4-7-11

2011 HMT-HPC Winter Weather Experiment

1. Introduction

The inaugural Hydrometeorological Testbed (HMT)-Hydrometeorological Prediction Center (HPC) Winter Weather Experiment was held January 10 – February 11, 2011 within the HMT area of HPC. Through subjective evaluation and real-time forecasting activities, the experiment aimed to answer the following questions:

- Can high-resolution models improve Day 1 forecasts of precipitation type and amount?
- Can high-resolution models assist forecasters in anticipating the placement and intensity of mesoscale snowbands for Day 1 forecasts?
- Can we better quantify and communicate winter weather uncertainty for Day 1-2 forecasts?

The experiment fostered unprecedented winter weather collaboration among 14 participants, from NCEP centers, WFOs, and NOAA research labs (Appendix A). This report summarizes the activities and findings of the experiment.

2. Experiment design

The experiment was active generally Monday-Friday from 8:30 am – 4:30 pm (Fig. 1). Given forecasts were made for the Day 1 (00-00UTC) and Day 2 (00-00 UTC) forecast periods, the Day 1 forecast made on Monday (the first day of the week) ended 00 UTC Wednesday and the Day 2 forecast ended 00 UTC Thursday. The Day 1 forecast made on Friday (the last day of the week) ended 00 UTC Sunday and the Day 2 forecast ended 00 UTC Monday. Thus, the only day of the week which was not covered by at least one forecast was the 24h period ending 00 UTC Tuesday.

Appendix B shows the daily schedule. During the morning participants were asked to subjectively rate experimental human and model forecasts (very poor, poor, fair, good, very good) via a web interface. The available experimental guidance is described in Appendix C. Verification focused on the ability of each experimental dataset to accurately predict precipitation type transitions, precipitation amount, mesoscale banding, and forecast uncertainty as applicable. The results of these ratings are shown in section 3.

During the afternoon participants used both operational and experimental guidance (Appendix C) to create experimental forecasts for snow/sleet accumulations and ice accumulations for storms of interest during the Day 1 (00–00 UTC) and Day 2 (00–00 UTC) forecast periods. The Day 1 forecast focused on the ability of the high-resolution models to improve the precipitation type and amount forecasts, and depict the location and intensity of mesoscale snow bands. Fields from the High Resolution Window runs (HRW-ARW and HRW-NMM) and the 4 km nest from an experimental version of the NAM (expNAM) were a focus on Day 1. The Day 2 forecast focused on uncertainty issues related to precipitation type and storm track, and relied on mean, spread, and probabilistic information from the NCEP SREF.

When selecting the forecast area, priority was given first to storms with a threat of significant freezing rain, then to storms with a potential for any type of precip type transition, and then to storms with the potential for heavy snow. In most cases, storms originally investigated during the Day 2 time period were investigated the following day in the Day 1 time period.

In addition to the accumulation forecasts, a forecast confidence discussion was prepared. This discussion detailed the forecaster's level of confidence (above average, average, or below average) in storm track, precipitation type, and mesoscale banding potential for the storm of interest as well as the reason for their level of confidence. Key questions are show in Appendix D.

HMT-HPC Winter Weather Experiment Days						
Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
January 9	10	11 D1	12 D1 D2	13 D1 D2	14 D1 D2	15 D1 D2
16 D2	17 HOLIDAY	18	19 D1	20 D1 D2	21 D1 D2	22 D1 D2
23 D2	24 AMS	25 AMS	26 AMS	27 AMS	28 AMS	29
30	31 ADMIN	February 1	2 D1	3 D1 D2	4 D1 D2	5 D1 D2
6 D2	7	8 D1	9 D1 D2	10 D1 D2	11 D1 D2	12 D1 D2
13 D2	14	15	16	17	18	19

Fig. 1. Calendar showing days which the experiment participants were on site (shaded). Day 1 (D1) and Day 2 (D2) experimental forecast valid periods are also denoted in gray text [shown at midpoint of valid period (12 UTC)].

2. Cases

The experiment time period was very active across the United States. Time-mean troughs were present just off the Pacific northwest coast and in the eastern U.S (Fig. 2a). Split flow was present across the intermountain west. The southern stream was active and contributed several storms over the central and eastern U.S.

In fact, three Northeast Storms Impact Scale (NESIS) (Kocin and Uccellini 2004) category 3 storms and one NESIS category 1 storm were verified by the experiment participants (Fig. 3). Although the pattern was more active in the east, several systems were verified in the Pacific northwest and high-plains (Table 1). The hemispheric flow was less predictable than usual during the experiment time period, as evidenced by 500 hPa height anomaly correlations (5 day forecast) for leading global models, which exhibit a "dip" during late January and early February (Fig. 2b).

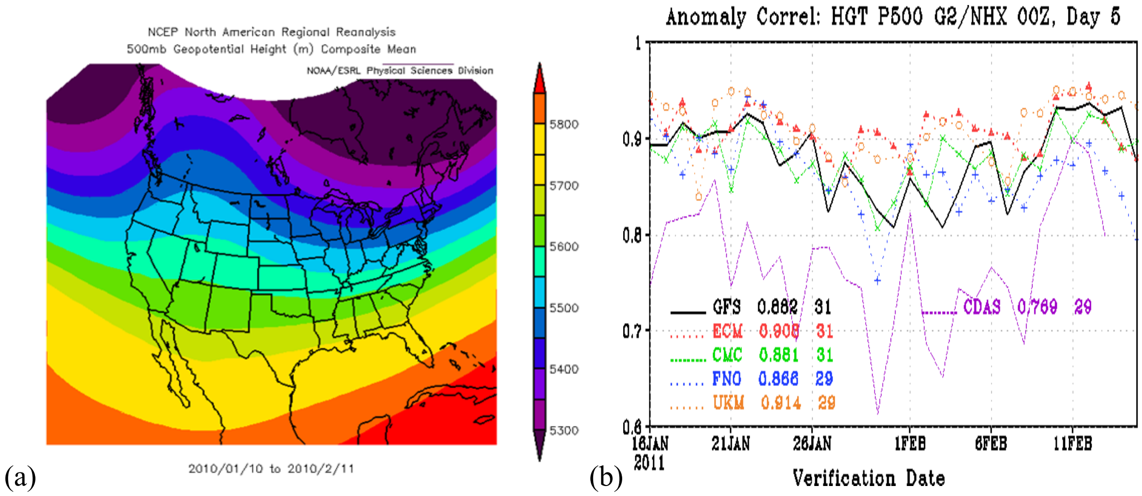


Fig. 2 (a) Composite 500 hPa height during the Jan 10 – Feb 11, 2011 period. (b) Northern Hemisphere 500-hPa anomaly correlation for the GFS (black), ECMWF (red), Canadian (green), FNMOC (Blue), and UKMET (orange) for the Day 5 forecast during the January 16 – Feb 15 period.

