verification called the Model Evaluation Tools (METs) (NCAR 2016b) was used. E-S uses postprocessed model output and observations as input. For postprocessing, the forecasts from the model grids are first interpolated onto a destaggered grid (from WRF native-staggered C grid), while keeping the same horizontal grid spacing as the WRF native input grid. This is done using the NCAR Unified Post Processor (UPP; [https://dtcenter.org/upp/users/docs/user\\_guide/V3/upp\\_users\\_guide.pdf](https://dtcenter.org/upp/users/docs/user_guide/V3/upp_users_guide.pdf)). The UPP software also interpolates the WRF output onto pressure surfaces (from native sigma), computes all additional variables of interest from WRF basic prognostic variables, and generates additional diagnostic levels used by MET (Dumais et al. 2018).

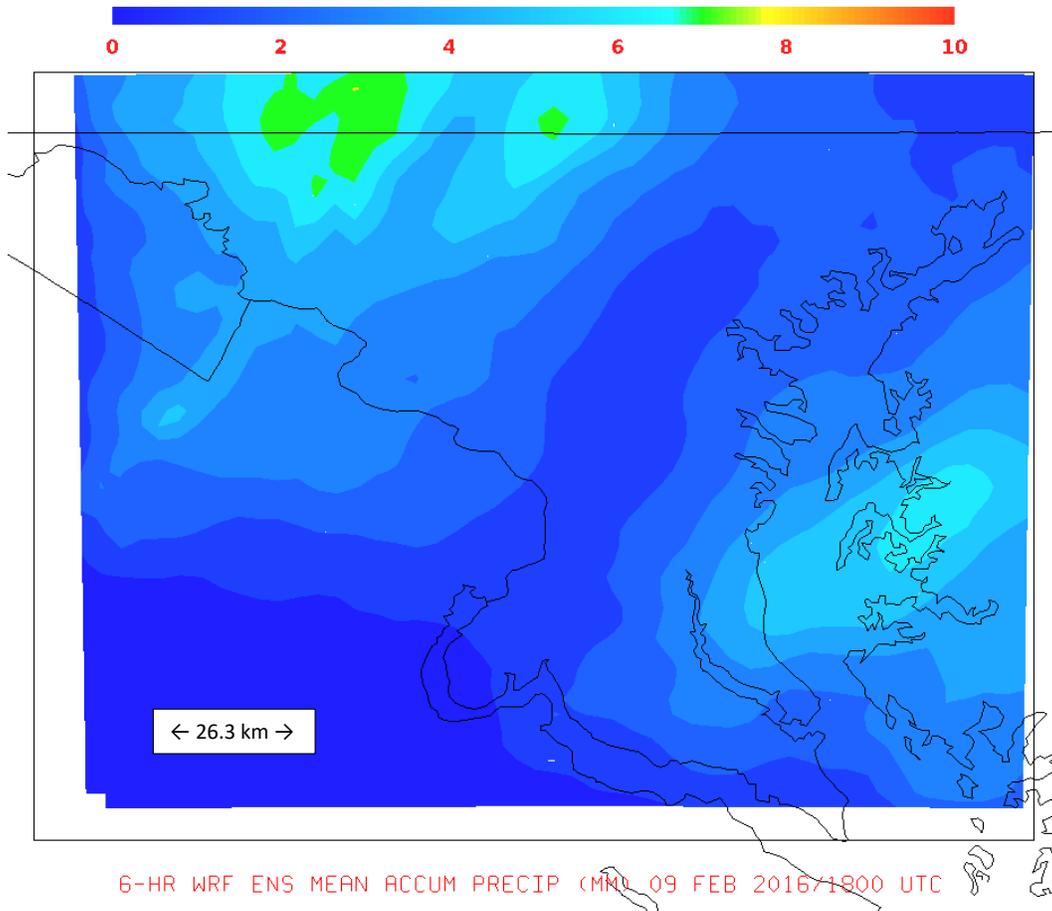
The observational data came from the NCEP hourly Stage IV Quantitative Precipitation Estimate national mosaic (4-km grid spacing) dataset (Nelson et al. 2016). The forecast and observations data required preprocessing to generate accumulated precipitation in such a way as to focus on the winter event's precipitation during the first 6 h of the free forecast, from 2016 09 Feb, 1200 UTC, to 09 Feb, 1800 UTC. This required the accumulated precipitation value be set to zero at 1200 UTC, then accumulated for each subsequent hour. To generate new sets of hourly forecast and observation files that contain these accumulations, the MET Pcp-Combine (P-C) tool was used. The next step involved the development of a new "verification" domain with the grid specifications to assure a match between the 1-km forecast grid over the inner domain and the 4-km observation grid over the continental United States (CONUS) domain. This process was accomplished iteratively using the MET Regrid\_data\_plane (R-D-P) tool, which performed an interpolation from the 1-km forecast grid to a "new" 4-km grid with the specifications set so that the new grid encompassed the same area as the 1-km grid. Normally, E-S can perform the regridding automatically, but for the 4-km observation grid over CONUS, E-S encountered an "out-of-memory" error due to the volume of data held in memory, even when running on the ARL High Performance Computer. After several iterations of the R-D-P tool, a good match was achieved that enabled E-S to run without error by directing the verification to the much smaller "new" grid, which significantly decreased the memory usage.

E-S generated gridded Network Common Data Format (NetCDF) files of several ensemble forecast values such as mean, spread, or uncertainty, and minimum and maximum from the 28 members. Using the observations and forecast data, E-S generated observation rank histograms, spread/skill variance statistics, Probability Integral Transform histograms and gridded NetCDF files containing the spatial (2-D) distribution of the observation ranks. For this study, we used the E-S mean

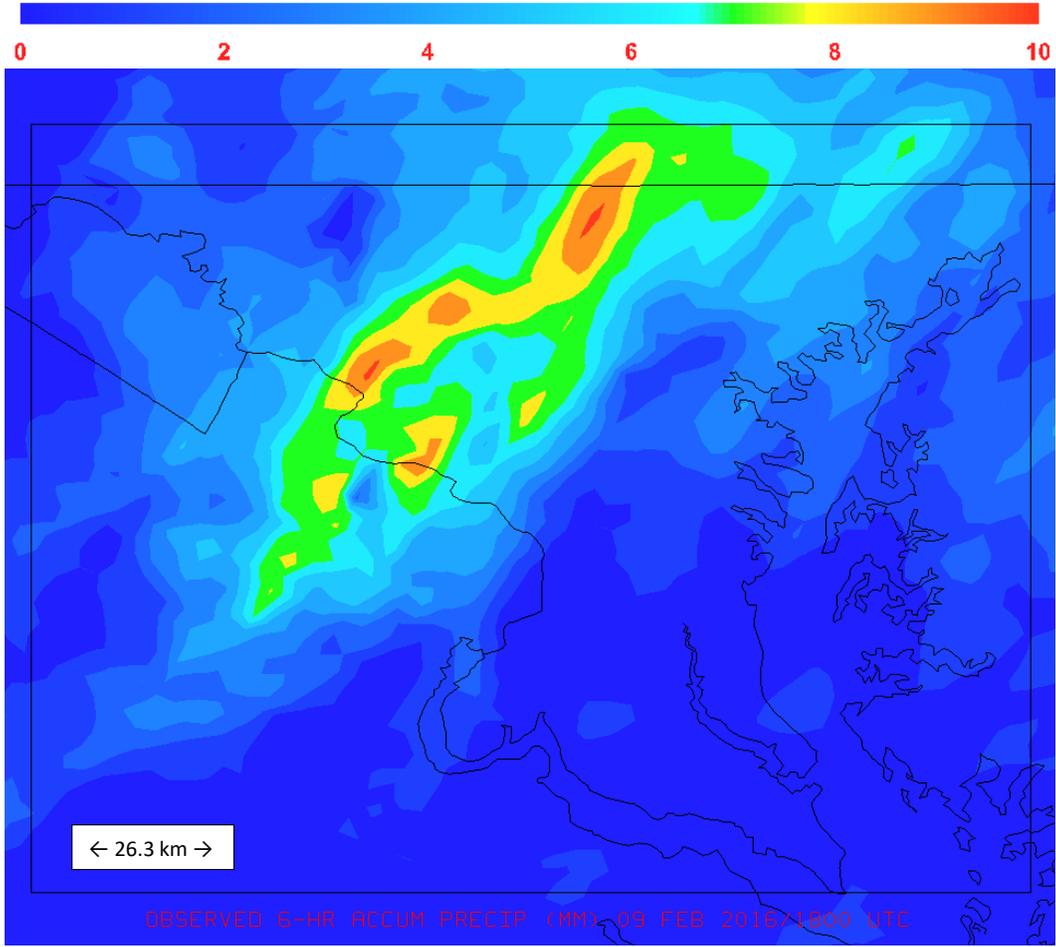
ensemble field from the 12-h lead time forecast (i.e.; the 6-h free forecast) of 6-h accumulated precipitation valid at 1800 UTC on 9 February 2016, the accumulated precipitation rank histogram valid at the same time, and the 2-D spatial distribution of observation ranks output.

For comparison with the E-S output, equivalent graphics of observed accumulated precipitation at 1800 UTC were generated from the 6-h accumulated precipitation fields produced by the P-C tool. In addition, time series plots depicting the accumulated precipitation at significant locations for the time period 1200 UTC to 1800 UTC for all ensemble members and from the observations were produced using the Unidata Integrated Data Viewer (IDV) visualization software. The IDV probe tool was used to extract the point values of accumulated precipitation at the desired locations and export as comma-separated value (CSV) files (UCAR 2018). R language for statistical computing was used to convert the CSV files into data structures that enabled generation of the time series plots (R Core Team 2019).

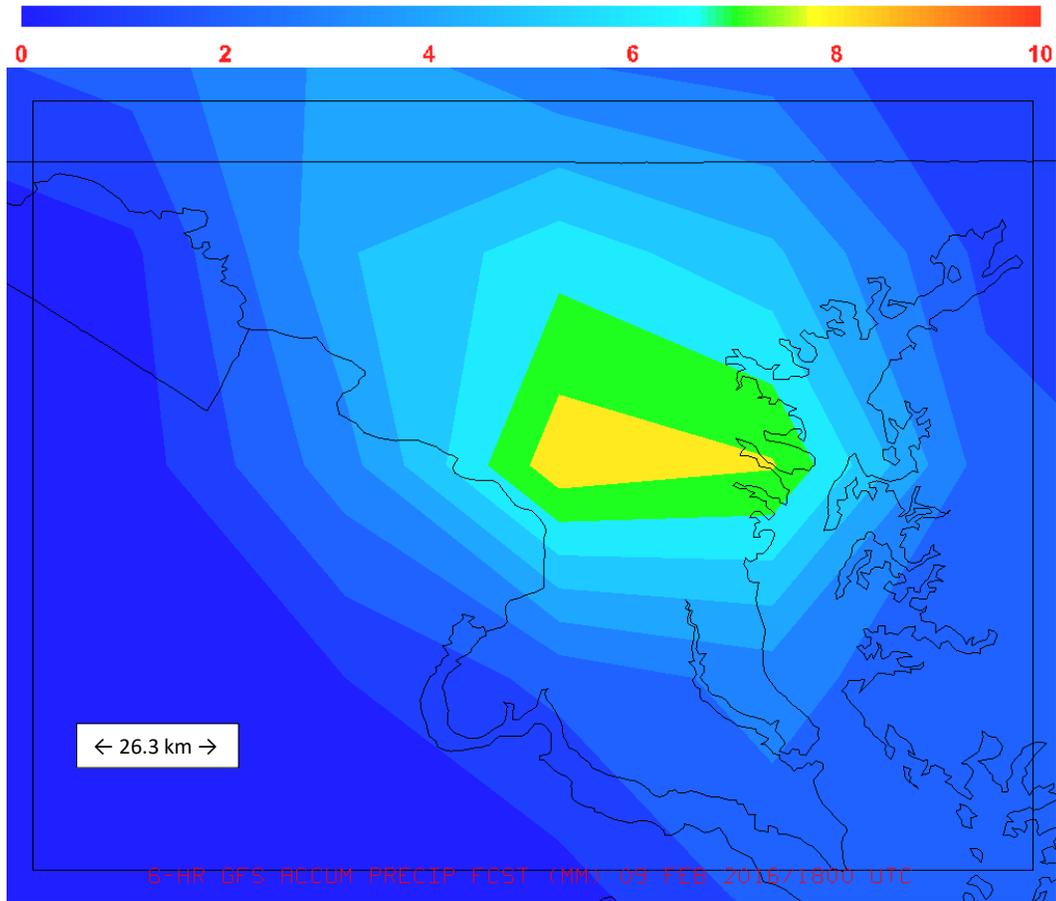
To characterize the uncertainty, the 6-h accumulated precipitation maximum feature for 9 February 2016 at 1800 UTC, situated northwest of the DC metro area, was the focus for analysis. This feature in the ensemble mean (Fig. 2) was displaced northwest of the observed feature (Fig. 3). For comparison, the same feature in the GFS forecast (Fig. 4) was situated southeast of the observed feature.



