89th AMS Annual Meeting, 16th Conference on Satellite Meteorology and Oceanography,
11-16 January 2009, Phoenix, AZ.

P1.15 AN EXAMINATION OF SIMULATED REFLECTIVITY FORECASTS FROM A 10-MEMBER STORM SCALE ENSEMBLE PREDICTION SYSTEM

Derek R. Stratman*[1], D. R. Bright[2], S. J. Weiss[2], J. S. Kain[3], J. J. Levit[2], M. C. Coniglio[3], K. W. Thomas[4], F. Kong[4], and M. Xue[4]
[1]*NOAA/NWS/SPC Hollings Scholar, Valparaiso University, Valparaiso, Indiana*
[2]*NOAA/NWS/Storm Prediction Center, Norman, Oklahoma*
[3]*NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma*
[4]*Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma*

1. INTRODUCTION

During the 2008 Hazardous Weather Testbed (HWT) Spring Experiment in Norman, OK, higher resolution modeling systems were subjectively analyzed by research scientists and weather forecasters from across the United States. The goal of this collaboration was to examine different storm scale models and their capabilities in operational meteorology. One of the ways the group of researchers and forecasters did this was by subjectively comparing the simulated reflectivities with the actual base reflectivities from the same time period on a one to ten scale with ten being the best. Two of the storm scale systems that were evaluated were the 10-Member WRF-ARW Storm Scale Ensemble Forecast (SSEF) system and the 2-km WRF-ARW deterministic forecast model. In order to explore some of the sensitivities associated with the SSEF system, the national mosaic of base reflectivity (BREF) was compared with the SSEF members' output of simulated reflectivities by using basic data analysis techniques. Also, the 2-km WRF-ARW deterministic forecast's simulated reflectivity was objectively compared to the 4-km control run's (with radar assimilation) simulated reflectivity with the intention of exploring the sensitivity of horizontal grid spacing size.

2. DATASETS

2.1 Models

During the 2008 HWT Spring Experiment, the 10-Member WRF-ARW SSEF system and the 2-km WRF-ARW deterministic forecast model were run for a 6-week period starting at the end of April and ending at the beginning of June. Both model forecasts were initiated everyday of the Spring Experiment at 00Z. Because of the large file sizes and computational times, this paper's study only used data from 10 days (e.g. 13, 14, 15, 20, 21, 22, 27, 28, 29, 30) in May 2008. All of these days had some type of precipitation, often convective, located in the area of focus.

* *Corresponding Author Address:* Derek R. Stratman, Valparaiso University, Department of Geography and Meteorology, 1809 Chapel Dr., Kallay-Christopher Hall, Valparaiso, Indiana 46383-6493; e-mail: Derek.Stratman@valpo.edu

The 10-member WRF-ARW SSEF system was comprised of eight ensemble members and two control forecasts (Table 1). The forecast system was computed with 4-km grid spacing on a domain that covered the eastern two-thirds of the United States (Figure 1). Each of the eight ensemble members had perturbed initial and lateral-boundary conditions with different physics schemes. All of these perturbed members were also assimilated with radar data. Both of the control runs were identical except that one control had radar assimilated into it and the other control did not have assimilated radar data. All ten members of the SSEF system did not have convective parameterization. The 2-km WRF-ARW was computed on the same domain as the SSEF system but with 2-km horizontal grid spacing.

member	IC	LBC	Radar data	mp_phy	sw-phy	pbl_phy
Cntl	00Z NAMa	00Z NAMf	yes	Thompson	Goddard	MYJ
C0	00Z NAMa	00Z NAMf	no	Thompson	Goddard	MYJ
N1	Cntl – em_pert	21Z SREF em_n1	yes	Ferrier	Goddard	YSU
P1	Cntl + em_pert	21Z SREF em_p1	yes	WSM 6-class	Dudhia	MYJ
N2	Cntl – nmm_pert	21Z SREF nmm_n1	yes	Thompson	Goddard	MYJ
P2	Cntl + nmm_pert	21Z SREF nmm_p1	yes	WSM 6-class	Dudhia	YSU
N3	Cntl – etaKF_pert	21Z SREF etaKF_n1	yes	Thompson	Dudhia	YSU
P3	Cntl + etaKF_pert	21Z SREF etaKF_p1	yes	Ferrier	Goddard	MYJ
N4	Cntl – etaBMJ_pert	21Z SREF etaBMJ_n1	yes	WSM 6-class	Goddard	MYJ
P4	Cntl + etaBMJ_pert	21Z SREF etaBMJ_p1	yes	Thompson	Dudhia	YSU

Table 1: Configurations of the SSEF system's members. For all members, ra_lw_physics = RRTM, sfc_physics = Noah, cu_physics = NONE.

