The 2015 HMT-WPC Winter Weather Experiment

12 January – 13 February 2015
Weather Prediction Center
College Park, MD

Final Report

Updated: April 6, 2015

1. INTRODUCTION

The Hydrometeorological Testbed at the Weather Prediction Center (HMT-WPC) hosted 28 forecasters, researchers, and model developers (Appendix A) at its fifth annual Winter Weather Experiment from January 12 – February 13, 2015. In addition to the on-site participants, a portion of the experiment was opened up to allow remote participation from both the operational and research meteorological communities via a daily forecast briefing webinar. This year's experiment continued focus on exploring the use of emerging short range microphysics-based snowfall forecasting techniques and further exploration of the extension of winter weather forecasts beyond 72 hours. Specifically, the goals of the experiment were to:

- Explore the utility of alternative microphysics-based snowfall forecasting methods, including their application to ensemble forecasts.
- Explore the utility of the parallel SREF and experimental NARRE for winter weather forecasting.
- Explore new datasets to improve the winter weather outlook forecast process.
- Gather feedback on the winter weather outlook forecasts.
- Enhance collaboration among NCEP centers, WFOs, and NOAA research labs on winter weather forecast challenges.

This report summarizes the activities, findings, and operational impacts of the experiment.

2. EXPERIMENT DESCRIPTION

Daily Activities

The 2015 experiment featured four types of activities. A detailed version of the daily schedule can be found in Appendix B.

a. Experimental Short Range Forecasts

Each morning, participants used a combination of operational and experimental model guidance to issue an experimental 24 hr deterministic snowfall forecast valid 00 – 00 UTC for a storm of interest during either the Day 1 (24 – 48 hr) or Day 2 (48 – 72 hr) period (Fig. 1a). For this forecast, participants were asked to draw 1", 2", 4", 8", 12", and 20" snowfall contours, depending on the event magnitude. Participants were also asked to designate their confidence for the forecast on a separate graphic (Fig. 1b). How the forecasters chose to communicate their confidence was free-form in order to promote discussion on ways to best communicate forecast confidence and uncertainty.

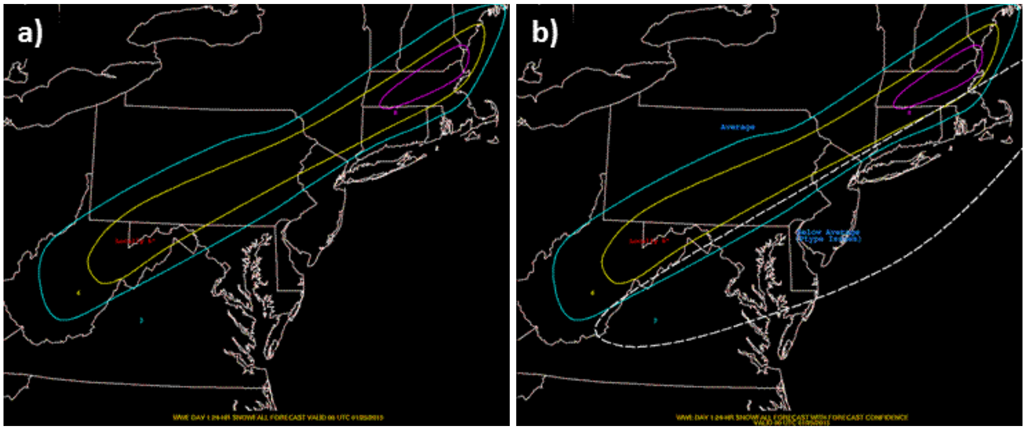

Figure 1. Example of an experimental Day 1 (a) 24 hr deterministic snowfall forecast and (b) accompanying confidence graphic issued during the 2015 HMT-WPC Winter Weather Experiment.

b. Experimental Medium Range Forecasts

Each afternoon, participants used a variety of derived guidance to issue experimental 24 hr probabilistic winter weather outlook forecasts (Fig. 2) for two days of their choice during the Day 4-7 (84-180 hr forecast) period. Three thresholds of experimental forecasts were created: 1) the probability of receiving at least 0.10" precipitation (liquid equivalent) in the form of snow, sleet, or freezing rain, 2) the probability of >0.50" liquid equivalent in the form of snow and/or sleet, and 3) the probability of >.01" of freezing rain. Participants were asked to draw probability contours indicating a 10%, 30%, 50%, 70%, and 90% chance of exceeding the threshold; the forecasts were valid 12 – 12 UTC.

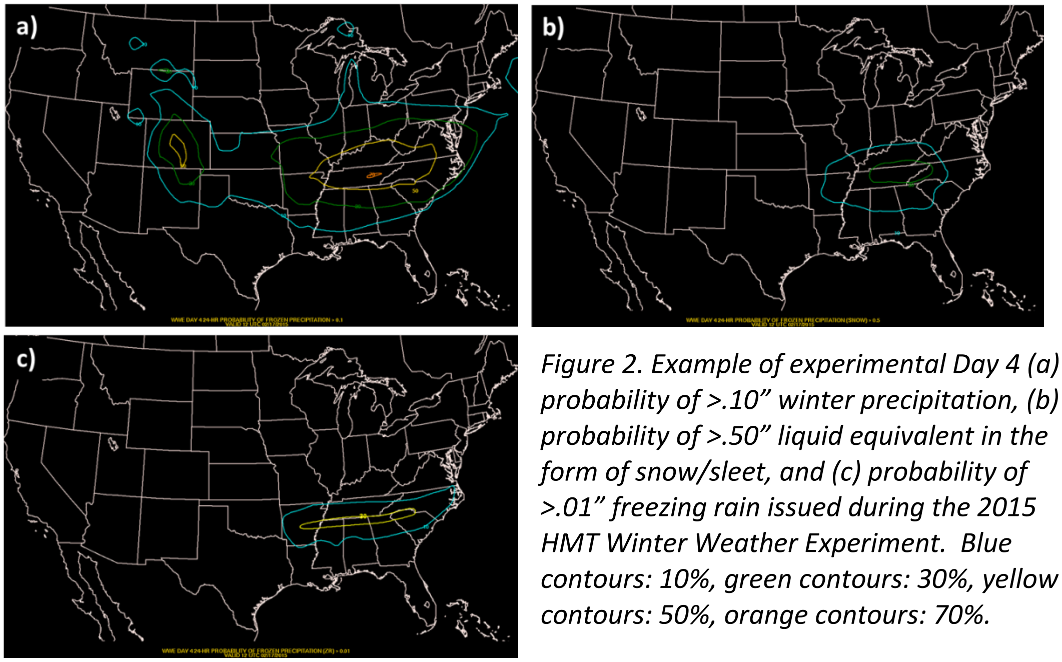

Figure 2. Example of experimental Day 4 (a) probability of >.10" winter precipitation, (b) probability of >.50" liquid equivalent in the form of snow/sleet, and (c) probability of >.01" freezing rain issued during the 2015 HMT Winter Weather Experiment. Blue contours: 10%, green contours: 30%, yellow contours: 50%, orange contours: 70%.

c. Forecast Discussion

For the first time, this year's Winter Weather Experiment opened access to the experiment to the wider meteorological community (e.g. operational forecasters, researchers and academia) via daily forecast briefings. While creating the short-term forecast each morning, participants created a presentation that outlined their forecast methodology and showed images of relevant data that were used in the forecast process. They then gave a brief webinar presentation each afternoon using to explain their forecast reasoning and highlight relevant data used. A daily invitation was sent to the meteorological community at-large, and those who opted to participate were able to see how experimental data sets were applied, ask questions, and engage in community discussion on datasets and forecast issues.

d. Subjective Model Evaluation

Twice daily, attendees worked together to subjectively evaluate the performance of the experimental forecasts and model guidance from the previous week of the experiment. The two subjective evaluation sessions consisted of a series of survey questions and associated graphics designed to spark discussion about the strengths, weaknesses, trends, biases, and overall effectiveness of the experimental forecasts and model guidance.

Evaluations of experimental short-range forecasts and the model guidance were conducted based on WPC's 20 km gridded snowfall analysis. To generate this analysis, precipitation type is determined based on surface observations. In regions where snow is observed, an initial analysis is generated using a combination of QPE from the Climatology-Calibrated Precipitation Analysis (CCPA; Hou et al. 2013) and climatological snow-to-liquid ratio (SLR) values (Baxter et al. 2005). This analysis is then modified based on COOP, CoCoRaHS, and METAR observations using a Barnes analysis. Snowfall observations are only retained in the analysis if all surrounding grid points also contain valid observations (i.e. no extrapolation is allowed). This strict requirement, combined with the relatively coarse grid size, often resulted in an analysis that experiment participants considered inadequate. To supplement the WPC analysis, participants turned to observations from the National Operational Hydrologic Remote Sensing Center[1] (NOHRSC).

Evaluations of experimental medium-range forecasts and guidance were conducted using two 4 km gridded analysis datasets developed by WPC. For the frozen precipitation greater than or equal to 0.10" threshold, the analysis is based on a

---

[1] http://www.nohrsc.noaa.gov/interactive/html/map.html

combination of hourly Stage IV precipitation data (Lin and Mitchell 2005) and the hourly 2.5 km Real-Time Mesoscale Analysis (RTMA; De Pondeca et al. 2011) 2 m temperature analysis. To generate the analysis, precipitation is accumulated each hour at grid points where the 2 m temperature is less than or equal to 0°C. If precipitation reaches a total of at least 0.10" during the 24 hr analysis period, the grid point is considered to have met the product definition of 0.10" (liquid equivalent) frozen precipitation. While this analysis relies on the assumption that snow, sleet, and/or freezing rain do not occur when 2 m temperatures are greater than 0°C, using gridded data allows for a much more coherent analysis than could be obtained from individual station observations alone. For the >0.50" liquid equivalent in the form of snow and >0.01" freezing rain thresholds, precipitation type from hourly RAP analysis is combined with hourly Stage IV data to identify areas that surpassed the >0.50" snow/sleet and >0.01" threshold. Due to a lack of hourly Stage IV data, these analyses are not available across the Northwest River Forecast Center's domain.