**Fig. 2** WRE-N ensemble mean 6-h accumulated precipitation (millimeters) at 1800 UTC from the 12-h forecast

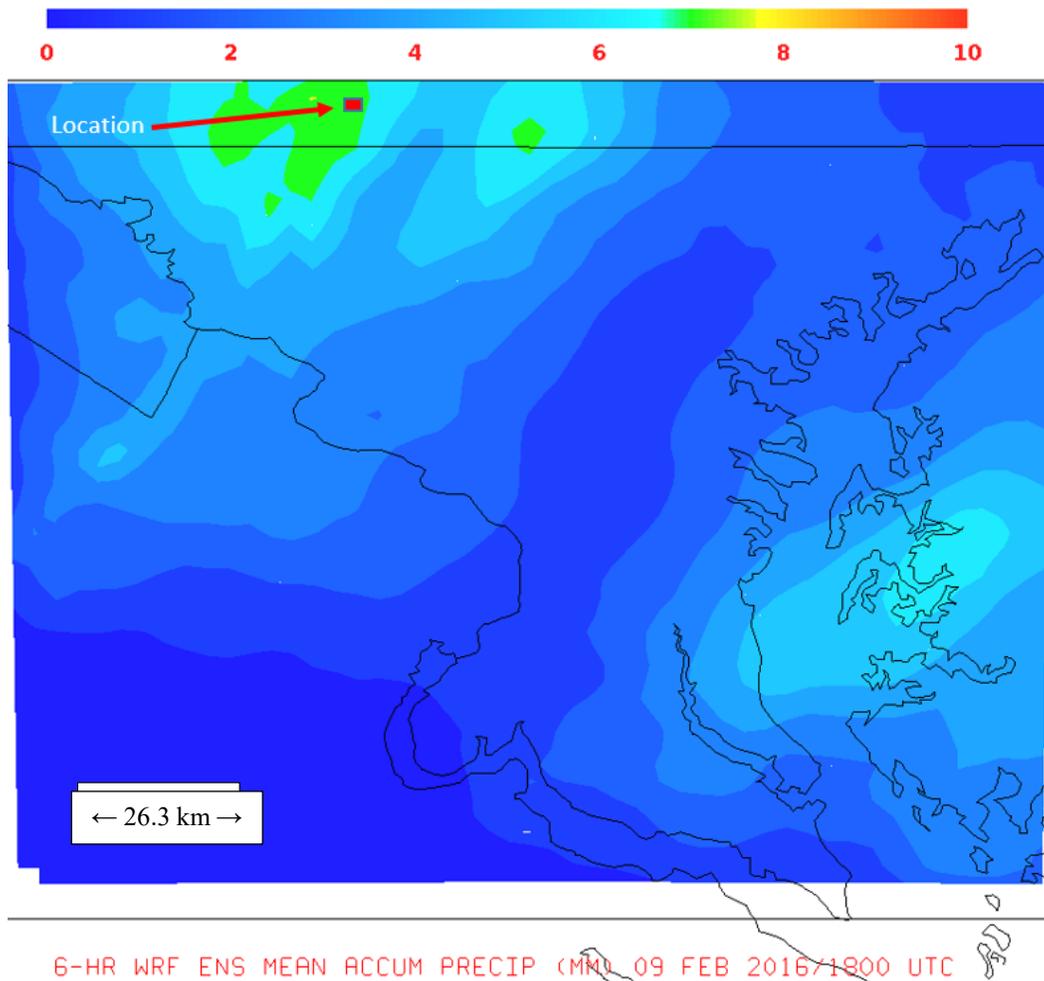


**Fig. 3** Observed 6-h accumulated precipitation (millimeters) at 1800 UTC



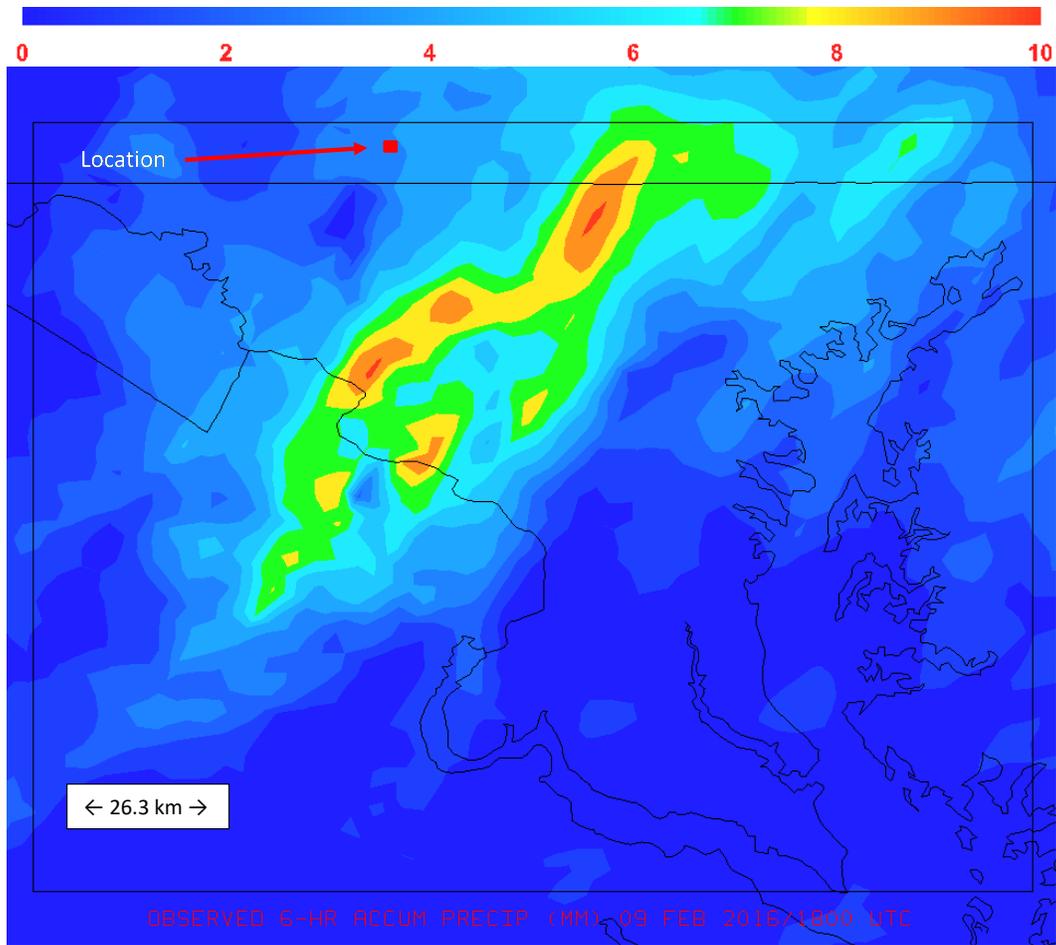
**Fig. 4** GFS deterministic 6-h accumulated precipitation (millimeters) at 1800 UTC from 12-h forecast

To compare the forecasts of accumulated precipitation from all ensemble members with the observations in a meaningful way, 6 h of data were extracted from a location (Fig. 5) within the maximum precipitation feature in the ensemble mean.



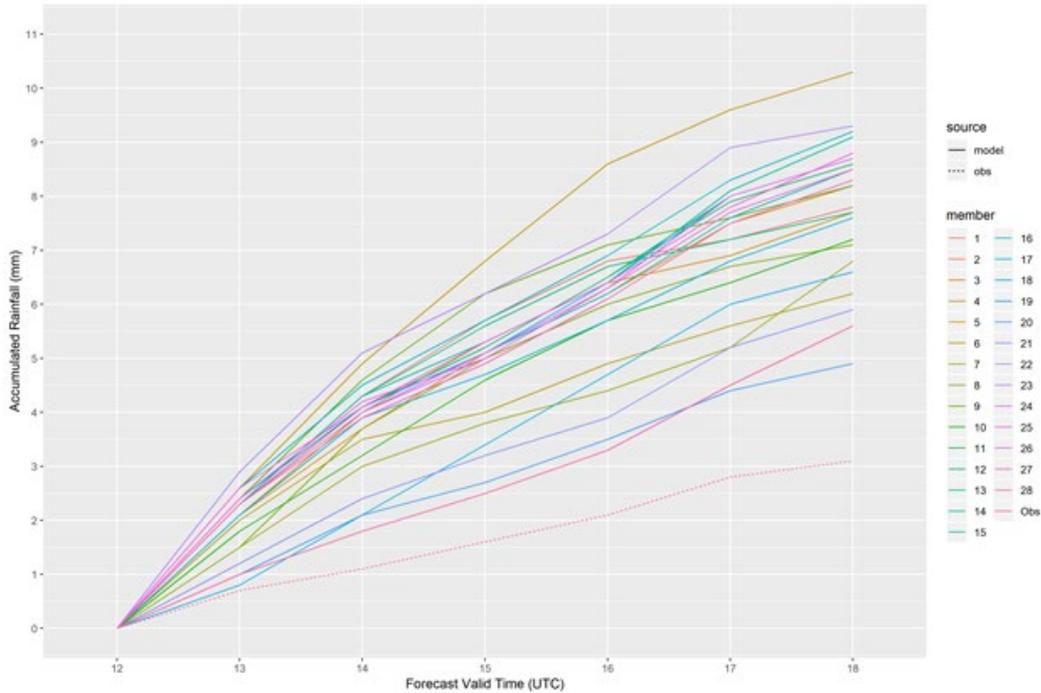
**Fig. 5** WRE-N ensemble mean 6-h accumulated precipitation (millimeters) at 1800 UTC from the 12-h forecast with location where time series data shown in the following plot were extracted

For the observed accumulated precipitation time series, the equivalent data were extracted from the same geographic location, as shown in Fig. 6. This position was well outside the observed precipitation maximum feature located to the south.



**Fig. 6** Observed 6-h accumulated precipitation (millimeters) at 1800 UTC. Location marked shows where Fig. 7 time series data were extracted.

