PRECIPITATION FORECASTS USING THE WRF-ARW AND WRF-NMM MODELS DURING THE HMT-WEST 2006 AND 2007 WINTER EXPERIMENTS

Huiling Yuan[1, *], Christopher Anderson[1, 2], Paul Schultz [1], Isidora Jankov[1, 2], and John McGinley[1]
1. NOAA/Earth System Research Laboratory (ESRL)/Global Systems Division (GSD)
2. Cooperative Institute for Research in the Atmosphere (CIRA), Colorado State University

1. INTRODUCTION

The NOAA Hydrometeorological Testbed (HMT) program (http://hmt.noaa.gov) is a long-term project, which links the research community with the operational weather and hydrological forecasting. The HMT-West winter experiments focus on the winter storms over the American River Basin (ARB) of Northern California. The NOAA/ESRL/GSD operated a time-lagged multimodel ensemble system for providing weather forecasting during the HMT-West 2006 and 2007 campaign.

Based on the HMT-West 2006 operations, ensemble forecasts were reconfigured for the HMT-West 2007 real-time forecasts. Multiple microphysical schemes and dynamic cores were used in weather research and forecasting (WRF, http://www.mmm.ucar.edu/wrf/users/) model. The four model runs include the WRF with Nonhydrostatic Mesoscale Model (NMM) dynamic core, which uses the Ferrier microphysics, and with the advanced research WRF (ARW) dynamic core, which uses the Ferrier, Thompson, and Schultz microphysics. All models were initialized per 6 h with the Local Analysis and Prediction System (LAPS) diabatic hot start and run to 30 h at 3-km resolution over Northern California (150 x 150 grids). Convective schemes were not used at 3–km resolution. Boundary conditions were from 40-km North American Mesoscale (NAM) model.

The 6-h/24-h quantitative precipitation forecasts (QPF) and probabilistic QPF (PQPF) from individual models and time-lagged multimodel ensembles were analyzed for the selected intensive operation periods (IOPs) during the two winters. The NCEP Stage IV 6-h precipitation analyses were selected as the observations. Both microphysical schemes and dynamic cores showed high impacts on precipitation forecasts over this mountainous area. Time-lagged multimodel ensembles can effectively use the previous runs, and the LAPS system avoids the "spin-up" problem. This ensemble system improves the QPF compared to individual models, especially for 6-h accumulations.

* Corresponding author address: Huiling Yuan, NOAA/ESRL, R/GSD7, 325 Broadway, Boulder, CO, 80305-3328. email: huiling.yuan@noaa.gov

2. DATA

The 4-km NCEP Stage IV 6-h precipitation analyses were bilinearly interpolated to the 3-km common grid of the model output and LAPS. Forecasts from the NMM-Ferrier, ARW-Ferrier, ARW-Thompson, and ARW-Schultz runs were analyzed for the four rerun IOPs during the HMT-West 2006 were rerun and four archived IOPs during the operational HMT-West 2007. The total 91 validation times are available for 6-h QPF accumulations during the eight IOPs.

3. METHOD

Verification metrics used in this study can be found in two books (Wilks 2006; Jolliffe and Stephenson 2003), including the root mean square errors (RMSE), equitable threat score (ETS), and Brier skill score (BSS). The RMSE measures forecast errors between the observations and corresponding forecasts. The ETS and BSS measure forecast quality for a dichotomous event.

The ETS is defined as:

$$ETS = \frac{a - a_r}{a - a_r + b + c} \quad (3.1),$$

where a is the number of correct forecasted events, b is the number of false alarm events, c is the number of missing events, d is the number of correct rejection, and n is the total number of events ($n = a + b + c + d$); and a_r is the number of hits for random forecasts, given by:

$$a_r = \frac{(a+b)(a+c)}{n} \quad (3.2).$$

The Brier score (BS) measures the deviation

between the forecast and observed probabilities and is computed by:

$$BS = \frac{1}{n}\sum_{j=1}^{n}(p_j - o_j)^2 \quad (3.3),$$

where p_j is forecast probability, and o_j =1 when the observation exceeds a selected threshold, otherwise o_j =0. The BSS is referenced to a forecast system or sample climatology (verification):

$$BSS = 1 - \frac{BS}{BS_{ref}} \quad (3.4).$$

Both the ETS and BSS are positively orientated, with the best value of 1 and skillful values of being positive. Five thresholds 0.01, 0.1, 0.5, 1, and 2 inches (i.e., 0.254, 2.54, 12.7, 25.4, 50.8 mm) were selected to calculate the ETS and BSS, in which sample climatology was computed on each grid pixel during the study period for deriving BS_{ref}.

