A Comparison of QPF from 4 km Grid Spacing WRF Simulations with Operational NAM and GFS Output Using Multiple Verification Methods

Haifan Yan and William A. Gallus Jr.
Geological and Atmospheric Sciences Department, Iowa State University

1. Introduction

- During summer, Quantitative Precipitation Forecast (QPF) skill is a minimum (Sukovich et al. 2014), which is unfortunate for Iowa where the heaviest rains and most flooding occurs then.
- QPF skill tends to be better when rainfall coverage is larger. When storm coverage is smaller than 10% of the whole domain, the largest displacement errors occur in rain systems. (Johnson and Olsen 1997)
- Equitable Threat Score (ETS) is higher for coarser grids, whether from the model configuration or when QPF is averaged from a finer grid. (Gallus 2002)

2. Model Setup

- 12-h long ARW-WRF simulations (version 3.5) were run every 6 hours (00, 06, 12 and 18 UTC) from March to November, 2013
- Domain had 200×200 points, with 4km horizontal grid spacing, and was centered over IA
- ARPS 3DVAR was used to assimilate radar data into the initial NAM background fields
- Thompson microphysics scheme was used with nonlocal MYJ PBL scheme.
- WRF and two operational models-- 12km NAM and 0.5° GFS – were verified using Stage IV data as ground truth
- WRF, NAM, GFS and STAGE IV were bilinearly interpolated to WRF grid (Hres) and GFS grid (Lres) to perform verification using MET tools.
- A normalized difference (D) between observation (O) and forecast (F) was used to calculate intensity sum difference:

$$D = (F-O) / [0.5\times(F+O)]$$

3. General Climatology Results

WRF under-predicted and NAM over-predicted the number of null precipitation cases – these cases have no skill when using FSS.
For flood cases, WRF was the only model able to suggest true magnitude of heavy rain potential; NAM and GFS largely underestimated the rainfall amount.

4. Fractions Skill Score (FSS)

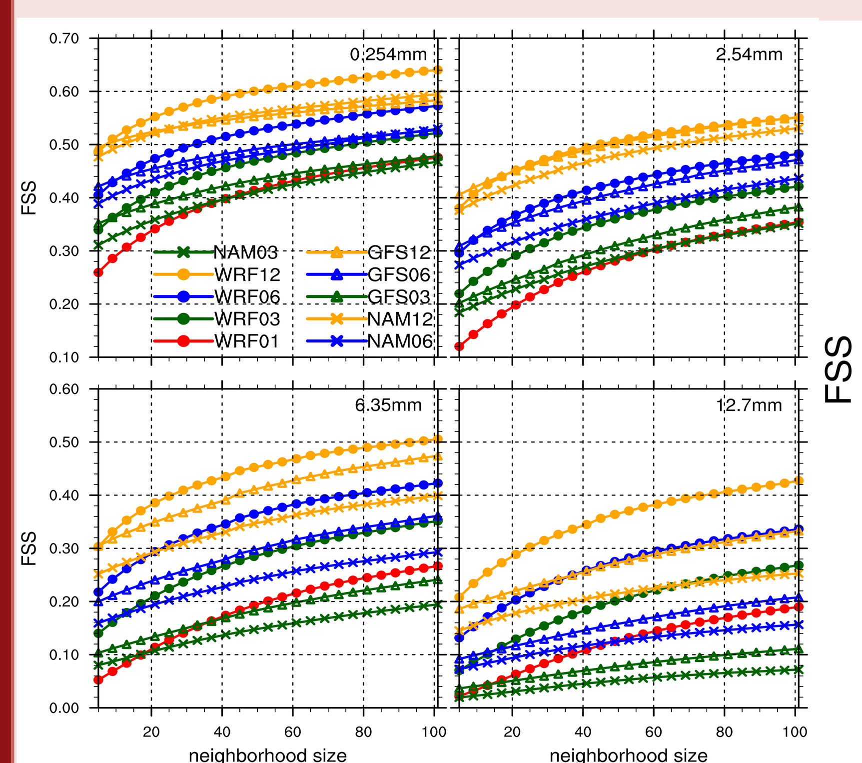

Fig. 1: Mean FSS of 1h, 3h, 6h and 12h accumulation intervals for the 3 models (colored curves) as a function of neighborhood size (in grid units) for 4 rainfall thresholds