Data

In addition to the full multi-center suite of operational deterministic and ensemble guidance, participants were asked to consider several different experimental guidance systems and microphysics-based snowfall forecasting techniques while preparing their Day 1-2 deterministic snowfall forecasts. Participants had access to a prototype version of the North American Rapid Refresh Ensemble (NARRE), an experimental system that is designed to represent a "next generation" version of the Short Range Ensemble Forecasting (SREF) System. Participants also had access to WPC's PWPF ensemble that is used to derive WPC's probabilistic winter precipitation forecasts. Table 1 summarizes the model guidance that was the focus of the short range portion of the experiment, and more information about each dataset is provided below.

a. Short Range Forecast Guidance

Snow to Liquid Ratio Schemes

Several snow-to-liquid ratio (SLR) schemes were used to generate post-processed snowfall forecasts for use and evaluation. The Roebber technique (Roebber et al. 2003) is a neural network approach that uses information about low, mid, and upper-level temperature and moisture to determine the SLR. SLR values are available every 3 hours at 40 km resolution, and are calculated using the North American Mesoscale Model (NAM) and Global Forecasting System (GFS). The Baxter SLR climatology (Baxter et al. 2005) is a 106 km resolution grid based on measured SLRs from NWS COOP data, and provides CONUS climatological SLR information for three seasons (October-November,

Table 1. Featured Day 1-2 guidance for the 2015 HMT-WPC Winter Weather Experiment. All models are initialized at 00 UTC except for the SREF and SREFP (21 UTC). Experimental guidance is shaded.

Provider	Model	Resolution	Forecast Hours	Notes
EMC	NAM	32 km	84	Operational NAM
EMC	NAM-RF	12 km	84	Rime factor modification of SLR
EMC	SREF (21 members)	16 km (32 km display)	87	Operational SREF; SLR derived from 2 m temperature such that $SLR = .5(273.15 - T_{2m}) + 8$ for temperatures < 5°C
EMC	SREFP (26 members)	16 km (32 km display)	87	Parallel version of SREF; separate 16 member ensemble available to test rime factor modification in ensembles
WPC	PWPF ensemble (32 members)	20 km	72	Operational PWPF ensemble prior to 3/2014; includes SREF members, GEFS mean, ECMWF mean, and deterministic NAM, GFS, CMC, and ECMWF; SLR is an average of multiple techniques
WPC	PWPF ensemble (57 members)	20 km	72	3/2014 update added 25 randomly selected ECMWF ensemble members
EMC/ESRL	NARRE (8 members)	13 km	48	Next generation version of SREF, WRF and NMMB cores; used same 2 m temperature-based SLR scheme as the SREF

December-January-February, March-April). The Environmental Modeling Center's (EMC) SLR algorithm, applied to the SREF, is defined as:

$$SLR = .5(273.15 - T_{2m}) + 8$$

in which 2 m temperature data from model forecasts to determine the SLR. Finally, WPC's operational SLR is a blend, generated at 6 hr intervals, consisting of equal weighting of the Roebber NAM SLR, Roebber GFS SLR, Baxter Climatological SLR, and an 11:1 ratio.

Experimental Model Guidance

Two experimental short-range ensemble systems were featured in experiment activities. The first was the parallel version of the SREF (SREFP). This parallel version has membership increased to 26 members (from 21), is reduced from 3 to 2 cores (ARW and NMMB), and has varying initial condition and physics schemes in an effort to increase diversity (Appendix C). Vertical resolution is also increased to 41 levels.

The second experimental ensemble was the NARRE. Provided by the EMC and Earth Systems Research Lab (ESRL), the ensemble was 13 km resolution and contained 8 members split evenly between WRF and NMMB cores, with additional diversity provided from varying microphysics, initial condition and boundary layer schemes. (Table 2).

Table 2. Membership characteristics of the NARRE used in the 2015 HMT Winter Weather Experiment.

Member	Initial Conditions	Microphysics
Ctl RAP	GFS	Thompson
RAP1	GEP01	Thompson
RAP2	GEP02	Ferrier
RAP3	GEP03	Ferrier
Ctl NMMB	GFS	Ferrier
NMMB1	GEP01	Ferrier
NMMB2	GEP02	Ferrier
NMMB3	GEP03	Ferrier

In addition to the two experimental ensemble systems, two versions of WPC's PWPF ensemble were available and evaluated. The WPC PWPF ensemble is currently a 57-member, 20 km ensemble that is generated internally by WPC and used extensively in the WPC Winter Weather Desk forecast process. The ensemble membership consists of all 21 SREF members, 5 randomly selected Global Ensemble Forecast System (GEFS) members, 25 randomly selected European ensemble (ECENS) members, and the latest operational NAM, GFS, GEFS ensemble mean, CMC, European Centre for Medium-Range Weather Forecasting (ECMWF), and ECMWF ensemble mean (ECENS) runs. Prior to March 2014, the PWPF ensemble consisted of 32 members (omitting the 25 random ECENS members) and was updated to improve diversity. Snowfall from the WPC PWPF ensemble is calculated using the operational WPC SLR.

Also featured were experimental model snowfall and snowfall accumulation techniques. This year's experiment continued previous investigation of the rime factor (RF) technique, which modifies the initial snow-to-liquid ratio (SLR) value (Table 3) by incorporating information from the model's microphysics about the amount of riming on ice particles. The modified SLR (SLR_{RF}) is then used in conjunction with an instantaneous percentage of frozen precipitation (POFP), a parameter describing the ratio of frozen to liquid hydrometeors in the lowest model level, to calculate the rime factor-modified snowfall:

$$Snowfall = (QPF) \times (POFP) \times (SLR_{RF}).$$

As in previous experiments, the RF technique was evaluated for the NAM. This year's experiment also began investigation into applying the RF in ensembles by expanding it to the parallel SREF. The RF-modified SREFP (SREFP_RF) ensemble consisted of the 16 SREFP members that contained the RF parameter, allowing for the evaluation of RF-modified snowfall ensemble products.

Table 3. Relationship between rime factor values and the resulting modification of the SLR.

Rime Factor	SLR Modification
1 < RF < 2 (fluffy snow)	$SLR_{RF} = SLR$
2 < RF < 5 (rimed snow)	$SLR_{RF} = \frac{SLR}{2}$
5 < RF < 20 (graupel)	$SLR_{RF} = \frac{SLR}{4}$
RF > 20 (sleet-like)	$SLR_{RF} = \frac{SLR}{6}$

Snowfall forecasts using the snow water equivalent (SWE) parameter in the NAM, SREFP and NARRE were also tested. SWE is an implicit parameter from the model's microphysics and land surface scheme that isolates the amount of liquid precipitation that has fallen in frozen form. SWE was examined because POFP is used continuously through the forecast time stamp as opposed to being applied instantaneously in the algorithm. Three-hour time differences of SWE were combined with a modification of the Baxter SLR to create a snowfall forecast. The Baxter SLR was modified using the aforementioned rime-factor methodology:

$$Snowfall = (SWE)x(SLR_{RF}).$$

b. Medium Range Forecast Guidance

A variety of experimental deterministic and probabilistic medium range guidance was also explored (Table 1). ESRL provided deterministic and probabilistic QPF products from their 2nd Generation Reforecast dataset[2] (Hamill et al. 2013), which is a dataset of historical (1985-2010) weather forecasts generated by re-running the version 9.0.1 of the GEFS. The ensemble used for the reforecast dataset features 11 members utilizing a 00 UTC initialization~~,~~ and provides a reference/training dataset for statistical post-

---

[2] http://www.esrl.noaa.gov/psd/forecasts/reforecast2/analogs/index2.html

processing of the current model forecast. The probabilistic and deterministic QPF products are created via analog approaches, using both the North American Regional Reanalysis (NARR) and Climatologically-Calibrated Precipitation Analysis (CCPA) as the observed precipitation analogs.

Furthermore, WPC generated 24 hr probabilistic guidance products for the three thresholds of interest for Days 4, 5, 6 and 7: the probability of ≥.10” of frozen precipitation (snow, sleet, freezing rain), the probability of >.50” of liquid equivalent falling as snow or sleet, and the probability of >.01” of freezing rain.

The QPF component of the guidance is created by disaggregating WPC’s operational day 4-5 and 6-7 QPF into 24 hour QPFs for each of the four days. A cumulative distribution function (CDF; Von Storch and Zwiers 1999; Wilks 2006) for amounts equal to or greater than 0.10 inches is then computed using the resultant 24 hour QPF is then used as mean, and the 24 hour QPF from each member of the GEFS, ECENS and CMCE as the variance.

The thermal component of the guidance is generated by using derived precipitation type fields from each of the 90 members of the GEFS, ECENS, and CMCE. A mosaic of snow, sleet, and freezing rain are used in computing the ensemble probability of frozen precipitation. The ensemble probability of frozen precipitation is then combined with the aforementioned probability of WPC QPF greater than 0.10 inches to create a probability of winter precipitation greater than 0.10 inches. This process was repeated to generate guidance for greater than 0.50 inches of liquid equivalent snow (using snow and sleet precipitation type), and 0.01 inches of freezing rain (using freezing rain precipitation type).

3. CASES

The four-week experiment period was generally characterized by a trough over the eastern two-thirds of the CONUS, with a ridge over the west coast (Fig. 3). This pattern allowed for multiple Arctic air incursions into the eastern United States, resulting in anomalously cold temperatures for much of the eastern CONUS (Fig. 4), particularly in the first two weeks of February (Fig. 4c, 4d).

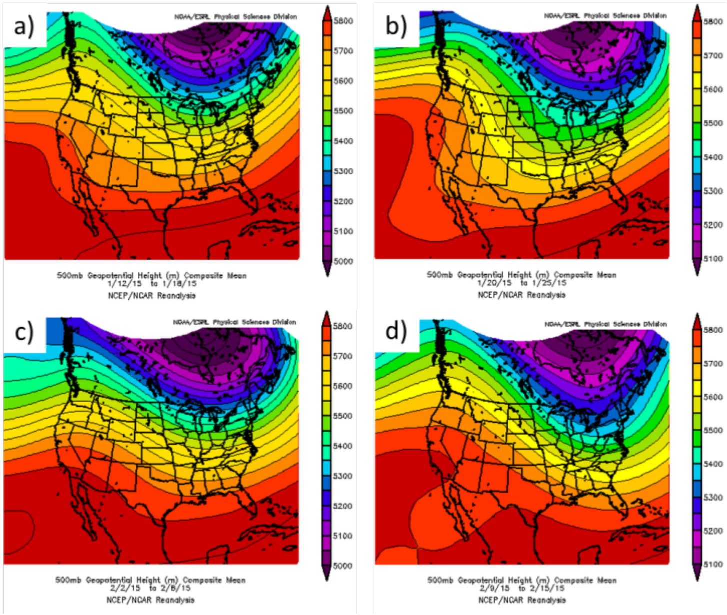

Figure 3. Composite mean 500 hPa heights for the 4 week experiment period: (a) 12-18 Jan, (b) 20-25 Jan, (c) 2-8 Feb, and (d) 9-15 Feb, 2015. Images generated from the NCEP/NCAR Reanalysis provided by NOAA/ESRL/PSD (http://www.esrl.noaa.gov/psd/data/composites/day/).