2.2 Radar Data

In order to objectively evaluate the generated simulated reflectivity values from these forecasts, a national mosaic of base reflectivities was used as a comparison tool. In this study, the operational UNISYS 2-km filtered radar mosaic product was used. The filtering is done with an algorithm that is designed to take into account the two bottom elevation slices in order to minimize the anomalous propagation (AP) at each individual WSR-88D radar site. However, since individual Doppler radars are used in the mosaic, some AP can make it through the filtering process due to the location and other factors causing a problem. Also, the radar data do not fill the entire domain of the models due to no radars being located over the ocean and non-U.S. areas.

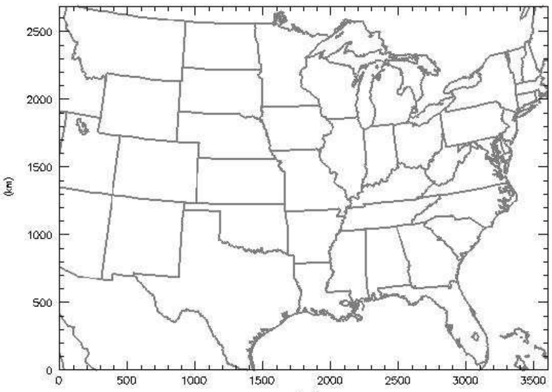

Figure 1: SSEF and 2-km WRF-ARW domain. Also, BREF was switched to this domain

3. METHODOLOGY

3a. Sensitivities

In this study, different sensitivities were explored by using the simulated reflectivity outputs from the SSEF system and the 2-km WRF-ARW. The first sensitivity explored involved the comparison of the different cloud microphysics schemes: Thompson, Ferrier, and WSM 6-Class. Another important sensitivity that was looked at has to deal with the radar assimilation into the ensemble. For this, the two 4-km control runs were compared since they were identical with the same physics schemes, but with and without assimilated radar data. It should be noted here that the control run without the assimilated radar data did not produce simulated reflectivities until forecast hour one (e.g. the first output time), while the control run with radar assimilation had simulated reflectivity output at the initial forecast hour. An additional sensitivity that will be discussed is the evolution of convection by the SSEF system's individual members.

The last sensitivity involved looking at the comparison of the simulated reflectivity from both the 4-km and 2-km horizontal grid spacing deterministic models runs. However, the 2-km WRF-ARW used a different cloud microphysics scheme (WSM 6-Class) than the 4-km control run with assimilated radar data (Thompson), which does not allow for a clean comparison between the two. Nevertheless, there was a member (e.g. n4) with the same physics schemes as the 2-km WRF-ARW. However, this member had a negative initial perturbation.

3b. Pairing Model Domain and Radar Domain

Since the SSEF system's output and the radar data are on two separate grids, a technique was used to match them up. It was determined that the best way to pair the two domains up was by converting the radar data over to the model domain grid. The best option was to do a scan of all of the actual radar reflectivity values within a 25 mile radius of the domain grid point and assign that grid point with the highest radar reflectivity value in that circle. However, this process took too long for what this study was supposed to entail, so a different option was used instead. The next best option was to select the actual radar reflectivity values closest to the model domain grid points. This process only took about a twentieth of the time, and the results subjectively looked similar (not shown). This process was used on all ten days of radar reflectivity data, so in other words, the actual radar reflectivity was converted to a 4-km grid spacing with the same grid point locations as the 4-km SSEF system's grid points. It should be noted that areas in the domain located over the oceans and non-U.S. land had no radar data, so the grid points were assigned values of zero.

3c. Data Analysis

In order to explore the sensitivities with simulated and actual reflectivity values, two basic data analysis techniques were used. The first technique that was used to compare the simulated reflectivities and the actual reflectivities from the ten days in May 2008 was fractional areal coverages. The areal coverages were found by counting the number of grid points with reflectivity values at or above the specified threshold (e.g. 20, 30, 40, and 50 dBz and greater) and dividing by the total number of grid points. This technique was done on all of the SSEF ensemble members, the 2-km WRF-ARW, and the actual radar reflectivity values.

