The 2014 HMT-WPC Winter Weather Experiment

Final Report

April 15, 2014

1. INTRODUCTION

The Hydrometeorological Testbed at the Weather Prediction Center (HMT-WPC) hosted 36 forecasters, researchers, and model developers (Appendix A) at its fourth annual Winter Weather Experiment from January 21 – February 21, 2014. In addition to 23 on-site participants, for the first time 13 forecasters participated in a portion of the experiment remotely. This year's experiment focused on exploring the use of emerging short range microphysics-based snowfall forecasting techniques while also continuing to explore 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 coupling with a land surface model.
- Explore the utility of parallel versions of the NAM, SREF, and GFS as well as the experimental ExREF system for winter weather forecasting.
- Explore new datasets to improve the winter weather outlook (Day 4-7) forecast process.
- Gather feedback about 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 2014 experiment featured three 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). Given the experiment emphasis on utilizing information provided by the model microphysics schemes, priority was given to events that featured precipitation type transitions when determining the daily forecast area. In addition to the 24 hr snowfall forecast, participants were also asked to issue a separate 6 hr snowfall forecast (Fig. 1b) for the period considered most critical to the overall evolution of the event. This forecast was designed to highlight the period of heaviest snow (ex: mesoscale banding), the expected timing of the transition from one type of precipitation to another, etc. For both forecasts, participants were asked to draw 2”, 4”, 8”, 12”, and 20” snowfall contours as well as indicate the highest snowfall amount they expected within the forecast domain.

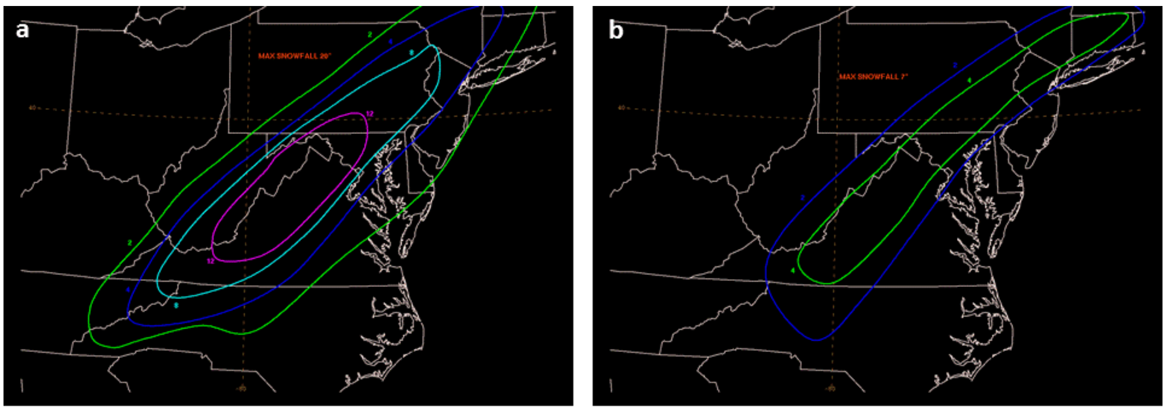

Figure 1. Example of an experimental Day 1 (a) 24 hr and (b) 6 hr snowfall forecast issued during the 2014 HMT-WPC Winter Weather Experiment.

b. Experimental Medium Range Forecasts

During the afternoon, participants used a variety of derived guidance to issue experimental 24 hr probabilistic winter weather outlook forecasts (Fig. 2) for the Day 4-7 (84-180 hr forecast) period. These forecasts were identical to those being prototyped on WPC’s operational Winter Weather Desk during the 2013-2014 winter season and highlighted the probability of receiving at least 0.10” precipitation (liquid equivalent) in the form of snow, sleet, or freezing rain. Participants were asked to draw probability contours indicating a 10%, 30%, 50%, 70%, and 90% chance of winter precipitation, and the forecasts were valid 12 – 12 UTC.

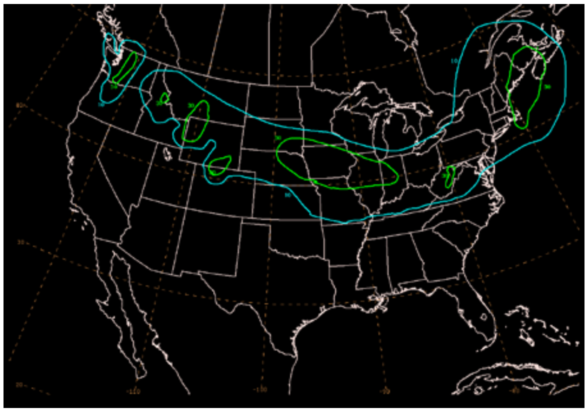

Figure 2. Example of an experimental Day 5 forecast issued during the 2014 HMT-WPC Winter Weather Experiment indicating the probability of frozen precipitation ≥ 0.10".

c. Forecast Discussion and Subjective Model Evaluation

For the first time, this year's Winter Weather Experiment offered the opportunity to participate remotely in the subjective model evaluation sessions. Although limited in scope, this option allowed for expanded participation among local NWS offices. On-site participants were asked to prepare a forecast discussion presentation that was used to provide a briefing about the day's forecasting activities to the remote participants at the start of each remote session. These forecast discussions used a combination of text and graphics to explain the synoptic and mesoscale forecast rationale, discuss any forecast challenges, and highlight any notable differences or interesting features in the experimental model guidance. In addition, participants were asked to indicate their confidence in the forecast as above average, average, or below average.

Following the daily forecast briefing, the on-site and remote participants worked together to subjectively evaluate the performance of both the experimental forecasts and the corresponding experimental model guidance for events from the previous week of the experiment. The subjective evaluations consisted of a series of survey questions and associated graphics designed to determine whether the experimental model guidance provided information to forecasters that was more or less useful than the information that could be gleaned from the traditional operational guidance alone.

The evaluations of the experimental short range forecasts and the associated 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; no extrapolation is allowed. This strict analysis 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 (NOHRSC, http://www.nohrsc.noaa.gov/interactive/html/map.html).

The evaluations of the experimental medium range forecasts were conducted using a 4 km gridded analysis of frozen precipitation greater than or equal to 0.10” developed by WPC. This analysis is based on a 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 relatively strict 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. Due to a lack of hourly Stage IV data, this analysis was not available across the Northwest River Forecast Center’s area.

Data

In addition to the full multi-center suite of numerical model guidance available to WPC forecasters, the 2014 experiment featured a variety of experimental short-term and medium range deterministic and ensemble guidance products (Table 1). The Short Range Ensemble Forecast (SREF) and North American Mesoscale Model (NAM) were used as the operational baselines during the experiment.

Table 1. Models and ensembles evaluated during the 2014 HMT-WPC Winter Weather Experiment. All models were initialized at 00 UTC except SREF and SREFP guidance (21 UTC). Experimental guidance is shaded.