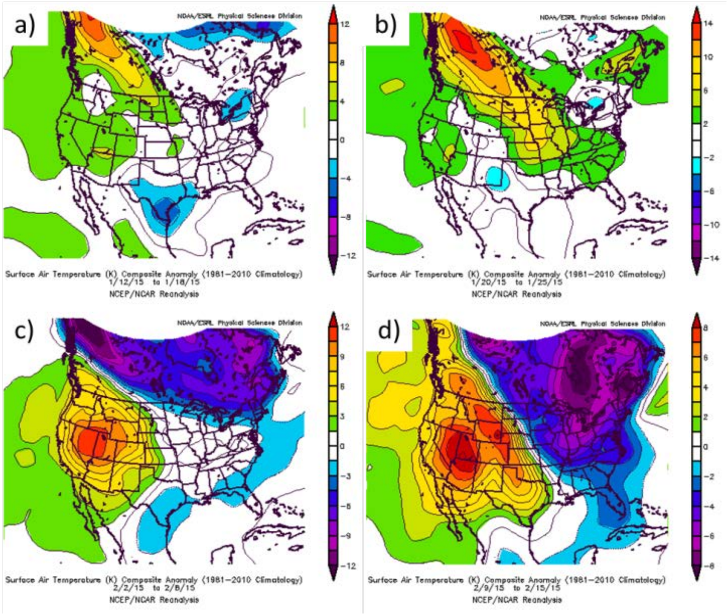

Figure 4. Composite mean surface temperature (K) anomalies for the 4 week experiment period: (a) 12-18 Jan, (b) 20-25 Jan, (c) 2-8 Feb, and (d) 9-15 Feb, 2015. Images generated from the NCEP/NCAR Reanalysis provided by NOAA/ESRL/PSD (http://www.esrl.noaa.gov/psd/data/composites/day/).

The amplified flow pattern resulted in several significant winter weather events, particularly across the southeastern and northeastern U.S. The first significant event of the experiment didn't occur until the end of the second week on January 23, where a digging 500 mb cut-off low brought a rare 12"+ of snowfall to the Amarillo, TX area. The focus of major events then shifted east, with a storm bringing over a foot of snow across IA and northern IL, including Chicago, around February 2. This storm then pushed east and produced over a foot of snow across NY and the New England region on February 3. The New England region was hit repeatedly throughout the last half of the experiment, with major snowfall events crippling the region on January 25-26, January 28, February 9-10, and February 16.

Table 4. Experimental short-range forecasts and subjective verification dates for the 2015 HMT-WPC Winter Weather Experiment. D1 and D2 refer to Day 1 (24 – 48 hr) and Day 2 (48 – 72 hr) forecasts, respectively. Supplemental verification was completed by WPC personnel in the week following the experiment to provide a more robust evaluation.

Forecast Valid Time	**Forecast**		**Verification**		**Forecast Area**
00Z 05 Jan 2015			D1	D2	Northeast
00Z 07 Jan 2015			D1	D2	Mid Atlantic
00Z 14 Jan 2015	D1		D1	D2	Four Corners
00Z 15 Jan 2015	D1		D1	D2	Mid Atlantic
00Z 17 Jan 2015		D2	D1	D2	Northeast
00Z 18 Jan 2015	D1	D2	D1	D2	Upper Midwest
00Z 22 Jan 2015	D1		D1	D2	Ohio Valley to Mid Atlantic
00Z 23 Jan 2015	D1		D1	D2	Southwest
00Z 25 Jan 2015	D1	D2	D1	D2	Ohio Valley to Northeast
00Z 28 Jan 2015			D1	D2	Northeast
00Z 03 Feb 2015			D1	D2	Northeast
00Z 04 Feb 2015	D1		D1	D2	Northeast
00Z 05 Feb 2015	D1		D1	D2	Central Midwest
00Z 06 Feb 2015	D1		D1	D2	Northeast
00Z 08 Feb 2015		D2			Northern Rockies
00Z 09 Feb 2015		D2	D1	D2	Northeast
00Z 10 Feb 2015			D1	D2	Northeast
00Z 11 Feb 2015	D1		D1	D2	Upper Midwest
00Z 12 Feb 2015	D1		D1	D2	Four Corners
00Z 13 Feb 2015	D1		D1	D2	Upper Midwest to Northeast
00Z 15 Feb 2015		D2	D1	D2	Upper Midwest
00Z 16 Feb 2015		D2	D1	D2	Northeast

A complete list of the snowfall events and time periods investigated for both the short term and medium range forecasts can be found in Table 4 and Table 5, respectively. Note that not all events listed in Table 4 were covered during experiment operations. Some events occurred during the off week (January 26-30) and/or occurred on weekend days where they were not in range of experimental short-term forecast periods. In cases where a major event occurred outside of the experiment, archived model data was saved so that the case could be included in verification exercises.

Table 5. Experimental CONUS medium range forecasts and subjective verification dates for the 2015 HMT-WPC Winter Weather Experiment. D4, D5, D6 and D7 refer to Day 4 (72 – 96 hr), Day 5 (96 – 120 hr), Day 6 (120 – 144 hr), and Day 7 (144 – 168 hr), forecasts, respectively. WPC personnel completed supplemental verification in the week following the experiment to provide a more robust evaluation.

Forecast Valid Time	Forecast			
12Z 17 Jan 2015		D5		
12Z 18 Jan 2015			D6	
12Z 19 Jan 2015	D4		D6	
12Z 20 Jan 2015	D4	D5	D6	D7
12Z 21 Jan 2015				D7
12Z 22 Jan 2015			D6	
12Z 24 Jan 2015	D4			
12Z 25 Jan 2015	D4	D5		
12Z 26 Jan 2015	D4	D5		
12Z 27 Jan 2015	D4	D5		
12Z 30 Jan 2015				D7
12Z 06 Feb 2015	D4			
12Z 08 Feb 2015		D5	D6	
12Z 09 Feb 2015	D4	D5	D6	
12Z 10 Feb 2015	D4	D5	D6	
12Z 11 Feb 2015		D5		
12Z 13 Feb 2015	D4			
12Z 15 Feb 2015		D5	D6	
12Z 17 Feb 2015	D4	D5	D6	D7
12Z 18 Feb 2015		D5	D6	D7

4. EXPERIMENTAL SHORT RANGE FORECASTS

During experiment operations 18 short term deterministic forecasts, 12 for Day 1 and 6 for Day 2, were created (Table 4). Through a combination of subjective evaluations completed during and after the experiment, a total of 42 cases were evaluated, 21 for both Day 1 and Day 2.

Figure 5 shows that the majority of the forecasts (67%) were rated as ‘good’ during evaluation, with 33% of cases being rated as ‘fair.’ A decrease in forecast performance occurred with time, with the percentage of ‘good’ ratings decreasing from 75% on Day 1 to 50% on Day 2, although a smaller sample size of Day 2 forecasts may affect the distribution. This decrease occurred despite lower standards for Day 2 forecasts; participants were often more hesitant to criticize Day 2 forecasts and willing to accept larger errors in placement and magnitude errors.

Common issues noted by evaluators, particularly in Day 1 forecasts, were the lack of spatial coverage of the lower magnitude (2” and 4”) contours (e.g. mis-located, not enough areal coverage), as well as spatial and magnitude errors with the heaviest snowfall (e.g. heaviest snow forecast in wrong location, under-forecast maximum snowfall amounts). For Day 2 forecasts, participants often gave the forecast credit for the having the “general idea” of the spatial coverage of snowfall, and were less concerned with accuracy of amounts. However, as in years past, participants noted that there were often significant discrepancies between the WPC and NOHRSC snowfall analysis, which made providing consistent and accurate evaluations difficult.

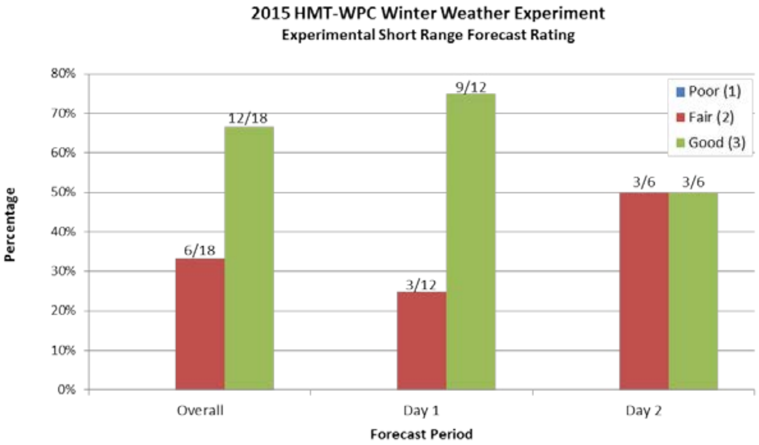

Figure 5. Percentage of ratings assigned to each experimental deterministic snowfall forecast during the subjective evaluation.

Regarding forecast confidence, the forecast team was allowed to create a graphic to communicate confidence in any way they saw fit. In most cases, groups resorted to outlining areas of higher or lower confidence, often using the terms 'below average,' 'average,' or 'above average' confidence. The topic of how to best communicate forecast confidence and uncertainty became a point of conversation amongst several of the groups. One popular sentiment was that confidence should not be confused with uncertainty; often times the terms are used interchangeably. Additionally, the point was raised that forecasters should never have low confidence in a forecast, and they should separate that from uncertainty when communicating with users. For example, during one forecast briefing a group expressed that "we are confident in our forecast, but there is still a high level uncertainty in the models regarding where the heaviest snow will fall." Examining ways to best communicate with users, particularly in regards to uncertainty, will become an increased point of emphasis in the future.

