

Analyzing Predictability and Communicating Uncertainty: Lessons from the Post-Groundhog Day 2009 Storm and the March 2009 “Megastorm”

NEIL A. STUART

NOAA/National Weather Service, Albany, New York

RICHARD H. GRUMM

NOAA/National Weather Service, State College, Pennsylvania

MICHAEL J. BODNER

NOAA/NWS/National Centers for Environmental Prediction, Camp Springs, Maryland

(Manuscript received 15 November 2012; review completed 26 March 2013)

ABSTRACT

Forecasting winter storms in the northeastern United States during the 2008–2009 season was very challenging owing to large uncertainty in the numerical weather prediction guidance prior to each storm. Forecasts for the February 2009 post-Groundhog Day event and the March 2009 “megastorm” featured significant spatial and timing errors in storm track, precipitation type, and areal extent.

Each storm’s impacts were communicated with considerable certainty, leading to confusion and misunderstanding of the actual uncertainty in each event. Both cases can serve as instructional examples for the forecast community to improve interpretation of levels of uncertainty, along with communication of uncertainty, during potentially high-impact events.

Examples of spatial and temporal uncertainties associated with both storms are presented. These uncertainties are illustrated using ensemble data from the National Centers for Environmental Prediction Global Ensemble Forecast System, Short Range Ensemble Forecast, and the higher-resolution deterministic models such as the North American Mesoscale model, Global Forecast System, and European Center for Medium-Range Weather Forecasts. In addition to standard ensemble output, forecast anomalies are presented because highly anomalous situations frequently have large societal impacts.

Techniques for analyzing ensemble probabilities for quantitative precipitation forecast (QPF) threshold values and ensemble plume QPF diagrams are demonstrated. In addition, various combinations of deterministic and ensemble means and spreads for mean sea level pressure and 500-hPa height will be presented to evaluate the predictability of surface low-pressure tracks. Several experimental techniques are presented to promote better understanding of predictability and better communication of uncertainty to users.

1. Introduction

Evaluating predictability in weather forecasting has been evolving rapidly in the past ten years with the advent and maturing of ensemble forecast systems such as the Short Range Ensemble Forecast (SREF) system and the Global Ensemble Forecast System [GEFS; formerly the Medium Range Ensemble Forecast]. Events, such as the “surprise” storm of January 2000 (Kleist and Morgan 2005; Tracton 2008) that affected the eastern United States with 30–60 cm

(1–2 ft) of snow, were studied using early research versions of ensemble forecast output to determine if any perturbations within the ensemble spread showed probabilities for the storms. Revelations of the success of some of the ensemble data accurately predicting the storm prompted the development and improvement of the aforementioned ensemble forecast output (SREF and GEFS). Improvements in the ensemble forecast output, including increasing resolution and number of perturbations within each dataset, are ongoing.

Since that landmark snowstorm of January 2000, techniques to analyze ensemble forecast guidance have improved the predictability of all types of high-impact weather events. Recent successes in extending warning lead times to as much as three days have proven the value of ensemble forecast guidance. Examples include snowstorms such as the Valentine's Day 2007 snowstorm (Grumm and Stuart 2007) and the active winters in the eastern United States since 2009, in addition to tropical cyclones such as Isabel (NWS 2004), Irene (NWS 2012), and Sandy (Blake et al. 2013). Forecasters continue to learn how to determine probabilities for the likelihood of various forecast scenarios by analyzing the spread in sources of deterministic and ensemble guidance.

However, failures in evaluating predictability can result from misuse and misunderstanding of ensemble guidance in determining probabilities for a spectrum of weather scenarios. These problems were quite apparent during less-predictable storms (i.e., the subjects of this study—the February 2009 and March 2009 snowstorms), where specific regions with the highest potential for high-impact snowfall amounts in each storm were not well-resolved in ensemble or deterministic guidance until 12–36 h prior to impact. The low predictability of these storms resulted in conflicting messages from many sources of weather information. Learning when to first notify the user community of the likelihood of a high-impact event can be related to the level of confidence determined through analysis of ensemble probabilities.

This study focuses on two challenging events that occurred during the 2008–2009 winter season in which anecdotal evidence suggests segments of the forecasting community misinterpreted and misused ensemble data, resulting in a much higher confidence of high-impact events than what was warranted based on the large spread depicted in ensemble guidance. In early 2009, two snowstorm events were promoted by the forecasting community [including broadcast and Internet media and the National Weather Service (NWS)] as high-impact events three or more days prior to onset. Both East Coast events, the non-event of 3 February 2009 (hereinafter the February '09 event) and the “megastorm” of 3 March 2009 (hereinafter the March '09 event), had significantly less impact than advertised.

The different levels of confidence expressed by most sources of weather forecasts, including the NWS and broadcast and Internet media, imply that forecasters have a wide range of understanding and

ability to evaluate predictability. Predictability is defined as “the extent to which future states of a system may be predicted based on knowledge of current and past states of the system” (Glickman 2000). Effective evaluation of predictability optimizes the potential for a consistent message from the forecasting community when communicating levels of confidence or uncertainty to the user community. Conflicting messages from the forecasting community (including the private sector¹ and NWS²)—from five days to as little as 24 h prior to storm impact—suggest that many forecasters do not always effectively use ensemble guidance to assess predictability.

In both events, forecast guidance provided by numerical weather prediction (NWP) models and ensemble prediction systems (EPS) showed considerable spread about the ensemble mean, indicating a high degree of uncertainty. In contrast, the high degree of uncertainty was not communicated effectively to the public by some sources of weather information in the February '09 storm, as indicated by the storm being described as “The Megastorm,” “Ground Hogzilla,” and “Big Daddy Storm.” It also was compared to the superstorm of March 1993 (Kocin et al. 1995) five days prior to the expected impact.

The use of these terms to describe the event suggested high confidence of a high-impact event when, in actuality, conflicting NWP guidance and considerable uncertainty associated with the event were present. Thus, it is clear that NWP and EPS data were misinterpreted and/or misunderstood. NWS offices in the northeastern United States also were suggesting a “significant storm likely” and the possibility of “significant rain and/or snow” in Hazardous Weather Outlooks (HWOs) and Area Forecast Discussions (AFDs) 3–5 days prior to storm impact. Terms used in HWOs and AFDs at adjacent NWS offices included “likely” ($\geq 55\%$ probability), “possible” (between 25 and 54% probability), “could,” and “may” (“could” and “may” are subjective terms that do not correspond to any specific probabilities but convey uncertainty). These words provided a

¹ The Washington Post Capital Weather Gang articles on [4 February 2009](#) and [6 February 2009](#) summed up issues and names that were assigned to this storm.

² NWS [Hazardous Weather Outlooks and Area Forecast Discussions](#) across the northeastern and mid-Atlantic United States highlighted varying degrees of confidence for a significant storm, suggesting a broad spectrum of understanding of ensemble data.

conflicting message of confidence to users and implied more uncertainty than the superlative labels applied to the storm.

The actual impact of the February '09 storm was much less than what was advertised, with 8–15 cm (3–6 in) of snow and isolated reports of up to 20 cm (8 in) from central and eastern Pennsylvania through Long Island, New York, and Massachusetts (Fig. 1). The storm did not even attain a Northeast Snowfall Impact Scale (NESIS) ranking (Kocin and Uccellini 2004). In contrast, the March '09 storm, which also was predicted to be a megastorm by some private sector and broadcast information sources, had a more widespread impact in terms of the areal extent of observed snowfall (Fig. 2). Nevertheless, the March '09 event was only rated a 1 out of 5 on the NESIS (the lowest ranking possible), and was significant perhaps in the southeastern United States but hardly a megastorm based on the minimal societal impacts represented by this lowest ranking. Again, there was a contrast in messages sent by different providers of weather information, ranging from megastorm to a widespread, accumulating snowfall of 25 cm (10 in) or less, suggesting different understandings of predictability and a broad spectrum of methods for communicating levels of confidence within the operational forecasting community.

The National Centers for Environmental Prediction (NCEP) GEFS, along with a comparison of individual deterministic models from different forecast centers, showed that the storms had contrasting degrees of uncertainty associated with them. There was much more spread in NWP model and ensemble forecast guidance in the details such as storm track and intensity in the February '09 event, indicating a lower degree of predictability. Conversely, there was less spread in the NWP model and ensemble guidance in the storm track and intensity in the March '09 event, indicating a higher degree of predictability.

