P1.40

OPTIMALLY SHIFTED QPF VERIFICATION SCORES

Ying Lin[1*], Barbara G. Brown[2], Keith Brill[1] and Geoffrey J. DiMego[1]
[1] National Centers for Environmental Prediction
[2] National Center for Atmospheric Research

1. INTRODUCTION

Precipitation forecast verification measures based on statistics obtained from model *vs.* observation (in the form of either point observation or analysis) comparisons of ***F****orecast*/***H****it*/***O****bserved* precipitation points over selected threshold values (hereafter referred to as the *FHO* approach) are used widely to gauge QPF skills. This is an objective, versatile approach, allowing many scores (equitable threat, probability of detection, false alarm ratio, *etc.*) to be computed from the *FHO* statistics. For example, at NCEP, monthly/month-to-date QPF scores for a number of operational models are displayed (http://www.emc.ncep.noaa.gov/mmb/ylin/pcpverif/scores), and precipitation scores are one of the criteria used to determine whether proposed model upgrades would have a positive impact on model forecasts (Mesinger, 1996).

There are a number of drawbacks to the *FHO* approach (see for example, Ebert and McBride, 2000; Baldwin and Kain, 2005), most noticeably its dependence on precipitation bias and sensitivity to model displacement error.

QPF scores such as equitable threat tend to favor forecasts with high bias. For this reason, equitable threat scores are often displayed alongside bias scores. Recently, Mesinger and Brill (2004) have developed a bias-modification of the *FHO*-based scores.

A small displacement error in model precipitation forecast can severely penalize the QPF scores (*e.g.* see Figures. 1 and 2). This problem is harder to address than the bias dependency – unlike bias, there is no standard measure within the *FHO* framework that quantifies displacement error to inform model users of the existence of the displacement and its possible impact on QPF scores.

A number of researchers have computed phase-shifted QPF scores for specific cases (Brown *et al.,* 2002 and Aves *et al.,*2002, among others). Several object-oriented verification approaches include phase error in their method: Ebert and McBride (2000) decomposes model QPF errors into components attributed to location/volume/pattern errors; Bullock *et al.* (2004) matches individual objects in the model precipitation forecast and observations and provide statistics on location/shape/intensity *etc.* for each matched pair.

*Corresponding author address: Ying Lin, NCEP/EMC, 5200 Auth Rd., Rm 207, Camp Springs, MD 20746.

We propose an approach to routinely and automatically compute the displacement error within the *FHO* framework. Precipitation scores computed before and after correcting for the displacement error can be placed side-by-side, to provide model users who are already familiar with the *FHO*-based scores an intuitive, more complete and balanced evaluation of model QPF performance.

2. THE METHOD

Figures 1 and 2 illustrate both the problem of the *FHO*-based scores' sensitivity to spatial displacement, as well as the proposed solution. The "forecast" field *(F)* in Figure 1, while visually quite close to the "observation" field *(O)* and is likely to be considered by human evaluators to be quite a good forecast, received poor equitable threat scores (red line in top panel in Figure 2). After shifting the "forecast" field by (28,35) in the (*x,y*) directions, the equitable threat score becomes much higher. The bias scores is unchanged here, because bias is defined as the ratio of the number of forecast points and observation points in the verification domain exceeding a given threshold, with no regard to spatial matching of *F* and *O* points.

In the illustration above, the optimal shift of (28,35) is obtained by maximizing area of overlap between *F* and *O* areas that exceed each field's median non-zero intensity. The idea was to define an "area of significant precipitation" *(ASP)* for *F* and *O*, shift *F* around until the maximum overlap is achieved, then re-compute the *FHO* statistics, to be used, along with the shift vector, as the optimally shifted scores. Defining *ASP* as the area(s) exceeding the median non-zero intensity works well for the conceptual case in Figure 1, but we find that it does not always work well in actual model forecasts. For example, some models are prone to produce large areas of very small precipitation amounts, and the resulting *ASP* does not look reasonable. We have experimented with defining *ASP* as areas exceeding the top 10^{th}/top 25^{th} percentile of non-zero precipitation; as areas exceeding the *average* non-zero amount of precipitation; as the areas of high precipitation intensity that produced 50%, 67%, 75% of domain-total precipitation. We found that while each of the method produced reasonable-looking *ASP*s in some forecasts/observations, there are always cases in which the resulting *ASP*s cannot be reasonably termed "area (s) of significant precipitation". Because of the difficulty in finding a single definition of *ASP* that would work well in most cases, we have decided to adopt the method used in Ebert and McBride (2000), in which the optimal shift is obtained by minimizing the total RMS error.

