Using the Mesoscale Model Evaluation Testbed (MMET) to test physics options in the Weather Research & Forecasting (WRF) model

Anthony Torres

SOARS, Fall 2014: Senior, University of Michigan Ann Arbor

SOARS® Summer 2014

Science Research Mentor: Jamie Wolff (NCAR)
Michelle Harrold (NCAR)
Cody Phillips (NCAR)

Writing and Communication Mentor: Wendy Abshire (UCAR)
Computing Mentor: Bill Anderson (NCAR)
Coach: Marcus Walter (WKYC-TV)
Peer Mentor: Erin Dougherty (SOARS, University of Virginia)

ABSTRACT

The Mesoscale Model Evaluation Testbed (MMET) has been set up by the Developmental Testbed Center (DTC) to assist the research community in efficiently testing and evaluating newly developed model techniques aimed to more accurately predict the weather and to potentially be implemented into operations. For this project, datasets available through the MMET were utilized to test the forecast performance of several configurations of the Weather Research & Forecasting model (WRF v3.5.1) using the Advanced Research WRF dynamical core for different physical parameterization schemes and grid-spacing. A significant derecho event that occurred on 29 June 2012 over the U.S. Midwest and Mid-Atlantic states was chosen for this case study.

Statistical analysis of each WRF configuration was conducted using the Model Evaluation Tools (MET), and results were plotted using R, a statistical package. Standard verification metrics were calculated for surface and upper-air temperature, dew point temperature, and wind speed, as well as precipitation. In addition, a more advanced spatial verification technique, known as the Method for Object-based Diagnostic Evaluation (MODE), was utilized to diagnose errors in forecast precipitation placement, coverage, and orientation.

Results indicated that model performance was sensitive to grid-spacing and physics options. Adjusting the grid-spacing from 15 to 5 km, while utilizing the same physics options, allowed for strong convective development that otherwise did not occur. Substituting different microphysics and radiation options at the 5-km grid-spacing had significant effects on developing the storm and resulting storm structure, while changing the planetary boundary layer scheme positively impacted the storm placement. The synoptic environment depicted by the WRF configurations that were able to simulate strong convective development is consistent with research on derecho formation.

This work was performed under the auspices of the Significant Opportunities in Atmospheric Research and Science Program. SOARS is managed by the University Corporation for Atmospheric Research and is funded by the National Science Foundation, the National Oceanic and Atmospheric Administration, the Cooperative Institute for Research in Environmental Science, the University of Colorado at Boulder, and by the Center for Multiscale Modeling of Atmospheric Processes.
1. Introduction

On 29 June 2012, a significant derecho event occurred over the U.S. Midwest and Mid-Atlantic states. Derichos are defined as widespread, long-lived wind storms associated with mesoscale convective systems (Johns and Hirt 1987). According to the comprehensive assessment of the 29 June 2012 derecho completed by NOAA (2013), the storm was responsible for significant damage including 13 deaths and 4.2 million people losing power. Sadly, 34 people died from heat-related illnesses due to a lack of power in the days following the storm (NOAA 2013). Filtered storm reports from the Storm Prediction Center (SPC) indicate that there were 675 reports of damaging winds in excess of 58 miles per hour, 49 reports of hail in excess of 1 inch in diameter, and 2 reports of tornadoes (Fig. 1a).

The North American Mesoscale (NAM) model and Global Forecast System (GFS) used operationally by forecasters failed to ever develop a derecho and the NOAA assessment (2013) suggested that the derecho event was not predicted until hours before the storm initiation. During the morning of 29 June 2012, the experimental High Resolution Rapid Refresh (HRRR) model did develop a derecho; however, as indicated in the report (NOAA 2013), some forecasters at impacted National Weather Service weather forecast offices were hesitant to trust the HRRR forecast due to past erratic performance in which the model developed many storms that never materialized. Thus, due to a lack of warning, many decision makers were caught by surprise. The NOAA report (2013) goes on to conclude that “fully supported, ensemble high-resolution model data extending further out in time are needed in order to improve the lead time in forecasting these extreme wind storms. Without these data, many of these derecho events will continue to have forecast lead times of 24 hours or less.”