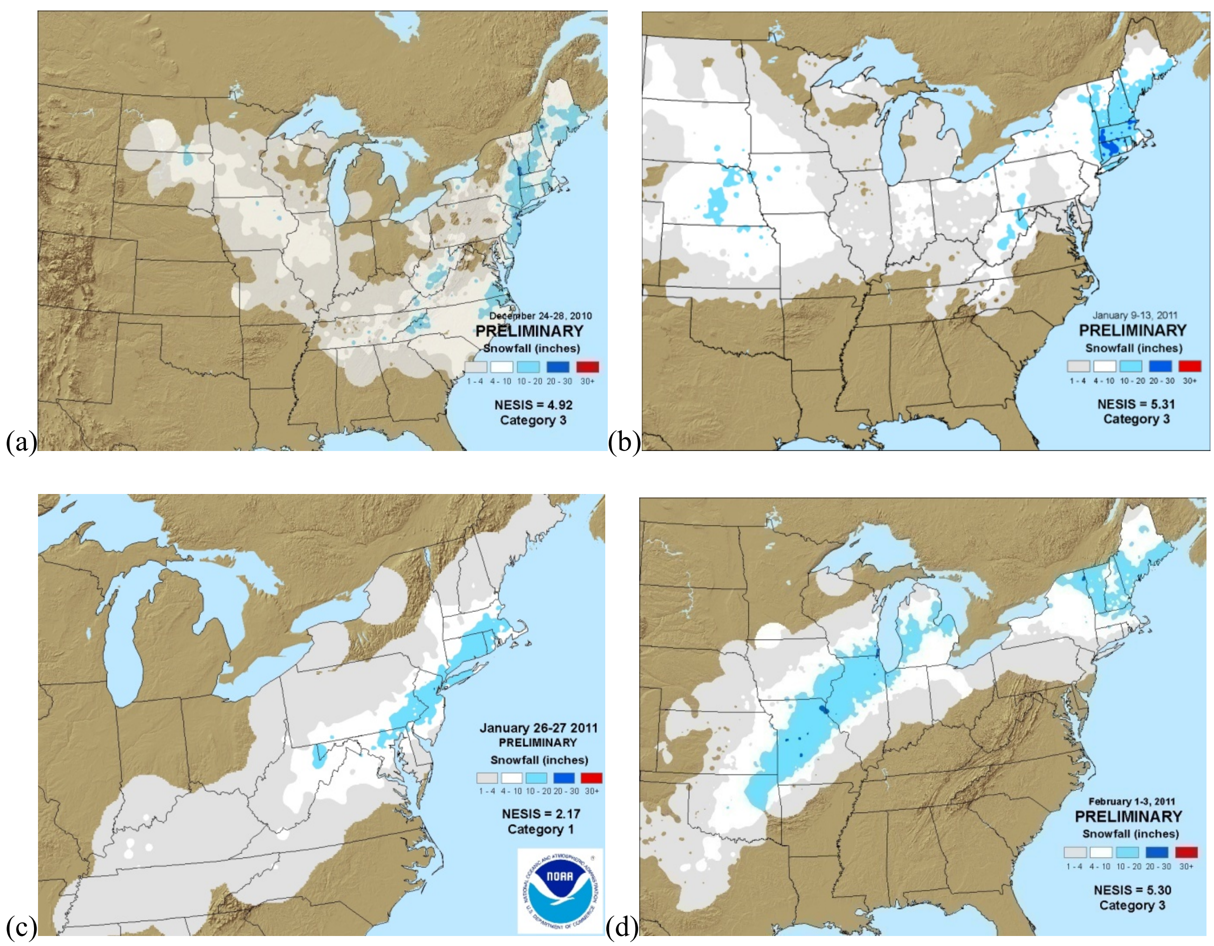


Fig. 3. Snowfall maps from four major storms verified by experiment participants, including (a) Dec 24-28, 2010, (b) January 9-13, 2011, (c) January 26-27, 2011, and (d) February 1-3, 2011. (source: NCDC)

Table 1. Cases forecast and verified by experiment participants.

Forecast Valid Time	D1 and/or D2 Forecast	Verification	Forecast Area	Notes
00Z 28 Dec 2010		x	Mid-Atlantic/Northeast	NESIS category 3 storm
00Z 8 Jan 2011		x	Mid-Atlantic/Northeast	
00Z 10 Jan 2011		x	Southern Plains to Southeast	
00Z 11 Jan 2011		x	Lower MS Valley to Southeast	Snow and freezing rain across the south
00Z 12 Jan 2011	x	x	Middle MS Valley to Mid-Atlantic	
00Z 13 Jan 2011	x	x	Lower Great Lakes and Northeast	NESIS category 3 storm
00Z 14 Jan 2011	x	x	Pacific NW to Northern Rockies	
00Z 15 Jan 2011	x	x	Northern Plains to Upper Midwest	
00Z 16 Jan 2011	x	x	Great Lakes	
00Z 17 Jan 2011	x	x	Northern Plains	
00Z 19 Jan 2011		x	Mid-Atlantic/Northeast	
00Z 20 Jan 2011	x	x	Northeast	
00Z 21 Jan 2011	x	x	Southern Plains to OH Valley	
00Z 22 Jan 2011	x	x	TN Valley to Northeast	
00Z 23 Jan 2011	x	x	Northern Great Basin to Central Plains	
00Z 24 Jan 2011	x	x	Central Plains to Middle MS Valley	
00Z 28 Jan 2011		x	Mid-Atlantic/Northeast	NESIS category 1 storm
00Z 2 Feb 2011		x	Southern Plains to Upper Great Lakes	
00Z 3 Feb 2011	x	x	Middle MS Valley to Northeast	NESIS category 3 storm, freezing rain from OH Valley to Mid-Atlantic
00Z 4 Feb 2011	x	x	Lower MS Valley/ Southeast	Freezing rain across Lower MS Valley
00Z 5 Feb 2011	x	x	Southern Plains to TN Valley	Snow in Dallas, freezing rain across Lower MS Valley
00Z 6 Feb 2011	x	x	Ohio Valley/Middle Atlantic	
00Z 7 Feb 2011	x	x	Mid-Atlantic/Northeast	
00Z 9 Feb 2011	x	x	Northern/Central Rockies to Central Plains	
00Z 10 Feb 2011	x	x	Southern Plains and Lower MS Valley	Substantial snowfall across northern OK
00Z 11 Feb 2011	x	x	Lower MS Valley to OH Valley and Mid-Atlantic	
00Z 12 Feb 2011	x		Lake Ontario	
00Z 13 Feb 2011	x		Pacific NW to Northern Rockies	
00Z 14 Feb 2011	x		Pacific NW to Northern Rockies	

3. Core Results

Participants were asked to subjectively rate human and model forecasts of precipitation type, snowfall amount, and ice amount for the Day 1 and Day 2 forecast. Participants also completed an exit survey. The results are organized according to the three core experiment questions.

Question: Can high-resolution models improve Day 1 forecasts of precipitation type and amount?

Overall 5 of the 14 participants (35.7%) thought that the high resolution guidance provided unique and valuable information for winter weather forecasts (Fig. 4). The majority [8 of the 14 participants (57.2%)] could neither agree nor disagree with this statement. One participant (7.1%) did not agree that the high resolution guidance provided unique and valuable information for winter weather forecasts.

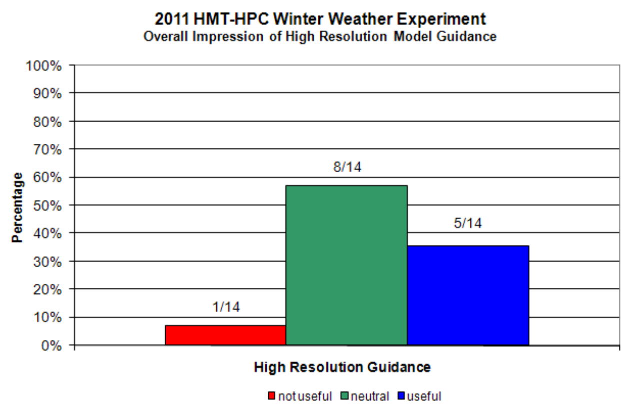


Fig. 4. Participants' overall impression of the experimental high-resolution guidance.

The mixed assessment of the high-resolution guidance was further evidenced by the model evaluations (Fig. 5). In general for precip type, snow, and ice amounts, the expNAM was comparable to the operational NAM for the Day 1 forecast, while the high-resolution windows were not as skillful as the operational NAM.

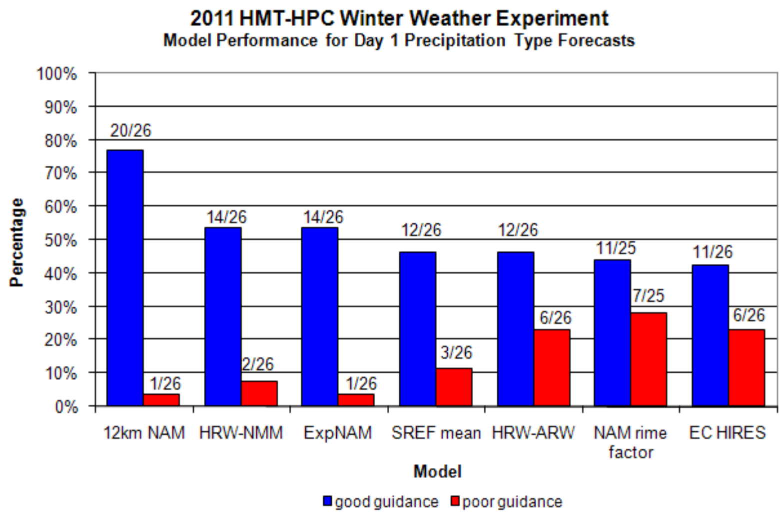
Participant discussions and surveys revealed the following pros and cons of the available high resolution winter weather guidance:

Pros:

- Topographic features (orographic precipitation, lake effect snow). For these features, detail can be added to the deterministic forecast.
- Visualizing temporal evolutions (ptype transitions, band evolution)
- Providing unique fields (simulated reflectivity).
- Depicting tight gradients (as observed). For example the back edge of the comma-head.