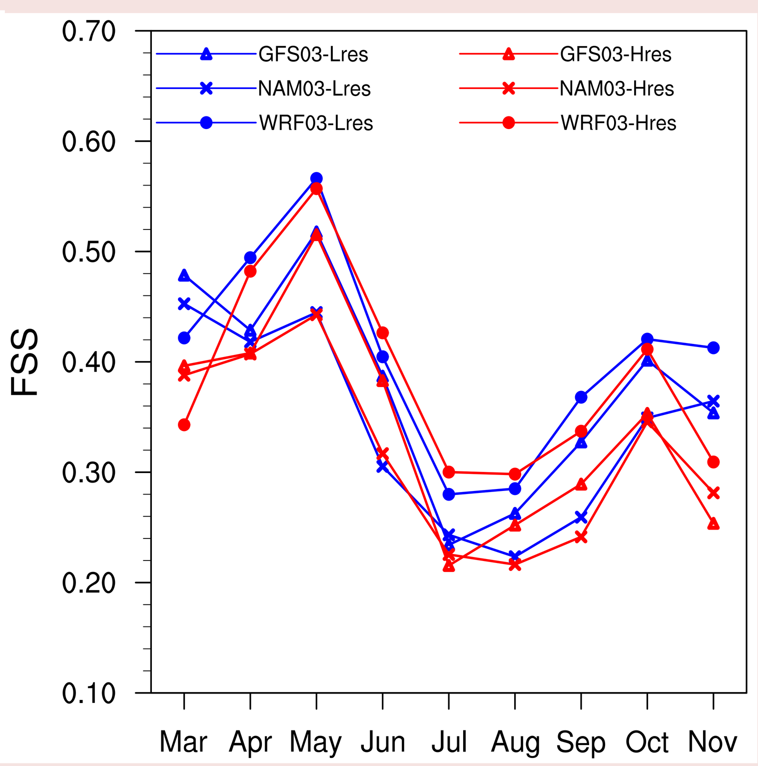

Fig. 3: Monthly variation of 3h mean FSS for Hres (red) and Lres (blue) verification grids at the smoothing size of 69 (Hres) and 5 (Lres) grid spacings

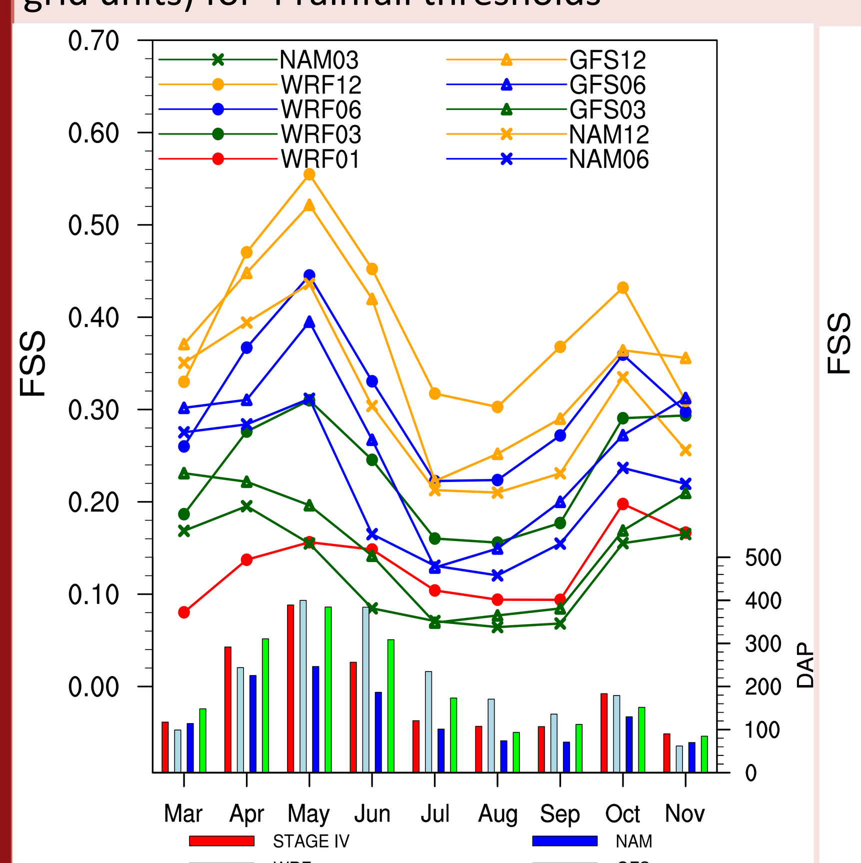

Fig. 2: Monthly variation of domain averaged rainfall volume (DAP) and 3h FSS at threshold of 6.35mm and 25 grid spacing neighborhood size

