

The 2012 HMT-HPC Winter Weather Experiment

Final Report

April 12, 2012



1. INTRODUCTION

The Hydrometeorological Testbed at the Hydrometeorological Prediction Center (HMT-HPC) hosted 21 forecasters, researchers, and model developers (Appendix A) at its second annual Winter Weather Experiment from January 9 – February 10, 2012. This year's experiment focused on using ensemble systems to help quantify and communicate uncertainty in winter weather forecasts. Specifically, the experiment aimed to answer 5 questions:

- What added benefits do storm scale ensembles provide for winter weather forecasting?
- Is the new NCEP SREF system an improvement over the current operational system?
- How can forecasters add value to probabilistic winter weather forecasts?
- How can post-processed products, such as ensemble clusters, be used in the forecast process?
- How can awareness of societal impacts improve forecast and decision support services?

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

2. EXPERIMENT DESCRIPTION

Data

The 2012 Winter Weather Experiment featured three experimental ensemble systems (Table 1). NCEP's Environmental Modeling Center (EMC) provided a pre-implementation version of the new Short Range Ensemble Forecast System (SREFP, scheduled for operational implementation in late spring 2012). The SREFP increases the resolution from 32 km to 16 km, replaces the ETA and Regional Spectral Model (RSM) members with members from the Non-hydrostatic Mesoscale Model on a rotated Arakawa B-grid (NMMB), uses more diverse initial conditions and physics schemes, and offers additional post-processing techniques. The Air Force Weather Agency (AFWA) provided a 10-member 4 km convection allowing ensemble. The AFWA ensemble features a multi-physics and multi-initial condition platform that incorporates initial conditions from the United Kingdom Meteorological Office's Unified Model (UKMET), NCEP's Global Forecast System (GFS), the Canadian Global Environmental Multiscale model (CMC), and the Navy Operational Global Atmospheric Prediction System (NOGAPS) (Table 2). Finally, the HPC provided the HPC Superensemble, an internally generated 28-member 32 km ensemble consisting of all 21 operational SREF members, 2 GFS Ensemble Forecast System (GEFS) mean members, one European Center for Medium Range Weather Forecasting (ECMWF) ensemble mean member (ECENS), and one member each from the deterministic North American Mesoscale model (NAM), GFS, CMC, and ECMWF. Data from the operational SREF were used for comparison to the experimental ensembles.

Each ensemble system featured a different snow-to-liquid ratio algorithm that was used to derive ensemble maximum, minimum, and mean snowfall accumulations as well as various accumulation probabilities. The methodology used by each ensemble is provided in Table 1.

Table 1. Ensemble systems used during the 2012 Winter Weather Experiment. Both the operational and parallel versions of the SREF were initialized at 21 UTC while the AFWA and HPC Superensemble were initialized at 00 UTC. SREFP snow-to-liquid ratios were capped at 28:1. This cap was lowered to 20:1 during the last week of the experiment.

Provider	Model	Resolution	Forecast Hours	Snow-to-Liquid Ratio
EMC	SREF (21 members)	32 km	87	10:1
EMC	SREFP (21 members)	16 km	87	For temperatures < 5°C: $SLR = (273.15 - T) + 8$
AFWA	WRF (10 members)	4 km	54	For temperatures < 4°C: $SLR = 5\sqrt{5 - (T - 273.15)}$
HPC	Superensemble (28 members)	32 km	72	Average of: 11:1 climatology Roebber technique* applied to the NAM Roebber technique applied to the GFS

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

Table 2. AFWA ensemble configuration during the 2012 Winter Weather Experiment.

Member	IC/LBC	Microphysics	PBL
1	UKMET Unified Model	WSM5	YSU
2	UKMET Unified Model	Goddard	MYJ
3	UKMET Unified Model	ETA	QNSE
4	GFS	Thompson	MYJ
5	GFS	Thompson	YSU
6	GFS	Goddard	QNSE
7	Canadian GEM	ETA	QNSE
8	Canadian GEM	WSM5	QNSE
9	NOGAPS	Thompson	MYJ
10	NOGAPS	WSM5	YSU

In addition to the standard ensemble mean and probabilistic fields, several new forecasting tools were available during the experiment. The SREFP featured ensemble clustering as part of its output. This technique uses the spatial correlation between 500 mb height fields in the individual ensemble members to identify groups of members that are similar to one another (Alhamed et al. 2002; Yussouf et al. 2004). The goal of ensemble clustering is to identify realistic solutions that are not well represented by the full ensemble mean or individual deterministic runs. A critical temperature probability tool was developed in order to provide an estimate of the uncertainty associated with an individual deterministic model solution. Using the deterministic model forecast as the mean and the corresponding ensemble system spread as the variance (ex: GFS and GEFS), a cumulative distribution function (CDF) was estimated that

indicates the probability of the temperature being below critical freezing and precipitation type thresholds at various levels throughout the atmosphere. The goal of this tool is to identify spatial uncertainty in critical aspects of the thermal field. Finally, a trend reverse, or “flip-flop”, tool was developed based on the work of Zsoter et al. 2009. This tool identifies changes between the current model initialization, the cycle 12 hours prior, and the cycle 24 hours prior. More details about each of these tools are available in Appendix B.

Snowfall verification was based on the gridded HPC snowfall analysis. This analysis combines climatological snow-to-liquid ratios (Baxter et al. 2005) with Stage IV quantitative precipitation estimates (QPE) to generate a first guess snowfall accumulation. This initial field is then modified by COOP, CoCoRaHS, and METAR observations using a Barnes analysis to produce the final snowfall analysis.

Daily Activities

The 2012 experiment featured three complementary activities. A detailed version of the daily schedule can be found in Appendix C.

a. Experimental Forecasts and Subjective Model Evaluations

Each morning, participants used a combination of operational and experimental 00 UTC guidance (21 UTC for the SREF and SREFF) to create 24 hr experimental probability of exceedance forecasts for a storm of interest during either the Day 1 (24 – 48 hr) or Day 2 (48 – 72 hr) period. These forecasts covered the 00 – 00 UTC period and outlined areas that had a slight (10%), moderate (40%), and high (70%) probability of exceeding 2 in, 4 in, and 8 in (4 in, 8 in, and 12 in for high-end events) of snowfall (Fig. 1). When choosing the daily forecast area, priority was given to storms with the highest potential societal impact.

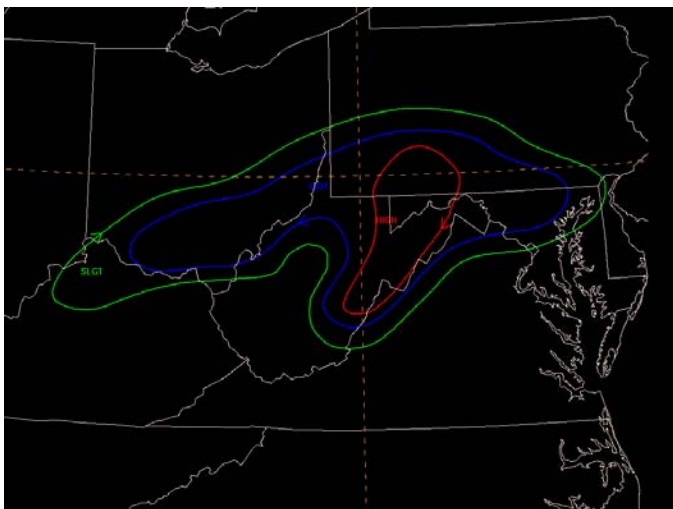


Figure 1. Experimental Day 1 forecast from 7 February 2012 indicating the probability of exceeding 2 inches of snowfall during the 24 hr period ending 00 UTC 9 February 2012.

In the afternoon, participants were asked to subjectively evaluate the performance of both the experimental forecasts and the experimental ensemble guidance for events from the previous week of the experiment. The subjective evaluations consisted of a series of survey

questions designed to determine whether the experimental ensemble guidance had provided better forecast guidance than the operational 21 UTC SREF.

b. Model Diagnostics Investigation

In addition to the experimental forecasting activities, a model diagnostics investigation was included in the experiment this year. The goal of this activity was to conduct a more rigorous investigation of model performance beyond the subjective evaluations of model snowfall forecasts as well as provide a more systematic view of potential issues that may impact forecast confidence. As part of the model diagnostics investigation, participants were asked to answer the following questions:

- Are there any meteorological features that the forecast team is unsure of?
- Can you diagnose any suspected model initialization errors that may influence these features?
- Is there run-to-run model continuity?
- Are there any outlier solutions or other unusual aspects of the model guidance?

