OBJECTIVE EVALUATION OF GLOBAL PRECIPITATION FORECAST

Yuejian ZHU

Environmental Modeling Center, NCEP/NWS/NOAA, Washington DC, USA

E-Mail: Yuejian.Zhu@noaa.gov

ABSTRACT

Numerical Weather Prediction (NWP) model quantitative precipitation forecasts (QPF) have been one of interest to forecasters (and public users) as they influence daily life. The rapid increases in model resolution are proportional to the increases in computer resources. The development of precipitation observation networks and the improvement in measuring techniques are greatly promoting the accuracy of observed precipitation. It is possible to build a model precipitation verification system based on high quality observation from the Continental United States (CONUS) area. The model precipitation forecast skills are more varied (sensitive) to different verification techniques, different spatial and temporal resolutions, especially model biases. Somehow, the forecast skills are misleading when the model forecasts are strongly biased. The subject of model bias and bias-free verification will be discussed in the separated paper. The past 4-year NCEP/GFS QPFs have been evaluated by different measures. The skills are improved gradually for past years when model increases the resolution, improves analysis system and physical processes. The winter seasons have more skills than summers. The lower spatial, temporal resolutions have more skills than the higher spatial, temporal resolutions.

1. INTRODUCTION:

NCEP Global Forecasting System (GFS) has made precipitation forecasts up to 16 days at each data assimilation cycle (00, 06, 12 and 18 UTC). The objective verifications have been made for past four years. There are many interesting aspects in the precipitation verification, such as verifying spatial and temporal resolutions, densities of observed stations, forecasting boxes, and so on. The most of our users are considering BIAS as very important verification index, some users are considering the SKILL is very useful. In the practice, the SKILL is really affected by BIAS (BI). For example, large BIAS will decrease the SKILL of Equitable Threat Score (ETS), on the other hand, large BIAS will increase the SKILL of True Skill Statistics (TSS). Therefore, the SCORE/SKILL is not quiet comparable each other when using different techniques. The BIAS-FREE verification scores/skills are very useful to users and model developers. In this paper, the verification methods will be introduced in the next section. The data of observed and forecasted for this objective evaluation will be described in section 3. Then, the varied skills of QPF are presented in section 4. Discussions and conclusions will be in section 5.

2. VERIFICATION TECHNIQUES:

There are many methods to evaluate precipitation forecast. The most commonly used by methodological research community is based on two by two contingency table for observed and forecast precipitation categories (see table). An h (hit) is a number of correct forecasts from specific precipitation threshold, f (false alarm) is a number of incorrect forecasts, m (missing) is a number of missing forecasts, and c is a number of correct rejections.

Precipitation	Observation (Yes)	Observation (No)
Forecast (Yes)	h	f
Forecast (No)	m	c

Table: Contingency table for forecast and observed precipitation categories, the meaning of letters h, f, m and c explained in the text.

Based on the table, if the precipitation threshold is chosen, following measurements will be formed to evaluate precipitation forecasts. Mainly, the Bias (BI) and Equitable Threat Score (ETS) are most useful, where,

$$BI = \frac{h+f}{h+m} \quad (1)$$

$$ETS = \frac{h-R(h)}{(h+f+m-R(h))} \quad (2)$$

where, $R(h) = \frac{h \cdot c - f \cdot m}{(h+m+f+c)}$

In additional to above two, hit rate (*HTR*) and false alarm rate (*FAR*) could be generated easily by defining:

$$HTR = \frac{h}{h+m} \quad (3)$$

$$FAR = \frac{f}{f+c} \quad (4)$$

By using *HTR* and *FAR*, True Skill Statistics (*TSS*) could be defined from the difference between *HTR* and *FAR*, that is:

$$TSS = \frac{h}{h+m} - \frac{f}{f+c} = \frac{h \cdot c - f \cdot m}{(h+m)\cdot(f+c)} \quad (5)$$