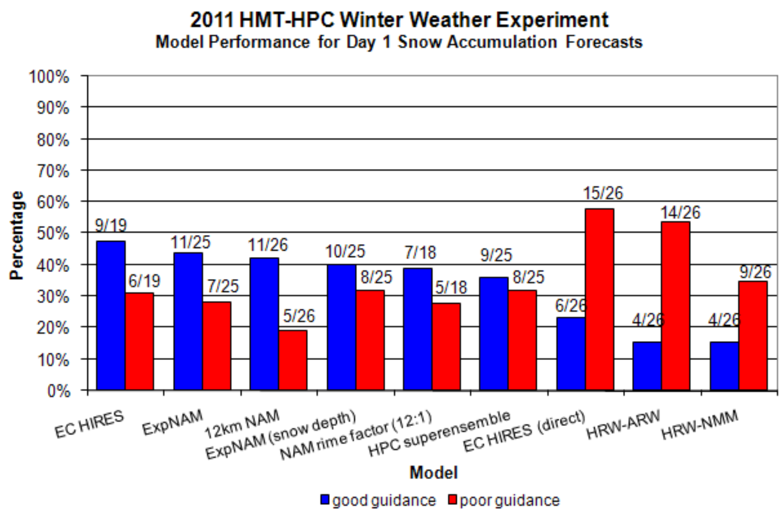
Cons:

- The amounts were not superior to the operational (e.g., Fig. 5)
- Without past experience, the runs can actually increase the uncertainty when taken in the context of conflicting, but trusted operational guidance.



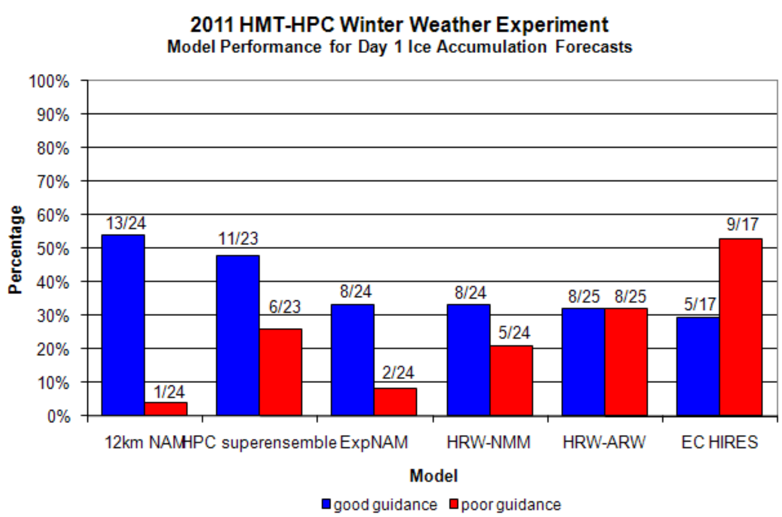
(a) Day 1 Precip Type

The current NAM consistently had the highest rated forecasts, with the HRW-NMM, and expNAM also doing well.



(b) Day 1 Snow Accumulations

The EC HIRES, expNAM, and current NAM had the highest rated forecasts. The High Res Windows received some of the lowest ratings.



(c) Day 1 Ice Accumulations

The current NAM and HPC superensemble had the highest rated forecasts.

Fig. 5 Summary ratings for (a) Day 1 precip type, (b) Day 1 snow accumulations, and (c) Day 1 ice accumulations. Blue bar is sum of "very good" and "good" ratings. Red bar is sum of "poor" and "very poor" ratings.

In summary, the expNAM 4 km nest had comparable skill to the operational NAM for the Day 1 precip type, snow, and ice amount forecasts, and appears ready for operational winter weather use. However the high-resolution windows were not as skillful as the operational NAM, and require additional development. The degree of improvement from high-resolution guidance during the winter appears less than that experienced during the warm season Spring Experiment (Barthold et al. 2011), and may reflect that the WWE was focused on derived fields of snow/ice from near-operational guidance (implemented within 1 year), while the Spring Experiment looked at QPF from the next generation guidance suite (implemented beyond 5 years).

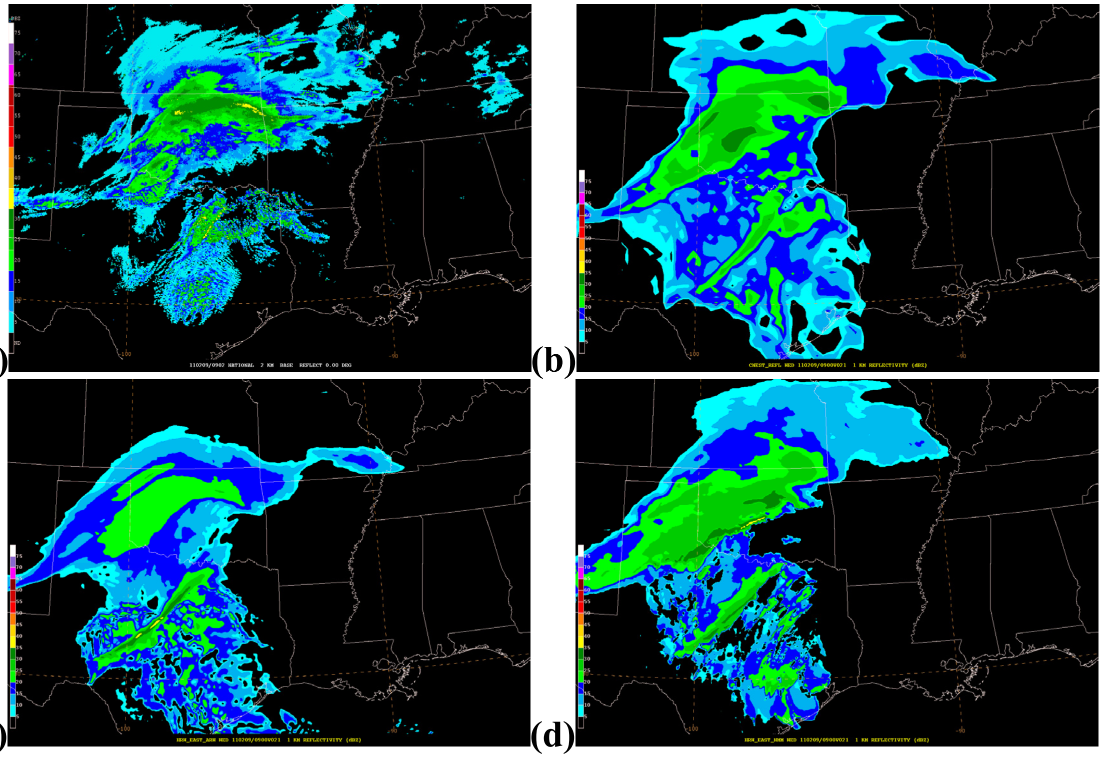
Question: Can high-resolution models assist forecasters in anticipating the placement and intensity of mesoscale snowbands for Day 1 forecasts?

A full assessment of this question was limited by participant's uncertainty in what defines a "band" and data loss of some of the simulated reflectivity products. However from the available data and team experience, predictions of mesoscale banding were challenging. Models did well for Jan 12th event, but were challenged for the Feb 2nd event, and completely missed the Feb 8th and Feb 9th bands. Although not a formal experimental dataset, the High Resolution Rapid Refresh did not forecast the observed bands until 1-2 hours before formation in 2 of these events. Teams could only add the band details in their deterministic product when there was model consensus.

An example of a notable mesoscale band in northeast OK during the Feb 9th event is shown in Fig. 6. Total snowfall accumulations exceeded 20" within the band, with much of the snow falling within 12 h. At 09 UTC 9 Feb, none of the high-resolution guidance provided a clear signal for an intense band. Also all runs were slow (too far west) with the precipitation shield. In general the solutions from the experimental guidance looked more similar to each other than the observations. Examination of the experimental High-Resolution Rapid Refresh (which was not available to the participants) showed that key aspects of the band were only evident ~2 h prior to formation, highlighting the limited predictability of this case.

Although there was mixed success with mesoscale band prediction, some participants noted that the availability of hourly simulated reflectivity (from the high-res window runs) was useful for visualizing the precipitation evolution an event. Many participants noted the generally lower values of simulated reflectivity from the HRW-ARW, as illustrated in Fig. 6c.

In summary, although hourly 4 km data can allow a forecaster to visualize explicit predictions mesoscale bands, the high-resolution guidance was only found useful in anticipating the placement and intensity of mesoscale snowbands when there was consensus among the guidance. In these high predictability cases, the added detail in the human forecast improved the forecast.



