Evaluating the Performance of RBMP-based Multi-model Ensemble Forecasts

Hong Guan[1,2], Bo Cui[2,3], Yuejia(Jiajia) Zhu[4], Yuejian Zhu[2], Wen Meng[2,3]

[1]SRG at EMC/NCEP/NWS, [2]EMC/NCEP/NWS, [3]IMSG at EMC/NCEP/NWS 5830 University Research Court, College Park, MD 20740, USA; [4]Anhui Meteorological Bureau/CMA, Hefei, China

1. Motivation

- The use of multi-model ensembles is becoming an attractive practice. Operationally, the NAEFS collaboration (Candille 2009) is a successful example that equally combines NCEP/GEFS and CMC ensemble.
- We expect to unequally blend three ensemble prediction systems by adding one more global ensemble FNMOC, which is supported by the NUOPC, a Tri-Agency (NOAA, Navy, Air Force) effort to improve interoperability.
- Proper scoring rules are sought to investigate the predictive performance.

2. Recursive Bayesian Model Process (RBMP)

- Bayesian Model Averaging (BMA) is a useful technique considering model uncertainty. Raftery et al. (2005) extended BMA from statistical models to dynamic models, greatly simplify the process of computing posterior probabilities.
- On account of computational/storage expensive and models upgrade over time in operation, we adapt the station-based BMA code from Bruce Veenhuis of MDL to global grid-based RBMP. For the *n*th iteration:

$$\hat{z}_{kij}^{(n)} = \frac{w_k^{(n-1)} g(y_{ij} \mid f_{kij}, \sigma^{(n-1)})}{\sum_{l=1}^{K} w_l^{(n-1)} g(y_{ij} \mid f_{lij}, \sigma^{(n-1)})} \qquad SD^{(n)} = \sqrt{\sum_{k=1}^{K} \hat{z}_{kij}^{(n)} (y_{ij}^{(n)} - f_{kij}^{(n)})^2}$$

$$w_k^{(n)} = (1-\alpha)\cdot w_k^{(n-1)} + \alpha\cdot \hat{z}_{kij}^{(n)} \qquad \sigma^{(n)} = (1-\alpha)\cdot\sigma^{(n-1)} + \alpha\cdot SD^{(n)}$$

- 2nd moment adjustment is also iteratively applied using decaying average method in RBMP.

$$\bar{S}(t) = \sqrt{(1-\beta)\cdot\bar{S}^2(t-1)+\beta\cdot s^2(t)} \qquad \bar{E}(t) = \sqrt{(1-\beta)\cdot\bar{E}^2(t-1)+\beta\cdot e^2(t)}$$

$$R_{ij}(t) = \frac{\bar{E}(t)}{\bar{S}(t)} (if\ \bar{S}(t)=0, R_{ij}(t)=1) \qquad D^m = (F_{ij}{*}^m(t) - \overline{F_{ij}{*}}(t))$$

$$F_{ij}^m(t) = F{*}_{ij}^m(t) + (R_{ij}(t)-1)\cdot D^m (m=1,\ldots,M)$$

1st moment adjusted forecast — 2nd moment adjustment

3. Verification Methodology

- Continuous ranked probability score (CRPS), which is negatively oriented, the smaller the better with the perfect value of 0. It can be viewed as a mean absolute error of a probability distribution.

$$CRPS(P,x_a) = \int_{-\infty}^{+\infty} (P(x) - 1_{\{x\ge x_a\}})dx$$

- Relative Operating Characteristic (ROC) area, the area under the ROC curve, is a useful measure of forecast skill, where ROC reflects the ability of the forecasts to discriminate between two alternative events, thus can be considered as measuring resolution. Perfect: ROC area=1, No skill: ROC area=0.5.
- The spread-skill relationship is also investigated to see how well the ensemble is being designed to represent uncertainty as realistic as possible. For a 'perfect ensemble', the ensemble spread should be equal to RMSE over the same period.

4. Data

- Ensemble forecasts: bias-corrected GEFS(v11), CMC(v10) and FNMOC(v10) for 2m temperature (every 12h out to 384h), initialized on 1 September 2013, and the verification period is from December 1, 2013 through November 30, 2014.
- Analysis/observation: ERA-interim/CONUS observations.