Experimental NAM Rime Factor-Modified Snowfall Accumulations

The RF modification (Table 3) was applied to the NAM by altering both the Baxter SLR climatology and the Roebber SLR to evaluate which method provided the best snowfall guidance. Participants were asked to use the WPC and NOHRSC snowfall analyses to score each NAM RF-modified output on a scale of 1 to 5, with 1 representing 'very poor' and 5 representing 'very good' guidance. Figure 6 shows that overall, the RF-modified Baxter SLR guidance (average rating: 3.00) performed better than the RF-modified Roebber SLR guidance (average rating: 2.84). However, this advantage was seen only in Day 1 forecasts, where the average Baxter rating (3.42) was higher than the average Roebber rating (3.10); both methods average a 2.57 score for Day 2, which represented sharp declines.

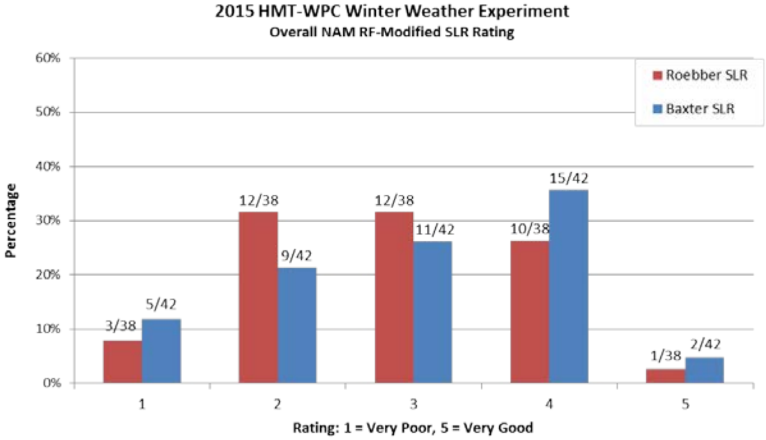

Figure 6. Percentage of ratings assigned to the NAM RF-modified Roebber and Baxter deterministic snowfall products during subjective evaluation.

Participants noted that the Roebber SLR appeared to have a high bias, especially along the warm boundary of systems and in areas where precipitation type was not purely snow. While the RF modification assisted in lowering Roebber SLR values (and corresponding snowfall amounts) in areas where riming and/or mixed phase precipitation took place, it did not work to offset any high biases seen in the Roebber SLR in deep cold air/pure snow environments. To the opposite effect, forecasters expressed concern that using the Baxter SLR also had limitations, and tended to have a low bias in areas of deep cold air. This was particularly noticeable in cases that focused on the upper Midwest and Northeast, where arctic air was entrenched for most of the experiment, as can be seen in Fig. 7. The RF-modified Roebber guidance (Fig. 7a) has a large swath of >18” stretching from Long Island north through the eastern portion of Maine, with a peak of >36” in central Massachusetts. The RF-modified Baxter guidance (Fig. 7b), however, has much lower values through this area (generally 12-20”) and max values of ~22-24”. While the WPC (Fig. 7c) and NOHRSC (Fig. 7d) 24 hour snowfall analyses show that isolated maximum amounts were upwards of 30”+, the Roebber guidance’s spatial coverage of >18” was overdone, and the Baxter guidance output is closer to what was analyzed.

To further investigate the effect of the RF modification and various SLR schemes on NAM snowfall output, participants were asked to evaluate the NAM 24 hour QPF that went into creating the snowfall (Fig 8). Similar to the snowfall evaluations, they ranked the NAM QPF on a scale of 1 to 5, with 1 representing ‘very poor’ and 5 representing ‘very good’ guidance, using

24-hr Stage IV precipitation data as the analysis. Overall, the NAM QPF scored an average of 3.05, with a majority of ratings clustering around the ‘fair’ rating, and experienced a decrease in forecast quality moving from Day 1 to Day 2 (Fig 8). A noticeable flaw in the NAM QPF was a high bias, as evaluators consistently noted that QPF values were high compared to the analysis, both in terms of spatial coverage of higher amounts and maximum QPF values. This can be seen in Figure 7 for a case in the northeast, where the NAM overproduces QPF (Fig. 7e) across portions of eastern Massachusetts, Rhode Island, eastern Connecticut, southern New Hampshire and southern Maine.

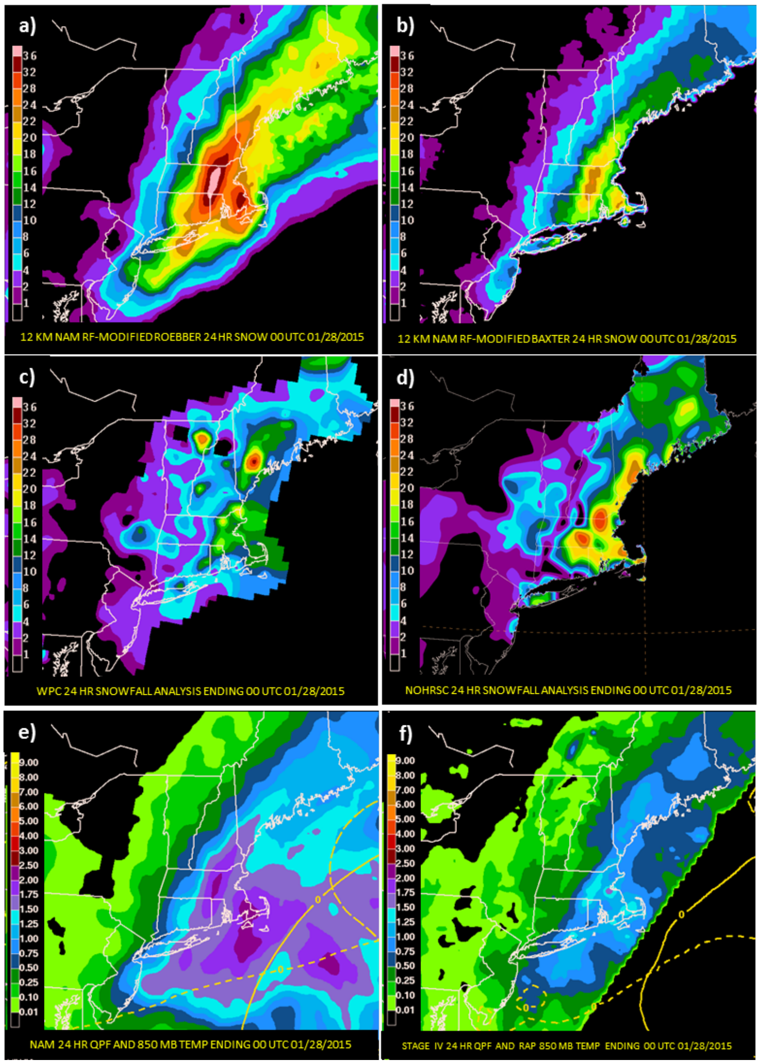

Figure 7. Showing the (a) NAM RF-modified Roebber 24 hour snowfall, (b) NAM RF-modified Baxter 24 hour snowfall, (c) WPC 24 hour snowfall analysis, (d) NOHRSC 24 hour snowfall analysis, (e) NAM 24 hour QPF and 850 mb 0^{o} C isotherms (yellow contours), and (f) StageIV 24 hour WPF analysis and RAP analysis of the 850 mb 0^{o} C isotherms (yellow contours), valid 00 UTC 28 Jan 2015. For panels (e) and (f), the short dashed contours represent the 850 mb 0^{o} C isotherm at 00 UTC 27 Jan 2015, the solid contour the 850 mb 0^{o} C isotherm at 12 UTC 27 Jan 2015, and the long dashed contour the 850 mb 0^{o} C isotherm at 00 UTC 28 Jan 2015.

A similar evaluation of 850 mb 0^{o}C isotherm forecasts revealed that spatial errors in the temperature field were common, but these were relatively small in scale when compared to errors in the QPF. In instances where there were large errors in the 850 mb temperature field, large spatial errors in the QPF field were also observed, and resulting large snowfall errors where mostly tied to QPF errors. Therefore, while the RF-modified Baxter SLR scheme provided better guidance than the Roebber during the experiment, it is possible that its suspected low bias worked to offset the NAM's high QPF bias in most cases, particularly at Day 1. This is likely given that a majority of the cases investigated in this year's experiment occurred in arctic air masses and featured a high NAM QPF bias. Given these trends, a statement that concludes the Baxter SLR is the best scheme to use for post-processing snowfall techniques should be used with caution, as it is possible that improving QPF forecasts would result in other SLR schemes, such as the Roebber, to be just as, if not more, effective.

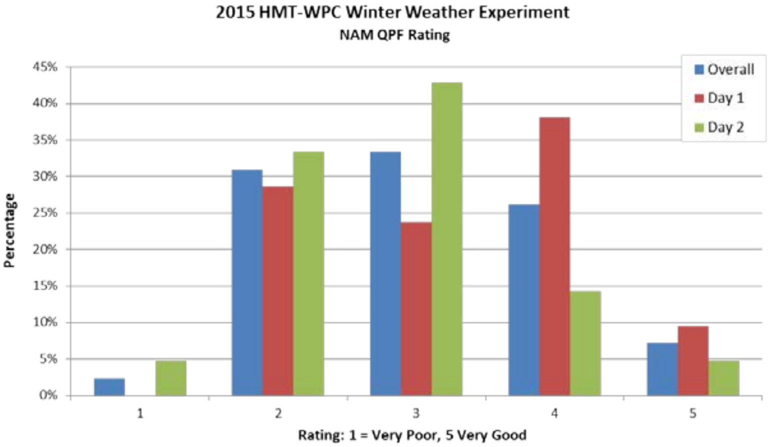

Figure 8. Percentage of ratings assigned to the NAM 24 hour QPF forecast during subjective evaluation.