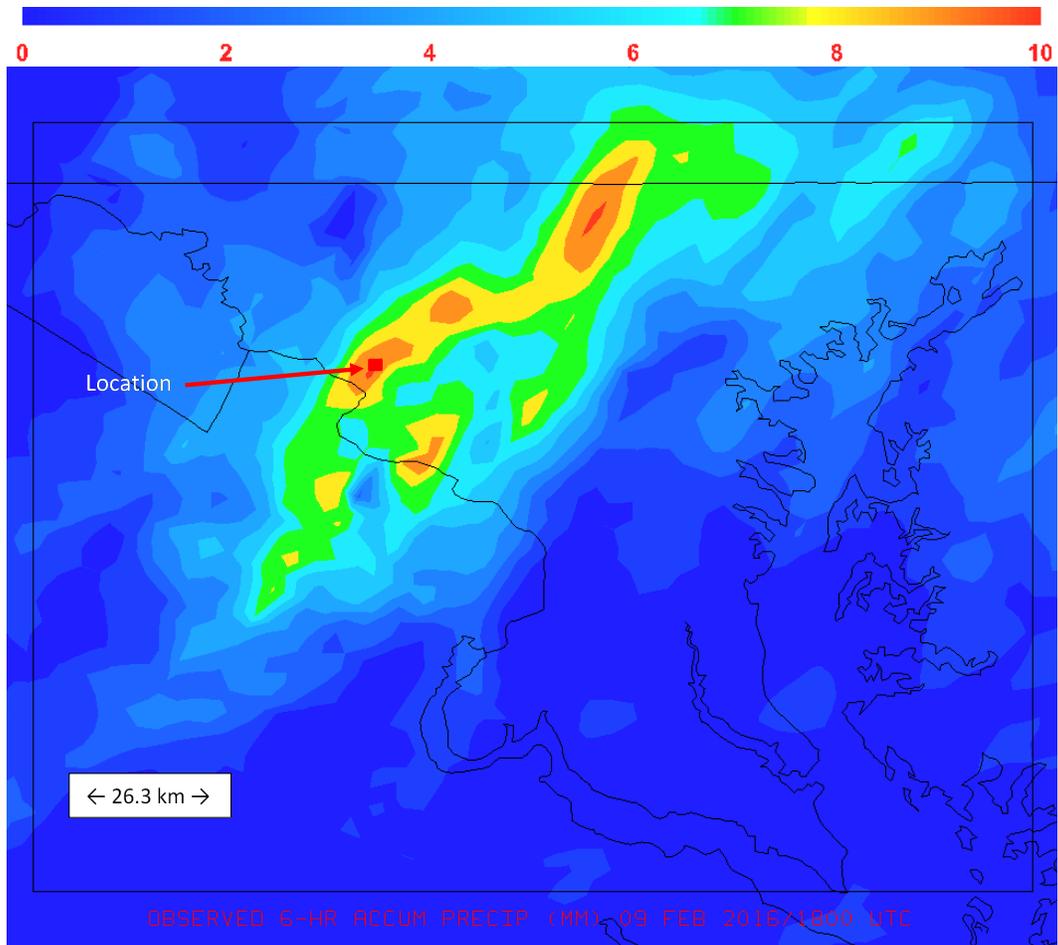
Both sets of time series data were plotted in the graphic shown in Fig. 7.



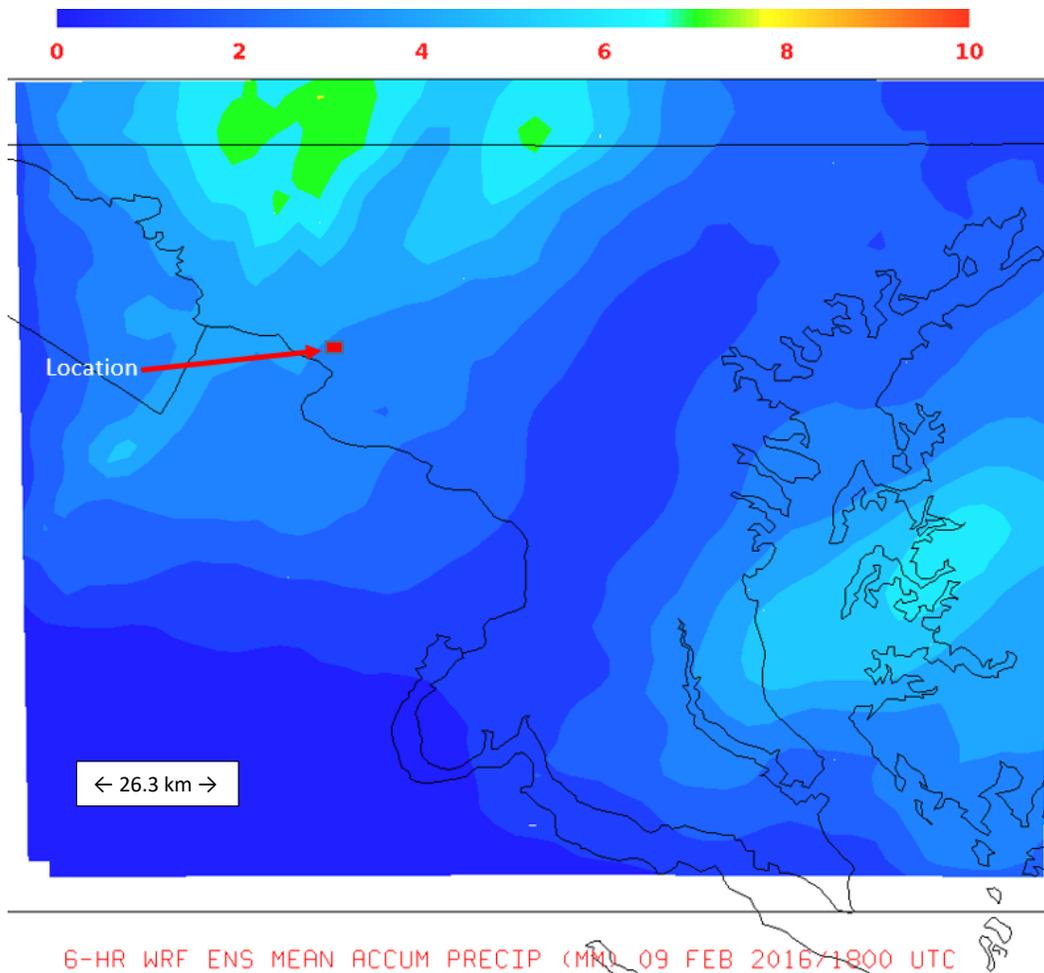
**Fig. 7 Time series of accumulated precipitation (millimeters) from the 28 WRE-N ensemble members and the observations for the period 1200 to 1800 UTC for the location shown in Figs. 5 and 6**

The time series shows the uncertainty of the precipitation forecasts at each hour during the period. The uncertainty increases with lead time. The greatest uncertainty occurs at 1800 UTC, when the accumulated precipitation ranges between approximately 5 and 10 mm with the entire ensemble overpredicting the precipitation compared with the observation. The overprediction is a consequence of the displacement of the forecast maximum feature north and west of the corresponding observed feature. In addition to the displacement error, there are errors in the structure of the maximum feature, which are difficult to assess using this analysis type.

The next time series focused on a geographic location inside the observed maximum feature shown in Fig. 8. The corresponding time series data from the ensemble was extracted from the same geographic location, shown in Fig. 9, but was well south of its corresponding maximum feature.

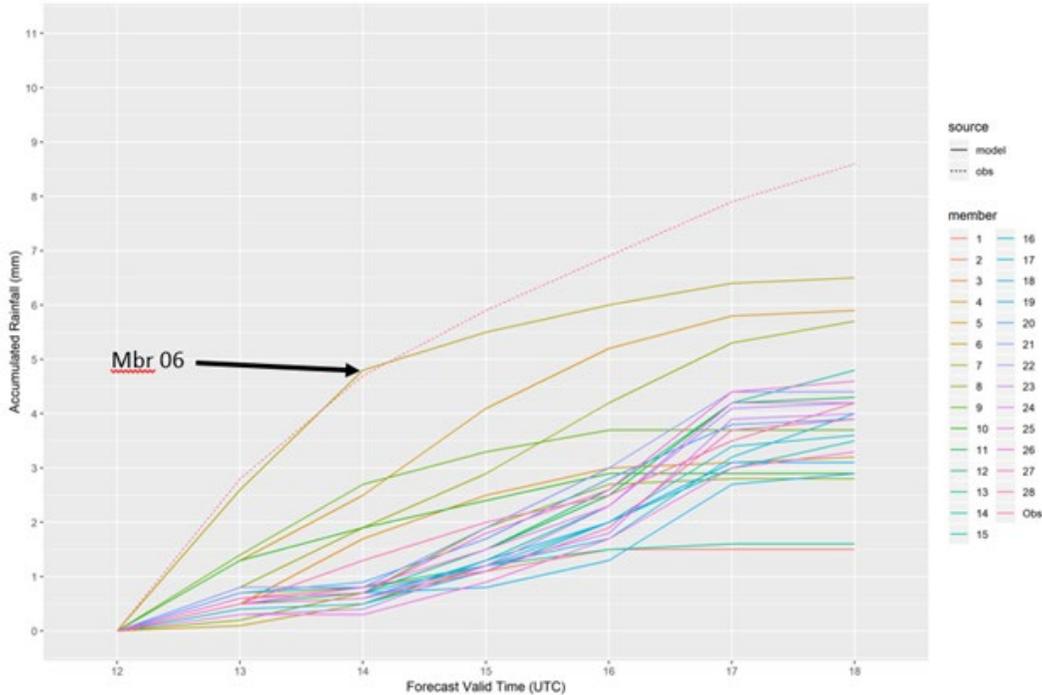


**Fig. 8** Observed 6-h accumulated precipitation (millimeters) at 1800 UTC from the 12-h forecast with location marked where time series data in the following plot were extracted



**Fig. 9** WRE-N ensemble mean 6-h accumulated precipitation (millimeters) at 1800 UTC with location marked where time series data in Fig. 10 were extracted

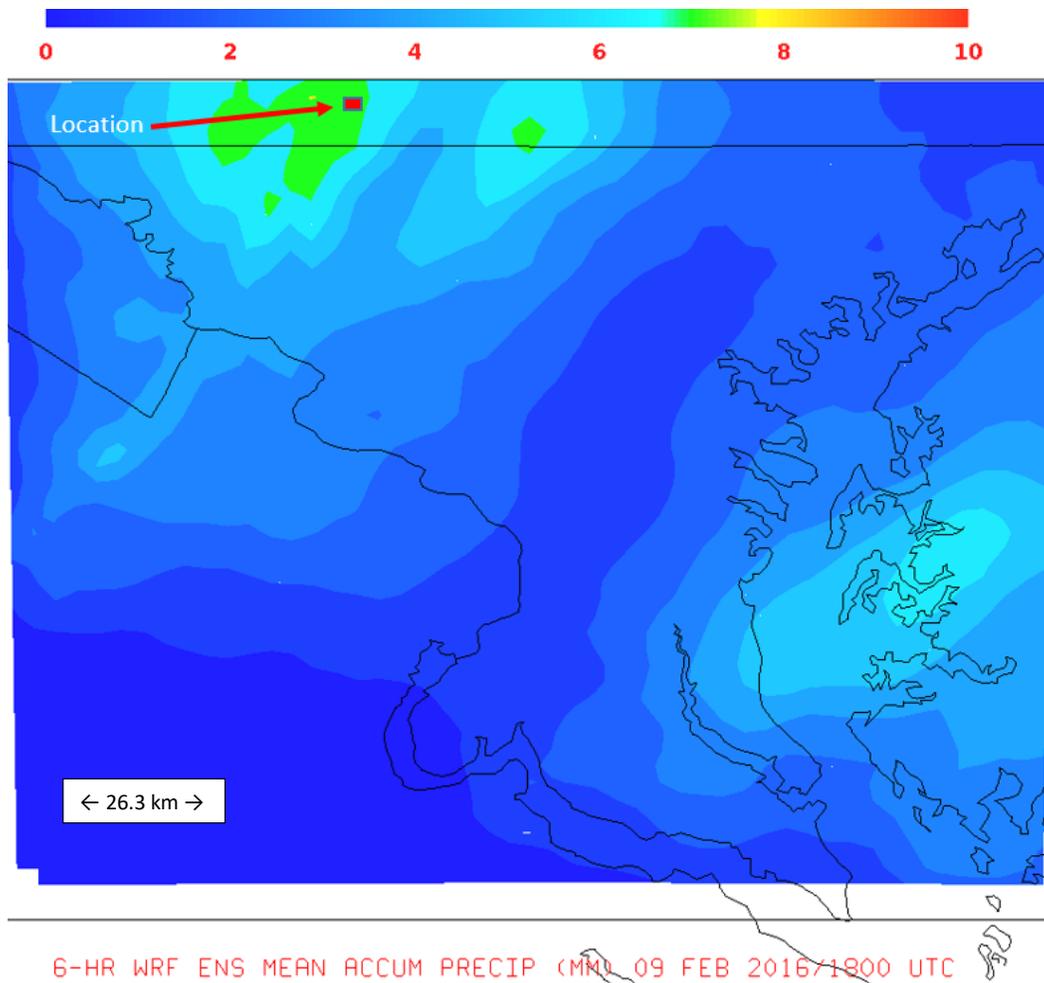
The time series in Fig. 10 shows the uncertainty of the precipitation forecasts at each hour during the period. The uncertainty increases with lead time, with the greatest uncertainty occurring at 1800 UTC, when the accumulated precipitation ranges between approximately 1.5 and 6.5 mm, with the entire ensemble underpredicting the precipitation compared with the observation. The underprediction is a consequence of the displacement of the forecast maximum feature north and west of the corresponding observed feature. The performance of ensemble member 06 is noteworthy because its prediction is nearly perfect from 1200 to 1500 UTC, and clearly superior to the other members, out to 1800 UTC. Further analysis of the performance of member 06 was accomplished and is reported in a later section of this report.