Provider	**Model**	**Resolution**	**Forecast Hours**	**Snow-to-Liquid Ratio (SLR) / Notes**
EMC	SREF (21 members)	16 km (32 km display)	87	For temperatures < 5°C: $SLR = (273.15 - T_{2m}) + 8$ capped at a maximum ratio of 28:1
EMC	SREFP (21 members)	16 km (32 km display)	87	For temperatures < 5°C: $SLR = (273.15 - T_{2m}) + 8$ capped at a maximum ratio of 28:1
EMC	SREFP-RF (21 members)	16 km (32 km display)	87	*SLR* = Rime factor modification of SREF SLR
ESRL	ExREF (8 members)	9 km	84	SLR based on 2 m temperature such that : *SLR* = 10:1 for $T_{2m} \geq 10°F$ *SLR* = 15:1 for $T_{2m} < 10°F$
EMC	NAM	12 km (32 km display)	84	*SLR* = Roebber Technique
EMC	NAM	12 km	84	*SLR* = Rime factor-modification to Roebber Technique
WPC	N/A	20 km	96-168	Probabilistic guidance of >.1" of winter precipitation falling in 24 hrs; precipitation type based on GEFS, ECENS and GEFS+ECENS
ESRL	GEFS Reforecast	32 km	180	2nd generation GEFS reforecast dataset; 24 hr QPF and PQPF guidance

a. Short Range Forecast Guidance

Two experimental short-range ensemble systems were featured. The first was the parallel version of the SREF (SREFP), which uses the Rapid Refresh (RAP) model instead of the GFS for the initial conditions of the seven WRF-ARW members. The second was the Experimental Regional Ensemble Forecasting Systems (ExREF), provided by NOAA's Earth System Research Laboratory (ESRL). The ExREF is a CONUS, 9 km resolution, multi-physics, multi-initial condition, multi-boundary condition ensemble system (Table 2). 7 of its 8 members feature use of the Local Analysis and Prediction System (LAPS; http://laps.noaa.gov) for their initial conditions, with the first member using the GFS analysis. Each member generates precipitation type (rain, snow, freezing rain, ice/mix) based on thickness thresholds. During the experiment, changes were implemented to the ExREF's initial condition perturbation scheme in order to create more diversity and spread between the members in the short-term (Bernardet et al. 2014).

Table 2. Membership characteristics of the ExREF. Member denoted by asterisk (*) denotes use of the "variational" version of the LAPS analysis; all others use the "traditional" version.

Member	Initial Conditions	Boundary Conditions	Microphysics
m00	GFS	GFS	Thompson
m01	LAPS	GFS	Thompson
m02	LAPS	GEFS 01	Ferrier
m03	LAPS	GEFS 02	WSM6
m04	LAPS	GEFS 03	Thompson
m05	LAPS	GEFS 04	Ferrier
m06	LAPS	GEFS 05	WSM6
m07	LAPS*	GFS	Thompson

Also featured were experimental snowfall and snowfall accumulation techniques. This year's experiment continued the 2013 experiment's investigation of the rime factor-modified (RF) snowfall by expanding it to the parallel SREF, as well as continuing to apply it to the operational and parallel versions of the NAM. The rime factor technique 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 (rime factor). The modified SLR (SLR_{RF}) is then used in conjunction with the percentage of frozen precipitation (POFP) in the lowest model level to calculate the rime factor-modified snowfall:

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

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)	$SLR_{RF} = \frac{SLR}{6}$

In the NAM, the rime factor modified the SLR derived using the Roebber Technique (Roebber et al. 2003). For the parallel SREF, the 2 m temperature-based SLR (Table 1) is modified in each ensemble member prior to calculating the ensemble mean value. In addition to the ensemble mean, mean snowfall forecasts for each of the model cores (ARW, NMM, NMMB) were created to allow for further investigation of core differences.

In addition to the rime factor-modified snowfall products, the utility of using model snow depth forecasts as a proxy for snow accumulation was evaluated. Three 24 hr change in snow depth forecasts were provided: one from the operational deterministic European Centre for Midrange Weather Forecasting (ECWMF) model, one from the operational NAM, and one from the parallel NAM. In the parallel NAM product, the rime factor output is coupled with the Noah Land Surface Model (LSM) to expand the range of possible snow densities (Fig. 3) such that higher rime factor values allow for a greater increase in the snow density, which in turn decreases the snow depth.

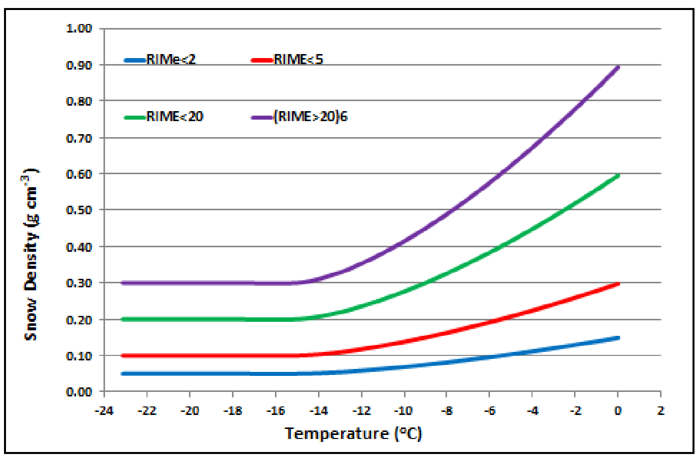

Figure 3. Relationship between temperature and snow density for different rime factor categories used by the Noah Land Surface Model

b. Medium Range Forecast Guidance

In addition to the short range 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 (Hamill et al. 2013), which is a dataset of historical (1985-2010) weather forecasts generated by re-running the version 9.0.1 of NCEP's Global Ensemble Forecast System (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-processing of the current model forecast. The probabilistic and deterministic QPF products are created via analog approaches, using the North American Regional Reanalysis as the observed precipitation analogs. More detailed information about the analog method used in the creation of the QPF guidance can be found on ESRL's 2nd Generation Reforecast website (http://www.esrl.noaa.gov/psd/forecasts/reforecast2/analogs/index2.html).

Additionally, WPC generated probabilistic guidance products denoting the probability of ≥.10" of frozen precipitation falling in a 24 hour period (probability of winter precipitation). To create this guidance, 6 hr precipitation forecasts from the GEFS mean are used to disaggregate WPC's operational Day 4-5 and 6-7 QPF in order to construct separate 24 hr QPFs for each of the four days. Using this 24 hr QPF as the mean value, a cumulative distribution function (CDF; Von Storch and Zwiers 1999; Wilks 2006) of QPF ≥.10" is then created using the members of the GEFS (20 members) and ECMWF Ensemble Prediction System (ECENS, 50 members) as the variance. The ensemble probability of frozen precipitation (snow, sleet, and freezing rain) is then calculated based on the precipitation type identified in each GEFS member (via NCEP dominant precipitation type; Manikin 2005) and/or each ECENS member (using a local WPC precipitation type algorithm); this is then combined with the probability of QPF ≥ 0.10" to produce the probability of winter precipitation. Three probability guidance products were created; one using the 20 GEFS members for precipitation type (denoted hereafter as GEFS-Based), one using the 50 ECENS members for precipitation type (ECENS-Based), and one combined product using both the GEFS and ECENS (70 members) for precipitation type (Combined).