However, it is important to emphasize that every storm exhibits different levels of predictability for different meteorological variables. In the February '09 storm, the track and intensity in the 2–5-day time frame were the most uncertain variables, possibly because of issues of poor initialization and resolution of multiple 500-hPa upper troughs that were predicted to phase. In contrast, in the March '09 storm the intensity, precipitation amounts, and placement of the precipitation maxima were the most uncertain variables in the 1–2.5-day time frame. The purpose of this study is to demonstrate how misinterpretation of

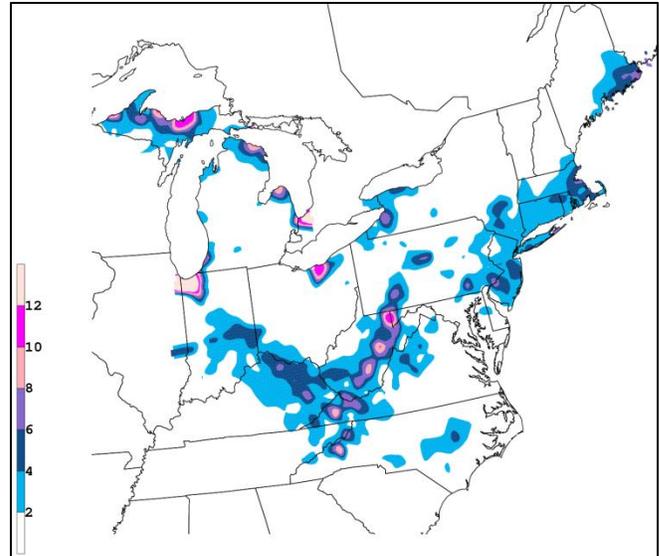


Figure 1. Observed snowfall in inches from the 3–4 February 2009 storm (1 in = 2.54 cm). *Click image for an external version; this applies to all figures hereafter.*

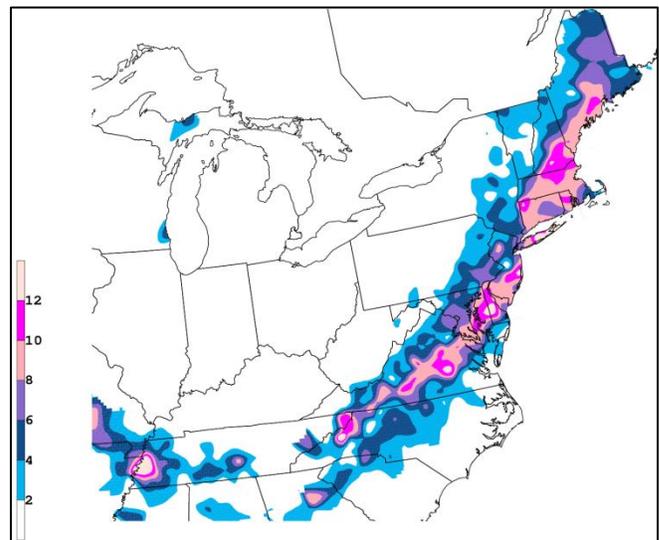


Figure 2. Observed snowfall in inches from the 1–3 March 2009 “megastorm” (1 in = 2.54 cm).

the predictability of different meteorological variables contributed to overforecasting the impacts of each storm. In addition, ensemble display techniques (e.g., Sivillo et al. 1997), including lagged average forecasts (LAFs), are used in order to demonstrate how the interpretation of predictability can be improved.

2. Methods

Datasets that were available to forecasters in real-time were archived and reviewed for both events. Displays featuring a poor man’s ensemble (PME)

technique, a LAF technique (Hoffman and Kalnay 1983), and a simple ensemble technique (to be discussed later in this section) are shown. Derived probabilities of 1.27 cm (0.5 in) and 2.54 cm (1 in) of liquid equivalent quantitative precipitation forecasts (QPFs), means and spreads of mean sea level pressure (MSLP), and 500-hPa heights and plume precipitation displays of multiple ensemble members also are presented.

All of the forecasts for the February '09 and March '09 events were based on a wide range of NWP models, including the NCEP Global Forecast System (GFS), the North American Mesoscale (NAM) model, the Meteorological Service of Canada Global Environmental Model (GEM), the United Kingdom Meteorological Office global model, and the European Centers (EC) for Medium-Range Weather Forecasts (ECMWF) high-resolution model. Many forecasters also have access to EPS output, including the GEFS and SREF, from one or more of these centers.

An experimental technique for evaluating forecast confidence used at the Hydrometeorological Prediction Center will be used. Referred to locally as the “flip-flop” technique, this version of a LAF uses three consecutive runs of both the GFS and ECMWF output to determine the changes in 500-hPa height over three successive guidance initializations times (i.e., 12 h between initializations for both the GFS and ECMWF). The difference in height at a given forecast hour for consecutive guidance initialization times is multiplied, and the square root of this quantity is calculated, giving the so-called flip-flop (FF):

$$FF = \sqrt{(H5_{t-12} - H5_{t-24})(H5_{t-12} - H5_{t00})}, \quad (1)$$

where $H5$ is the 500-hPa geopotential height, t is time, and FF is displayed in units of dekameters (dam).

Positive values of FF represent inconsistent trends in the models over the three consecutive forecast cycles (i.e., flip-flops). For example, a positive value will result if the 500-hPa heights fall and then rise at a grid point over three successive forecast cycles. Larger values imply greater uncertainty in the 500-hPa geopotential height field. Negative values are not depicted, as they indicate model consistencies from run to run and imply steady or improving forecast consistency. This technique applies only to deterministic operational NWP models, as EPS mean-height forecasts often exhibit errors of much less

magnitude than the deterministic operational NWP models (Zsoter et al. 2009).

Additionally, operational forecasters rely heavily on deterministic operational NWP model output for short- and long-range forecasting. Forecasters often examine run-to-run changes in model forecast output. The experimental 3-cycle LAF technique and the ensemble-based techniques presented in this study provide both a quantitative method to track these changes and an estimate of forecast uncertainty.

Standardized anomalies (Grumm and Hart 2001) also were evaluated to gauge the climatological significance and potential magnitude of these events. The departure of the ensemble means from normal depicted in the NWP model and ensemble guidance are displayed in standard deviations (SDs) from normal. A convergence of forecasts is required to produce larger anomalies (Zhu 2005), meaning that more ensemble members depicting anomalies of relatively high magnitude will result in larger ensemble mean anomaly values.

3. Results

a. The February '09 storm

The storm was given superlative labels by some weather information sources at lead times beyond three days. It is shown that, at these lead times, significant spreads were depicted by deterministic LAFs and ensemble guidance with regard to key forecast parameters such as storm timing, track, and thermal profiles that determine precipitation type. The LAFs will be presented with different sequential combinations of GFS and ECMWF initializations. Additionally, two different valid times are depicted because the timing of the storm varied with each set of guidance and each initialization, which often occurs 3–5 days or more prior to an event because of increasing errors with time. This contributes to the spread and uncertainty in the guidance. Within three days of the storm, the predictability increased as the onset of the storm drew closer, and it became apparent the storm would produce less snowfall than was forecasted at longer ranges.

1) DETERMINING UNCERTAINTY WITH THE 3-CYCLE LAF

The 3-cycle LAFs shown in Fig. 3 comprise the 0000 UTC 29 January, 1200 UTC 29 January, and 0000 UTC 30 January cycles valid at 0000 UTC 4

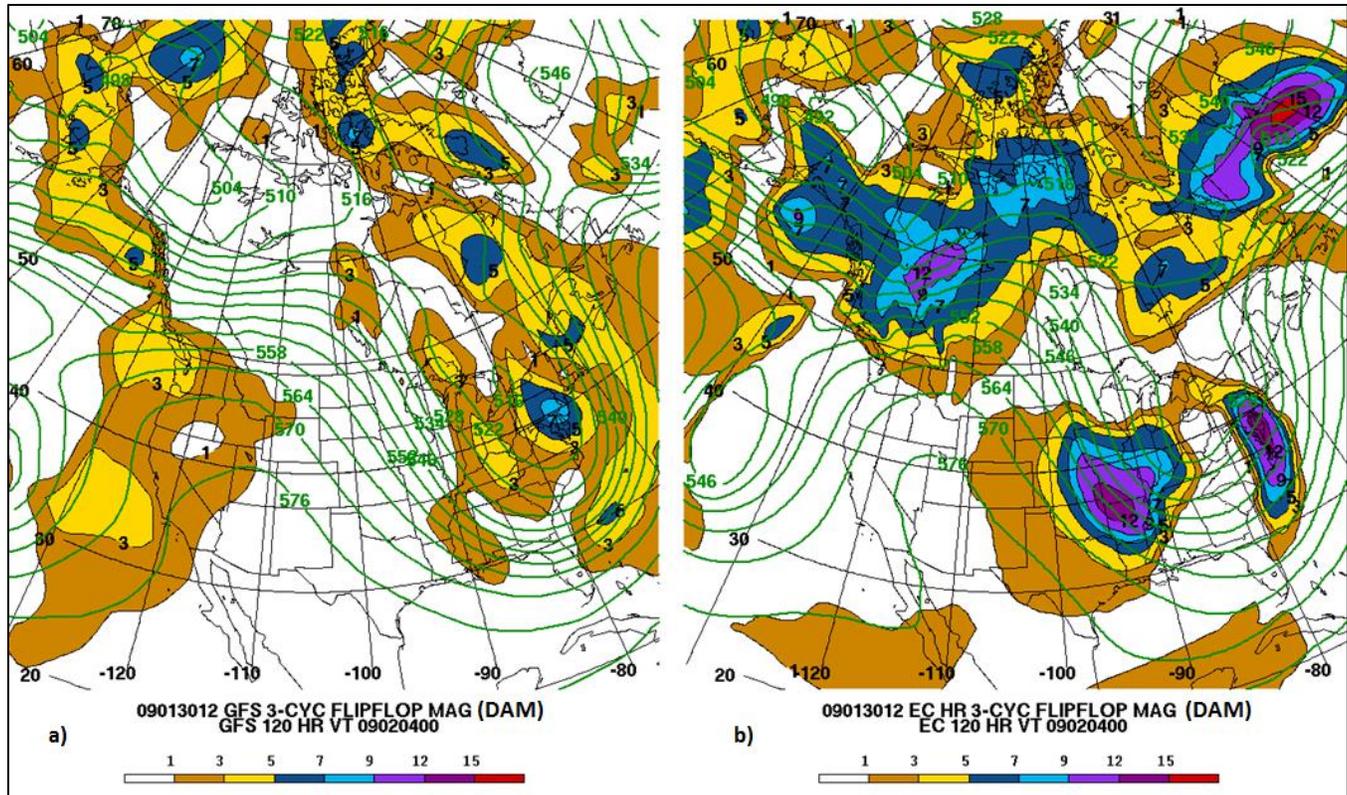


Figure 3. LAF 500-hPa heights (contour, dam) and flip-flop magnitude (shaded, dam) comprising the 1200 UTC 29 January, 0000 UTC 30 January, and 1200 UTC 30 January 2009 runs of a) GFS and b) ECMWF. Valid time is 0000 UTC 4 February 2009.