The other technique that was used to compare the different reflectivities was areal biases. The biases at the four reflectivity thresholds were found for the SSEF system's ten individual members and the 2-km WFR-ARW. The biases were computed by taking the forecasted reflectivity divided by the actual reflectivity. On this scale, a one would be a perfect match.

4. RESULTS

4a. Model Cloud Microphysics Schemes

Simulated reflectivity outputs by the individual SSEF members were highly dependent on the cloud microphysics schemes. Therefore when all of the members were plotted on bias plots for each of the reflectivity thresholds, there was no surprise that clustering of the members took place (not shown). This clustering was due to the different cloud microphysics schemes in each of the members. After grouping the members that used the Thompson scheme (cn, c0, n2, n3, and p4), the Ferrier scheme (n1 and p3), and the WSM 6-Class scheme (p1, p2, and n4), it became clear which cloud microphysics scheme performed the best and which one performed the worst.

In Figure 2, the 40 dBz biases are shown. The ensemble members that had biases closest to one (i.e. no bias or perfect) all had the WSM 6-Class cloud microphysics scheme. The ensemble members with the Ferrier cloud microphysics scheme were just a little bit further away from one. However, the five ensemble members with the Thompson cloud microphysics scheme all had a substantial low bias. This means the Thompson scheme did not produce as much simulated reflectivity as the other cloud microphysics schemes or as much as what the actual reflectivity depicted. This was evident at all threshold levels. Also, since these biases include no radar data over the data void areas in the domain, the biases would be even smaller.

It should also be noted that all of the biases decreased as the reflectivity threshold values were increased. At the 50 dBz and greater threshold, the Ferrier cloud microphysics scheme had biases close zero (not shown). This means the Ferrier cloud microphysics scheme did not produce many reflectivity values above 50 dBz, which can be associated with severe weather, but this was expected.

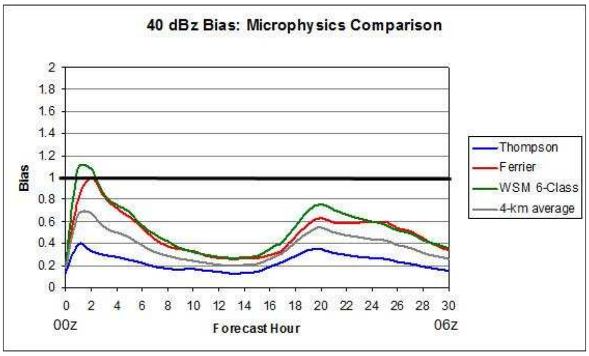

Figure 2: Average biases of the SSEF members with the same microphysics scheme at the 40 dBz threshold.

4b. Assimilation of Radar Data

The sensitivity of the assimilation of radar data into higher resolution models can be explored with the SSEF system's two control runs because they had the same initial and lateral-boundary conditions and physics schemes, but only one of them had radar data assimilated into it at the beginning. In Figure 3, both control runs were plotted on the fractional areal coverage graph along with the ensemble average and the actual data. The low values of the controls can be attributed to their use of the Thompson cloud microphysics scheme. However, some conclusions can still be drawn. The control run with assimilated radar data (e.g. cn) outperformed the control run without assimilated radar data (e.g. c0) until around forecast hour nine. After this point, the difference between the two controls was negligible. This indicates the assimilation of radar data into models is beneficial for the short term forecast, but for longer term forecasts, it does not matter if radar assimilation is used in the model run.

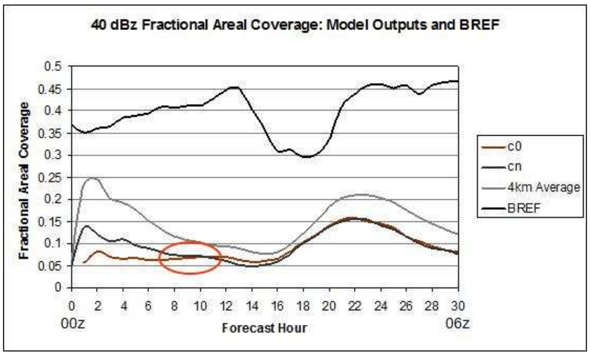

Figure 3: Graph of the two controls' (e.g. with and without radar assimilation) percent coverages along with the BREF's percent coverages at the 40 dBz threshold. The red circle highlights where the difference between the two controls begins to get small.