**Fig. 10** Time series of accumulated precipitation (millimeters) from the 28 WRE-N ensemble members and the observations for the period 1200 to 1800 UTC for the location shown in Figs. 8 and 9

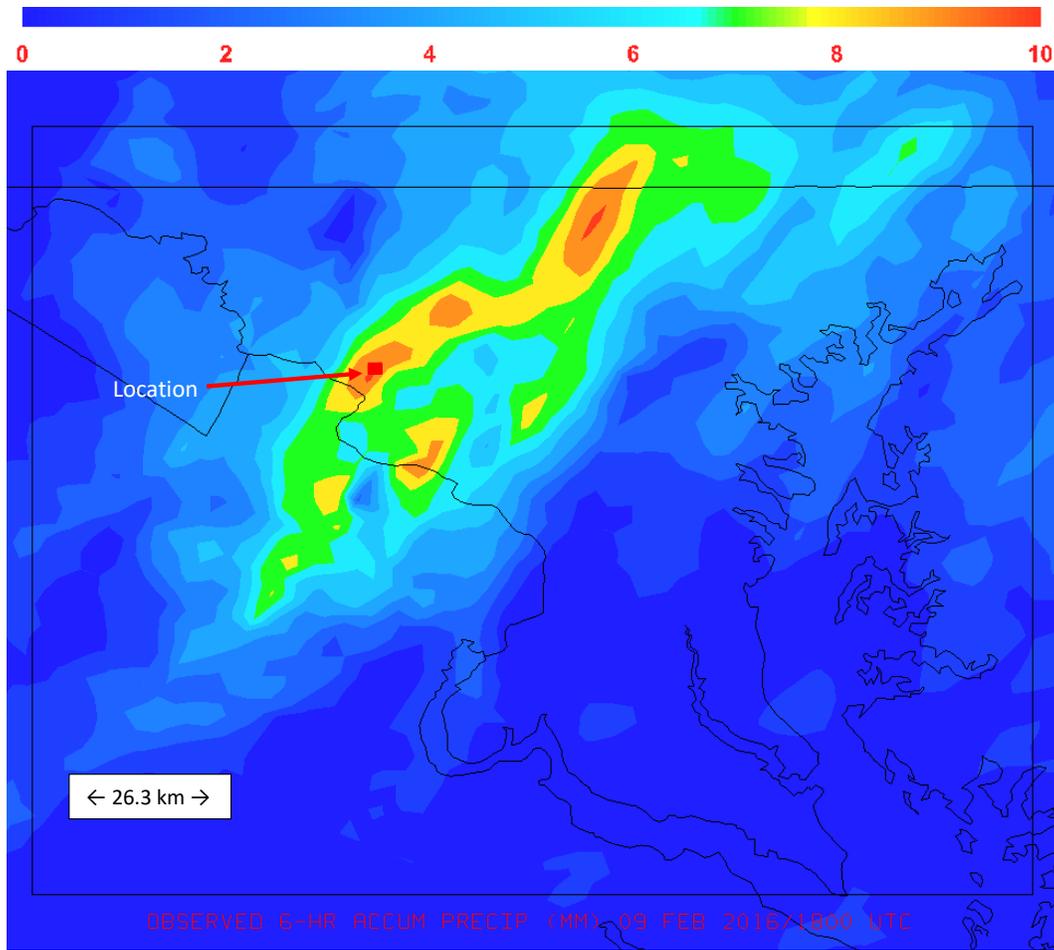
While this approach revealed spatial variability in model bias and displacement errors, there was no information about the skill of the model in predicting the maximum feature itself. To shed some light on this aspect of performance, it was decided to conduct an analysis of time series data for geographic points that were located inside the maximum features of both the ensemble and the observations. This way, a comparison could be made of the model’s ability to predict its maximum feature in a manner that attempts to remove the contribution from the displacement error.

Figure 11 shows the WRE-N ensemble mean 6-h accumulated precipitation (millimeters) at 1800 UTC with the location within the maximum forecast precipitation feature from which the time series data were extracted.



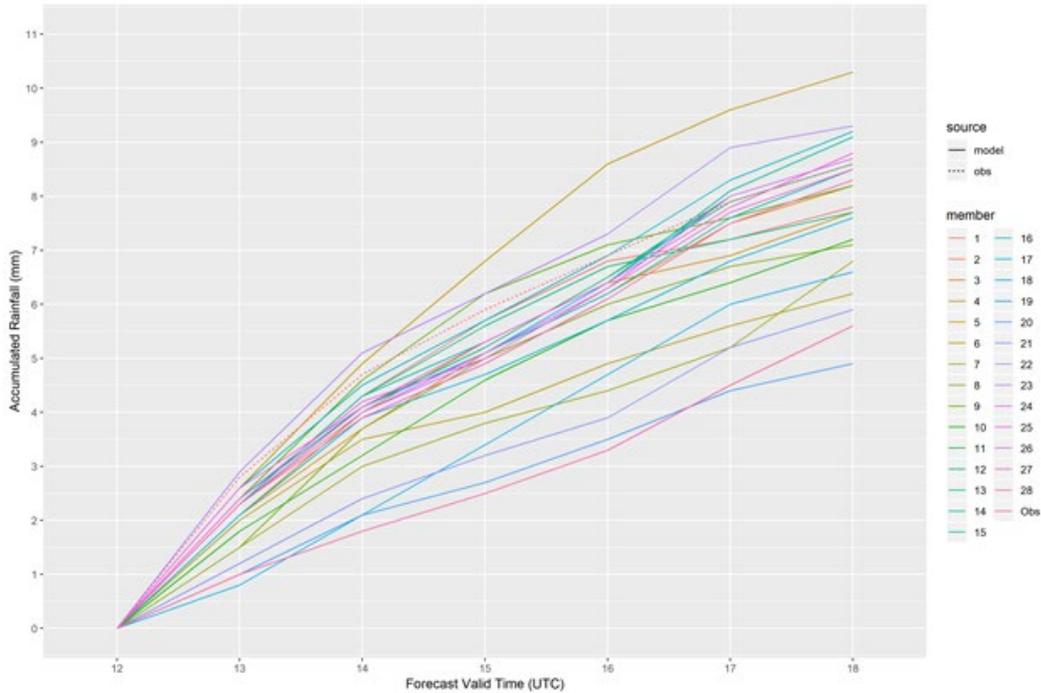
**Fig. 11** WRE-N ensemble mean 6-h accumulated precipitation (millimeters) at 1800 UTC from the 12-h forecast with location marked where WRE-N time series data in the following plot were extracted

Figure 12 shows the observed 6-h accumulated precipitation (millimeters) at 1800 UTC with the location within the maximum precipitation feature from which the observed time series data were extracted.



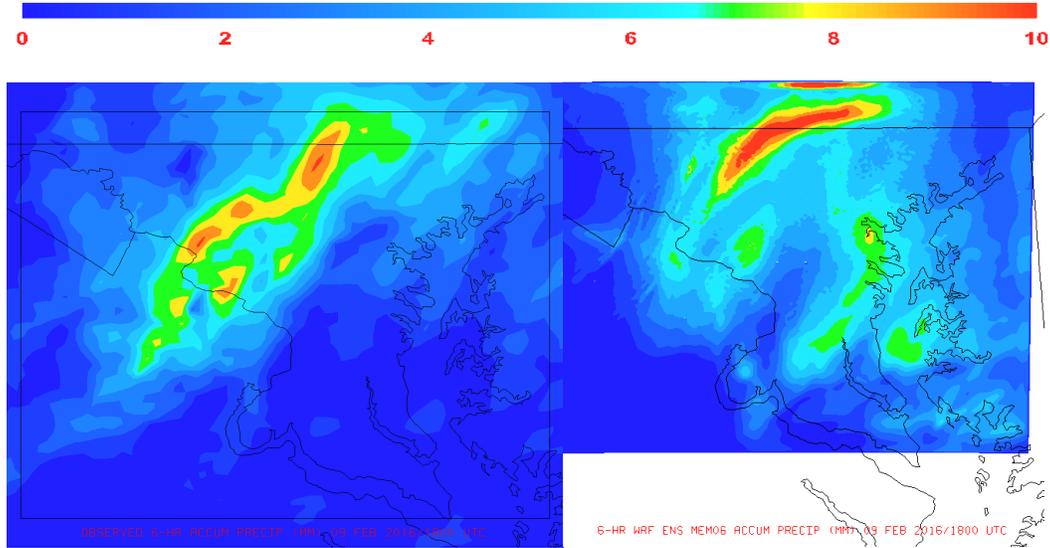
**Fig. 12** Observed 6-h accumulated precipitation (millimeters) at 1800 UTC with location marked where time series data in Fig. 13 were extracted

The time series shown in Fig. 13 shows the uncertainty of the precipitation forecasts at each hour during the period. The uncertainty increases with lead time, with the greatest uncertainty occurring at 1800 UTC, when the accumulated precipitation ranges between approximately 5 and 10 mm, with most members underpredicting the precipitation compared with the observation. The bias magnitude was judged to be less than the previous time series analyses. A few of the members show an overprediction tendency, but again, of less magnitude. The closer agreement between the ensemble and the observations appears to be a consequence of the removal of the displacement error. It also shows that the ensemble’s handling of the maximum feature itself is reasonable. Without a more quantitative way to assess the ensemble skill, it is difficult to judge how well it performs using only this analysis.