3. CASES

Like much of the winter, the experiment period was characterized by a trough over the eastern two-thirds of the country with a ridge over the west coast (Fig. 4a). This pattern allowed for multiple Arctic incursions into the eastern United States, resulting in anomalously cold temperatures all the way to the Gulf coast. At the same time, the ridge along the west coast resulted in temperatures several degrees above normal across much of the southwestern U.S. (Fig. 4b).

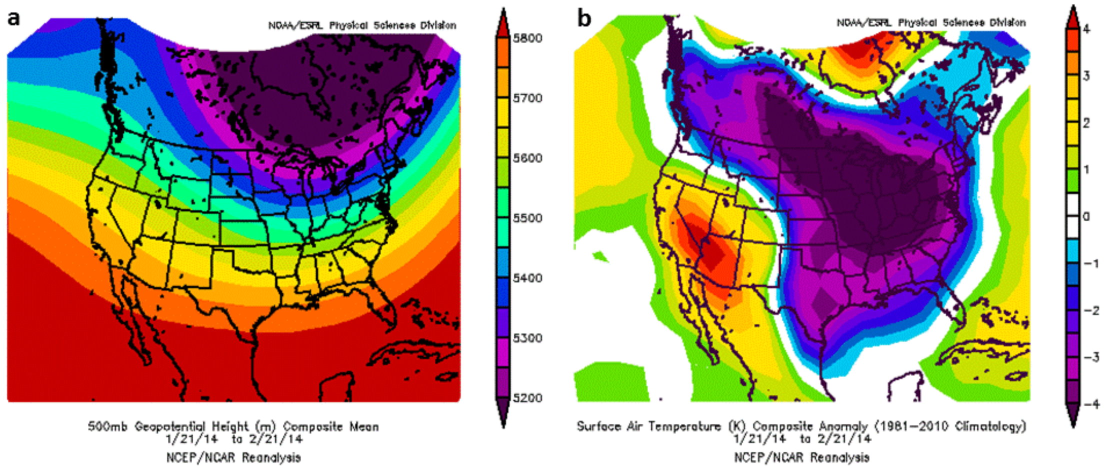

Figure 4. Composite (a) mean 500 hPa heights and (b) surface temperature anomalies for the 21 January – 21 February 2014 period. Images generated from the NCEP/NCAR Reanalysis provided by NOAA/ESRL/PSD (http://www.esrl.noaa.gov/psd/data/composites/day/).

The below normal temperatures across much of the country during the experiment period resulted in numerous significant winter weather events. The experiment got off to a fast start with an event during the first day of the experiment across the Mid-Atlantic and Northeast that brought 12” of snow to parts of New York and southern New England. With cold air entrenched in the east, much of the southeastern United States got a taste of winter at the end of January with a storm that brought snow and freezing rain to areas from central Mississippi to southeastern Virginia. Although only a few inches of snow fell, this event resulted in widespread power outages and traffic gridlock across the Atlanta, GA metro area. Farther north, more than 0.10” freezing rain and 1-3” of snow fell across coastal North and South Carolina. The Mid-Atlantic was revisited by a second major winter storm in mid-February, when at least 12” of snow fell from northwestern North Carolina into southern Pennsylvania. Finally, the experiment wrapped up with a significant snowfall event in the upper Midwest that brought 12” of snow to parts of northern Minnesota and Wisconsin. A complete list of the snowfall events investigated during this year’s experiment can be found in Table 4.

Table 4. Experimental short-range forecasts and subjective verification for the 2014 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 forecasters in the weeks following the experiment to provide a more robust evaluation.

Forecast Valid Time	Forecast		Verification		Forecast Area	Notes
00Z 4 Jan 2014			D1	D2	Mid Atlantic to Northeast	
00Z 6 Jan 2014			D1	D2	Midwest to Ohio Valley	
00Z 22 Jan 2014			D1	D2	Mid Atlantic to Northeast	Significant snowfall from Mid Atlantic into Northeast
00Z 23 Jan 2014	D1		D1	D2	Mid Atlantic to Northeast	Significant snowfall from Mid Atlantic into Northeast
00Z 26 Jan 2014	D1	D2	D1	D2	Great Lakes and Ohio Valley	
00Z 30 Jan 2014	D1	D2	D1	D2	Southeast to Mid Atlantic	Significant snow and ice across southeastern U.S.; widespread power outages
00Z 1 Feb 2014		D2	D1	D2	Rockies to Central Plains	
00Z 2 Feb 2014	D1	D2	D1	D2	Central Plains to Great Lakes	
00Z 13 Feb 2014		D2	D1	D2	Southeast to Mid Atlantic	Significant snowfall from Mid Atlantic into Northeast
00Z 14 Feb 2014	D1	D2	D1	D2	Mid Atlantic to Northeast	Significant snowfall from Mid Atlantic into Northeast
00Z 21 Feb 2014	D1	D2	D1	D2	Upper Midwest	Significant snowfall across Upper Midwest
00Z 22 Feb 2014	D1		D1	D2	Upper Midwest	Significant snowfall across Upper Midwest
00Z 23 Feb 2014	D1		D1	D2	Central and Northern Rockies	

4. EXPERIMENTAL SHORT RANGE FORECASTS

During experiment operations 15 short term deterministic forecasts, 8 for Day 1 and 7 for Day 2, were created. Through a combination of subjective evaluations completed during the experiment and supplemental evaluations completed after, a total of 26 cases were evaluated, 13 on both Day 1 and Day 2.

Figure 5a shows the confidence that was assigned to each forecast by the forecast team. A majority (10/15) of the forecasts were issued with 'average' confidence; there was a general decline in confidence from Day 1 to Day 2, as no Day 2 forecasts were issued with 'above average' confidence.

A majority of the forecasts (9/15) were rated as 'fair' during evaluation, with 5 cases being rated as 'good' and only 1 'poor.' Some of the common issues noted by evaluators 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). However, participants noted that there were often significant discrepancies between the WPC and NOHRSC snowfall analysis, which made providing consistent and accurate evaluations difficult.

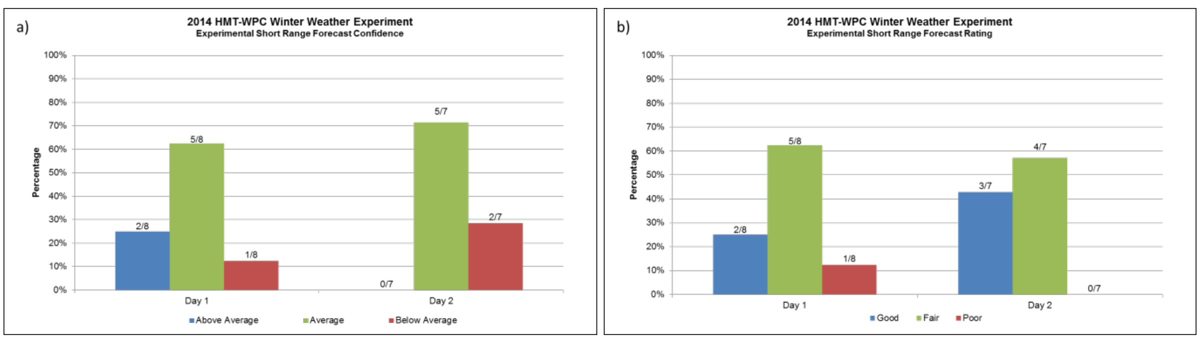

Figure 5. (a) Forecast confidence assigned to the short range deterministic forecasts, and (b) the rating assigned to each forecast during the subjective evaluation.