(a) (b) (c) (d)

Fig. 6. Comparison of mosaic reflectivity valid 09 UTC 9 Feb 2011 from (a) observed base reflectivity mosaic, (b) experimental NAM (1 km AGL), (c) HRW-ARW (1 km AGL), and (d) HRW-NMM (1 km AGL).

Question: Can we better quantify and communicate winter weather uncertainty for Day 1-2 forecasts?

Perhaps surprisingly, short range forecast busts occurred relatively frequently during the experiment. These included Cape Hatteras snowfall (Jan 22), Dallas snowfall (Feb 4), and the northeast OK heavy snowband (Feb 9) (Fig. 6). These events highlighted the predictability challenges winter weather presents, given sensitivity to slight QPF and temperature differences.

Participants were asked to write confidence discussions to help focus thought on predictability (see Appendix D). An example discussion for the 24 h period ending 00 UTC 3 Feb 2011 is shown below. During this time blizzard conditions were observed in the Chicago metro (see also Fig. 3d), stranding cars on Lake Shore Drive. Approximately 17" of snow was accumulated during the 24 h period. The exercise of writing these discussed helped forecasters focus on the predictability of each scenario, and highlighted confidence information that could be integrated into the current HPC Heavy Snow and Icing Discussion.

Within the discussion, participants were asked to rate their confidence in their snow and ice forecasts relative to the expected confidence for a given lead time. Participants tended to favor "average" confidence for snow accumulation forecasts, while confidence was evenly distributed for ice accumulation forecasts (Fig. 7).

Day 1 (valid 00Z 2 Feb – 00Z 3 Feb)

Storm track: Fairly confident of storm track, though high resolution models seem to deepen low center more and pull it further north and west into the cold air. P-type: Influenced by storm track and extent to which warm air can move northward. Fairly certain within say 100-150 km where the rain/ice/snow line will set up. Maximum precipitation axis is fairly consistent, though shifting north-south depending on storm track differences, with higher res models further north and west.

Meso banding potential: Exists because of negative EPV values, conditionally unstable potential temperature profiles and frontogenetic forcing forecast in areas of high RH, in the area of interest (Kirksville MO through northern IL) to predicted low position.

Confidence: above average confidence that an area near Chicago will receive up to 20" of snow. Also above average confidence that south-central IL to central PA will see glaze to a thickness of 0.25-0.5". Confident that New York to New England mountain areas will see 1'-2' of snow, including Adirondack, Catskills, Green and White Mts., and perhaps Berkshires. Uncertainty that exists results from differences in storm track among models, particularly high res models which intensify storm and pull it farther west and north. ExpNAM Snow Depth across areas in northern OH and central PA showed much greater snow accumulation than other guidance. We think it might be the result of freezing rain falling onto snow, increasing the water equivalent, but not changing the mean density of the snow. The resulting increase in water equivalent would result in increased diagnosed snow depth, without snow actually accumulating.

Snow probs: Chicago IL 12": 80%, 20": 20%   Albany NY 6": 70%, 12": 30%
Ice probs: Harrisburg PA: 0.1": 80%, 0.5": 20%   Albany NY 0.01": 70%

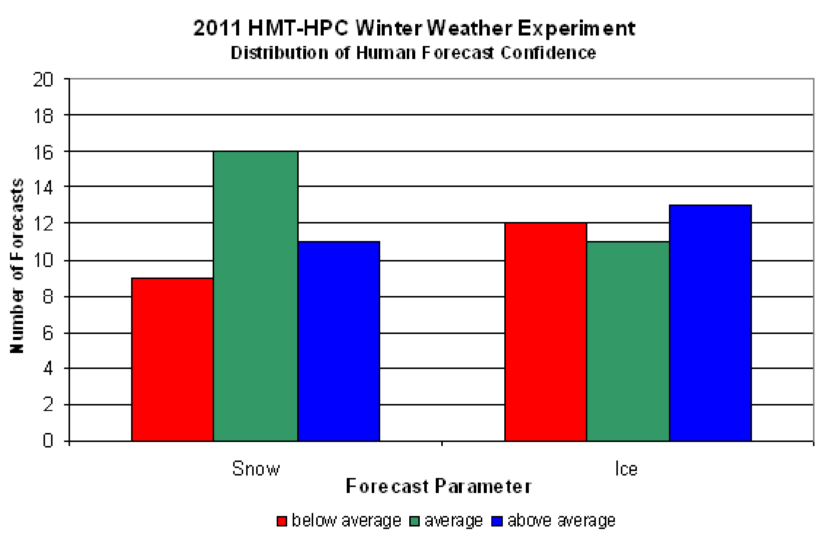


Fig. 7. Distribution of participants' forecast confidence for snow and ice.

As part of the verification process, the experimental forecasts were evaluated. The results of this evaluation show that when forecast teams rated their snow and forecasts as low (high) confidence, there were generally larger (smaller) errors (Fig. 8a). Thus, the forecast team's confidence was qualitatively correlated to the errors for snow accumulations. This confidence-skill relationship is not as evident for the ice accumulation forecasts (Fig. 8b). Thus, forecasters may have more skill at anticipating errors for snow accumulations than for ice accumulations, although more cases are needed. Also, the quality of the ice analysis was questionable, and may have affected results. This issue will be discussed in more detail in Section 5.

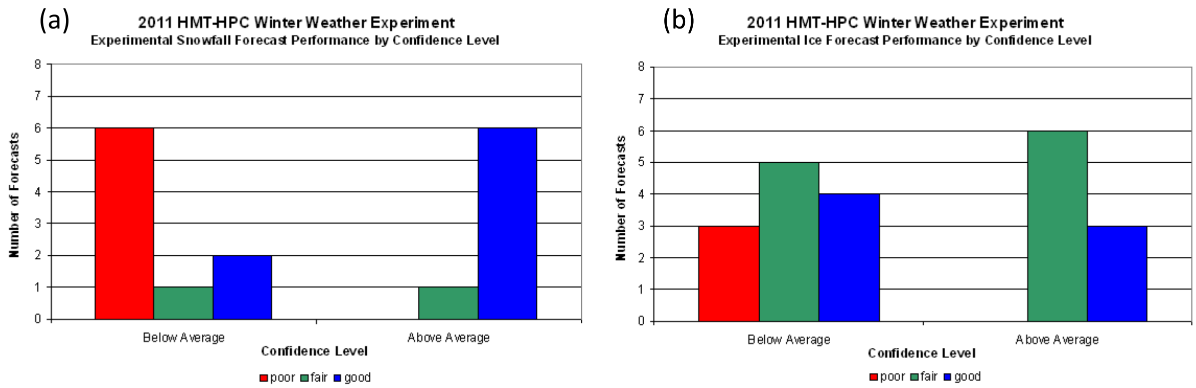


Fig. 8. Distribution of subjective ratings of human forecasts when confidence was below and above average for (a) snow and (b) ice accumulations.

Although a majority of participants found the SREF probabilistic and mean guidance useful (61% of participants), the observed amounts fell outside the bounds of the SREF max/min 46.2% (12/26) of the time for snow and 19.2% (5/26) of the time for freezing rain. The survey question did not discriminate between Day 1 or Day 2. Such underdispersion is a known weakness of ensembles and affects forecast confidence in ensemble information (e.g. Novak et al. 2008). The experiment's focus on high-impact events (storms) during a time of reduced hemispheric predictability (Fig. 2b) may contribute to the high frequency of missed events.

In summary, there is room for improvement in the quantification and communication of winter weather uncertainty for the Day 1-2 forecasts. The relative frequent "busts" and frequent occurrence of solutions falling outside the SREF envelope during the experiment (even at Day 1 forecast projections) highlights the inherent practical predictability challenges of winter weather, and the need for continued ensemble improvements. Forecasters exhibited a confidence-skill relationship for snow amount, which may be useful information for users. The overall value of confidence discussions is unclear, and will likely require user assessment in the future.

4. Additional Results

Besides results from the three core experiment questions, additional results can be gleaned from the experiment data. This section identifies these additional results.

Derived Fields

Snow and ice accumulations are not direct model output. Thus algorithms are needed to derive the amounts (Appendix C). For snow amount, the experiment tested four different approaches:

- Roebber: Precipitation type is derived using the NCEP Dominant Precip Type method (Manikin 2005). Snow amount is derived using a time-average ptype and the model QPF. The Roebber Technique (Roebber 2003) applied to the model temperature and humidity vertical profiles is used to determine the snow ratio. This is the method used operationally at HPC (applied to 6 h data operationally).
- Rime Factor: Diagnostic part of the Ferrier microphysics scheme that indicates the degree of riming on hydrometeors. Can derive an "ice accumulation rate" using rime factor value in the lowest layer of the model and assuming an appropriate snow to liquid ratio (12:1 in this case). The hourly data can be summed.
- Snow Depth: Snowfall determined from the change in model snow depth. Snow-to-liquid ratio of new snow calculated as a function of 2 m temperature.
- ECMWF Direct: Snow amount field derived directly from the ECMWF microphysics scheme.

