DTC Visitor Program Final Report

Use of the CLUE to examine importance of mixed physics in ensembles

William A. Gallus, Jr., Iowa State University

Introduction

A well-known problem in ensembles designed to predict convection on fine grids has been the lack of spread often present in the ensembles. Some prior works have suggested that the use of mixed physics or mixed models increases the spread and may lead to better forecasts (e.g., Chen et al. 2009, Li et al. 2017). Clark et al. (2010) suggested the role of mixed physics in an ensemble will vary depending on the variables examined. However, there is a danger of clustering with some of these approaches (e.g., all members with the same microphysics scheme will look more alike despite using different planetary boundary layer schemes than members using different microphysics), and one quality of a good ensemble, that each member be equally likely to verify, can be violated (e.g., Stensrud et al. 2000, Johnson et al. 2011). Thus, an ideal approach might be one where sufficient spread is present from the techniques creating the perturbed initial conditions/lateral boundary conditions (IC/LBCs) while avoiding the use of mixed physics.

The Community Leveraged Unified Ensemble (CLUE) offered a good opportunity to explore the impact of using mixed physics in an ensemble, because two of the sub-ensembles, Core and S-Phys, were identical except for Core also including mixed physics. In addition, a subset of members allowed the examination of changes in microphysics scheme alone. The two goals of this particular project were (1) to determine the impact of adding mixed physics to an ensemble that already included mixed IC/LBCs, and (2) to determine systematic differences in the use of different microphysics. The original plan was to use the output of CLUE from only 2016. However, because the time period of the project extended through March 2018, I was able to also look at the same sub-ensembles for 2017 for task 1. Because a few small changes were made in the ensembles between the two years, the use of output from both years allowed some extra insight into the impacts of a few other model changes. In addition to impacts on forecasts of reflectivity and rainfall, the project also examined impacts on convective initiation.

Methodology

The primary focus of the study was on task 1, the impact of adding mixed physics to an ensemble, and to address this question, 9 members of the Core (Table 1) and S-Phys (Table 2) ensembles were examined in 2016 with all 10 members used in 2017 for Core (Table 3) and S-Phys (Table 4). In 2016, since S-Phys was missing member 6, member 2 was also eliminated from Core to allow an equal number of members to be compared. Member 2 was chosen to be eliminated so that all nine members used NAM output for IC/LBCs prior to the addition of any perturbations. In 2016, the S-Phys ensemble used the Thompson microphysics with the NOAH LSM and MYJ PBL schemes. These were also the schemes used in the Control member within Core. In Core, the varied microphysics included the P3, Milbrandt-Yau (MY) and Morrison schemes, and PBL scheme variations included MYNN and YSU. Both ensembles used a mixture of initial conditions and lateral boundary conditions with most initialized using the NAM and radar data assimilation via the ARPS 3DVAR system, but with variations from the control member coming through use of perturbations from the SREF added to the NAM initialization. Core member 2 (which was only used in 2017) differed in its IC/LBCs by using RAP analyses with GFS supplying the LBCs. In 2017, the primary changes were in the switch to the MYNN PBL scheme and RUC land surface scheme in the control member (1) of Core, and thus as well in all S-Phys members. In addition, ICs in members 7-10 of both Core and S-Phys were supplied by the RAP analysis instead of NAM, and the perturbations were applied to different members in 2017 than in 2016.

Member	IC	LBC	Microphysics	LSM	PBL	Model
1	NAMa+3DVAR	NAMf	Thmopson	NOAH	MYJ	arw
3	1+arw-p1_pert	arw-p1	P3	NOAH	YSU	arw
4	1+arw-n1_pert	arw-n1	MY	NOAH	MYNN	arw
5	1+arw-p2_pert	arw-p2	Morrison	NOAH	MYJ	arw
6	1+arw-n2_pert	arw-n2	P3	NOAH	YSU	arw
7	1+nnmb-p1_pert	nmmb-p1	MY	NOAH	MYNN	arw
8	1+nmmb-n1_pert	nmmb-n1	Morrison	NOAH	YSU	arw
9	1+nmmb-p2_pert	nmmb-p2	P3	NOAH	MYJ	arw
10	1+nmmb-n2_pert	nmmb-n2	Thompson	NOAH	MYNN	arw

Table 1: Specifications for the 2016 Core mixed physics ensemble. NAM refers to 12 km NAM output with "a" being analysis and "f" forecast. 3DVAR refers to ARPS3DVAR and cloud analysis. Model names appended with "pert" refer to perturbations extracted from a 16 km grid-spacing SREF member.

Member	IC	LBC	Microphysics	LSM	PBL	Model
1	NAMa+3DVAR	NAMf	Thompson	NOAH	MYJ	arw
2	1+arw-p1_pert	arw-p1	Thompson	NOAH	MYJ	arw
3	1+arw-n1_pert	arw-n1	Thompson	NOAH	MYJ	arw
4	1+arw-p2_pert	arw-p2	Thompson	NOAH	MYJ	arw
5	1+arw-n2_pert	arw-n2	Thompson	NOAH	MYJ	arw
7	1+nnmb-p1_pert	nmmb-p1	Thompson	NOAH	MYJ	arw
8	1+nmmb-n1_pert	nmmb-n1	Thompson	NOAH	MYJ	arw
9	1+nmmb-p2_pert	nmmb-p2	Thompson	NOAH	MYJ	arw
10	1+nmmb-n2_pert	nmmb-n2	Thompson	NOAH	MYJ	arw

Table 2: Specifications for the 2016 S-Phys single physics ensemble. Notations as in Table 1.

Member	IC	LBC	Microphysics	LSM	PBL	Model
1	NAMa+3DVAR	NAMf	Thompson	NOAH	MYJ	arw
2	RAPa+3DVAR	GFSf	Thompson	RUC	MYNN	arw
3	1+arw-p1_pert	arw-p1	P3	NOAH	YSU	arw
4	1+arw-n1_pert	arw-n1	MY	NOAH	MYNN	arw
5	1+nmmb-p1_pert	nmmb-p1	Morrison	NOAH	MYJ	arw
6	1+nmmb-n1_pert	nmmb-n1	P3	NOAH	YSU	arw
7	2+arw-p2_pert	arw-p2	MY	NOAH	MYNN	arw
8	2+arw-n2_pert	arw-n2	Morrison	NOAH	YSU	arw
9	2+nmmb-p2_pert	nmmb-p2	P3	NOAH	MYJ	arw
10	2+nmmb-n2_pert	nmmb-n2	Thompson	NOAH	MYNN	arw

Table 3: Specifications for the 2017 Core mixed physics ensemble. Notation as in Table 1, with RAPa referring to 13 km RAP analysis, and GFSf referring to 18 UTC initialized GFS forecasts.

