EVALUATION OF THE NATIONAL SEVERE STORMS LABORATORY MESOSCALE ENSEMBLE

Brianna M. Lund

National Weather Center Research Experience for Undergraduates Program, Norman, Oklahoma, St. Cloud State University Department of Atmospheric and Hydrological Sciences, St. Cloud, Minnesota

Dustan M. Wheatley

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, and NOAA/National Severe Storms Laboratory, Norman, Oklahoma

Kent H. Knopfmeier

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, and NOAA/National Severe Storms Laboratory, Norman, Oklahoma

ABSTRACT

Accurate short-term forecasts are critical for forecasters when anticipating severe weather events and improving such forecasts has long been a focus for meteorologists. The recent emergence of ensemble based data-assimilation systems has proven to be a promising step toward the improvement of these vital forecasts. The National Severe Storms Laboratory Mesoscale Ensemble (NME) is a 36-member ensemble that provides hourly forecasts and analyses of a variety of products used for severe weather forecasting, such as soundings and 2-m temperature fields. This project seeks to quantitatively evaluate said products through comparison to observations from a number of sources (surface stations, rawinsondes, etc.), including the Oklahoma and Texas mesonets.

1. INTRODUCTION

A detailed representation of mesoscale environments is key for accurately forecasting the potential threat of severe weather. More specifically, precise, up-to-date forecasts and analyses for these environments on time scales as little as an hour (if not smaller) are important for forecasters in order to get warnings out to the public in a timely fashion for them to take appropriate precautions. The improvement of such forecasts has long been an important goal of the meteorological community.

Operational numerical weather prediction (NWP) models as well as mesoscale observations are the basis for severe weather forecasting. However, they currently tend to lack the ability to give detailed information about critical mesoscale features such as drylines and cold pools (Stensrud et al. 1999). The recent emergence of ensembles and ensemble-based data-assimilation systems, though, has proven to be a promising step toward filling in the data gaps about these features. Ensembles have been shown to improve forecasts in relation to single deterministic models, such as the Rapid Refresh (RAP) model or the North American Mesoscale (NAM) model. One example of such improvement is a more accurate location and intensity of the dryline, which is often crucial to storm initiation (Fujita et al. 2007). Furthermore, Wheatley et al. 2012 demonstrates that incorporating data assimilation with ensembles can yield even greater improvement, such as with the significant tornado parameter (STP). The National Severe Storms Laboratory Mesoscale Ensemble (NME; NOAA 2013) is a recently developed, ensemble-based data-assimilation system that builds upon this previous work, and its performance is evaluated alongside the Rapid Refresh (RAP) model and surface observations.

The NME is a Weather Research and Forecasting (WRF) model-based, 36-member ensemble whose initial and boundary conditions are derived from the 12Z RAP forecast cycle. It

---

Corresponding author address: Brianna M. Lund, St. Cloud State University, 720 4th Ave S. St. Cloud, MN 56301. Email: lubr0903@stcloudstate.edu.

uses the WRF variational data-assimilation (WRF-Var) software to generate random samples of error to create the slightly different initial conditions for each member in order to account for uncertainties in those conditions. Member diversity is maintained through the use of differing combinations of physical parameterization schemes, including cumulus, planetary boundary layer, and radiation. For the assimilation process, an ensemble Kalman filter is used for its capability to produce realistic mesoscale environments (Wheatley et al. 2012).

The primary objective of this study is to evaluate the NME for its ability to reproduce pre- and near-storm environments. More specifically, the project seeks to quantitatively evaluate the severe weather forecasting products produced by the NME using special soundings launched on severe days, as well as observations from the Oklahoma/Texas mesonets and METAR sites.

Section 2 describes how the experiment is implemented for the ensemble and evaluation methods used. Section 3 includes the results found from the soundings and surface products of 2-m temperature and specific humidity. Lastly, section 4 consists of overall conclusions drawn from the experiment and provides possible options for future work.

2. METHODS

The NME ran daily during the 2013 Hazardous Weather Testbed Spring Experiment from May 6th to June 10th; however, the focus of this project is centered on the last two weeks, May 27th to June 10th, because the RAP had problems with its land surface model during the first two weeks. The process begins at 12Z each day using the 12Z RAP analysis to construct the model's initial conditions, and then a 1-h forecast is made from each ensemble member. From there, all of the member forecasts are averaged to form an ensemble-mean forecast. Then, routinely available observations from land and marine surface stations, rawinsondes, aircraft and satellite are assimilated using the following variables: temperature, dew point, pressure, and horizontal wind components. Each ensemble member is updated and a new 1-h forecast is made. This process continued with assimilation then forecast for each hour at the top of the hour from 14Z to 03Z into the next day for the entirety of the two-week period.

