

4B.6

FORECAST GUIDANCE FOR SEVERE THUNDERSTORMS BASED ON IDENTIFICATION OF EXTREME PHENOMENA IN CONVECTION-ALLOWING MODEL FORECASTS.

Ryan Sobash^{*1}, J. S. Kain², D. R. Bright³, A. R. Dean³, M. C. Coniglio², S. J. Weiss³, and J. J. Levit³

¹University of Oklahoma, School of Meteorology, Norman, OK

²NOAA/OAR/National Severe Storms Laboratory, Norman, OK

³NOAA/NCEP/Storm Prediction Center, Norman, OK

1. INTRODUCTION

In an ever-increasing stream of high-resolution numerical model output, the development of new forms of forecast guidance is critical to give operational forecasters the ability to assimilate these large datasets in a timely manner. Forecasts of convective weather have benefited from the ability of convection-resolving models to provide useful guidance of convective mode, intensity, and motion (Done et al. 2004; Kain et al. 2006; Weisman et al. 2008). Guidance associated with convective mode is particularly useful because mode is often associated with the likelihood of different severe weather phenomena (e.g. tornadoes, large hail, damaging winds). Forecasters at the Storm Prediction Center routinely interrogate simulated reflectivity fields from high-resolution model output (among other forms of model guidance) to infer model predictions of convective mode (Weiss et al. 2007; Kain et al 2008).

Since the high-resolution models are beginning to explicitly resolve convective circulations, we have the potential to further interrogate high-resolution NWP output to produce new forms of forecast guidance. In particular, we have the capacity to mine the output for extreme convective phenomena, which could then be examined to determine the degree of correspondence to severe phenomena in the real atmosphere. Although convective circulations are rather coarsely resolved in the current suite of real-time convection-allowing models, subjective assessments suggest that models with grid spacing as coarse as 4 km are able to predict the occurrence of distinct phenomena such as supercells (Kain et al. 2008), which are associated with a disproportionate share of severe weather, including tornadoes. If the correspondence between extreme phenomena predicted by the models, such as strong mesocyclones (supercells), and observations of severe weather can be quantified, relationships could be developed as a tool for using output from convection-allowing models in unique ways as guidance for the forecasting of severe weather.

This concept is fundamentally different from traditional NWP-based assessments of severe weather potential because it relies on identification of explicit convective phenomena rather than environmental conditions that might support severe thunderstorms. Automated assessments of model output have been used for years to characterize convective environments

for the purpose of assessing the likelihood of severe convection, but this is the first time (to our knowledge) that such an assessment has been based on the explicit development of convective-scale phenomena in models. This study builds on an initial proof of concept presented by Sobash et al. (2008) (hereafter SOB08) and explores some of the potential research issues raised therein. The goal of this work is twofold: to determine if predicted convective features can be related to observed severe weather and to examine methods to produce a diagnostic tool that can be used as a guidance product in an operational forecasting setting.

2. DEVELOPMENT OF CONCEPT

The surrogate-severe concept was initially explored during the 2008 NOAA HWT Spring Experiment (hereafter SE2008). Here, extreme phenomena associated with convection were identified in output from a 10-member convection-allowing ensemble. The output was mined for the presence of low-to-mid level mesocyclones, strong low-level winds, and moderately strong low-level winds associated with bowing and linear simulated reflectivity structures. These fields were selected in an ad hoc manner based on subjective assessments which indicated a possible relationship with observed severe weather. A guidance product was produced from the output and analyzed daily alongside experimental probabilistic forecasts and Storm Prediction Center (SPC) Convective Outlooks. On a number of days during SE2008, the agreement in location between the surrogate-severe guidance and observations was surprisingly good. To explore the concept further, SOB08 identified extreme values in five unique 2-D output fields originating from a single daily real-time convection-allowing model forecast. They analyzed various forms of forecast guidance using this technique and demonstrated the utility of the guidance products from one deterministic model.