Member	IC	LBC	Microphysics	LSM	PBL	Model
1	RAPa+3DVAR	GFSf	Thmopson	RUC	MYNN	arw
2	NAMa+3DVAR	NAMf	Thompson	RUC	MYNN	arw
3	1+arw-p1_pert	arw-p1	Thompson	RUC	MYNN	arw
4	1+arw-n1_pert	arw-n1	Thompson	RUC	MYNN	arw
5	1+nmmb-p1_pert	nmmb-p1	Thompson	RUC	MYNN	arw
6	1+nmmb-n1_pert	nmmb-n1	Thompson	RUC	MYNN	arw
7	2+arw-p2_pert	arw-p2	Thompson	RUC	MYNN	arw
8	2+arw-n2_pert	arw-n2	Thompson	RUC	MYNN	arw
9	2+nmmb-p2_pert	nmmb-p2	Thompson	RUC	MYNN	arw
10	2+nmmb-n2_pert	nmmb-n2	Thompson	RUC	MYNN	arw

Table 4: Specifications for the 2017 S-Phys single physics ensemble. Notations as in Table 3.

To evaluate the impacts of using mixed physics, MET (Model Evaluation Tools) and METviewer were used to verify the ensembles. Verification was performed using MRMS observations. Three fields were evaluated, 1-hour precipitation, 3-hour precipitation, and composite reflectivity (CREF). Several types of verification were performed. Traditional point-to-point measures such as Gilbert Skill Score and Bias were computed for each member of the two ensembles, and averages were taken of the members to evaluate how the mixed physics might be impacting general skill and areal coverage within its members compared to the S-Phys members. In addition, object-based spatial verification was performed on each member through the Method for Object-based Diagnostic Evaluation (MODE; Davis et al. 2006a, 2006b, 2009), and averages of MODE system parameters were computed for each ensemble, loosely following Gallus (2010). These parameters included area of the objects, median and 90th percentile values, intensity sum, and counts of objects. Finally, standard ensemble verification was performed on the probabilistic forecasts using measures such as ROC areas, reliability, and Brier skill scores. Statistical significance testing was performed for some average comparisons using the bootstrapping approach available in METviewer.

Originally, verification was performed on a subset of the full CLUE domain for 2016 alone to accelerate the running of MET. Later, however, the full domain MET output became available through the efforts of DTC personnel, and verification was then performed over the full domain for both 2016 and 2017.

To accomplish task (2), five CLUE member configurations that used different microphysics but were otherwise the same were examined for 2016 alone. These members were all initialized with the NAM model and radar data assimilation via the ARPS3DVAR system, with no perturbations, and used the NOAH LSM and MYJ PBL. The different microphysics among the five configurations were the Thompson, Morrison, MY, P3, and WSM6 schemes.

The SFE2016 ran from May 2 through June 3, with model output only available from the weekday portions of that period. Similarly in 2017, the project ran from May 1 through June 2. During the project, I discovered that a problem existed in some of the composite reflectivity data from runs using the MY microphysics scheme in 2016, and this reduced the size of the dataset. Likewise, in 2017, a problem prevented S-Phys from being run during the first part of the project. In the end, a total of 24 cases were available for comparisons of precipitation data in 2016, 17 cases for comparison of CREF, and only 11 cases in 2017 for both fields (the case size represents events for which output was available from both ensembles).

In additional to the comparisons that could be performed using MET, two other comparisons were made for the 2016 ensembles. First, since it was discovered that the Factor Separation Approach could not be performed because of the design of the members, impacts were isolated using configurations where only a single change had been made in the physics schemes, with no differences in the IC/LBCs. Fourteen such comparisons were possible and made using domain total 3 hour precipitation. In addition, a subset of 10 cases with relatively pristine convective initiation were examined to evaluate differences in the ensemble prediction of this initiation. Location and timing were studied using each member of both ensembles.

Results

The impacts of the use of mixed physics were determined using multiple verification strategies including point-to-point measures applied to individual members, the same metrics averaged for all ensemble members, MODE attributes, and traditional ensemble metrics making use of probability values. In the discussion below, emphasis will be on 1 hour precipitation and CREF. In general, 3 hour precipitation behaved similarly to 1 hour precipitation, except results exhibited more skill as would be expected for a longer time period.

a) Point-to-point metrics

Gilbert Skill Score (GSS) applied to both a threshold of 0.254 mm for 1 hour precipitation and 20 dBZ for CREF for the 2016 output reveals some interesting trends (Fig. 1). First, for precipitation, the spread is slightly larger in Core than in S-Phys, and the control member usually has the highest skill at all times (red curve). Ideally, each member of an ensemble should be equally likely to verify, so there should be very little spread in a parameter like GSS. However, for CREF, the spread is noticeably larger in Core, and this increased spread comes about by having several members that are performing much more poorly than any member of S-Phys. Again the control member usually has the highest skill. In 2017 (not shown) the differences in spread for CREF were greatly reduced.

Bias (not shown) for 1 hour precipitation thresholds of 2.54 and 6.35 mm shows more noticeably an increase in spread in the Core ensemble, along with all members of both ensembles usually having too large of areal coverage. The same trends occur in both 2016 and 2017. Bias for CREF clearly indicates more spread among the members of Core than S-Phys, with some differences between the 2016 output (Fig. 2) and 2017 output (Fig. 3). For a 20 dBZ threshold, P3 members all have a low bias while all other members of Core have biases greater than 1.0 with the control member usually having the highest bias. At this same threshold, all S-Phy members are more consistent with the control members. For a 40 dBZ threshold in 2016, the MY members have a very large bias and behave very differently from the other members, many of which have low biases at most times.

In 2017, Core again has a much larger spread among members than S-Phys, and for 20 dBZ, the control member is no longer usually the highest, implying a reduction in areal coverage due to the use of the different land surface and PBL scheme. The P3 members also do not have the low bias problem that was present in 2016, and all S-Phys members have lower bias values than in 2016. For 40 dBZ, the MY members still have a very high bias, and are joined by the P3 members. A large change is evident in S-Phys at 40 dBZ where a high bias is now present in most members at most times.