Participants were also asked for their thoughts on the utility of the RF and other microphysical tools (such as POFP) in the snowfall forecast process. Feedback shows that most participants felt that the RF and POFP are helpful tools, and can help forecasters understand where riming or mixed precipitation may impact snowfall amounts. There was also positive response for using the rime factor to make modifications to SLR schemes, as nearly all participants thought this provided enhancement to the guidance. However, the current method of using instantaneous (hourly or 3 hourly) RF and POFP parameters to post-process cumulative snowfall is lacking. Additional benefit would be gained from having RF and POFP applied to QPF and SLR parameters at each model time-step. This would produce a snowfall, or at the very least, a frozen QPF parameter, that is continuous in time and prevents the need to assume instantaneous conditions apply to an entire 1 or 3 hour block of cumulative QPF.

Experimental Ensemble Performance

To evaluate the performance of the operational SREF, SREFP, and experimental NARRE, participants rated the 24 hour mean snowfall forecasts on a scale of 1 to 5, with 1 representing 'very poor' and 5 representing 'very good' guidance, compared against the WPC and NOHRSC snowfall analysis (Fig. 9). Evaluators accounted for the fact that the means were not deterministic forecasts and adjusted ratings accordingly. All of the mean snowfall forecasts used for this evaluation were generated with the same snowfall post-processing technique, using the EMC 2 m temperature SLR algorithm deployed in the operational SREF. Note that the experimental NARRE (available for only Day 1 forecasts) and SREFP were subject to data outages; the NARRE was available for 17 of the 42 total cases, and the SREFP for 32 of the 42 cases.

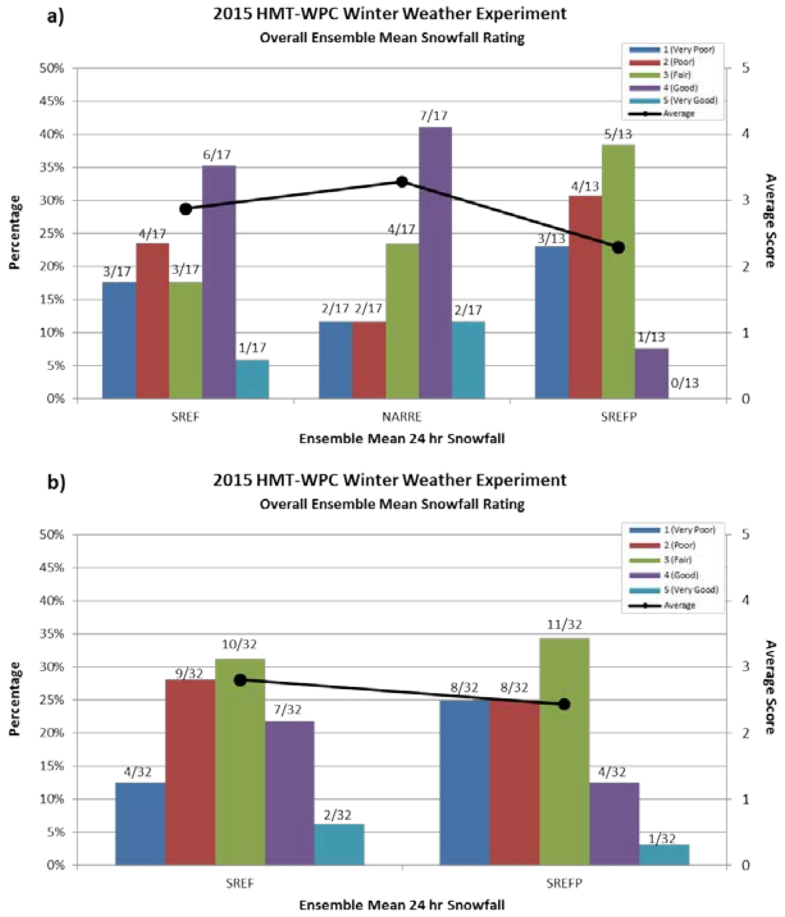

Figure 9. Percentage of ratings assigned to the 24 hour ensemble mean snowfall guidance during subjective evaluation for (a) the 17 cases in which the NARRE was available, and (b) the 32 cases the SREP was available. The black dots and solid black contour represent the overall average score for each ensemble system during the experiment.

Despite its smaller sample size, the NARRE mean performed the best of the ensembles (Fig. 9a; 3.29 average), with ~76% of available evaluations being scored as “fair” or better (scores of 3 and above). The SREFP scored the worst ; in the 32 cases it was available for evaluations (Fig. 9b), it consistently scored lower (2.44 average) than the operational SREF (2.86 average), with 50% of its ratings being scored as “poor” or “very poor” (scores of 1 or 2). When compared against the operational SREF on a case-by-case basis, the parallel SREF was rated worse guidance on ~40% of cases (13/32), while being rated better guidance in 21% of cases (7/32).

One of the main benefits of the NARRE was its resolution (Fig. 10). With all members and the mean at 13 km, the ensemble was able to resolve smaller-scale features, such as lake effect and topographical enhancements (Fig. 10c), which the two SREF ensembles (displayed at 32 km) could not (Fig. 10a, 10b). Also, its smaller membership (8 members) often resulted in a more coherent mean field, which provided benefit when compared to the snowfall analyses. However, participants noted that this was often the result of the 8 members being in relative agreement, signaling under-dispersion, and probabilistic forecasts were often over-confident.

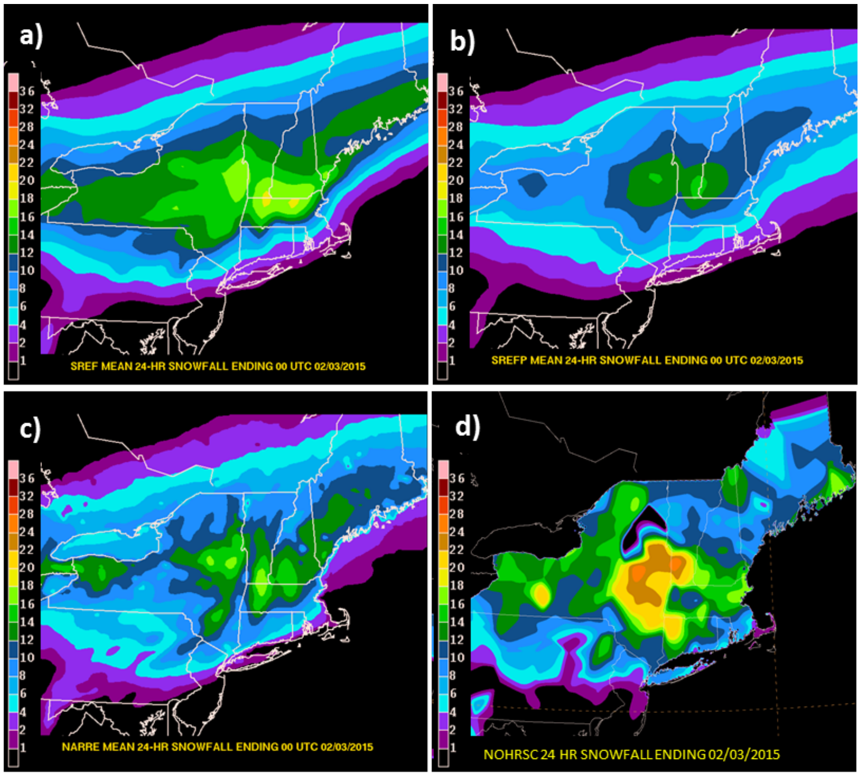

Figure 10. Showing the (a) SREF 24 hour mean snowfall, (b) SREFP 24 hour mean snowfall, (c) NARRE 24 hour mean snowfall, and (d) NOHRSC 24 hour snowfall analysis valid 00 UTC 03 Feb 2015.

Evaluation showed that the mean QPF and snowfall fields from the SREFP could be underdone, or produce “washed out” guidance that lacked any detail, particularly for Day 2 forecasts.

Limited investigation into this suggests it is a product of the 5 extra members and additional boundary layer, initial condition and microphysics diversity. Additionally, members of the SREFP tended to cluster according to their core (NMB or ARW); in cases where each core had different solutions regarding the spatial, temporal or magnitude details of a system, the mean was underdone with QPF amounts and overdone with spatial coverage. While these issues were noted in evaluating the mean fields, there did not appear to be systematic issues with the probabilistic fields, although further investigation is needed.

WPC Probabilistic Winter Precipitation (PWPF) Ensemble

Two versions of WPC's multi-member ensemble (Table 1) were used and evaluated: the 32-member PWPF (PWPF32), and the 57-member PWPF (PWPF57). During evaluation exercises participants rated probabilistic guidance for two snowfall thresholds (2" and 8" in 24 hours), as well as the mean snowfall, for both ensembles on a scale of 1-5, where 1 represented a 'very poor' forecast and 5 represented a 'very good' forecast. To do this, the WPC and NOHRSC 24 hr snowfall analysis was shown, and the 2" and 8" observed areas were outlined and plotted over the corresponding probabilistic forecasts for each ensemble. Figure 11 shows an example of PWPF performance for the 24 hour period ending 00 UTC 28 January 2015, with the PWPF32 plotted across the top and the PWPF57 plotted across the bottom.

