19B.1 Application of object-oriented verification techniques to ensemble precipitation forecasts

William A. Gallus, Jr.
Iowa State University

1. Introduction

Mass et al. (2002), among others, have discussed several problems, such as a double penalty, that occur with the use of traditional point-to-point verification measures to evaluate precipitation forecasts from fine grid spacing models. To provide more informative measures of forecast performance that better reflect the quality of finer grid forecasts, several new spatial verification techniques have been proposed including object-oriented verification. Object-oriented approaches compare the properties of matched forecast and observed objects, where the object may be, for instance, a precipitation system determined using rainfall or reflectivity data. Object oriented techniques verify the location, size, shape, intensity, and other attributes of the object, and are therefore very intuitive in their interpretation (Ebert and Gallus 2009). One of the first object-oriented approaches developed was the Contiguous Rainfall Area (CRA) method (Ebert and McBride 2000) which was later used to explore systematic model biases in prediction of Central U.S. mesoscale convective systems (Grams et al. 2006). More recently, the Method for Object-based Diagnostic Evaluation (MODE; Davis et al. 2006a, b), was developed and included as part of a community verification system known as MET (Model Evaluation Tools; http://www.dtcenter.org/met/users).

These object-oriented approaches have been appliled to deterministic forecasts, but as of yet, little work has been done to explore how they might best be used with ensemble forecasts. In this paper, both CRA and MODE are applied to two different sets of ensembles to examine how closely the behavior of the object parameters matches results found from traditional ensemble spread and skill measures applied to these two sets of ensembles. The first set was used by Clark et al. (2008) to compare the temporal evolution of skill and spread in an ensemble using mixed physics and mixed models along with no perturbation of initial conditions (ICs) or lateral boundary conditions (LBCs) with an ensemble having fixed physics but perturbed ICs and LBCs. The second set of ensembles, examined by Clark et al. (2009), was used to compare skill and spread between a relatively coarse grid spacing ensemble with 15 members and a finer grid spacing ensemble with only 5 members.

2. Data and Methodology

To examine the use of object-oriented verification methods on ensemble 6-hourly accumulated precipitation forecasts, two different sets of ensemble forecasts from the Weather Research and Forecasting (WRF; Skamarock et al. 2001) model were evaluated. The first set included an 8 member 15 km grid spacing ensemble that used unperturbed ICs and LBCs with mixed physics and dynamic cores (hereafter Phys), and another 8 member 15 km ensemble using a fixed dynamic core and physics package but perturbed ICs and LBCs (hereafter IC/LBC). Clark et al. (2008) compared the precipitation forecasts of these two ensembles, integrated for 120 hours for 72 cases, using traditional verification metrics and found that the spread and skill of the two ensembles was initially comparable but after roughly 24 hours, the lack of perturbed LBCs reduced the growth of spread in Phys so that better spread and skill were found at later times in the IC/LBC ensemble. In addition, Clark et al. (2008) noted that a diurnal signal reflecting heavier nocturnal precipitation in the region could be seen in some of the traditional verification measures. In the present study, both MODE and CRA were applied to the first 60 hours of the forecasts from both ensembles to determine if the object parameters identified by both approaches behaved in a similar manner to the traditional spread and skill measures, and to examine how the diurnal precipitation cycle influenced the object parameters. As in Clark et al. (2008), both the WRF forecasted precipitation and Stage IV observations used for verification were remapped to a 10 km grid before being input to CRA and MODE.

In both CRA and MODE, a threshold of 6.25 mm was used to evaluate the 6 hour precipitation forecasts. In CRA, correlation coefficient maximization was used to obtain the best fit of the forecast with observations. The object parameters of rain rate, rain area, rain volume, and displacement error were analyzed.

The second set of ensembles evaluated included 5 members of a 10 member 4 km grid spacing WRF ensemble (hereafter ENS4) run by the Center for the Analysis and Prediction of Storms (CAPS; Xue et al. 2007) for the 2007 National Oceanic and Atmospheric Administration (NOAA) Hazardous Weather Testbed Spring Experiment, and 5 members of a 15 member 20 km grid spacing WRF ensemble (hereafter ENS20) run for the same 23 cases (Clark et al. 2009). All of these ensembles were constructed using both mixed physics and perturbed ICs and LBCs. For ENS4, the control member (CN) used the 2100 UTC analysis from NCEP's 12 km grid spacing North American Model (NAM; Janjic 2003) for ICs, and the 1800 UTC NAM forecasts for LBCs. The other members used perturbations extracted from the 2100 UTC NCEP SREF WRF-ARW (Advanced Research WRF) and WRF-NMM (Nonhydrostatic Mesoscale Model) members which were added to the 2100 UTC NAM analysis with the corresponding SREF forecasts used for LBCs (3-hr updates). The 5 members analyzed from the 20 km ensemble were those members having the best statistical consistency, as in Clark et al. (2009). The physical parameterizations varied among the ensemble members.

Corresponding author address: William A. Gallus, Jr., ISU, 3010 Agronomy Hall, Ames, IA. Email: wgallus@iastate.edu.