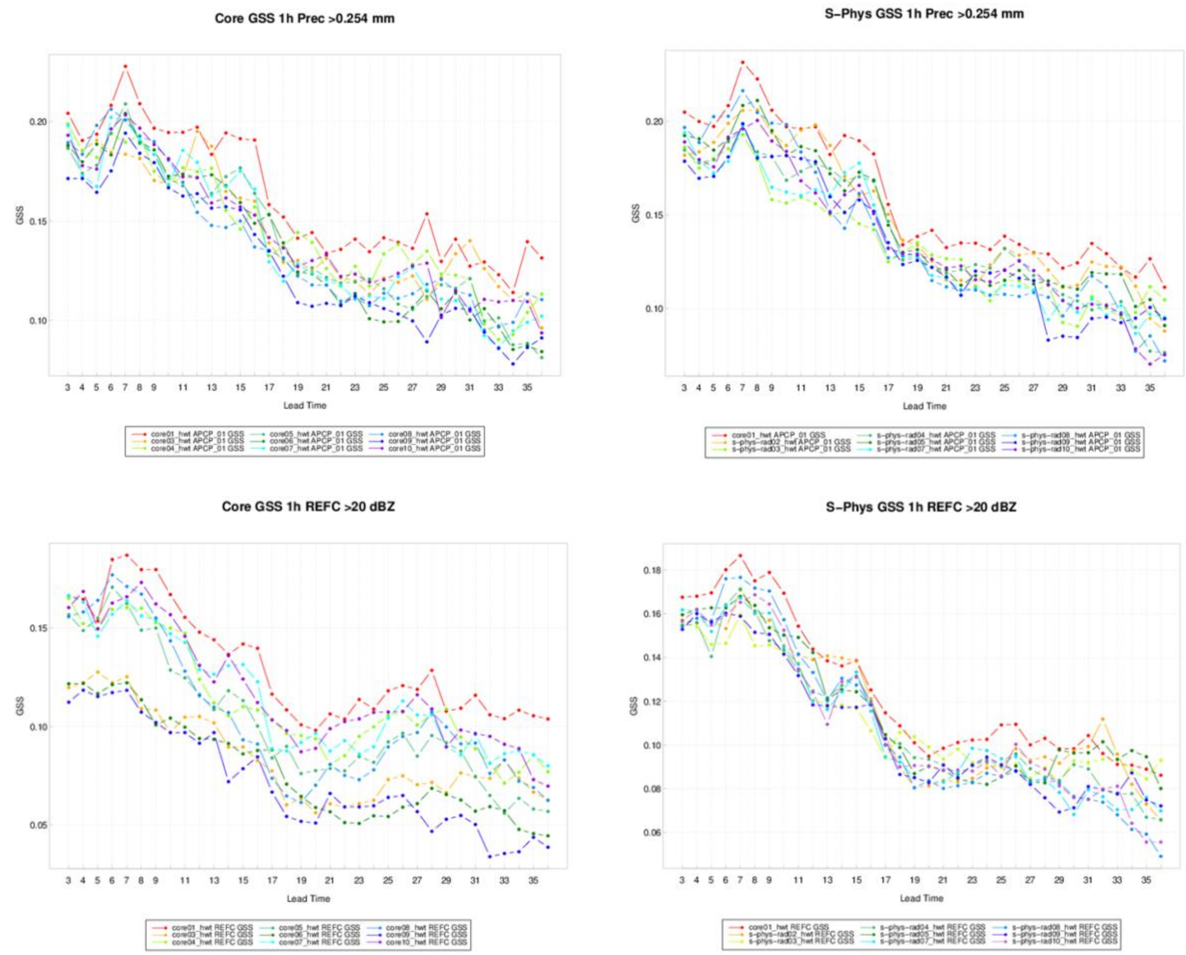

Figure 1: GSS for each member of Core (left) and S-Phys (right) for 2016 for 1 hour precipitation exceeding 0.254 mm (top) and CREF greater than 20 dBZ (bottom). In Core, red indicates member 1, orange 3, light green 4, medium green 5, dark green 6, cyan 7, blue 8, dark blue 9, and purple 10 (see Table 1 for configuration details). In S-Phys, red is member 1, orange 2, light green 3, medium green 4, dark green 5, cyan 7, blue 8, dark blue 9, and purple 10 (see Table 2 for configuration details).