NAM and SREFP Rime Factor-Modified Snowfall Accumulations

The benefit of the RF-modified snowfall guidance differed depending on the parent model to which it was applied (Fig. 6). For the NAM, participants found the RF-modified guidance provided information that was either similar or slightly more useful in the forecast process as the NAM Roebber. There were no instances where the RF snowfall was deemed less useful than the Roebber in a Day 1 forecast.

The most common benefit of the NAM RF snowfall noted by participants was the increased resolution at which it is displayed (12 km compared to the NAM Roebber at 32 km). While the increased resolution allowed for improved spatial detail, it also had the tendency to overdo snowfall amounts in topography. In many cases it was considered an improvement over the

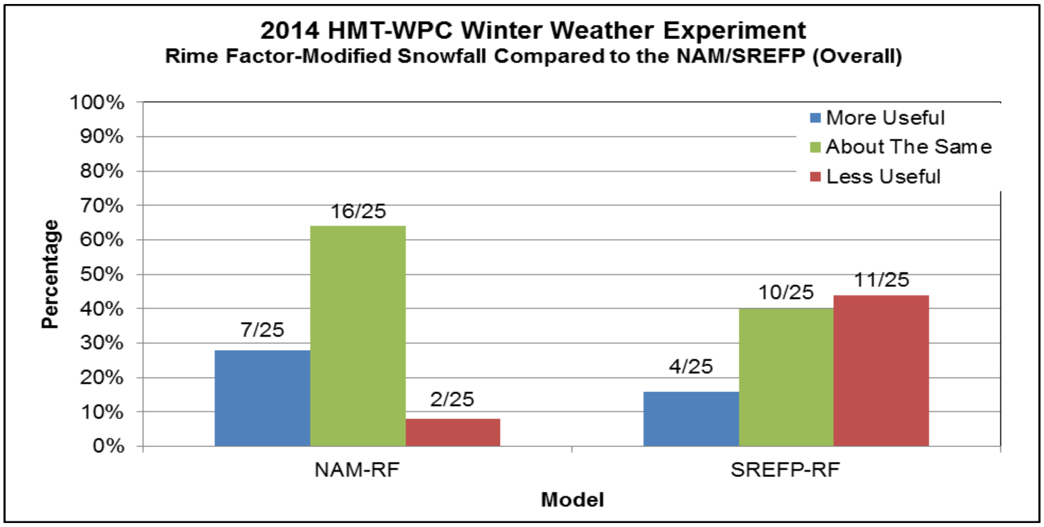

Figure 6. Experimental model performance for NAM rime factor-modified snowfall (left) and the SREFP rime factor-modified snowfall (right) based on participant feedback from subjective model evaluations. Participants were asked whether the rime-factor modified forecasts more useful, less useful, or about the same when compared to their corresponding operational baselines; the 00 UTC NAM rime factor was compared to the operational 00 UTC NAM Roebber snowfall, and 21 UTC SREFP rime factor was compared to the 21 UTC SREFP snowfall.

Roebber product in specific areas, but when considering the entire domain the improvement was not enough to justify a 'more useful' rating in the evaluation.

Aside from resolution differences, the utility of the NAM RF snowfall product was maximized on the warmer boundary of systems and in precipitation-type transition zones (e.g. rain changing to snow). Figure 7 shows an example encompassing the 24 hour period ending 00 UTC 30 January, when a winter storm brought a mix of rain, sleet, freezing rain and snow to the southeast United States. The RF snowfall (Fig. 7d) correctly increases accumulations in central GA by identifying that snow would fall across this region when the Roebber technique (using NCEP dominant precipitation type) did not (Fig. 7c). In addition, the RF snowfall reduces amounts on a small scale in eastern NC in a region where mixed precipitation occurred (not shown). Accordingly, participants noted that using the actual rime factor (Fig. 7e) and POFP (Fig. 7f) data helped alert them that a mix of frozen precipitation featuring heavily rimed particles would occur during the event. While participants were encouraged by the potential of this new approach, they found the actual values of the rime factor (Table 3) to be difficult to apply physically, and generally preferred using the percent of frozen precipitation in the forecast process.