3. OBSERVATION AND FORECAST DATA:

The observed precipitations used in this study are NCEP stage II (GAGE only) and stage IV (Mosaicked from the regional multi-sensor precipitation estimator analyses (MPEs) generated by the River Forecast Centers (RFCs) which covered the Continental Unites States (CONUS) area (see: http://wwwt.emc.ncep.noaa.gov/mmb/ylin/pcpanl). There are about 7,000-8,000 automated gauge reports from CONUS per every 24 hours from 1200UTC to next 1200UTC. The hourly precipitation analysis from NCEP stage IV (radar + gauge) is on 4.7625 km HRAP grid (AWIPS #240). Varied NCEP/GFS daily quantitative precipitation forecast (QPF) data are archived, but most of them are low resolution (2.5 degree by 2.5 degree latitude, longitude grids, about 250km). Currently, there are three different resolutions of NCEP/GFS QPF production available to users, but all of them are interpolated from highest model resolution. The resolutions are 2.5 X 2.5 degree global latitude, longitude grid resolution (about 250 km around mid-latitude), 1.0 x 1.0 degree global resolution and 0.5 x 0.5 degree global resolution (since May 2003).

The verifications are either on grid boxes (resolution dependent) or at observation location (GAGE only). In this study, the verification is on latitude-longitude grid box only, there are several grid resolutions are considered which are 2.5-degree grid, 1.0 degree grid, 80km (AWIPS GRID 211) and 0.5-degree. To examine the interpolation scheme, figure 1 is an example of NCEP stage IV analysis for 6-hour accumulated precipitation that interpolated to three different resolutions. Top one is a 15km resolution interpolated from 4.8km that considered as a very high resolution, showing meso-scale weather band and other phenomena. The middle one is a 0.5-degree resolution from 4.8km. The bottom one is a 1.0-degree resolution. When reducing resolution, the details will be lost by interpolation or smoothing. Currently, the 0.5-degree resolution is a highest global model forecast production available from NCEP GFS. Apparently, the middle one of figure 1 is the closest precipitation analysis we could get. By comparing to 15km analysis, 0.5-degree resolution analysis keeps most of the precipitation futures in this example. The analysis from lower resolutions, such as 1.0-degree and 2.5-degree (not shown here), will miss all meso-scale futures.

4. PRECIPITATION FORECAST SKILLS:

The precipitation forecast skills are highly depended on the resolutions of verified grids/boxes (spatial) and time period (temporal). There is higher skill if the verified grids/boxes are very large or the time period is very long. Figure 2 and figure 3 are the results from same verified period (18 days, 72 sample sizes from 6-hour accumulation period of 0000UTC, 0600UTC, 1200UTC and 1800UTC), but the verification spatial resolutions, which are 1.0-degree (figure 1) and 0.5-degree (figure 2), are different. The four different forecasts (00-06 hours, 06-12 hours, 12-18 hours and 18-24 hours, leading time) have been verified by calculating ETS, TSS and BI. The x-axis is 6-hour threshold precipitation amounts in mm. The numbers above x-axis are total observed grids/boxes in the verified period for that threshold. The BIs (figure 1) are very similar at level of 1.7 (70% over-forecast) for all lead time forecasts up to 2.0 mm thresholds. There is a big jump (spin-up) after 6-hour forecast from 5.0 mm to 25 mm. There are not enough samples to support extreme precipitation (beyond 25 mm). The TSS and ETS are reasonable decreasing their skills when increasing lead time. There is the same future for figure 2 except the BI are larger, TSS and ETS are less. Therefore, it is not comparable if the spatial resolutions are difference. The similar results will be for different temporal resolutions (not shown). For example, there is more skill for 24-hour accumulated precipitation forecast than for 6-hour.

For consistence and comparison, the 2.5-degree grids/boxes, 24-hour accumulated precipitation observations (1200UTC to next 1200UTC) against forecasts (12-36 hours, 36-60 hours, and etc … from 0000UTC), have been used to evaluate GFS QPFs. The statistics (not shown here) are based on seasonal (winter-December, January and February. spring-March, April and May. summer-June, July and August. fall-September, October and November) average. In general, the precipitation forecast has more skill in winter when comparing to summer.

5. DISCUSS AND CONCLUSION:

The verification of precipitation forecast is very important to users and model developers. ETS is a very reasonable measure to evaluate over-all performance of QPF when considering BI as additional to. Bias-free verification is a future consistence measurement to users if we could apply. There are not comparable if the spatial or temporal resolutions are different. There is always challenge to our model developers when observation is improved, and spatial and temporal resolutions are increased.

REFERENCES:

Cherubini, T., A. Ghelli, and F. Lalaurette, 2001: Verification of precipitation forecasts over the Alpine region using a high density observing network. ECMWF Technical Memorandum, 340, 18pp

Hamill, T. M, 1999: Hypothesis tests for evaluating numerical precipitation forecasts. Weather and. Forecasting, 14, 155-167.