After answering these questions, participants were then asked to use their answers to these questions and their understanding of the forecast situation to rate their overall forecast confidence as below average, average, or above average. This information was then translated into a forecast confidence discussion that combined elements of a traditional forecast discussion with information about forecast confidence in various aspects of the event (timing, precipitation type, snowfall amounts, etc.).

c. Societal Impacts Scenario

Another new aspect of this year's experiment was a focus on improving communication of forecast information to non-meteorological decision makers. As the role of decision support services expands within the National Weather Service (NWS), it becomes imperative to be able to communicate forecast information clearly. Each day, participants were asked to prepare a mock briefing for a non-meteorological decision maker focused on the potential impacts of the upcoming winter weather event. The focus of these briefings varied depending on the situation, but they were typically aimed at a regional transportation or emergency manager.

In order to prepare for the briefings, participants were introduced to the concept of inverting the typical scientific communication pyramid (Fig. 2) to better reach a public audience (Somerville and Hassol 2011). This communication paradigm stresses providing the bottom line first instead of detailed background information. In addition to the concept of the inverted communication pyramid, participants were provided with a list of forecast issues that are potentially important from a societal impacts perspective. These issues included the timing of the event (ex: rush hour impacts), reductions in visibility, snowfall rates, and the climatological rarity of the event.

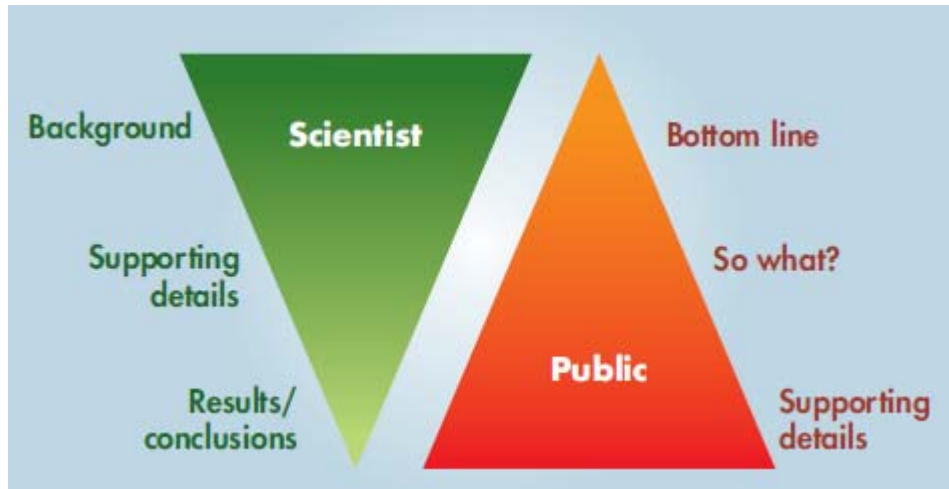


Figure 2. Comparison of the typical communication paradigms employed by scientists and the public (Somerville and Hassol 2011, their Fig. 3).

3. CASES

The experiment period was characterized by a broad mean trough over the central and eastern U.S. and a ridge over the west coast (Fig. 3a). More importantly, this winter was characterized by anomalously warm temperatures over much of the entire country. The experiment period was no different, with temperatures 3-4°C above average across much of the eastern half of the country (Fig. 3b). These anomalously warm temperatures contributed to lower than normal snowfall totals across much of the country.

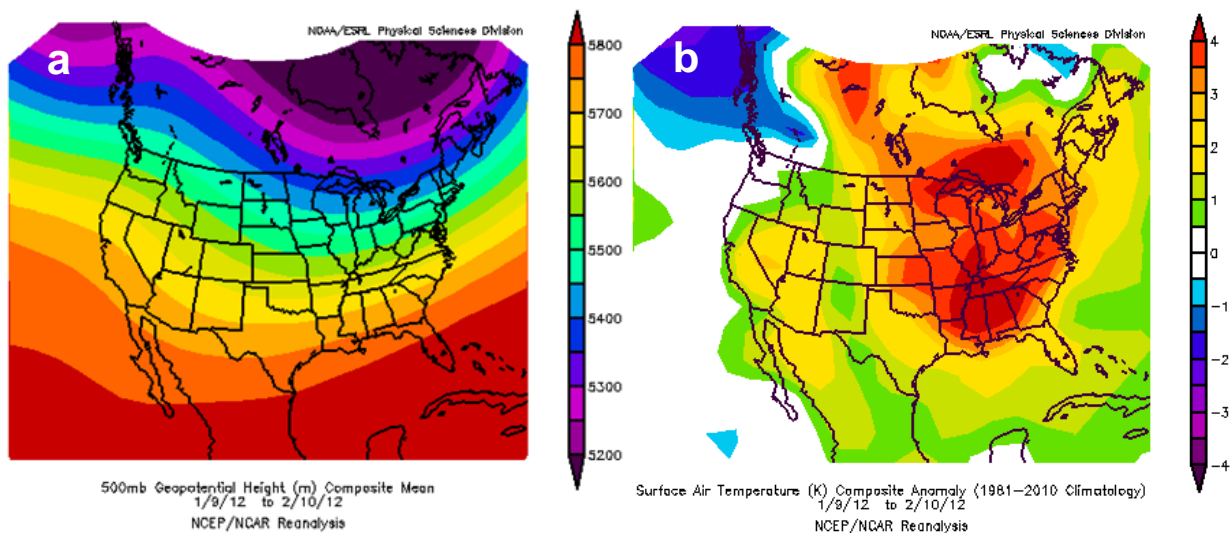


Figure 3. (a) Composite mean 500 mb heights and (b) surface air temperature anomalies during the January 9 – February 10, 2012 period. Images generated from the NCEP/NCAR Reanalysis provided by NOAA/ESRL/Physical Sciences Division (<http://www.esrl.noaa.gov/psd/data/composites/day/>).

While the 2012 experiment featured a less active pattern overall compared to the 2011 experiment, there were three notable storms during the experiment period. In mid-January, a series of storms brought accumulating snowfall to Seattle, including a record 6.8 inches on January 18th. Mid-January also brought widespread moderate snowfall to the northern Mid Atlantic and New England, with totals approaching 10 inches in parts of coastal Connecticut and Massachusetts. Finally, the beginning of February brought blizzard conditions to parts of the central U.S., with heavy snowfall extending from the Denver, CO area east into western Iowa. A complete list of the snowfall events investigated during this year’s experiment can be found in Table 3.

Table 3. Experimental forecasts and subjective verification for the 2012 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 HPC forecasters in the weeks following the experiment to provide a more robust evaluation.

Forecast Valid Time	Forecast		Verification		Forecast Area	Notes
00Z 1 Jan 2012			D1		Upper Midwest	
00Z 2 Jan 2012			D1		Upper Midwest to Great Lakes	
00Z 3 Jan 2012				D2	Great Lakes	
00Z 10 Jan 2012				D2	Southern Plains	
00Z 12 Jan 2012		D2	D1	D2	Northern/Central Rockies	
00Z 13 Jan 2012	D1	D2	D1	D2	Northeast	
00Z 14 Jan 2012	D1		D1	D2	Great Lakes	
00Z 16 Jan 2012		D2	D1	D2	Pacific Northwest to Northern Great Basin	
00Z 19 Jan 2012	D1		D1	D2	Pacific Northwest to Northern Great Basin	Accumulating snow in Seattle
00Z 21 Jan 2012		D2	D1	D2	Upper Midwest to mid Mississippi Valley	
00Z 22 Jan 2012	D1	D2	D1	D2	Mid Atlantic to Northeast	
00Z 26 Jan 2012	D1	D2	D1	D2	Pacific Northwest to Northern Great Basin	
00Z 27 Jan 2012	D1		D1	D2	Northern Great Basin to Northern Rockies	
00Z 28 Jan 2012	D1		D1	D2	Northeast	
00Z 29 Jan 2012	D1		D1	D2	Upper Great Lakes	
00Z 2 Feb 2012	D1	D2	D1	D2	Northeast	
00Z 4 Feb 2012	D1	D2	D1	D2	Central Plains	Significant snow in Denver/Front Range
00Z 5 Feb 2012	D1		D1	D2	Central Plains	
00Z 9 Feb 2012	D1	D2	D1	D2	Ohio Valley to Mid Atlantic	
00Z 11 Feb 2012		D2	D1	D2	Upper Great Lakes	
00Z 12 Feb 2012	D1	D2	D1	D2	Mid Atlantic to Northeast	

4. EXPERIMENTAL MODEL PERFORMANCE

Through a combination of evaluations completed during the experiment and supplemental evaluations completed after the experiment ended, a total of 38 cases were evaluated, 19 Day 1 forecasts and 19 Day 2 forecasts. All of the results shown below are based on these subjective evaluations.