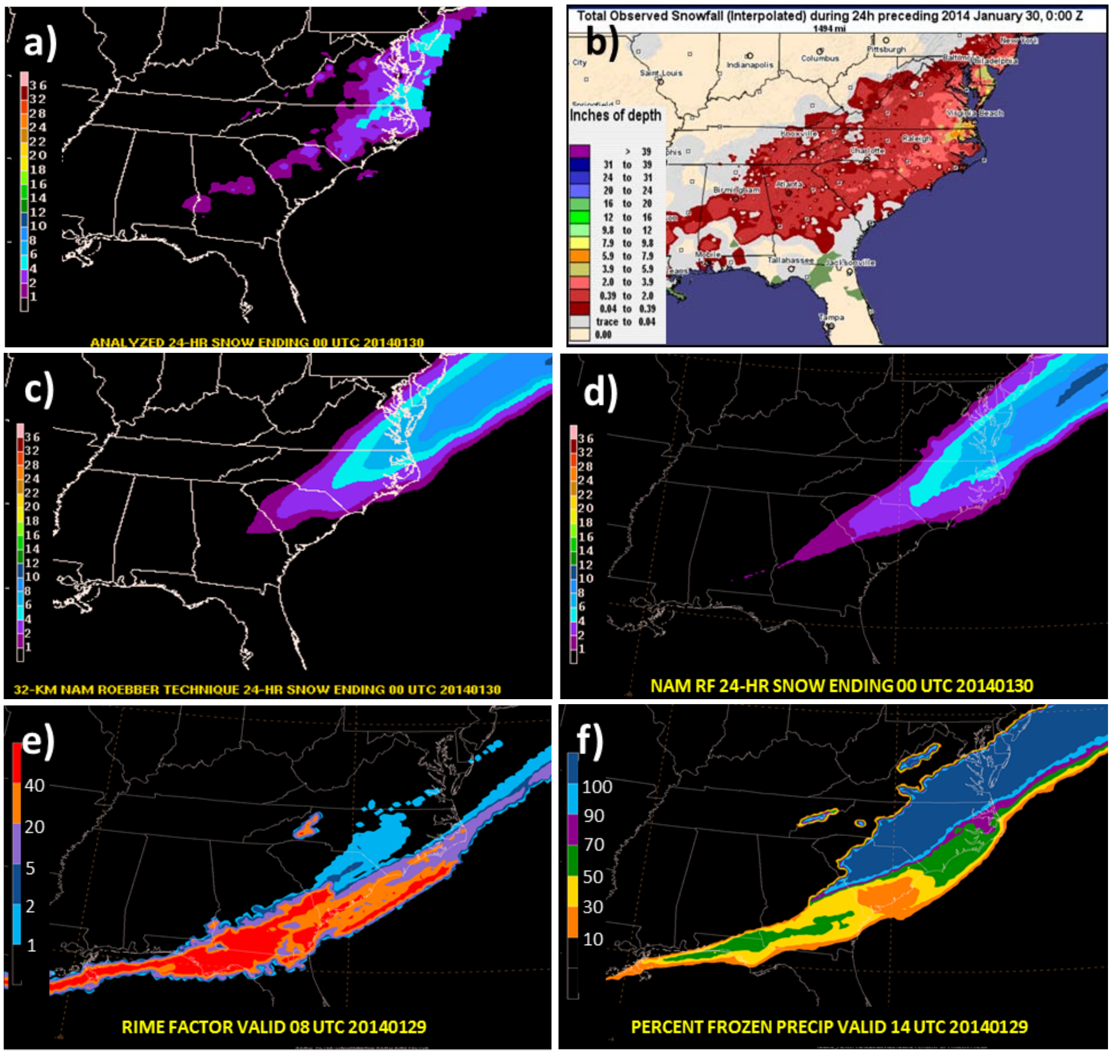

Figure 7. Observed 24 hour snowfall from ending 00 UTC 30 January 2014 from the (a) WPC snowfall and (b) NOHRSC snowfall analysis, the Day 1 24 hour snowfall forecast valid 00 UTC 30 January 2014 from the (c) NAM Roebber technique and (d) NAM RF-modified snowfall, the (e) NAM RF forecast valid 08 UTC 29 January 2014 and the (f) NAM percent of frozen precipitation forecast valid 14 UTC 29 January 2014.

For the SREFP, the rime factor-modified mean snowfall generally provided worse guidance (Fig. 6) than the 2 m temperature based algorithm (Table 1); this effect was magnified in Day 1 forecasts, where 58% of cases (7/12) were ruled less useful than the SREFP snowfall (compared to 31% of Day 2 forecasts). The main issue with the rime-factor modified SREFP was its tendency to produce whole-scale reductions in the snow field; instead of reducing snowfall amounts only in local areas such as transition zones, it noticeably (and often incorrectly) reduced the snowfall forecast over an entire region, sometimes by several inches (Fig. 8). Participants found this was consistently a degradation of the forecast when compared the

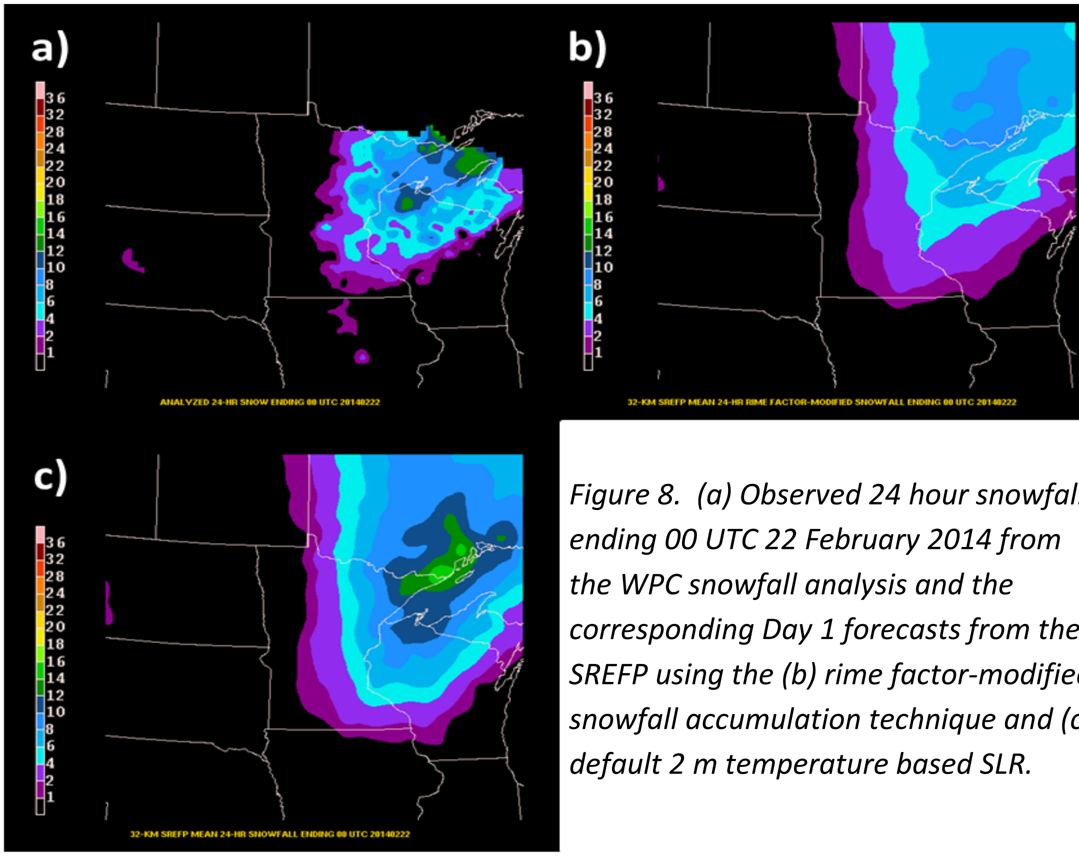

Figure 8. (a) Observed 24 hour snowfall ending 00 UTC 22 February 2014 from the WPC snowfall analysis and the corresponding Day 1 forecasts from the SREFP using the (b) rime factor-modified snowfall accumulation technique and (c) default 2 m temperature based SLR.

SREFP, even in cases where the reduced amounts were not enough to downgrade the subjective rating. These reductions continue to be investigated and are believed to be due to the POFP and rime factor values in each individual ensemble member.

SREFP and ExREF

The SREFP forecasts were consistently very similar to those from the operational SREF. In many cases, the magnitude of the differences was small enough that participants felt there were no tangible differences between the forecasts (Fig. 9, left side). In cases where larger magnitude differences were observed, the SREFP tended to provide more useful guidance; this was consistent across Day 1 (4 of 13 cases rated as more useful) and Day 2 (5 of 13 cases).

