Extended-range Probabilistic 500hPa Geopotential Height Forecasting over Northern Hemisphere Using Bayesian Model Averaging

Lu-ying JI¹ and Xie-fei ZHI¹,²,*

¹Key Laboratory of Meteorological Disasters, Ministry of Education (KLME) / Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC-FEMD), Nanjing University of Information Science & Technology, Nanjing, China
²Nanjing Joint Center for Atmospheric Research (NJCAR), Nanjing, China
*Corresponding author

Keywords: 500hPa geopotential height, Bayesian model averaging (BMA), Extended-range weather forecasts.

Abstract. ECMWF, NCEP and UKMO EPS data from the TIGGE datasets were used to conduct the Bayesian model averaging (BMA) extended-range (10-15 days) probabilistic forecasts for 500hPa geopotential heights over Northern Hemisphere. The results show that the multimodel BMA method performed better than raw ensemble and the forecast skill varied as season changes. The forecast skill was the best in summer and worst in winter. However, the skill score of BMA with respect to the raw ensemble indicated that the improvement of the forecast in winter was the highest. Hence, the BMA method can improve the worse raw ensemble to a larger extent. In addition, BMA provided a full probability distribution, which depicted the quantitative uncertainty of the forecasts. With the increase of latitude, the uncertainty of the forecast was increasing.

Introduction

According to the time scale, the forecasting systems of the meteorological department can be divided into short-term weather forecasts (1-3 days), mid-term weather forecasts (4-10 days), extended-range weather forecasts (10-30 days) and short period climate predictions above month scale. The development of 10-30d extended-range weather forecast relatively lags behind, but it is of great importance in the precaution and reduction of meteorological disasters. Extended-range weather is difficult to forecast because its lead times have exceeded the theoretical limit of daily weather forecasts [1].

Madden-Julian Oscillation (MJO) is one of the most widely used measures for extended-range forecasts [2,3], but it has a poor performance in high latitudes. If we focus on the daily weather forecasts, it still mainly depends on the numerical models. From the perspective of the numerical weather prediction (NWP), the improvement of extended-range forecasts mainly lies on the extension of the lead times. Ensemble forecasts can effectively realize it, so that they have become an important method for extended-range forecasting. Miyakoda et al. [4] successfully used the numerical models to forecast 10-30d averaged blocking high pressure, whose work is viewed as the beginning of the extended-range numerical forecasts and facilitates the emergence of extended-range weather forecasts at major forecast centers around the world. Further researches were continually conducted [5-7]. Recently, multimodel ensemble techniques like super-ensemble and bias-removed ensemble mean based on TIGGE datasets are proved to be effective methods to improve the extended-range forecasting skill [8,9].

Taking consideration of the long lead times of extended-range forecasts and the limits of deterministic forecasts, it may be better to pay more attention to the probabilistic forecasts for extended-range weather. Bayesian model averaging (BMA) described by Raftery et al. [10] is a statistical postprocessing method for producing probabilistic forecasts in the form of a predictive probability density function (PDF). Studies applying the BMA method to weather forecasts or climate predictions have demonstrated that the BMA forecasts were well calibrated and performed better than the raw ensemble forecasts. [11-15]. To date, BMA has mainly been used for short-term temperature
and precipitation forecasts or climate predictions. However, little previous work has been done on the application of BMA to extended-range 500hPa geopotential heights and little attention has been paid to the difference in the BMA’s improvement with respect to the raw ensemble as the season changes.

In the present study, the BMA method was applied to the multimodel ensemble systems in which ECMWF, NCEP and UKMO EPS from TIGGE datasets were incorporated. And then we mainly focus on the BMA’s ability for probabilistic forecasts in various seasons.

Data and Methods

(a) Data

This research used the ECMWF ERA-Interim reanalysis 500hPa geopotential height products at 1200 UTC on a 2.5°×2.5° horizontal resolution as the verifying data and ECMWF, NCEP and UKMO EPS multimember 500hPa geopotential height forecasts with lead times of 10-15 days with initial time 1200 UTC obtained from TIGGE datasets as the forecast data. The test period lasted from 1 January 2011 to 31 December 2013. The study area is the whole North Hemisphere. Detailed information of the three single-center EPSs is listed in Table 1.

| Center | Country     | Model            | Ensemble members (perturbed) | Lead time (in days) |
|--------|-------------|------------------|------------------------------|--------------------|
| ECMWF  | Europe      | T339L62/T255L62  | 50                           | 15                 |
| NCEP   | United States| T216L28          | 20                           | 16                 |
| UKMO   | United Kingdom|                | 23                           | 15                 |

(b) Bayesian Model Averaging

Bayesian model averaging is a statistical postprocessing method that can combine multiple statistical models for joint inferences and predictions, and also can provide a highly sharp probability distribution. Raftery et al. [10] applied the BMA method to 48-h forecasts of surface air temperature and sea level pressure in the Pacific Northwest using ensemble forecasts, showing BMA could yield calibrated and sharp predictive PDFs. According to the BMA theory, the BMA predictive PDF $p(y)$ can be written as

$$p(y) = \sum_{k=1}^{K} p(y|M_k)p(M_k|y^T)$$

where $y$ is the predictive variable, $K$ is the number of models being combined, $p(y|M_k)$ is the component forecast PDF based on model $M_k$, $p(M_k|y^T)$ is the posterior probability of model $M_k$ reflecting how well model $M_k$ fits the training data such that it can be viewed as the weight of model $M_k$. Thus $p(M_k|y^T)$ is nonnegative and $\sum_{k=1}^{K} p(M_k|y^T) = 1$.