**Fig. 13 Time series of accumulated precipitation (millimeters) from the 28 WRE-N ensemble members for the location shown in Fig. 11 and the observations for the location shown in Fig. 12 both for the period 1200 to 1800 UTC**

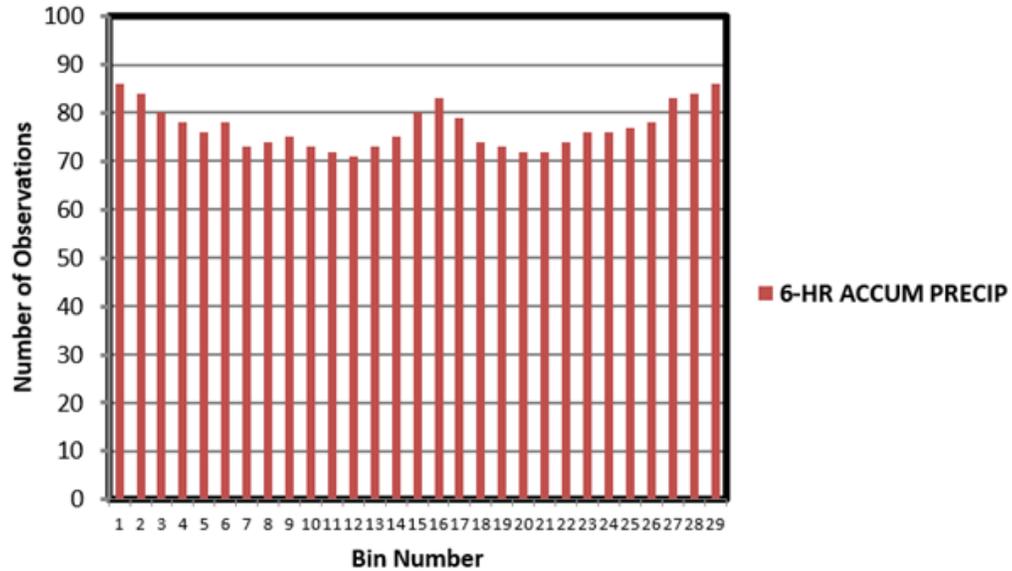
The noteworthy performance of ensemble member 06, which used the GEFS 0.5° member 05 for initial and lateral boundary conditions, was compared with the other members revealed in the time series analysis (see Fig. 10). Figure 10 was based on where the point of interest was located to inside the observed maximum feature, which was considerably south and east of the corresponding feature in the ensemble mean forecast. Figure 14 shows the closeness on the placement of the maximum feature between the ensemble mean and the observations at 1800 UTC. Even the structure of the highest of the maximum precipitation appears quite similar. This particular case was subjected to further analysis using a spatial verification technique that provided a more quantitative assessment of the member 06 skill. The analysis and results will follow later in this report.



**Fig. 14 Ensemble member 06 and observed 6-h accumulated precipitation (millimeters) at 1800 UTC**

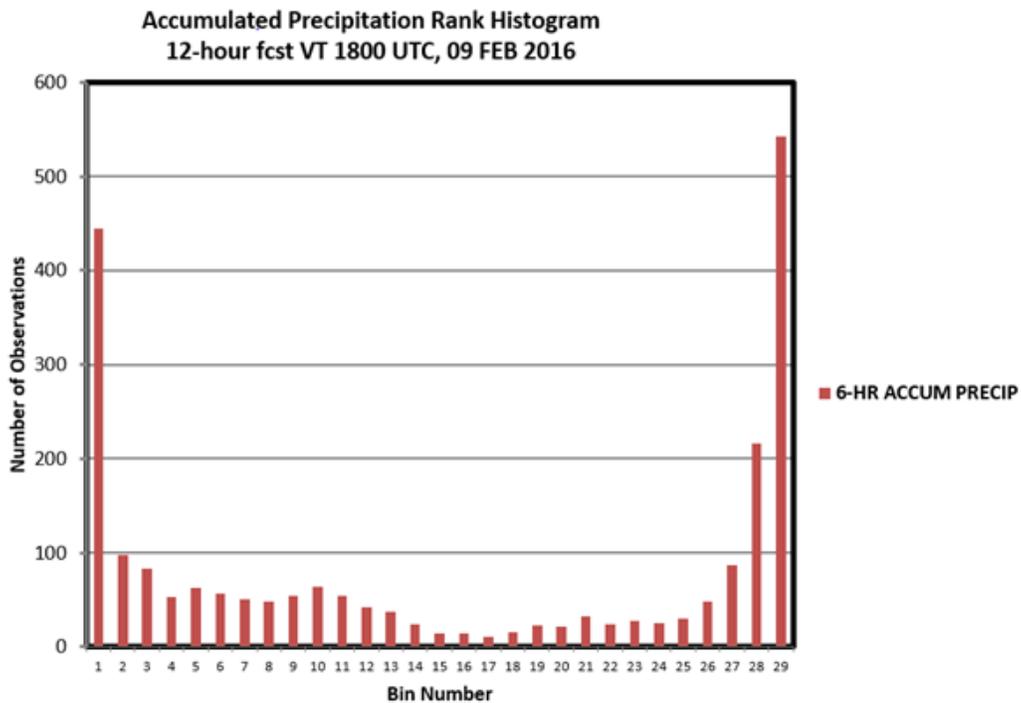
Continuing the analysis of the ensemble’s skill as a whole, a rank histogram, also known as a “Talagrand diagram”, was generated using MET E-S. Rank histograms check where the verifying observation usually falls with respect to the ensemble forecast data, which is arranged in increasing order at each grid point using bins. The number of bins is equal to the number of ensemble members, with the lowest bin representing the value from the ensemble member having the lowest value at a given observation point, and the highest bin representing the value from the ensemble member having the highest value at that observation point. The number in the lowest bin (bin 1) represents the number of observations that were lower than any of the ensemble members. In an ensemble with a uniform rank histogram, each member represents an equally likely scenario, so the observation value is equally likely to fall into any of the bins. In this case the histogram shows a flat appearance with all bins being of approximately equal height (Hamill 2001). The implication of a uniform rank histogram is that the spread of forecast accumulated precipitation from the ensemble replicates the range of accumulated precipitation values in the observations. Figure 15 shows an example of a uniform rank histogram.

**Example of a Uniform Rank Histogram**



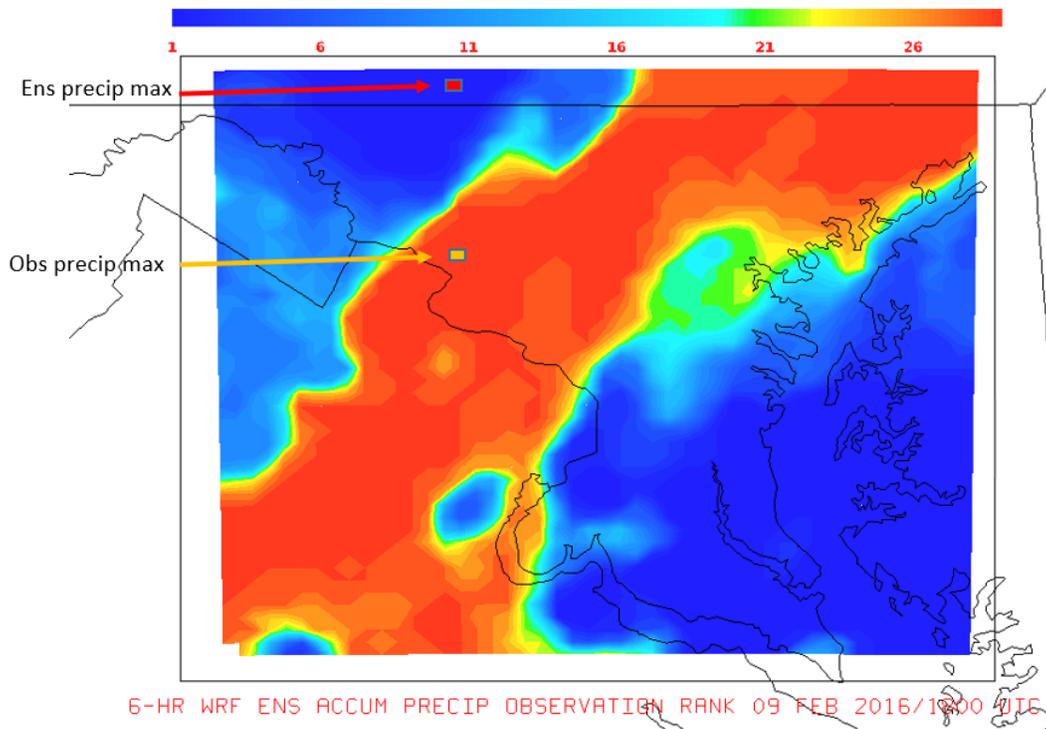
**Fig. 15 Uniform rank histogram example**

The rank histogram for the WRE-N ensemble shown in Fig. 16 was produced for the 12-h (6-h free forecast) forecast and shows the observation ranks for 6-h accumulated precipitation



**Fig. 16 Rank histogram for the WRE-N ensemble 6-hr accumulated precipitation forecast at 1800 UTC**

The rank histogram shows a U-shape, which is indicative of an ensemble with insufficient spread. In this case, there are many 6-h accumulated precipitation observations that fall outside the extremes of the ensemble accumulated precipitation. This pattern can also happen when there are negative and positive biases present. (Hamill 2001). From the previous analysis of the time series at two different locations, the presence of negative and positive biases was confirmed and is the most probable cause of the U-shaped rank histogram. A particularly useful way to depict this spatial bias variation is to plot the observation ranks over the ensemble domain. Figure 17 shows a plot of observation ranks (1–29) for 6-h accumulated precipitation, with the locations where the time series data were extracted corresponding to the ensemble mean maximum feature and the observation maximum feature along with the rank histogram.



**Fig. 17 WRE-N 6-h accumulated precipitation observation ranks at 1800 UTC**

The color scale shows the observation ranks assigned so that the dark red color is for the highest ranks (near 29) corresponding to the far right end of the rank histogram, and the dark blue color is for the lowest ranks (near 1) corresponding to the far left end of the rank histogram. Thus the red areas indicate the presence of precipitation underforecasting and the dark blue areas indicate overforecasting. Note the lack of extensive areas with the colors between red and dark blue, which represent the middle ranks. This sparseness was caused by the lack of observations whose values fall in the central ranks, which is evident from the middle part of the

rank histogram. This “couplet” pattern of negative and positive bias juxtaposed in the domain is indicative of a significant displacement error in the location of the precipitation maximum feature and is corroborated by the results from the previous analysis of the ensemble mean and observed precipitation and the time series.

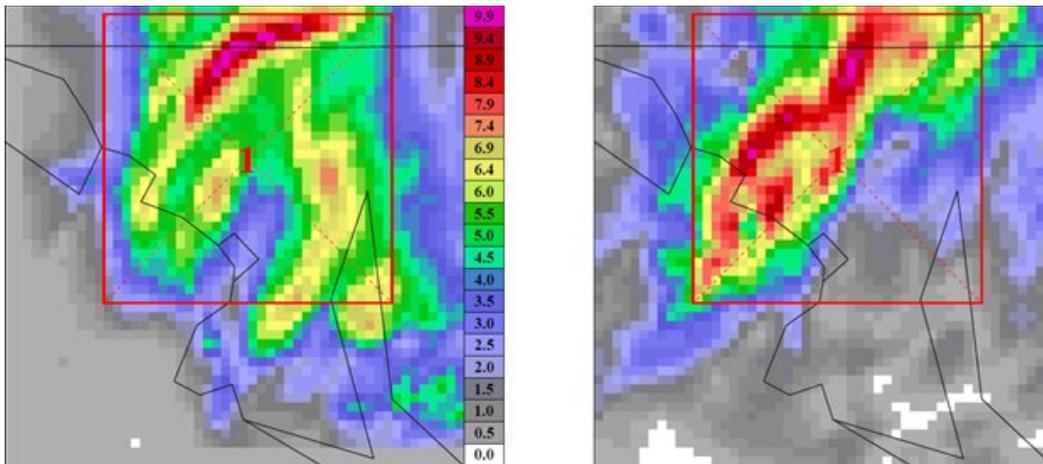
### **3. Application of the Scale-Decomposition Technique**

---

To obtain more-quantitative evidence of the ensemble’s skill in predicting accumulated precipitation, a spatial method for verification was considered. Spatial verification approaches, in contrast to grid-to-point verification approaches, focus on features and coherent spatial structures that characterize meteorological fields. Since these approaches account for the intrinsic spatial correlation existing between nearby grid points, they do not suffer from double penalty errors that impact grid-to-point verification. Spatial verification approaches take into account the observation and forecast time–space uncertainties, and seek to provide feedback on the forecast error in physical terms. Within the spatial verification approaches, the Intensity-Scale technique belongs to the scale-decomposition (or scale-separation) verification approaches. The scale-decomposition approach enables users to perform the verification on different spatial scales. Weather features on different scales (e.g., frontal systems versus convective showers) are often driven by different physical processes. Verification on different spatial scales can therefore provide deeper insights into model performance. The spatial scale components are usually obtained by applying a single band spatial filter to the forecast and observation fields (e.g., Fourier decomposition, wavelets). The scale-decomposition approaches measure error, bias, and skill of the forecast on each different scale component. The scale-decomposition approach provides feedback on the scale dependency of the error and skill and on the capability of the forecast for reproducing the observed scale structure (NCAR 2016b).