SREFP

Overall, the SREFP provided better winter weather forecast guidance than the operational SREF (Fig. 4), with the bulk of the improvement occurring during the Day 2 forecast period (not shown). Participants found that the increased resolution in the SREFP resulted in realistic forecast details that allowed them to better target areas expected to receive heavy snowfall. In topography, this additional detail correctly confined snow to higher elevations.

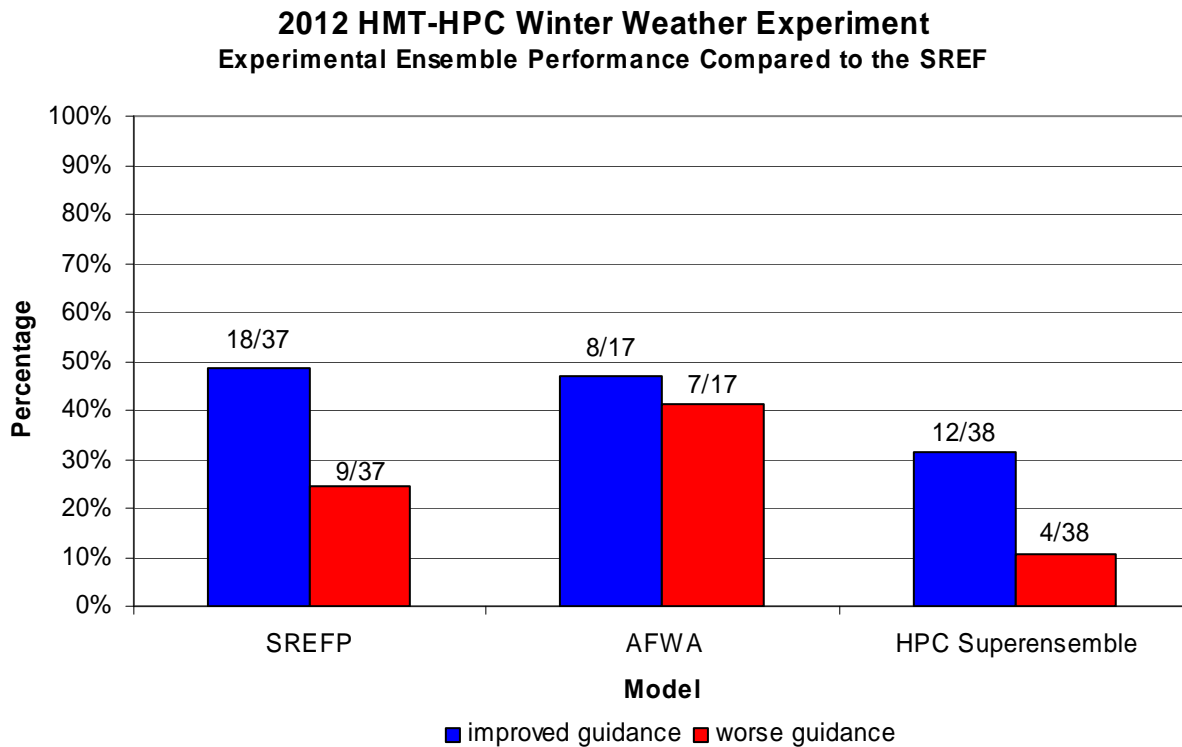


Figure 4. Experimental ensemble performance based on participant feedback from subjective model evaluations conducted during the 2012 HMT-HPC Winter Weather Experiment. Participants were asked to determine whether the ensemble mean snowfall forecasts from the 00 UTC experimental guidance (21 UTC SREFP) were much better, better, about the same, worse, or much worse than the guidance provided by the operational 21 UTC SREF, based on observations from the gridded HPC snowfall analysis. The AFWA ensemble was only available for the Day 1 (24 – 48 hr) forecast period.

Figure 5 shows an example of the benefit of the increased resolution in the SREFP in a Day 2 forecast of a lake-effect snow event. The HPC snowfall analysis indicated regions of heavier snow downwind of both Lake Michigan and Lake Erie as well as lake-enhanced upslope snowfall in the mountains of West Virginia. While the operational SREF mean identified three areas of snowfall, the SREFP mean provided much more detail and correctly enhanced snowfall amounts in all three regions. The difference is most noticeable in West Virginia, where the SREFP indicated a much greater threat for heavy snow than the operational SREF.

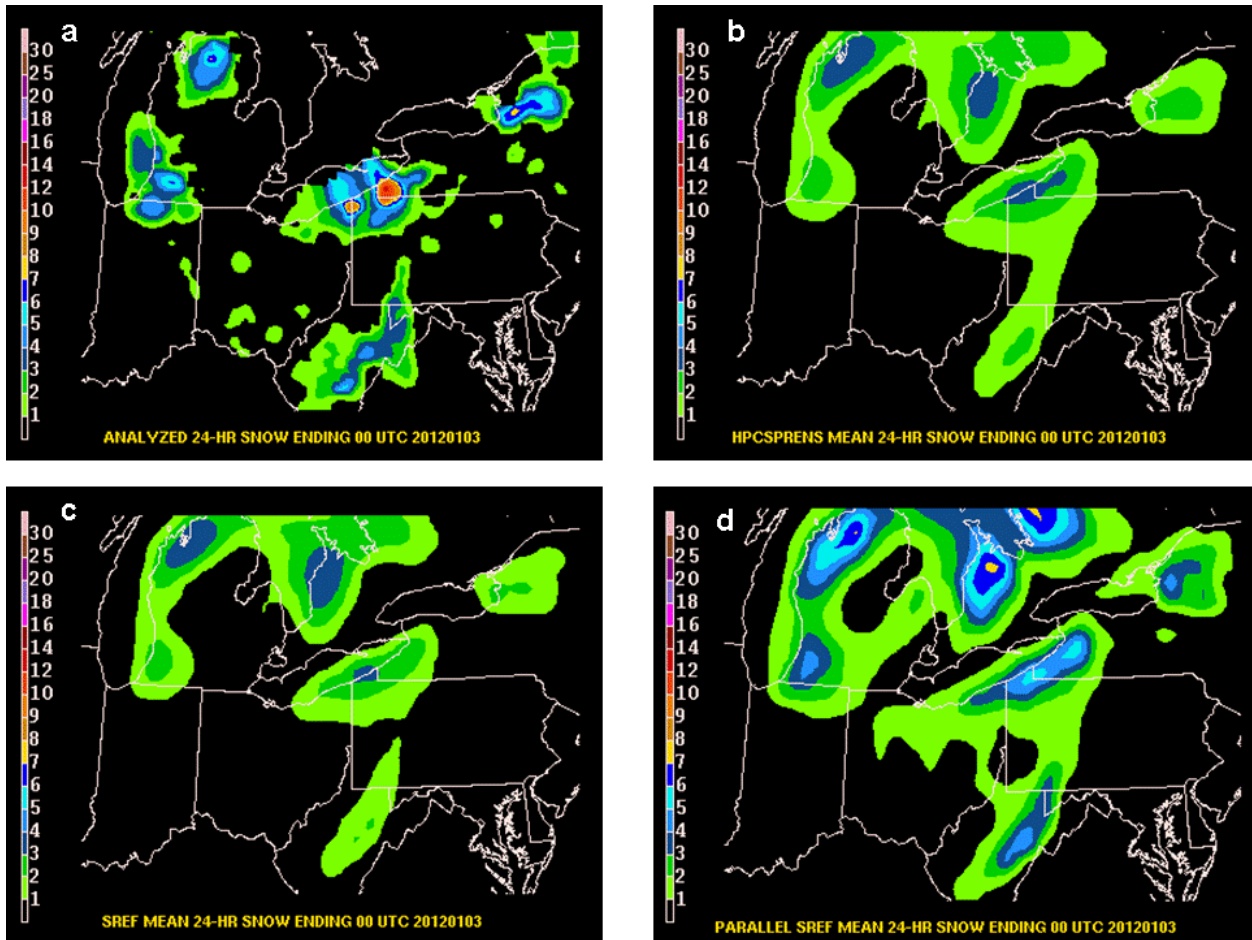


Figure 5. (a) Observed 24 hr snowfall ending 00 UTC 3 January 2012 from the HPC snowfall analysis and the corresponding Day 2 ensemble mean snowfall forecasts from the (b) HPC Superensemble, (c) operational SREF, and (d) SREFP.