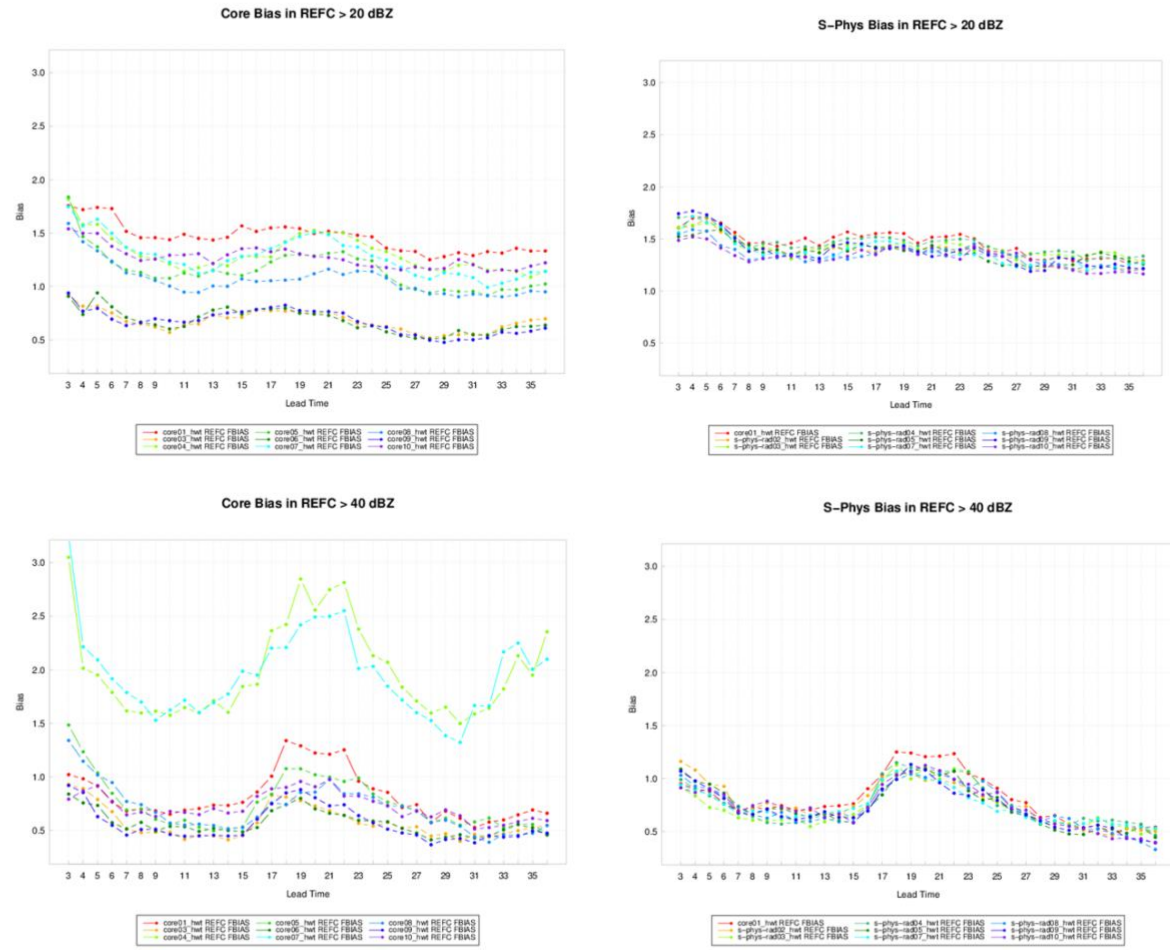

Figure 2: Bias for 2016 output for Core (left) and S-Phys (right) for 20 dBZ CREF threshold (top) and 40 dBZ threshold (bottom). Individual Core member physics schemes as indicated in Fig. 1.

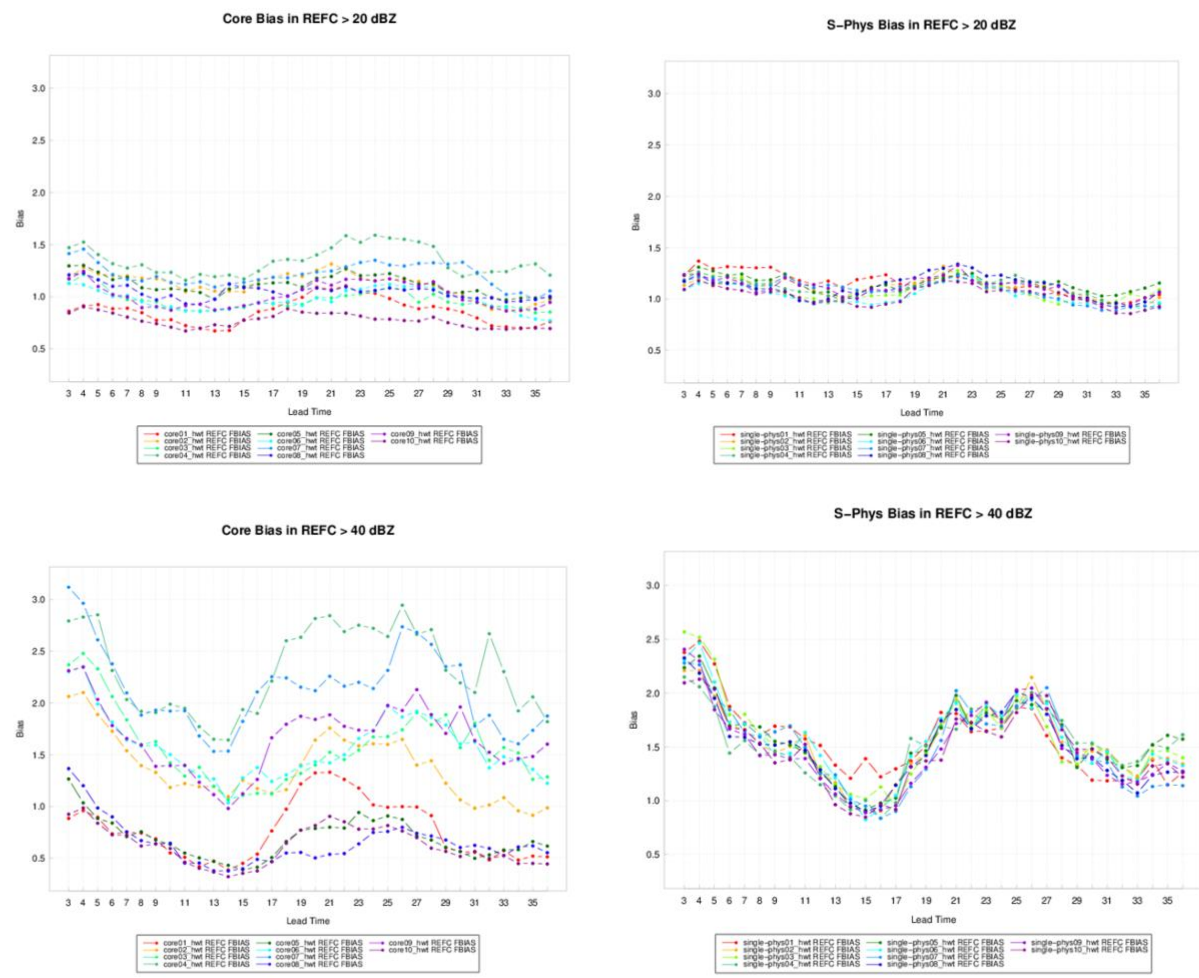

Figure 3: As in Fig. 2 except for 2017

b) MODE verification

Several different system attributes were compared using MODE for the two ensembles in both years. For 1 hour precipitation, the median value based on a threshold of 2.54 mm to define the object areas (Fig. 4) shows spread is greater for Core than S-Phys. Of note, for 1 hour precipitation, almost all members in both ensembles in 2016 are greater than the observations (black curve). The same is true in 2017 (not shown). In both years, the increased spread in Core does not translate into a better capture of the observations within the envelope. For heavier precipitation, the 90th percentile values behaved similarly in both years (not shown). When averaged together (not shown), the median value for both ensembles was too high compared to observations, with Core at most times having a worse error of roughly 0.1 mm. The 90th percentile values were also high at most times compared to the observations, with Core continuing to have a worse error. As can be inferred from Fig. 4, the problem was less severe for median precipitation during the afternoon and evening hours when convection was usually most intense (forecast hours 21-25).

A similar trend is apparent in averaged values for median and 90th percentile for CREF in Core (Fig. 5) with the average being too high compared to observations. However, for S-Phys, the median value is now often less than observations. Errors are therefore of comparable magnitude but opposite sign at most times for the two ensembles. Both ensembles in both their median and 90th percentile values show an afternoon peak in values that really does not show up in observations for median. It does show up for the 90th percentile, but the models are too intense with the maximum.