The particular tool which was used to perform scale-decomposition verification was the MET Wavelet-Stat (W-S) tool. This tool uses the intensity-scale technique developed by Casati et al. (2004), the metrics of which appear in the Appendix. The Intensity-Scale technique evaluates the forecast skill as a function of the intensity or magnitude of the field values and of the spatial scale of the error. The scale components are decomposed by applying a 2-D Haar wavelet filter (Casati et al. 2004). Note that wavelets, because of their locality, are suitable for representing discontinuous fields such as precipitation. The technique uses a categorical approach, which is a long-established and reliable approach, suitable for non-normally distributed variables such as precipitation. The Intensity-Scale technique was specifically designed to handle the difficult characteristics of precipitation fields and for the verification of spatial precipitation forecasts (NCAR 2016b).

To evaluate the use of the Intensity-Scale method as a spatial technique for quantifying the skill of the WRE-N ensemble, the W-S tool was used to apply that method to the 6-h accumulated precipitation field from ensemble member 06 and the Stage IV precipitation observations. The W-S tool allows the user to adjust the position of the tile to focus the verification on a specific feature of interest. In this case, the feature was the precipitation maximum. Figure 18 shows the accumulated precipitation forecast from member 06 (left image) and the observed precipitation (right image) present in the inner domain. Also shown is a  $32 \times 32$  grid square tile, outlined in red, where a 2-D Haar wavelet filter was applied for decomposition into orthogonal component images for the different scales. The tile was positioned to encompass both maxima. Note that member 06 spreads the maximum feature out over a larger area to the northwest and southeast of the observed feature, which leads to errors of a larger scale. These errors are discussed later.



**Fig. 18** Graphic produced by W-S showing the 6-h precipitation (millimeters) from (left) member 06 and (right) observations, along with a  $32 \times 32$  grid square tile where a 2-D Haar wavelet filter was applied

The scale-decomposition process of W-S involves application of the thresholds to the forecast and observed precipitation fields, which transforms them into binary fields by assigning a value of 1 where there is an event, and a value of 0 where there is no event. Figure 19 graphically shows the difference (forecast minus observation, or F-O) between the binary forecast field and the binary observations field at the original 4-km resolution. The particular threshold value for this graphic is  $\geq 3.0$  mm of 6-h accumulated precipitation. The value of the differences for the events are either +1, 0, or -1, which accounts for the lack of differences with lighter color shades leaving only the darkest red or blue shaded features in the domain. The value of the frequency bias ( $< 1$ ) indicates the prevalence of event underforecasting. The Intensity Skill Score (ISS) ranges from  $-\infty$  (no skill, if negative) to +1 (good skill, if positive). Note the large (red) area of overforecast precipitation near the

southeast corner of the domain and a smaller area of overforecast precipitation to the northwest and an area of underforecast (blue) precipitation along the western domain boundary. The two overforecast areas are an indication of a spatial displacement error.

APCP\_06(\*,\*) >=3.0 vs APCP\_06(\*,\*) >=3.0  
 Tile 1, Binary, Difference (F-0)

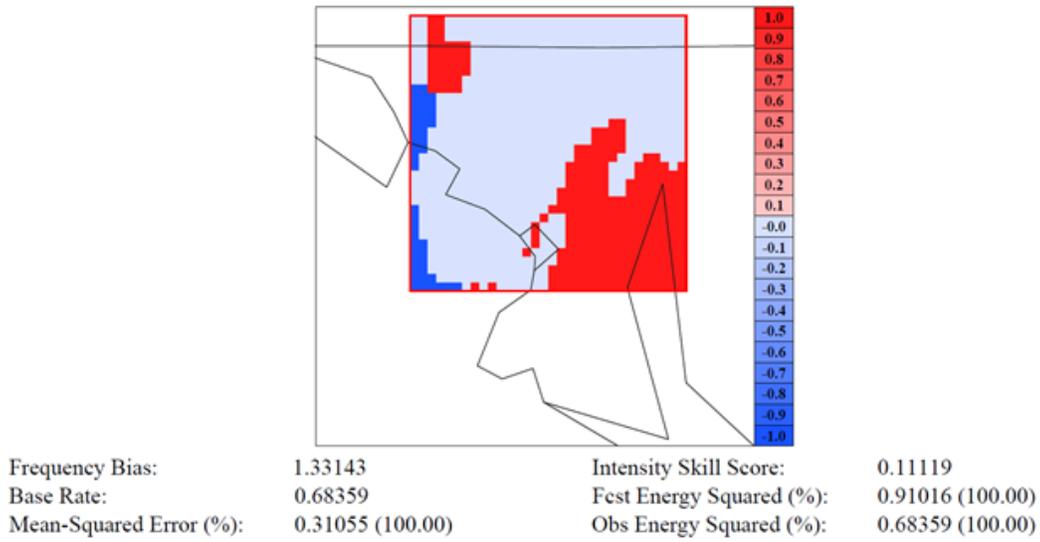
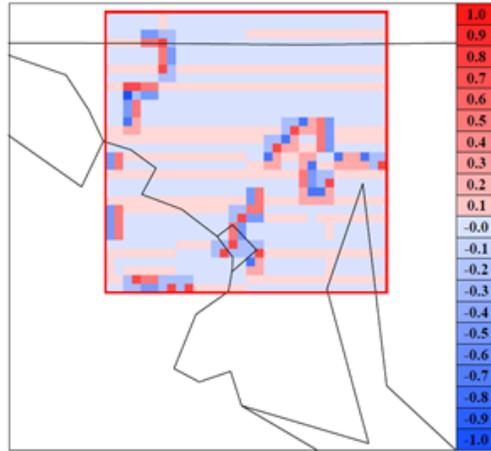


Fig. 19 W-S output graphic of the binary difference field for threshold value  $\geq 3.0$  mm

W-S computes the same statistics for each decomposed scale. Figure 20 shows the first spatial scale component (Scale 1) of the binary field difference that was obtained by taking the difference (F-O) between the spatial scale components of the binary forecast and observation fields per Fig. 19. Because the wavelet transform is a linear operator, decomposing the binary difference field into spatial scale components is equal to taking the difference between spatial scale components of the forecast and observation fields (Casati 2010). The decomposition starts by computing the mean component function (MCF) of the accumulated precipitation field after spatial averaging over an  $8 \times 8$ -km subarea of the binary difference field. This is sometimes referred to as the “father” wavelet. The variation-around-the-mean-component function (VCF), sometimes called the “mother” wavelet, is also computed. The sum of the MCF and VCF produces the original binary difference field (Casati 2004). Scale 1 (Fig. 20) is the first VCF wavelet component with a corresponding resolution of 4 km. The magnitude of the VCF wavelet is thus  $\leq |1|$ , which accounts for the presence of the full range of blue and red color shades present in the difference field in the graphic. The value of the frequency bias (1.33) indicates the prevalence of event overforecasting, which is identical in value at all scales. The ISS indicates good skill at 4-km resolution.

APCP\_06(\*,\*) >=3.0 vs APCP\_06(\*,\*) >=3.0  
 Tile 1, Scale 1, Difference (F-0)

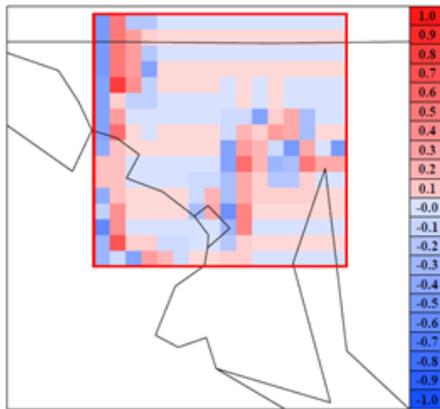


Frequency Bias:	1.33143	Intensity Skill Score:	0.42982
Base Rate:	0.68359	Fcst Energy Squared (%):	0.01270 (1.39)
Mean-Squared Error (%):	0.03320 (10.69)	Obs Energy Squared (%):	0.02246 (3.29)

Fig. 20 W-S output graphic of the first spatial scale (4 km) for threshold value  $\geq 3.0$  mm

Figures 21–25 show the graphical representation of the accumulated precipitation error for the next five scales (2–6), which are 8, 16, 32, 64, and 128 km.

APCP\_06(\*,\*) >=3.0 vs APCP\_06(\*,\*) >=3.0  
 Tile 1, Scale 2, Difference (F-0)



Frequency Bias:	1.33143	Intensity Skill Score:	0.28727
Base Rate:	0.68359	Fcst Energy Squared (%):	0.02759 (3.03)
Mean-Squared Error (%):	0.04150 (13.36)	Obs Energy Squared (%):	0.02002 (2.93)

Fig. 21 W-S output graphic of the second spatial scale (8 km) for threshold value  $\geq 3.0$  mm

APCP\_06(\*,\*) >=3.0 vs APCP\_06(\*,\*) >=3.0  
 Tile 1, Scale 3, Difference (F-0)

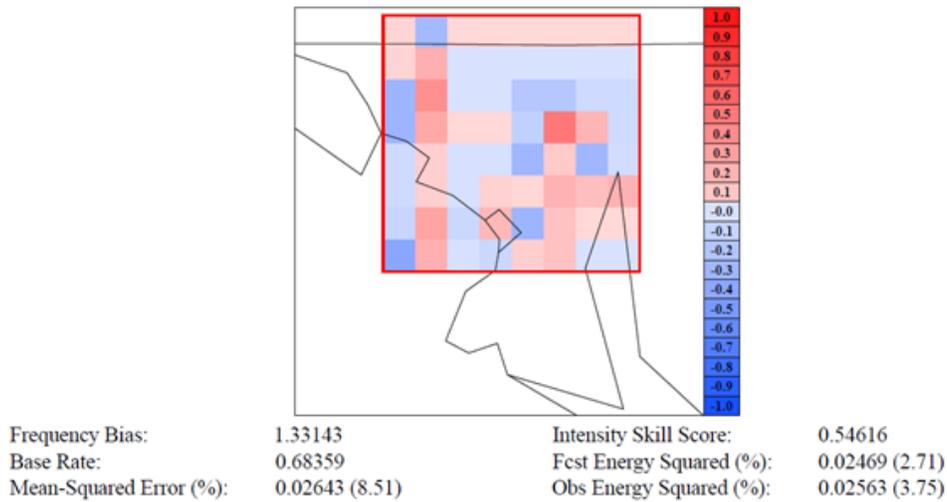


Fig. 22 W-S output graphic of the third spatial scale (16 km) for threshold value  $\geq 3.0$  mm

APCP\_06(\*,\*) >=3.0 vs APCP\_06(\*,\*) >=3.0  
 Tile 1, Scale 4, Difference (F-0)