Although overall the mean snowfall forecasts from the SREFP provided improved forecast guidance, there were numerous instances where the snowfall amounts appeared to be too high, especially as surface temperatures decreased. Similarly, while the SREFP generally provided improved probabilistic forecast guidance (Fig. 6), at times the snowfall probabilities appeared to extend too far back into the cold air. Unlike the current operational SREF, which uses a fixed 10:1 snow-to-liquid ratio, the SREFP uses a snow-to-liquid ratio that varies based on 2 m temperature (see Table 1). While snow-to-liquid ratio typically increases as temperatures decrease, it is dependent on more processes than the new algorithm represents, such as the

**2012 HMT-HPC Winter Weather Experiment
Ability to Capture 2"/24hr Snowfall Events**

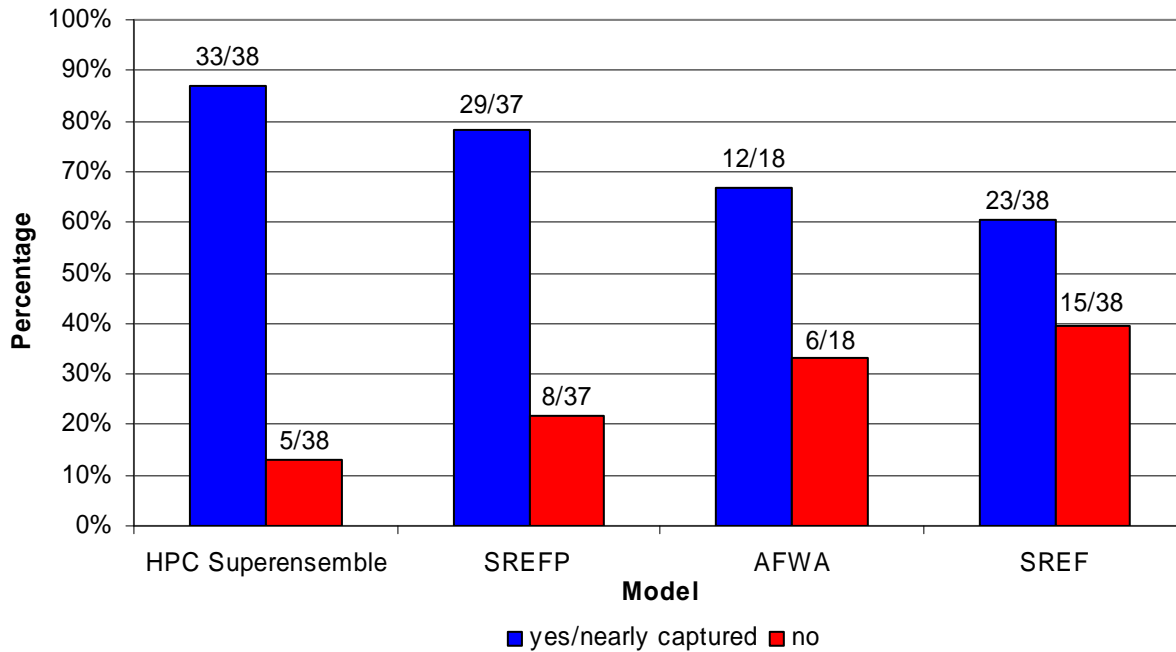


Figure 6. Ability of the experimental ensembles to capture the 2 in/24 hr snowfall events with the model 1% probability contour. Participants were asked to determine whether the observed 2 inch snowfall area fell entirely within the 1% probability contour from the model. “Nearly captured” represents cases in which there were only very small areas of observed 2 inch snowfall outside of the 1% probability contour. The AFWA ensemble was only available for the Day 1 (24 – 48 hr) forecast period.

vertical profiles of temperature and humidity and the vertical velocity. A more complete approach may require considering processes occurring above the surface, similar to the Roebber technique (Roebber et al. 2003).

While increased dispersion likely contributed to the SREFP’s ability to better capture the range of possible solutions than the operational SREF, participants noted several instances in which the SREFP failed to resolve subtle features in the mean pressure and height fields (ex: weak shortwaves). Although ensemble mean features are expected to be less defined than those from individual deterministic solutions, the difference in the depiction between the operational SREF and SREFP was surprising. These concerns are significant because the SREF is used extensively by the HPC’s Winter Weather Desk.

AFWA Ensemble

The AFWA ensemble provided the first opportunity to test the utility of a high resolution convection-allowing ensemble for winter weather forecasts. While the performance was mixed

overall, it was comparable to the performance of the SREFP during the Day 1 period (Fig. 7). Like the SREFP, participants found that the increased resolution provided useful forecast details that allowed them to better target areas expected to receive heavy snowfall. The improvements were especially noticeable in terrain, where the AFWA ensemble was able to confine the heaviest snowfall amounts to the tops of mountain ranges while keeping any significant snow out of the warmer valleys. Additionally, the AFWA ensemble generally performed well in cold air damming situations.

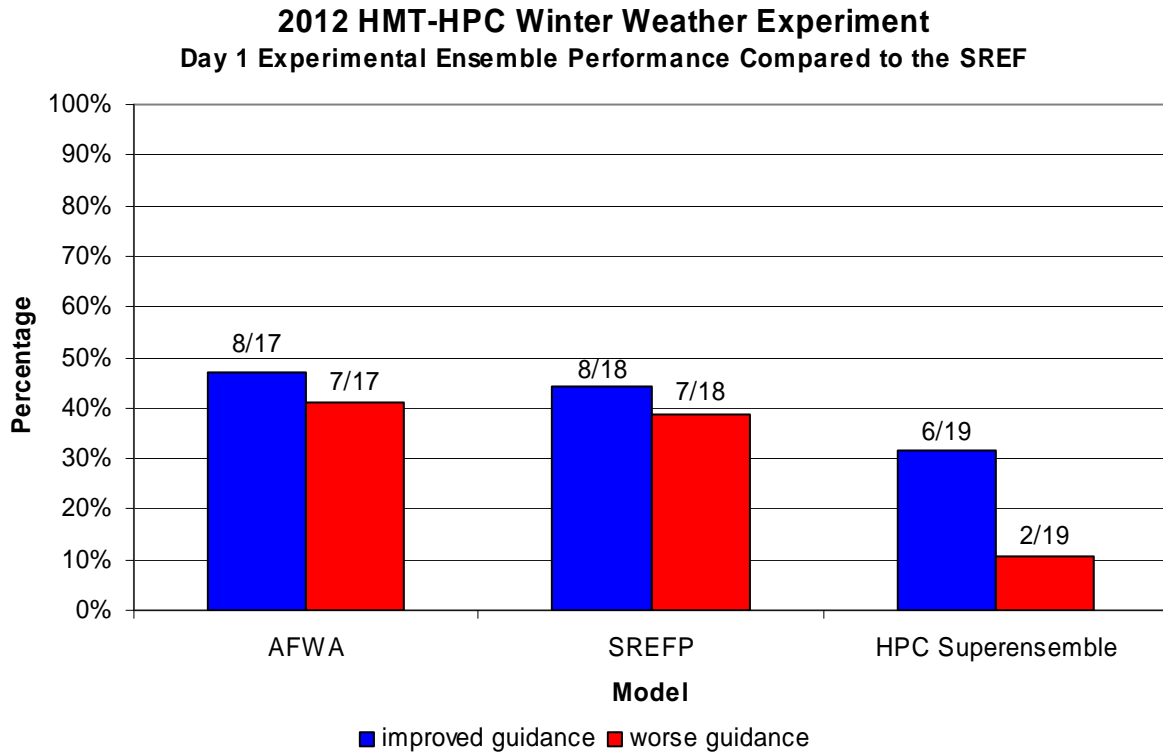


Figure 7. Experimental ensemble performance for the Day 1 period based on participant feedback from subjective evaluations conducted during the 2012 HMT-HPC Winter Weather Experiment. Participants were asked to determine whether the ensemble mean snowfall forecasts from the 00 UTC experimental guidance (21 UTC SREFP) were much better, better, about the same, worse, or much worse than the guidance provided by the operational 21 UTC SREF, based on observations from the gridded HPC snowfall analysis.