Mason, I. B., 2003: Binary Events, in book of “Forecast Verification: A practitioner’s guide in atmospheric science”. Edits: I. T. Jollife and D. B. Stephenson. Wiley, 37-76

Stephenson, D. B., 2000: Use of the “odds ratio” for diagnosing forecast skill. Weather and Forecasting, 15, 221-232.

Zhu, Y., and Z. Toth, 1999: Objective evaluation of QPF and PQPF forecasts based on NCEP ensemble. 3rd International scientific conference on the Global Energy and Water Cycle Experiments (GEWEX), 16-19 June 1999, Beijing, China, 47-48

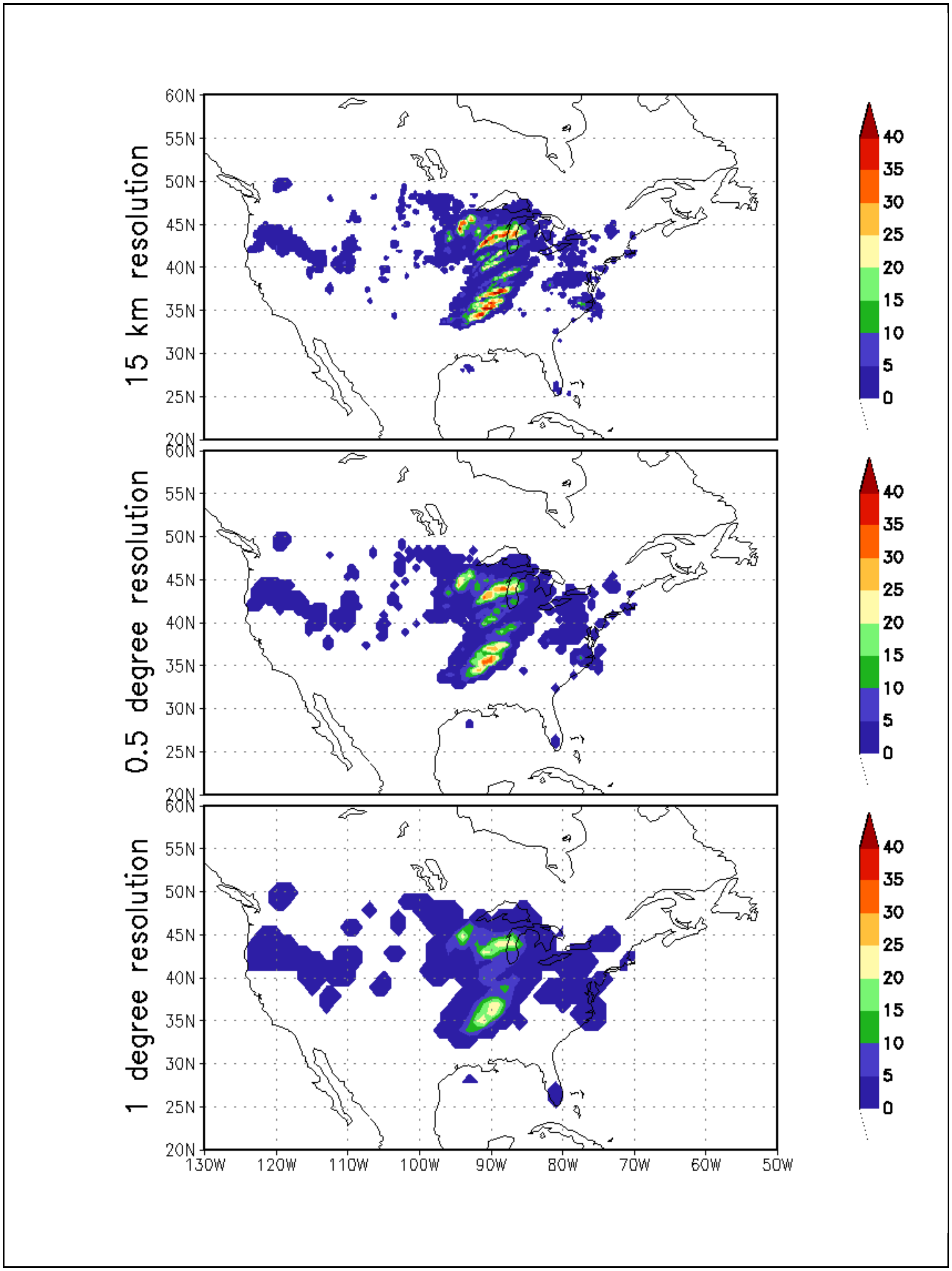

Figure 1. Observed 6-hour accumulated precipitations are for 00UTC – 06UTC of 11 May 2003 from NCEP stage IV analysis interpolated to 15-km (top), 0.5-degree (middle) and 1-degree (bottom) resolutions.

