

A Comparison of Three Kinds of Multimodel Ensemble Forecast Techniques Based on the TIGGE Data

ZHI Xiefei^{1*}(智协飞), QI Haixia¹(祁海霞), BAI Yongqing¹(白永清), and LIN Chunze²(林春泽)

¹ Key Laboratory of Meteorological Disaster of Ministry of Education,
Nanjing University of Information Science & Technology, Nanjing 210044

² Wuhan Institute of Heavy Rain, China Meteorological Administration, Wuhan 430074

(Received November 23, 2010; in final form December 12, 2011)

ABSTRACT

Based on the ensemble mean outputs of the ensemble forecasts from the ECMWF (European Centre for Medium-Range Weather Forecasts), JMA (Japan Meteorological Agency), NCEP (National Centers for Environmental Prediction), and UKMO (United Kingdom Met Office) in THORPEX (The Observing System Research and Predictability Experiment) Interactive Grand Global Ensemble (TIGGE) datasets, for the Northern Hemisphere (10° – 87.5° N, 0° – 360°) from 1 June 2007 to 31 August 2007, this study carried out multimodel ensemble forecasts of surface temperature and 500-hPa geopotential height, temperature and winds up to 168 h by using the bias-removed ensemble mean (BREM), the multiple linear regression based superensemble (LRSUP), and the neural network based superensemble (NNSUP) techniques for the forecast period from 8 to 31 August 2007.

The forecast skills are verified by using the root-mean-square errors (RMSEs). Comparative analysis of forecast results by using the BREM, LRSUP, and NNSUP shows that the multimodel ensemble forecasts have higher skills than the best single model for the forecast lead time of 24–168 h. A roughly 16% improvement in RMSE of the 500-hPa geopotential height is possible for the superensemble techniques (LRSUP and NNSUP) over the best single model for the 24–120-h forecasts, while it is only 8% for BREM. The NNSUP is more skillful than the LRSUP and BREM for the 24–120-h forecasts. But for 144–168-h forecasts, BREM, LRSUP, and NNSUP forecast errors are approximately equal. In addition, it appears that the BREM forecasting without the UKMO model is more skillful than that including the UKMO model, while the LRSUP forecasting in both cases performs approximately the same.

A running training period is used for BREM and LRSUP ensemble forecast techniques. It is found that BREM and LRSUP, at each grid point, have different optimal lengths of the training period. In general, the optimal training period for BREM is less than 30 days in most areas, while for LRSUP it is about 45 days.

Key words: multimodel superensemble, bias-removed ensemble mean, multiple linear regression, neural network, running training period, TIGGE

Citation: Zhi Xiefei, Qi Haixia, Bai Yongqing, et al., 2012: A comparison of three kinds of multimodel ensemble forecast techniques based on the TIGGE data. *Acta Meteor. Sinica*, **26**(1), 41–51, doi: 10.1007/s13351-012-0104-5.

1. Introduction

As the atmosphere is a nonlinear dissipative system, the numerical weather predictions are restricted by the model physical parameterizations, initial errors, boundary problems, etc. Therefore, it may take quite a long time to improve the weather forecast skill for a mature single model and that is why scientists have long before put forward the idea of ensemble fore-

casting (Lorenz, 1969; Leith, 1974; Toth and Kalnay, 1993).

Nowadays, numerical weather prediction is developing from traditional deterministic forecast to ensemble probabilistic forecast. Along with the rapid development of communication, networking, and computers as well as other technologies, international cooperation in weather forecasting has become much closer, especially when The Observing System Research and

Supported by the China Meteorological Administration Special Public Welfare Research Fund (GYHY(QX)2007-6-1) and National Key Basic Research and Development (973) Program of China (2012CB955204).

*Corresponding author: zhi@nuist.edu.cn.

©The Chinese Meteorological Society and Springer-Verlag Berlin Heidelberg 2012

Predictability Experiment (THORPEX) Interactive Grand Global Ensemble (TIGGE) data are available. TIGGE is a key component of THORPEX, and the latter is contained in the WMO (World Meteorological Organization) World Weather Research Programme. THORPEX aims to accelerate the improvements in the accuracy of 1-day to 2-week high-impact weather forecasts. The TIGGE project has been initiated to enable advanced research and demonstration of the multimodel ensemble concept and to pave the way toward operational implementation of such a system at the international level (Park et al., 2008; Bougeault et al., 2010).

Krishnamurti et al. (1999) proposed a so-called multimodel superensemble forecasting method, which is a very effective post-processing technique able to reduce direct model output errors. In his subsequent multimodel superensemble forecast experimentation of 850-hPa winds, precipitation, and track and intensity of tropical cyclones, it was revealed that the superensemble forecast significantly reduced the errors compared with the individual models and the multimodel ensemble mean (Krishnamurti et al., 2000a, b, 2003, 2007a). The 24–144-h superensemble forecasts of 500-hPa geopotential height indicate that the superensemble achieved a higher ACC (Anomaly Correlation Coefficient) skill than the best single model forecast. Rixen and Ferreira-Coelho (2006) conducted a superensemble of multiple atmosphere and ocean models by utilizing linear regression and nonlinear neural network techniques, and made the short term forecasting of sea surface drift along the west coast of Portugal. Their results indicate that the superensemble of the atmosphere and ocean models significantly reduced the error of 12–48-h sea surface drift forecast. Cartwright and Krishnamurti (2007) pointed out that the 12–60-h superensemble forecast of precipitation in the southeastern America during summer 2003 is more accurate than that of each single model. In the superensemble forecast of precipitation during the onset of the South China Sea monsoon, Krishnamurti et al. (2009a) found that the superensemble forecasts of precipitation and extreme precipitation during landfalling typhoons exhibited a higher forecast skill

than the best single model forecast. Further studies by Zhi et al. (2009a, b) show that the forecast skill of the multimodel superensemble forecast with running training period is higher than that of the traditional superensemble forecast for the surface temperature forecast in the Northern Hemisphere midlatitudes during the summer of 2007. After various forecast experiments, it was proved that the superensemble method may significantly improve the weather and climate prediction skills (Stefanova and Krishnamurti, 2002; Mishra and Krishnamurti, 2007; Krishnamurti et al., 2007, 2009; Rixen et al., 2009; Zhi et al., 2010).