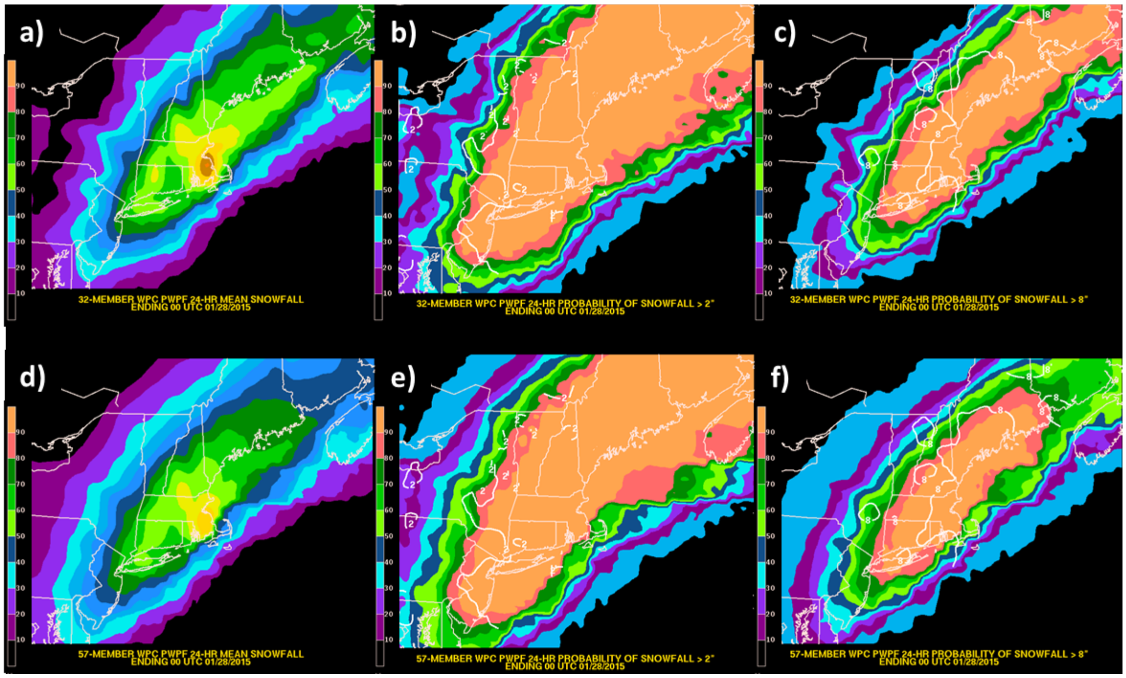

Figure 11. Showing ensemble guidance from the PWPF32 (top row) and PWPF57 (bottom row) valid 00 UTC 28 January 2015. Ensemble mean 24 hour snowfall is shown in (a) and (d), ensemble probability of receiving greater than 2" of snow in 24 hours is shown in (b) and (e), and ensemble probability of receiving greater than 8" of snow in 24 hours is shown in (c) and (f). Observed areas receiving >2" (panels b and e) and 8" (panels c and f) are outlined in white.

The two ensembles performed similarly throughout the experiment, with the PWPF32 rating slightly higher (Fig 12). The average ratings across all cases of the PWPF32 and PWPF57 for the probability of 2” threshold were 3.67 and 3.54, respectively; for the 8” probabilities the average scores were 3.72 and 3.69. However, the PWPF32 was given a higher rating than the PWPF57 in ~26% (10 of 39 cases) for the 2” threshold and 21% (8 of 39) for the 8” threshold; correspondingly, the PWPF57 was given a higher rating in just 13% of cases (5 or 39) for both thresholds.

As is reflected in these results, participants noticed that the probabilistic fields often looked similar (Fig. 11). In most instances where different ratings were assigned, small differences in the spatial coverage or magnitude of the probabilities led to the group choosing to rate one ensemble higher than the other. An example of this can be seen in Fig. 11c and Fig. 11f, where, despite similar solutions for the probability of >8” over the entire domain, evaluators rated the PWPF32 (Fig. 11c) higher than the PWPF57 (Fig. 11f) due to lower probabilities in western/southern New Jersey and eastern Pennsylvania.

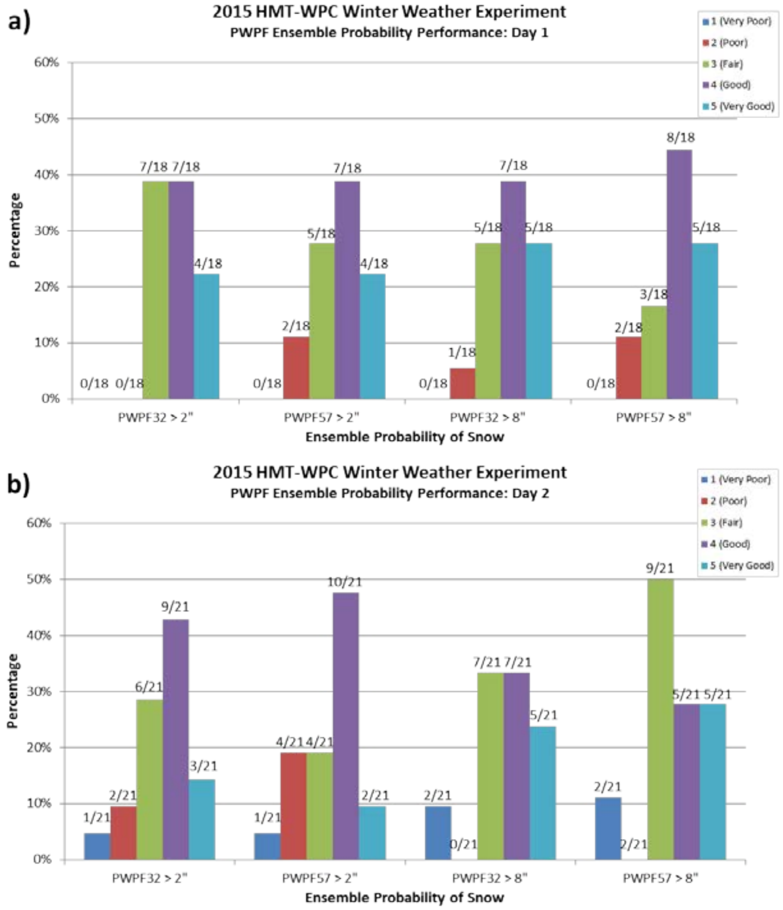

Figure 12. Percentage of ratings assigned to the ensemble probability of greater than 2” and 8” in 24 hours for (a) Day 1 forecasts and (b) Day 2 forecasts during subjective evaluation.

Use of Microphysical Parameters in Ensembles

The use of experimental microphysical parameters in an ensemble was also evaluated. Evaluation focused on three parameters using the SREFP: rime factor (RF), percent of frozen precipitation (POFP) and snow water equivalent (SWE). SWE was combined with the Baxter SLR to create one experimental snowfall product (SREFP_B), while another experimental product used RF, SWE and the Roebber SLR. Since data for the RF and POFP was available in only 16 of the SREFP members, this second snowfall product was created using a separate ensemble consisting of those 16 members (SREFP_RF).

Table 6. Percentage and average ratings of the 24 hour snowfall guidance from the operational SREF, SREFP, RF-modified SREFP using SWE and the Roebber SLR, and SREFP using SWE and the Baxter SLR.

	Rating (1 = Very Poor, 5 = Very Good)					**Total Cases**	**Average**
	1	**2**	**3**	**4**	**5**		
SREF	14.29%	26.19%	23.81%	30.95%	4.76%	42	2.86
SREFP	25.00%	25.00%	34.37%	12.50%	3.12%	32	2.44
SREFP_RF	18.75%	46.87%	21.87%	12.50%	0.00%	32	2.28
SREFP_B	25.71%	45.71%	20.00%	5.71%	2.85%	35	2.14

Table 6 shows that the ensemble mean snowfall products which used the SWE instead of QPF (SREFP_RF and SREF_B) provided inferior guidance when compared to the standard means from the operational SREF and SREFP. It was found that the SWE parameter was problematic, particularly in ARW members, as it often provided unrealistic solutions that 1) appeared "splotchy," and/or 2) provided liquid water values that were higher than the model QPF output over the same time period.

While ensemble snowfall products that use combinations of experimental microphysical parameters need further development, participants felt other displays from an ensemble perspective were promising. Showing probabilistic POFP guidance in the form of probabilities of POFP exceeding 50% and 80%, as well as the average POFP values, were found helpful in identifying areas of mixed precipitation. Fig 13 shows an example from the 24 hour period ending 00 UTC 11 February. The mean snowfall field from the SREFP suggests that snow will accumulate in north-central North Dakota and west-central Minnesota; however, the probability of POFP >80% (Fig 13a, values of 10-40%) and average POFP (Fig. 13b, values of 10-50%) suggests precipitation in these areas is not likely to be all snow. However, since the POFP parameter is currently instantaneous and conditionally dependent on precipitation at the surface, viewing these probabilities could be misleading. There is no way to tell if members are signaling a temperature profile with no frozen precipitation, or one that supports frozen precipitation but doesn't have any precipitation falling at that instantaneous moment, as both instances result in no POFP data. One suggestion was to include POFP at various model levels near the surface to reduce the dependency on precipitation.

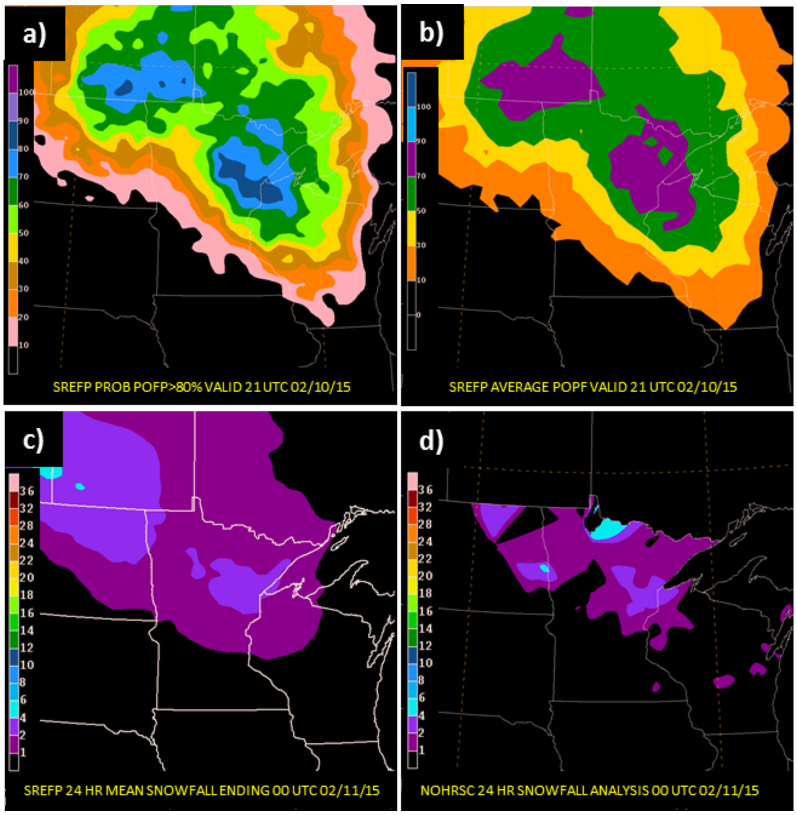

Figure 13. Showing the (a) SREFP ensemble probability of the percent of frozen precipitation >80% valid 21 UTC 10 Feb 2015, (b) average SREP percent of frozen precipitation valid 21 UTC 10 Feb 2015, (c) SREP 24 hour mean snowfall valid 00 UTC 11 Feb 2015, and (d) NOHRSC 24 hour snow analysis valid 00 UTC 11 Feb 2015. Note that the NOHRSC analysis does not contain data north of the U.S./Canadian border.