3. REAL LIFE EXAMPLES

Examples of how the optimally shifted scores might work within the *FHO* frame work are shown in Figures 3-6, on the QPF performance of Eta and GFS models during Hurricane Frances at the beginning of September 2004. The verifying analysis is the NCEP/CPC's satellite-based CMORPH (Janowiak, 2005). Note that we decided to use CMORPH in this test case because of its availability over the ocean and its high spatial and temporal frequency, even though CMORPH is not necessarily at its best in tropical storm situations (Janowiak, personal communication).

3-hourly precipitation totals from Eta, GFS and CMORPH ending at 2004090315 and 2004090403 are shown in Figures 3 and 5 (all forecasts are from model cycle 2004090218). Equitable threat and bias scores with and without optimal shift of forecast fields (by minimizing RMS error), along with shift vectors indicating the directions and amount of shifts on the forecast fields, are shown in Figures 4 and 6. The scores give model users a clear idea of how the models perform under the traditional *FHO* verification system, the amount of displacement error, and how the QPF performance might be different if there were no displacement errors.

4. WHY OPTIMAL SHIFTING?

This is a reasonable question to ask, given the availability of object-oriented approaches such as Ebert and McBride (2000) and Bullock *et al.* (2004). We argue that while the optimally shifted scores do not provide as comprehensive a description of model QPF errors as the object-oriented approaches, its ability to work within the *FHO* framework and its straight-forward, intuitive display of results make it easy for model users to add to their repertoire of model QPF evaluation tools.

While the optimally shifted scores deal only with the displacement error issue (it can be interpreted as both spatial and temporal displacement error), this is the hardest error to address within the *FHO* framework and hence most likely to benefit from a complementary approach.

We envision the optimally shifted scores as a useful complement to the traditional *FHO* scores, giving model users an easy way to more completely evaluate model performance.

5. FUTURE PLANS AND PRACTICAL CONSIDERATIONS

At NCEP, routine precipitation verifications are performed over ConUS and its fourteen sub-regions, for accumulation periods of 24h (using the NCEP/CPC 1/8-degree daily gauge analysis) and 3h (using NCEP's Stage II analysis – Lin and Mitchell, 2005). As displacement errors would be more evidently seen over small regions and short time periods – over the entire ConUS there might be several storm systems being forecast, each with its own phase errors - we plan to compute the optimally shifted scores for the 3-hourly verifications over the 14 sub-regions of ConUS.

5. ACKNOWLEDGEMENT

The impetus of this work was from a discussion between Stephen Lord (NCEP/EMC) and Bob Gall (NCAR/DTC) in spring 2004. DTC provided travel/visiting support to the lead author.

REFERENCES

Aves, S.L., W.A. Gallus, E. Kalnay, and M. Miller, 2002: The use of a phase shifted verification score to evaluate warm season QPF. *Abstract, 19th Conf. on Weather Analysis and Forecasting/15th Conf. on Numerical Weather Prediction.* Amer. Meteor. Soc.. Paper JP2.7.

Baldwin, M.E. and J.S. Kain, 2005: Sensitivity of several performance measures to displacement error, bias and event frequency. Submitted to *Wea. Forecasting.*

Brown, B., Bullock, R., Davis, C., Manning, K., Morss, R., Mueller, C., 2002a: An object-based diagnostic approach for QPF and convective forecast verification (http://www.rap.ucar.edu/research/verification/pdf/barb_vx_wrkshp_poster.pdf). Verification workshop: making verification more meaningful.