The Mesoscale Model Evaluation Testbed (MMET, www.dtcenter.org/eval/meso_mod/mmet/) has been set up by the Developmental Testbed Center (DTC) to assist the research community in efficiently testing and evaluating newly developed model techniques aimed to more accurately predict the weather and to potentially be implemented into operations. For this project, datasets available through the MMET were utilized to test the forecast performance of several configurations of the Weather Research and Forecasting model (WRF v3.5.1) using the Advanced Research WRF (ARW) dynamic core with different physical parameterization schemes and grid spacing. The initial conditions and lateral boundary conditions for each model configuration were derived from the NAM model initialized at 12 UTC on 28 June 2012. Convective initiation of the derecho began around 18 UTC on 29 June 2012, allowing for a 30-hour forecast spin-up of the model configurations. Statistical analyses were performed on each of the WRF configurations by computing traditional verification metrics as well as a more advanced spatial verification technique.

In addition to assessing model performance through objective verification metrics, a more subjective assessment of model performance was also conducted in order to provide a more thorough understanding of why certain model configurations performed better than others. Parameters that were evaluated most closely were surface and upper-air temperature and wind speed, 700 mb vertical velocity, 3-hour accumulated precipitation, 1000-500 mb thicknesses, and composite reflectivity. In order to perform a qualitative analysis of model performance, it is important to understand the environment in which derechos form. A study conducted by Evans and Doswell (2001) examined observed upper air soundings that occurred within 2 hours and 167 kilometers of 67 derechos. The finding of this study indicated that derechos can form in a large range of shear and instability environments. Evans and Doswell (2001) consistently found that events with weak synoptic forcing, similar to the 29 June 2012 derecho, were characterized by weak midlevel winds and strong low-level inflow; the fast propagation of the derechos was associated with strong outflow caused by intense cold pools in the wake of the derechos.
The methods section will feature in-depth discussion of each of the verification metrics used in the evaluation of each model configuration ran for this case study. In addition, the results and discussion sections will include subjective analysis of model performance as a compliment to the objective evaluation performed.

2. Methods

Prior to running the model, all of the input data was preprocessed using the WRF Preprocessing System (WPS). The steps in WPS included first defining the domain of interest and interpolating static geographic information to that domain using `geogrid.exe`. The NAM input data was preprocessed using `ungrib.exe` and then interpolated to the domain defined with `geogrid` using `metgrid.exe`. After preprocessing the data, the next step was running the WRF v3.5.1 model itself. Since this is a case study of a actual event, `real.exe` was first executed in order to initialize WRF for this real data case. After initialization, the WRF ran using `wrf.exe`.

The initial conditions and lateral boundary conditions for each configuration were established by the 12Z NAM that ran on 28 June 2012. The NAM provided lateral boundary conditions in 3-hour increments out to 84 hours. The computational domain used for each configuration included the entire continental United States outlined in black (Fig. 2).

The observed derecho initiated around at 18 UTC on 29 June 2012 near South Bend, Indiana (Fig 1b). Thus, initializing each model configuration at 12 UTC on 28 June 2012 provides a forecast lead time of around 30 hours prior to the derecho formation (Fig. 3).

Each configuration of the WRF model is compared to meteorological observations from the time of the model run. The baseline configuration of the WRF model utilized the Air Force Weather Agency’s (AFWA) operational physics suite at 15- and 5-km horizontal grid spacing. The 15-km configuration parameterized convection, whereas the 5-km configuration resolved convection explicitly. Table 1 indicates the physics options used in the baseline model configuration.