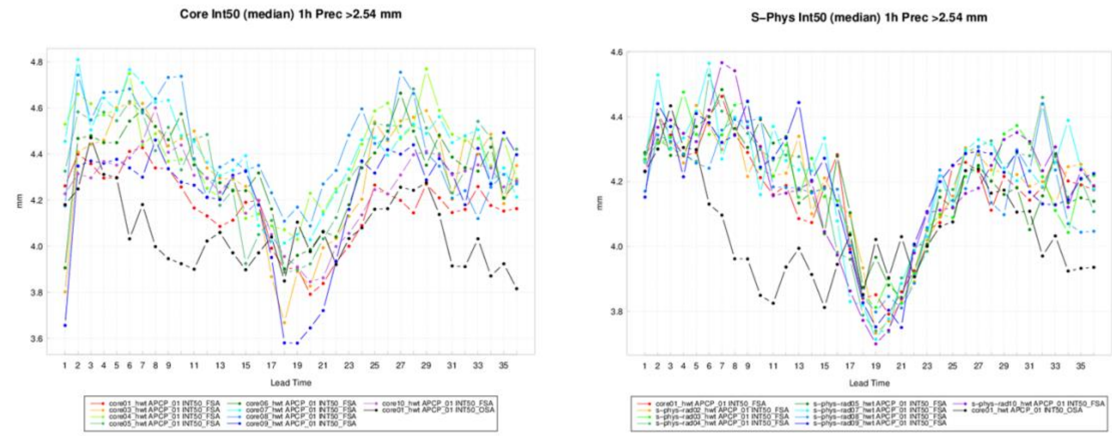

Figure 4: Median 1 hour precipitation value (mm) for both ensembles in 2016 (Core left, S-Phys right), with the observed value shown in black.

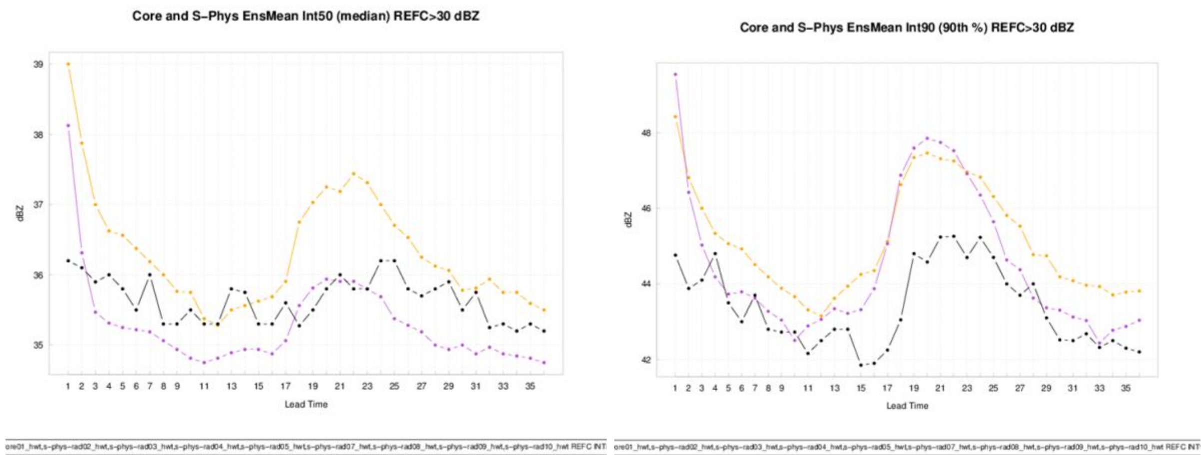

Figure 5: Median (left) and 90th percentile (right) values of CREF (dBZ) averaged among the 9 ensemble members for the Core (orange) and S-Phys (purple) ensembles in 2016. Observed value shown in black.

The median CREF values for each member of both ensembles are displayed in Fig. 6. The increased spread in Core is apparent in these plots for both 2016 and 2017, and unlike with 1 hour precipitation, the increased spread results in a much better capturing of the observations within the envelope of Core. In fact, in both years the observations fell outside any member prediction in S-Phys nearly all (2016) or all (2017) of the time. It should be noted, however, that although the observations are better captured in Core, some of its members greatly overestimate the reflectivity values. It can be seen that the median values are especially high in the MY members during the afternoon. Of note, for S-Phys, the members were usually less than observations in 2016 but greater than observations in 2017. This might indicate that the use of the MYNN PBL scheme and the RUC land surface scheme resulted in more intense reflectivity, but further work is needed to be sure. Similar results were obtained for the 90th percentile values except that S-Phys members in both years were usually too high compared to observations.