Clark et al. (2009) found in a comparison of these two ensembles using traditional measures that the explicitly-resolved convection in ENS4 led to a much better representation of the diurnal cycle than in ENS20 whose members used convective parameterizations. Possibly because of the better diurnal signal, ENS4 was more skillful than ENS20, even when the 4 km ensemble's 5 members were compared to the full 15 members of the 20 km ensemble. Spread was also found to increase more rapidly with time in ENS4 than in ENS20. MODE was used (again with a threshold of 6.25 mm) to evaluate the rainfall systems in these 33 h forecasts to determine if object parameters also reflected the improved forecast of the diurnal signal in ENS4, and showed the same differences in spread growth. As in Clark et al. (2009), the comparison of the two ensembles was performed on a 20 km grid that was basically a 2000 x 2000 km subset of the coarser ensemble domain.

3. Comparison of the mixed physics ensemble and mixed IC/LBC ensemble

Using the first 60 hours of the forecasts from Clark et al. (2008), the standard deviations (SDs) within the 8 member ensembles for several object parameters computed by CRA and MODE were compared. SDs were used as a measure of spread, and data were examined as a function of forecast hour from the first 6 hours through the 54-60 h forecast. The standard deviations were computed for parameters valid for individual objects (systems). For CRA and MODE, a system had to be depicted in at least 4 of the 8 ensemble members to be included in the analysis. A Welch two-sided t-test was used within the R statistical package to test for statistical significance of differences in the SDs between the two ensembles. No differences identified by MODE were statistically significant, but for some parameters in CRA, the differences were significant.

Standard deviations for rain rate show a more consistent tendency for increases with time in IC/LBC than in Phys (Fig. 1). In fact, the nearly 11% difference in the relative rate of increase (slope normalized by the average rain rate in all curves at all times) of a best-fit line between the two ensembles in the CRA data is the largest difference among the four parameters examined in the present study. An analysis of covariance was performed to test for the statistical significance of differences in the slopes of the best-fit lines for the two ensembles. The differences in the slopes for the two ensembles were stastically significant for both the CRA and MODE results. During at least the first five forecast periods, both CRA and MODE indicate greater SDs in Phys than in IC/LBC, but by the last forecast period, both techniques show a larger SD in IC/LBC than in Phys. Some of these differences in the CRA results were statistically significant at the 95% confidence level. This result implies that a mixture of different physical schemes is necessary to result in more variability in rain rates until enough time has passed that differences in LBCs likely affect the atmospheric conditions contributing to precipitation systems, and hence rain rates, in the IC/LBC members. Although not shown, it should be pointed out that observed rain rates evidenced a diurnal cycle with maxima in the 00-06, 24-36, and 48-60 hour periods, and minima during hours 06-18, and 36-48. The model forecasts (not shown) missed the first diurnal peak, possibly evidence of spin up problems during the 00-12 h period, but did capture the other extrema in the diurnal cycle, albeit with less amplitude than that observed (e.g., peak variation in Phys of 20.1 mm, and IC/LBC of 18.0 mm compared to observed variations of 20.6 mm and 31.5 mm). The diurnal signal is most apparent in the SDs for Phys from CRA (Fig. 1), and is stronger in the Phys ensemble than in the IC/LBC ensemble using both techniques.

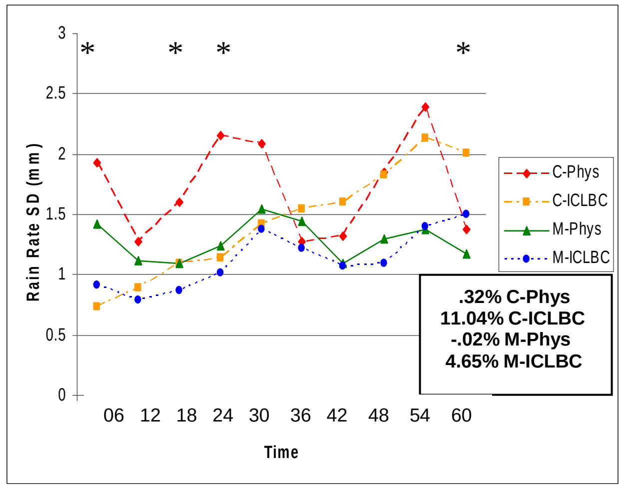

Figure 1: Standard deviation of 6-hr rainfall (mm) among the 8 ensemble members of Phys (CRA results in red, MODE in green) and IC/LBC (CRA orange, MODE blue) as a function of forecast hour (time). Differences in CRA results between Phys and IC/LBC statistically significant with p values less than .05 shown with asterisks. Slope of best-fit line for each set of data, expressed as percentage change relative to average rain rate SD, is shown in inset (boldfaced when differences between Phys and IC/LBC are significant with $p < .05$).

Standard deviations for rain volume from both CRA and MODE do show some diurnal signals in both Phys and IC/LBC (Fig. 2). In addition, the MODE output clearly shows a faster rate of growth for SDs in IC/LBC than in Phys, and the MODE results are statistically significant. Unlike with rain rate, no differences between IC/LBC and Phys were statistically significant. Clark et al. (2008) showed a strong diurnal cycle in observed 3-hourly rain volume (maxima at hours 00-06, 24-30, and 48-54) that also occurred with much smaller amplitude in both ensemble forecasts. That study also showed that the Phys ensemble members had larger 3-hourly rain volumes than the members of IC/LBC and that Phys possessed larger spread for rain volumes compared to IC/LBC. The larger volumes in Phys in Clark et al. (2008) correspond with larger SDs for Phys found in the present study.