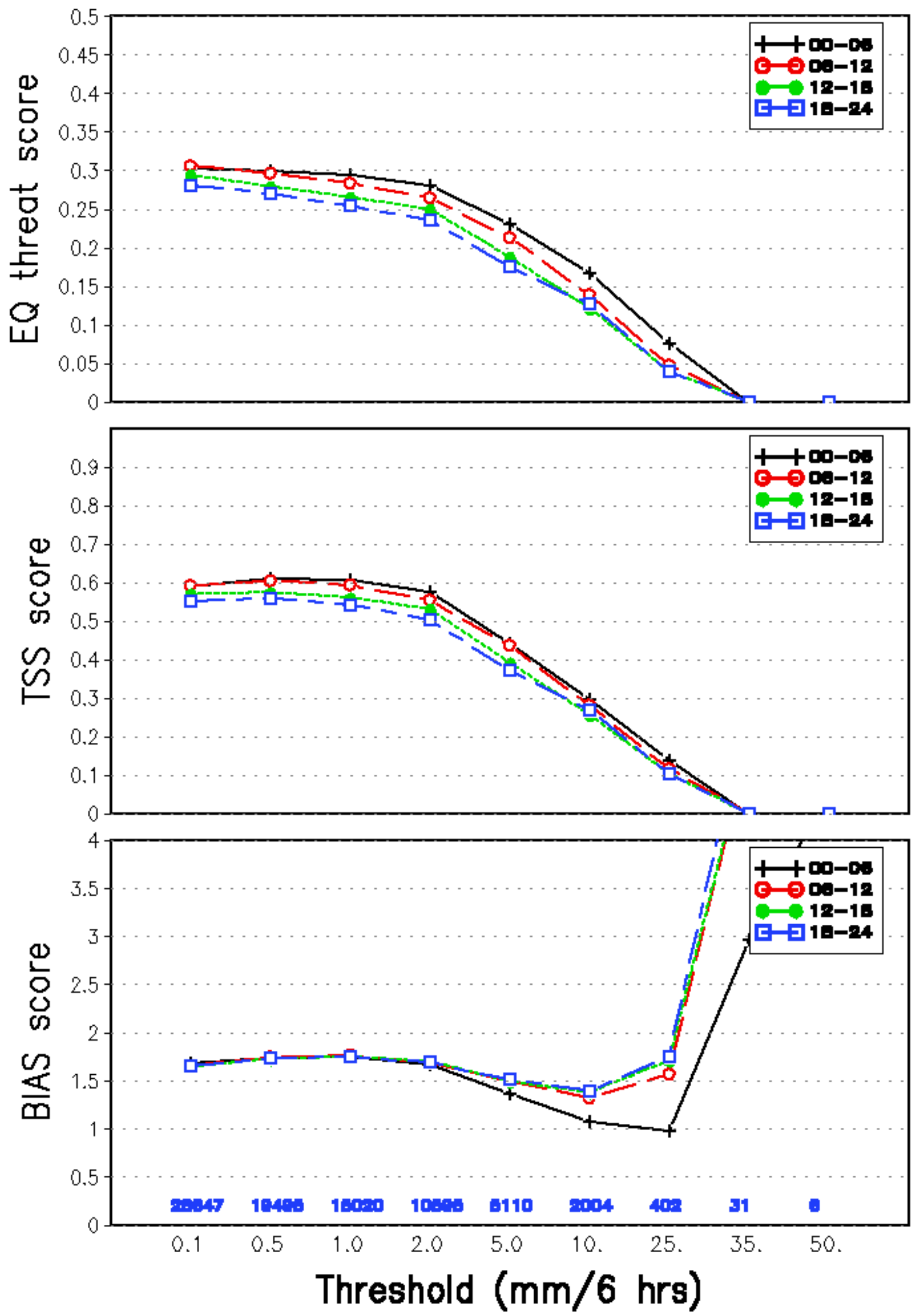

Figure 2. The verification scores are for 0000UTC of 24 May 2003 – 1800UTC of 22 June 2003 (totally 120 6-hour verification periods-samples) from 1-degree resolution. Solid line is 00-06 hours accumulated forecasts, long dash lines are 06-12 hours, doted lines are for 12-18 hours and long-short dashed lines are for 18-24 hours. ETS (top), TSS score (middle) and BIAS score (bottom).