Bullock, R., B.G. Brown, C.A. Davis, M. Chapman, K.W. Manning and R. Morss, 2004: An object-oriented approach to the verification of quantitative precipitation forecasts. Part I – methodology. *Preprints, 17th Conf. on Probability and Statistics in Atmospheric Sciences/20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction.* Amer. Meteor. Soc.. Paper J12.4.

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

Janowiak, J.J., R.J. Joyce, P.A. Arkin and P. Xie, 2005: CMORPH: A new high-resolution global precipitation analysis system with potential use for hydrologic model data assimilation. *Abstract, 19th Conf. on Hydrology,* Amer. Meteor. Soc., San Diego, CA. Paper 1.4.

Lin, Y., K.E. Mitchell, 2005: The NCEP Stage II/IV precipitation analyses: development and applications. *Preprints, 19th Conf. on Hydrology,* Amer. Meteor. Soc., San Diego, CA. Paper 1.2.

Mesinger, F., 1996: Improvements in quantitative precipitation forecasts with the Eta regional model at the National Centers for Environmental Prediction: the 48-km upgrade. *Bull. Amer. Meteor. Soc.,* **77,** 2637-2649.

Mesinger, F., and K. Brill, 2004: Bias normalized precipitation scores. *Preprints, 17th Conf. on Probability and Statistics,* Amer. Meteor. Soc., Seattle, WA. Paper J12.6.

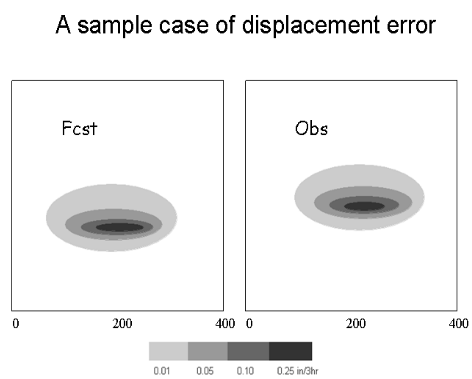

Fig. 1. A conceptual example of forecast displacement error.

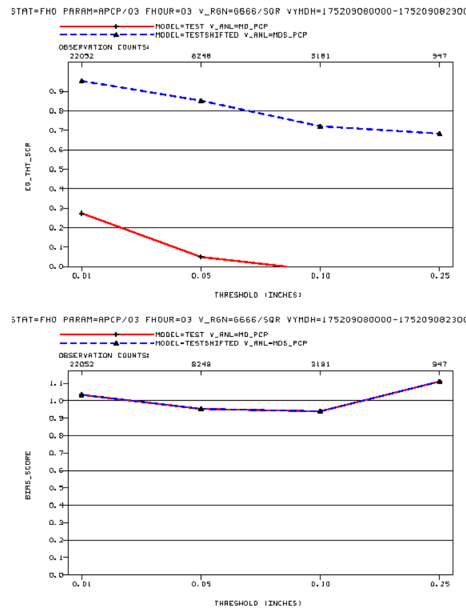

Fig. 2. Equitable threat (top) and bias (bottom) scores for the Fig. 1. Solid red lines: scores for the forecast as it is. Dashed blue lines: scores for the forecast, after the forecast field has been shifted by rightward, and upward, for 28 and 36 grid points, respectively (the verification domain is 400x400).