Standard deviation of areal coverage of rainfall in terms of 10 x 10 km grid boxes is shown in Fig. 3. The MODE results differ from the CRA results in the first 18 hours with CRA showing relatively constant SDs with time while the MODE results depict a decrease. After the 12-18 h period, though, both techniques show a general increase with time, with the bigger growth in standard deviation happening in the IC/LBC ensemble. Both MODE and CRA show a statistically significantly larger growth in spread for IC/LBC. In the CRA results, some of the differences between the two ensembles are also statistically significant at the later times. Overall, SDs for rain area increase more rapidly with time than for the other parameters examined. It should be noted that the observed areas strongly reflected a diurnal cycle with maxima/minima at roughly the same times as the peaks in rain rate (not shown). Maximum areal coverage was 2-3 times that of the minimum coverage. Similar to rain rate, and even moreso rain volume, the amplitude of the cycle in the forecasts was greatly damped, especially in the IC/LBC ensemble (not shown). The SDs in Phys depict more of a diurnal cycle than those of IC/LBC. It is possible the cycle in IC/LBC is hidden somewhat by the faster rate of growth of SDs with time in that ensemble than in Phys.

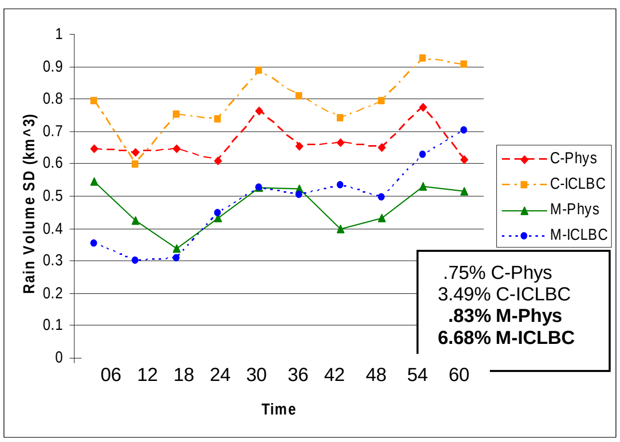

Figure 2: As in Figure 1 except for rain volume (km^3).

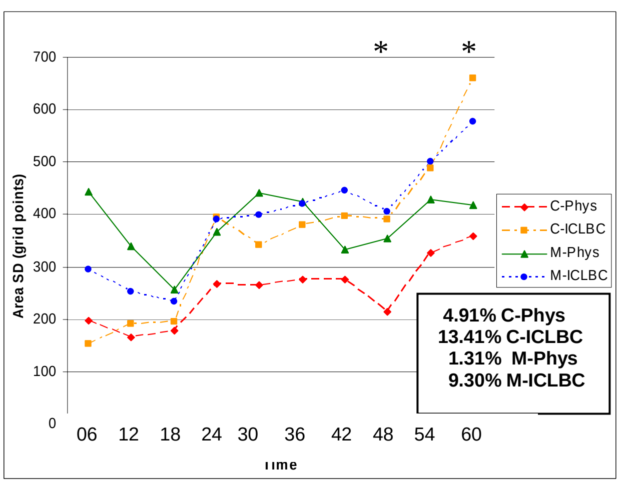

Figure 3: As in Fig. 1 except for areal coverage (number of 10 x 10 km grid boxes) of rainfall.

As with the other parameters, SDs of displacements from both techniques (Fig. 4) suggest a faster rate of growth for IC/LBC than for Phys, and for the MODE results, this difference is significant. In addition, at most times, the SDs are larger in IC/LBC than in Phys. Perhaps most noticeable in Fig. 4 is the large difference in the magnitudes of the SDs between the two techniques. It is likely that these large differences in values are due to the fact that systems are only matched in CRA when they are contiguous, whereas in MODE, systems are matched if the interest parameter is greater than some threshold. Thus, some systems are matched in MODE even though they are not contiguous and may be separated by some distance. Analysis of the first and last 6-h forecast periods from the 72 cases for a few of the ensemble members suggests that nearly half of the matched objects in MODE do not exhibit overlap. Average displacements for the two techniques (not shown) support this theory with CRA values generally between 100-150 km, and MODE values between 200-250 km at all times. The differences in the way the schemes operate should lead to larger average displacement errors and standard deviations in MODE than in CRA. In all of the analyses performed in this study, false alarm and missed systems are not included in the computation of parameters.

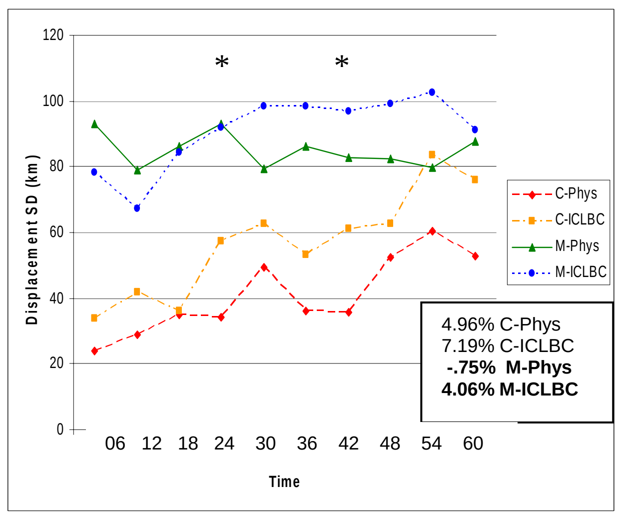

Figure 4: As in Fig. 1 except for displacement error (km).