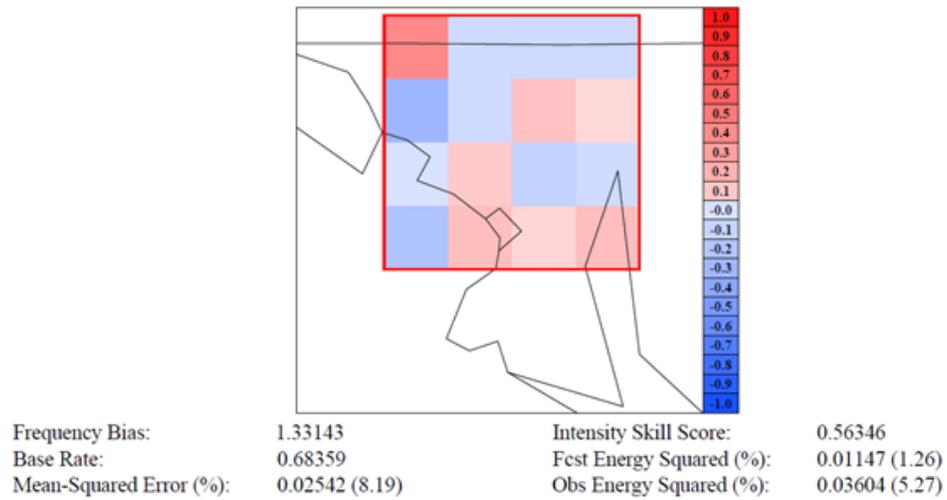


Fig. 23 W-S output graphic of the fourth spatial scale (32 km) for threshold value  $\geq 3.0$  mm

APCP\_06(\*,\*)  $\geq 3.0$  vs APCP\_06(\*,\*)  $\geq 3.0$   
 Tile 1, Scale 5, Difference (F-0)

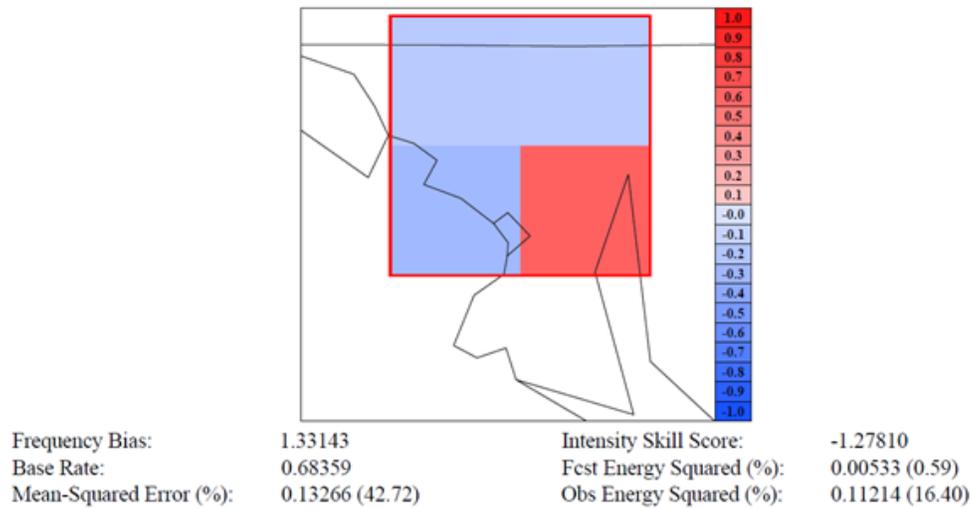


Fig. 24 W-S output graphic of the fifth spatial scale (64 km) for threshold value  $\geq 3.0$  mm

APCP\_06(\*,\*)  $\geq 3.0$  vs APCP\_06(\*,\*)  $\geq 3.0$   
 Tile 1, Scale 6, Difference (F-0)

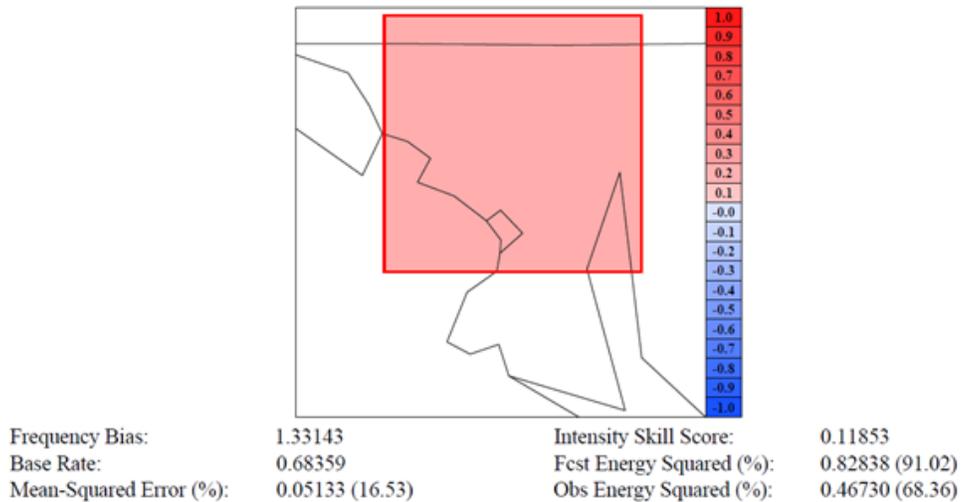


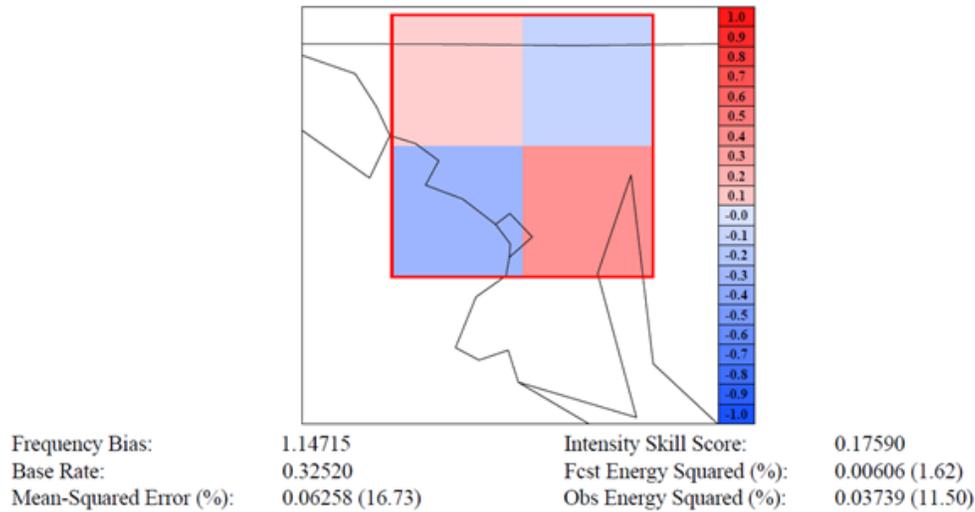
Fig. 25 W-S output graphic of the sixth spatial scale (128 km) for threshold value  $\geq 3.0$  mm

At each scale, the VCF wavelet captures the variation component of the accumulated precipitation binary difference field, and the MCF wavelet captures the mean component of the accumulated precipitation field after spatial averaging over a subarea of the binary difference field given by squaring the value of the scale. The magnitude of the VCF wavelet is thus  $\leq |1|$ , which accounts for the

presence of the full range of blue and red color shades present in the difference field in each graphic. The sum of the VCF and MCF wavelets produces the original binary difference field. The value of the frequency bias (1.33) indicates the prevalence of event overforecasting, which is identical in value at all scales. The ISS value for 8-, 16-, and 32-km scales shows good skill. At the 64-km scale there is a noteworthy change in the ISS from positive to negative despite the low threshold value and large scale, which goes against the expected trend (Fig. 24). This trend is most likely the result of the detection of the large error (64 km) due to the displacement of the forecast maximum accumulated precipitation feature to the northwest and southeast of the observed maximum. Figure 24 displays this pattern via the juxtapositioned squares in the lower part of the domain, which exhibit a sharp color contrast difference, with overforecasting in the right square and underforecasting in the left square. This is also evident in the images in Fig. 18 when comparing the forecasted and observed precipitation maxima locations. Casati et al. (2004) noted a very similar occurrence of an isolated negative ISS value surrounded by positive values in their tabular results, which was attributed to displacement error. At the 128-km scale (Fig. 25), the ISS changes back to a positive value indicating good skill, though not at the level which might have been expected given the higher ISS values at Scales 1–4 before the reversal at Scale 5. This is probably due to the impact of the displacement error at this scale.

A similar pattern is present in the results for Scales 1–6 at the next threshold value of  $\geq 6.0$  mm. The overforecast tendency is still present, but the magnitude is smaller due to the increase in threshold value. The displacement error at Scale 5 (64 km) is still evident, but the signal is not as strong as the lower threshold value of  $\geq 3.0$  mm. Figure 26 shows the spatial scale Component 5, which is the fifth VCF wavelet component with a corresponding resolution of 64 km. The ISS score at this threshold value and spatial scale would be expected to have been even more positive had this displacement error not been so dominant.

APCP\_06(\*,\*) >=6.0 vs APCP\_06(\*,\*) >=6.0  
 Tile 1, Scale 5, Difference (F-0)



**Fig. 26 W-S output graphic of the fifth spatial scale (64 km) for threshold value  $\geq 6.0$  mm**

The results for Scales 1–6 at the highest threshold value of  $\geq 9.0$  mm showed a more consistent trend of increasing ISS score with scale size. The value of the frequency bias increased over those present at the lower thresholds, indicating an increase in the overforecast tendency. The size of the precipitation features defined by the  $\geq 9.0$  mm threshold, as shown in the binary difference field, are significantly smaller than those at the lower thresholds, which is expected at higher accumulated precipitation amounts. At the 64-km scale (Fig. 27), the high positive value of the ISS would suggest that the displacement error present at the lower threshold values, is not present at this higher threshold, thus indicating that there was no significant displacement of the forecast higher threshold precipitation features from the observed features. It is possible that there were no features with an area of 64 km<sup>2</sup> whose accumulated precipitation value met or exceeded the 9-mm threshold value. If this is the case, it would be difficult to rule out the possibility that there may have been significant displacement of features defined by this threshold, since they are not detectable at this scale. This situation needs further investigation to clarify if, and how much, displacement actually occurred.

APCP\_06(\*,\*) >=9.0 vs APCP\_06(\*,\*) >=9.0  
 Tile 1, Scale 5, Difference (F-0)

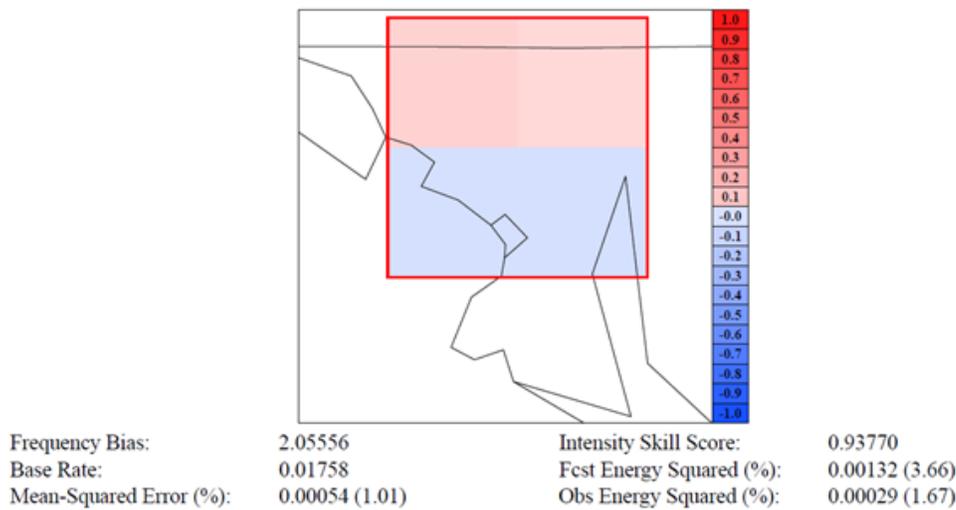


Fig. 27 W-S output graphic of the fifth spatial scale (64 km) for threshold value  $\geq 9.0$  mm

Table 1 shows the ISS scores for spatial scales 4–128 km for the three threshold values. The negative ISS value at  $\geq 3.0$  mm for the 64-km scale size is anomalous relative to positive scores at higher and lower scales, and for the same scale at the next higher threshold. There is a similarly anomalous depression of the ISS value at the same scale for the  $\geq 6.0$ -mm threshold relative to the ISS values at the next lower and higher scales (32 and 128 km). The combination of the noted anomalous negative ISS and the depressed ISS values shows the presence of the displacement error at the 64-km scale size.