The results from the ExREF were bimodal, with the guidance being considered more useful and less useful than the operational SREF in 7 of 19 cases (Fig. 9, right side). Participants noted that the increased resolution was generally a benefit, particularly in areas of topography and the handling of mesoscale features (e.g. lake-effect snow). However, it was noted that the guidance tended to be somewhat inconsistent; in some cases the heaviest snowfall axis was

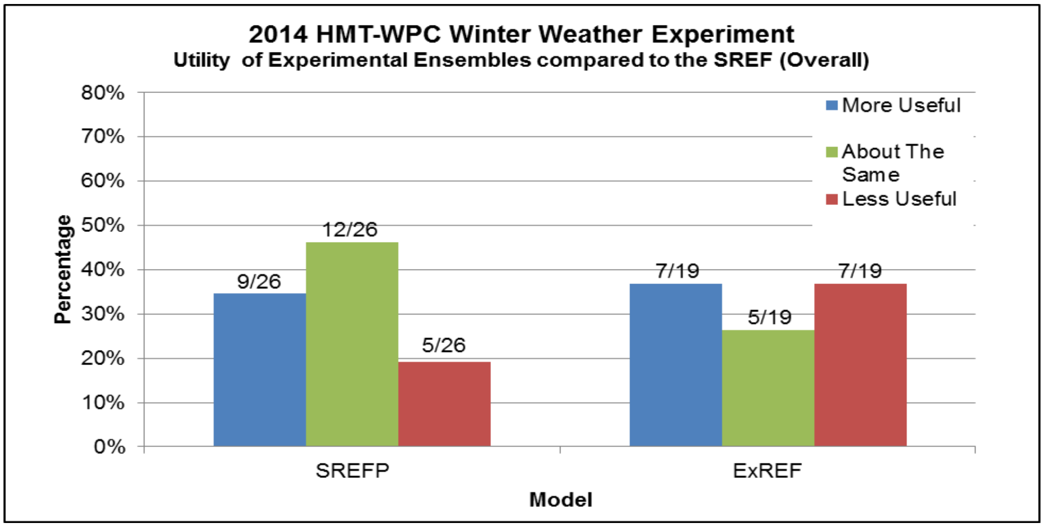

Figure 9. Experimental model performance for the parallel SREF (SREFP) mean and ExREF mean snowfall based on participant feedback from subjective model evaluations. Participants were asked to determine whether the SREFP and ExREF forecasts were more useful, less useful, or about the same when compared to operational SREF mean; the 00 UTC cycle was used for the ExREF, and the 21 UTC cycle was used for the parallel and operational SREF.

displaced too far to the north and west, and participants noted that it was equally likely to either under or over-forecast snowfall amounts throughout the course of the experiment. The ExREF did undergo changes to its initial condition perturbation scheme in early February, and this change appears to have switched the model's snowfall from having a high bias (before change) to a low bias (after change), but this warrants further investigation.

Snow Depth Parameters (NAM, ECWMF, NAM parallel rime factor)

Overall, the subjective evaluations revealed that the 24 hr change in snow depth parameters showed a slight bias towards being too small when compared to snowfall observations; this bias was magnified for the parallel NAM's rime factor-Noah LSM coupled product. Participants noted that the snow depth guidance did not provide additional value to the forecast process when compared to model snowfall output. However, feedback from evaluations showed that many feel that the idea of coupling model snowfall with a land surface model is a promising step toward improving snowfall accumulations, but that the process is currently too unrefined to provide any benefit.

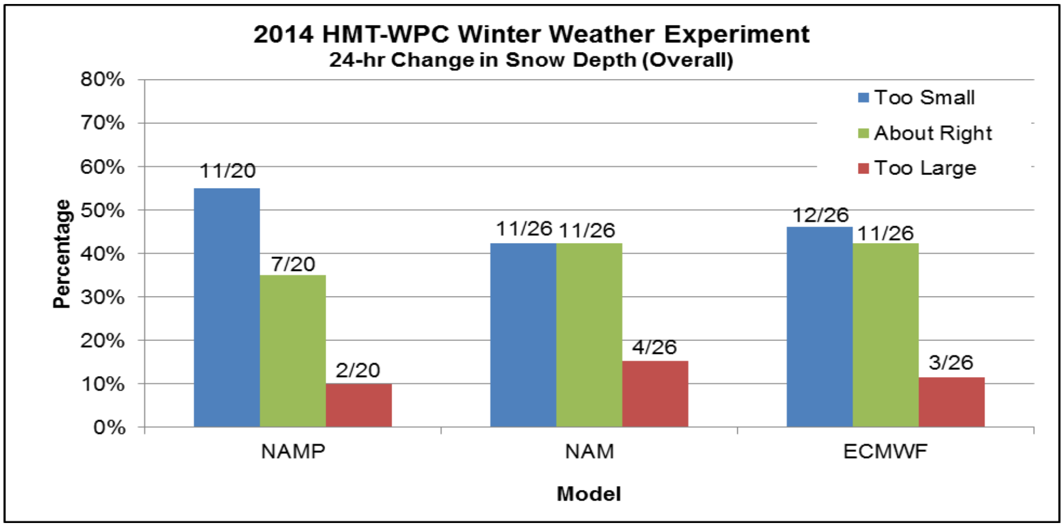

Figure 10. Experimental model performance for the 24 hr change in snow depth from the parallel NAM, operational NAM and operational ECMWF models. Participants were asked to determine whether the magnitude of the forecasted 24 hr change in snow depth from each model was too large, too small, or about right when compared to the WPC snowfall analysis.

5. EXPERIMENTAL MEDIUM RANGE FORECASTS

In addition to the short-term snowfall forecasts, participants were also asked to issue winter weather outlook forecasts for the Day 4-7 period. Featuring these forecasts in the Winter Weather Experiment provided an opportunity to gather more direct feedback about both the potential utility of these forecasts to local WFO operations and the quality of some of the available guidance datasets.

Depending on the forecast day, 10-13 experimental forecasts were available for evaluation. The percentage of experimental forecasts rated as “good” is highest on Day 4 and generally decreases with increasing forecast lead time (Fig. 11). The opposite trend is seen for forecasts rated as “poor”.

In general, experiment participants thought that expanding winter weather forecasts into the Day 4-7 time period would be a good addition to WPC’s product suite. However there were differing views of the product among participants based on their geographical and climatological perspective. For example, participants from the southern U.S. found the current precipitation threshold of 0.10” useful since it alerts them to the possibility of almost any frozen precipitation, which is generally a fairly unusual event in that part of the country. Participants from the northern U.S., however, indicated that the 0.10” precipitation threshold