Upon completion of the experiment, the upper air products are analyzed. First, the actual sounding images are created from the RAP, NME individual ensemble members, NME-mean profiles from the 1-h forecast and analysis, and observed soundings to get a qualitative sense of how the NME behaves. An example of which is shown in Figure 1.

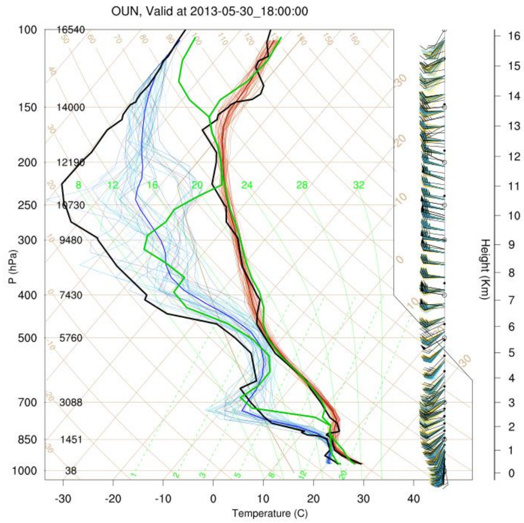


Figure 1. 1-h forecast sounding diagram valid at 1800 UTC 30 May 2013. The black lines are the observed sounding from KOUN. The green lines are the RAP sounding from the model gridpoint closest to KOUN. The thin red and blue lines are the NME individual member soundings, and the bold red and blue lines are the NME-mean sounding also from the model gridpoint closest to KOUN.

Next, 1800 UTC soundings are quantitatively analyzed, but only on days that had at least two soundings launched at distinct locations. Days and times with multiple soundings were chosen in order to gather a more informative representation of the mesoscale environment. The following three days at 1800 UTC are examined: May 29th for Amarillo and Fort Worth, Texas, Dodge City, Kansas, North Platte, Nebraska, and Norman, Oklahoma, May 30th for Davenport, Iowa, Norman, Oklahoma, and Springfield, Missouri, and May 31st for Norman, Oklahoma and Springfield, Missouri. Vertical error profiles for winds, temperature and relative humidity for these three days are created for the RAP and NME to compare them relative to one another and to the observed soundings at the mandatory levels (925, 850, 700, 500, 400, 300, 250, 200 hPa). Model data is extracted from both of the models at the same date, time and geographical location as the observed soundings. A simple absolute error calculation is made at each of the mandatory levels by averaging all values at each level across the spatial domain and temporal two-week period.

Finally, to acquire a more general sense of the performance of the NME, vertical error profiles are also produced at 0000 UTC from the entire two-week period for 47 upper-air stations located in the eastern two thirds of the United States.

For the surface products, radar scans and color-filled 2-m temperature plots from the 14Z to 03Z time frame of the two-week period are examined in order to find a few cases that are representative of storm-induced cold pools. The criterion is to look for local minima in temperature near the surface that are convectively induced. A particular hour and, in a few cases, multiple hours are selected from each of the following event days: May 28th at 02Z, May 30th at 03Z, 15Z and 23Z, May 31st at 02Z and 15Z, June 8th at 23Z, June 9th at 00Z, 02Z and 20Z. A four-paneled image is created from the NME and RAP 1-hour forecast and analysis of 2-m temperatures at the aforementioned dates and times with the observations overlaid. The forecast and analysis figures generated from the NME are qualitatively analyzed and compared relative to the RAP and the observations to see how well the NME determined the position and strength of the storm-induced cold pools. For quantitative analyses of these figures, a root mean square difference (RMSD) is calculated for the selected dates and times.

Similarly, specific humidity fields are plotted in much the same way as the 2-m temperature fields are in order to examine how the NME produced the location of the dryline. Specific humidity is chosen as the variable to analyze the dryline because it is less sensitive to changes in elevation (Coffer et al. 2013). The 2100 UTC and the 0000 UTC hours for the 28th of May to the 10th of June are examined, but the focus is on the earliest five days due to the most pronounced presence of a dryline and its role in convective initiation on those days. After June 1st at 21Z, the dryline does not act as a focus for convective initiation.