<table>
<thead>
<tr>
<th>Microphysics</th>
<th>WRF Single-Moment 5 scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiation SW and LW</td>
<td>Dudhia/RRTM schemes</td>
</tr>
<tr>
<td>Surface Layer</td>
<td>Monin-Obukhov similarity theory</td>
</tr>
<tr>
<td>Land Surface Model</td>
<td>Noah</td>
</tr>
<tr>
<td>Planetary Boundary Layer</td>
<td>Yonsei University scheme</td>
</tr>
<tr>
<td>Convection (15-km only)</td>
<td>Kain-Fritsch scheme</td>
</tr>
</tbody>
</table>

*Table 1. Physics parameterization options used in the baseline model configuration.*

In addition to the 15- and 5-km baseline model configurations, five other model configurations were tested by substituting different physical parameterizations, generally altering only one scheme at a time, with one exception noted below. For microphysics, several parameterization schemes were substituted including a primitive single-moment scheme that uses only warm hydrometeors developed by Kessler (1969) and more sophisticated double-moment schemes including NSSL2 (Mansell et al. 2010) and Thompson (Thompson et al. 2008) that use both warm and frozen hydrometeors. Other substitutions...
included swapping in the RRTMG (Iacono et al. 2008) short-wave and long-wave radiation schemes and adjusted the planetary boundary layer parameterization to use the MYNN2 (Nakanishi and Niino 2006) in conjunction with the Thompson microphysics scheme. The output from each WRF run was post-processed using the National Centers for Environmental Prediction’s (NCEP) Unified Post Processor (UPP v2.1). Output from the post-processing was used to visually display model fields through the use of the NCAR Command Language (NCL).

Statistical analysis of each WRF configuration was conducted using the Model Evaluation Tools (MET) (Fowler et al. 2010) and was plotted using METViewer, a tool that queries MET verification output from a database and creates plots using the R statistical package. The domain used for verification covers the East region in Figure 2. Model verification was performed for surface and upper-air temperature, dew point temperature, and wind as a function of forecast lead time and vertical level, while precipitation was plotted as a function of accumulation threshold. For surface and upper-air verification of temperature, dew point temperature, and wind, MET was used to calculate mean error (ME) (Eq. 1) and bias-corrected root mean square error (BCRMSE) (Eq. 2). For the precipitation evaluation, traditional verification metrics of frequency bias (Eq. 3) and Gilbert Skill Score (GSS) (Eq. 4) were computed. In addition, a more advanced spatial technique known as the Method for Object-based Diagnostic Evaluation (MODE) (Davis et al. 2006) (Davis et al. 2009) was utilized for evaluating the quantitative precipitation forecasts.

a. Equations

1.) MEAN ERROR (OR BIAS)

The mean error (ME) is the overall average difference between the forecast and the observed value.

$$ME = \frac{1}{N} \sum_{i=1}^{N} (f_i - o_i) \quad \text{Eq. 1}$$

The mean error can range from a score of $-\infty$ to $+\infty$ with 0 being a perfect score.

2.) BIAS-CORRECTED ROOT MEAN SQUARE ERROR (BCRMSE)

The BCRMSE is essentially the square root of the average square error of the forecasts with the bias term removed. This is calculated with the following equation:

$$BCRMSE = \sqrt{ME^2 + s_{f-o}^2} \quad \text{Eq. 2}$$

Where $s_{f-o}^2$ is the variance of the error and is calculated as:

$$s_{f-o}^2 = s_f^2 + s_o^2 - 2s_fs_or_{f-o}$$

The range of values is between 0 and $\infty$, with 0 being a perfect score.