However, it could be possible for an individual model ensemble to outperform a multimodel ensemble containing poor models (Buizza et al., 2003). Therefore, the multimodel ensemble forecast technique and its applications need to be further investigated. It is necessary to study comparatively the characteristics of different multimodel ensemble forecasting schemes.

2. Data and methodology

2.1 Data

The data used in this study are the daily ensemble forecast outputs of surface temperature, and 500-hPa geopotential height, temperature, and winds at 1200 UTC from the European Centre for Medium-Range Weather Forecasts (ECMWF), the Japan Meteorological Agency (JMA), the US National Centers for Environmental Prediction (NCEP), and the United Kingdom Met Office (UKMO), in the TIGGE archive. The characteristics of the four models involved in the multimodel ensemble forecast are listed in Table 1 in accordance with Park et al. (2008).

The forecast data of each model cover the period of 1 June to 31 August 2007, with the forecast area in the Northern Hemisphere (10° – 87.5° N, 0° – 360°), the horizontal resolution of $2.5^{\circ} \times 2.5^{\circ}$, and the forecast lead time of 24–168 h. The NCEP/NCAR reanalysis data for the corresponding meteorological variables were used as “observed values”. Note the area and the horizontal resolution of the NCEP/NCAR reanalysis data are consistent with those of the TIGGE data.

Table 1. Characteristics of the four TIGGE ensembles

Operation center	Initial perturbation method (area)	Horizontal resolution	Forecast length (day)	Perturbation members
UKMO	BVs (globe)	T213	10	15
JMA	BVs (NH+TR)	T319	9	51
ECMWF	SVs (globe)	TL399/TL255	0–10/10–15	102
NCEP	BVs (globe)	T126	16	84

Note: BVs stand for Bred Vectors and SVs for Singular Vectors. NH stands for Northern Hemisphere and TR for tropics.

2.2 Methodology

2.2.1 Linear regression based superensemble forecasting

The multimodel superensemble forecast is formulated after Krishnamurti et al. (2000a, 2003). At a given grid point, for a certain forecast time and meteorological element, the superensemble forecasting model can be constructed as:

$$S_t = \bar{O} + \sum_{i=1}^n a_i (F_{i,t} - \bar{F}_i), \quad (1)$$

where S_t represents the real-time superensemble forecast value, \bar{O} the mean observed value during the training period, $F_{i,t}$ the i th model forecast value, \bar{F}_i the mean of the i th model forecast value in the training period, a_i the weight of the i th model, n the number of models participating in the superensemble, and t is time.

The weight a_i can be calculated by minimizing the function G in Eq. (2) below with the least square method. The acquired regression coefficient a_i will be implemented in Eq. (1), which creates the superensemble forecasts in the forecasting period.

$$G = \sum_{t=1}^{N_{\text{train}}} (S_t - O_t)^2. \quad (2)$$

It should be noted that the traditional superensemble employs a fixed training period of a certain length, while an improved superensemble proposed by Zhi et al. (2009a) applies a running training period, which chooses the latest data of a certain length for the training period right before the forecast day. The linear regression based superensemble using running training period will be abbreviated as LRSUP hereafter.

2.2.2 Nonlinear neural network based superensemble forecasting

In addition to the linear regression method, the three-layer back propagation (BP) of nonlinear neural network technique (Geman et al., 1992; Warner

and Misra, 1996) is implemented for the superensemble forecast (hereafter abbreviated as NNSUP). During the training period, the output from each model is taken as the input for the neural network learning matrix. During the forecast period, the well-trained network parameters are carried into the forecast model to obtain the multimodel superensemble forecasting (Stefanova and Krishnamurti, 2002; Zhi et al., 2009b).

2.2.3 Bias-removed ensemble mean and multimodel ensemble mean

Bias-removed ensemble mean (hereafter abbreviated as BREM) is defined as

$$\text{BREM} = \bar{O} + \frac{1}{N} \sum_{i=1}^N (F_i - \bar{F}_i), \quad (3)$$

where BREM is the bias-removed ensemble mean forecast value, and N the number of models participating in the BREM. The running training period is also adopted in the BREM technique.

In addition, the multimodel ensemble mean (hereafter abbreviated as EMN) is performed and used as a cross-reference for the superensemble forecasts.

$$\text{EMN} = \frac{1}{N} \sum_{i=1}^N F_i. \quad (4)$$

In the verification of the single model forecasts and evaluation of the multimodel ensemble forecasts, the root-mean-square error (RMSE) is employed.

$$\text{RMSE} = \left[\frac{1}{n} \sum_{i=1}^n (F_i - O_i)^2 \right]^{\frac{1}{2}}, \quad (5)$$

where F_i is the i th sample forecast value, and O_i is the i th sample observed value.

3. Results

3.1 Comparative analyses of linear and nonlinear superensemble forecasts

Based on the ensemble mean outputs of the 24–

168-h ensemble forecasts of surface temperature in the Northern Hemisphere from the ECMWF, JMA, NCEP, and UKMO, the multimodel superensemble forecasting was carried out for the period of 8–31 August 2007 (24 days). The length of the running train-

ing period was set to be 61 days.