4. RESULTS

The LAPS initialization avoids the "spin-up" in the model and the forecast errors grow with the lead time. With the increasing forecast lead time, the RMSE of 6-h precipitation accumulations increases for each model (Fig. 1). The NMM-Ferrier model has the largest error and the ARW-Thompson model shows the smallest error among single models. The multimodel ensemble mean from the four models has smaller error than any individual model at each lead time. The spatial correlation coefficients (Fig. 2), which were computed for corresponding observation and forecast fields at each validation time and averaged, show much better results for the multimodel ensemble mean. The correlation coefficients dramatically decrease after the 6-h lead time, slight vary from the 12-h to 24-h lead times, and increase from the 30-h lead time. The study domain is marked with orographic precipitation. After the 6-h lead time, the spatial patterns are confined to the topography and slightly vary.

Similarly, the multimodel ensemble mean outperforms individual models for the ETS (Fig. 3), especially for lower thresholds (0.01-0.5 inches, Figs. 3a-3c). Generally, the ETS decreases with the increasing lead time. At the extreme high threshold (2 inches/6-h, Fig. 3e), the skill largely drops after the 6-h lead time. The Ferrier scheme has similar performance as the Thompson scheme at lower thresholds (Figs. 3a-3c), but not for higher thresholds (Figs. 3d, 3e). The Schultz microphysical scheme has the lowest skill at low thresholds (Figs. 3a, 3b) and better skill at higher thresholds (Figs. 3c-3e). The ETS of 24-h QPF has similar skill for the 6-h and 12-h lead times (Figs. 4a, 4b). The NMM-Ferrier model has the best 24-h ETS for the thresholds up to 1 inch/24-h. The multimodel ensemble mean does not gain big skill improvements for 24-h QPF as the 6-h QPF.

The BSS with the 6-h lead time is calculated for each model using five time-lagged members (QPF with the five lead times). The performance is similar for individual models at five selected thresholds. The multimodel ensemble uses twenty members including four models with five time-lagged members, which shows higher skill than all individual models. Considering the 90% confidence intervals, the BSS for the time-lagged multimodel system is skillful up to 1 inch/6-h (Fig. 5a-5d).

5. SUMMARY

Based on the WRF models with the NMM and WRF dynamic cores and multiple microphysics, the forecasts during the HMT-West 2006 and 2007 were assessed for eight IOPs. The 6-h QPF from two dynamic cores with the same microphysics are comparable at all lead times, while the NMM shows better skill for 24-h QPF at lower thresholds than the ARW. Using the same dynamic core, forecasts with different microphysics show discrepancies in terms of different verification metrics. The 6-h PQPF for using time-lagged members from each model are comparable, while time-lagged multimodel ensembles show better results than any individual model. More tests are needed for constructing an optimal ensemble system.

6. ACKNOWLEDGEMENTS

This research was performed while the first author held a National Research Council Research Associateship Award at NOAA/ESRL.

7. REFERENCES

Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification. A Practitioner's Guide in Atmospheric Science*. Wiley and Sons Ltd.

Wilks, D. S., 2006: Statistical Methods in the Atmospheric Sciences. 2d ed. Elsevier Academic Press, 627 pp.

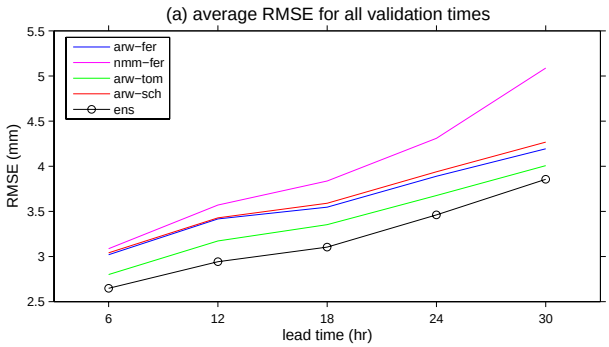

Fig. 1. (left) The RMSE with increasing lead times for 6-h QPF from ARW-Ferrier (blue), NMM-Ferrier (pink), ARW- Thompson (green), ARW-Schultz (red), and the multimodel mean at each lead time (black with circles).

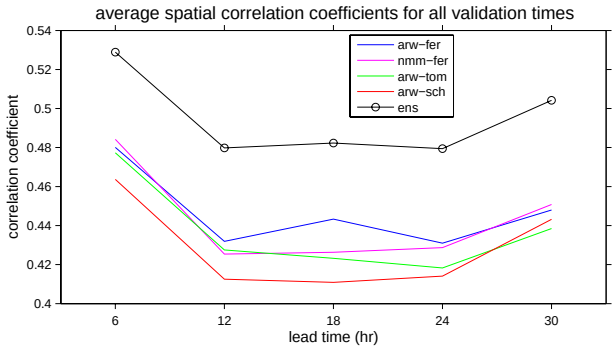

Fig. 2. (right panel) Same as Fig. 1 but for averaged 6-h spatial correlation coefficients.