5. EXPERIMENTAL MEDIUM RANGE FORECASTS

In addition to the short-term snowfall forecasts, participants were also asked to issue various winter weather outlook forecasts for the Day 4-7 period. In addition to WPC's 24 hour probability of >.1" of winter precipitation forecasts, the forecast group also created 24 hour probabilities of >.50" liquid equivalent in the form of snow (including sleet), >.01" freezing rain. In evaluations, participants were asked to rank each experimental forecast on a scale of 1-3, choosing between "good" (3), "fair" (2), and "poor" (1).

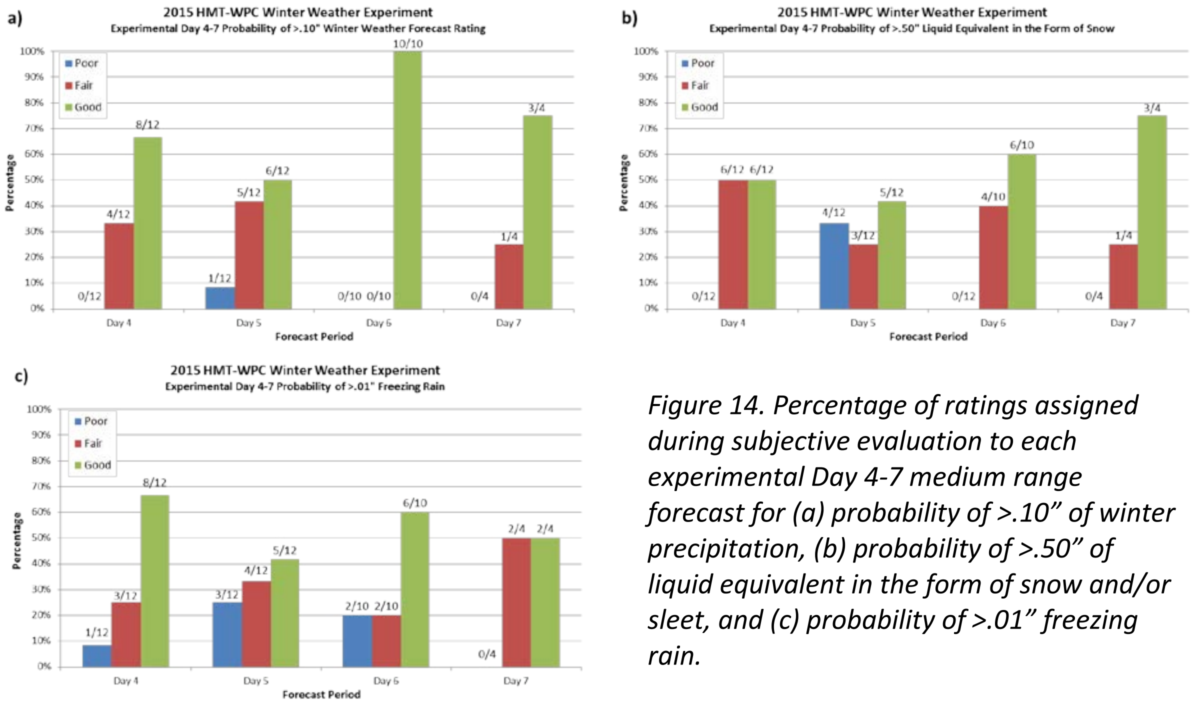

Figure 14. Percentage of ratings assigned during subjective evaluation to each experimental Day 4-7 medium range forecast for (a) probability of >.10" of winter precipitation, (b) probability of >.50" of liquid equivalent in the form of snow and/or sleet, and (c) probability of >.01" freezing rain.

Overall, forecasts performed well (Fig. 14); approximately 97% of forecasts for the >.10" winter precipitation were rated as "good" or "fair", this number dropped slightly to ~89% for the >.50" liquid equivalent of snow product and ~84% for the freezing rain product. Participants noted that freezing rain had less predictability than the winter weather and snow products, especially at longer lead times. However, participants often took this into account during ratings, and were often hesitant to downgrade freezing rain forecasts to "poor" given their understanding of the lower predictability and verification issues associated with freezing rain.

Results of the model guidance evaluations are seen in Table 7. Participants evaluated model guidance from the two days associated with the experimental forecasts (e.g. Day 4 and 6). For the first day (e.g. Day 4), they evaluated model guidance for all three (>.10" winter weather, >.50" liquid equivalent snow, >.01" freezing rain) products from all four guidance systems (combined, ECENS, GEFS, CMCE. Due to time constraints, they only evaluated guidance for the >.1" winter weather product for the second day (e.g Day 6). Because of this, there were no evaluations for the >.50" liquid equivalent of snow and >.01" freezing rain forecasts at Day 7. Evaluations were done on a scale of 1-5, with 1 representing a "very poor" forecast, 3 a "fair" forecast, and 5 a "very good" forecast.

No guidance was superior across all forecasts and time periods, but the combined (90 member multi-ensemble) guidance was the best when considering overall performance for all three forecast products.

Table 7. Average rating from subjective evaluation, on a scale of 1-5 (1="very poor", 5="very good"), of probabilistic model guidance for the three Day 4-7 probabilistic thresholds for the four guidance systems.

24 hour Probability of Winter Precipitation >.10"

	Number	Day 4	Day 5	Day 6	Day 7	Overall
Combined	50	3.43	3.21	3.42	3.63	3.52
ECENS	50	3.35	3.00	3.35	3.75	3.14
GEFS	50	3.57	3.14	3.5	3.35	3.20
CMCE	50	3.57	3.07	3.57	3.38	3.22

24 hour Probability of Liquid Equivalent (snow/sleet) >.50"

	Number	Day 4	Day 5	Day 6	Day 7	Overall
Combined	25	3.57	4.00	4.2	--	3.80
ECENS	25	3.57	3.83	4.2	--	3.76
GEFS	25	3.62	3.83	4.2	--	3.78
CMCE	25	3.50	3.83	4.2	--	3.72

24 hour Probability of Freezing Rain >.01"

	Number	Day 4	Day 5	Day 6	Day 7	Overall
Combined	25	3.79	3.83	3.8	--	3.80
ECENS	25	3.86	3.83	3.6	--	3.80
GEFS	25	3.97	3.66	4.0	--	3.68
CMCE	25	2.53	3.16	4.0	--	2.88

Evaluators noticed that the guidance seemed to be under-dispersed overall, especially in regard to the >.10" winter weather product. There were several instances when participants expressed concern that probabilities provided by the model guidance were too high, particularly at Days 6 and 7, where 50%, 70% and even 90% contours were produced. Additionally, model guidance tended to miss capturing large areas of observed precipitation in their probabilities, although a dominant trend was for observed areas to fall near, but not completely in, probability contours. Discussion during evaluation sessions highlighted that these errors appeared to be due to timing issues in the guidance (i.e. ensemble systems tended to be too fast with system progression throughout the Day 4-7 period). In several instances this trend was then applied to Day 4-7 forecasts later in the week, as forecast teams would increase the spatial coverage of contours, or pull the edge of the model guidance contours farther west, when making their forecast.

Evaluations showed the >.50" liquid equivalent of snow product was successful at identifying the temporal and spatial areas of high impact events. Participants were again concerned with model guidance generating high probabilities at longer lead times (>50% at Days 6 and 7), but major events were regularly observed in instances where higher probabilities were shown, although often with associated spatial and temporal errors. Additional discussion during evaluations focused on the .50" threshold itself. Many participants felt that .50" may be too high a threshold, as several high magnitude snowfall events fell shy of this amount. There were several suggestions to reduce the .50" threshold in order to provide more utility, or to include a separate, lower threshold (e.g. .25" liquid equivalent).

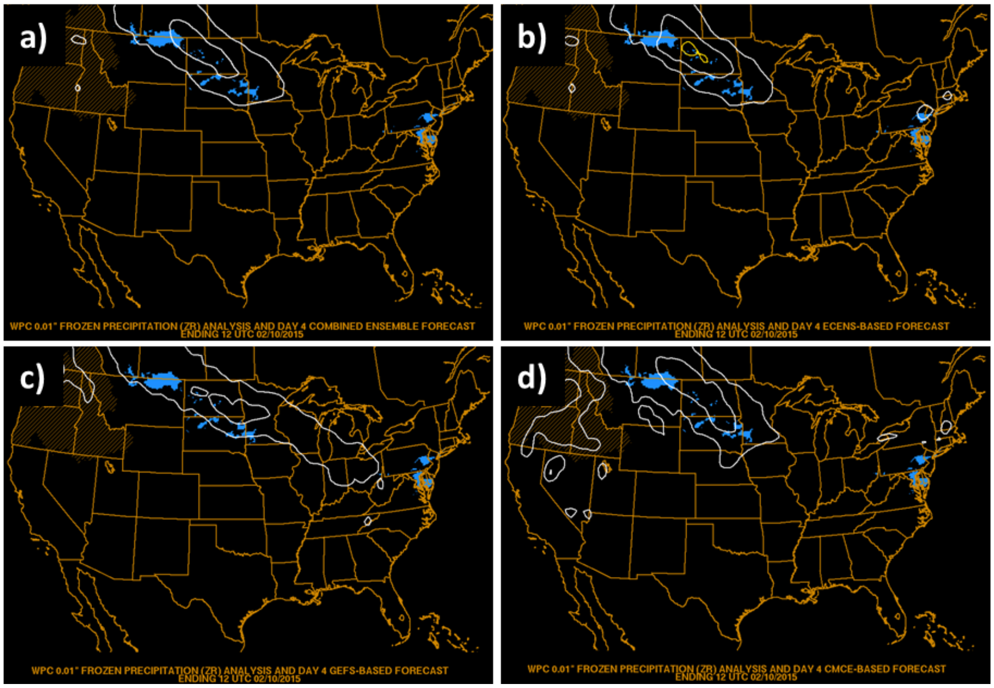

Figure 15. Showing the 24 hour probability of >.01" freezing rain from the a) combined (90 member multi-model ensemble) guidance, (b) ECENS guidance, (c) GEFS guidance and (d) CMCE guidance for a Day 4 (96 hour) forecast valid 12Z 10 Feb 2015. The first two white contours represent >10% and >30% probability, and the yellow contour represents >50% probability. The blue shading indicates areas of analyzed freezing rain, using a combination of hourly StageIV QPE data and hourly RAP precipitation type analysis.