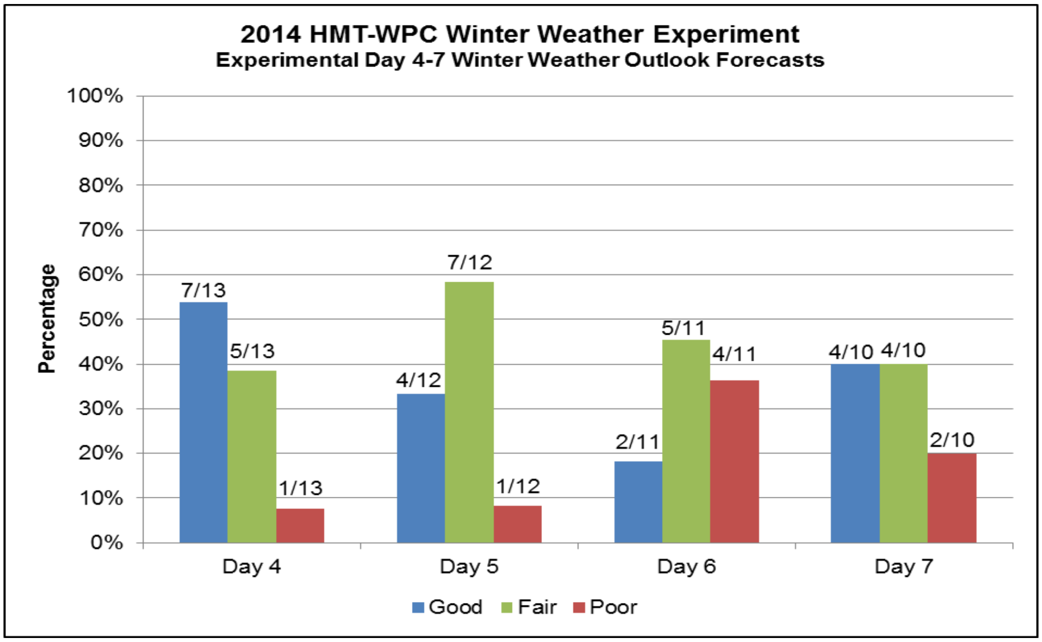

Figure 11. Subjective evaluation of experimental Day 4-7 forecasts based on the WPC analysis of frozen precipitation greater than or equal to 0.10" (liquid equivalent). Participants were asked to rate each forecast as good, fair, or poor.

wasn't meaningful to them, since frozen precipitation occurs fairly routinely during the winter. One solution may be to generate the product for several different precipitation thresholds, similar to WPC's operational probabilistic winter precipitation forecasts, instead of trying to identify a single precipitation threshold that provides forecast value everywhere. Participants also suggested adjusting the values of the contours and including additional information such as the corresponding forecasts of fronts and pressures to provide meteorological context.

Participants were also asked to provide feedback about the utility of a number of different datasets for making these winter weather outlook forecasts. In particular, participants were asked to rank the probabilistic Day 4-7 guidance developed by WPC from most useful to least useful for both the Day 4 and the Day 7 winter weather outlook forecasts (Fig. 12). In cases for which all three versions of the guidance were available (GEFS-Based, ECENS-Based, and Combined), the guidance based on precipitation type from the European ensemble members (ECENS-Based) was considered most useful for Day 4 forecasts. None of the guidance options stood out at Day 7, but it is important to note that the guidance based on a combination of GEFS and ECENS precipitation type information (Combined) was never considered the least useful dataset.

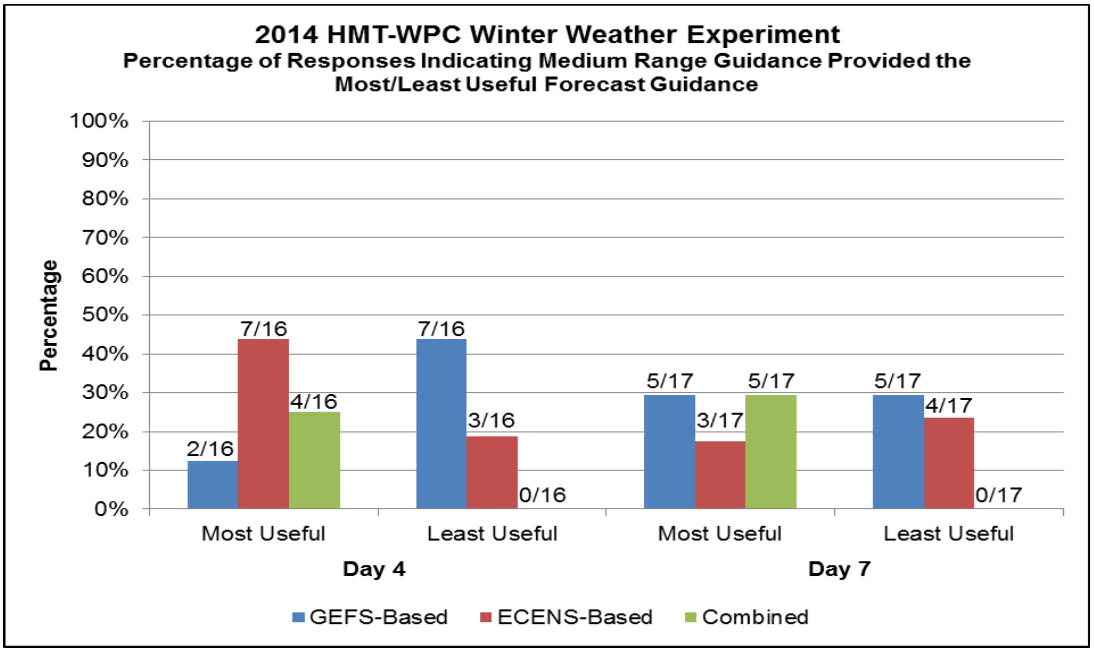

Figure 12. Subjective evaluation of forecast guidance utility for Day 4-7 winter weather outlook forecasts at both Day 4 and Day 7 in cases for which all three guidance variations were available. Participants were asked to rank the three datasets from most useful to least useful and were allowed to assign the same ranking to datasets if they determined that there was no difference in forecast utility.

While the subjective ranking indicates some preference for the ECENS-Based guidance, participants noted that in general the three guidance options were very similar to one another. In addition, there were several winter weather events during the experiment that were not indicated by any of the three WPC probabilistic guidance options during the Day 4-7 time period. This implies that as currently constructed, none of the available guidance is capable of fully representing the range of possible forecast solutions.

As a result of both these findings as well as input from other WPC forecasters who have participated in the prototype on the operational Winter Weather Desk during the 2013-2014 winter, HMT-WPC has recommended a number of changes to the Day 4-7 winter weather outlook forecast procedure:

- **Combine ensemble systems to generate an improved starting point for the Day 4-7 forecasts.** As noted above, the guidance based on a combination of GEFS and ECENS precipitation type information was never found to be the least useful dataset. In addition, expanding the starting point ensemble system should also help address the issue of observed events falling outside the ensemble envelope.

- **Investigate methods to increase diversity in the available guidance datasets.** Currently, each of the three guidance datasets tested during the experiment use the operational WPC QPF as the mean of the ensemble precipitation distribution, which is thought to constrain the range of possible solutions. Diversity can be improved by generating separate guidance based on each ensemble's respective QPF.
- **Use the same precipitation type methodology across all ensemble systems.** During the experiment, different precipitation type algorithms were used in different ensemble systems. This made it more difficult to compare results across ensemble systems since some of the algorithms were conditional on precipitation and others were not.

Finally, in addition to the ensemble-based guidance developed by WPC, participants were also asked to consider guidance from the GEFS reforecast dataset. While this guidance wasn't formally evaluated, the general consensus among participants was that the reforecast data typically provided useful guidance. Participants noted that they would be interested in obtaining longer term verification statistics about the reforecast dataset to better understand its overall performance. The positive reaction to the reforecast data further supports HMT-WPC's position that this dataset should be supported and upgraded operationally. Operational support would allow for greater use among WPC forecasters and would help foster the development of additional forecast tools.