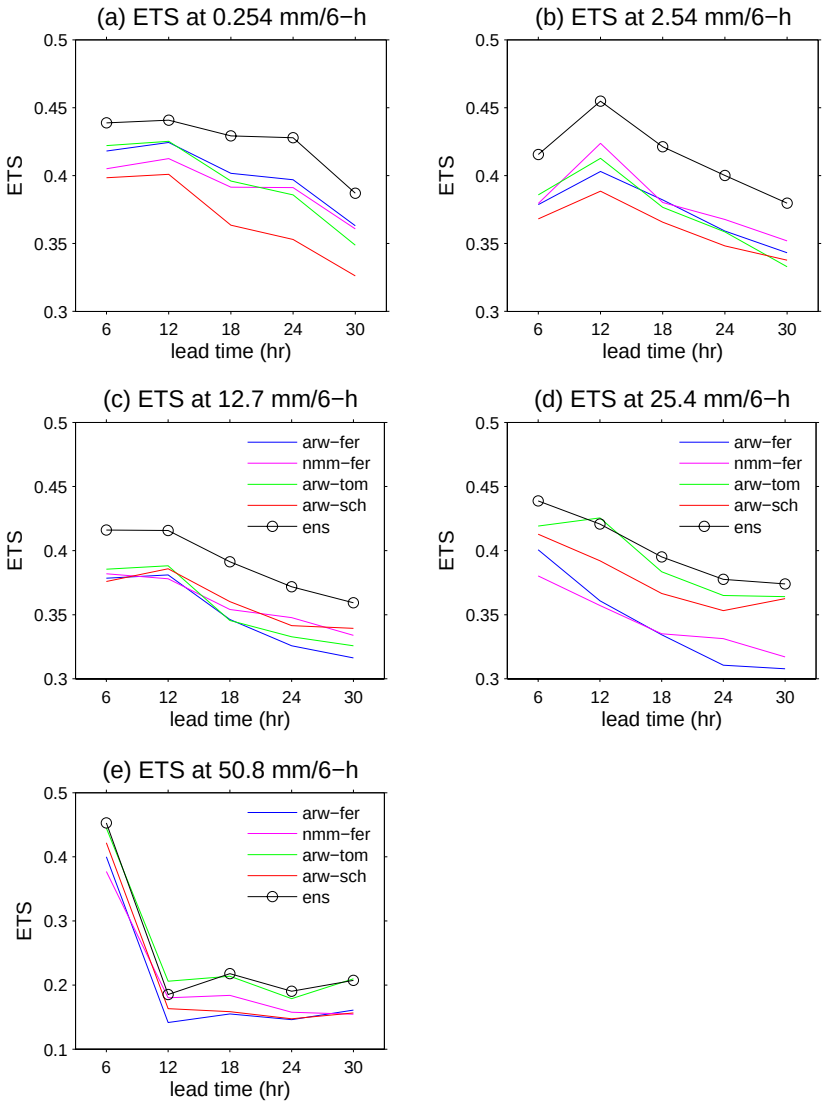

Fig. 3. The ETS with increasing lead times for 6-h QPF from the ARW-Ferrier (blue), NMM-Ferrier (pink), ARW- Thompson (green), ARW-Schultz (red), and multimodel mean at each lead time (Black with circles) at five thresholds: 0.254 mm (a), 2.54 mm (b), 12.7 mm (c), 25.4 mm (d), and 50.8 mm (e) per 6 h.