4c. Evolution of Convection

The best way to analyze the evolution of convection was by looking at fractional areal coverages for threshold values at or above 40 dBz. The thresholds that were less than this showed the same result. In Figure 4, the SSEF's ensemble members clearly fostered a diurnal trend, but the actual data from the BREF showed a secondary peak in the percentage of intensities around 12z. In this same time frame, the fractional areal coverages were decreasing toward a minimum around 14z to 15z. After 15z, the areal coverages started to match the actual areal coverage in shape, which was due to the diurnal cycle.

This nocturnal convective peak was most likely due to such features as mesoscale convective systems (MCS), which are common during the month of May. In Figure 5, the convective nocturnal peak in the actual areal coverage ceased to exist because there were not as many core reflectivity values greater than 50 dBz. For example on 29 May at 12Z, the national mosaic of BREF shows a MCS with a large area of dBz values above the 40 dBz threshold but not above the 50dBz threshold (Figure 6a). Also, the control run for the same time is shown (Figure 6b). Notice the control run was hinting at some convection, but the coverage of that simulated convection is not as nearly intense or as widespread as what the actual BREF mosaic depicts.

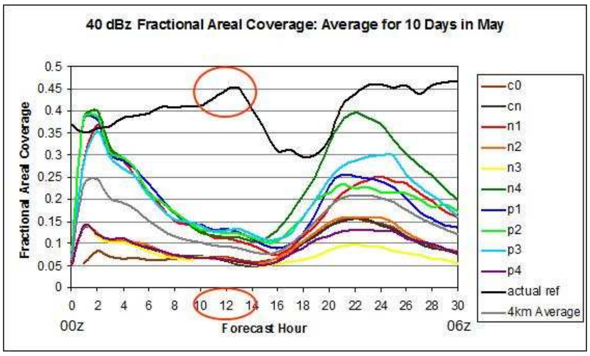

Figure 4: Graph of all SSEF members' percent coverages along with BREF's percent coverage at the 40 dBz threshold. The red circles highlight the time of the convective peak, which was around 12Z.

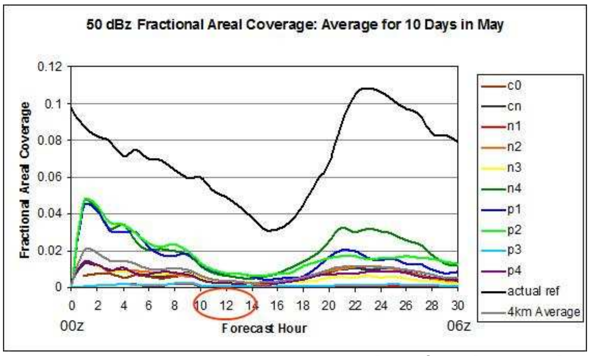

Figure 5: Same as in Figure 4 but for the 50 dBz threshold.

4d. 4-km vs. 2-km Horizontal Grid Spacing

As mentioned before, because the 4-km control with assimilated radar data did not have the same cloud microphysics scheme as the 2-km WRF-ARW, one of the perturbed members (n4) was used for the comparison due to having all of the same physics schemes as the 2-km WRF-ARW. In Figure 7, although n4 is slightly different than the 2-km WRF-ARW due to the perturbation, the difference between the two was negligible for most of the forecast period until the very end, where the initially perturbed member actually did better in forecasting the 40 dBz and greater fractional areal coverage. Since the comparison is not perfect, no definitive result can be concluded from this finding, but nonetheless, it is interesting to consider.

5. SOURCES OF ERROR

Due to time constraints, several sources for error were possible and likely occurred with each step of this research. However, this study was meant to be a new way of looking at the SSEF system, which is still in the research mode. First of all, only ten days worth of data were used in this study. Ideally, for any significant results to be justified, at least thirty days of data should be used. Second of all, as mentioned before, the BREF radar data did not fill the entire domain of the models. If this would be fixed, the biases would be even smaller. Conversely, some ground clutter still managed to get through the filtering process, so if this would be fixed, the biases would be even larger, especially for the smaller reflectivity thresholds.