While participants found the AFWA ensemble helpful for identifying the location of the heaviest snowfall, they found that the mean snowfall amounts themselves were often too high. Figure 8 shows an example of the high snowfall accumulations that were often observed in the AFWA ensemble mean. While the AFWA ensemble does an excellent job distinguishing between the snowfall minimum in the Hudson Valley and the potential for higher amounts in the mountains of New York, Vermont, and New Hampshire, the observed snowfall amounts were about half of the AFWA forecast values. This tendency for the AFWA ensemble to generate snowfall

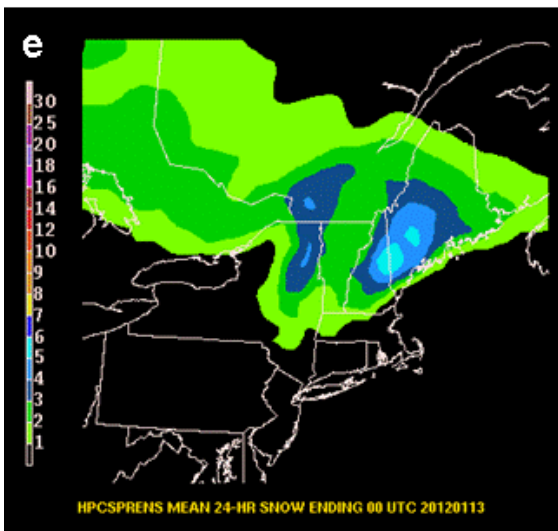
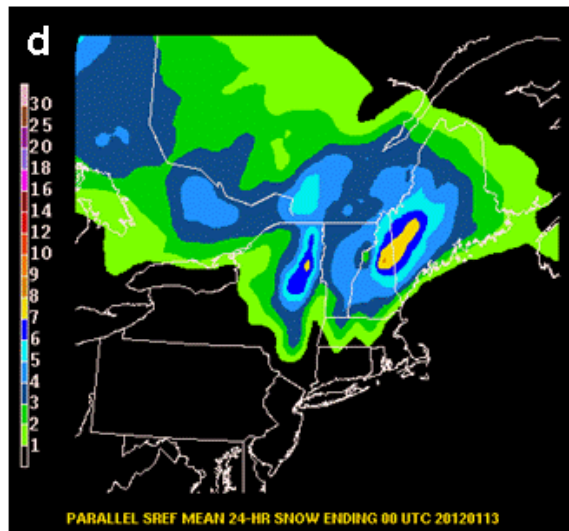
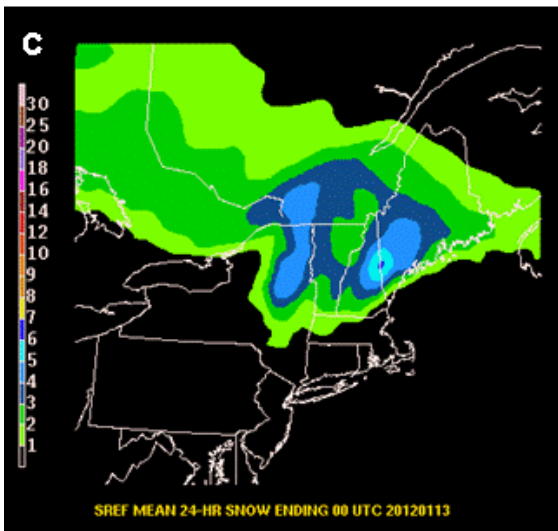
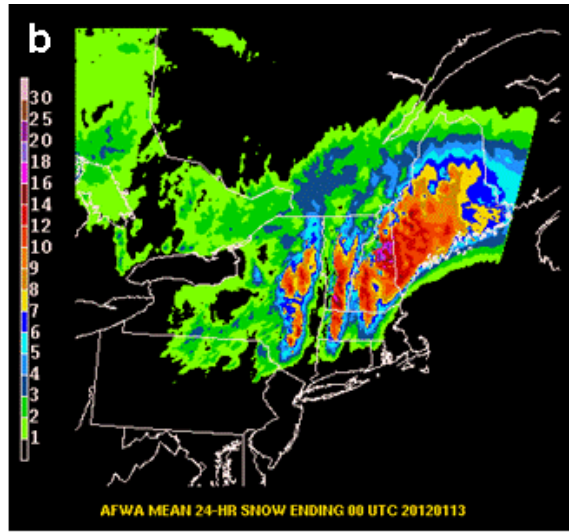
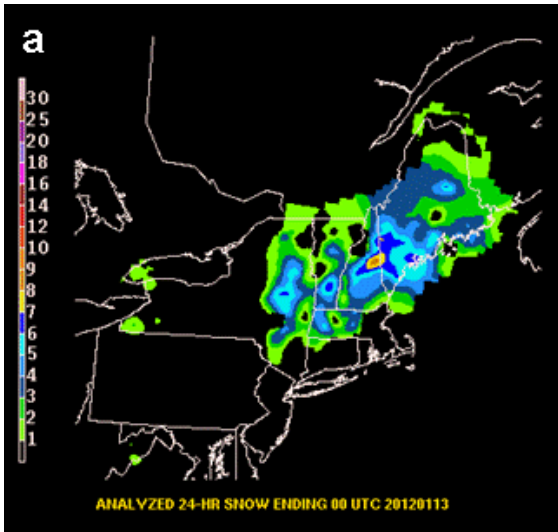


Figure 8. (a) Observed 24 hr snowfall ending 00Z 13 January 2012 from the HPC snowfall analysis and the corresponding Day 1 ensemble mean snowfall forecasts from the (b) AFWA, (c) operational SREF, (d) SREFF, and (e) HPC Superensemble.

amounts that appeared to be unrealistically high accounted for several of the “worse” and “much worse” ratings in the subjective model evaluation surveys.

The high mean snowfall amounts from the AFWA ensemble are likely due to a combination of the snow-to-liquid ratio algorithm and the ensemble mean quantitative precipitation forecast (QPF). Similar to the SREFP, the AFWA ensemble calculates snow-to-liquid ratio based on 2 m temperatures (see Table 1), which may not adequately account for processes occurring above the surface. Additionally, previous work has found that precipitation forecasts from high resolution models and ensembles often have a high bias (e.g. Schwartz et al. 2009). Although QPF was not specifically evaluated during this experiment, Figure 9 shows that in this case, the QPF from the AFWA ensemble was higher than the other ensemble guidance, which likely further contributed to the high mean snowfall amounts.

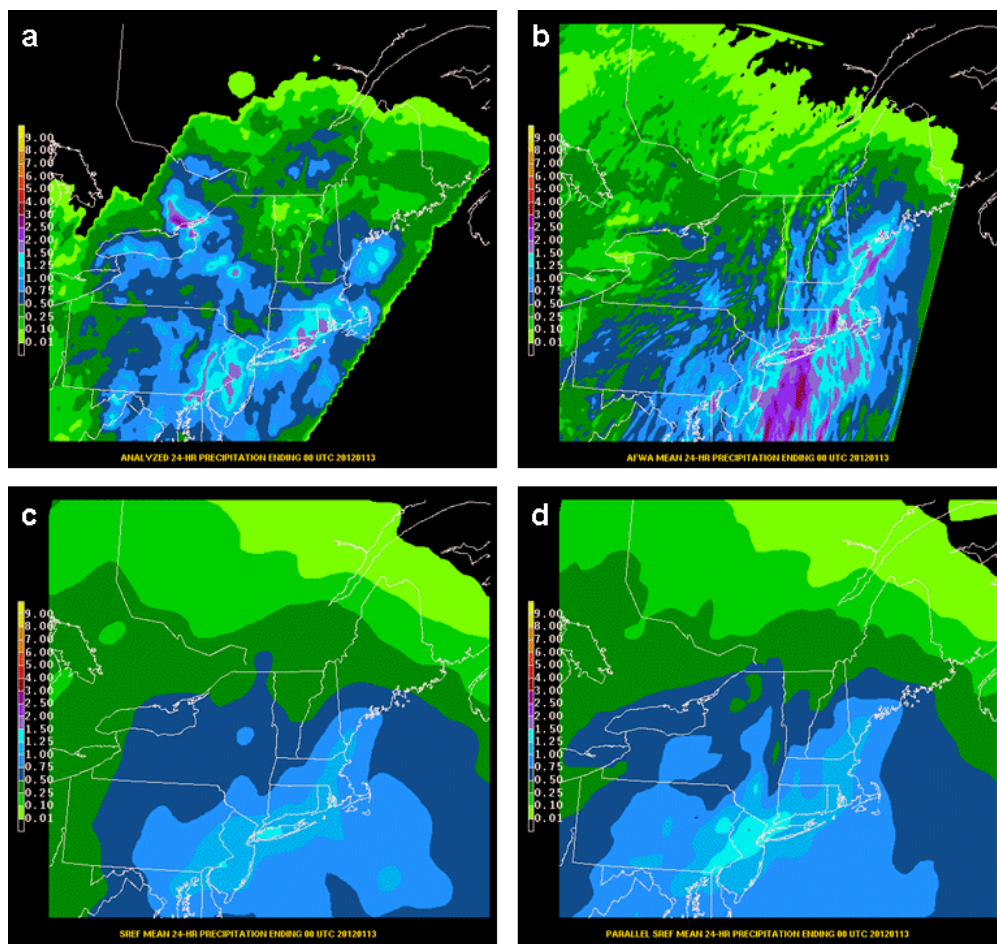


Figure 9. (a) Observed 24 hr precipitation (Stage IV) valid 00 UTC 13 January 2012 and the corresponding Day 1 ensemble mean precipitation forecasts from the (b) AFWA, (c) operational SREF, and (d) SREFP. QPF was not available from the HPC Superensemble.