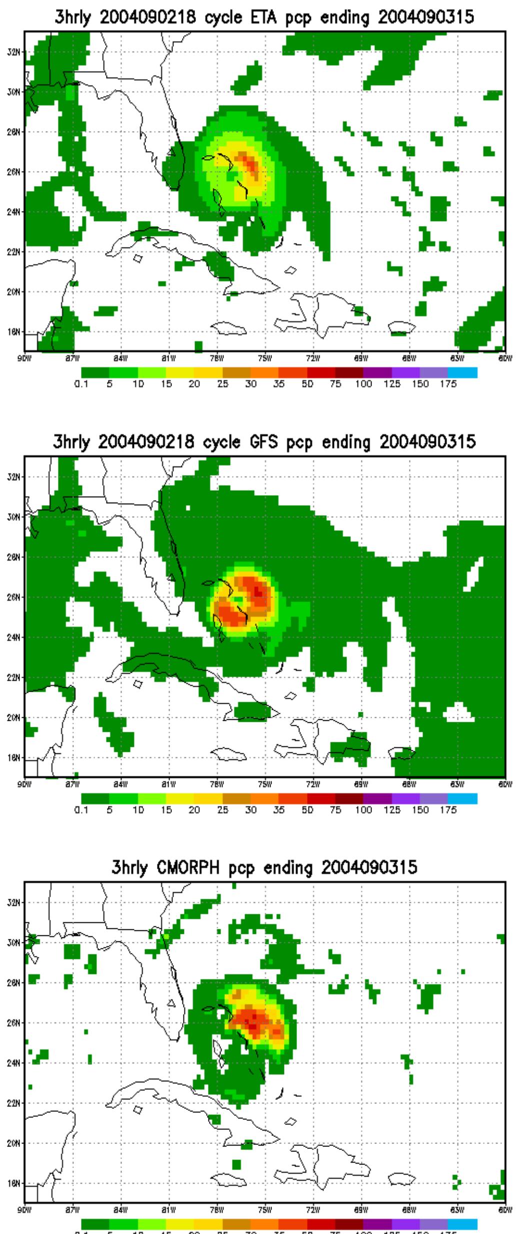

Figure 3. 3-hourly precipitation accumulation (mm) ending 2004090315 (UTC), from the 2004090218 cycle forecast of Eta (top) and GFS (center) models, and from CMORPH analysis (bottom).

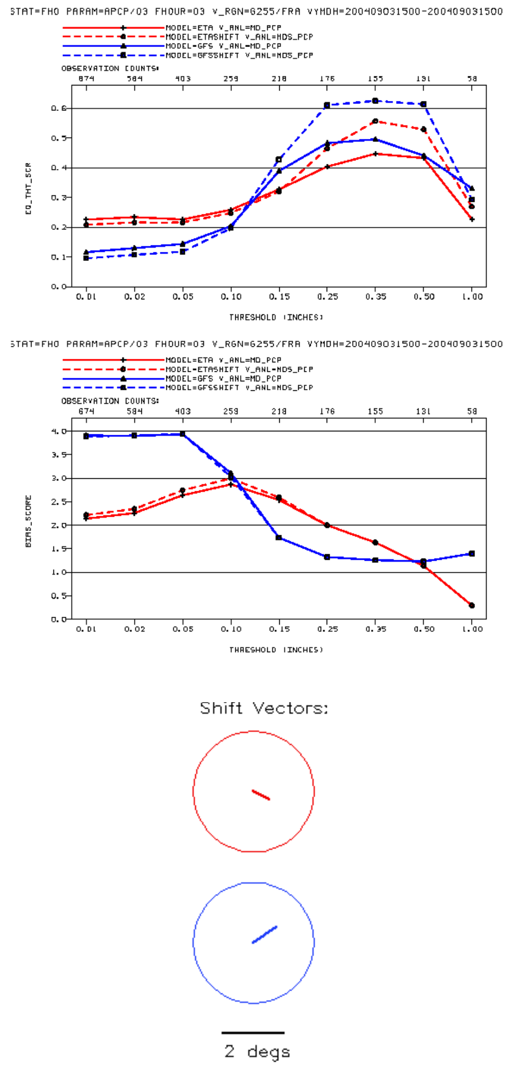

Figure 4. Equitable threat (top) and bias (bottom) scores for Eta (red) and GFS (blue) QPF verification, along with the shift vectors, for the case in Figure 3. Solid lines are scores without the shift, dashed lines are scores after optimal shift.