As shown in Fig. 1, for the 24–168-h forecasts in the entire forecast period, the superensemble forecasts (LRSUP and NNSUP) together with the multimodel EMN and BREM reduced the RMSEs by some means

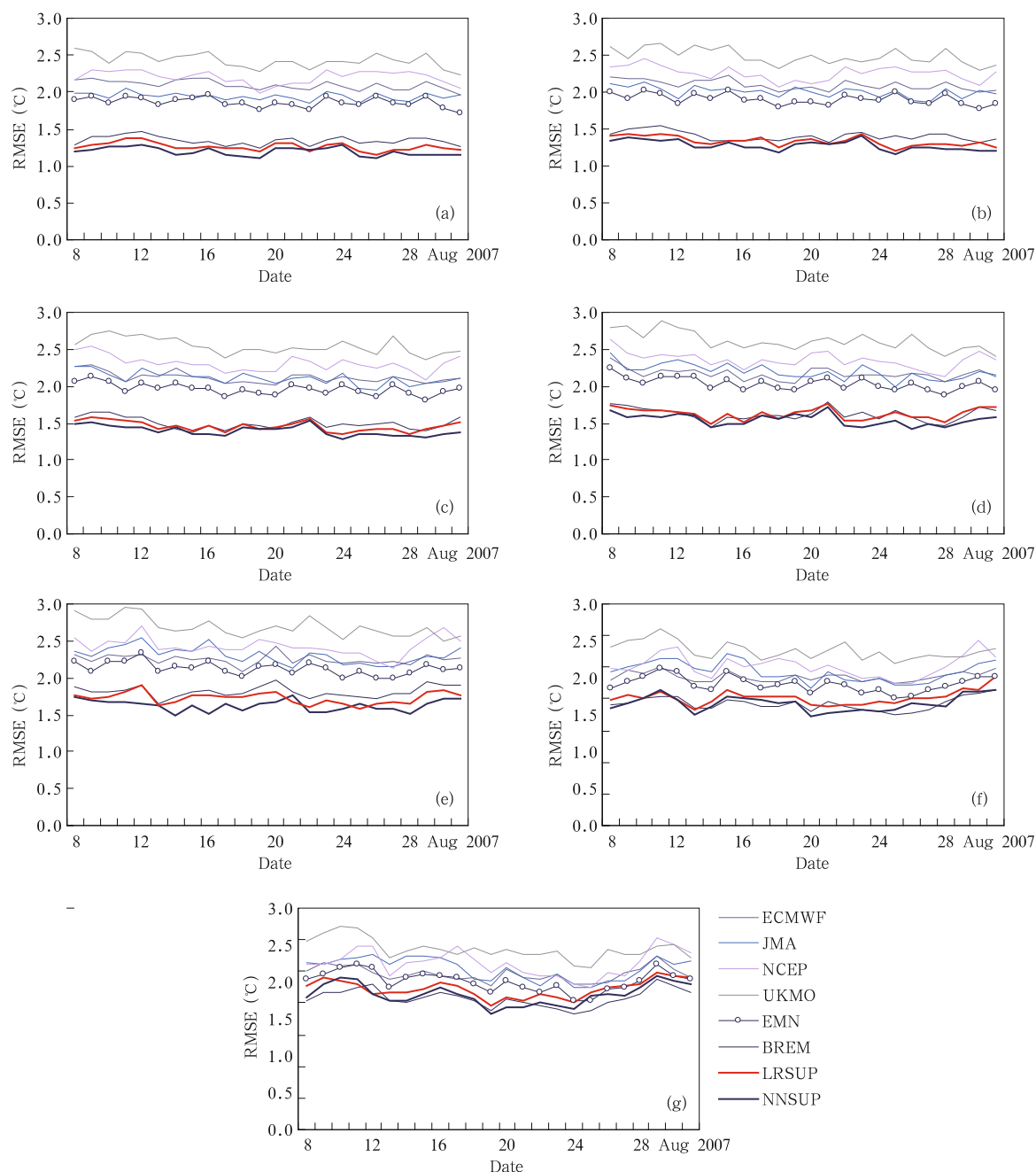


Fig. 1. RMSEs of the surface temperature forecasts from the ECMWF, JMA, NCEP, and UKMO together with the multimodel ensemble mean (EMN), bias-removed ensemble mean (BREM), linear regression based superensemble (LRSUP), and neural network based superensemble (NNSUP) at (a) 24 h, (b) 48 h, (c) 72 h, (d) 96 h, (e) 120 h, (f) 144 h, and (g) 168 h from 8 to 31 August 2007 in the Northern Hemisphere (10° – 80° N, 0° – 360°).

compared with the single model forecasts. With the extension of forecast lead time, the forecast skill improvement decreased. For the 24–120-h forecasts, the RMSEs of the LRSUP and NNSUP were much smaller than those of the single model forecasts, and the BREM also had improved forecast skills to some extent. When the forecast lead time was longer, e.g., 144–168 h, the BREM forecast skill caught up with that of LRSUP and NNSUP in terms of the RMSEs. Therefore, for the summer Northern Hemisphere surface temperature, the multimodel ensemble forecast performed better than the single model forecast. Although the forecast improvement decreased with the forecast lead time, the forecast results remained stable. The NNSUP was skillful and outperformed the BREM and LRSUP for the 24–120-h forecasts. But for the 144–168-h forecasts, the errors of the BREM, LRSUP, and NNSUP were approximately equal.