Probabilistically, the AFWA ensemble was similar to both the operational SREF and the SREFP during the Day 1 period (Fig. 10). Participants found that it was generally able to capture the range of possible solutions, but that the amount of detail resulted in discontinuous probability

areas that are likely unrealistic. Smoothing the probability fields using a neighborhood technique may help produce more physically realistic results, although some of the limitations may simply be related to underdispersion associated with a small 10-member ensemble.

**2012 HMT-HPC Winter Weather Experiment
Ability to Capture 2"/24hr Snowfall Events (Day 1)**

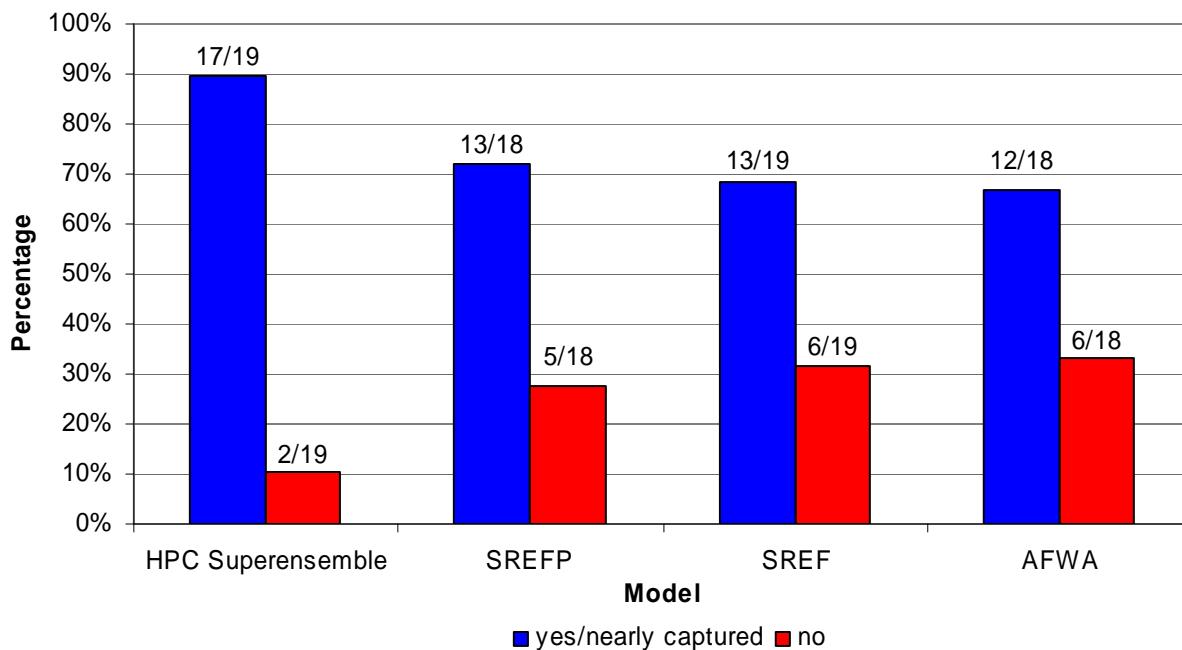


Figure 10. Ability of the experimental ensembles to capture the range of possible forecast solutions during the Day 1 period. Participants were asked to determine whether the observed 2 inch snowfall area fell entirely within the 1% probability contour from the model. “Nearly captured” represents cases in which there were only very small areas of observed 2 inch snowfall outside of the 1% probability contour.

HPC Superensemble

In most cases, the mean snowfall forecasts from the HPC Superensemble were similar to those from the operational SREF (not shown). This result is not surprising since the individual operational SREF members account for 21 of the 28 members of the HPC Superensemble. When noticeable differences did exist, the mean snowfall forecast from the HPC Superensemble generally improved upon the forecast from the operational SREF (Fig. 4). This is encouraging since it indicates that the addition of individual deterministic runs and other ensemble means into the HPC Superensemble results in improved forecast guidance. Probabilistically, the HPC Superensemble provides much better guidance than the operational SREF, rarely missing observed snowfall events. Given the overall positive performance of the SREFP, the HPC Superensemble is expected to improve further when the SREF is upgraded.

5. EXPERIMENTAL FORECAST TOOLS

In addition to the standard deterministic and ensemble model output, a number of derived forecast tools were available for use during the experiment. After using these tools for several days, participants were asked to provide comments about the utility of each forecast tool within an operational forecasting environment.

Ensemble Clusters

In addition to the standard mean and probability fields, the SREFP also offered the opportunity to explore the utility of ensemble clusters. In an effort to gauge whether the ensemble clusters were identifying viable forecast solutions, participants were asked to determine whether the individual ensemble cluster means provided better forecast guidance than the full SREFP ensemble mean for both 500 mb heights and 3 hr QPF. In general, there were rarely major differences between the cluster means and the full ensemble mean, and any improvements were very subtle. This result may not be surprising, however, since the goal of ensemble clustering is simply to identify reasonable alternatives to the ensemble mean. It is important to note that the cluster evaluation performed during the experiment was limited by both the “snapshot” nature of the 500 mb evaluation (00 UTC or 12 UTC only) and the reliance on radar-based QPE data for the evaluation of 3 hr precipitation forecasts. In addition, this winter was characterized by a lack of strong upper level systems, potentially making it more difficult to discern differences between the solutions at 500 mb.

Despite these limitations, most participants liked the concept of ensemble clustering and saw the potential value of the technique, although they were able to identify many areas that need improvement in order for clusters to become an operationally viable forecast tool. As they are currently designed, the ensemble clusters are calculated based on 500 mb height correlations over the entire SREFP domain. In many cases, the feature that the forecast team was investigating was not the feature that was highlighted by the clustering algorithm, making it difficult to apply the cluster information to the forecast. A regional or feature-based approach to clustering may be more useful because it would allow forecasters to focus specifically on the system or area of concern. Additionally, using the “most representative” cluster member instead of the cluster mean may provide a more well-defined solution.

Another hurdle to the operational use of clusters is the time-dependent nature of the cluster calculation. Currently, clusters are calculated independently at each 3 hr time stamp. This makes it impossible to follow the evolution of a particular solution through time and limits the utility of some of the cluster output (ex: QPF is limited to 3 hr increments). Using an approach similar to the ECMWF, which calculates clusters over 2-3 day periods, may help mitigate this issue.

Trend Reverse (“Flip-Flop”) Tool

Participants generally found the trend reverse tool useful, particularly for providing increased awareness of potential model discrepancies. However, many participants found it to be incomplete since significant trends in the models (ex: rapid deepening of a surface low) are not depicted. Since the tool was otherwise useful for identifying potential areas of concern, participants suggested adding information about model trends, since information about both trends and trend reversals are needed to provide a complete depiction of model continuity issues.

Critical Temperature Probabilities

Overall, participants found the critical temperature probabilities useful since they provided an indication of the uncertainty associated with a single deterministic solution. These probabilities are currently displayed as single plan views for each level and temperature threshold. Participants found that it was difficult to construct an overall picture of the temperature profile and the uncertainty associated with it from the current visualization method and suggested combining information from multiple vertical levels on a single display.

6. MODEL DIAGNOSTICS INVESTIGATION

The model diagnostics investigation revealed that while there were generally few notable model initialization errors, run-to-run model continuity could be a significant issue. In both cases, once initialization or continuity errors were identified, it was difficult to determine whether those errors would ultimately have a significant impact over the experimental forecast area. The difficulty in translating these errors into meaningful information for the daily forecast meant that it was hard to use this knowledge to provide better information about forecast confidence.