February. Positive values represent inconsistent trends in the models over three consecutive forecast cycles, referred to as flip-flops. Both the GFS (Fig. 3a) and the ECMWF (Fig. 3b) show relative maxima, indicative of greater forecast inconsistency over the mid-Atlantic coastal region, with the ECMWF showing the greatest changes over the three forecast cycles. The maxima concentrated near the 500-hPa trough indicate that successive runs of both models exhibit inconsistent trends in the position and depth of the 500-hPa trough. The ECMWF also suggests significant changes in magnitude in the upstream side of the 500-hPa trough and with the next upstream trough moving over the western North American upper ridge. This feature suggests that the ECMWF exhibits even greater uncertainty in the 500-hPa geopotential height field than the GFS. Additional LAFs comprising different 3-cycle periods between 1200 UTC 29 January and 1200 UTC 30 January also showed differences in positions of the 500-hPa trough, or flip-flops of more than 10 dam (328 ft) associated with the 500-hPa system. These differences decreased in magnitude at ≤ 3.5 days prior to the expected impact, and thus implied increasing predictability or increasing

certainty in the forecast; this is typical as the storm impact gets nearer, but varies from storm to storm.

2) PREDICTABILITY OF PRECIPITATION TYPE AND AMOUNT

Model forecast trends in precipitation types and amounts became apparent across the northeastern United States after the 1200 UTC 29 January set of guidance. Displays of plume diagrams at Albany, New York (Fig. 4), illustrate the significant changes in the GEFS forecasts for rain, snow, sleet, and freezing rain in the 1200 UTC guidance between 29 January and 2 February. The initial trends between 1200 UTC 29 January and 1200 UTC 31 January show an abrupt shift to snow as the predominant precipitation type at Albany, New York, along with a reduction in the overall liquid equivalent precipitation as seen in the clustering of most members below ~ 0.8 cm (0.31 in, Fig. 4b). By 1200 UTC 2 February, the predicted amounts of snow were greatly reduced, as seen by the clustering of most of the members below ~ 0.3 cm (0.12 in) of liquid equivalent precipitation (Fig. 4d).

Another effective method for evaluating liquid equivalent precipitation is to analyze probabilities for

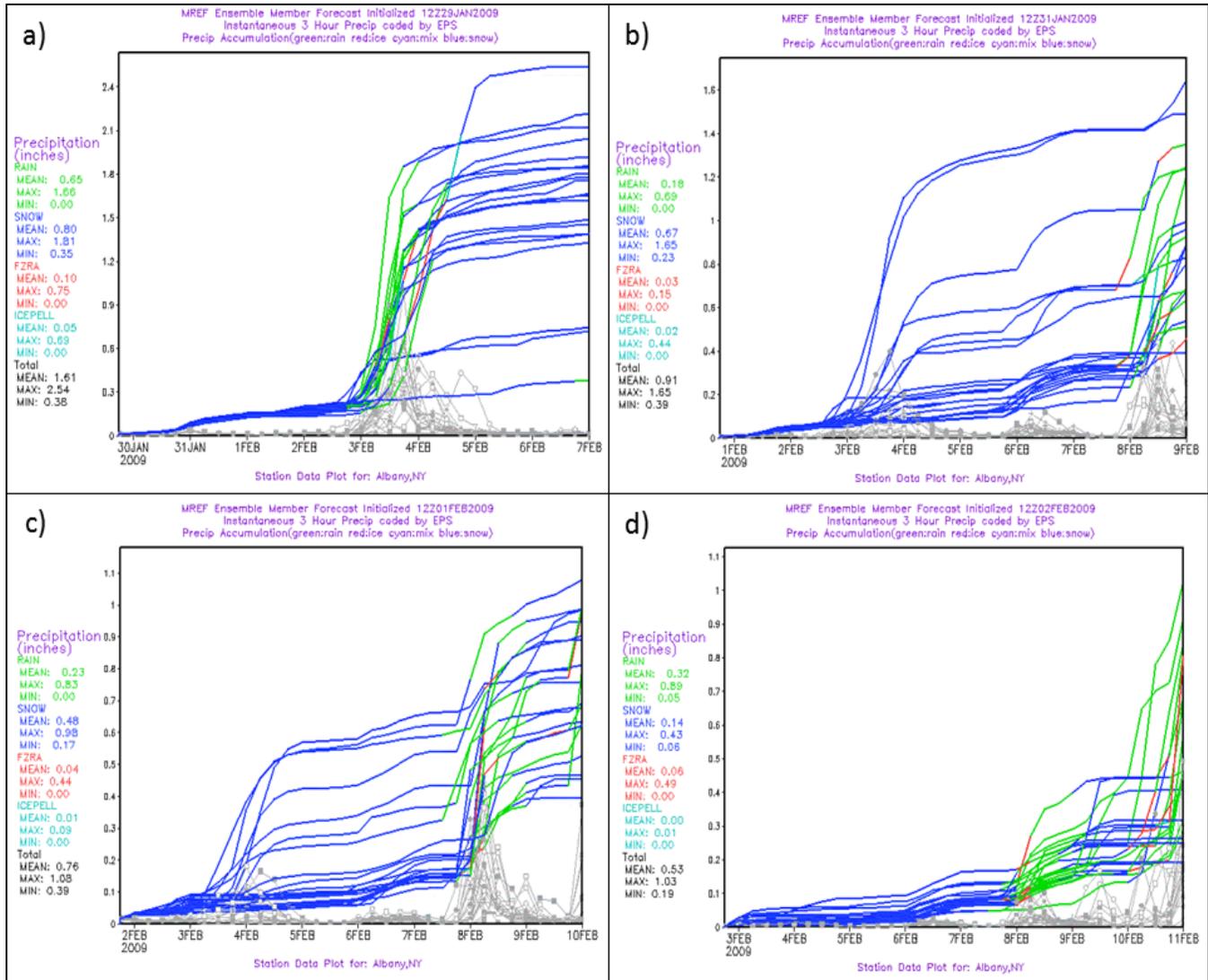


Figure 4. Plume diagram of 8-day forecasts for Albany, NY, from the GEFS (rain, liquid equivalent snow, sleet, and freezing rain depicted as green, dark blue, light blue, and red, respectively) initialized at a) 1200 UTC 29 January 2009, b) 1200 UTC 31 January, c) 1200 UTC 1 February, and d) 1200 UTC 2 February. The gray filled squares and open circles represent instantaneous 3-h precipitation for each ensemble member.

different amounts of liquid equivalent precipitation in the ensemble guidance. Depictions of the 50% or greater probability for ≥ 2.54 cm (≥ 1 in) of liquid equivalent precipitation in 24 h were analyzed using the 1200 UTC 29 January through the 1200 UTC 2 February GEFS model guidance. The NWS defines probabilities $\geq 55\%$ as likely. Two and a half cm (1 in) of liquid equivalent precipitation would convert roughly into 25 cm (10 in) or more of snow, which justifies this type of an event to be mentioned in an HWO from the NWS.

The 50% probabilities shifted from the interior northeastern United States in the 1200 UTC 29

January guidance to southeastern coastal New England in the 1200 UTC 30 January guidance (Fig. 5)—a significant shift in 24 h. This was another clear signal that the greatest impacts from the storm were rapidly trending to the immediate coast, confirming the uncertainty in the storm track with each successive set of guidance. The 1200 UTC 31 January run of the GEFS showed that even the small chance (30%) probabilities were completely offshore (not shown). This low-impact scenario of only lesser amounts of snowfall confined mainly to coastal areas continued to be predicted in successive sets of guidance until the storm’s impact on 3–4 February.

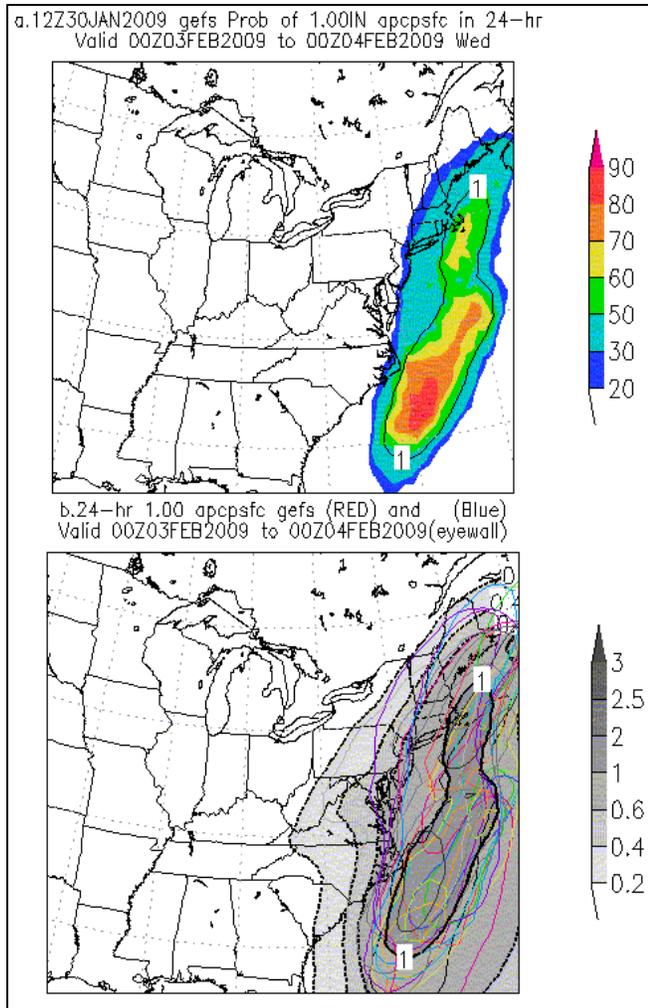


Figure 5. GEFS forecast of 1 in (2.54 cm) or more QPF (thick black contour) from forecasts initialized at 1200 UTC 30 January 2009, valid at 0000 UTC 4 February. Upper panel (a) shows probability (shaded) and the mean 1-in (2.54 cm) contour. Lower panel (b) shows the mean QPF for the period, the 1-in (2.54 cm) contour for individual ensemble members (colors), and the ensemble mean 1-in (2.54 cm) contour (thick black contour).