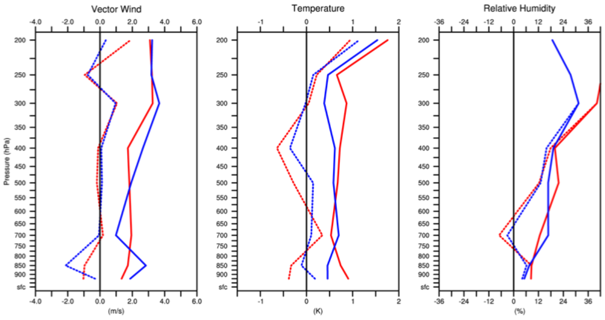
3. RESULTS

a. Soundings (Upper Air)

1) 1800 UTC Soundings

There are several key features indicated in the 1800 UTC error profiles (Fig. 2). The NME errors are smaller by about 0.5-1 °C for both forecast and analysis temperature profiles. The relative humidity error profiles show the largest difference in performance. Both models behave similarly below 400 hPa in the forecast, but the NME has much smaller errors by about 18% above that level. Throughout the entirety of the analysis profile, the NME relative humidity error is smaller, on average, by about 8-10%. Although both models have virtually identical biases for the forecast vector winds from 700-250 hPa, the RAP bias is closest to zero from 900-700 hPa and 250-200 hPa. Vector wind errors, are similar for the forecast and analysis, but the NME bias is closer to zero. The RAP had about a 1 °C cold bias.

PRIOR:



POSTERIOR:

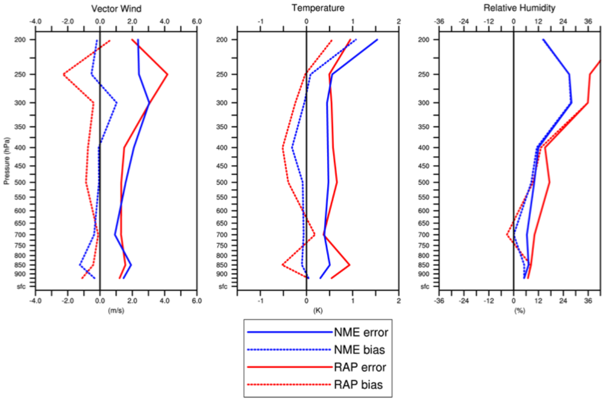
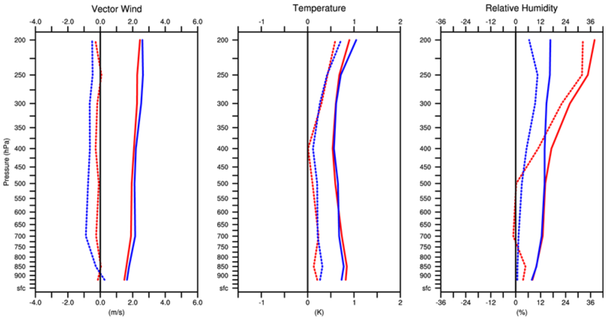


Figure 2. Spatially and temporally averaged error profile for 1800 UTC soundings. Forecast wind, temperature and relative humidity profiles are on top and analysis profiles are on the bottom. Blue lines correspond to the NME and the red lines correspond to the RAP. The dashed lines are the bias and the solid lines are the absolute error.

2) 0000 UTC Soundings

In this case, Figure 3 shows that the RAP has smaller errors for the forecast vector wind error profile by as much as 0.2 m s^{-1} and roughly 0.25 m s^{-1} in the analysis. The differences in error and bias for both forecast and analysis between the NME and RAP are still relatively small, although the NME has a low bias of roughly 1 m s^{-1}. Likewise, the temperature errors are noticeably similar, although, the RAP has a slightly lower bias in the analysis. The forecast profile does show that both models have a small degree of warm bias, but the analysis shows the RAP bias was closer to zero and the NME was still a little warm. Like the 1800 UTC soundings, relative humidity errors are virtually identical below 500 hPa, but above, the NME has a much smaller error and bias for both forecast and analysis profiles. The NME was consistently moist by about 6%, but the RAP bias became increasingly too moist above 500 hPa at most by 30% at the top of the profile.