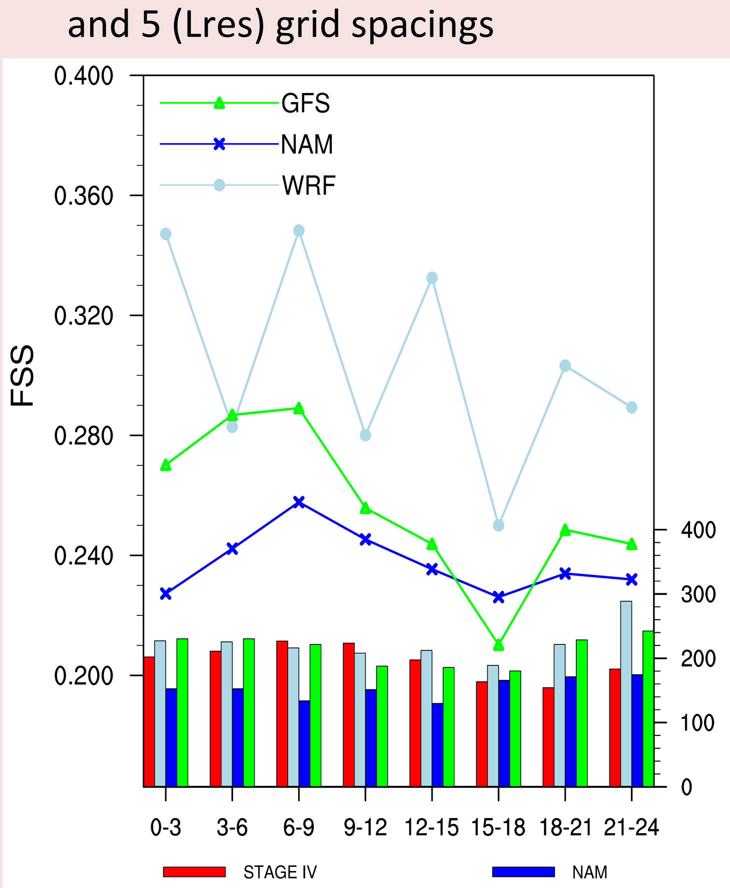

Fig. 4: Same as Fig. 2, but for diurnal mean FSS at threshold of 2.54mm

- High-res WRF had a large advantage for high thresholds, but GFS with the coarsest res had a better performance than NAM due to the dry bias of NAM.
- Skill increases more as accumulation interval increases, than for spatial scale increasing.
- For this sample of cases, FSS varied more due to wetness of the month than due to cold season vs warm (unlike other metrics in other studies)
- Both finer grid spacing (WRF) and longer accumulation time can further reduce the difference between warm and cold seasons.
- WRF tended to predict heavy precip volume during warm season even though obs showed much less rain in June-Sept.
- At the same neighborhood scale, FSS is not influenced much by grid on which verification is done
- The lowest skill happened when the rain volume had only small bias errors (late morning), so displacement or area/shape errors may be the main cause. But during late afternoon, the intensity errors are larger, and likely reduce the mean FSS.