In some cases, the uncertainties between these methods were larger than the meteorological uncertainties. Thus, it is important to learn the skill of different techniques. Fig. 9 shows results of three different comparisons which can help answer this question:

- Operational Roebber vs. experimental Rime Factor technique applied to the NAM.
- Operational Roebber vs. experimental snow depth technique applied to the expNAM.
- Operational Roebber vs. direct technique from the ECMWF.

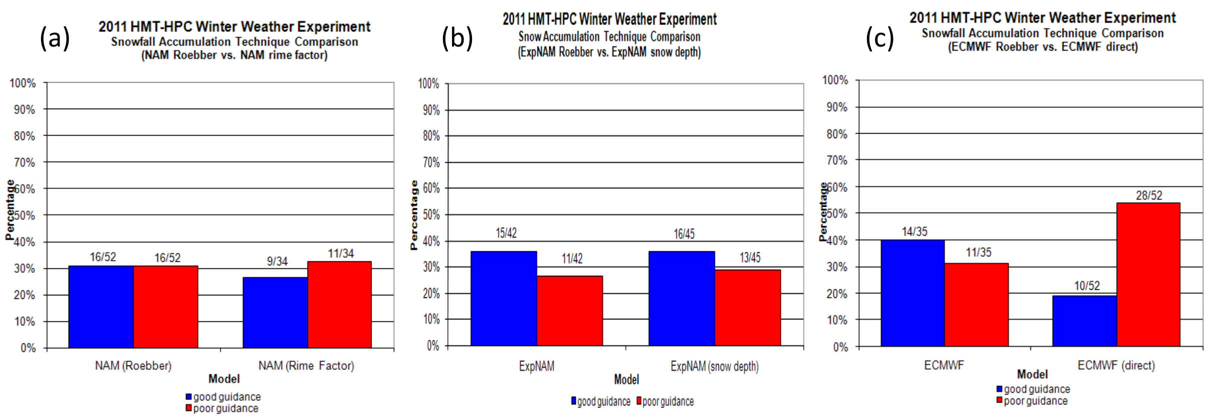


Fig. 9 Summary ratings for techniques to derive snow amount. (a) Comparison of operational technique (Roebber) and Rime Factor technique applied to the NAM (b) Comparison of operational technique (Roebber) and Snow Depth technique applied to the ExpNAM (c) Comparison of operational technique (Roebber) and direct technique from the ECMWF.

In all three comparisons the operational Roebber technique is similar or superior to the experimental techniques (Fig. 9). Although the snow depth parameter appears similar to the Roebber technique in the overall ratings (Fig. 9b), participants noted egregious errors in select cases. For example, since snowfall was calculated from the change in snow-water equivalent, rainfall or melting during the period caused erroneous snowfall values. Figure 10 shows an example where several inches of snow were predicted in the Cascades using the Roebber technique applied to the ExpNAM output, while no snow was predicted using the snow depth parameter. Several inches were observed. Melting and rainfall occurring after the initial snowfall likely accounts for this difference over the 24 h period. Thus a revised algorithm accounting for rainfall and melting processes may be promising.

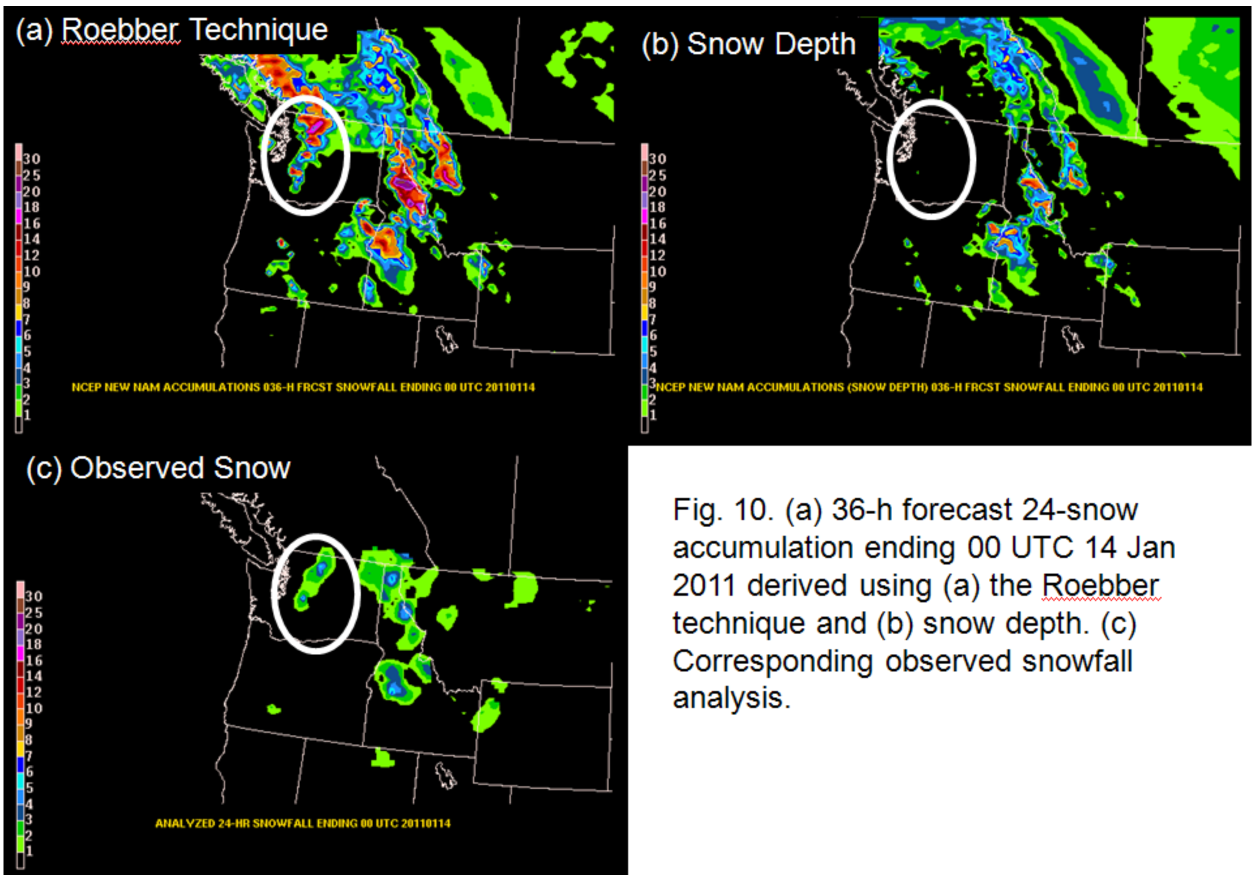


Fig. 10. (a) 36-h forecast 24-snow accumulation ending 00 UTC 14 Jan 2011 derived using (a) the Roebber technique and (b) snow depth. (c) Corresponding observed snowfall analysis.

Boundary Layer Temperatures

The experiment also highlighted how small differences in the boundary layer temperatures can affect winter weather. Table 2 shows an example of the observed and predicted (30 h forecast) 2 m temperatures for selected sites in the mid-Atlantic valid 06 UTC 2 Feb 2011. All forecasts had precipitation falling at the sites. The NAM and all experimental high-resolution guidance was too cold (2–6° F), while the GFS was within 2° F. Given the similar physics used in the NAM and experimental guidance, this case may suggest boundary layer temperature biases. Additional cases would be necessary to identify such an issue.