In addition, the above analysis shows that the NNSUP was reasonably better than the other forecast schemes, because the NNSUP scheme might have reduced the forecast errors caused by the nonlinear effect among various models. However, repetitive adjustment of the neural network parameters had to be performed to obtain the optimal network structure, which slowed down the operation efficiency. Meanwhile, BREM and LRSUP had the advantage of being computationally simple with reasonable accuracy; they were easier to be implemented for forecasters in their daily operation. In the following, the multimodel ensemble forecast schemes of BREM and LRSUP (hereafter further abbreviated as SUP) will be employed to give a multimodel consensus forecast of

500-hPa geopotential height and temperature as well as the zonal and meridional wind fields for comparative analysis.

3.2 SUP and BREM forecasting of 500-hPa geopotential height

The SUP method for 24–72-h forecasts of the 500-hPa geopotential height had a high forecast skill. Especially, for 24-h forecast, it performed much better than the optimal single model forecasting (figure omitted). Figure 2a shows that the average RMSE of the 96-h SUP forecasts was very close to that of the best single model ECMWF forecast, while the BREM had a lower forecast skill than the ECMWF forecast most of the time. For longer forecast lead time, this was also the case (figure omitted).

The overall low forecast skill of the SUP and BREM at longer than 96-h forecast lead time may be attributed to the difference in the forecasting capability of each model in different latitudes as well as the systematic errors of the models. In addition, it is unreasonable that the length of the training period at all grid points is fixed at 61 days for the SUP and BREM. In the following, the optimal length of the running training period will be examined at each grid point before the SUP and BREM forecasts are conducted.

As shown in Fig. 2, the RMSEs of the superensemble with optimized training (O-SUP) are smaller to some extent than those of the superensemble without optimized training (SUP). The optimal BREM (O-BREM) forecast is also better than the BREM forecast.

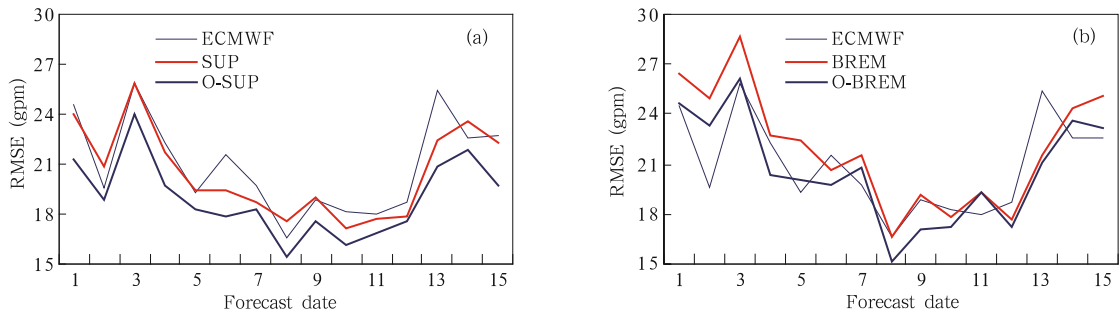


Fig. 2. RMSEs of the 96-h forecasts of geopotential height at 500 hPa from the ECMWF by using (a) the superensemble with (O-SUP) and without (SUP) optimized training and (b) the bias-removed ensemble mean with (O-BREM) and without (BREM) optimized training at each grid over the area 10° – 60° N, 0° – 360° . The ordinate denotes the RMSE and the abscissa denotes forecast date.

To sum up, the 24–72-h forecast experiments of 500-hPa geopotential height in the Northern Hemisphere show that the improvement of SUP over the individual models was more obvious, and the BREM forecast skill was somehow inferior to that of SUP. But for longer than 96-h forecast lead time, both SUP and BREM forecast schemes might not well reduce the overall errors in the region. However, after the length of the running training period was optimized at each grid point, the forecast errors declined to some extent.

Zhi et al. (2009b) indicated that for the 24–168-h superensemble forecasts of the surface temperature in the Northern Hemisphere, the optimal length of the training period is about two months. Since the forecast skill becomes lower when taking a longer training period in the BREM forecast, for shorter forecast lead time of 24–72 h, the most appropriate training period

is about half month. For 96–168-h forecasts, it is suitable to select about one month as the optimal length of the training period. This shows that the selection of the length of the training period is essential for the SUP and BREM forecasts. Only when appropriate length is selected, can the forecast error be reduced to a minimum. Too long or too short training periods may influence the forecast skill. For forecasts at different lead time, SUP and BREM forecasts also need different optimal training periods.

In order to obtain the best forecast skill, the optimal length of the training period should be determined for the SUP forecast. As the models involved in the multimodel ensembles contribute differently for different forecast regions, the length of the training period for each grid point should be tuned. As shown in Fig. 3, for most areas in the Northern Hemisphere, the