3) PREDICTABILITY OF THE SURFACE-BASED STORM TRACK

i. Operational deterministic and ensemble guidance

The center of the surface low pressure in the 1200 UTC 29 January ECMWF was over the interior northeastern United States (Fig. 6a), similar to that of the GFS and GEFS (not shown). Forecasts of MSLP from the ECMWF from successive runs (Figs. 6b, c, and d) showed a significant eastward progression of the forecasted surface low-pressure center from just inland of the coast to offshore, with significant spread also seen in the spaghetti plots in the GEFS forecasts

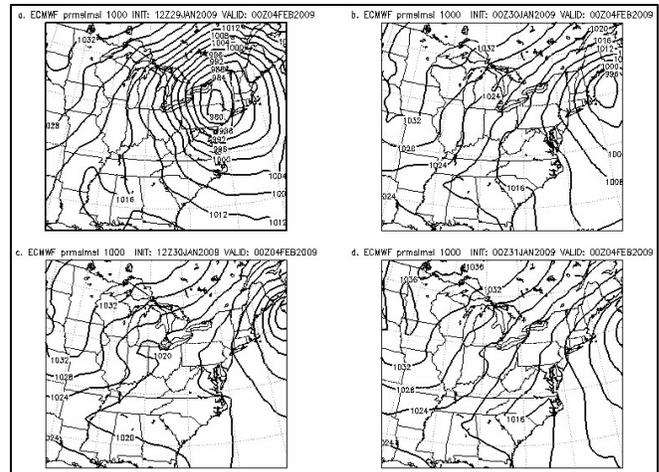


Figure 6. ECMWF MSLP forecast (hPa) valid at 0000 UTC 4 February 2009, initialized at a) 1200 UTC 29 January, b) 0000 UTC 30 January, c) 1200 UTC 30 January, and d) 0000 UTC 31 January.

(Figs. 7a and b). The GEFS depicted relatively weak low pressure, with MSLP anomalies at 1–2 SD below normal (Figs. 7c and d). The LAF from the GFS (Fig. 8) and ECMWF (Fig. 9) showed large spreads over the northeastern United States, reflecting the tendency of the surface low to be forecast farther east in each successive run. The valid times depicted in Figs. 6 and 7 are 12-h apart because each model resolved the storm differently, resulting in different timing of the storm development along the East Coast.

ii. Poor man’s ensemble

A “poor man’s ensemble” (PME; Bowler et al. 2008) of deterministic operational forecast models indicated high uncertainty for many aspects of the predicted cyclone. Figure 10 shows the PME mean MSLP forecast and spread of the ECMWF, GEM, and GFS initialized 1200 UTC 29 January and valid at 1200 UTC 3 February. Based on the position of the averaged surface low center in the PME, it is clear that a coastal storm was predicted by all three models. However, the track of the cyclone and its intensity varied considerably. The PME mean shows a broad cyclone center and a large spread (>6 hPa), indicating high uncertainty. The greatest uncertainty was northeast of the mean cyclone center, with a secondary maximum on the west side of the cyclone. The varied solutions that contributed to the large spread would indicate many possibilities in the predicted surface low track, resulting in significant differences in sensible weather—including temperatures, wind speed and direction, and precipitation type and amounts.

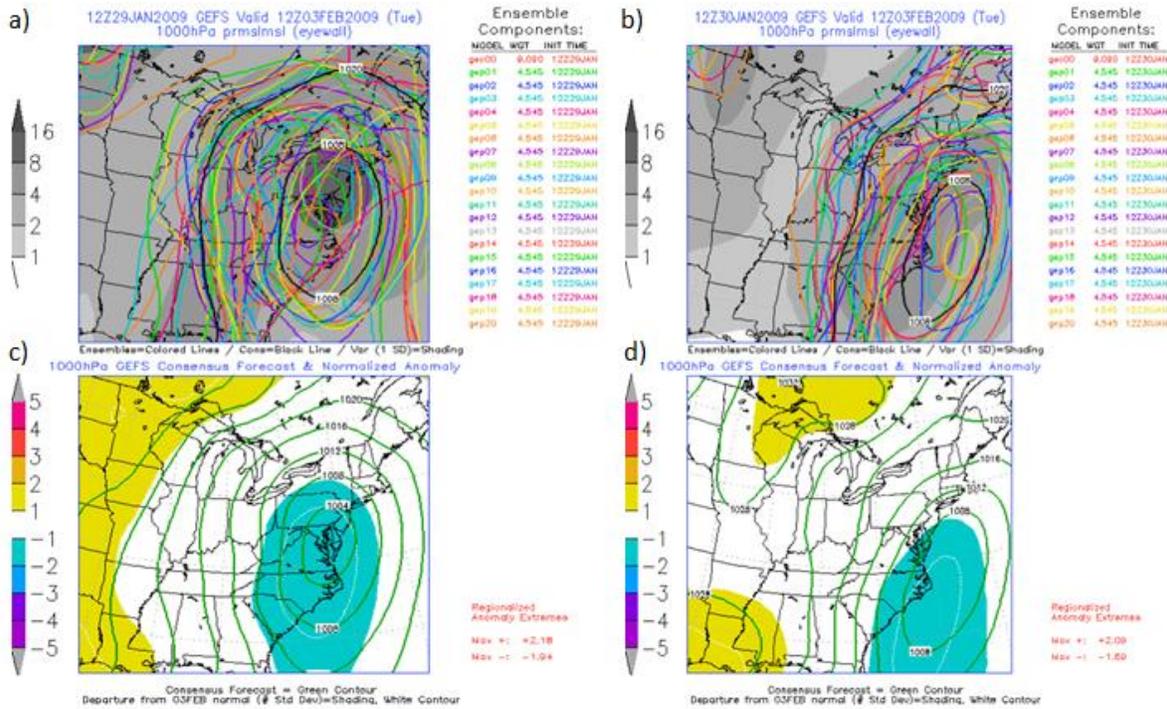


Figure 7. Spaghetti plot of MSLP (hPa) (upper panels) and anomalies (lower panels) valid 1200 UTC 3 February 2009 from the GEFS initialized on (left) 1200 UTC 29 January and (right) 1200 UTC 30 January.

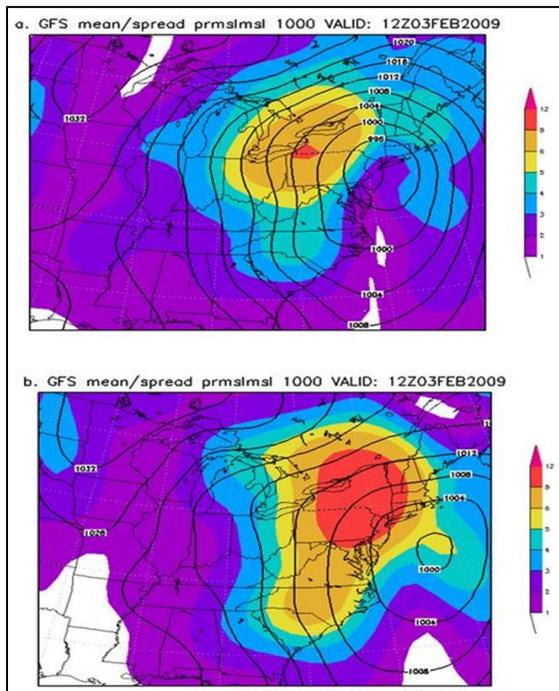


Figure 8. NCEP GFS LAF valid at 1200 UTC 3 February 2009 from forecasts initialized at a) 1200 and 1800 UTC 29 January and 0000 and 0600 UTC 30 January 2009 and b) 1200 UTC on 29, 30, and 31 January and 1 February 2009. Contours show the MSLP (hPa), and shading shows the spread about the mean.

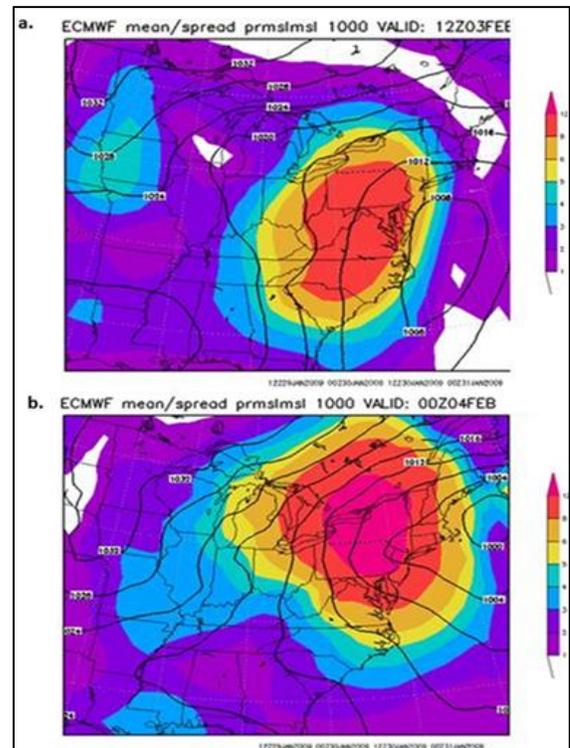


Figure 9. Same as in Fig. 8 except LAF from ECMWF forecasts valid at 1200 UTC 3 February and 0000 UTC 4 February 2009.