3.) FREQUENCY BIAS

The frequency bias is the ratio of the total number of forecasts of an event to the total number of observations of the event, where the event is defined by a particular threshold.

$$f_{bias} = \frac{\text{hits} + \text{false alarms}}{\text{hits} + \text{misses}} = \frac{\text{total forecasted}}{\text{total observed}} \quad \text{Eq. 3}$$
A hit indicates that an event was both observed and forecasted by the model. A false alarm occurs when the model forecasts an event that is not observed. A Miss indicates that an event was observed, but not forecasted by the model. (Table 2)

<table>
<thead>
<tr>
<th>Forecast at Threshold</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Hits (YY)</td>
</tr>
<tr>
<td></td>
<td>False Alarms (YN)</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Misses (NY)</td>
</tr>
<tr>
<td></td>
<td>Correct Rejections (NN)</td>
</tr>
<tr>
<td></td>
<td>YY+NY</td>
</tr>
<tr>
<td></td>
<td>YN+NN</td>
</tr>
</tbody>
</table>

**Table 2.** Contingency table used for calculating the frequency bias and gilbert skill score.

A perfect frequency bias score is 1. If the value is greater than 1, the event was over-forecast, and if the value is less than 1, it indicates an under-forecast. The minimum frequency bias is 0 if no events were predicted and the maximum value is $\infty$.

4.) **GILBERT SKILL SCORE (GSS)**

GSS is the ratio of the number of times an event was correctly forecast to the number of times an event was either forecast or occurred, corrected for random hits that are expected by chance. The GSS is calculated with the following equation:

$$GSS = \frac{hits - hits_{rand}}{hits + misses + false alarms - hits_{rand}} \quad Eq. 4$$

where:

$$hits_{rand} = \frac{(Total \ Forecast \ Area)(Total \ Observed \ Area)}{Total \ Area}$$

A perfect score in this case is 1. Since the GSS takes into account random hits by chance, it is possible to get a score as low as -0.33, with a no-skill forecast having a score of 0.

5.) **BASE RATE**

The base rate is the number of grid points that have observed precipitation at a particular threshold over the total number of grid points in the model’s domain.

$$base \ rate = \frac{total \ observed}{total \ area} \quad Eq. 5$$

Thus, the base rate ranges between 0 and 1.

6.) **METHOD FOR OBJECT-BASED DIAGNOSTIC EVALUATION (MODE)**

In addition to the traditional verification metrics used for the precipitation evaluation, MODE was also applied to provide a spatial verification approach. Through a variety of highly configurable options, MODE is able to evaluate model performance by taking into account the size and shape of observed precipitation objects and comparing them to forecasted precipitation objects. MODE was designed to mimic the subjective evaluation of a human forecaster in an automated, objective manner. In this way, MODE is physically meaningful and intuitive while providing information on location and structural errors.

SOARS® 2014, Anthony Torres, 5
The MODE analysis was conducted by first acquiring all of the raw precipitation data from observed and forecast sources. The precipitation data is then smoothed through a process called convolution. The convolution radius was 3 for configurations with 15 km grid spacing and 9 for 5 km grid spacing. After applying the convolution operator, a mask field was produced by using a convolution threshold of >0.254 mm. From this, the precipitation objects were generated for analysis.

3. Discussion

a. Composite Reflectivity

The 29 June 2012 derecho initiated at approximately 18 UTC near South Bend, Indiana. Due to the fast propagation of the storm, the leading edge of the squall line was located over eastern Virginia by 03 UTC 30 June 2012 (Fig. 3). When compared to composite reflectivity output from each WRF configuration, the predicted forecasts diverge significantly. The 15-km baseline (Fig. 4a) and 5-km Kessler (Fig. 4c) configurations do not develop any storm at all. Other configurations, such as the 5-km RRTMG (Fig. 4f) and 5-km Thompson (Fig. 4g) develop the storm but place it too far north and west. The 5-km RRTMG configuration’s error is not a result of a poor storm track but delayed initialization of the storm at 00 UTC on 30 June 2012. The 5-km Thompson configuration storm placement errors are a result of storm initiation too far north and west (in southern Wisconsin). While no WRF configuration placed the derecho over eastern Virginia at 03 UTC on 30 June 2012 (forecast lead time of 39 hours), configurations that were closest were the 5-km baseline (Fig. 4b), 5-km MYNN2 (Fig. 4d), and 5-km NSSL2 (Fig. 4e). It is important to note the structural differences embedded within each of the model configuration’s solution including the extent of aerial coverage of the rainfall associated with the derecho. As the derecho matured, a significant rain shield developed and caused rain to fall over a large area west of the leading edge of the storm. The 5-km MYNN2 (Fig. 4d), 5-km RRTMG (Fig. 4f), and 5-km Thompson (Fig. 4g) configurations showed a pronounced stratiform rain shield behind the main convection. Models that developed a strong linear or bowing structure include the 5-km NSSL2 (Fig. 4e), 5-km RRTMG (Fig. 4f), and 5-km Thompson (Fig. 4g), whereas the 5-km baseline model configuration showed disorganized areas of strong convection in the model-derived composite reflectivity product throughout the duration of the derecho event.