Table 1 Intensity Skill Scores for spatial scales 4 to 128 km and threshold values of  $\geq 3.0$ ,  $\geq 6.0$ , and  $\geq 9.0$  mm

Threshold (mm)	Scale (km)					
	4	8	16	32	64	128
$\geq 3.0$	0.43	0.29	0.55	0.56	-1.28	0.12
$\geq 6.0$	-0.09	0.00	-0.28	0.30	0.18	0.97
$\geq 9.0$	-1.04	-0.95	-0.53	0.45	0.94	0.96

Summarizing the results from applying the scale-decomposition method of verification show that ensemble member 06 performed well at  $\geq 3.0$ -mm for scales 4–32 km, then poorly at the 64-km scale, with a only a modest recovery to a good score at 128 km. Increasing the threshold value to  $\geq 6.0$  mm showed poor scores for

scales 4–16 km, then improved scores for scales 32–128 km, though the score at 64 km looks anomalously lower than would otherwise be expected given the good score at the next lower scale (32 km) and the excellent score at the next higher scale (128 km). At the highest threshold value, the scores at the smallest scales (4–16 km) were poor but increased significantly for scales 32–128 km. The anomalous negative score for the  $\geq 3.0$ -mm threshold and 64-km scale and, to a lesser extent, the 128-km scale was most likely the result of the detection of the large error at both scales due to the displacement of the precipitation maximum feature. This displacement error may have similarly impacted the score at the  $\geq 6.0$ -mm threshold and 64-km scale, which was also anomalously lower than nearby scores at higher and lower scales. There is no evidence of the displacement error at the highest threshold value judging from the scores alone. Based on these results, member 06 showed good skill at all scales between 4 and 32 km at the lowest threshold (3.0 mm). This good performance was repeated at the next two threshold values (6 and 9 mm) but at larger scales of 32–128 km. There is an error in the placement of the precipitation maximum feature, which is manifest mainly in the score for the lowest threshold at the 64-km scale and to a limited extent in the score at the same scale at the threshold value of  $\geq 6.0$  mm. Further investigation is needed to clarify whether the displacement error at the highest threshold value was simply not detected for features with areas  $64 \text{ km}^2$  or larger due to the lack of such features with accumulated precipitation values that met or exceeded this threshold.

#### **4. Conclusions and Summary**

---

---

Uncertainty quantification of the WRE-N ensemble using 1) the mean ensemble and observed 6-h accumulated precipitation graphics, rank histogram, and 2-D observation ranks from MET E-S, and 2) time series of 6-h precipitation from key locations in the inner model domain provided a qualitative confirmation of the presence of uncertainty in the location of the precipitation maximum feature that was attributable to a spatial displacement error. These tools also confirmed that the spread in ensemble-accumulated precipitation was due to either underdispersion relative to the range of accumulated precipitation values from the observations or the presence of spatial variations in the model bias resultant from the displacement error. These results provide some insight into the ensemble performance as a whole but are not very quantitative. The application of the scale-decomposition technique to quantify the uncertainty of a single member of the ensemble provided significantly more details on the performance of that member, which was not achievable using results from E-S. The MET W-S tool uses the Intensity-Scale method and provides an assessment of model skill as a function of threshold values

and spatial scales by providing an ISS for each combination of threshold and scale. The tool enabled the separation of the larger displacement errors from the smaller scale errors attributable to smaller-scale processes. The separation was characterized by negative ISS and pointed out the strengths of the model forecast, where the ISS is positive. The tool enabled the determination of frequency bias at all thresholds, which showed a dominant tendency for overforecasting 6-h accumulated precipitation. Despite the displacement error at the 64- and 128-km scales, the model showed good skill at all scales between 4 and 32 km at the lowest threshold (3 mm). This good performance was repeated at the next two threshold values (6 and 9 mm). At larger scales, 32–28 km, there is some evidence impact due to the displacement error at 64 km and a possibility of the impact at the highest threshold that cannot be detected using these results. This would imply that for the higher thresholds, the model lacks some skill in predicting smaller-scale convection features with scales less than 16 km. These conclusions about model performance are only based on one case study and not intended as a final assessment. More case studies are needed to obtain conclusive assessment results.

In summary, the results obtained through the use of the scale-decomposition technique look promising for quantifying the uncertainty of each ensemble member individually as a function of the spatial scale and the threshold value. The next step toward quantifying the ensemble uncertainty is to apply this technique to all members of the ensemble. Results from all members may allow the identification of specific model configurations to use in an ensemble that would enable the ensemble to better characterize the true uncertainty.

## 5. References

---

- Casati B, Ross G, Stephenson D. A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteor Appl.* 2004;11:141–154.
- Casati B. New developments of the intensity-scale technique within the Spatial Verification Methods Intercomparison Project. *Wea Forecast.* 2010;25:113–143.
- Dumais R, Cai H, Raby J, Reen B. Use of a high resolution WRF-ARW ensemble to provide short-range guidance for a complex mixed precipitation event near the Washington, DC, area – experimental design. White Sands Missile Range (NM): CCDC Army Research Laboratory (US); 2018 Oct. Report No.: ARL-TN-0921.
- Dumais Jr RE, Kirby SF, Passner JE, Knapp DI. Development of a nowcast modeling system to support Army aviation forecasts. Presented at the 16th Conference on Aviation, Range, and Aerospace Meteorology, 93rd Annual American Meteorological Society Conference; 2013 Jan 6–10; Austin, TX.
- Hamill TM. Interpretation of rank histograms for verifying ensemble forecasts. *Mon Wea Rev.* 2001;129:550–560.
- [UCAR] Integrated data viewer (IDV). Boulder (CO): University Corporation for Atmospheric Research; 2018 [accessed 2018 July]. <https://www.unidata.ucar.edu/software/idv/>.
- Mittermaier MP. Using an intensity-scale technique to assess the added benefit of high resolution model precipitation forecasts. *Atmos Sci Lett.* 2006;7:36–42.
- [NCAR] National Center for Atmospheric Research. Unified post processor (UPP). Boulder (CO): NCAR; 2016a [accessed 2018 Oct 15]. [https://dtcenter.org/upp/users/docs/user\\_guide/V3/upp\\_users\\_guide.pdf](https://dtcenter.org/upp/users/docs/user_guide/V3/upp_users_guide.pdf).
- [NCAR] National Center for Atmospheric Research. Model evaluation tools (MET). ver. 5.2. Boulder (CO): NCAR; 2016b [accessed 2019 Aug 15]. <https://dtcenter.org/community-code/model-evaluation-tools-met>.
- Nelson BR, Prat OP, Seo D-J, Habib E. Assessment and implications of NCEP stage IV quantitative precipitation estimates for product intercomparisons. *Wea Forecast.* 2016;31:371–394.
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2019 [accessed 2019 Sep 30]. <https://www.R-project.org/>.

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Duda MG, Huang X-Y, Wang W, Powers JG. A description of the advanced research WRF version 3. Boulder (CO): National Center for Atmospheric Research; 2008. Report No.: NCAR/TN-4751STR.

## **Appendix. Wavelet-Stat (W-S) Metrics**

---

---

The following provides a brief overview of the metrics that appear in the Wavelet-Stat (W-S) output graphics.\*

Frequency Bias is the ratio of the total number of forecasts of an event to the total number of observations. A value of 1.0 indicates a perfect forecast. A value of <1.0 indicates the event was not forecasted frequently enough, while a value of >1.0 indicates the event was forecasted too frequently.

The Intensity Skill Score (ISS) evaluates the forecast skill as a function of the intensity or magnitude of the field variable and of the spatial scale of the error. Positive ISS skill score values are associated with a skillful forecast, whereas negative values are associated with no skill.

Base Rate is the overall proportion of grid points with observed events, to total grid points in the domain.

The Percent Mean Squared Error (%) is computed from the Mean Squared Error (MSE), which is the average of all the differences indicated by the pixels in the domain (Mittermaier 2006). The MSE for a perfect forecast is 0. Percent MSE is the percentage of the total MSE at a specific scale size and threshold.

The Percent Forecast Energy Squared (%) (En2%), Percent Observation Energy Squared (%) (En2%), and their corresponding energy values (En2) are used for computing bias as a function of scale and threshold. The energy is computed by taking the average of the difference field grid-point squared values.

---

\*National Center for Atmospheric Research (NCAR). Model evaluation tools (MET). ver. 5.2. Boulder (CO): NCAR; 2016b [accessed 2019 Aug 15]. <https://dtcenter.org/community-code/model-evaluation-tools-met>.

## List of Symbols, Abbreviations, and Acronyms

---

2-D	2-dimensional
4-D	4-dimensional
ARL	Army Research Laboratory
ARW	Advanced Research Weather Research and Forecasting
CONUS	continental United States
CSV	comma-separated value
DC	Washington, DC
Delmarva	peninsula occupied by Delaware and parts of Maryland and Virginia
E-S	Ensemble-Stat
F-O	forecast minus observation
FDDA	4-D data assimilation
GEFS	Global Ensemble Forecast System
GFS	Global Forecast System
IC/LBC	initial conditions/lateral boundary condition
IDV	Integrated Data Viewer
ISS	Intensity Skill Score
MCF	mean component function
MET	Model Evaluation Tools
MSE	Mean Squared Error
NCAR	National Center for Atmospheric Research
NCEP	National Center for Environmental Prediction
NetCDF	Network Common Data Form
NWP	Numerical Weather Prediction
P-C	Pcp-Combine
R	a language and environment for statistical computing
R-D-P	Regrid_data_plane

UPP	Unified Post Processor
UTC	Coordinated Universal Time
VCF	variation-around-the-mean-component function
W-S	Wavelet-Stat
WRE-N	Weather Running Estimate-Nowcast
WRF	Weather Research and Forecasting
WRF-ARW	Weather Research and Forecasting-Advanced Research WRF

1 DEFENSE TECHNICAL  
(PDF) INFORMATION CTR  
DTIC OCA

1 CCDC ARL  
(PDF) FCDD RLD CL  
TECH LIB

1 GOVT PRINTG OFC  
(PDF) A MALHOTRA

1 US NAVY RSRCH LAB  
(PDF) DR J MCLAY

1 US AIR FORCE 557TH WEATHER WING  
(PDF) R CRAIG

1 DCGS-A WEATHER EET LEAD  
(PDF) J CARROLL

4 UCAR  
(PDF) T JENSEN  
T FOWLER  
J H GOTWAY  
B BROWN

2 ESRL  
(PDF) K FENTON  
M WANDISHIN

10 CCDC ARL  
(PDF) FCDD RLC E  
B MACCALL  
T JAMESON  
FCDD RLC ED  
G VAUCHER  
FCDD RLC EM  
H CAI  
J SMITH  
J PASSNER  
R DUMAIS  
B REEN  
L DAWSON  
FCDD RLC NB  
E CHIN