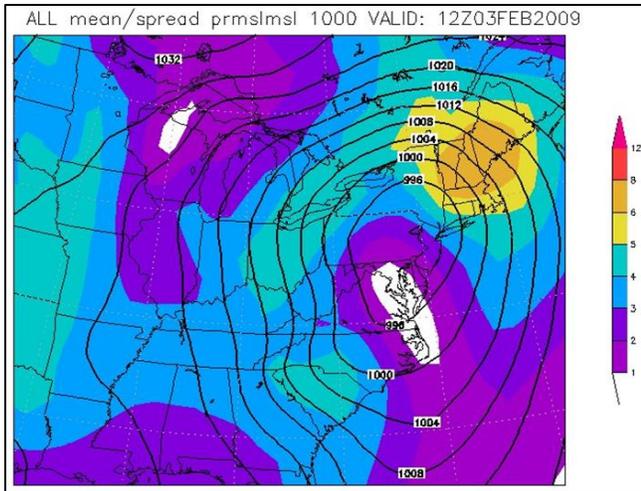


Figure 10. The poor man's ensemble (comprising the GFS, ECMWF, and GEM) initialized at 1200 UTC 29 January and valid 1200 UTC 3 February 2009. Isobars are in hPa, and shading shows the spread about the mean.

iii. Verifying cyclone

The verifying MSLP analysis showed a 1000-hPa cyclone well offshore by 1200 UTC 3 February (Fig. 11). Comparing the forecast and verifying intensity and location of the surface low center, the storm was not only weaker but farther offshore than the 1200 UTC 29 January guidance suggested (Fig. 6a).

It is evident through the analysis of the multiple sets of ensembles that predictability can change with time and can depend on how different sources of guidance at different resolutions depict different aspects of a storm system. Consequently, ensembles must increase in resolution to give us the information on uncertainty that is needed when mesoscale dynamics are more important.

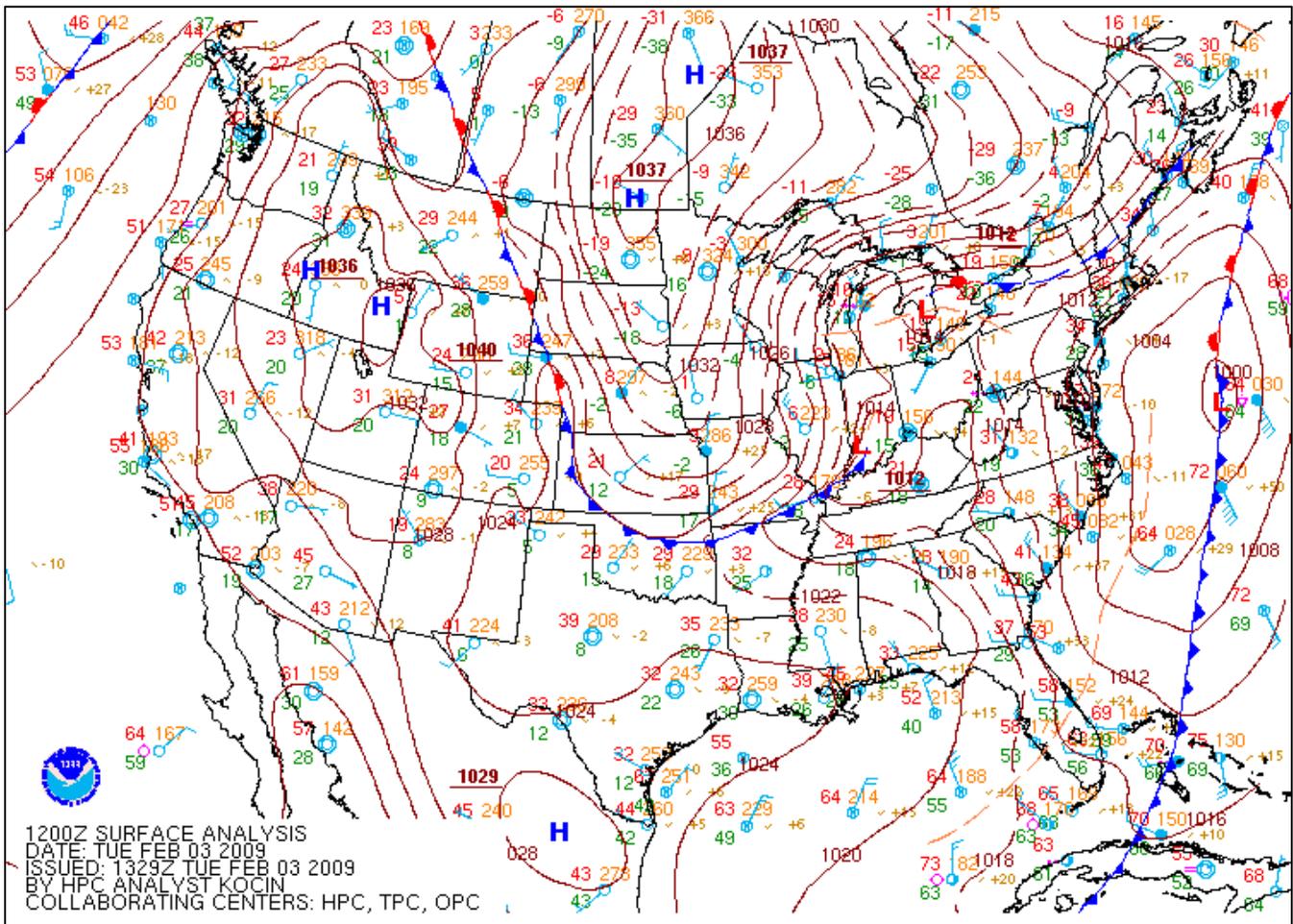


Figure 11. MSLP analysis (hPa) valid 1200 UTC 3 February 2009.

b. The March '09 storm

The degree of uncertainty associated with this storm was different than in the February '09 storm. The same caveats apply to this storm as in the February '09 storm in terms of the initializations and valid times varying in the 3–5-day forecasts. The spread in the storm track and thermal profiles depicted in the deterministic and ensemble guidance were less than those of the February '09 storm in the eastern United States. This higher level of confidence may have contributed to the superlative labels attributed to the March '09 storm more than three days prior to the onset. There was a high level of confidence that snow would be the dominant precipitation type over most of the eastern United States. However, there was considerably less certainty in the precipitation amounts.

1) DETERMINING UNCERTAINTY WITH THE 3-CYCLE LAF

The 3-cycle LAFs (comprising the 1200 UTC 25 February, 0000 UTC 26 February, and 1200 UTC 26 February cycles) are shown in Fig. 12. Both the GFS (Fig. 12a) and ECMWF (Fig. 12b) featured flip-flops in the 500-hPa height field over eastern Canada in the region north of the upper low. The ECMWF suggests more uncertainty in the area of cyclogenesis over South Carolina. Compared to the February '09 event, these values were less in magnitude and areal extent near the base of the 500-hPa trough. As the placement of the 500-hPa trough is a key factor in determining the surface low track, the 3-cycle LAFs implied more certainty in the storm track for the March '09 event (compared to February '09).

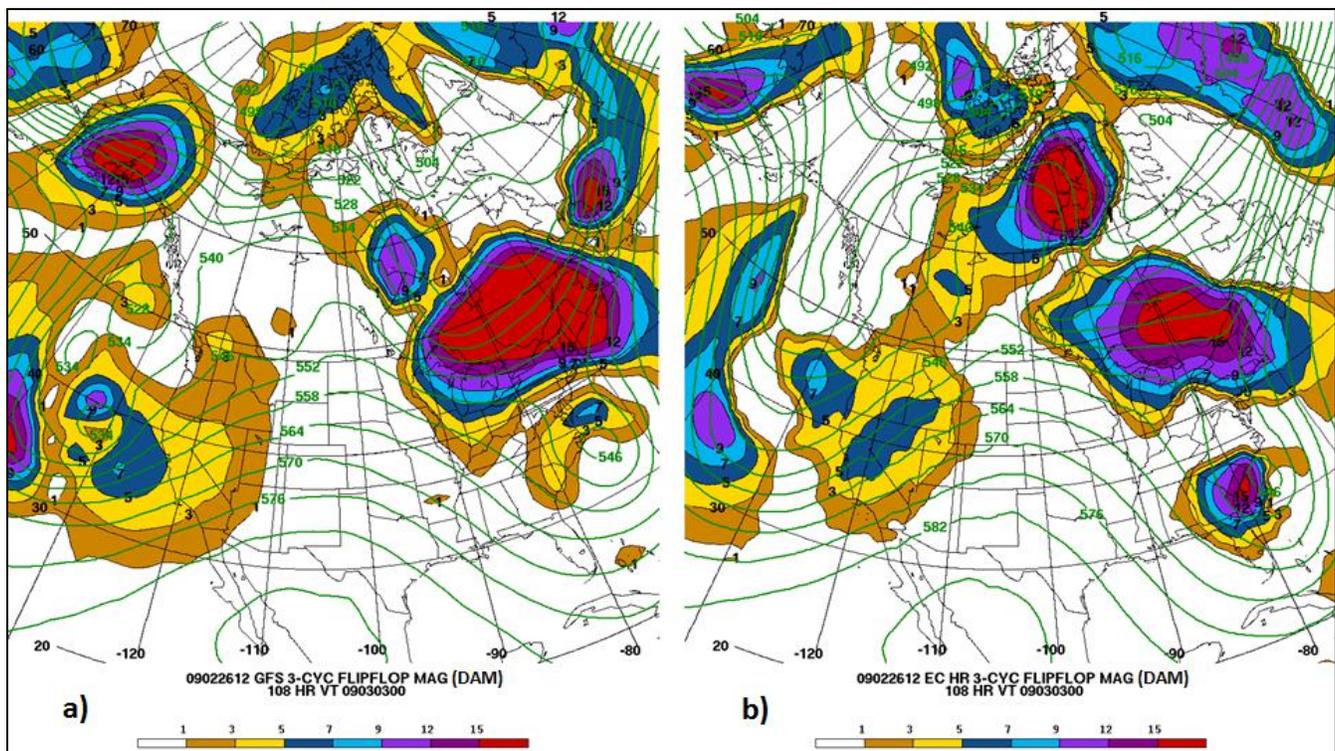


Figure 12. LAF 500-hPa heights (contoured, dam) and flip-flop magnitude (shaded, dam) comprising the 1200 UTC 25 February, 0000 UTC 26 February, and 1200 UTC 26 February 2009 runs of the a) GFS and b) ECMWF. Valid time is 0000 UTC 3 March 2009.