PRIOR:



POSTERIOR:

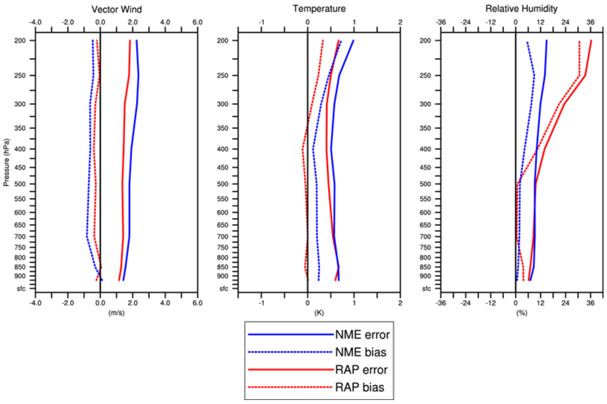


Figure 3. Spatially and temporally averaged error profile for 0000 UTC soundings. Forecast wind, temperature and relative humidity profiles are on top and analysis profiles are on the bottom. Blue lines correspond to the NME and the red lines correspond to the RAP. The dashed lines are the bias and the solid lines are the absolute error.

b. 2-m Temperature Fields (Cold Pools)

1) 30 May 2013 (0300 UTC)

For this event, the cold pool is found straddling the Kansas/Oklahoma border as indicated by the red circle in Figure 4. The RAP produces generalized cooling in the analysis, which occasionally results in overcooling. An example of this can be found in northwest Oklahoma where the RAP was 1-2 °C cooler than the observations. On the other hand, the NME also produces cooling in the analysis, but tried to localize changes around the observations more. This sometimes results in a slight warm bias of 1-2 °C, such as in southeast Kansas.

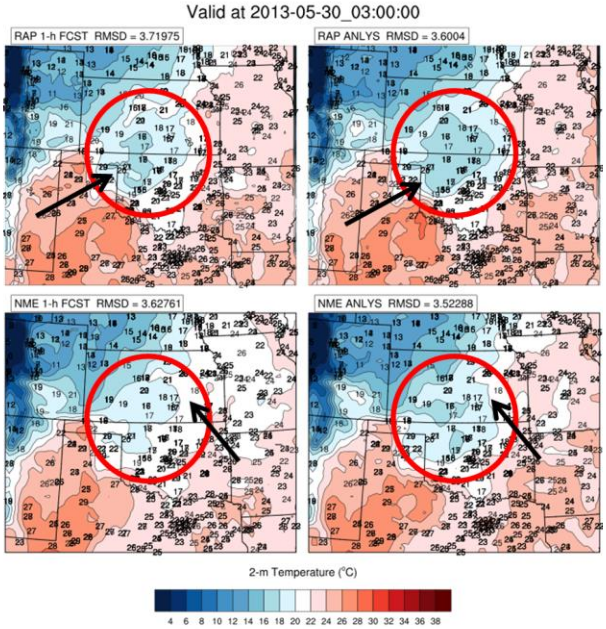


Figure 4. 2-m temperature (°C; see label bar) fields for the RAP and NME 1-hr forecast and analysis valid at 0300 UTC on 30 May 2013. The red circle indicates the location of the cold pool and the arrows point to the regions where the models had cold biases (the RAP) and warm biases

2) 9 June 2013 (0200 UTC)

For this event, the cold pool is found in central Kansas as indicated by the red circle in Figure 5. As was the case with most of these plots, the RAP 1-h forecast produces overcooling. This is evident in northeastern Kansas where the model 2-m temperature is in the 16-18 °C range, but a number of observations were 19 °C. This occurs again in central Kansas and just east of the Oklahoma panhandle in which the RAP captures a few observations reporting 20 °C in its 18-20 °C contour. The cool bias that is shown in the NME and RAP 1-h forecasts continues to be present in the analyses. The NME does not have as large of a cold bias as the RAP, but it also cools further in the analysis.