Station	**Observation**	**GFS**	**NAM**	**Exp NAM**	**HRW-ARW**	**HRW-NMM**
MDT (Harrisburg, PA)	30 (ZR)	31	27	28	26	27
PHL (Philadelphia, PA)	31 (ZR)	33	29	28	27	27
BWI (Baltimore, MD)	33 (RA)	34	30	30	30	30
IAD (Dulles, VA)	34 (RA)	34	32	32	32	32

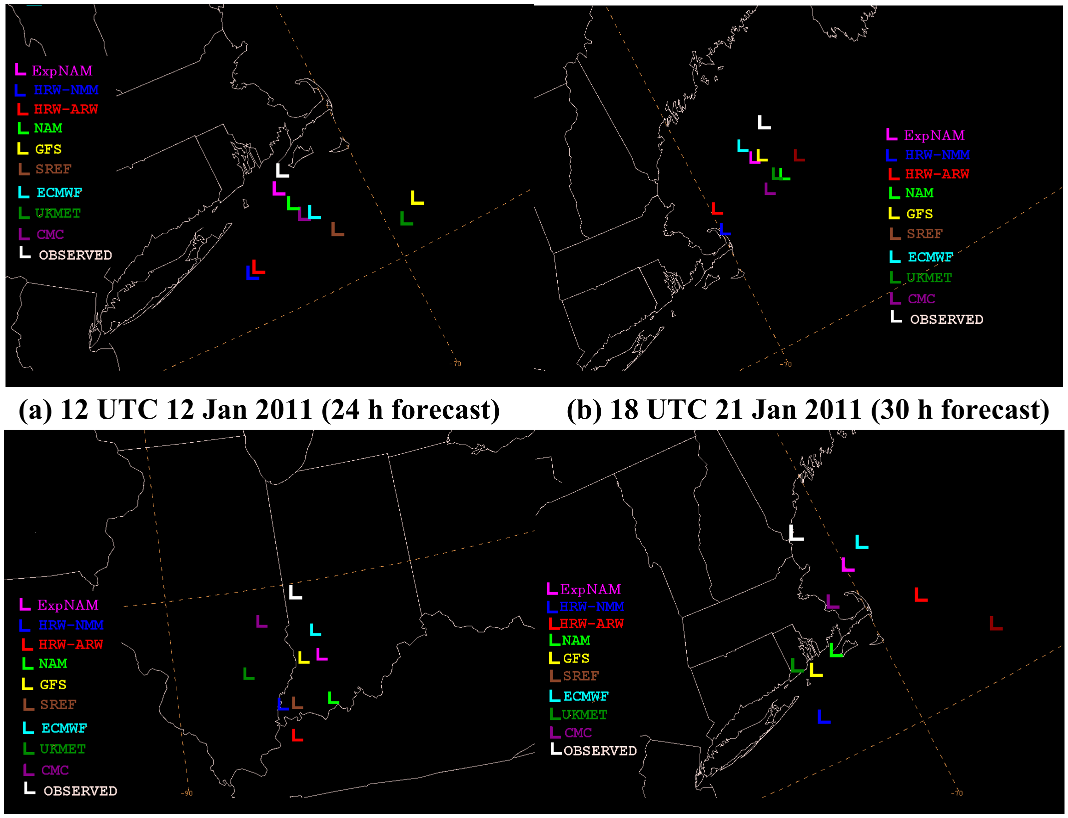
Table. 2. Observed and forecasted 2 m temperatures valid 06 UTC 2 Feb 2011.

Surface Low Positions and Depth

Surface low positions and depth were examined on an ad-hoc nature for major cyclones. Example low position forecasts are show in Fig. 11 while corresponding example cyclone depth forecasts are shown in Fig. 12.

A common finding was that the High-Res Window runs (both ARW and NMM) were substantially slower than most guidance (i.e., Fig. 11). This slow bias should be monitored in future model upgrades. On the other hand, the expNAM (12 km domain) improved on the operational NAM low position in all four cases (Fig. 11).

In terms of cyclone depth, operational guidance under predicted the depth of the cyclones in 3 of the 4 selected cases (Fig. 12). Among these four cases, the expNAM had the best average cyclone depth, and was closer to the observed depth than the operational NAM in all four cases, and closer to the observations than the ECMWF in three cases. There was no clear signal in the performance of the High-Resolution Windows relative to the operational NAM.



(a) 12 UTC 12 Jan 2011 (24 h forecast) **(b) 18 UTC 21 Jan 2011 (30 h forecast)**

(c) 06 UTC 2 Feb 2011 (42 h forecast) **(d) 06 UTC 6 Feb 2011 (42 h forecast)**

Fig. 11 Forecast and observed low positions for guidance from select storms.

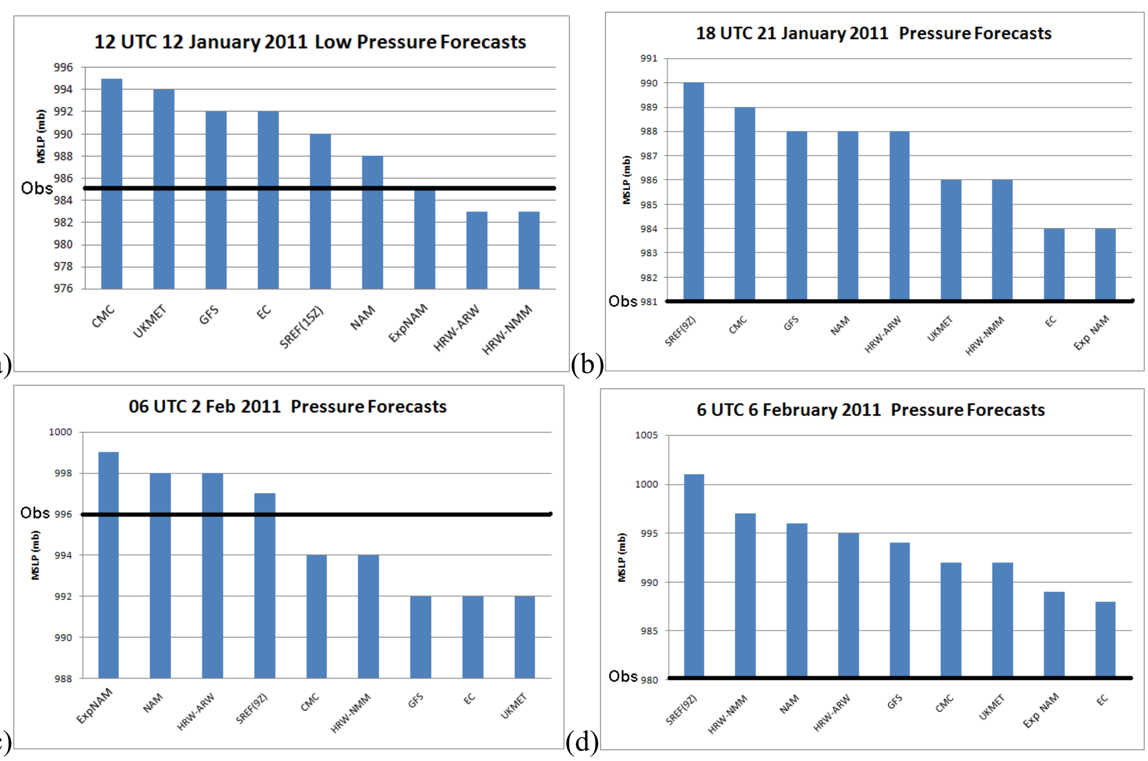


Fig. 12 *Forecast and observed MSLP for guidance from select storms corresponding to Fig. 10.*

5. R2O Actions

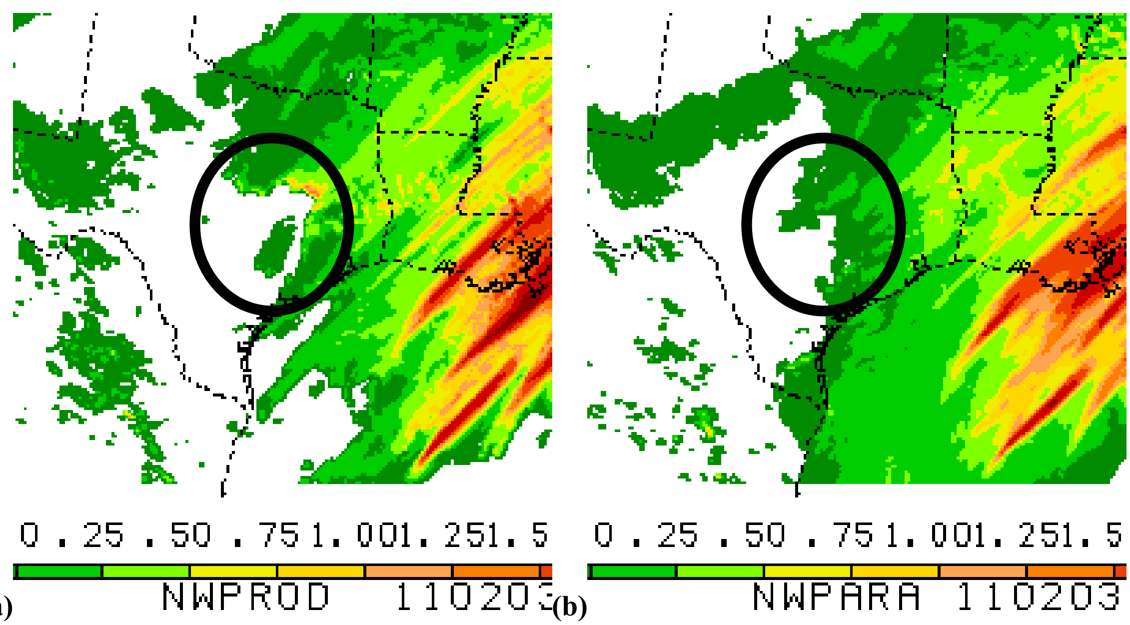
The experiment served as an evaluation period for numerous guidance datasets not typically utilized by HPC's Winter Weather Desk.Research-to-operations actions are divided into immediate impacts, and planned activities.