For geopotential heights, $p(y|M_k)$ can be seen as a normal distribution. Then the details of parameter estimation can refer to Raftery et al. [10]. In our implementation, the training period is set as a sliding window, the result of which is better than that of a fixed one. After several experiments, we chose 45 days as the sliding training period.

(c) Verification Methods

In this study, we used mean absolute error (MAE), continuous ranked probability score (CRPS) and their corresponding skill score (MAESS and CRPSS) to evaluate BMA’s forecasts [16]. The first two metrics are negatively oriented, namely, better performance means smaller values. The specific formula is as follows.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |f_i - o_i|$$

where $f_i$ is the forecast value and $o_i$ is the observation, $N$ is the total number of samples. MAE is used for deterministic forecasts.
CRPS = \frac{1}{n} \sum_{i=1}^{n} \int [f(y_i) - H(y_i - o_i)]^2 dx \quad (3)

where \(H(y_i - o_i)\) is a Heaviside function that jumps from 0 to 1 at the observed value. If \(y_i < o_i\), \(H(y_i - o_i) = 0\), otherwise, \(H(y_i - o_i) = 1\). CRPS is defined to measure the overall probabilistic performance.

Skill score is a measure of the relative improvement of the BMA forecasts over low-skilled raw ensemble forecasts. The higher skill score means the better forecasts. For MAE and CRPS, the skill score is calculated as:

\[ SS = 1 - \frac{\text{Score(forecast)}}{\text{Score(raw ensemble)}} \quad (4) \]

Results

(a) PDF

Unlike methods that can only predict the probability of exceeding a certain threshold, BMA gives a full probability distribution, as is shown in Fig. 1. As can be seen, BMA prediction PDF is a weighted sum of the component PDFs. The observation fell within the 90% prediction interval. The reason why extended-range forecasts is not as accurate as short-to-medium-range forecasts is due to the larger forecast uncertainty. One of the biggest advantages of BMA method is its ability to quantify the forecast uncertainty according the full PDF. The BMA error, namely, forecast uncertainty, is defined as the BMA 90% prediction interval (dotted lines).

Figure 1. BMA predictive PDF (black curve) and its three components (the rose red curve is ECMWF, the green curve is NCEP and the blue curve is UKMO) for 500hPa geopotential height at 120°E, 35°N on 21 July 2012. Also shown are the BMA 90% prediction interval (dotted lines) and the verifying observation (solid vertical line).

The spatial distribution of BMA forecast uncertainty in different seasons was shown in Fig. 2. Taking the 10-day forecast as an example, the BMA error, in various seasons is different. Generally, the uncertainty in summer and autumn was relatively smaller than that in spring and winter. It can be seen that the forecast uncertainty in polar regions was larger in summer, but the uncertainty in low and middle latitudes were the smallest in the four seasons. In addition, the uncertainty of BMA forecast increased from low to high latitudes whatever the season was.
In order to compare the differences between BMA performances in various seasons, Fig.3 illustrates the forecast skill of BMA in various seasons with lead times from 10 to 15 days. The differences of BMA model between various seasons were fairly obvious in extended-range forecasts. The BMA performance in summer was the best with the smallest MAE and CRPS value, then autumn and spring followed, and winter was the worst, which is similar to Fig.2. This is because the forecast skill of raw ensemble in summer was much superior to that of the other three seasons and then autumn, spring and winter followed (figure omitted). BMA model is constructed on the basis of raw ensemble, thus the performance differences of BMA model in four seasons are as same as those of the raw ensemble. In addition, and the differences seemed to become larger as lead time increased.

Figure 3. Mean verification metrics for BMA forecasts in different seasons with lead time of 10 days. (a) MAE; (b) CRPS.

Figure 4 shows the skill score of BMA model with respect to the raw ensemble. The skill score of MAE and CRPS with the lead times of 10-15 days are both above 0, illustrating the BMA’s capacity to improve extended-range weather prediction skill. Although the BMA forecast skill was the worst in winter, the improvement in winter was the highest for the whole extended-range period. It means BMA method can play an important role in improving the prediction skill for the poor raw ensemble. For 10-12-day forecast, the improvement in spring was the smallest, while for 13-15-day forecast the improvement in summer was the smallest. The skill score of CRPS was much higher than that of MAE, which demonstrated the BMA’s advantages as a statistical probabilistic method of postprocessing raw ensemble.
Summary

In this study, we applied the BMA method to TIGGE multi-center EPSs, and obtained BMA predictive systems for 500hPa geopotential heights with lead times of 10-15 days. BMA can yield a full prediction PDF, which is a weighted sum of the members’ PDFs. Based on the PDF, BMA can also describe the forecast uncertainty quantitatively. The forecast uncertainty increased from low to high latitudes over Northern Hemisphere.