3. METHODOLOGY

As in SOB08, this study focuses on daily model forecasts from the NSSL-WRF model. This system uses the Advanced Research WRF (ARW – Skamarock et al. 2005) model to produce daily, 36-hr, 4 km forecasts at the NSSL (National Severe Storms Laboratory) over the eastern three-fourths of the CONUS (CONTinental U.S. – see Fig. 1). The model is initialized at 00 UTC and is run to 12 UTC the next day (36 h forecast), in a timely enough manner to be used as guidance for forecasts of the next day's convective cycle. The initial and lateral boundary conditions come

^{*}Corresponding author address: Ryan Sobash,
120 David L. Boren Blvd Suite 5642
Norman, OK 73072
e-mail: rsobash@ou.edu

NSSL-WRF Configuration	
Horizontal Grid	4.0 km
Vertical Levels	35
PBL/Turb. Param.	MYJ
Microphysics	WSM6
Radiation (SW/LW)	Dudhia/RRTM
Init. Conditions	40 km NAM

Table 1: Configuration of the NSSL-WRF.

from the operational North American Mesoscale (NAM) model. No convective parameterization is used; all precipitation originates from the microphysical scheme.

The configuration of the model is summarized in Table 1. This model configuration has remained frozen since early 2007 (over two years). The NSSL-WRF system has been enhanced to supplement standard WRF output with five unique 2-D output fields (Kain et al. 2008). These five fields are hourly-maximum values of 1) 10-meter wind speed (UU), 2) 1 km AGL reflectivity (RF), 3) maximum column (below 400 hPa) upward vertical velocity (UP), 4) maximum column (below 400 hPa) downward vertical velocity (DN), and 2-5 km (AGL) updraft helicity (UH). Each field represents the maximum that occurred in the previous hour (i.e. maximum at each of the model's timesteps in the past hour). This strategy permits the capture of storm-scale features in the model that have short lifetimes and might be missed using the hourly values. In addition, it allows one to track extreme phenomena in between output times.

To sample the behavior of the five 2-D fields across a range of threshold values, a set of 10 thresholds was chosen based on each field's frequency distribution. The distribution was composed of each day's 12-36 hour forecasts over the 51 days of SE2008. The thresholds were chosen at the percentiles between 99.99th and 99.999th, in increments of 0.001%. For each 00 UTC model run in the SE2008 period, surrogate severe reports were accumulated from 12 UTC to 12 UTC (the 13-36 hour output times) for each of the five fields at the model grid points where the field threshold was exceeded at any output time in the 24-hour period, i.e. these grid points were marked as ones in a background field of zeros. The 12 UTC to 12 UTC period was chosen to allow for direct comparison to the SPC's archive of preliminary observed daily storm reports (see <http://www.spc.noaa.gov/climo/>). These observed reports (based on tornadoes, hail greater than or equal to $\frac{1}{4}$ in., wind gusts greater than or equal to 50 knots) were also mapped to the NSSL-WRF's 4 km grid for each day in the experiment.

The final guidance product was created using a procedure originally applied to these observed reports by Brooks et al. (1998). Their approach was designed to be consistent with the SPC's operational probabilistic convective outlooks, in which the end product expresses the likelihood of severe convective weather within 25 miles of a point. Thus, the first step is to tag (with value of one) all grid points within 40 km (~ 25 mi.) of a

surrogate report point as a "hit" (to approximate the field on an 80 km grid). Next, a spatial smoother (i.e. filter) with a Gaussian kernel is applied to the binary report field to produce a field of surrogate severe density (SSD). Specifically, at each grid point, the surrogate severe density (SSD) is given by

$$SSD = \sum_{n=1}^N \frac{1}{2\pi\sigma^2} \exp\left(-\frac{d_n^2}{2\sigma^2}\right)$$

where d_n is the distance from the grid point to the point marking the nth report, N is the total number of reports, and sigma is the spatial smoothing parameter, which is the same in both the x and y directions (i.e. isotropic smoothing). For all the results in this study, $\sigma = 120$ km. To reduce computational time, the results shown here only included grid points within a distance of 5σ . Since the Gaussian filter is acting on a binary field (1s and 0s), the resultant density value can be interpreted as the sum of the Gaussian weights at each grid point within 600 km of the point. The density field is then multiplied by 100 for ease of comparison with probabilistic forecasts, which are usually expressed in percentages. To provide a direct comparison with observed reports, an observed severe density (OSD) field was created in a directly analogous manner, by mapping the observed