b. Synoptic Environment

The synoptic environment depicted in each of the model configurations exhibited noticeable differences at the 30 hour forecast lead time that may have played a role in developing the derecho. When analyzing the 1000-500mb thickness, the model configurations that did not develop any significant convection showed a weakness in the ridging aloft over Missouri. That is, there was a split in the 1000-500 mb thickness contours between the eastern and western U.S. This was seen in the 15-km baseline (Fig. 5a) and 5-km Kessler (Fig. 5c) models. Other model configurations that did develop significant convection, such as the 5-km baseline (Fig. 5b), 5-km MYNN2 (Fig. 5d), 5-km NSSL2 (Fig. 5e), 5-km RRTMG (Fig. 5f), and 5-km Thompson (Fig. 5g) showed stronger ridging aloft with continuous 1000-500 mb thickness contours between the eastern and western U.S.

c. Mesoscale Features

Another key factor in assessing model performance is investigating whether predicted mesoscale features in each configuration are consistent with derecho formation and structure. Figure 8 examines mesoscale-level updrafts and downdrafts at the 700-mb level at the 39 hour forecast lead time. The model configurations that show significant updrafts, such as the 5-km baseline (Fig. 6b), 5-km MYNN2 (Fig. 6d), 5-km NSSL2 (Fig. 6e), 5-km RRTMG (Fig. 6f), and 5-km Thompson (Fig. 6g) also depicted the derecho in composite reflectivity output (Fig. 4).

Model configurations that failed to pick up strong sustained updrafts at the 700-mb level for any forecast lead time included the 15-km baseline (Fig. 6a) and 5-km Kessler (Fig. 6c). These models, again, did not pick up on significant convection in the composite reflectivity (Fig. 4).
b. Verification

All of the objective verification was performed using MET over the Eastern U.S. domain (Fig. 2).

1.) 2M DEW POINT TEMPERATURE

A strong diurnal trend is evident with the smallest errors seen around 12 UTC valid times and the largest errors near 00 UTC valid times for the 2m dew point temperature BCRMSE (Fig. 7a). BCRMSE values are clustered for early lead times and spread becomes at later forecast lead times.

Similar to the BCRMSE, a diurnal signal was also evident in the 2m dew point temperature bias (Fig. 7d). In general, each model configuration started with a warm bias and transitioned to a cold bias at later forecast lead times. The 5-km RRTMG configuration showed the smallest cold bias, whereas the 5-km Kessler configuration showed the strongest cold bias. At forecast lead times of greatest interest (27 to 42 hours, prior to and during the storm), BCRMSE and bias of each configuration were within 2°C of observations.

2.) 2M TEMPERATURE

Each model configuration showed increasing error with forecast lead time when looking at the BCRMSE (Fig. 7b). A diurnal trend was also noted with each configuration with the smallest errors seen around 12 UTC valid times and the largest errors near 00 UTC lead times. The BCRMSE values were generally clustered throughout all forecast lead times, with the largest error coming from the 5-km Kessler configuration.

The diurnal trend was less pronounced in the plot featuring the 2m temperature bias (Fig. 7e). Instead, most of the model configurations showed a slight cool bias throughout each of the lead times. The 5-km RRTMG configuration had the weakest temperature bias, and even showed a slight warm bias around 24, 36, and 48 hour forecast lead times. There was strong general agreement for 2m temperature between most model configurations. Within the models in agreement, there was a cold bias as low as -2°C. The 5-km Kessler configuration had by far the largest cold bias, with values as low as -6°C.

