JP 2.7 THE USE OF A PHASE SHIFTED VERIFICATION SCORE TO EVALUATE WARM SEASON QPF

Steven L. Aves[1], William A. Gallus, Jr.[1], Eugenia Kalnay[2] and Matthew Miller[3]

[1] Iowa State University, Ames, IA
[2] University of Maryland, College Park, MD
[3] ABC-TV 28/CBS-TV 13, Lubbock, TX

1. INTRODUCTION

While skill scores for precipitation forecasts have shown improvements over the years, they are still poor for warm-season convective events. This has been attributed to poor resolution of subgrid-scale processes and poor resolution of convection forced by mesoscale features (Stensrud and Fritsch 1994). Additionally, a spatial bias caused by the misrepresentation of boundaries or inability to handle the speed or evolution of the system can cause a degradation of forecasting skill, even if the forecasted precipitation is similar in structure to the observed fields. The effect of this spatial bias was ascertained by testing model sensitivity to the spacing between gridpoints, and by executing a shifting algorithm with the goal of removing the spatial bias and finding the forecasting skill of the unbiased result.

2. DATA AND METHODOLOGY

The data used for this study was the same as that in Gallus and Segal (2001). The methodology was also similar, with an ensemble of Eta model initializations utilizing different convective schemes and initial data. The shifting algorithm used first and second derivatives of forecasted and observed precipitation fields in an attempt to better match the minima and maxima in each field within the bounds of a maximum shift radius, and then shifted the forecast data accordingly. Equitable Threat Scores (ETS; Schaefer 1990) were then calculated to find the improvement of skill of the shifted data.

3. RESULTS

3.1 Dependence of Threat Score on Verification Grid Box Size

One of the problems with current objective verification techniques is the inability to provide useful information about potentially small space or time errors. Spatial errors may particularly penalize high resolution models verified on fine resolution grids. Tustison et al. (2001) discuss the contribution of this scale-related error (representativeness error) to total error, showing that threat scores will worsen because of it if verification is done using point-to-area conversion, as is often done.

To quantify the impact of verification grid box size, 10 km Eta model output was chosen from 11 warm season convective events (a subset of the full sample discussed later).

Using NCEP Stage IV 6-hour rainfall data, forecasts were verified (using equitable threat score, ETS) in three ways: (i) observations were averaged to the model's 10 km grid, (ii) model output was averaged to a 30 km grid and verified on that coarser grid, and (iii) all cases were rerun with a 30 km version of the Eta, and verified on that 30 km grid (same grid used in option ii).

The ETSs rose significantly when the 10 km model output was averaged onto and verified on the 30 km grid. The increases over ETSs computed on the 10 km grid were often comparable to the largest average improvements found in Gallus and Segal (2001) when initial conditions were modified to better represent mesoscale features. Even higher ETSs were obtained when cases were rerun with the 30 km Eta. Since ETSs for the 30 km model were even higher than those from the 10 km version averaged onto 30 km, it seems possible that forecasts of these convective events are adversely affected by the use of convective schemes at 10 km grid spacing, finer than that for which the schemes were designed. It is also possible that this result reflects another flaw with the use of ETS, as suggested by Mass et al. (2002).

3.2 Verification Using a Spatial Shift Technique

Figure 1 shows the improvement in ETSs after shifting along the x-axis. The similar graph for y-axis shifting is almost identical and has been omitted. Figure 1 shows an improvement for all precipitation thresholds, with the largest improvement in intermediate thresholds. Skill scores for lower thresholds were not improved as much because they have the largest unshifted skill scores, leaving less room for improvement. Also, there are typically many more points of light

Corresponding Author Address: Steven L. Aves, Atmospheric Science, Colorado State University, Fort Collins, CO 80521; e-mail: aves@atmos.colostate.edu

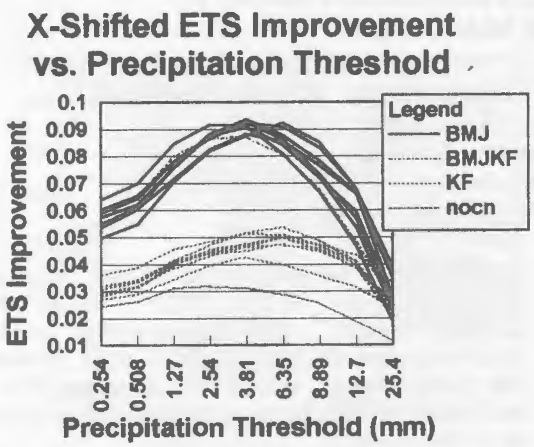

Figure 1. ETS improvement after shifting for each ensemble member and precipitation threshold.