5. Performance against Analysis

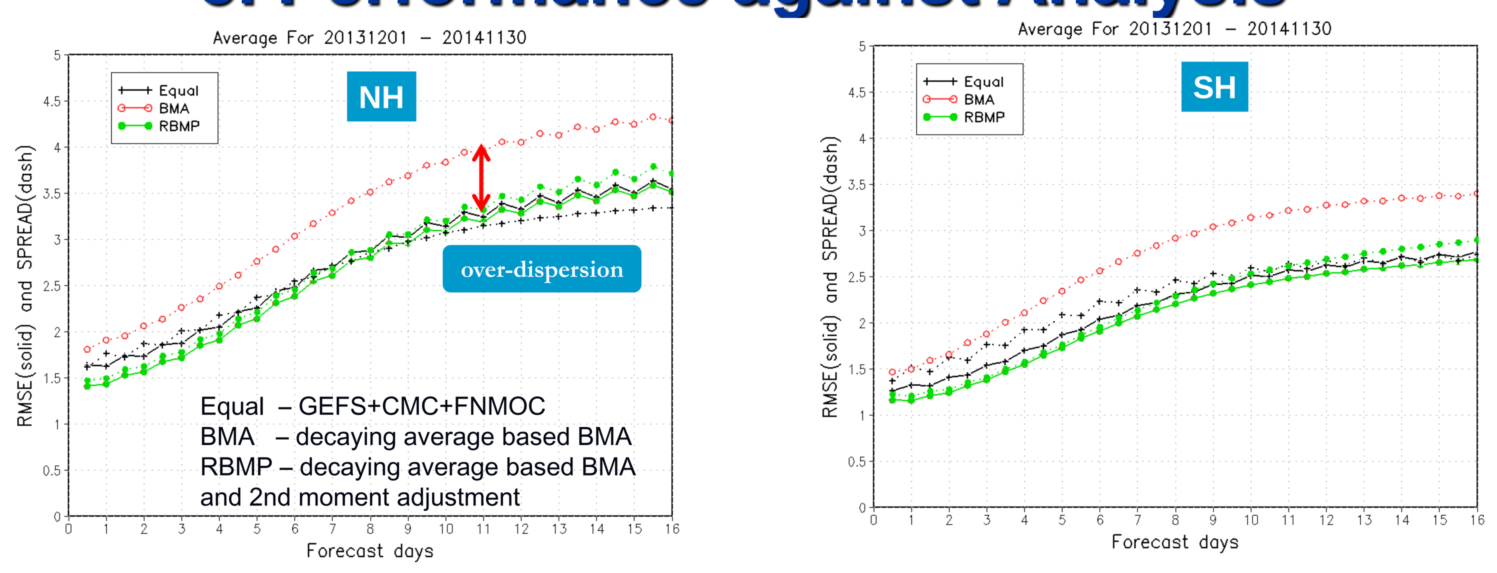

FIG.1. RMSE and spread averaged for one year over NH (left) and SH (right).