3.) 10M WIND SPEED

Overall, there was strong general agreement within each of the model configurations when looking at the BCRMSE (Fig. 7c) and the bias (Fig. 7d). There were low BCRMSE values of under 1.7 m/s during earlier forecast lead times of 3 to 30 hours. After a forecast lead of 30 hours, the BCRMSE values increased and began to show a strong diurnal signal with the largest errors occurring around 12 UTC valid times and the smallest errors occurring around 00 UTC valid lead times.

There was generally a high 10m wind speed bias through all lead times and model configurations (Fig. 7f). The weakest bias came from the 5-km Kessler configuration with a slight low-to-neutral bias occurring around 33, 57, and 81 hour forecast lead times. Each model configuration showed a strong diurnal trend with the highest biases occurring at around 06 UTC valid times and the lowest biases occurring around 21 UTC forecast lead times.

4.) 2M TEMPERATURE BIAS BY STATION

Figure 8 shows the temperature bias between each model configuration output and observed temperatures at a forecast lead time of 27 hours (3 hours prior to storm initiation). Each of the model configurations showed a cold bias over the Midwest and Great Lakes. The 5-km Kessler configuration showed the strongest cold bias (lower than -4°C).
5.) **Frequency Bias**

Figures 9a and 9b show the frequency bias at forecast lead times of 36 and 39 hours, respectively. The base rate decreased as precipitation threshold increased. At a forecast lead time of 36 hours (Fig. 6a), the frequency bias score closest to one is from the 15-km baseline model. This contradicts the interpretation of the results from the spatial analysis performed by MODE and the convective reflectivity which never developed the derecho. This suggests that grid-to-grid verification approaches benefit coarser resolution models and penalize higher resolution models more severely for displacements in time and position.

6.) **Gilbert Skill Score (GSS)**

Figures 9c and 9d show the GSS at forecast lead times of 36 and 39 hours, respectively. With this metric, the best performing model configurations are the 5-km MYNN2 and 5-km RRTMG. Additionally, the GSS for each configuration decreased as the base rate decreased. This suggests that the model configurations had a poor handle on heavier precipitation events at the 5-km grid scale. No model configuration received a GSS higher than 0.12.

7.) **MODE Analysis**

At 03 UTC on 30 June 2012, the observed derecho was in a mature stage, and the leading edge had advanced eastward to the Mid-Atlantic coast with an extensive rain-shield extending as far west as the Ohio River (Fig. 3b). From the MODE analysis, none of the model configurations accurately picked up how fast the derecho was moving. The configurations that were farthest east were the 5-km baseline (Fig. 10b), 5-km MYNN2 (Fig. 10d), and 5-km NSSL2 (Fig. 10e). The track errors in the 5-km RRTMG (Fig. 10f) and 5-km Thompson (Fig. 10g) configurations continued through the duration of the derecho event due to poor timing and initiation placement, respectively. The precipitation analyzed from the MODE analysis in the 15-km baseline (Fig. 10a) and 5-km Kessler (Fig. 10c) are from short term convective events unrelated to the main derecho event.

c. **Future Work**

In an effort to find a model configuration that best captures the location, timing, and intensity of this damaging derecho and potentially other high impact derecho events in the future, it would be interesting to run one additional model configuration that combines the best performing schemes that have already been examined here. This physics suite would include the combination of RRTMG for short- and long-wave radiation, Thompson microphysics, and the MYNN2 planetary boundary layer scheme. Similar analysis methods would be applied as were discussed above.

In addition to running another model configuration, continuation of a deeper investigation will be performed by completing vertical analysis of model performance. One possible way to accomplish this is by comparing observed soundings in areas near the derecho to forecast soundings from each of the configurations.