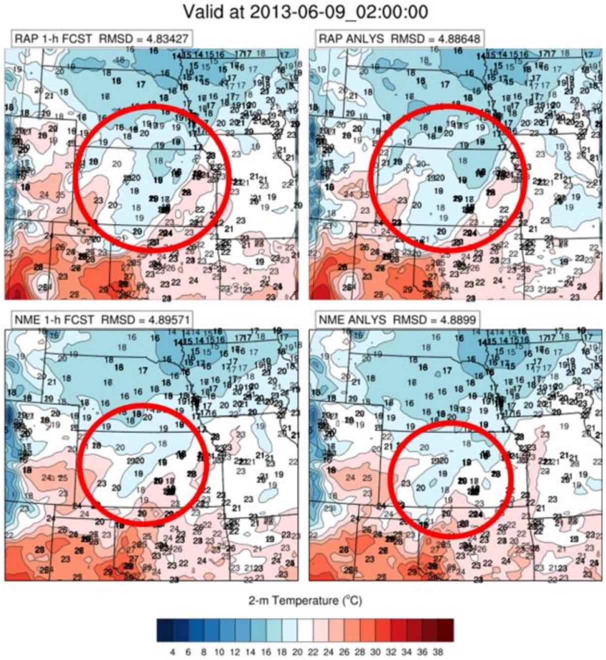


Figure 5. Same as in Fig. 4 but valid at 0200 UTC on 9 June 2013. The red circle indicates the location of the cold pool.

c. Specific Humidity Fields (Dryline)

1) 28 May 2013 (0000 UTC)

Both models position the dryline very similarly and accurately through northwest Oklahoma into the Texas panhandle (Fig. 6). The RAP tends to have high-amplitude localized pockets that are too moist. This is most visible in southwest Oklahoma, where the highest observed value is 15g kg^{-1}, but the model produces values in the 18-20 g kg^{-1} range. The NME, though absent of the high amplitudes, tended to be too moist by 2-3 g kg^{-1} over a larger region. Both models, though, are too moist overall across the central Kansas/Oklahoma border.

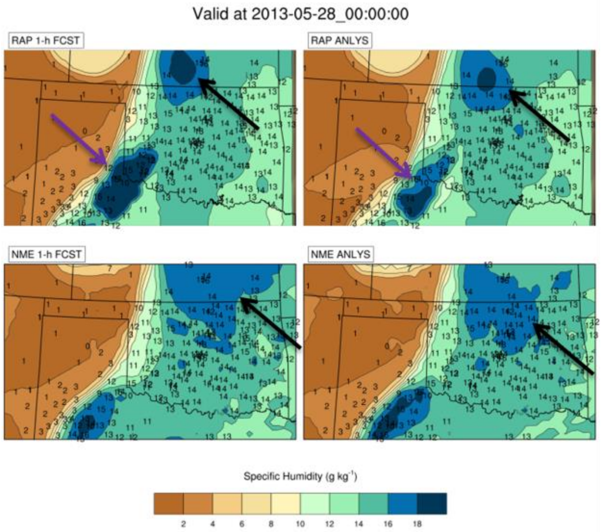


Figure 6. Specific humidity (g kg^{-1}; see label bar) fields for the RAP and NME 1-h forecast and analysis valid at 0000 UTC 28 May 2013. The purple arrows point to the region of values of moisture that are too high from the RAP and the black arrows point to regions of high moisture values that both models produced.

2) 1 June 2013 (0000 UTC)

Figure 7 shows that both models perform similarly to the May 28th case in their ability to correctly simulate the dryline strength/position. However, the NME tends to be too moist by about 1-2 g kg^{-1} in much of Oklahoma ahead of the dryline. On the other hand, the NME 1-h forecast and analysis shows a high degree of low-level moistening associated with light precipitation in SW Kansas in both forecast and analysis that is not shown in the RAP. The RAP does have the same feature in the forecast as indicated by the red circle in Figure 7, but smoothes it out in the analysis. Due to the lack of observations in the area, it is difficult to say which model is correct for this feature.

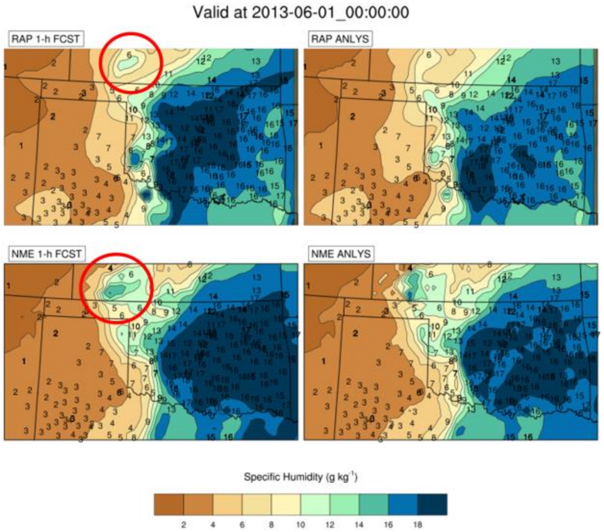


Figure 7. Same as in Fig. 6, but valid at 0000 UTC 1 June 2013. The red circles highlight the low-level moistening due to light precipitation.