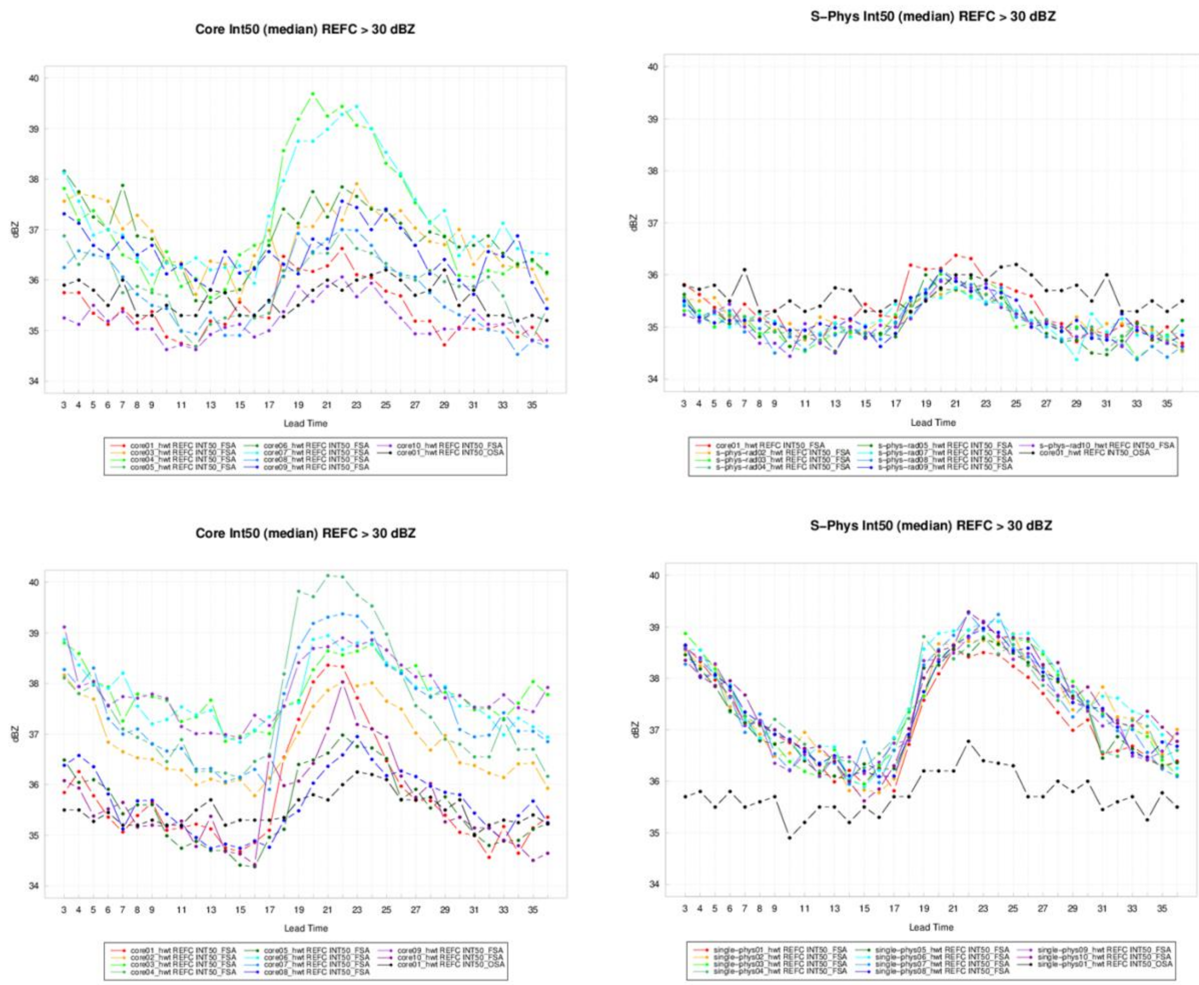

Figure 6: Median CREF value (dBZ) for all members of Core (left) and S-Phys (right) for 2016 (top) and 2017 (bottom). Observations are shown in black.

Areas within the MODE objects for CREF (greater than 30 dBZ) are shown in Fig. 7. As would be expected, these results should be somewhat similar to the point-to-point metric of bias. Much more spread exists in the Core ensemble compared to S-Phys. In 2016, the control member (in red) lies closest to the observed value (black) at nearly all times. In 2017, this is not the case, implying a worsening in the forecasts of areal coverage when the PBL scheme and land surface schemes are switched to MYNN and RUC, respectively. The Core ensemble does a better job of capturing the observed value within the envelope of members. This is especially true in 2017 when Core always had the observations within the envelope, usually around the median value, while the observations were almost always outside the envelope of S-Phys, as nearly all members overestimated the areal coverage, in contrast to 2016 when nearly all members (except the control) underestimated areal coverage. Again, this implies a potentially large impact from the change made in the PBL and land surface scheme in 2017. It should be noted in both years that a distinct clustering occurs in Core with curves rarely crossing each other. This suggests that different physics combinations have very systematic differences in the amount of echo above 30 dBZ with limited variability over time (i.e., one member will always have broader areas of echo; another member will always have much less). Such behavior again is concerning and implies the ensemble is not well-designed as each member would not be equally likely to verify.

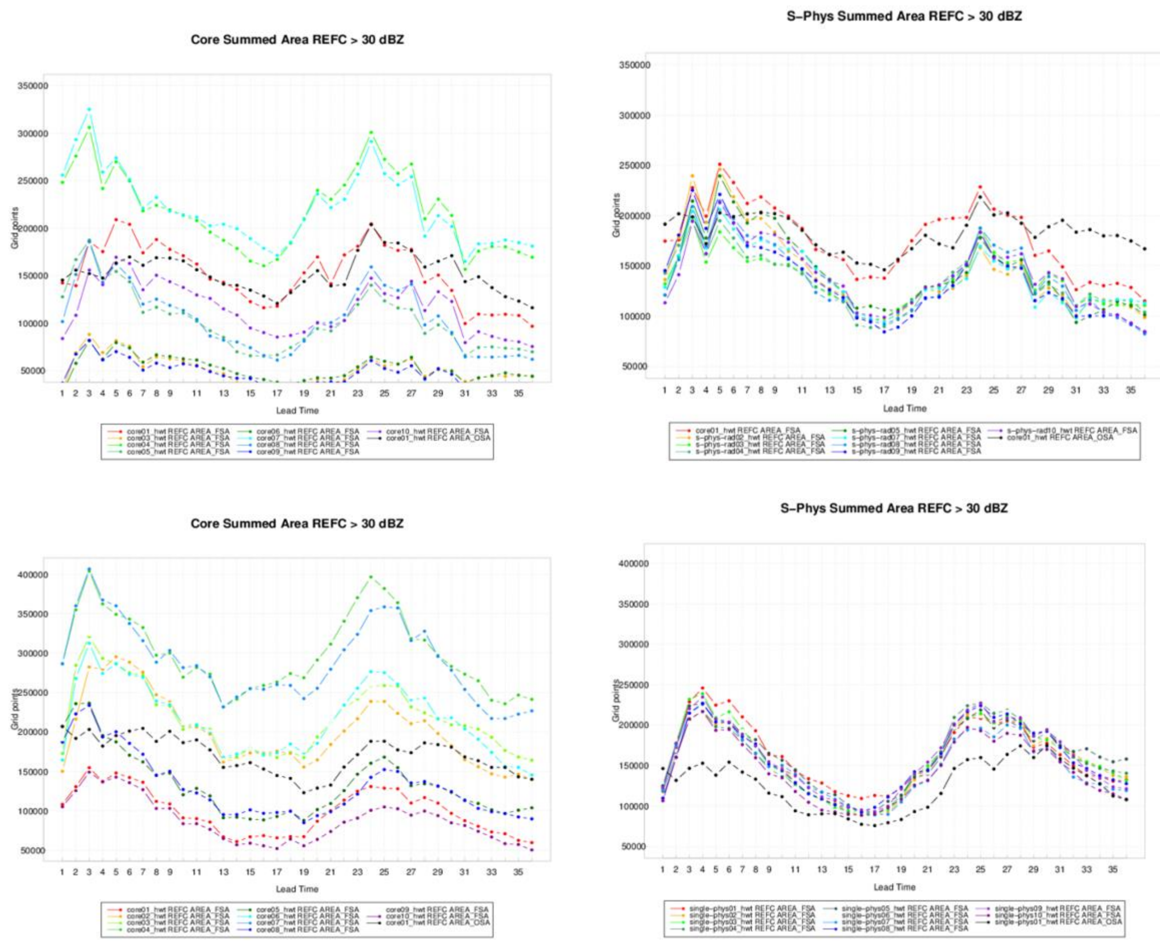

Figure 7: Areas (grid points) within the MODE objects for Core (left) and S-Phys (right) members in 2016 (top) and 2017 (bottom). Observations are in black.