In summary, the trends toward increasing spread with time in the IC/LBC ensemble with less increase in the Phys ensemble seen in Clark et al (2008) are observable in the four object parameters examined, and are especially noticeable in both the MODE and CRA results for SDs of areal coverage. This result suggests that spread in areal coverage of forecasted precipitation systems is more sensitive to whether or not perturbed LBCs are used than spread in average rain rates, rain volumes, or displacement errors. For all four parameters, MODE finds spread to be greater in Phys at the early times and greater in IC/LBC at the later times; CRA only shows this for two parameters. Although the two techniques do not produce identical results, they do show the same general trends. CRA better shows differences between the two ensembles for the SDs of rain rate, while MODE shows more of a difference for SDs of rain volume. The diurnal cycle with precipitation being more abundant during the overnight hours has some influence on the SDs of these parameters, especially for rain rate, volume, and areal coverage.

4. Comparison of 4 km and 20 km grid spacing ensembles

To determine whether or not object parameters show the trends found in Clark et al. (2009) in a comparison of two ensembles using different grid spacing, MODE was used on 6 hour accumulation periods covering 00-06, 06-12, 12-18, 18-00 and 00-06 UTC, or the 3-9, 9-15, 15-21, 21-27, and 27-33 hour forecast periods, for the cases evaluated in that study. The comparisons used five members for each ensemble, with the rainfall input to MODE on a 20 km grid. Because the results discussed in section 3 showed that both object-oriented verification techniques depicted the same general trends, the CRA method was not used on ENS4 and ENS20 output.

The rain area (for amounts exceeding 6.25 mm) in terms of grid boxes (20 x 20 km) averaged for all objects identified in all ensemble members and SD of rain area, from both ENS4 and ENS20, and the observed rain area for each 6 hr period are shown in Fig. 5. The diurnal minimum during the 12-18 UTC period can be seen in the observations, with higher values during the 00-12 UTC period. Both ensembles incorrectly show a peak during the 18-00 UTC period, and both show an overestimate (high bias) at all times. However, ENS4 has less of a high

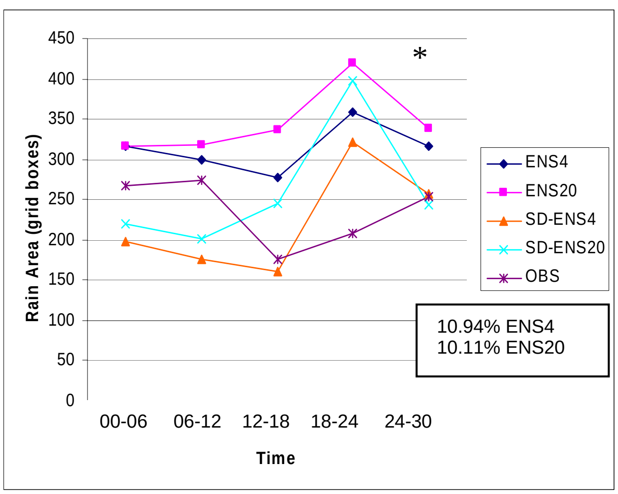

Figure 5: Rain area (in 20x20 km grid boxes) averaged for all objects in all ensemble members from MODE as a function of forecast hour (time) for ENS4 (dark blue) and ENS20 (pink), along with the observed value (purple) and the standard deviations for ENS4 (orange) and ENS20 (cyan). Slope of best-fit line for SD data, expressed as percentage change relative to average rain area SD, is shown in inset. Asterisk at top indicates that errors compared to observations for ENS4 were statistically significantly less (p <.05) than for ENS20.

bias is statistically significantly closer to observations than ENS20 during the final forecast period, and does show a minimum during the 12-18 UTC period, unlike the 20 km ensemble. Both ensembles disagree most with observations during the daylight hours (12-00 UTC).

SDs for the ensembles generally follow the same trends as the average rain area, with SDs for ENS20 larger than those for ENS4. During the last two periods, the rate of growth of spread is slightly larger in ENS4 than in ENS20. Additionally, the slope of the best fit line applied to the data from all valid times is slightly greater for ENS4 than for ENS20. This faster growth of spread is consistent with Clark et al's (2009) finding of faster growth in ensemble variance in ENS4 compared to ENS20. Unlike the Equitable Threat Score (ETS) results discussed in Clark et al., the biggest improvements in the 4 km depiction of rain area relative to ENS20 occurred during the 12-18 and 18-00 UTC periods, and not in the 06-12 UTC period.

Average rain rates for the ensembles and observations, along with the SDs are shown in Fig. 6. Both ensembles tended to predict the rain rate to within 10% of the observed value, much better agreement than found for rain area (Fig. 5). At all times except for the diurnal minimum (12-18 UTC), the two ensembles slightly overestimated the rates. The models correctly depicted the times of maxima and minima. The 4 km results were closer to the observed rates during the 00-06 UTC period, and then again in the last 12 hours of the simulation. At all times, ENS4 had more spread than ENS20, with a hint of faster growth of spread during the last 6-12 hours of the forecast. However, the slopes of the best fit lines for all of the data indicated less growth with time than for the SDs of rain area, and slightly faster growth in spread for ENS20 than for ENS4, a result opposite to that for rain area (Fig. 5) and what was found in Clark et al. (2009). SDs were no more than 25% of the magnitude of rain rates, a much smaller fraction than for rain area (Fig. 6) where the SDs always exceeded 50% of the magnitudes of the average areas.