The additional analysis is expected to continue through at least the next year in effort to properly rank model performance. The configurations that show the greatest improvement to the baseline configuration will be recommended for more extensive analysis under a wide variety of weather events to assess the applicability for use in 24x7 operations.

4. **Conclusion**
On 29 June 2012, a significant derecho traversed over the U.S. Midwest and Mid-Atlantic states. This event was not well predicted by many operational models at the time and, thus, caught many people by surprise. In effort to prevent similar storms from being missed by models in the future, I ran 7 different configurations of the WRF-ARW v3.5.1. Each model run was initialized at 12 UTC on 28 June 2012, giving a forecast lead time of at least 30 hours prior to the derecho development.

Grid spacing had a clear effect on model performance. The 15-km baseline configuration did not allow for resolving convection explicitly and, rather, used a convective parameterization scheme, which did not pick up on the storm. However, the 5-km baseline configuration, which did resolve convection explicitly, showed significant convective development in the general area of the observed derecho.

Model performance was also sensitive to the microphysics schemes used in each run of the WRF. The objective analysis confirmed that the primitive single-moment microphysics Kessler scheme was unable to properly develop a derecho, or even any significant convection. The NSSL2 and Thompson microphysics schemes both showed significant convective development in the area of the observed derecho.

In addition to the objective analysis provided by MET, a few additional conclusions can be made from the subjective analysis of the model output. The model configurations that performed well during the derecho event displayed both synoptic and mesoscale features that were consistent with what one would expect with such an event. The 1000-500 mb thickness indicated strong ridging extending continuously from the Eastern to Western U.S. Derechos tend to travel along the periphery of a ridge, and this analysis showed that the model configurations with the strongest ridging, such as the 5-km baseline, 5-km MYNN2, 5-km NSSL2, 5-km RRTMG, and 5-km Thompson, were the configurations that tended to develop significant convection. Related, the two model configurations that did not pick up the storm, the 15-km baseline and 5-km Kessler, showed a weakness in the ridge over Missouri. When analyzing mesoscale features, such as updrafts, the model configurations that did pick up the storm also showed long, sustained updrafts at the 700mb level.
5. Figures

**Figure 1.** Overview of the derecho event including the temporal radar composite and the SPC storm reports. Each wind report represents winds in excess of 58 mph, and each hail report represents hail larger than 1 inch in diameter.

**Figure 2.** Full domain used in each WRF-ARW configuration. The EAST domain was used for all of the verification techniques.
Figure 3. Regional composite reflectivity mosaic of the 29 June 2012 derecho over the Midwestern and Mid-Atlantic regions.

Figure 4. Composite reflectivity output from each WRF configuration at a forecast lead time of 39 hours (valid at 03:00 UTC 30 June 2012).
Figure 5. 1000-500 mb thickness (m), mean sea-level pressure (mb), and 3 hour accumulated precipitation at a forecast lead time of 30 hours.
Figure 6. 700 mb vertical velocity (m/s) at a forecast lead time of 39 hours.

Figure 7. Bias-corrected root mean square error (top) and mean error (bottom) for 2-m dew point temperature (left), 2-m temperature (middle), and 10-m wind speed (right) by forecast lead time.
Figure 8. Spatial plots of 2-m temperature mean error at each station for each WRF configuration at a forecast lead time of 27 h.
Figure 9. Frequency bias (top) and Gilbert Skill Score (bottom) over the CONUS-EAST domain at forecast lead times of 36 (left) and 39 (right) hours. The base rate is indicated in black and is on the y-2 axis.
Figure 10. Example illustrating MODE objects created from WRF 3-h accumulated precipitation fields and the associated Stage II analysis fields for the 39 hour forecast from the 28 June 2012 12 UTC initialization. In this example, the configuration being evaluated is using the RRTMG long- and short-wave radiation schemes. A convolution threshold of 0.254 mm was used to define the objects; the shaded regions represent objects in the forecast field, and outlined areas represent objects in the observed field.
References