Similar plots for 1 hour precipitation areas (not shown) indicate more spread in Core as well, but the differences are far less than for CREF, with a roughly 60-100% variation from the median in Core for CREF but only a 10-20% variation in Core for 1 hour precipitation. For S-Phys at most times, variations are only roughly 10% for both CREF and 1 hour precipitation (the one exception is in 2016 where the control run deviated more from the other 8 members). Nonetheless, Core still does a better job of capturing the observations within its envelope.

c) Traditional Ensemble Verification

ROC curves, areas under the curves, reliability diagrams, and Brier skill scores were examined for both ensembles, and generally indicated only a slight advantage at best for the Core ensemble. ROC curves for both years can be seen in Fig. 8 for two rainfall thresholds. In 2016, the two curves are very similar for 2.54 mm, but Core has a noticeable advantage for 6.35 mm.

Although not shown, Core had a bigger advantage in area under the ROC curve at most times for 0.254 mm. The improvement of Core over S-Phys is more obvious in 2017. In both years, skill (area under the ROC curve > 0.7) only existed at the majority of times through 2.54 mm. Skill was only present for the first 6-12 hours of the forecast for the 6.35 mm threshold (not shown).

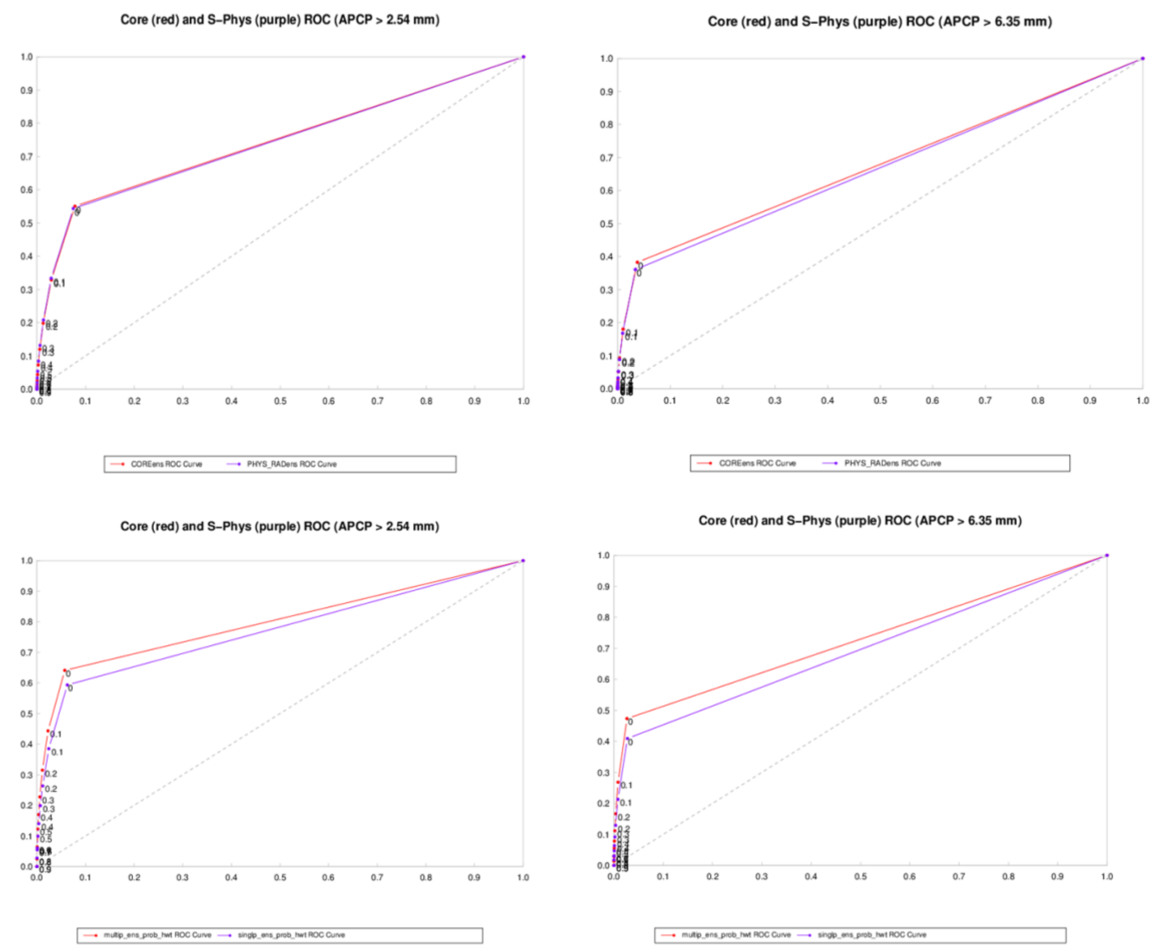

Figure 8: ROC curves for Core (red) and S-Phys (purple) in 2016 (top) and 2017 (bottom) for 2.54 mm (left) and 6.35 mm (right) 1 hour rainfall thresholds.

Reliability diagrams suggested a similar small advantage for the Core ensemble (not shown), but both ensembles overestimated the probabilities except for 0%, with curves lying well to the right of the diagonals. Skill relative to climatology only existed in 2016 for both ensembles for 1 hour precipitation above 0.254 mm. Some skill was present for 3 hour rainfall at the 2.54 mm threshold. The Core ensemble performed better in 2017 and showed some skill for 1 hour precipitation at the 2.54 mm threshold. For 3 hour precipitation, the Core ensemble was relatively reliable with its curve close to the diagonal. The difference in performance between Core and S-Phys was much greater in 2017 than in 2016, perhaps suggesting again that the change in PBL scheme and land surface scheme harmed the S-Phys ensemble in 2017.

d) Other Comparisons

The analysis of 14 comparisons when only a single change was made to a member configuration revealed a few items of note. A change in the microphysics scheme alone from Thompson to Morrison had a 27% larger impact on domain total 3 hour precipitation than a change made to the IC/LBCs alone. A change from Thompson to P3 microphysics had a 9% larger impact than changes in IC/LBCs alone. A change in PBL scheme in addition to a change in IC/LBCs resulted in a 9% increase in the impacts on precipitation. A change in microphysics in addition to a change in IC/LBCs increased the impacts on precipitation amount by 26%.

The investigation of convective initiation found there was less spread in the location of the initiation in S-Phys than in Core, but also smaller peak errors on average among the members for the 10 cases (Fig. 9). Both ensembles had the observed location within the envelope of member solutions in 6 of the 10 cases.