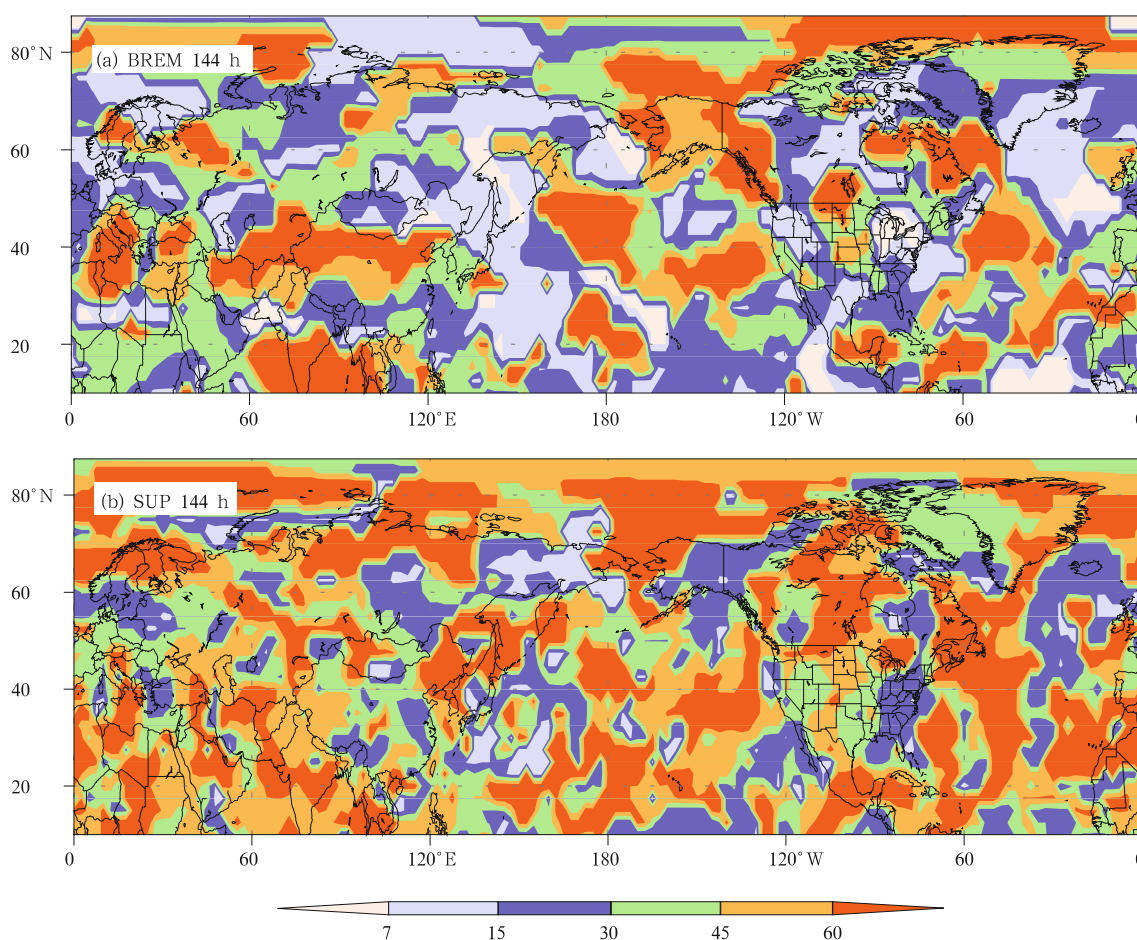


Fig. 3. Distributions of the optimal length (days) of the running training period for the 144-h forecasts of 500-hPa geopotential height using the (a) BREM and (b) SUP techniques over the Northern Hemisphere.

optimal length of the BREM training period is less than that of the SUP training period. Generally speaking, for both of the BREM and SUP schemes, the length of the training period changes significantly from one area to the other, which may be caused by different forecasting system errors associated with each model involved in the ensemble in different geographical regions. At present, due to lack of data, it is difficult to analyze the features of the changes in the optimal length of the training period for different seasons within each region.

3.3 Further comparison between SUP and BREM

Figure 4 shows the forecast RMSEs of the 500-hPa geopotential height, zonal wind, meridional wind, and temperature of the best single model, the EMN, the optimal BREM, and the optimal SUP averaged in the forecast period in the Northern Hemisphere excluding high latitudes (10° – 60° N, 0° – 360°). As shown in Fig. 4a, the RMSEs of the 24–168-h best single model forecasts of the 500-hPa geopotential height range from 10.4 to 37.7 gpm. The RMSE of SUP has always been the lowest for 24–120-h forecasts. Overall, the average error of the 24–120-h SUP forecasts is about 15 gpm, which reduces the RMSE by 16% compared with the

best single model forecasts. For 24–120-h forecasts, the BREM forecast has a lower skill than the SUP forecast.

However, when the forecast lead time is extended to 144–168 h, BREM has an approximately equal forecast skill as SUP. As shown in Figs. 4b–4d, similar conclusions are found for other variables at 500 hPa. For 144–168-h forecasts of 500-hPa temperature, the RMSEs of the SUP and BREM forecasts can still be reduced respectively by 8% and 10% compared with the best single model forecasts (Fig. 4d).

Figure 4 shows that SUP can effectively improve the forecast skill of all the studied variables at 500 hPa. For the 24–120-h forecasts, SUP is superior to BREM, EMN, as well as the best single model forecast. Especially, the 24–120-h forecast error of the 500-hPa geopotential height from SUP is 16% less than that of the best single model forecast, while that of BREM error is about 8%. For the 144–168-h forecasts, the SUP and BREM forecast skills are approximately equal.