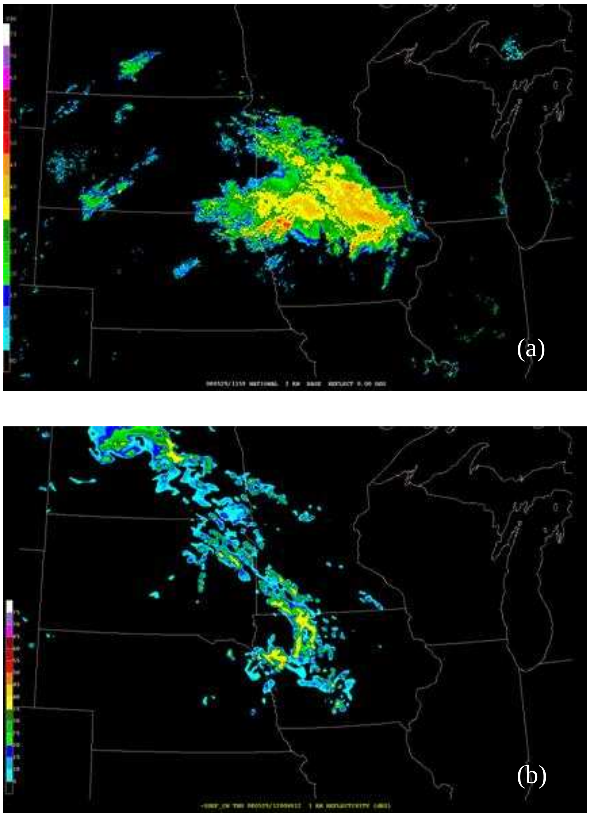

Figure 6: (a) 29 May 2008 12z BREF; (b) 00Z control (cn) member's 12-hour simulated reflectivity forecast.

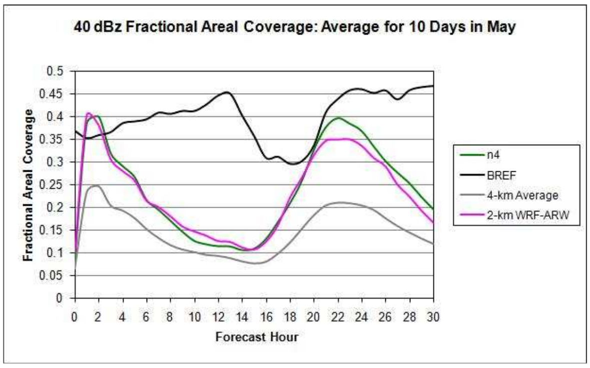

Figure 7: Graph of the percent coverages of the 4-km and 2-km model runs at the 40 dBz threshold.

6. CONCLUSIONS AND FUTURE WORK

The goal of this study involved the exploration of a few sensitivities by using the simulated reflectivity output from the 10-member SSEF system and the 2-km WRF-ARW and the actual BREF values from the national mosaic. The simulated reflectivities and the BREF data were used together in two basic data analysis techniques (e.g. fractional areal coverages and biases) to evaluate the sensitivities. It was found that the Thompson cloud microphysics scheme did not produce as much areal coverage of simulated reflectivity as the WSM 6-Class and Ferrier schemes. Also, the conclusion can be drawn that the assimilation of radar data into the model runs produces a more accurate areal coverage for about the first nine forecast hours. After that point, the difference was negligible. For this data set, the 10-member SSEF system had difficulty producing the overnight convective trends. Finally, it was shown that based on this data set the fractional areal coverages of the almost identical 4-km negatively perturbed member (e.g. n4) and the 2-km WRF-ARW were very similar to each other through most of the forecast period.

The first area of this study to improve upon would be the sample size. A sample size of at least thirty days would be ideal rather than just ten. Also, it would be advantageous to examine in even more depth the advantages and disadvantages of the higher resolution WRF-ARW. In addition, it would be beneficial to explore the possible reasons as to why the WRF-ARW members had difficulty producing the nocturnal convective activity and if and how this can be fixed in future storm scale models. Lastly, other output fields, such as precipitation forecasts, should be used to further evaluate the sensitivities of storm scale models.

Acknowledgement: A special thanks to David Bright, my mentor, Steve Weiss, Jack Kain, Mike Coniglio, Jason Levit, Keli Tarp, and any additional SPC staff not mentioned for their help this past summer with the data and equipment. Thanks to the Ernest F. Hollings Undergraduate Scholarship Program for this wonderful opportunity.