2) PREDICTABILITY OF PRECIPITATION TYPE AND AMOUNT

Figure 13 shows the GEFS 24-h probability of 1.27 cm (0.5 in) of liquid equivalent precipitation initialized at 1200 UTC 27 February and valid at 0000

UTC 2 March and 0000 UTC 3 March, respectively. This forecast was made approximately 120 h before the snow fell in the mid-Atlantic United States. It is clear that the area of likely or greater probabilities (>50%) was consistent in both forecast cycles along and east of the immediate eastern United States coast.

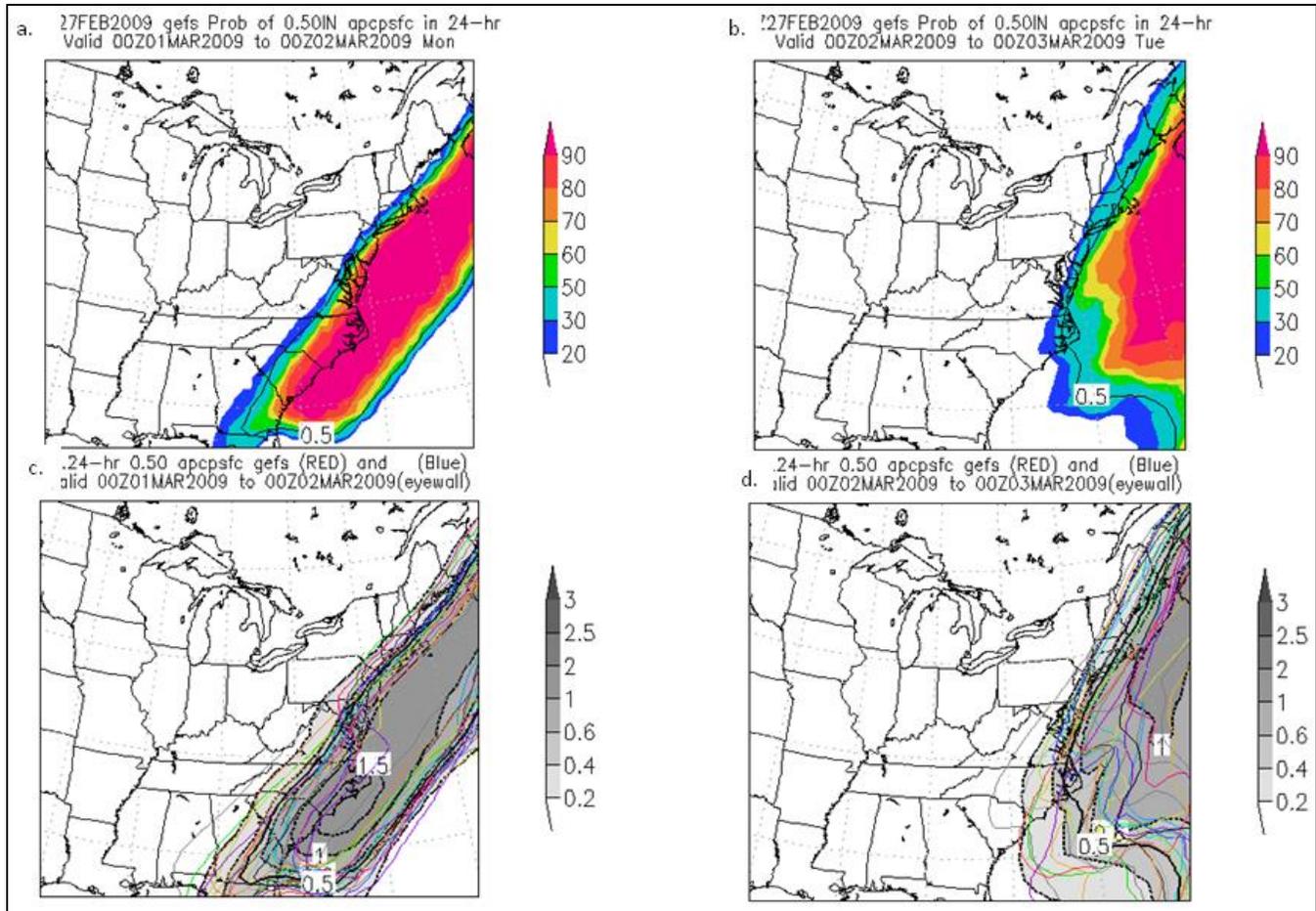


Figure 13. GEFS forecast of 0.5 in (1.27 cm) or more QPF from forecasts initialized at 1200 UTC 27 February 2009 and valid for the 12 hr periods ending at (left) 0000 UTC 2 March and (right) 0000 UTC 3 March. Upper panels (a and b) show probability (shaded) and the mean 0.5-in contour. Lower panels (c and d) show the mean QPF for the period, each member’s 0.5-in contour (colors), and the ensemble mean 0.5-in contour (thick black contour).

Precipitation probabilities from successive runs indicated an increasing threat of widespread heavy precipitation. The likely probabilities for 2.54 cm (1 in) of liquid equivalent precipitation extended through immediate coastal areas of the eastern United States through New England. However, there were differences in the position of the northwestward extent of the likely probabilities for 2.54 cm (1 in) of liquid equivalent precipitation, seen in a comparison of the 0000 UTC 1 March GEFS (Fig. 14) and the 0300 UTC 1 March SREF (Fig. 15). This contributed to uncertainties in determining whether the heaviest precipitation and highest impact of the storm would be confined to the immediate coast or if instead it would include the majority of the heavily populated eastern United States corridor. Regardless of the region of highest impact, the relatively modest ensemble snowfall predictions with maxima up to 2.54 cm (1 in)

of liquid equivalent, which translates up to 30 cm (1 ft) of snow depending on liquid-to-snow ratios, still did not support the label of “megastorm.” This is verified by the observed snowfall amounts (refer back to Fig. 2) and the ranking of 1 on the NESIS.

3) PREDICTABILITY OF THE SURFACE-BASED STORM TRACK

Deterministic forecast models and ensemble guidance for the storm track generally exhibited less spread prior to the March ‘09 storm compared to the February ‘09 storm. The GEFS MSLP forecasts initialized at 1200 UTC 26 February (Figs. 16a and c) and 1200 UTC 27 February (Figs. 16b and d) and valid at 0000 UTC 2 March both feature a weak 1004-hPa low off the United States East Coast. The SREF was similar (not shown). The GEFS and SREF showed run-to-run consistency and *little* spread in the spaghetti

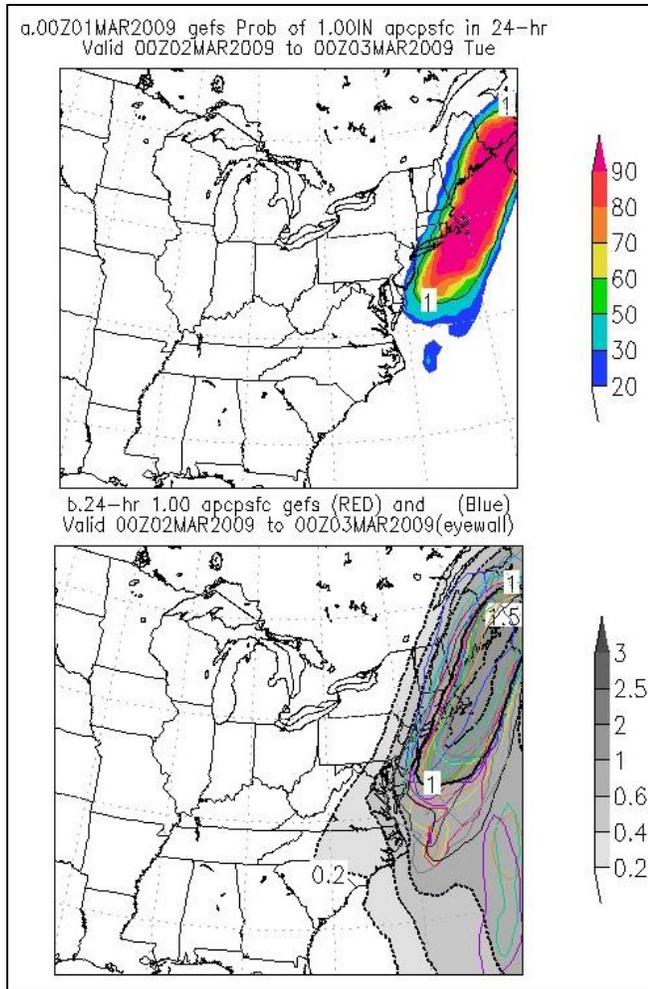


Figure 14. GEFS forecast of 1 in (2.54 cm) or more QPF from forecasts initialized at 0000 UTC 1 March 2009 and valid at 0000 UTC 3 March. Upper panel (a) shows probability (shaded) and the mean 1-in contour. Lower panel (b) shows the mean QPF for the period, the 1-in contour for individual ensemble members (colors), and the ensemble mean 1-in contour (thick black contour).

plots for the strength and track of the surface low-pressure center over the eastern United States. This is in contrast to the February ‘09 storm in which the guidance showed *more* spread and *less* run-to-run consistency.