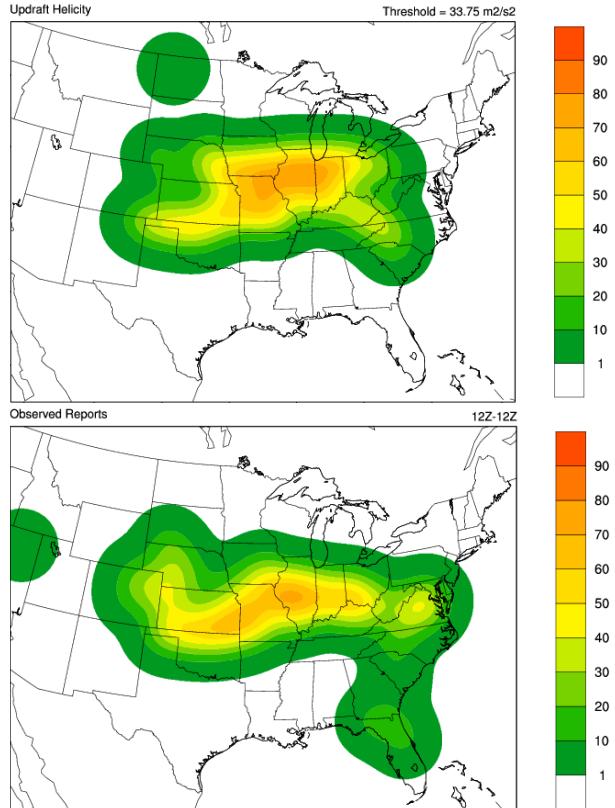
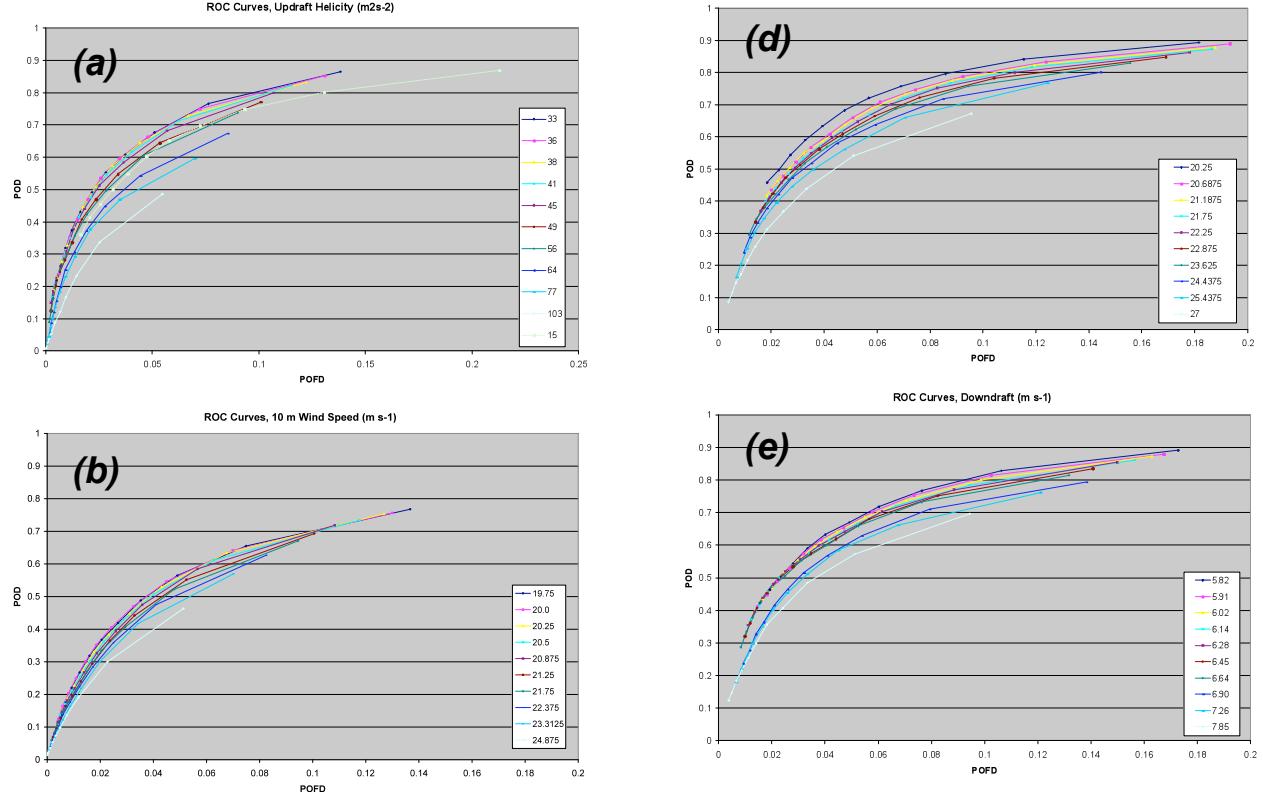


Fig. 1: Examples of (a) Surrogate-severe density UH field and (b) Observed-severe density field for 3 June 2008.

severe reports to the NSSL-WRF's 4 km grid for each day in the experiment and following the same procedure. Variations of this procedure have also been tested, and some of these are described below.