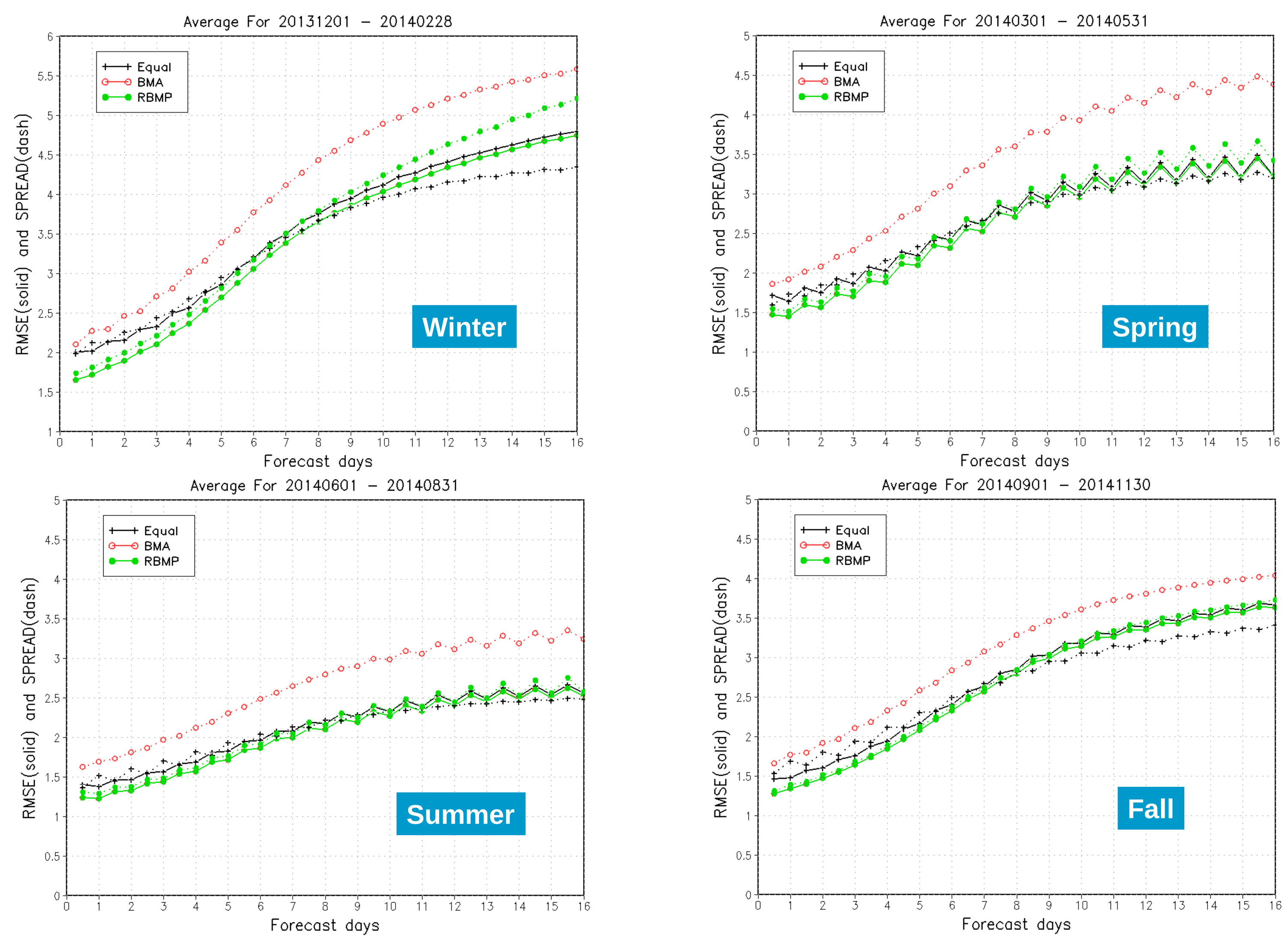

FIG.2. RMSE and spread averaged for each season over NH.

- BMA could improve ensemble mean forecast at all lead times, especially for short lead times, but it could increase ensemble spread to be over-dispersive.
- RBMP keeps the feature of BMA, the ensemble spread could be properly modified by 2nd moment adjustment as well.