4) VERIFYING CYCLONE

At 0000 UTC 1 March only a weak surface cyclone was present over Georgia. By 0000 UTC 2 March the cyclone showed a double structure over the Carolinas (Fig.17), which is what the NAM had been predicting. The high resolution of the NAM contributed to its ability to resolve this mesoscale structure in its predictions. However, the NAM is just

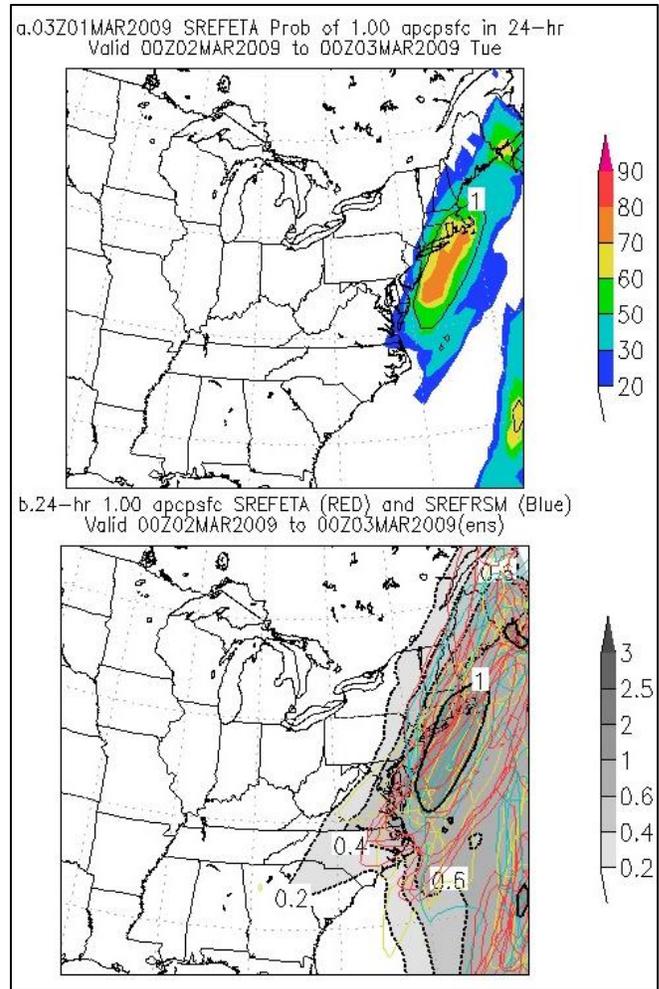


Figure 15. Same as Fig. 14 except SREF forecast of 1 in (2.54 cm) or more QPF from forecasts initialized at 0300 UTC 1 March 2009 and valid at 0000 UTC 3 March.

one model, and forecasters typically evaluate model skill subjectively based on verification of predictions, siding with guidance from the model perceived to be depicting the most skill. This may have reinforced the perception of high predictability within the forecast community, but it was just one model among multiple deterministic models and ensembles. Based on the over-predicted impacts, it can be argued that communicating a range of possibilities and/or a probabilistic approach could have aided the forecast community in more effectively communicating predictability of the associated hazards.

Between 1200 UTC 2 March and 0000 UTC 3 March, the double structure evolved into a modest surface cyclone that was present off the coast of New Jersey and accelerated to the east of Maine. The event was clearly in its last stages over New England.

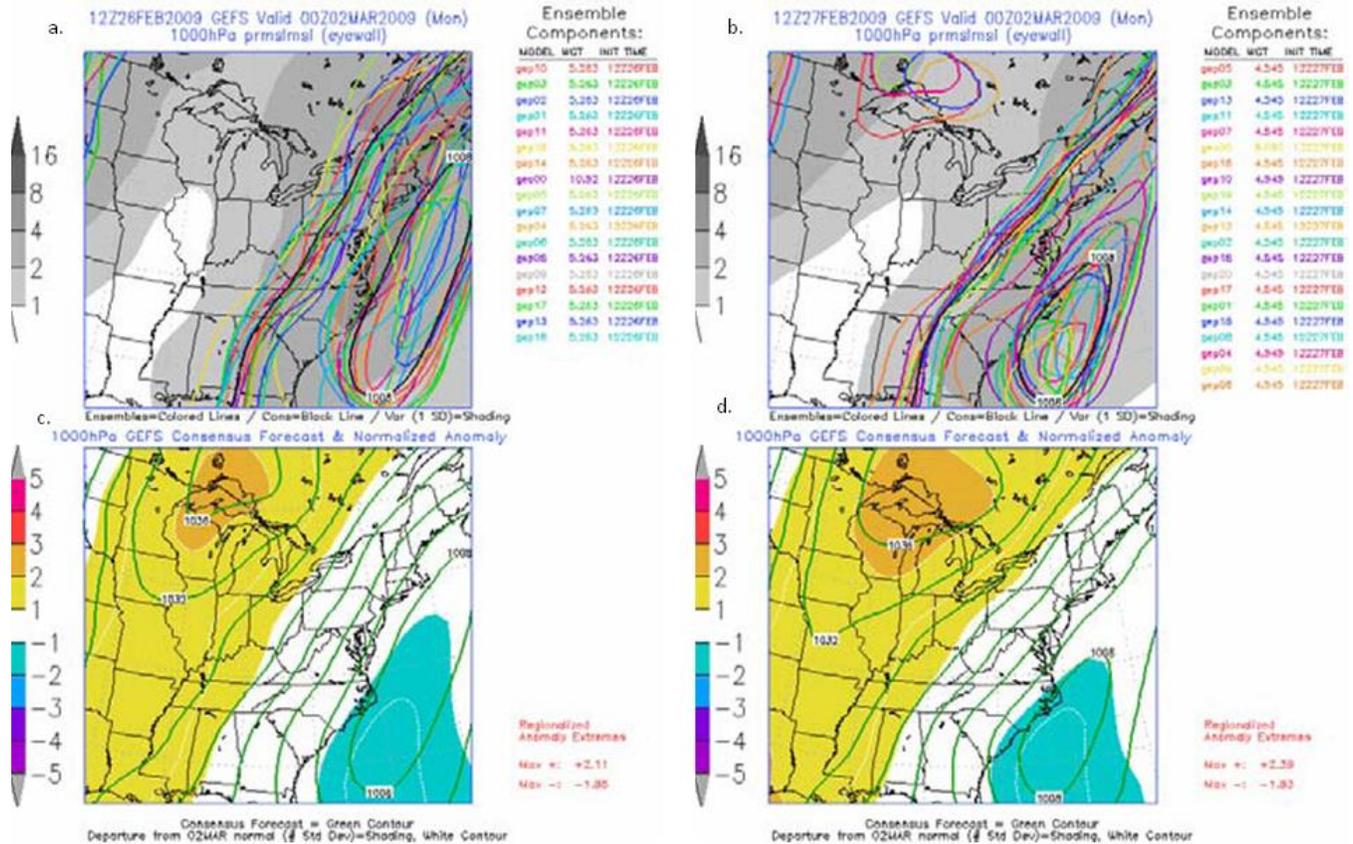


Figure 16. Spaghetti plot of MSLP (hPa) (upper panels) and anomalies (lower panels) valid 0000 UTC 2 March 2009 from the GEFS initialized on (left) 1200 UTC 26 February and (right) 1200 UTC 27 February.

4. Conclusions

The February ‘09 and March ‘09 snow events highlighted the need for forecasters to understand predictability and convey uncertainty information more effectively. Weather forecast information from weather information sources provided to the public for these two winter storms reflected relatively high confidence in the potential for significant events. It should be stated that not all forecasters in the media and the NWS exhibited the high-confidence forecasts (e.g., evidenced in part by the use of terms “likely,” “possible,” “could,” and “may” in some NWS HWOs), suggesting varying levels of understanding of predictability within the forecast community. As the impact of each storm got closer in time, the various sources of information modified the expectations downward but at different rates depending on the sources of information and their decision on when there was enough confidence in significantly changing the forecast. The conflicting information from weather forecast providers during these two events is evidence that a problem of misunderstanding predictability

exists in the forecasting community and needs to be resolved (Novak et al. 2008).

Predictability of various aspects of these storms changed as the potential onset of the storm drew nearer. This commonly occurs when the atmospheric features that later become responsible for East Coast storms become better resolved in the data network when entering North America from the relatively data-sparse Pacific Ocean. Consequently, the pertinent atmospheric features are better resolved in NWP model and ensemble member initializations. Even at time ranges <12 h, uncertainty arises owing to the chaotic nature of the atmosphere and model imperfections. However, the range of forecast possibilities is much narrower, resulting in much better guidance and forecast accuracy at shorter ranges.