3.4 Effect of the model quality on SUP and BREM

Krishnamurti et al. (2003) indicated that the best model involved in the superensemble contributes to

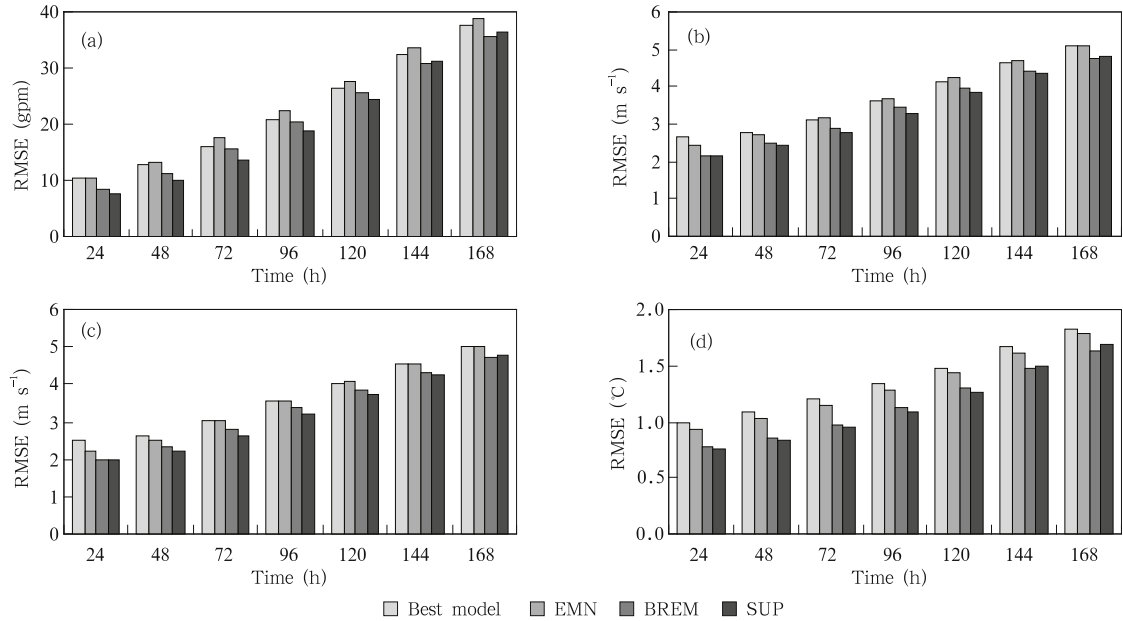


Fig. 4. RMSEs of (a) 500-hPa geopotential height, (b) zonal wind, (c) meridional wind, and (d) temperature forecasts for the best individual model, the EMN, the optimal BREM, as well as the optimal SUP averaged for the forecast period 17–31 August 2007 in the area 10° – 60° N, 0° – 360° .

approximately 1%–2% improvement for the superensemble forecast of the 500-hPa geopotential height, while the overall improvement of the superensemble over the best model is about 10%. This improvement in the superensemble is a result of the selective weighting of the available models during the training period. That is to say, the weight distribution of all the models contributes a lot to the improvement of the superensemble forecasting techniques. Then how will a poor model affect the SUP and BREM? From a detailed analysis of forecast errors in Fig. 1, we have found that the UKMO model has the largest errors among the four models participating in the ensemble. Now, two kinds of forecast schemes are designed to investigate the effect of model quality on the multimodel ensemble forecasting.

Scheme I: multimodel forecast including the UKMO forecast data.

Scheme II: removing the UKMO forecast data from the multimodel ensemble.

The training period and the forecast period are the same as the above for the two schemes. Figure 5 gives the RMSEs of the SUP and BREM using the original results of the multimodel data (4-SUP and 4-BREM), as well as SUP and BREM after removal of

the UKMO data (3-SUP and 3-BREM).

As shown in Fig. 5, the 3-BREM (without the UKMO model) forecast error is less than that of the 4-BREM (with UKMO). Take 24- and 168-h forecasts of 500-hPa geopotential height as examples, RMSEs have been reduced from 10.4 and 37.7 gpm to 7.8 and 34.6 gpm, respectively. The results show that the BREM forecast is sensitive to the performance of each model involved in the ensemble. The better the model involved is, the higher skill the multimodel ensemble forecast will have. Therefore, it is necessary to examine the performance of each model involved before conducting the BREM forecast.

However, as shown in Fig. 5, the skill of the 24–168-h SUP forecast of each variable differs from that of the BREM forecast. The forecast errors of 3-SUP without the UKMO model and 4-SUP including the UKMO model are approximately equal, i.e., the SUP forecast is not very sensitive to the poor model involved. The reason for this is that the SUP method itself requires the models participating in the ensemble to have certain spread. For the UKMO model, although the forecast error of this model is large, it is still within the spread. In addition, the poor model will be assigned a small weight in the SUP forecast.