BMA is originally designed to produce probabilistic forecasts, but as a by-product it also produces deterministic forecasts (median forecasts). Whatever for probabilistic forecasts or deterministic forecasts, the BMA model outperformed the raw ensemble forecasts in our experiments. The forecast skill decreased with the increase of lead times.

As the lead time increased, there was a growing significantly difference between various seasons. The performance was the best in summer, and the worst in winter both for BMA model and raw ensemble forecasts. However, the extended-range skill score in winter being the highest indicated that BMA method can improve the prediction skill of the poor raw ensemble to a larger extent.

Generally, the BMA forecasts based on multimodel EPSs greatly improved the quality of extended-range 500hPa geopotential height forecasts. With the development of probabilistic forecast methodologies and the use of multimodel ensemble forecast, extended-range forecasting will become well developed.

In this study, we only chose ECMWF, NCEP and UKMO EPS from TIGGE datasets to implement the experiments. This is due to the limited number of forecast centers providing extended-range numerical weather prediction (NWP). It is widely acknowledged that the model’s prediction ability in a week is the key to the extended-range weather forecasts [17]. Naturally how to improve numerical model may be the major problem of increasing the forecast skill of extended-range weather forecasts. As a result, more EPSs providing extended-range products can be used as the BMA members to achieve a better performance.

Acknowledgement

This research was financially supported by the National Natural Science Foundation of China (41575104) and NJCAR key project (NJCAR2016ZD04).

References

[1] E.N. Lorenz, Atmospheric predictability experiments with a large numerical model, Tellus. 34 (1982) 505-513.
[2] M.B. Galin, Study of the low-frequency variability of the atmospheric general circulation with the use of time-dependent empirical orthogonal functions, Izvestiya Atmos and Oceanic Physics. 43 (2007) 15-23.

[3] B.N. Goswami, P.K. Xavier, Potential predictability and extended range prediction of Indian summer monsoon breaks, Geophys. Res. Lett. 30 (2003), doi: 101029/2003GL017810.

[4] K. Miyakoda, T. Gordon, R. Caverly, W. Stern, J. Sirutis, and W. Bourke, Simulation of a Blocking Event in January 1977. Mon. Wea. Rev. 111 (1983) 846-869.

[5] M. Déquá, J.F. Royer, The skill of extended-range extratropical winter dynamical forecasts, J. Climate. 5 (1992) 1346-1356.

[6] T.N. Palmer, Extended-Range Atmospheric Prediction and the Lorenz Model, Bull. Amer. Meteor. Soc. 74 (1992) 49-65.

[7] M.S. Tracton, K. Mo, W. Chen, E. Kalnay, R. Kistler, and G. White, Dynamical Extended Range Forecasting (DERF) at the National Meteorological Center, Mon. Wea. Rev. 117 (1989) 1604-1635.

[8] H.H. Cui, X.F. Zhi, Multi-model ensemble forecasts of surface air temperature in the extended range using the TIGGE dataset (in Chinese), Trans. Atmos. Sci. 36 (2013) 165-173.

[9] X.F. Zhi, H.X. Qi, Y.Q. Bai, C.Z. Lin, A comparison of three kinds of multimodel ensemble forecast techniques based on the TIGGE data, Acta. Meteor. Sinica. 26 (2012) 41-51.

[10] A.E. Raftery, T. Gneiting, F. Balabdaoui, M. Polakowski, Using bayesian model averaging to calibrate forecast ensembles, Mon. Wea. Rev. 133 (2005) 1155-1174.

[11] L.Y. Ji, X.F. Zhi, S.P. Zhu, Extended-range probabilistic forecasts of surface air temperature over East Asia during boreal winter (in Chinese), Trans. Atmos. Sci. 40 (2017) 346-355.

[12] S.K. Min, A. Hense, A Bayesian approach to climate model evaluation and multi-model averaging with an application to global mean surface temperatures from IPCC AR4 coupled climate models, Geophys. Res. Lett. 330 (2006) 153-172.

[13] S.K. Min, D. Simonis, A. Hense, Probabilistic climate change predictions applying Bayesian model averaging, Philos. Trans. Roy. Soc. 365 (2007) 2103-2116.

[14] L.J. Wilson, S. Beauregard, A.E. Raftery, R. Verret, Calibrated surface temperature forecasts from the Canadian ensemble prediction system using bayesian model averaging, Mon. Wea. Rev. 135 (2007) 4226-4230.

[15] J.M. Sloughter, A.E. Raftery, T. Gneiting, C. Fraley, Probabilistic quantitative precipitation forecasting using bayesian model averaging, Mon. Wea. Rev. 135 (2007) 3209-3220.

[16] D.S. Wilks, Statistical methods in the atmospheric sciences, second ed., Academic Press, 2006.

[17] M.P. Baldwin, D.B. Stephenson, D.W. Thompson, T.J. Dunkerton, A.J. Charlton, A. O'Neill, Stratospheric memory and skill of extended-range weather forecasts, Science, 301(2003) 636-640.