The run-to-run continuity question was typically answered using information from the trend reverse tool. While there are certainly other aspects of forecast continuity to consider, the responses to this question revealed that the GFS and the NAM struggled the most with maintaining run-to-run continuity (Fig. 11). As expected, ensembles were typically more consistent than individual deterministic runs, although participants noted a number of instances in which the ensemble means exhibited problems with run-to-run continuity as well. Among the ensemble solutions considered, the GEFS mean struggled the most to maintain forecast continuity.

Participants were asked to combine the information they learned during the model diagnostics investigation with their understanding of the forecast situation from the experimental forecast period to rate their overall forecast confidence as below average, average, or above average. Figure 12 summarizes the forecast team’s overall confidence in their forecasts. Forecast confidence tended to be average on Day 1, and more evenly distributed between the three categories on Day 2. In general, forecast confidence tended to stay the same or increase

2012 HMT-HPC Winter Weather Experiment
Run-to-Run Model Continuity

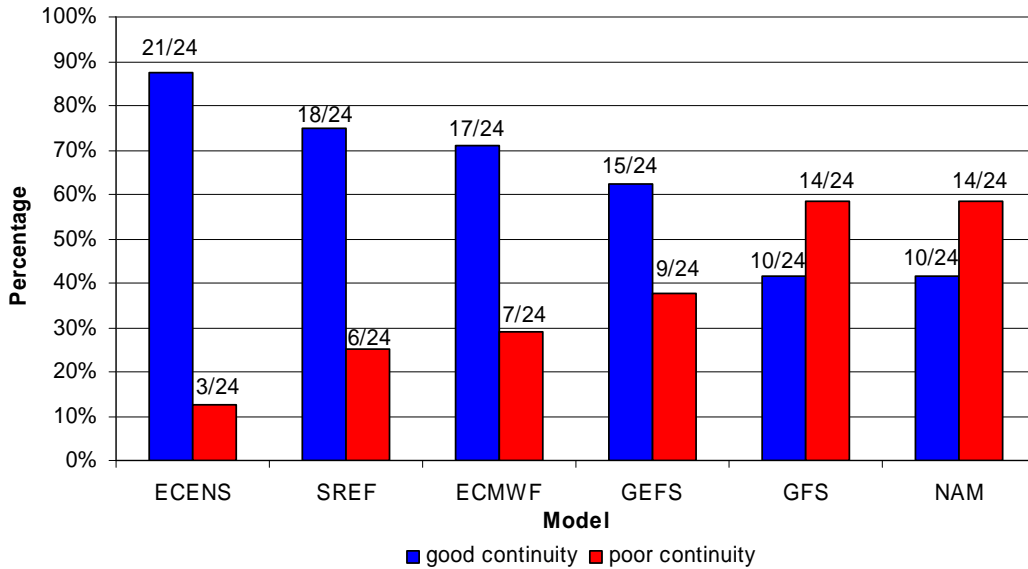


Figure 11. Model run-to-run continuity analysis based on information from the trend reverse tool during the 2012 HMT-HPC Winter Weather Experiment. Participants were asked to determine whether there were model continuity issues that would impact their forecast area of concern. Since the three most recent model cycles are needed to calculate trend reversals, the NAM and the SREF were restricted to 60 hr forecasts and did not cover the entire Day 2 forecast period.

2012 HMT-HPC Winter Weather Experiment
Forecast Confidence

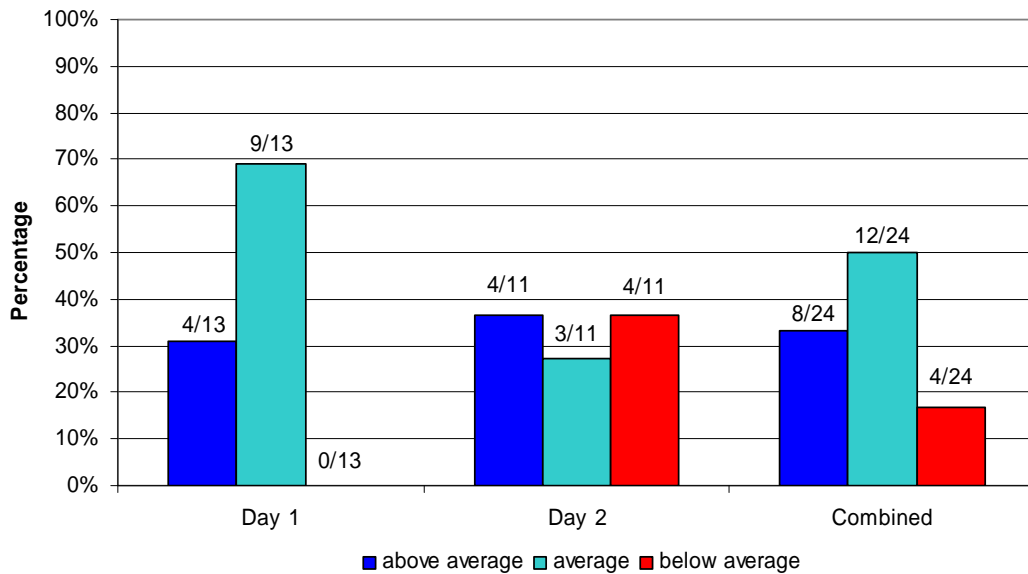


Figure 12. Overall forecast confidence as rated by participants during the 2012 HMT-HPC Winter Weather Experiment.

between Day 2 and Day 1 forecasts for the same event. There was only one instance in which forecast confidence decreased between a Day 2 and a Day 1 forecast.

7. SOCIETAL IMPACTS SCENARIO

The mock briefings conducted during the societal impacts component exposed participants to the challenges of communicating information about complex meteorological situations to non-meteorological decision makers. Participants generally found the concept of the inverted communication pyramid (Somerville and Hassol 2011) useful for structuring their briefings. The main challenge was often how to best portray forecast confidence to the public. Participants sometimes struggled to differentiate between projecting confidence and authority in the delivery of their briefing and conveying their level of forecast confidence.

This portion of the experiment was significantly enhanced by the collaboration that developed with researchers from the University of North Carolina (UNC) Institute for the Environment. Starting during the second week of the experiment, three scientists working to improve communications between NWS forecasters and emergency managers listened to the briefings and provided feedback about their effectiveness. At the end of the experiment, the UNC group summarized their feedback into a document with suggested briefing guidelines for HPC forecasters (Appendix D). The guidelines stressed that the speaker should:

- Focus on the final message (what, when, where, and how much)
- Know the audience and their specific decision thresholds
- Convey confidence in the forecast
- Avoid meteorological jargon
- Explain forecast details only if necessary

HPC continues to work with the UNC group and others to improve communication of complex forecast information.

8. SUMMARY AND OPERATIONAL IMPACTS

The second annual HMT-HPC Winter Weather Experiment was conducted January 9 – February 10, 2012. This year's experiment focused on using ensemble systems to help quantify and communicate uncertainty in winter weather forecasts. During the experiment, participants issued experimental probabilistic forecasts of exceeding 2 in, 4 in, and 8 in of snow over a 24 hr period. In addition to the experimental forecasts, participants subjectively evaluated the experimental ensemble guidance, explored potential model diagnostics issues, rated their overall forecast confidence, and participated in mock societal impacts briefings.

The experiment revealed that while probabilistic winter weather forecasting remains a challenge, higher resolution ensembles can provide important forecast details. A number of the experiment findings are directly relevant to HPC operations and the broader winter weather forecasting community:

- **Overall, the SREFP provided better winter weather forecast guidance than the current operational SREF.** The SREFP provided improved forecast details and probabilistic information. However, a more sophisticated snow-to-liquid ratio technique may be necessary.
- Although the AFWA ensemble tended to have a high bias with snowfall amounts, particularly in topography, it provided useful and realistic forecast details. **HPC is working with SPC, AWC, and NCO to explore the feasibility of providing routine access to the AFWA ensemble.**
- **Determining snow-to-liquid ratio and precipitation type remains a challenge, and forecasters must be aware of the limitations of the current methods.** For example, the NCEP dominant precipitation type method is a “yes/no” decision, which may not be appropriate in marginal environments. HMT-HPC will be working with EMC to explore the utility of using information from the models’ microphysics schemes to explicitly predict snowfall accumulations and precipitation type.
- The SREFP ensemble cluster guidance appears promising, **but several improvements are necessary in order for clusters to become an operationally viable forecast tool.** Participants suggested a regional or feature-based approach to clustering, as well as calculating clusters in a way that allows for temporal continuity.
- Effectively communicating forecast information to non-meteorological decision makers takes practice. Instead of starting with background information, **invert the scientific communication pyramid and provide the bottom line first and supporting details last.** UNC provided a useful summary of suggested briefing guidelines.