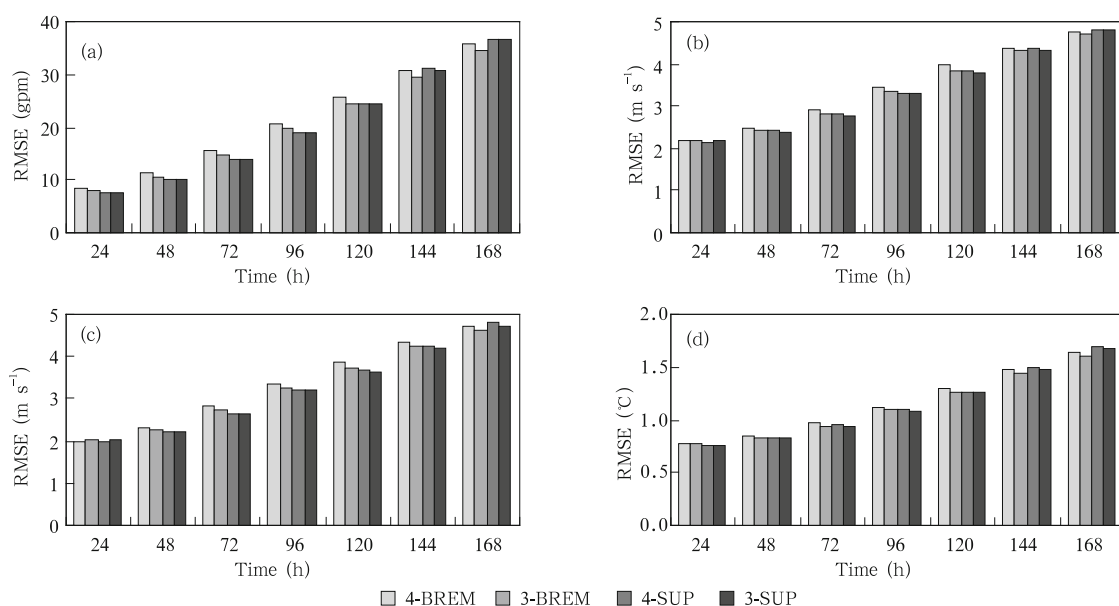


Fig. 5. RMSEs of (a) geopotential height, (b) zonal wind, (c) meridional wind, and (d) temperature at 500 hPa from the optimal 4-SUP and 4-BREM and the optimal 3-SUP and 3-BREM with the worst model data excluded from the multimodel suite. The results were averaged for the period 17–31 August 2007 over the area 10°–60°N, 0°–360° for 24–168-h forecasts.

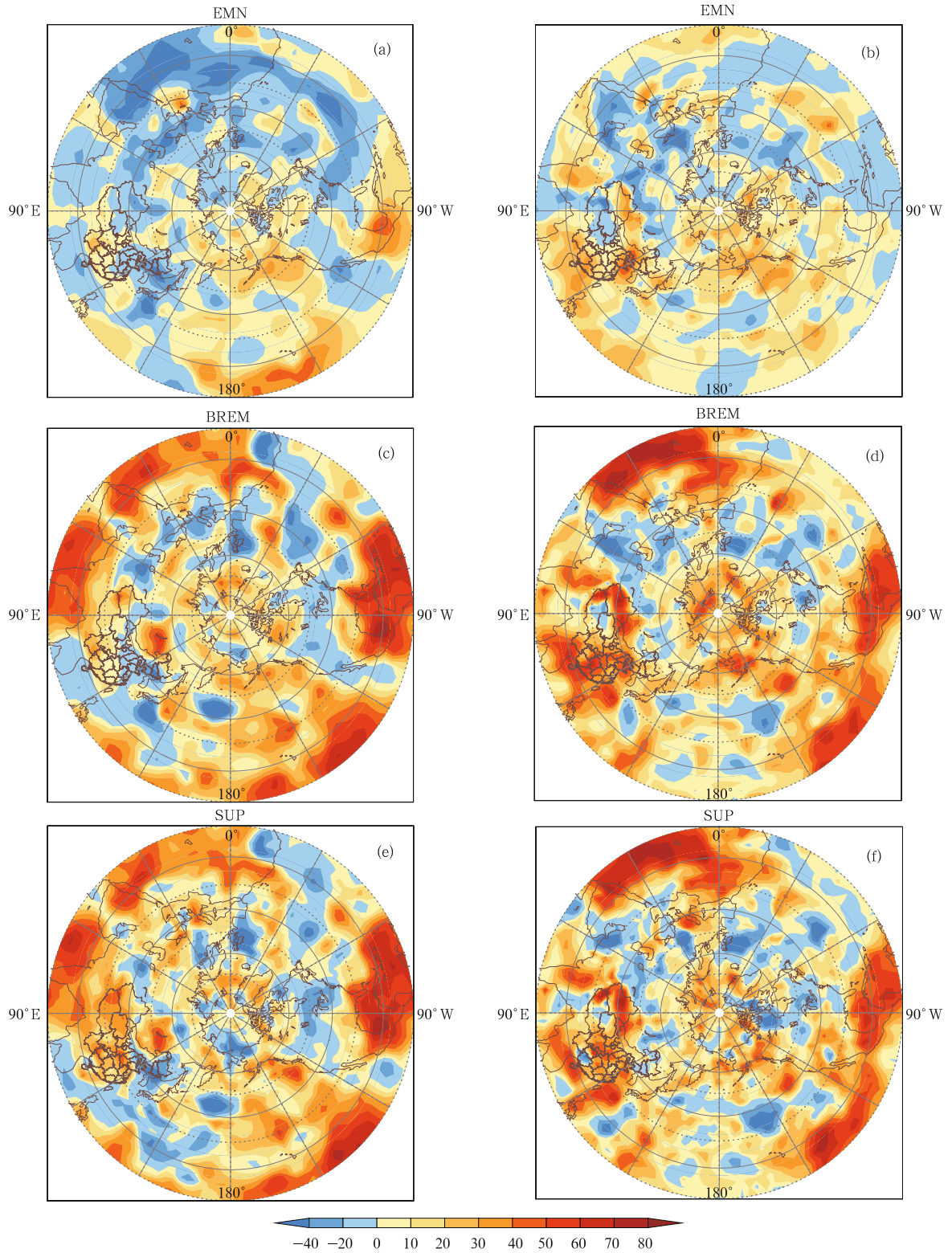


Fig. 6. Geographical distributions of the reduction percentage (%) of the mean RMSEs of 500-hPa geopotential height (left panels) and temperature (right panels) over the best model by the EMN (top panels), the optimal BREM (middle panels), and the optimal SUP (bottom panels) for 144-h forecasts from 17 to 31 August 2007 in the Northern Hemisphere (10° – 87.5° N, 0° – 360°).

Thus, the inferior model has a small impact on the SUP results.