For the same event, the features highlighted by the purple circles in Figure 8 that the RAP generated are worth noting. It is concluded that the model is likely generating convection where those small pockets of moisture are located, but since the reflectivity indicates no such feature occurring, the model attempts to smooth out those pockets in the analysis. In general, the RAP tends to produce too much convection.

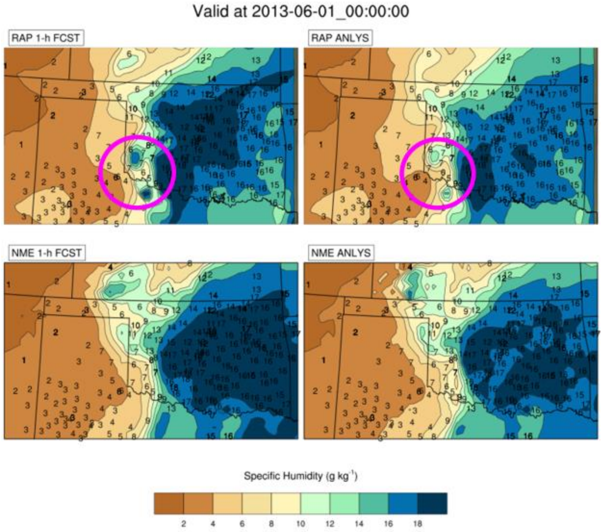


Figure 8. Same as in Fig. 7. The purple circles highlight high moisture pockets that the RAP produces.

4. CONCLUSIONS

General findings from this experiment are that, except for the wind, the NME errors and biases for temperature and moisture in both the 1800 UTC forecast and analysis sounding error profiles are smaller. However, later at 0000 UTC, the differences in error between the RAP and NME are very small. The errors appear to converge in the late afternoon/early evening hours. Cold pool analysis also shows no real discernible differences. Both models are fairly accurate in reproducing cold pool strength (i.e. temperature characteristics) and location. However, the RAP tends to produce too much cooling over too large an area and the NME does not cool enough.

A similar result is also found in regards to dryline location, which the RAP and NME both position accurately. Besides the differences in how the RAP and NME produces certain features in the moisture fields, the position of the dryline is similar and quite accurate. Though they perform comparably to one another, an important note is that the NME did so with fewer computational resources than the RAP.

A consideration for future work is to expand the time frame to include a more representative sample of warm-season events. In addition, subsets of ensemble members with like parameterization schemes can be examined to see how each impacts the NME behavior.

5. ACKNOWLEDGMENTS

This work was prepared by the authors with funding provided by National Science Foundation Grant No. AGS-1062932, and NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA11OAR4320072, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, NOAA, or the U.S. Department of Commerce.

The author would also like to give a special thanks to the REU students for their support and camaraderie throughout the duration of this project.

6. REFERENCES

Coffer, B., L. Maudlin, P. Veals, A. Clark, 2013: Dryline position errors in experimental convection-allowing NSSL-WRF model forecasts and the operational NAM. *Wea. and Forecasting*, **28,** 746-761.

Fujita, T., D. Stensrud, D. Dowell, 2007: Surface data assimilation using an ensemble kalman filter approach with initial condition and model physics uncertainties. *Mon. Wea. Rev*., **135,** 1846-1868.

NOAA, cited 2013: Mesoscale ensemble description. [Available on-line from http://rapidrefresh.noaa.gov/WoFMeso/mesoscale_ensemble_description.html]

Stensrud, D., H. Brooks, J. Du, M. Tracton, E. Rogers, 1999: Using ensembles for short-range forecasting. *Mon. Wea. Rev.,* **127,** 433-446.

Stensrud D., N. Yussouf, D. Dowell, M. Coniglio, 2009: Assimilating surface data into a mesoscale model ensemble: cold pool analyses from spring 2007. *Atmos. Res*., **93,** 207-220.

Wheatley, D., D. Stensrud, D. Dowell, N. Yussouf, 2012: Application of a WRF data assimilation system to springtime severe weather events 2007-09. *Mon. Wea. Rev.,* **140,** 1539-1557.