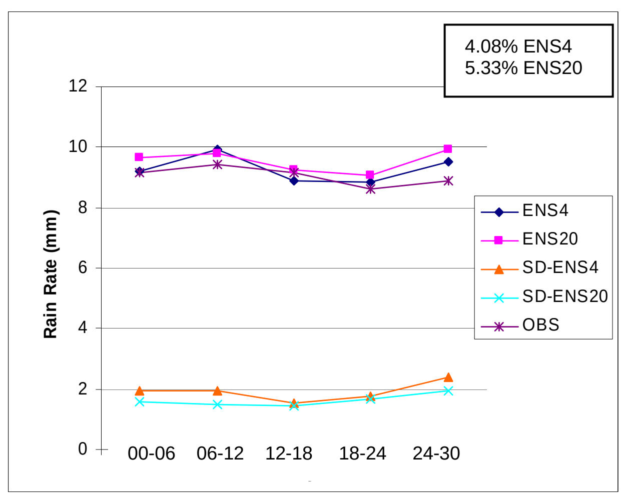

Figure 6: As in Fig. 5, except for rain rate (mm)

For rain volume (Fig. 7), as with rain area, both ensembles showed a high bias at all times, with particularly large errors during the 12-00 UTC period. The diurnal cycle is apparent in the observations, and ENS4 does a better job of showing the diurnal cycle, although both ensembles are too quick to increase the rain volume during the afternoon (18-00 UTC). The SDs follow the same trends as with rain area, with ENS20 having a greater SD at all times until the last 6 hours, when ENS4 may be evidencing the faster error growth discussed in Clark et al. (2009). The slopes of the best fit lines show the biggest difference in rate of growth with time for this parameter, with ENS4 having a noticeably larger slope. Clark et al. also noted that in Hovmoller diagrams averaged over the model domain, the ENS20 mean computed using probability matching (PM; Ebert 2001) appeared to generate the diurnal maximum of day 2 too early and too intensely, and this result may be reflected in the peak in volume occurring in the ENS20 data between 18 and 00 UTC (Fig. 7). Best agreement between both ensembles and the observations does occur during the 06-12 UTC period, in agreement with ETS values for the PM means in Clark et al. (2009). Because errors and SDs were relatively small for rain rate (Fig. 6), the larger values occurring for rain volume indicate that areal coverage of rainfall is a more troublesome forecasting challenge for the ensemble members.

Average displacements and SDs of displacement for the two ensembles are shown in Fig. 8. ENS4 has less displacement error at all times. Both ensembles have comparable SDs, roughly 50% of the magnitude of the displacements, and unlike the other three parameters examined, growth in SDs with time is negligible. The displacement errors do not appear to reflect the diurnal cycle. Displacements in ENS4 were usually in the S or SW direction (not shown) through the 27 h forecast (00 UTC) and then toward the NNE after that time (i.e., mean position of forecasted objects usually was SW of the observed one prior to hour 27). For ENS20, there were no systematic trends in the displacement direction.

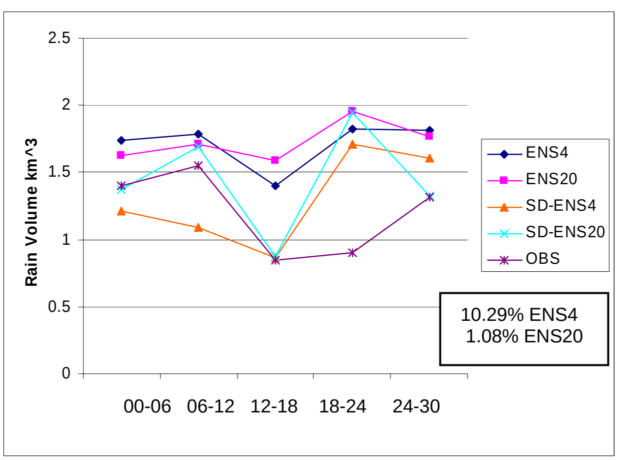

Figure 7: As in Fig. 5, except for rain volume in km^3.

In summary, the object parameters evaluated from MODE often supported the conclusions based on traditional verification approaches applied to the 4 and 20 km ensembles. The object parameters indicated that ENS4 better depicted the diurnal cycle, and generally had smaller errors at most times for most parameters than ENS20. There was also some evidence of the faster error growth found by Clark et al (2009) with SDs growing more rapidly in the ENS4 parameters than in the ENS20 ones. The analysis of object parameters suggests that average rain rate is forecasted better than area, location and volume. Also, variability in rain rate among members is less than it is for the other parameters.