Unlike in SOB08, the SSD product is created for each of the five fields and thresholds individually. This will presumably isolate the impact from the different input fields in order to determine those that correspond most closely with observed severe weather. Also, this will allow for an in-depth examination of the sensitivity to field threshold. Thus, the focus of this paper is to begin the investigation into questions 5.1 (What are the best surrogate fields?) and 5.2 (What are the appropriate thresholds for identifying surrogates?) in SOB08. To do so, a standard forecast verification approach is applied to the SSD forecasts. Contingency-table based metrics (e.g. POD, FAR) are computed using different SSD thresholds to create the binary forecast. For these metrics, the observed binary grid is composed of SPC observed severe reports. As for the surrogate reports, these reports are expanded by flagging grid points within 40 km. To examine the skill of the SSD forecasts, Relative Operating Characteristic (ROC) curves are created using the verification scores. These curves plot the probability of detection (POD) against the false alarm rate (POFD) using a set of probability thresholds. SSD thresholds between 1 and 50 were chosen for each ROC curve, although this gives an incomplete representation of the entire ROC curve space. This precludes a more quantitative evaluation of ROC curve areas, since extrapolation would likely result in invalid results. For now, ROC curve area differences are

Fig. 2: ROC curves for (a) UH (b) UU (c) RF (d) UP and (e) DN. The x-axis (POFD) is expanded to increase separation between the curves.



determined visually to assess the comparative skill at different thresholds and between different fields.

A fractions skill score (FSS) is also computed from the SSD using the OSD field for verification. The FSS is a skill score based on the mean squared error, relative to a low-skill reference forecast (Murphy and Epstein 1989, Schwartz et al. 2009), and is defined as

$$\text{FSS}_{(n)} = \frac{\text{MSE}_{(n)} - \text{MSE}_{(n)\text{ref}}}{\text{MSE}_{(n)\text{perfect}} - \text{MSE}_{(n)\text{ref}}} = 1 - \frac{\text{MSE}_{(n)}}{\text{MSE}_{(n)\text{ref}}},$$

Where the MSE of the forecast and the reference forecast are defined as

$$\text{MSE}_{(n)} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [O_{(n)i,j} - M_{(n)i,j}]^2.$$

$$\text{MSE}_{(n)\text{ref}} = \frac{1}{N_x N_y} \left[\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} O_{(n)i,j}^2 + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} M_{(n)i,j}^2 \right].$$

where both scores are summed over all grid points in the domain. The MSE of the reference forecast can be interpreted as the largest possible MSE that can be obtained from the forecast and the observations.

4. INTRA-FIELD COMPARISON

First, verification scores of each field's ten thresholds will be compared individually using ROC curves and fractions skill scores, beginning with UH. The ROC curves for the UH field (Fig. 2a) give an indication of the usefulness of the UH SSD forecasts. All ten curves lie above and to the left of the "no skill" line (a diagonal from the point where POD and POFD are 0 to where they are 1), indicating that the forecasts can discriminate between events and non-events. Note: The horizontal scales for the ROC curves presented herein are expanded to improve the separation between the curves for each field threshold.

As the threshold increases, the areas under the ROC curves decrease. The difference in areas is negligible for the lowest four thresholds, but the decrease is more apparent for higher thresholds, although the overall difference between the lowest and highest thresholds is relatively small. Thus, it appears that the usefulness of the UH SSD forecasts reaches a maximum for thresholds near $33 \text{ m}^2\text{s}^{-2}$. To further test this result and examine the behavior of the ROC curves below this lowest threshold, SSD UH fields and scores were computed using a threshold of $15 \text{ m}^2\text{s}^{-2}$. This is plotted in Fig. 2a as the dashed line. The area under the ROC curve is smaller (moreso than the thresholds near $33 \text{ m}^2\text{s}^{-2}$), lending further confidence that a threshold near $33 \text{ m}^2\text{s}^{-2}$ is an optimal value for the UH field for this forecast model. Fractions skill scores for the UH thresholds are shown in Table 2. The maximum FSS (0.753) is attained with the $33 \text{ m}^2\text{s}^{-2}$ threshold. FSS decreases with increasing threshold, as well as with decreasing threshold below $33 \text{ m}^2\text{s}^{-2}$ ($15 \text{ m}^2\text{s}^{-2}$ FSS of 0.727, not shown in Table 2).