Regarding the .01" freezing rain guidance, there was a general consensus that the guidance did well overall in identifying the larger scale/higher impact icing events (Fig. 15), but also that the guidance was too expansive on the spatial coverage of the freezing rain threat overall. Several participants expressed that this wasn't necessarily the fault of the model guidance, but spoke more toward the general low predictability of freezing rain at such long lead times. There was, however, additional concern over the applicability of the CMCE guidance, which consistently (and erroneously) signaled a freezing rain threat in areas in the intermountain west and along the Mexican border of New Mexico and Texas. Figure 15 highlights one example of a Day 4 forecast that captured a large event in the northern Great Plains well, but also missed a significant event in the Mid-Atlantic. The CMCE guidance created spurious probabilities in Nevada. Participant feedback suggested raising the freezing rain threshold to .05" or .10", which would create a product that highlighted events of higher impact, but may also help to alleviate spurious guidance.

6. SUMMARY AND OPERATIONAL IMPACTS

The fifth annual HMT-WPC Winter Weather Experiment was conducted January 12 – February 13, 2015. This year's experiment focused on exploring emerging short range microphysics-based snowfall forecasting techniques as well as continuing to expand and test winter weather forecasts for the Day 4-7 period. Over the course of the four week experiment, 28 participants issued experimental short range deterministic snowfall forecasts and probabilistic medium range winter weather outlook forecasts, and conducted subjective evaluations of experimental forecasts and model guidance. For the first time, participants gave a daily forecast webinar to members of the operational, research and academic communities to expand exposure of the experimental forecasts and guidance featured during the experiment.

The experiment confirmed that new microphysics-based parameters and snowfall forecasting techniques provide useful information to the forecaster, particularly in helping identify areas of mixed precipitation and/or transition zones. However, the effectiveness of these products is limited by the need for post-processing procedures, in which instantaneous parameters are applied to 1, 3 or 6 hour blocks of time. The effectiveness of these products is also tied to the accuracy of their parent numerical model. Expansion of advanced microphysical output into additional deterministic models and ensembles will further improve their effectiveness in improving winter weather forecasts.

This year's experiment also continued to demonstrate the viability of Day 4-7 winter weather outlook forecasts, including adding additional thresholds and precipitation types. Feedback for the Day 4-7 probabilistic winter weather suite was overwhelmingly positive among experiment participants. New forecast products focusing on heavy snowfall and freezing rain potential showed promise, but likely need further consideration and/or development.

A number of the experiment findings are directly relevant to operational winter weather forecasters and future forecasting experiments:

- **Microphysics based parameters provide added value to the forecaster.** The rime factor (RF) and percent of frozen precipitation (POFP) parameters provide valuable information on the type of hydrometeors in the lowest model level, helping provide clarity on if precipitation will fall as graupel, wet snow, dry snow or even mixed precipitation. Continued investigation is needed on how to best display these parameters, especially in ensembles, in order to provide streamlined information to forecasters. Additionally, Training materials on how to access and use these instantaneous fields in the NAM will be provided to WFO forecasters.

- **Future development of numerical models should include focus on providing advanced microphysical parameters in their output**, including expansion into ensembles and global models. Increasing the suite of guidance that provides advanced microphysics information will offer increasingly detailed information regarding the uncertainty associated with winter weather forecasts.

- **Continued movement toward eliminating post-processing is necessary to improve model snowfall forecasts.** The first step is to **provide frozen QPF in the model output**, which would represent the amount of precipitation that reached the ground in a frozen state. This would eliminate the need to apply instantaneous fields (e.g. precipitation type, POFP) to the QPF in post-processing. The second development is to **work toward a model-implicit snowfall**, which would use frozen QPF and an appropriate SLR to generate snowfall continuously through time in the model run.

- **Advancements to WPC's Day 4-7 probabilistic winter weather guidance suite were well received.** Preliminary results suggest that new developmental work, including using a 90 member multi-center ensemble, increased the effectiveness of the probabilistic guidance and helped to increase diversity. However, model guidance still remains under-dispersed overall.

- **Experimental heavy snow and freezing rain products showed promise,** with feedback showing that operational forecasters would find these products useful. WPC will continue investigating proper thresholds, including a proposed .25" threshold for the liquid equivalent of snow and .10" threshold for freezing rain, to maximize their effectiveness as potential forecast products.

The 2015 HMT-WPC Winter Weather Experiment provided an opportunity to bring the forecasting, research, and model development communities together to explore the challenges associated with both short-term and medium range winter weather forecasting. The experiment identified several potential areas for improvement, which will continue to be explored by HMT-WPC both in the coming months and in future experiments.

ACKNOWLEDGEMENTS

The HMT-WPC Winter Weather Experiment would not be possible without the dedication of a host of individuals, including Tom Workoff, Faye Barthold, Mike Bodner, Tony Fracasso, Mark Klein, Dave Novak, Mike Musher, Dan Petersen, Paul Kocin, and Frank Pereira. We would like to thank Brad Ferrier (EMC) and Eric Aligo (EMC), who developed the rime factor-modified snowfall accumulation technique, Jun Du (EMC) for providing the parallel SREF, and Isidora Jankov (ESRL) for providing the NARRE.

REFERENCES

Baxter, M. A., C. E. Graves, and J. T. Moore, 2005: A climatology of snow-to-liquid ratio for the contiguous United States. *Wea. Forecasting*, **20**, 729-744.

De Pondeca, M. S. F. V., G. S. Manikin, G. DiMego, S. G. Benjamin, D. F. Parrish, R. J. Purser, W-S. Wu, J. D. Horel, D. T. Myrick, Y. Lin, R. M. Aune, D. Keyser, B. Colman, G. Mann, and J. Vavra, 2011: The Real-Time Mesoscale Analysis at NOAA's National Centers for Environmental Prediction: Current status and development. *Wea. Forecasting*, **26**, 593-612.

Hamill, T. M., G.T. Bates , J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau, Y. Zhu and W. Lapenta, 2013: NOAA's second-generation global medium-range ensemble reforecast dataset. *Bull. Amer. Meteor. Soc*., **94**, 1553-1565.

Lin, Y. and K. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: Development and applications. Preprints. *19th Conf. on Hydrology*, San Diego, CA., 1.2.

Roebber, P. J., S. L. Bruening, D. M. Schultz, and J. V. Cortinas, 2003: Improving snowfall forecasts by diagnosing snow density. *Wea. Forecasting*, **18**, 264-287.

Von Storch, H., and F.W. Zwiers 1999. *Statistical Analysis in Climate Research*. Cambridge University Press, 496 pp.

Wilks, D.S., 2006. *Statistical Methods in the Atmospheric Sciences, 2nd Edition*. International Geophysics, 648 pp.

APPENDIX A

Participants

Week	WPC Forecaster	NCEP/WFO	Research/Academia/ Private Sector	EMC
Jan 12 -16	Frank Pereira	Brian LaSorsa (LWX) Greg Heavener (PHI) Nathan Marsili (IWX)	Jennifer Tate (NCSU) Isidora Jankov (ESRL) Sara Ganetis (SBU) Matt Sienkiewicz (SBU)	Yihua Wu Geoff Manikin
Jan 20-23	Mike Musher	Frank Nocera (BOX) Julie Malingowski (GJT) Chris Gibson (MSO) Jaret Rogers (SPC)	Ellen Sukovich (ESRL)	Corey Gaustini
Feb 2-6	Paul Kocin	Pete Banacos (BTV) David Hotz (MRX) Mike Bettwy (AWC)	Marty Baxter (CMU) Jim Steenburgh (Utah) Nathan Korfe (SBU)	Eric Aligo
Feb 9-13	Dan Petersen	Michael Fries (PBZ) Tim Gingrich (AKQ) Tom Dang (STO)	Ed Szoke (ESRL) Bruce Veenhuis (MDL) Brian Kolts (FirstEnergy)	Jeff McQueen

APPENDIX B

Daily Schedule

8:00am – 10:00am	Determine forecast area of interest and time period (Day 1 or Day 2); issue experimental 24 hr forecasts; make forecast presentation
10:00am – 10:30am	WPC-CPC map discussion
10:30am – 11:30am	Subjective short-term model and forecast evaluation
11:30am – 12:30pm	Lunch
12:30pm – 1:00pm	Weather briefing*
1:00pm – 2:00pm	Subjective Day 4-7 model and forecast evaluation
2:00pm – 3:30pm	Issue experimental Day 4-7 winter weather outlook forecasts
3:30pm – 4:00pm	Group discussion

APPENDIX C

Parallel SREF Configuration

Member	Initial Conditions	Physics	Convection
NMMB_CTL	NDAS	Ferrier_hires	BMJ old shallow
NMMB_N1	NDAS	WSM6	SAS
NMMB_P1	NDAS	Ferrier_hires	BMJ new shallow
NMMB_N2	NDAS	Ferrier	SAS
NMMB_P2	NDAS	WSM6	BMJ old shallow
NMMB_N3	GFS	Ferrier_hires	SAS
NMMB_P3	GFS	WSM6	BMJ new shallow
NMMB_N4	GFS	WSM6	SAS
NMMB_P4	GFS	Ferrier_hires	BMJ old shallow
NMMB_N5	RAP	WSM6	SAS
NMMB_P5	RAP	Ferrier_hires	BMJ new shallow
NMMB_N6	RAP	Ferrier_hires	SAS
NMMB_P6	RAP	WSM6	BMJ old shallow
ARW_CTL	RAP	WSM6	KF
ARW_N1	RAP	Ferrier	BMJ
ARW_P1	RAP	Thompson	Grell
ARW_N2	RAP	Ferrier	KF
ARW_P2	RAP	Thompson	BMJ
ARW_N3	GFS	WSM6	Grell
ARW_P3	GFS	Thompson	KF
ARW_N4	GFS	WSM6	BMJ
ARW_P4	GFS	Ferrier	KF
ARW_N5	NDAS	Ferrier	Grell
ARW_P5	NDAS	WSM6	KF
ARW_N6	NDAS	Thompson	BMJ
ARW_P6	NDAS	Thompson	Grell