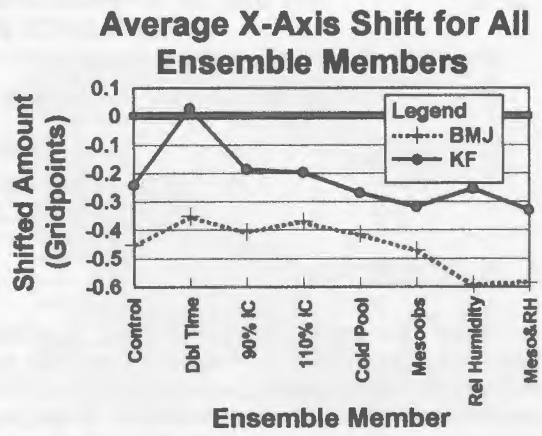

Figure 2. Average shift for each ensemble member as calculated by the shifting algorithm.

forecasted precipitation as compared to heavy precipitation, and these amounts are created by weaker meteorological forcings that may be more difficult to accurately predict. Thus, simply shifting the points did not have as large of an impact on the ETSs of the lower thresholds because of the larger number of incorrect points. Skill scores for higher thresholds were not improved as much because scores at these high precipitation amounts were biased towards zero, meaning there was often no precipitation predicted. Since the shifting algorithm can only cause spatial displacement, these values were not changed and the improvement was therefore diminished, on average.

Figure 1 also shows that runs that used the Betts-Miller-Janjic (BMJ; Janjic 1994) convective parameterization showed a larger ETS improvement than did runs using the Kain-Fritsch (KF; Kain and Fritsch 1993) scheme, and the difference between the two regimes is statistically significant above the 99.9% confidence level. This indicates that the spatial bias of the BMJ scheme may be larger than that of the KF scheme, and if so, the BMJ scheme shows more skill of producing the structure of the precipitation fields. Interestingly, the BMJ scheme already showed more skill than the KF scheme in the unshifted data, and since no significant correlation was found between unshifted ETSs and ETS improvement, the larger ETS improvement of the BMJ over the KF scheme cannot be attributed to some sort of nonlinearity.

While the difference of shifted ETS improvements was significant between the different convective parameterizations, it was not significant among ensemble members with similar convective schemes but varied initial conditions (not labeled in figure). This indicates that the different initial conditions, which were the only differences among the members, did not cause a large difference in the forecasted precipitation fields. Both precipitation amounts and spatial location were not significantly different among the ensemble members, which the small difference in ETS improvements accurately reflected.

Figure 2 demonstrates the average shift along the x-axis required for BMJ-type and KF-type runs with different initial conditions (90% and 110% IC refer to fully perturbed initial conditions with percentage of grid point perturbation from domain average, others are explained in Gallus and Segal, 2001, with Dbl Time indicating altered moist physics). As expected, the BMJ-type initializations were shifted farther than the KF-type initializations, which likely resulted in the larger ETS increase after shifting. However, the difference between the two types of runs is not statistically significant. The y-axis average shift has been omitted because it was not significantly different from zero for all ensemble members.

A few less than desirable qualities were found with the shifting algorithm. It appears that the algorithm does not shift the axis under consideration uniformly, which can lead to the narrowing or broadening of precipitation regions. Narrowing was the preferred mode of distortion, which led to a general decrease in domain-averaged precipitation after shifting. Future work will involve refining the shifting algorithm to better conserve precipitation amounts, and to facilitate shifting in any direction instead of only along the x- and y-axis.

4. ACKNOWLEDGEMENTS

This work was supported in part by NSF/USWRP grant ATM-9908932.

5. REFERENCES

Available on the web at http://cumulus.geol.iastate.edu/sat01.html