Immediate Impacts

- **The experimental NAM is ready for operational use in winter weather forecasting**
    - Comparable ptype, snow, and ice amount forecasts to operational, while superior low track guidance.
- **Development work is needed for the High-Res Windows**
    - The HRW-ARW and HRW-NMM were among the poorest scoring guidance sources for the Day 1 snow and ice accumulations.
    - HRW-ARW and HRW-NMM were slow with extratropical systems.
    - Of all the high resolution models, the HRW-ARW seemed to consistently have the most limited instantaneous precipitation type coverage, which affected its amounts.

- **Diagnosed and fixed "strange band" bug**

  The experiment had direct contribution to diagnosis and fixing of "strange band" bug in the HRW-NMM. Narrow (20 km) bands of heavy QPF occurred along the cloud ice/cloud water interface in the HRW-NMM. An example is shown in Fig. 13a. Cloud water drifted to the cloud ice and water loaded the crystals (Personal Communication, Ferrier and Pyle). EMC changed the maximum temperature at which at which ice nucleation (the first initiation of small ice crystals) occurs from -15C to -5C, and this change appears to have fixed the problem (Fig. 13b). This change is scheduled to be implemented March 29, 2011.



(a) (b)

Fig. 13. Comparison of a 48-h forecast of the 24 h accumulated precipitation (shaded according to scale, in) ending 12 UTC 3 February, 2011.

- **Examined utility of direct model output**
  It was hypothesized that more direct output from the models such as the NAM Rime Factor or NAM snow depth would be superior to operational methods used to derive snow and ice accumulations. The experiment results do not support this hypothesis. Thus, these datasets are not ready for operational use. However, future work on improving snow and ice accumulations from direct model output is encouraged.

- **Enhanced collaboration and discussions of winter weather issues**
  In addition to collaboration among the dedicated participants, increased incorporation of the experiment activities into operations was accomplished through daily wrap up discussions and map discussion presentations. These activities helped broaden the participation and provided a forum for educating users/researchers on National Center operational forecast issues, while learning about user needs and the latest science.

Recommended Actions

HPC

- **Consider incorporating expNAM (4 km) model guidance into the HPC QPF and Winter Weather blenders**
  Although the high-resolution guidance was comparable or less skillful than the NAM for snow and ice accumulations, participants noted a major advantage of the high-resolution guidance is the improved topographic depictions. Thus, incorporation of the more skillful expNAM into the HPC QPF and winter weather blenders can provide a more realistic starting point in areas of complex terrain.

- **Improve the HPC ice analysis**
  The current HPC ice accumulation analysis is poor, which makes verifying freezing rain forecasts very difficult. The current approach relies on METAR observations, which are sparse and of poor quality. Possible mitigating actions include:
    - Use observations from WFO LSR product or WFO RMR product
    - Implement manually drawn analyses

- **Create sounding display from the high resolution windows**
  It was difficult to take full advantage of the improved model resolution offered by the high resolution window runs because of limited information about vertical profiles. The high resolution window upgrade scheduled for Spring 2011 will include the data needed to display soundings in NSHARP.

- **Consider including standard confidence information in Heavy Snow and Icing Discussion**
  To help forecasters focus on the predictability of each scenario, standard confidence information could be integrated into the current HPC Heavy Snow and Icing Discussion. For example, the overall forecaster confidence for each system could be noted, and the rationale for the confidence discussed.

NCEP

- **Improve precip type and amount methods**

  The HRW-ARW precip type field was sparse and not realistic, and contributed to poor derived snow amounts in some cases. Also, the uncertainty in the method used to derive snowfall from model guidance was as large as the meteorological uncertainty in some cases. Given the snow depth parameter was well rated, except in rain-on-snow situations, or when melting occurred during the time period, consider development work on a modified snowdepth that accounts for such processes.

- **Establish NCEP standard for calculating snow ratio**

  Currently several different approaches are used within NCEP to calculate snow ratio including:

  - SPC SREF is based on sfc temp
  - EMC SREF is a static 10:1
  - NAM snow depth is based on sfc temp
  - HPC is based on Roebber neural net

  Although a direct comparison of these methods was not attempted during the experiment, the positive results of the Roebber technique (e.g., Fig. 9) suggests it is a skillful method.

NWS

- **Implement a standard gridded snowfall analysis**

  Currently there are three primary gridded analyses available:

  - HPC analysis (uses RFC QPF and Barnes analysis of METAR, COOP, and COCORAHS)
  - CR GIS analysis
  - NOHRSC analysis (surface obs and snow model)

  Examples of these analyses are shown in Fig. 13. Each analysis has separate pros and cons, and thus collaborative development is encouraged. A standardized snowfall analysis would allow for a consistent answer of what fell, and allow intercomparison of verification statistics.

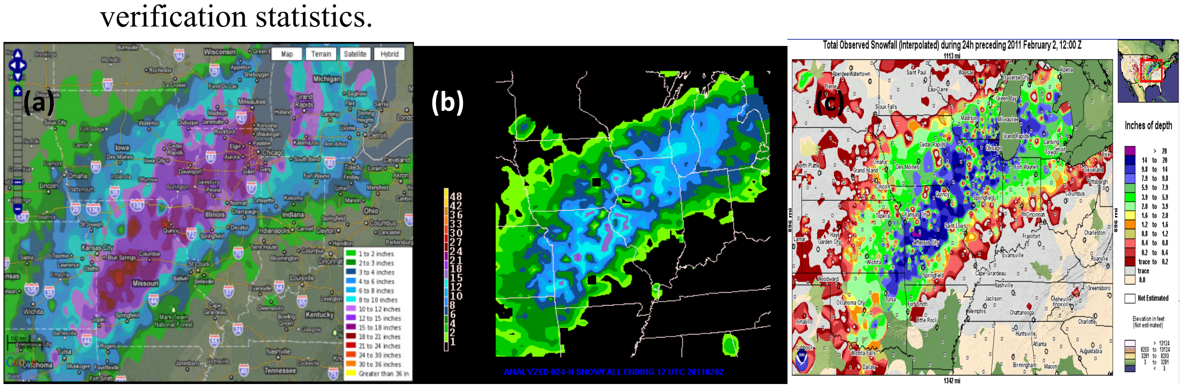


Fig. 13 24-h accumulated snowfall ending 12 UTC 2 Feb, 2010 from (a) Central Region GIS analysis, (b) HPC analysis, and (c) NOHRSC analysis.

6. Future Winter Weather Experiments

Based on the success of the 2011 HMT-HPC Winter Weather Experiment, a future experiment is encouraged. From the 2011 experiences, future experiments should continue to focus on a storm of interest within a national domain with the theme of evaluating new high resolution models and ensembles. Additional items to consider in future experiments include:

- Adding QPF and storm track to forecasts and/or evaluations
- Consider replacing freezing rain assessment with freezing line assessment
- Consider societal impacts aspects
- Provide more context with evaluation of model diagnostics fields
    - look at initializations
    - 500 mb comparisons
    - soundings
    - derived fields such as frontogenesis / wet bulb zero height
    - model trends
- Test improved derived snow and ice methods.

Expanded participation from WFOs and others is also encouraged. To facilitate enhanced participation, improvements to the testbed facility are encouraged including:

- Additional workstations to expand participation
- Upgraded workstations to improve the speed of display of large datasets
- Larger monitors to facilitate group discussions
- Redundant data backups to assure reliability of data.

7. Summary

The inaugural Winter Weather Experiment of the HMT-HPC was conducted January 10 - February 11, 2011, at HPC. Participants included forecasters from HPC, other National Centers, and Weather Forecast Offices, as well as modelers from the EMC and researchers from NOAA HMT. The experiment explored the use of convection-allowing (~4-km resolution) models for improving near term forecasts of snow and freezing rain accumulations, and investigated different methods of quantifying and communicating the uncertainty associated with winter weather forecasts.