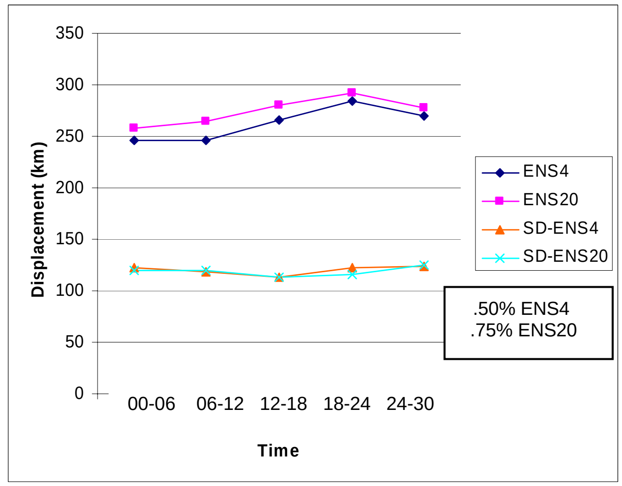

Figure 8: As in Fig. 5, except for displacement in km, with no observations plotted.

5. Use of object-oriented verification parameters in ensemble forecasting

To investigate how information from object-oriented techniques might assist forecasters, a series of tests were performed. In one test, the percentage of time that the observed rate, volume, and areal coverage fell within the spread of the two 8 member ensembles was examined during all 6-h forecast periods, which were grouped into day 1, 2 and 3 forecasts. Figure 9 shows that this approach worked much better for predicting rain rate and volume than for areal coverage. For rate and volume, the two ensembles captured the observed values roughly 50% of the time; for areal coverage the figure was closer to 10%. For all three parameters, increasing skill with time was more pronounced in IC/LBC than in Phys, a result likely influenced by the faster growth in spread in IC/LBC (Clark et al. 2008).

In a second test, output was examined to see if a spread-skill relationship where large spread was associated with less skillful forecasts existed in the object parameters from the ensembles. Fig. 10 shows the time evolution for the Phys ensemble of mean absolute error for rain volume, rate, and areal coverage for those systems that had large SDs among the members and those that had small values. Large and small were defined to be greater than 150% of the average SD or less than 50% of the average, respectively. Rain volume and areal coverage show a clear separation with much more accurate forecasts in the events where spread is relatively small. Rain rate does not show the relationship as definitively, although for a majority of the time, it still applies. However, at hours 00-06, 18-24, and 48-60, the forecast skill either does not vary much with SD or the more accurate forecast is associated with larger SDs. It is not entirely clear why rain rate would behave differently, although it should be pointed out that the SDs for rain rate are a much smaller fraction of the average magnitudes than those for the other parameters (not shown). Average

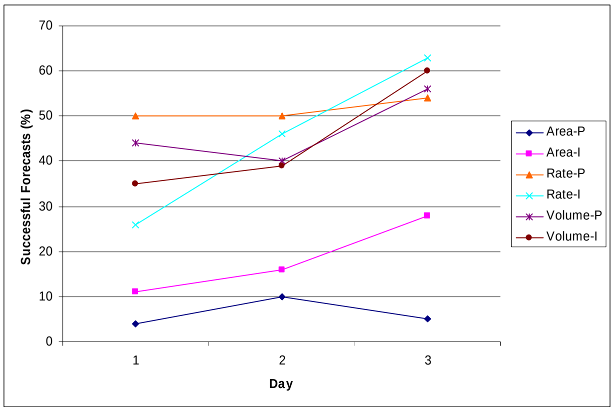

Figure 9: Percentage of time that observed values of rain volume (purple for Phys, brown for IC/LBC), rate (orange for Phys, cyan for IC/LBC), and areal coverage (blue for Phys, pink for IC/LBC) fell within the minimum and maximum predicted by the two ensembles, based on CRA output. Day 1 refers to 6h forecast periods in the first 18 hours of the forecast, day 2 refers to those in the 18-42 h period, and day 3 those in the 42-60 h period.

rain rates at all times for both ensembles are around 12 mm, so that the SDs (Fig. 1) are roughly 10% of the average values. Rain volumes are typically around 1 km^3 (not shown), so that SDs (Fig. 2) are often over 50% of these values. Likewise, SDs for areal coverage (Fig. 3) and displacement (Fig. 4) are often 50% or more of the typical values (not shown) of around 700 (MODE) – 800 (CRA) points and 100 km (CRA) - 200 km (MODE). These results in general imply that forecasters may be able to establish a confidence level for their forecasts of some object parameters using the SDs from ensembles. The same general tendencies were found for IC/LBC (not shown), with clearly better forecasts of rain volume and areal coverage when spread was small, but mixed results for rain rate.

Finally, a third test was performed to investigate if the average of a parameter value from the set of ensemble members would be a better forecast than the parameter value from the Probability Matched (PM) ensemble mean (Ebert 2001). Probability matching is a technique that looks at the full distribution of forecast values from any ensemble of *n* members, and then assigns every *n*th value to the simple ensemble mean. The PM approach generally reduces the usual problem of high biases for light amounts that occurs when a simple mean is taken for ensemble forecasts of parameters like precipitation.