Some conclusions from the interpretation of the

UH ROC curves are applicable for the ROC curves from the other four SSD fields. In each case, the lowest threshold produced the largest ROC curve area. For UU and RF, the largest FSS was associated with the lowest threshold (0.615 and 0.675, respectively). For UP, the largest FSS occurred for the threshold associated with the 99.995th percentile. For DN, FSS was maximized at the three thresholds in the middle of the range. Overall, the UP and DN FSS did not vary substantially for all but the largest two thresholds. Examination of other thresholds to test the behavior of the verification scores below the lowest threshold was not performed, but will be reported in future work. Further refinement is necessary, but both the ROC curve analysis and the FSSs suggest that it is possible to select an optimal threshold for application of each SSD field.

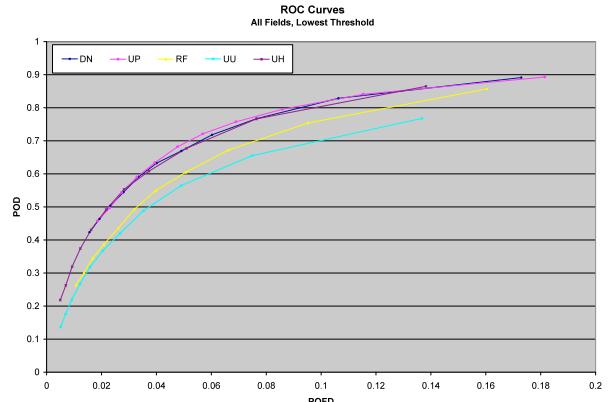


Fig. 3: ROC curves for the lowest threshold in each of the five surrogate fields. The x-axis (POFD) is expanded to increase separation between the curves.

5. INTER-FIELD COMPARISON

Verification scores and FSS will be used to draw comparisons between the relative ability of each field to successfully forecast severe weather reports using the threshold which had the best verification score. In terms of ROC curve areas, the highest areas

Table 2: Fraction skill scores for the five surrogate fields.

percentile	UH	UU	RF	UP	DN
99.99 th	0.753	0.615	0.675	0.715	0.727
99.991 st	0.746	0.611	0.671	0.72	0.73
99.992 nd	0.734	0.601	0.667	0.723	0.732
99.993 rd	0.717	0.586	0.66	0.726	0.736
99.994 th	0.696	0.564	0.649	0.728	0.737
99.995 th	0.665	0.545	0.631	0.729	0.737
99.996 th	0.616	0.512	0.609	0.726	0.737
99.997 th	0.548	0.457	0.572	0.709	0.729
99.998 th	0.454	0.381	0.514	0.663	0.701
99.999 th	0.295	0.266	0.406	0.551	0.623

occurred with the 99.99th percentile threshold, as indicated in the previous section. Using these thresholds, the ROC curve areas for UH, UP, and DN are very similar (Fig. 3). The RF and UU fields have the fourth and fifth largest areas, respectively. UH has the highest FSS score (0.753), followed by DN (0.727) and UP (0.715). Again, RF and UU have the fourth and fifth largest FSS, respectively.

6. DISCUSSION

Preliminary results of the verification scores indicate that the UH, UP, and DN field have superior skill in forecasting the areas of observed severe weather compared to RF and UU. Although in some cases these fields provide duplicate information when used together, in other cases each field provides independent information. The UH field is primarily a diagnostic indicator of supercell (i.e. rotating) thunderstorms. For other types of convection, the UH field may be an extremely poor predictor. For example, “pulse”-type thunderstorms, while not rotating, occasionally do produce severe weather and would potentially be better predicted using the UP or DN fields.

UP and DN provide an alternative picture to the development of convection in the model compared to the other fields, especially UH and UU. Whereas UH and UU often produce “swaths” of surrogate severe reports along the track of model generated convection, UP and DN typically produce a more diffuse, scattershot, field of surrogate reports. This likely impacts the character of the resultant SSD fields.

7. APPLICATIONS AND FUTURE WORK

This study focuses on an initial comparison of the behavior of five surrogate severe storm report fields across a range of thresholds using output from the NSSL-WRF model. During the 2009 Spring Experiment, the SSD product was examined on a daily basis from a set of three convection-allowing WRF models. Preliminary indications are that significant sensitivities exist with different models, even with similar horizontal resolutions. A more quantitative description of these differences is necessary in the future in order to determine appropriate threshold values from different models.