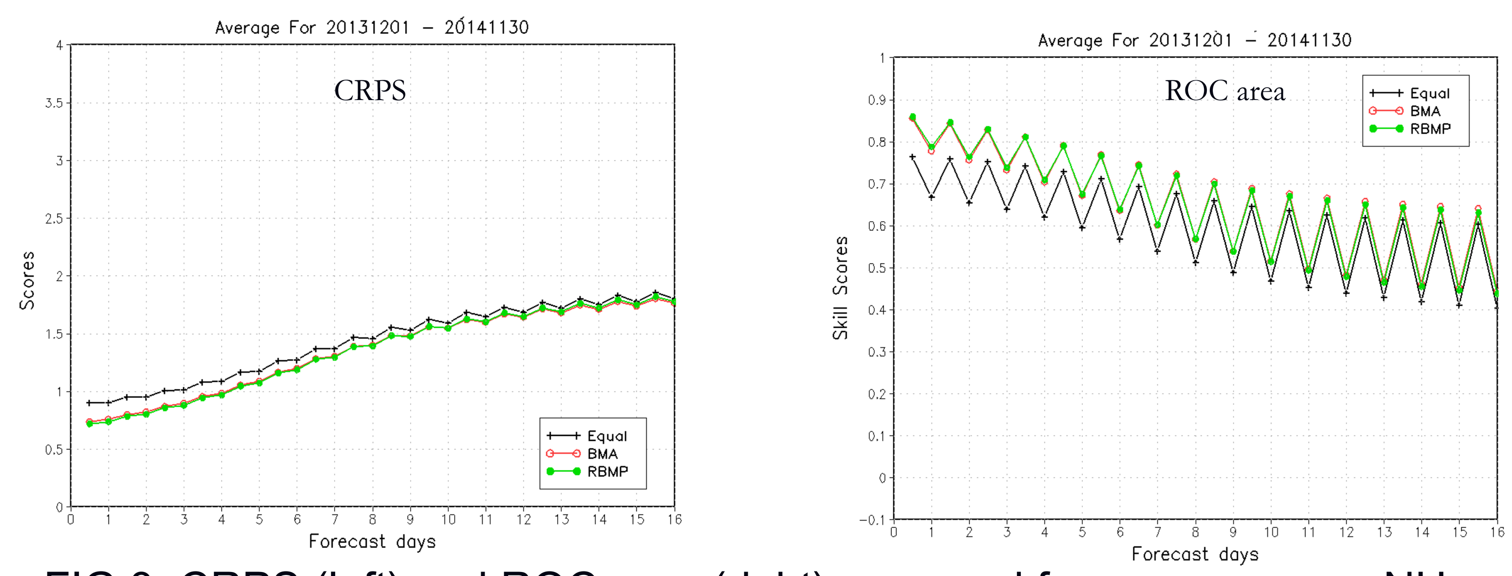

FIG.3. CRPS (left) and ROC area (right) averaged for one year over NH.

- In terms of CRPS and ROC area, forecast skill is improved by both BMA and RBMP with more effectiveness in short lead times.

6. Performance against Observation

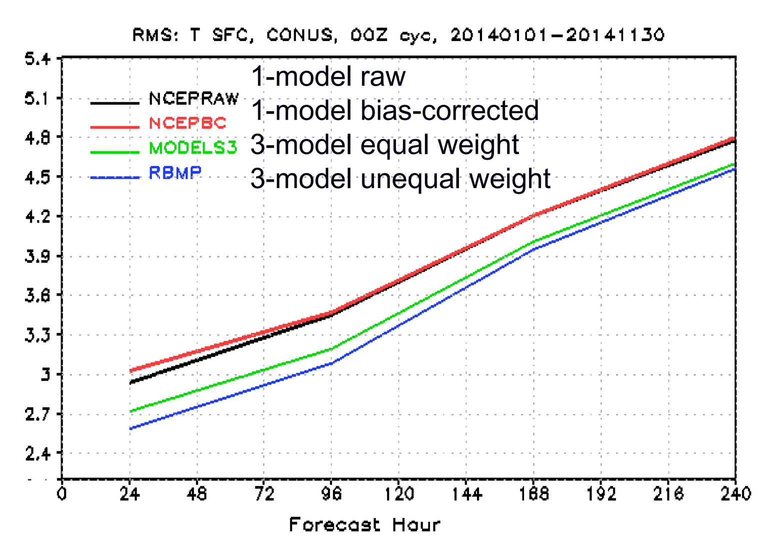

FIG.4. RMSE based on CONUS observations averaged for 11 months.

- It also confirms that RBMP could outperform equal weighted mean of ensembles.

7. Discussion and Conclusion

- RBMP is an effective way to conduct multi-model ensembles with larger improvement in short lead times.
- Further work should explore if RBMP or similar process can be applied to non-normal variables such as wind or precipitation.
- Contacts: Hong.Guan@noaa.gov