For the 8 member ensembles, the PM approach resulted in similar skill to the approach of applying CRA to each member and then averaging the output for object parameters (results not shown). For ENS4 and ENS20, however, MODE results showed that for some parameters

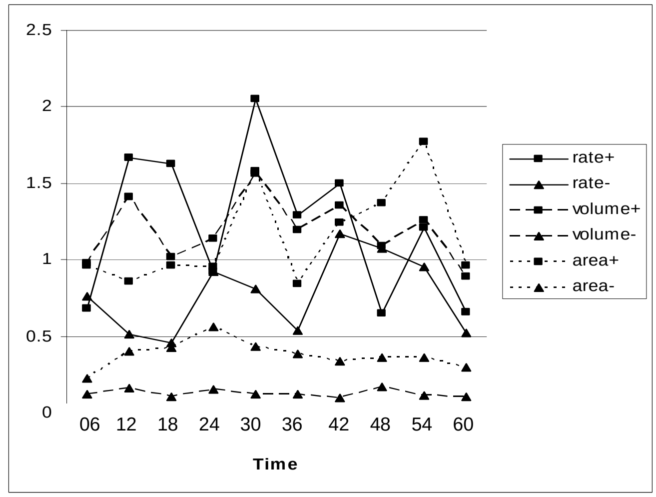

Figure 10: Mean absolute error for volume (km^3), rate (*2.54 mm) and areal coverage (*1000 grid boxes) of precipitation from CRA applied to Phys as a function of time for cases with standard deviations over 150% of the mean (squares with solid line for rate, dashed line for volume, and dotted line for area), and less than 50% of the mean (triangles with solid line for rate, dashed line for volume, and dotted line for area).

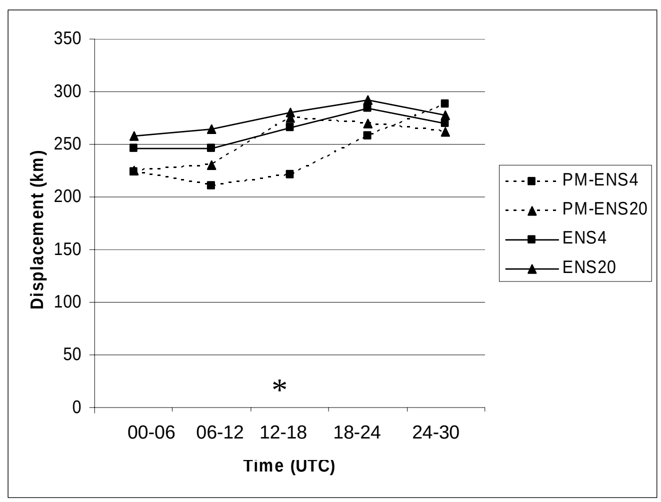

Figure 11: Displacement error (km) for the average of 4km ensemble members (solid line with squares) and the PM ensemble mean (dotted with squares), and for the average of the 20 km ensemble members (solid with triangles) and the 20 km PM ensemble mean (dotted with triangles), based on MODE output. Asterisks at top indicate times when PM approach at 20 km was statistically significantly better ($p < .05$) compared to observations than ENS20 approach; asterisks at bottom indicate the same but for 4 km results.

the PM approach might work better, and for others, the averaging of MODE output from each member would be better. In Fig. 11, the results for displacement can be seen. For ENS20, the PM ensemble mean precipitation field always yields a smaller displacement error than the average of the ensemble members. For ENS4, the probability matched ensemble mean value is better in the first 4 evaluation periods, but shows much faster error growth over time, so that the average of the individual members' displacements becomes a better forecast during the final 6 hour period. Such a trend is not as apparent in the 20 km ensemble output.

Different behavior is noted for the rain areas (Fig. 12) and rates and volumes (not shown). For both ensembles, the forecasted rain areas are much closer to the observed values when the average of the areas from the individual ensemble members are used compared to the areas determined from the PM ensemble mean. For the 4km results, the differences are statistically significant at almost all times; for the 20km results, p values are less than .05 only at the 6-12 h period, but are less than .10 (not shown) during periods 12-18 and 24-30 as well. As was the case with Phys and IC/LBC, the areas in the PM mean forecast were larger than the average areas from the ensemble members. Because the ENS4 and ENS20 forecasts had a more persistent problem with overestimates of rainfall coverage, the PM forecasts of area were always worse than those obtained from averages of the individual ensemble members. The temporal trends in the average values of area still differ substantially from the observed diurnal cycle but do appear to be slightly more realistic than the trends associated with the PM forecast.

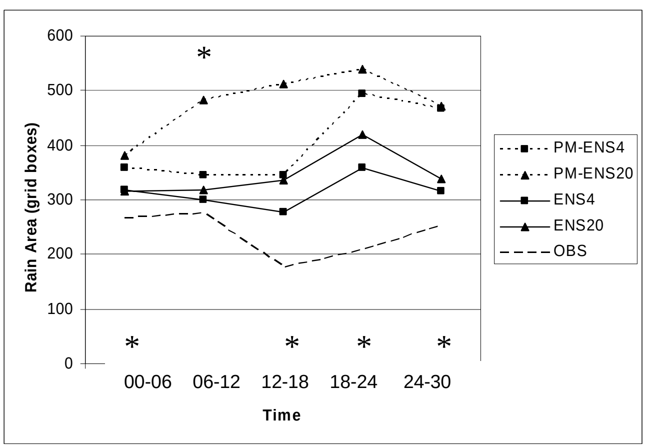

Figure 12: As in Fig. 11 except for rain area (20 x 20 km grid boxes) with observations also shown (dashed line). Asterisks at top indicate times when PM approach at 20 km was statistically significantly worse ($p < .05$) compared to observations than ENS20 approach; asterisks at bottom indicate the same but for 4 km results.