This study was conducted during the period from mid-April to early-June when severe convection was prevalent. Experimental guidance products were made available to the Spring Experiment participants as well as the operational forecasting community, with SSD plots generated in real-time for NSSL-WRF forecasts at <http://www.nssl.noaa.gov/wrf>. Preliminary assessments suggest that the SSD concept is very promising, having significant potential as a guidance product that effectively identifies and highlights areas of model-generated “severe” convective activity that are predicted by convection-allowing models.

A more robust examination of this technique over a longer time period and over different seasonal weather regimes is planned. Also, it is important to examine possible statistical relationships between the surrogate forecast fields and different types of severe weather reports (hail, wind, tornadoes). A variety of questions remain which could be the focus of further work on this subject. First, this work has shown that traditional verification metrics can be used to find optimal thresholds for the surrogate fields. The behavior of the fields near and below the lowest thresholds needs to be analyzed to determine if these are appropriate thresholds. In addition, other surrogate fields may exist which could improve skill; for example, preliminary work using a vertically integrated graupel field from the NSSL-WRF model microphysics scheme is ongoing. Finally, some combination of surrogate severe storm fields, while often providing similar information, could possibly improve these forecasts across a larger range of events.

8. ACKNOWLEDGEMENTS

We are grateful Harold Brooks for many useful discussions related to this work and to the many individuals at CAPS, NCAR, and NCEP/EMC who worked to contribute high-resolution model output to the 2008 Spring Experiment. We thank Scott Dembek of NSSL/CIMMS and NASA/USRA for writing and implementing the code to extract hourly-maximum values of selected fields from the WRF model.

9. REFERENCES

- Brooks, H. E., M. Kay, and J. A. Hart, 1998: Objective limits on forecasting skill of rare events. Preprints, *19th Conference on Severe Local Storms*, Minneapolis, Minnesota, Amer. Meteor. Soc., 552-555.
- Done, J., C. A. Davis, and M. L. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecast (WRF) model. *Atmos. Sci. Lett.*, 5, 110–117, doi:10.1002/asl.72.
- Kain, J. S., S. J. Weiss, J. J. Levit, M. E. Baldwin, and D. R. Bright, 2006: Examination of convection-allowing configurations of the WRF model for the prediction of severe convective weather: The SPC/NSSL Spring Program 2004. *Wea. Forecasting*, 21, 167-181
- Kain, J. S., S. J. Weiss, D. R. Bright, M. E. Baldwin, J. J. Levit, G. W. Carbin, C. S. Schwartz, M. L. Weisman, K. K. Droegemeier, D. B. Weber, K. W. Thomas, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, 23, 931-952.

Murphy, A. H., and E. S. Epstein, 1989: Skill scores and correlation coefficients in model verification. *Mon. Wea. Rev.*, **117**, 572–581.

Schwartz, C. S., J. S. Kain, S. J. Weiss, M. Xue, D. R. Bright, F. Kong, K. W. Thomas, J. J. Levit, M. C. Coniglio, 2009: Next-day convection-allowing WRF model guidance: A second look at 2 vs. 4 km grid spacing. *Mon. Wea. Rev.*, in press.

Skamarock, W.C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, J. G. Powers, 2005: A Description of the Advanced Research WRF Version 2. NCAR Tech Note, NCAR/TN-468+STR, 88 pp. [Available from UCAR Communications, P. O. Box 3000, Boulder, CO 80307].

Sobash, R. A., D. R. Bright, A. R. Dean, J. S. Kain, M. Coniglio, S. J. Weiss, and J. J. Levit, 2008: Severe storm forecast guidance based on explicit identification of convective phenomena in WRF-model forecasts. Preprints, 24th Conference on Severe Local Storms, Savannah, GA, Amer. Meteor. Soc., 11.3.

Weisman, M. L., C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0-36 h Explicit convective forecasts with the WRF-ARW model. *Wea. Forecasting*, **23**, 407-437.

Weiss, S. J., and J. S. Kain, D. R. Bright, J. J. Levit, M. Pyle, Z. I. Janjic, B. Ferrier, J. Du, M. L. Weisman, and M. Xue, 2007: The NOAA Hazardous Weather Testbed: Collaborative testing of ensemble and convection-allowing WRF models and subsequent transfer to operations at the Storm Prediction Center. Preprints, 22nd Conference on Weather Analysis and Forecasting, Park City, UT, Amer. Meteor. Soc., 6B.4.