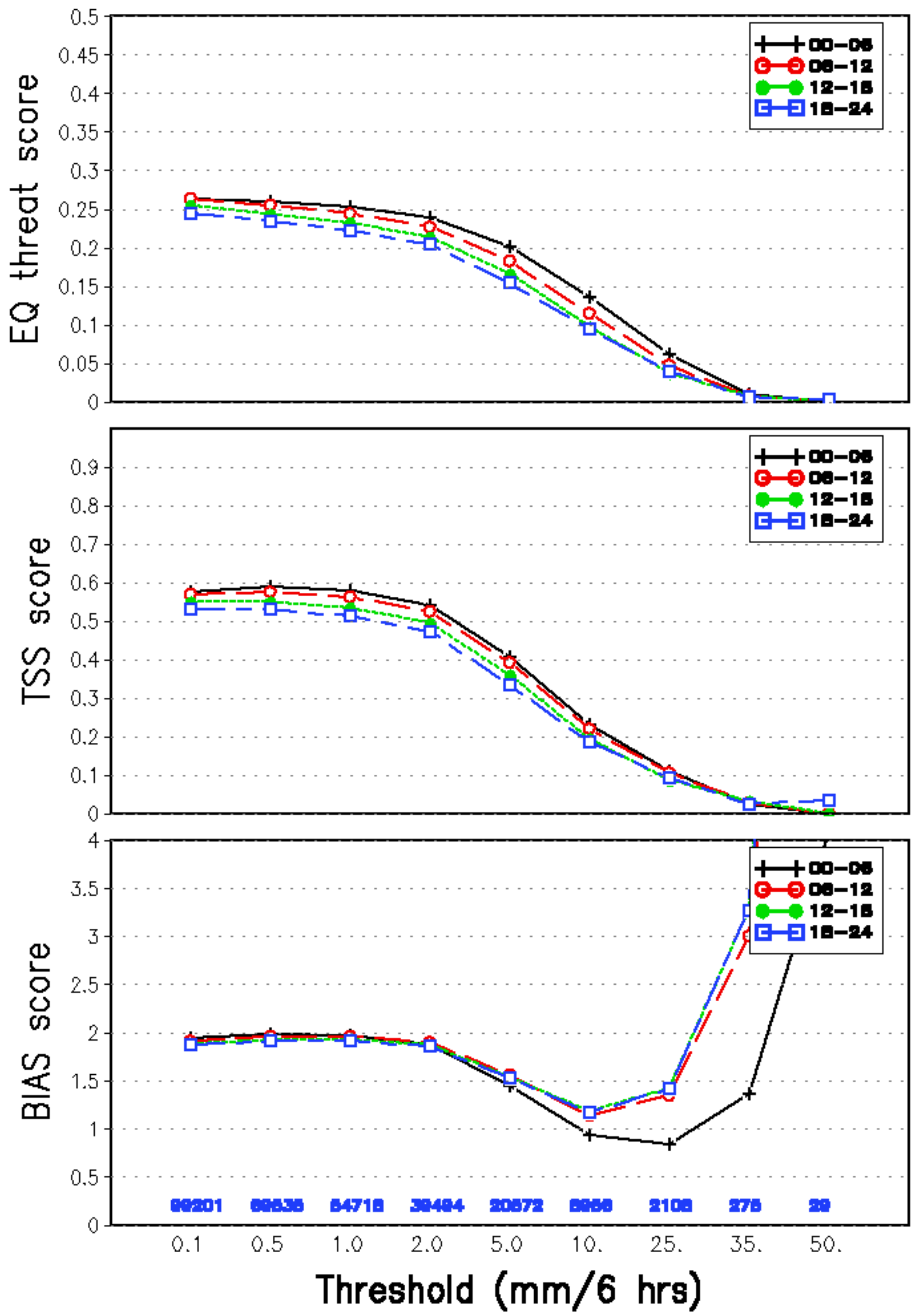

Figure 3. The plots are the same as figure 2 except for 0.5-degree resolution verification.