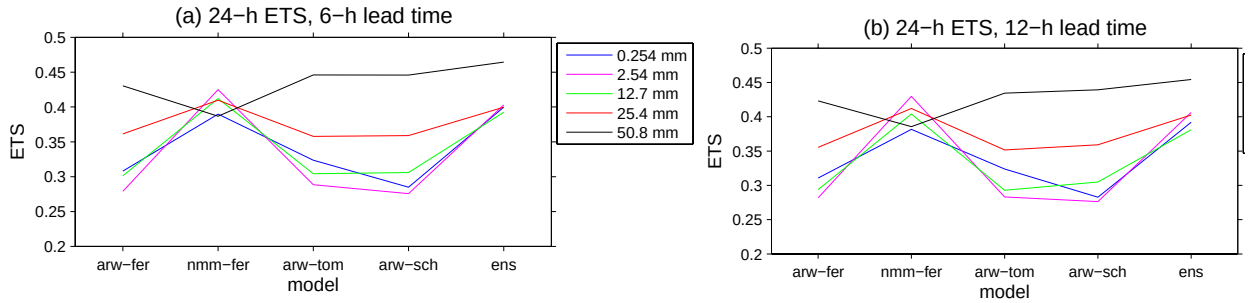

Fig. 4. The ETS for 24-h QPF at five thresholds: 0.254 mm (blue), 2.54 mm (pink), 12.7 mm (green), 25.4 mm (red), and 50.8 mm (black) per 6 h, with the 6-h lead time (a) and 12-h lead time (b). The abscissa shows the names of the models - arw-fer: ARW-Ferrier, nmm-fer: NMM-Ferrier, arw-tom: ARW-Thompson, arw-sch: ARW-Schultz, and ens: multimodel ensemble mean at each lead time.

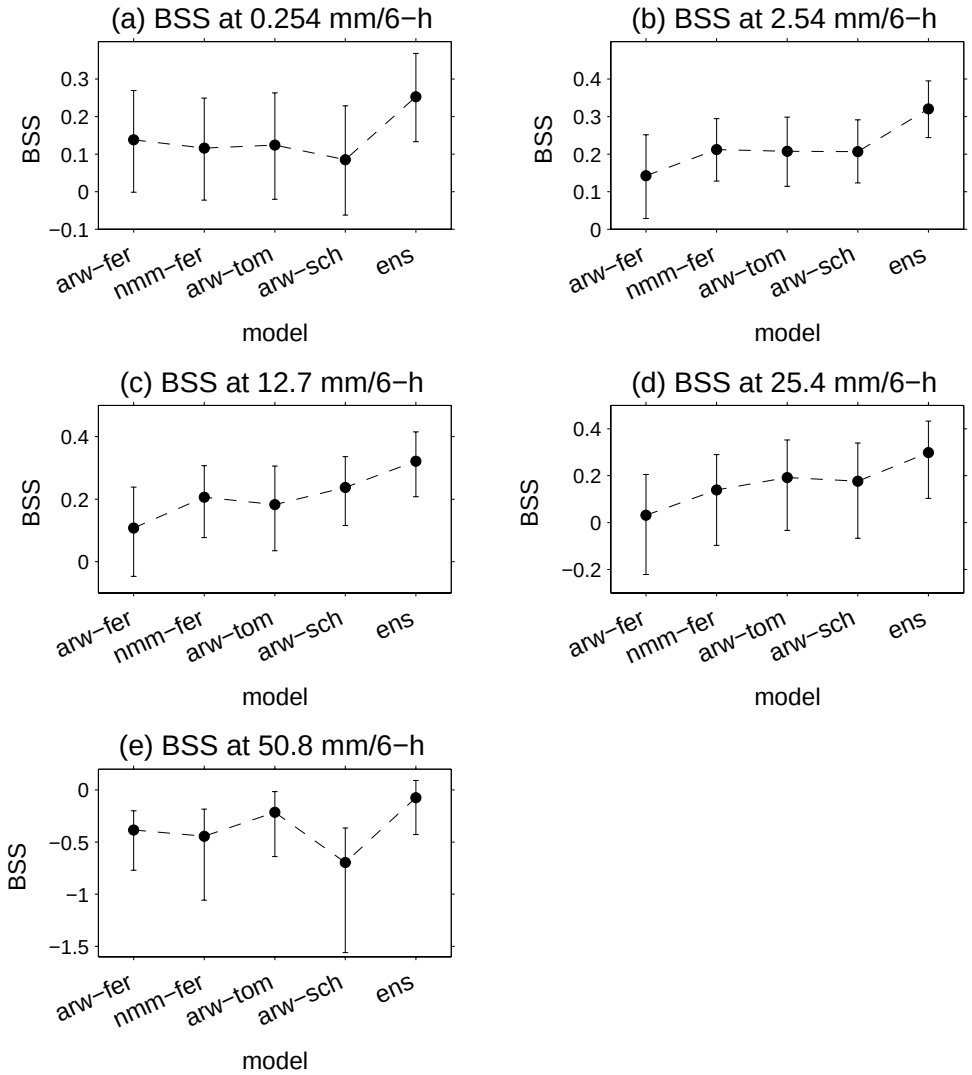

Fig. 5. The BSS of 6-h PQPF with a 6-h lead time for five selected thresholds: 0.254 mm (a), 2.54 mm (b), 12.7 mm (c), 25.4 mm (d), and 50.8 mm (e) per 6 h. The abscissa shows the names of the models - arw-fer: ARW-Ferrier, nmm-fer: NMM-Ferrier, arw-tom: ARW-Thompson, arw-sch: ARW-Schultz, and ens: multimodel ensemble. The error bars indicate the 90% confidence intervals.