Uncertainty information is readily obtained using LAFs, PME techniques, and specific EPSs. Data provided by these methods illustrate how each modeling system is sensitive to small changes in initial conditions, an important consideration highlighted by Kleist and Morgan (2005) in their study of the 24–25 January 2000 surprise East Coast snowstorm. Better

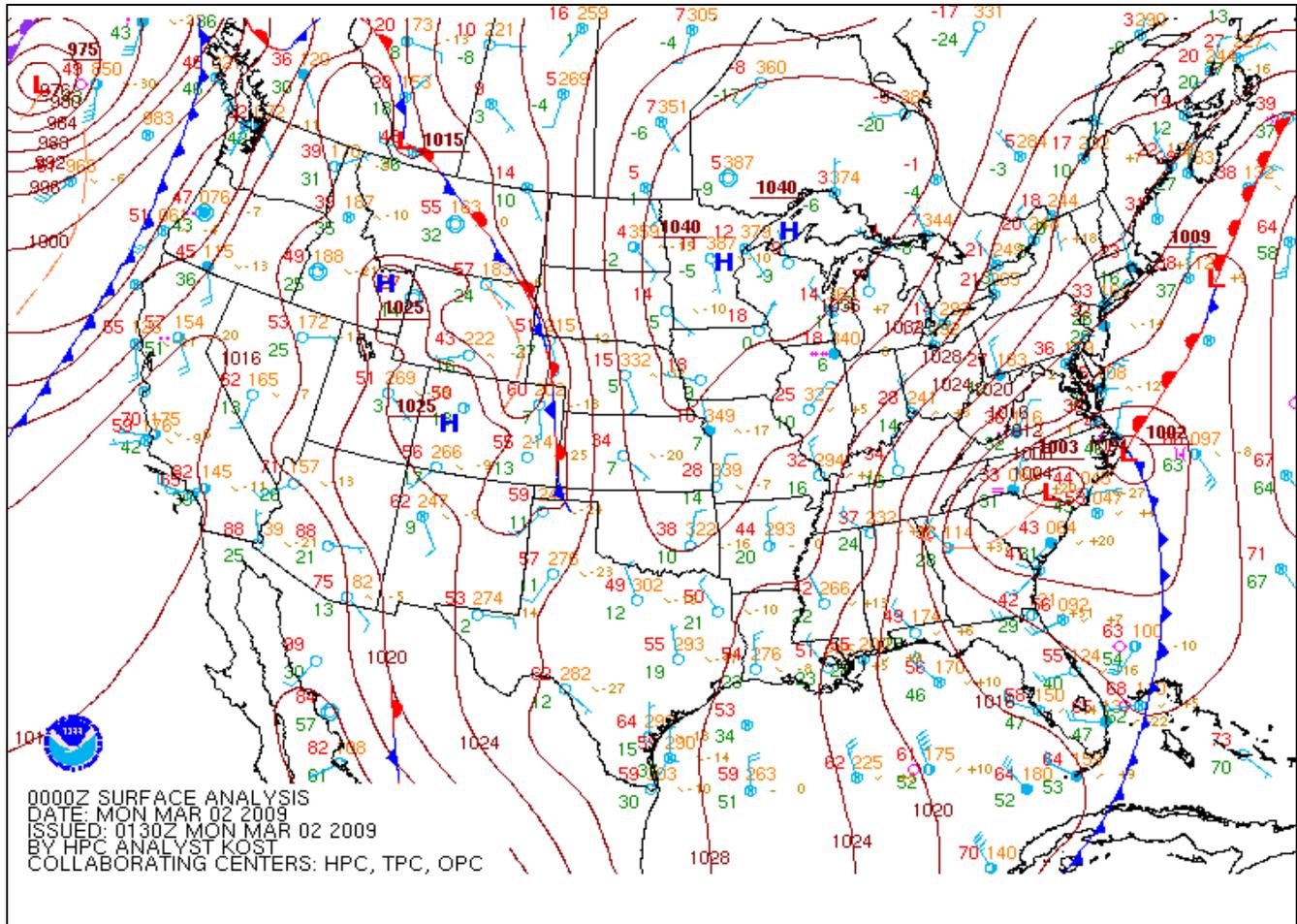


Figure 17. MSLP analysis (hPa) valid 0000 UTC 2 March 2009.

methods of ensemble generation to express forecast uncertainty will continue to be developed.

The ensemble mean should not be used as the best forecast and should always be used in conjunction with the spread. One particular ensemble mean solution can result from a wide variety of spreads depending on the magnitudes of the extremes, yet the range of possible outcomes is greater when the spread is greater. The uncertainty information is the critical part in forecasting, and using the mean without the spread loses this critical component. The large spread in both cases, but especially the February '09 case, indicated a high degree of uncertainty and thus should have resulted in less forecaster confidence.

Conveying uncertainty accurately by the forecasting community also is important when considering the implications on the user community and how they perceive the forecast information we provide. The increasing number of information sources available to society leads to easier choosing of

information sources that satisfy their own preferences, beliefs, and understanding of issues (Manjoo 2008)—including hyped or exaggerated weather information. Therefore, it is imperative that forecasters constantly improve their skill in extracting information from EPS data to become the authority among the variety of media sources broadcasting weather information.

These two storms raise questions relative to individual forecast biases, human emotion in the forecast process, and ineffective use of uncertainty and predictability information. It is likely that all factors played a role in these events, which not only affects how forecasters communicate information, but how the public understands and responds to the information. Uncertainty is and always will be both an essential and non-negotiable part of the forecast (Silver 2012, 407). Forecasters failing to deal with this uncertainty in weather forecasting are avoiding the complex issues and may be over-simplifying the process. More recent predictability challenges, such as

the 26–27 December 2010 (Kocin et al. 2011) and 8–9 March 2013 snowstorms in the northeastern United States, highlighted the significant differences in predicted-versus-observed weather from the Baltimore/Washington D.C. region (e.g., Freedman 2013; Samenow 2013) through New England (e.g., Sampson 2013). These recent examples illustrate the ongoing need to determine methods to improve and communicate predictability.

Acknowledgments. This study was inspired by the (i) numerous difficult winter storm forecasts the authors have forecasted in their 20-y or greater careers in the public, private, and armed forces sectors and (ii) humbling lessons learned from their numerous forecast successes and failures. The authors thank Dr. Martin Baxter of Central Michigan University for his thoughtful reviews and the snowfall graphics. The authors also thank the Scientific Services Division of the NWS Eastern Region Headquarters, Dr. Paul Croft of Kean University, and numerous anonymous reviewers for their assistance. The views, opinions, and findings contained in this study are those of the authors, not their employer, the National Weather Service.

REFERENCES

- Blake, E. S., T. B. Kimberlain, R. J. Berg, J. P. Cangialosi, and J. L. Bevin II, 2013: Tropical cyclone report: Hurricane Sandy (AL182012), 22–29 October 2012. National Hurricane Center Tropical Cyclone Report, 157 pp. [Available online at www.nhc.noaa.gov/data/tcr/AL182012_Sandy.pdf.]
- Bowler, N. E., A. Arribas, A., and K. R. Mylne: 2008: The benefits of multianalysis and poor man's ensembles. *Mon. Wea. Rev.*, **136**, 4113–4129.
- Freedman, A., cited 2013: D.C. Snow no-show a lesson in forecasting uncertainty. *Climate Central*, 7 March 2013. [Available online at cstar.cestm.albany.edu/PostMortems/CSTARPostMortems/2013/NWApaperfiles/ClimateCentralarticle.htm.]
- Glickman, T. S., 2000: *Glossary of Meteorology*. American Meteorological Society, 855 pp.
- Grumm, R. H., and R. Hart, 2001: Standardized anomalies applied to significant cold season weather events: Preliminary findings. *Wea. Forecasting*, **16**, 736–754.
- _____, and N. Stuart, 2007: Ensemble predictions of the 2007 Valentine's Day winter storm. Preprints, *22nd Conf. on Weather Analysis and Forecasting*, Park City, UT. Amer. Meteor. Soc., P2.49. [Available online at ams.confex.com/ams/pdfpapers/123823.pdf.]
- Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus A*, **35A**, 100–118.
- Kleist, D. T., and M. C. Morgan, 2005: Application of adjoint-derived forecast sensitivities to the 24–25 January 2000 U.S. East Coast snowstorm. *Mon. Wea. Rev.*, **133**, 3148–3175.
- 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.
- _____, P. N. Schumacher, R. F. Morales Jr., and L. W. Uccellini, 1995: Overview of the 12–14 March 1993 superstorm. *Bull. Amer. Meteor. Soc.*, **76**, 165–182.
- _____, L. W. Uccellini, J. Alpert, B. Ballish, D. Bright, R. Grumm, and G. Manikin, 2011: The blizzard of 25–27 December 2010: Forecast assessment. Hydrometeorological Prediction Center Report, 57 pp. [Available online at www.hpc.ncep.noaa.gov/winter_storm_summaries/event_reviews/2010/December25_27_2010_Blizzard.pdf.]
- Manjoo F., 2008: *True Enough: Learning to Live in a Post-Fact Society*. Wiley and Sons, 256 pp.
- NWS, 2004: Service Assessment: Hurricane Isabel, September 18–19, 2003. National Weather Service, 63 pp. [Available online at www.nws.noaa.gov/os/assessments/pdfs/isabel.pdf.]
- _____, 2012: Service Assessment: Hurricane Irene, August 21–30, 2011. National Weather Service, 129 pp. [Available online at www.nws.noaa.gov/om/assessments/pdfs/Irene2012.pdf.]
- Novak, D. R., D. R. Bright, and M. J. Brennan, 2008: Operational forecaster uncertainty needs and future roles. *Wea. Forecasting*, **23**, 1069–1084.
- Samenow, J., cited 2013: Snowquester: when forecast information fails. *The Washington Post*, 7 March 2013. [Available online at cstar.cestm.albany.edu/PostMortems/CSTARPostMortems/2013/NWApaperfiles/CapWxGang.htm.]
- Sampson, Z. T., cited 2013: 'Unusual' snowstorm dumped unexpected snowfall. *The Boston Globe*, 9 March 2013. [Available online at cstar.cestm.albany.edu/PostMortems/CSTARPostMortems/2013/NWApaperfiles/TheBostonGlobe_N.mht.]
- Silver, N., 2012: *The Signal and the Noise: Why So Many Predictions Fail—but Some Don't*. The Penguin Group, 544 pp.
- Sivillo, J. K., J. E. Ahlquist, and Z. Toth, 1997: An ensemble forecasting primer. *Wea. Forecasting*, **12**, 809–818.
- Tracton, M. S., 2008: Must surprise snowstorms be a surprise?. *Meteorological Monographs*, **33**, 251–268.
- Zhu, Y., 2005: Ensemble forecast: A new approach to uncertainty and predictability. *Adv. Atmos. Sci.*, **22**, 781–788.
- 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.