6. Summary and Conclusions

The use of object-oriented verification approaches to evaluate and enhance ensemble forecasts was tested by using both CRA and MODE on two sets of ensembles. The first set included an 8 member ensemble with mixed physics and dynamic cores with unperturbed ICs and LBCs (Phys) and an 8 member ensemble with fixed physics and perturbed ICs and LBCs (IC/LBC). Clark et al. (2008) found using traditional spread and skill measures that spread increased much faster in IC/LBC compared to Phys so that both spread and skill were better in IC/LBC than in Phys after roughly 24-30 hours. The second set of ensembles included 5 members of a 4 km ensemble (ENS4) and 5 members of a 20 km ensemble (ENS20). Clark et al. (2009) found that a diurnal signal in precipitation was better depicted in ENS4, and this may have partly accounted for better skill measures for ENS4 compared to ENS20. In addition, spread grew faster with time in ENS4.

Both CRA and MODE showed that in four object parameters studied, rain rate, volume, areal coverage and displacement error, greater increases in spread with time occur in IC/LBC than in Phys, agreeing with Clark et al. (2008). This trend was particularly noticeable in areal coverage of precipitation systems, suggesting that spread in areal coverage of forecasted precipitation systems is more sensitive to whether or not perturbed lateral boundary conditions are used than spread in average rain rates, rain volumes, or displacement errors. Although the two object-oriented techniques do not produce identical results, they do show the same general trends. The diurnal cycle with precipitation being more abundant during the overnight hours has some influence on the SDs of the object parameters examined, especially rain rate, volume, and areal coverage.

In a comparison of object parameters derived from MODE for ENS4 and ENS20, ENS4 was found to better depict the diurnal cycle, and generally had smaller errors at most times for most parameters than ENS20. There was also some evidence of the faster error growth found by Clark et al (2009) with standard deviations growing more rapidly in the ENS4 parameters than in the ENS20 ones. Standard deviations were much smaller for rain rate than for other parameters, and agreement with observations was also best for rain rate.

Several tests were performed also to examine methods of using object-oriented verification output to assist forecasters. It was found that predictions of areal coverage of precipitation systems are more accurate when based on the average of the predicted areas from each ensemble member as opposed to using a PM ensemble mean forecast input into the object-oriented techniques. For the other parameters, rain rate, volume, and displacement, differences in the skill of both approaches were less substantial.

7. Acknowledgments

This research was funded in part by NSF grant ATM-0537043 and the WRF Developmental Testbed Center. Special thanks are given to John Halley Gotway and Randy Bullock at NCAR for assistance with MODE and methods of analyzing MODE output, and to Jon Hobbs for help with statistical analysis.

8. References

Clark, A. J., W. A. Gallus, Jr., and T.-C. Chen, 2008: Contributions of mixed physics versus perturbed initial/lateral boundary conditions to ensemble-based precipitation forecast skill. *Mon. Wea. Rev*., 136, 2140-2156.

Clark, A. J., W. A. Gallus, Jr., and M. Xue, 2009: A comparison of precipitation forecast skill between small convection-resolving and large non-convection-resolving ensembles. *Mon. Wea. Rev.* (in press).

Davis, C., B. Brown, and R. Bullock, 2006a: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev*., 134, 1772-1784.

Davis, C., B. Brown, and R. Bullock, 2006b: Object-based verification of precipitation forecasts. Part II: Application to convective rain systems. *Mon. Wea. Rev.*, 134, 1785-1795.

Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.*, 129, 2461-2480.

Ebert E. E., and J. L. McBride, 2000: Verification of precipitation in weather systems: determination of systematic errors. *J. Hydrology*, 239, 179-202.

Ebert, E. E., and W. A. Gallus, Jr., 2009: Toward better understanding of the contiguous rain area (CRA) method for spatial forecast verification. Submitted to *Wea. Forecasting*.

Grams, J.S., W.A. Gallus, L.S. Wharton, S. Koch, A. Loughe, and E. E. Ebert, 2006: The use of a modified Ebert-McBride technique to evaluate mesoscale model QPF as a function of convective system morphology during IHOP 2002. *Wea. Forecasting*, 21, 288-306

Janjic, Z. I., 2003: A nonhydrostatic model based on a new approach. *Meteorol. Atmos. Phys*., 82, 271-285.

Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bull. Amer. Meteor. Soc*., 83, 407-430.

Skamarock, W. C., J. B. Klemp, and J. Dudhia, 2001: Prototypes for the WRF (Weather Research and Forecasting) model. Preprints, *Ninth Conf. on Mesoscale Processes*, Fort Lauderdale, FL, Amer. Meteor. Soc, J15. [Available online at **http://ams.confex.com/ams/pdfpapers/23297.pdf**].

Xue, M., F. Kong, D. Weber, K. W. Thomas, Y. Wang, K. Brewster, K. K. Droegemeier, J. S. Kain, S. J. Weiss, D. R. Bright, M. S. Wandishin, M. C. Coniglio, and J. Du, 2007: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2007 spring experiment. *22nd Conf. Wea. Anal. Forecasting/18th Conf. Num. Wea. Pred.*, Park City, UT, Amer. Meteor. Soc., CDROM 3B.1.