5. Method for Object-based Diagnostic Evaluation (MODE)

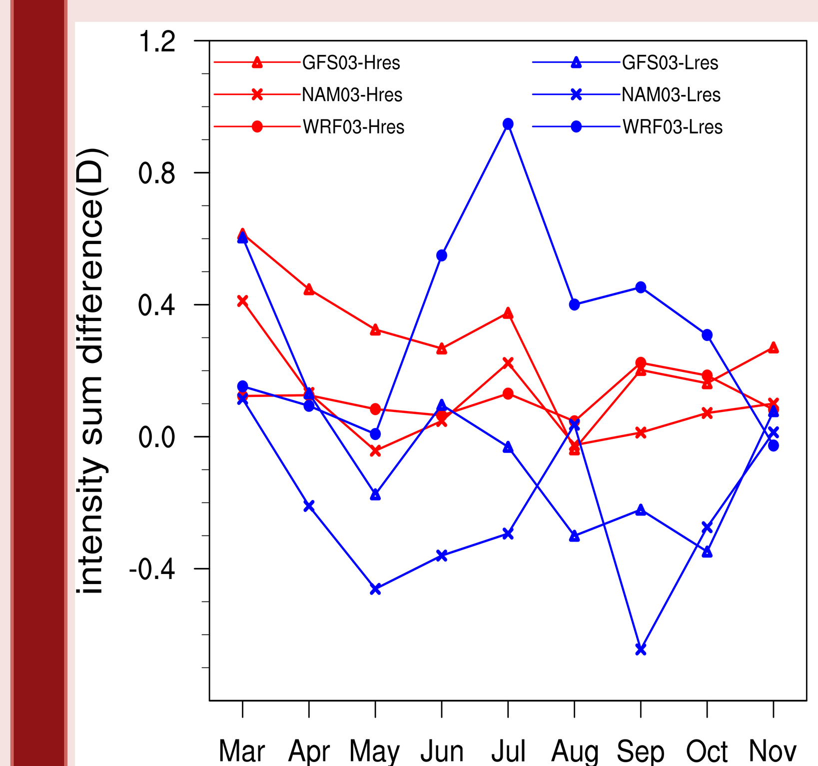

Fig. 5: Monthly mean intensity sum diff. (from MODE) between matched pairs of Hres and Lres

- Lres largely increased the variation between models and months in Hres
- Models on their own grid had a smaller wet bias.
- In Hres, all the models overpredict the intensity for matched storms

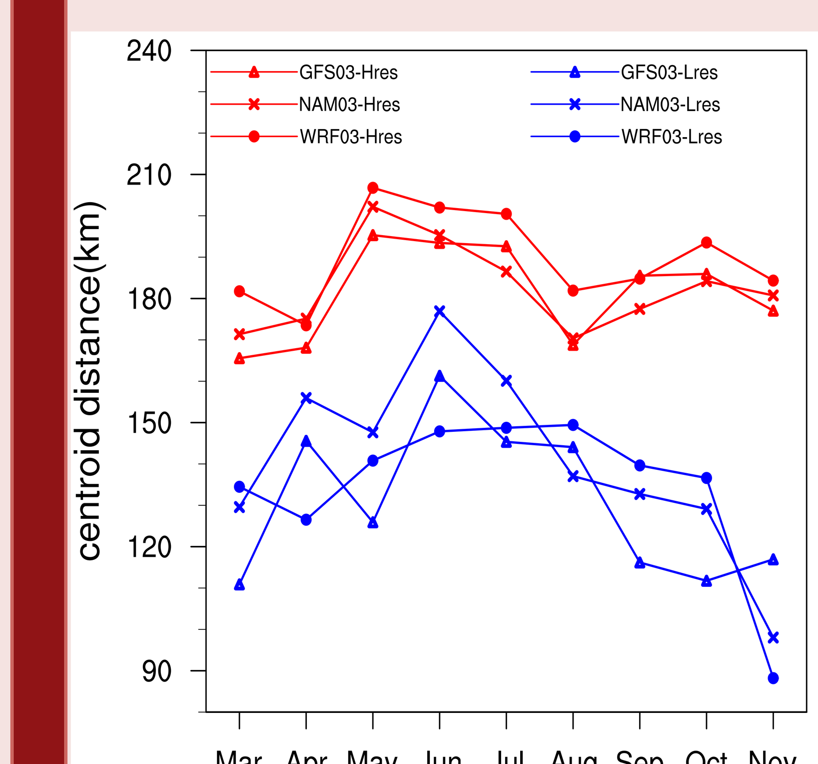

Fig. 6: Same as Fig.5, but for centroid errors

- Displacement error was evident during May and June even with high skill scores
- Choice of verification grid had large impact in MODE. Coarser grid largely reduces displacement bias

6. Conclusions

- High resolution WRF model runs with ARPS radar assimilation had higher skill in flood cases. NAM had an obvious dry bias resulting in the lowest skill among the three models.
- FSS strongly correlated to observed intensity sum, especially for higher resolution models and longer accumulation times.
- Choice of verification grid did not affect general FSS scores, but coarse resolution largely reduces displacement errors and increases intensity sum errors indicated by MODE.

Acknowledgements

This research was supported by the Iowa Flood Center. The rainfall observations (Precipitation NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data) are provided by NCAR/EOL under sponsorship of the National Science Foundation. http://data.eol.ucar.edu/.