During the experiment, participants issued experimental snow and freezing rain forecasts for two 24-hour periods (Day 1 and Day 2). After completing the forecasts, participants were asked to write a forecast confidence discussion in which they detailed the uncertainties in the forecast, rated their overall confidence, and provided probabilities of a key winter weather event (such as snowfall greater than 4 inches) occurring at a location of interest. In addition to the experimental forecasts, participants subjectively evaluated the quality of the available model guidance for forecasting both snow and freezing rain accumulations as well as precipitation-type transitions.

The experimental NAM, which is scheduled to become operational in Spring 2011, was found to provide useful winter weather forecast guidance. However the high-resolution windows were not as skillful as the operational NAM, and require additional development. The experiment also revealed the difficulty of deriving snow and freezing rain accumulations from the models, with significant differences sometimes observed between techniques to derived amounts. Although hourly 4 km data can allow a forecaster to visualize explicit predictions mesoscale bands, the high-resolution guidance was only found useful in anticipating the placement and intensity of mesoscale snowbands when there was consensus among the guidance.

A majority of participants found the SREF probabilistic and mean guidance useful (61% of participants), however, the observed amounts fell outside the bounds of the SREF max/min 46.2% (12/26) of the time for snow and 19.2% (5/26) of the time for freezing rain. This result highlights the need for additional ensemble development. An analysis of the forecast confidence discussions shows that snowfall forecasts rated as low (high) confidence by the forecast team tended to be associated with larger (smaller) forecast errors. This information may be useful for users.

The HMT-HPC Winter Weather Experiment provided a unique opportunity to foster winter weather collaboration between participants from the research and operational forecasting communities. The issues discovered during the course of the experiment are being explored through collaboration with partners, including EMC. The HMT-HPC plans to build on these discoveries and the overall success of this year's experiment in 2012.

8. References

Barthold, F. E., and CoAuthors, 2011: The quantitative precipitation forecasting component of the 2010 NOAA hazardous weather testbed spring experiment. Preprints, *24th Conference. on Wea. Anal. and Forecasting/ 20th Conference on NWP*, Amer. Meteor. Soc., Seattle, WA, 9A.4.

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

Eckel, F. A., C. F. Mass, 2005: Aspects of effective mesoscale, short-range ensemble forecasting. *Wea. Forecasting*, **20**, 328–350.

Kocin, P. J., and L. W. Uccellini, 2004: A snowfall impact scale derived from northeast storm snowfall distributions. *Bull. Amer. Meteor. Soc.*, **85**, 177–194.

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.

Novak, D. R., D. R. Bright, M. J. Brennan, 2008: Operational forecaster uncertainty needs and future roles. *Wea. Forecasting*, **23**, 1069–1084.

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

APPENDIX A

Testbed Support

Project Supervisor: Ed Danaher

Project Leader: David Novak

Facilitator: Faye Barthold

Science and Technical Support: Mike Bodner

Participants

Week	HPC Forecaster	Visitor	Visitor
Jan 10-14	Dan Petersen	Dan Baumgardt (WFO ARX)	Jon Racy (SPC)
Jan 18-21	Frank Pereira	Steve Zubrick (WFO LWX)	Bruce Entwistle (AWC)
Feb 1-4	Chris Hedge Mike Musher	Bill Bua (UCAR/COMET)	Ying Lin (EMC)
Feb 7-11	Rich Bann	Ellen Sukovich (HMT) Ben Moore (HMT)	Geoff Manikin (EMC)

APPENDIX B

Daily Schedule

8:30am-9:30am—Create ptype analysis at key synoptic time of interest. Use resulting analysis to evaluate experimental guidance performance as it relates to precipitation type. Use this information to complete ptype survey questions. On the first day of each week, a brief orientation session will be held to explain the motivation and organization of the experiment as well as some of the data being evaluated.

9:30am-11:00am—Use HPC 24-h snow and ice analysis to evaluate experimental guidance performance as it relates to precipitation amounts and indication of forecast uncertainty. Use this information to complete snow/ice amounts survey questions.

11:00am-11:30am—HPC-CPC Map Discussion

11:30-Noon—Lunch

Noon-12:30pm—Synoptic overview and determine the storms of interest for the Day 1 and Day 2 time periods.

12:30am-2:30pm—Create Day 1 24-h snow/sleet and ice accumulation forecasts based on 12 UTC guidance. Prepare a forecast confidence discussion covering the uncertainties in the Day 1 forecast.

2:30pm-4pm—Create Day 2 24-h snow/sleet and ice accumulation forecasts based on 12 UTC guidance. Prepare a forecast confidence discussion covering the uncertainties in the Day 2 forecast.

4pm-4:30pm—Daily wrap-up discussion

APPENDIX C

Experimental Data

The following experimental datasets were available:

NCEP NAM Rime Factor: Diagnostic of degree of riming directly from the Ferrier microphysics scheme. Can be used to derive snow accumulations.

High-Resolution Window Run ptype, snow, and ice accumulations: Precipitation type and snow, sleet, and ice amounts are derived from model output. Precipitation type is derived using the NCEP Dominant Precip Type method (Manikin 2005). Ice and sleet amounts are determined using a time-average ptype and the model QPF. A 1:1 ratio is assumed. Snow amount is derived using a time-average ptype and the model QPF. The Roebber Technique (Roebber 2003) applied to the model temperatures and humidity profiles are used to determine the snow ratio.

ExpNAM ptype, snow, and ice accumulations: Pre-implementation version of the expNAM (NMMB model), including the 12 km parent domain and 4 km CONUS nests. Precipitation type and snow, sleet, and ice amounts are derived from the 4 km CONUS nest as with the high-res window runs (see above). The 12 km parent domain data are used for MSLP and the synoptic environment.

ECMWF deterministic snow amount: Direct model output of snowfall determined from the convective cloud and stratiform cloud parameterizations.

HPC Super Ensemble snow and ice accumulations: Ensemble mean of derived snow and ice accumulations. Membership is comprised of the SREF members, NAM, GFS, ECMWF deterministic, CMC deterministic, SREF mean, and GEFS mean, for a total of 28 members. Precipitation type is derived using the NCEP Dominant Precip Type method (Manikin 2005) for NCEP guidance, and a simplified thermal profile technique for foreign model guidance. Ice and sleet amounts are determined using the ensemble relative frequency of ptype and the ensemble average QPF. A 1:1 ratio is assumed. Snow amount is derived using the ensemble relative frequency of ptype and the ensemble average QPF. The snow ratio is a 4 member average of the Roebber Technique applied to the NAM, Roebber Technique applied to the GFS, climatology (Baxter 2005), and a fixed 11:1 ratio.

NCEP SREF probabilities and derived fields: Unconditional ptype probabilities, max/min snow and ice amounts, probability of persistent freezing rain, and probability of a dendritic layer exceeding threshold depths.

HMT-West ensemble: Ensemble mean QPF from a nine member nested ensemble centered over northern California (9 km parent domain surrounding a 3 km nest). This model is run at GSD and is initialized with GFS data.

APPENDIX D

Model/Human Forecast Evaluation Survey Questions

*question was also asked for Day 2

*Using your precipitation type analysis, please rate the ability of each Day 1 model forecast to correctly predict the observed precipitation type transitions at the chosen synoptic time.

*Using the gridded snow and ice analyses, please rate the overall quality of the experimental snow/sleet and ice accumulation forecasts for the Day 1 forecast period.

*Using the gridded snow analysis, please rate the ability of each model to correctly predict the observed snow/sleet accumulation amounts for the Day 1 forecast period.

*Using the gridded ice analysis, please rate the ability of each model to correctly predict the observed ice accumulation amounts for the Day 1 forecast period.

Using a combination of radar data and the gridded snow analysis, please rate the ability of each model to predict any observed mesoscale snowbands during the Day 1 forecast period.

Do the observed snow and freezing rain amounts fall within the maximum and minimum amounts indicated by the SREF? Please consider both the Day 1 and Day 2 forecasts.

Forecast Confidence Discussion Questions

Overall, how would you rate your confidence in the forecast (above average, average, or below average)? (i.e., is there a large spread between the SREF max/min amounts?)

What features are the forecast team members unsure of? Why?

Where is the uncertainty in the forecast coming from?

What guidance datasets were outliers?

Can you quantify the probability of a key event occurring?
(i.e, 70% prob of >4” at DCA / 1 in 3 chance of northerly storm track)?