The HMT-HPC Winter Weather Experiment provided a unique opportunity to bring forecasters, researchers, and model developers together to explore the challenges of probabilistic winter weather forecasting. The experiment identified important issues that will continue to be explored and raised awareness of the challenges associated with both forecasting and verifying winter weather.

ACKNOWLEDGEMENTS

The HMT-HPC Winter Weather Experiment would not be possible without the dedication of a host of individuals including Faye Barthold, Mike Bodner, Dave Novak, Tom Workoff, Dan Petersen, Rich Otto, Mike Musher, Rich Bann, and Andrew Orrison. Jun Du (EMC) was instrumental in providing consistent access to the SREFP in its pre-implementation phase, while Scott Rentschler filled the same role at AFWA. Ken Galluppi (UNC), Jessica Losego (UNC), and Burrell Montz (East Carolina University) provided valuable insight during the societal impacts component.

REFERENCES

- Alhamed, A., S. Lakshmiarahan, D.J. Stensrud, 2002: Cluster analysis of multimodel ensemble data from SAMEX. *Mon. Wea. Rev.*, **130**, 226-256.
- 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.
- Roebber, P. J., S. L. Bruening, D. M. Schultz, and J. V. Cortinas, 2003: Improving snowfall forecasting by diagnosing snow density. *Wea. Forecasting*, **18**, 264-287.
- Schwartz, C. S., and Coauthors, 2009: Nest-day convection-allowing WRF model guidance: A second look at 2-km versus 4-km grid spacing. *Mon. Wea. Rev.*, **137**, 3351-3372.
- Somerville, R. C. J., and S. J. Hassol, 2011: Communicating the science of climate change. *Phys. Today*, **64**(10), 48-53.
- Yussouf, N., D.J. Stensrud, S. Lakshmiarahan, 2004: Cluster analysis of multimodel ensemble data over New England. *Mon. Wea. Rev.*, **132**, 2452-2462.
- Zsoter, E., R. Buizza, and D. Richardson, 2009: "Jumpiness" of the ECMWF and Met Office EPS control and ensemble-mean forecasts. *Mon. Wea. Rev.*, **137**, 3823-3836.

APPENDIX A
Participants

Week	HPC Forecaster	WFO	Research/Academia/ Private Sector	EMC
Jan 9 – 13	Dan Petersen	Chris Smallcomb (WR) Jared Guyer (SPC)	Brian Colle (SBU) Matt Souders (SBU)	
Jan 17 – 20	Rich Otto	Jeff Tongue (OKX)		Jun Du
Jan 23 – 27	Mike Musher	Rob Cox (CYS)	Pete Manousos (First Energy) Brian Cerruti (MDL)	
Jan 30 – Feb 3	Rich Bann	Walt Drag (PHI)	Ellen Sukovich (ESRL)	Brad Ferrier
Feb 6 – 10	Andrew Orrison	Brian Lasorsa (LWX) Michael Scotten (AMA)	Phil Shafer (MDL)	DaNa Carlis

APPENDIX B

Experimental Forecast Tools

Ensemble Clusters

The ensemble clustering technique identifies solutions within the full ensemble mean that are similar to one another based on the spatial correlation between the 500 mb height fields in the individual SREF members every 3 hours. A cluster is identified if the correlation is at least 0.93, and the SREFP can generate a maximum of 6 clusters. Clusters 1 and 2 represent extreme solutions, while Cluster 3 represents the middle of the distribution. Clusters 4-6 are formed from any members that remain after Clusters 1-3 have been determined.

Models:

- SREFP

Available forecast fields:

- 500 mb heights
- Mean sea level pressure
- 1000-500 mb thickness
- 3hr precipitation

Critical Temperature Probabilities

Critical temperature probabilities indicate the probability of the temperature being below critical freezing and precipitation type temperatures. These values are derived by estimating a cumulative distribution function (CDF) using the deterministic model forecast as the mean and the spread from the corresponding ensemble system as the variance in this equation:

$$F(x) = \Pr\{T \leq x\} = \int_{-\infty}^{T \leq x} f(x) dx,$$

where T is the ensemble mean temperature field and x is the critical temperature value. These probabilities are then plotted along with the deterministic model forecast to indicate the spatial uncertainty in the thermal field.

Models:

- ECMWF
- GFS
- NAM (with operational SREF spread)

Available forecast fields:

- 700 mb $T < 0^{\circ}\text{C}$ and $T < -7^{\circ}\text{C}$
- 850 mb $T < 0^{\circ}\text{C}$
- 700-850 mb layer $T < 0^{\circ}\text{C}$
- 2 m $T < 32^{\circ}\text{F}$ (GFS and NAM only)

Trend Reverse (“Flip-Flop”) Tool

The trend reverse tool identifies changes between the current model initialization, the cycle 12 hours prior, and the cycle 24 hours prior. Trend reversals are calculated at each grid point based on the following equation:

$$F = \sqrt{(X_{t-12} - X_{t-24})(X_{t-12} - X_{t_0})},$$

where X is the forecast field of concern (ex: 500 mb heights) and the subscripts indicate the difference in forecast run time from that of the most recent forecast. Positive trend reversals in the models over the previous 3 forecast cycles (i.e. “flip-flops”) indicate the potential for forecast uncertainty, with greater values indicating greater uncertainty. Consistent trends result in imaginary values of F , which indicate no flip-flop.

Models:

- GFS
- GEFS
- ECMWF
- ECENS
- NAM
- SREF

Available forecast fields:

- 500 mb heights
- 850 mb heights
- Mean sea level pressure

APPENDIX C

Daily Schedule

A brief orientation session will be held at 8:15am on the first day of each week to explain the motivation and organization of the experiment as well as some of the data being evaluated.

- | | |
|--------------------------|---|
| 8:30am – 9:00am | Discuss yesterday's forecast; determine today's forecast time period (Day 1 or Day 2) |
| 9:00am – 11:00am | Draw contours for the probability (slight-10%, moderate-40%, high-70%) of exceeding 2 in, 4 in, and 8 in of snow (4 in, 8 in, and 12 in for high-end events) for one 24 hr period (00 – 00 UTC); forecasts are based on 00 UTC guidance |
| 11:00am – 11:30am | HPC-CPC map discussion |
| 11:30am – 12:30pm | Lunch |
| 12:30pm – 1:30pm | Model diagnostics investigation and forecast confidence discussion |
| 1:30pm – 2:30pm | Societal impacts scenario |
| 2:30pm – 4:00pm | Subjective verification using the gridded HPC snowfall analysis to evaluate experimental guidance performance as it relates to snowfall accumulation and the indication of forecast uncertainty |
| 4:00pm – 4:30pm | Group discussion |

APPENDIX D
Guidance for HPC Briefings
WxEM Team: Ken Galluppi, Jessica Losego, Burrell Montz

1. **Make your “bottom line” the top line.** Partners want to know what your final message is: what is your best guess of what the hazard is, when it will happen, where, and how much. What you would think of putting at the end of your briefing to summarize is what you should be putting first. Do not discuss “why” (see item #4 for more information).
2. **Make sure you know who your audience is.** Avoid the use of nebulous, subjective terms unless you really know your audience. What you may consider to be low impact event could be a high impact event for the partner you’re briefing. For example, you may say that a winter event will be “low impact” because 1 inch of snow will fall. However, this may be a high impact event to a school transportation director who is taking into consideration a 16-year old driver who may have never driven in snow.

If you know your audience you can localize your information to help the partner relate to what you’re telling him. This helps the partner manage risk. For example, a DOT official may want to know the amount of rain that is forecast to fall before a winter event because he needs to know if a costly road pre-treatment will wash away and be ineffective. Knowing your audience concerns will help you address the information they need.

3. **Convey confidence in your forecast.** It is important to sound confident in your forecast by not saying “it could do this or it could do that”. Give your best guess first (see #1) because a partner wants to be on the same page with you before you get into your explanation. A partner would rather hear your best guess interpretation then doing their own interpretation and guess.

Don’t use language such as, "I'm not sure, but this is what I think" because it places doubt over everything you just said. It is better to say "Right now, given what I know, this is my best answer," or "I don't know, but will get back to you."

4. **Partners are not as interested in “why”.** You should explain the weather only if the partner needs to 1) know what may cause change (what to watch for) until they are briefed again; 2) build confidence. For example, you could tell a partner to expect x amount of snow, but if it moves faster less snow will fall.

Do not use jargon such as "the storm will deepen" or “the snow to liquid ratio is x”. Instead you should be practical and say "the winds will get stronger" or “the snow will be very wet and heavy” because it relates to what the partner cares about.