6. SUMMARY AND OPERATIONAL IMPACTS

The fourth annual HMT-WPC Winter Weather Experiment was conducted January 21 – February 21, 2014. This year's experiment focused on exploring emerging short range microphysics-based snowfall forecasting techniques as well as continuing to test winter weather forecasts for the Day 4-7 period. Over the course of the four week experiment, 23 on-site participants issued experimental short range deterministic snowfall forecasts, probabilistic medium range winter weather outlook forecasts, an prepared a brief forecast discussion. For the first time, the on-site participants were joined by 13 additional remote participants to evaluate the utility of the available experimental guidance.

The experiment revealed that while new microphysics-based snowfall forecasting techniques can provide useful information to the forecaster in precipitation type transition zones, discerning the details of winter weather events remains a challenge even 48-72 hours into the future. Building off of the 2013 experiment, while there are still numerous opportunities for improvement this year's experiment once again demonstrated the viability of Day 4-7 winter weather outlook forecasts. A number of the experiment findings are directly relevant to operational winter weather forecasters and future forecasting experiments:

- **The current WPC snowfall analysis is inadequate.** In order to provide meaningful verification information, it is imperative for WPC to use an analysis that incorporates a more comprehensive set of snowfall observations. If available in a gridded form, the snowfall analysis produced by NOHRSC appears to be a reasonable option.
- **A number of potential improvements to both the available Day 4-7 guidance and the resulting forecast product have been identified** and will be transitioned to WPC operations. In addition to the proposed changes to the forecast guidance, the precipitation threshold and probability contours used in the product should be revisited to ensure that it provides value to WPC's customers.
- **The utility of the rime factor-modified snowfall is maximized in three situations: on the warmer boundary of systems (south/east), in areas of mixed or transitioning precipitation type, and areas where the NCEP dominant precipitation type incorrectly classifies sleet and/or freezing rain as snow.** However, rime factor and percent of frozen precipitation may hold more value as stand-alone parameters and may not be applicable to ensemble mean solutions at this time. HMT-WPC will continue to work with EMC to fine tune the usability and display of these parameters.
- **The GEFS reforecast dataset was well-received, but needs to be operationally supported in order to be used to its full potential and further objective verification is needed to better quantify the potential benefits.** Operational support would allow these data to be more fully incorporated into the forecast process as well as foster the development of additional forecast tools.
- **Remote participation was considered a success, but further advancements can be made to improve the experience and increase the value for remote participants.** Remote participants appreciated the information provided in the daily forecast discussions, but more effort should be made to include experimental data and/or include them directly in the forecast process. Investigating new technology that would allow for direct sharing from a Linux workstation (such as Webex) could provide a significant advantage.

The 2014 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 in the coming months.

ACKNOWLEDGEMENTS

The HMT-WPC Winter Weather Experiment would not be possible without the dedication of a host of individuals including Faye Barthold, Tom Workoff, Mike Bodner, Wallace Hogsett, Dave Novak, Rich Otto, Mike Musher, Dan Petersen, and Jim Hayes. Brad Ferrier (EMC) and Eric Aligo (EMC) developed the rime factor-modified snowfall accumulation technique and provided support throughout the experiment, while Mike Ek (EMC) coupled the rime factor output with the Noah Land Surface Model. Jun Du (EMC) provided the parallel SREF. Ligia Bernardet (ESRL) and Isidora Jankov (ESRL) provided the ExREF. Tom Hamill (ESRL) and Gary Bates (ESRL) were instrumental in providing the GEFS reforecast dataset.

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.

Bernardet, L. R., I. Jankov, S. Albers, K. Mahoney, T. Workoff, F. Barthold, W. Hogsett, D. Reynolds, and J. Du, 2014: The Experimental Regional Ensemble Forecast System (ExREF). *6th Conference on Weather Analysis and Forecasting / 22nd Conference on Numerical Weather Prediction,* Atlanta*, GA*, Amer. Meteor. Soc., 126.

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.

Hou, D., M. Charles, Y. Lou, Z. Toth, Y. Zhu, R. Krzysztofowicz, Y. Lin, P. Xie, D. –J. Seo, M. Pena, and B. Cui, 2013: Climatology-Calibrated Precipitation Analysis at fine scales: Statistical adjustments of Stage IV towards CPC gauge based analysis. In press, *J. Hydrometeor*.

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.

Manikin, G. S., 2005: An overview of precipitation type forecasting using NAM and SREF data. *21st Conference on Weather Analysis and Forecasting/17th Conference on Numerical Weather Prediction*, Washington, D.C., 8A.6.

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 21 – 24	Rich Otto	Melissa Ou (CPC) Al Cope (PHI)* Rich Kinney (BIS)* Darren VanCleave (SAC)* Jack Settelmaier (SRH)*	Dave Kingsmill (NOAA HMT) John Wagner (MDL) Pete Manousos (First Energy)	Jacob Carley
Jan 27 – 31	Mike Musher	Frank Nocera (BOX)* David Barjenbruch (BOU)* Justyn Jackson (AMA)* Tony Fuentes (REV)*	Pat Market (Missouri) Isidora Jankov (NOAA HMT)	Brad Ferrier
Feb 10 – 14	Dan Petersen	Adrienne Leptich (OKX) Josh Boustead (OAX)* Jeff Vitale (LUB)* Chauncy Schultz (BYZ)*	Paul Stokols (NWSHQ) Matt Sienkiewicz (SBU) Sarah Ganetis (SBU)	Eric Aligo
Feb 18 – 21	Jim Hayes	Brandon Smith (AWC) Brian LaSorsa (LWX) Brian Pettegrew (AWC) Chris Bowman (EAX) Trevor Alcott (WRH) Greg Heavener (PHI)* Kurt Buffalo (AMA)*		Jun Du

*remote participant

APPENDIX B

Daily Schedule

A brief orientation session will be held at 8:00am on the first day of each week to explain the motivation and organization of the experiment as well as the data being evaluated.

8:30am – 11:00am Determine forecast area and time period (Day 1 or Day 2).

Using 00 UTC guidance, issue an experimental 24 hr deterministic snowfall forecast (2”, 4”, 8”, 12”, and 20” contours) for the 00 – 00 UTC period. Within this period, identify the most critical 6 hr period (based on precipitation type transitions, heaviest snowfall, etc.) and issue a separate 6 hr deterministic snowfall forecast. For both forecasts, indicate the highest expected snowfall amount within the forecast domain.

Prepare the short-term portion of the forecast discussion presentation.

11:00am – 11:30am WPC-CPC map discussion

11:30am – 12:30pm Lunch

12:30pm – 2:00pm *remote access* Subjectively evaluate the performance of the experimental guidance using the WPC snowfall analysis.

2:00pm – 4:00pm Issue experimental CONUS Day 4-7 winter weather outlook forecasts, valid 12 – 12 UTC.

Prepare medium range portion of the forecast discussion presentation.

4:00pm – 4:30pm Group discussion