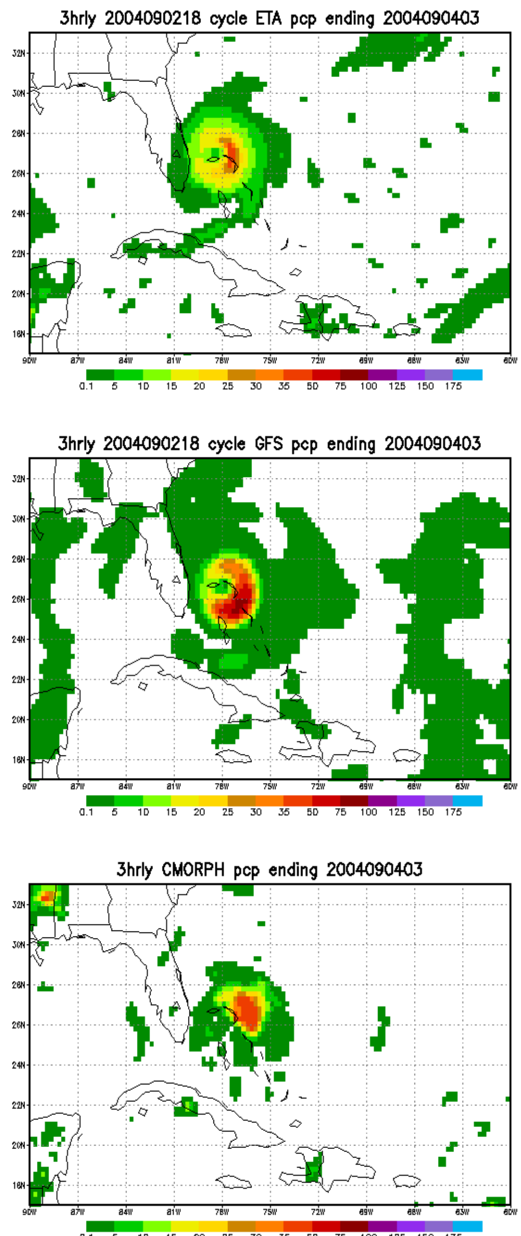

Figure 5. 3-hourly precipitation accumulation (mm) ending 2004090403 (UTC), from the 2004090218 cycle forecast of Eta (top) and GFS (center) models, and from CMORPH analysis (bottom).

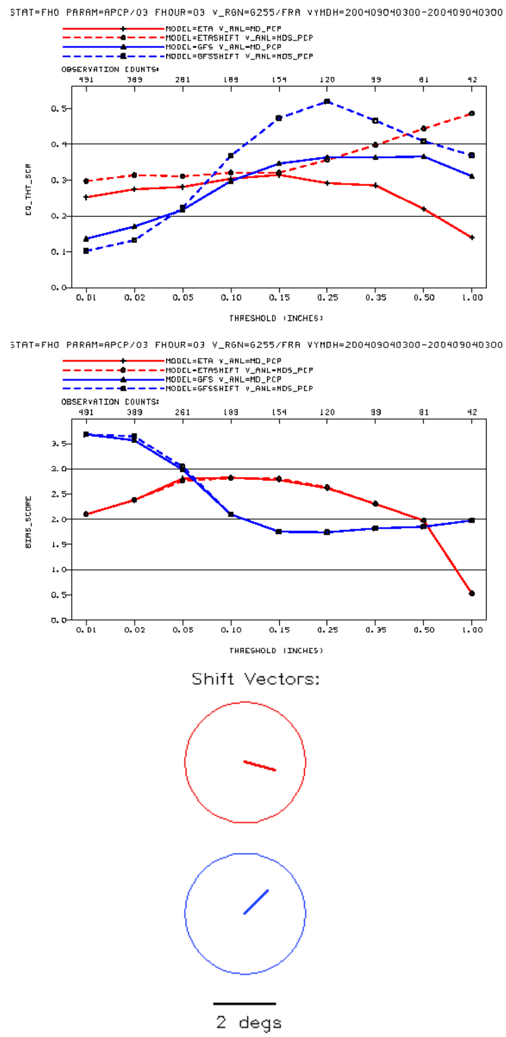

Figure 6. Equitable threat (top) and bias (bottom) scores for Eta (red) and GFS (blue) QPF verification, along with the shift vectors, for the case in Figure 5. Solid lines are scores without the shift, dashed lines are scores after optimal shift.