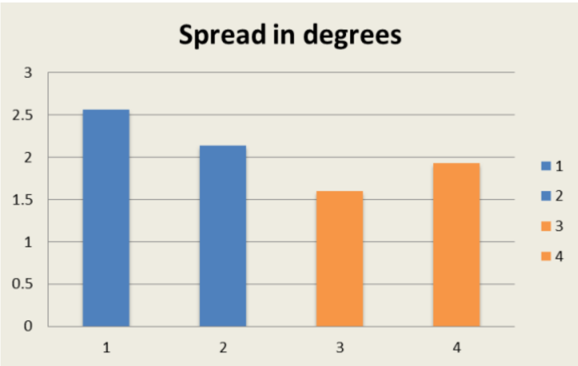

Figure 9: Variation in average spread (left) among the 9 ensemble members for the 10 cases (blue 1 represents latitude spread in Core, blue 2 longitude in Core, orange 3 is latitude for S-Phys, 4 is longitude for S-Phys), with box and whisker plots (right) of maximum errors in latitude (degrees) (leftmost 2 bars) and longitude (rightmost 2 bars) among the 9 members for the 10 cases. Boxes 1 and 3 are for Core, 2 and 4 are for S-Phys.

e) Microphysical Run Comparison

The comparison of the impacts of microphysics alone were consistent with the results shown earlier for the impacts of mixed physics, with MY always having the greatest areas above 30 dBZ, largest median and 90th percentile reflectivity values. P3 had the smallest areas, but was above average on the median and 90th percentile values, implying small but relatively intense cores of reflectivity. Thompson usually had an area most closely matching that observed. At most times, all schemes were too intense with the median reflectivity, and thus the schemes that had relatively lower values performed better (Thompson, Morrison). The same was generally true for the 90th percentile, except that Morrison often was lower than observations, while other schemes were higher. Morrison and WSM6 were often best for this parameter.

Of note, when convective initiation was examined for this 5-member microphysical ensemble, initiation was captured only one time in 6 cases (due to some missing data for these members, only 6 of the full set of 10 cases could be examined), whereas the Core and S-Phys ensembles captured it 4 times out of the 6 cases. This implies despite microphysical scheme choice having the single biggest impact on the model solutions, either a larger ensemble member size, or the added impacts of mixed IC/LBCs and other physics nonetheless were important.

Summary and Discussion

Two CLUE sub-ensembles were examined in detail to study the impact of including mixed physics in an ensemble that already used mixed IC/LBCs. In addition, a 5 member sub-ensemble consisting only of mixed microphysics was studied to understand systematic differences in these schemes. Comparisons were made using 24 cases of 1 and 3 hour precipitation from 2016 CLUE output, 17 cases of CREF in 2016, and 11 cases of both precipitation and CREF from 2017 CLUE output. Multiple verification metrics were examined.

In most cases, the mixed physics ensemble (Core) had noticeably more spread than the Single physics (S-Phys) ensemble. Differences in spread were surprisingly large when using CREF instead of precipitation, with much more spread showing up in the reflectivity fields. In most cases, but not all, the increased spread in Core better captured the observed value, and S-Phys appeared to be substantially underdispersive at most times. The average value of all members agreed better with observations a small amount for Core compared to S-Phys. However, especially for reflectivity, this average value came about from members like those that used MY microphysics that had large positive errors in intensity and areal coverage, which tended to balance many negative errors found in many of the other configurations. A summing of the errors from individual members would reveal Core to be worse. Traditional ensemble measures gave a slight advantage to the mixed physics ensemble, but suggested very little skill for 1 hour precipitation. More skill was present for 3 hour precipitation.

A similar increase in spread was shown in an evaluation of 10 cases of pristine convective initiation from the 2016 sample of cases. However, despite the increased spread in latitude and longitude positioning of initiation in Core, both ensembles correctly captured the observed location within their envelope of solutions in 60% of the cases. Thus, the performance of the two ensembles might be regarded as equal.

Finally, a comparison of 5 members of an ensemble using only mixed microphysics showed that MY was systematically too intense with high reflectivity values and too extensive with areal coverage above 30 dBZ. P3 had the smallest areas of all schemes but was often second most intense at the 90th percentile value, suggesting it creates small but intense cores. Despite evidence that changes in the microphysics schemes lead to the biggest changes among the mixed physics members, an examination of 6 cases of convective initiation suggests that an ensemble made up only of members with different microphysics schemes performs much more poorly at capturing the location of observed initiation within its envelope of solutions. This result implies that variations in PBL schemes are also important and/or variations in IC/LBCs.

Deliverables

In addition to a seminar given at NCAR during August 2017, research results were presented at the European General Assembly in Vienna, Austria in April 2018, and at the American Meteorological Society Numerical Weather Prediction Conference held in Denver, CO in June 2018. The PI is continuing to work with scientists at the DTC to develop a manuscript to be submitted to a refereed journal in the coming months.

References

Chen, C. H., T. Wang, Y. K. Tan, et al., 2009: Research of multi-model short-range ensemble forecasting techniques in forecasting rainy season over Changjiang–Huaihe basin in 2003. *J. Trop. Meteor.*, **25**, 449–457. DOI:10.3969/j.issn.1004-4965.2009.04.010.

Clark, A. J., W. A. Gallus, Jr., M. Xue, and F. Kong, 2010: Growth of spread in convection-allowing and convection-parameterizing Ensembles. Wea. Forecasting, **25**, 594-612.

Davis, C., B. Brown, and R. Bullock, 2006a: Object-based verification of precipitation forecasts. Part I: Methods 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.

Davis, C., B. Brown, R. Bullock, and J. Halley-Gotway, 2009: The Method for Object-based Diagnostic Evaluation (MODE) applied to numerical forecasts from the 2005 NSSL/SPC Spring Program. Wea. Forecasting, **24**, 1252-1267.

Gallus, W. A., Jr., 2010: Application of object-oriented verification techniques to ensemble precipitation forecasts. Wea. Forecasting. **25**, 144-158.

Johnson, A., X. Wang, M. Xue, and F. Kong, 2011: Hierarchical cluster analysis of a convection-allowing ensemble during the Hazardous Weather Testbed 2009 Spring Experiment. Part II: Ensemble clustering over the whole experiment period. Mon. Wea. Rev., 139, 3694-3710.

Li, X., H. R. He, C. H. Chen, 2017: A convection-allowing ensemble forecast based on the breeding growth mode and associated optimization of precipitation forecast. J. Meteor. Res., 31 (5), 955-964. Doi: 10.1007/s13351-017-6695-0.

Stensrud, D. J., J. Bao, and T. T. Warner, 2000: Using initial conditions and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. Mon. Wea. Rev., 128, 2077–2107.