3.5 Geographical distribution of the improvement by SUP and BREM over the best model

A detailed examination in Fig. 6 shows that the 144-h forecast skills of the 500-hPa geopotential height and temperature are improved significantly by using the SUP and BREM forecast techniques in most areas, especially in the tropics. In the extratropics, the SUP and BREM forecast skills have been improved by more than 20% in the areas of Ural Mountains, Lake Baikal, and the Sea of Okhotsk compared with that of the best single model forecast. It is well known that Eurasian blockings frequently occur over the Ural Mountains, Lake Baikal, and the Sea of Okhotsk (Zhi and Shi, 2006; Shi and Zhi, 2007), which has a significant impact on the persistent anomalous weathers in the upstream and downstream regions. The above analysis indicates that the multimodel ensemble forecasts may significantly improve the forecast skills of the variables at 500 hPa in mid-high-latitudes. Therefore, it is helpful for improving the forecast of some high-impact mid-high latitude weather systems by using the multimodel ensemble forecast techniques.

4. Conclusions

The superensemble forecasting takes full advantage of multimodel forecast products to improve the forecast skill. Through series of comparative analysis, the following conclusions are obtained.

(1) Comparative analysis of linear and nonlinear multimodel ensemble forecasts shows that for 24–120-h forecasts, the NNSUP forecast performs better than LRSUP and BREM forecasts. However, for 144–168-h forecasts, the forecast errors of BREM, LRSUP, and NNSUP are approximately equal.

(2) Both LRSUP (or SUP) and BREM forecasts of 500-hPa geopotential height have different optimal lengths of training period at each grid point. The optimal length of the training period for SUP is more than one and a half months in most areas, while it is less than one month for BREM.

(3) The SUP forecasts using the optimal length of the training period at each grid point have roughly a 16% improvement in the RMSEs of the 24–120-h forecasts of 500-hPa geopotential height, temperature, zonal wind, and meridional wind, while the improvement of the BREM is only 8%. But for 144–168-h forecasts, the forecast skill of the SUP is comparable to that of the BREM.

(4) For 24–168-h forecasts of the 500-hPa geopotential height, temperature, and winds, the BREM forecast without the UKMO model is more skillful than that with the UKMO model, while the SUP forecast error without the UKMO model is equivalent to that with the UKMO model. Hence, it is necessary to verify each model involved before conducting the BREM forecasting.

Acknowledgments. We are grateful to Drs. Zhang Ling and Chen Wen for their valuable suggestions.

REFERENCES

- Bougeault, P., and Coauthors, 2010: The THORPEX Interactive Grand Global Ensemble. *Bull. Amer. Meteor. Soc.*, **91**, 1059–1072, doi: 10.1175/2010BA MS2853.1.
- Buizza, R., D. Richardson, and T. N. Palmer, 2003: Benefits of increased resolution in the ECMWF ensemble system and comparison with poor-mans ensembles. *Quart. J. Roy. Meteor. Soc.*, **129**, 1269–1288.
- Cartwright, T. J., and T. N. Krishnamurti, 2007: Warm season mesoscale super-ensemble precipitation forecasts in the southeastern United States. *Wea. Forecasting*, **22**, 873–886.
- Geman, S., E. Bienenstock, and R. Doursat, 1992: Neural networks and the bias/variance dilemma. *Neural Computation*, **4**, 1–58.
- Krishnamurti, T. N., C. M. Kishtawal, T. LaRow, et al., 1999: Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, **285**, 1548–1550.
- , —, Z. Zhang, et al., 2000a: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, **13**, 4197–4216.
- , —, D. W. Shin, et al., 2000b: Improving tropical precipitation forecasts from multianalysis superensemble. *J. Climate*, **13**, 4217–4227.

- , K. Rajendran, and T. S. V. Vijaya Kumar, 2003: Improved skill for the anomaly correlation of geopotential height at 500 hPa. *Mon. Wea. Rev.*, **131**, 1082–1102.
- , C. Gnanaseelan, and A. Chakraborty, 2007a: Forecast of the diurnal change using a multimodel superensemble. Part I: Precipitation. *Mon. Wea. Rev.*, **135**, 3613–3632.
- , S. Basu, J. Sanjay, et al., 2007b: Evaluation of several different planetary boundary layer schemes within a single model, a unified model and a multimodel superensemble. *Tellus A*, **60**, 42–61.
- , A. D. Sagadevan, A. Chakraborty, et al., 2009a: Improving multimodel forecast of monsoon rain over China using the FSU superensemble. *Adv. Atmos. Sci.*, **26**(5), 819–839.
- , A. K. Mishra, and A. Chakraborty, 2009b: Improving global model precipitation forecasts over India using downscaling and the FSU superensemble. Part I: 1–5-day forecasts. *Mon. Wea. Rev.*, **137**, 2713–2734.
- Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418.
- Lorenz, E. N., 1969: A study of the predictability of 28-variable atmosphere model. *Tellus*, **21**, 739–759.
- Mishra, A. K., and T. N. Krishnamurti, 2007: Current status of multimodel super-ensemble and operational NWP forecast of the Indian summer monsoon. *J. Earth Syst. Sci.*, **116**, 369–384.
- Park, Y.-Y., R. Buizza, and M. Leutbecher, 2008: TIGGE: preliminary results on comparing and combining ensembles. *Quart. J. Roy. Meteor. Soc.*, **134**, 2029–2050.
- Rixen, M., and E. Ferreira-Coelho, 2006: Operational surface drift forecast using linear and nonlinear hyper-ensemble statistics on atmospheric and ocean models. *J. Mar. Syst.*, **65**, 105–121.
- , J. C. Le Gac, J. P. Hermand, et al., 2009: Superensemble forecasts and resulting acoustic sensitivities in shallow waters. *J. Mar. Syst.*, **78**, S290–S305.
- Shi Xiangjun and Zhi Xiefei, 2007: Statistical characteristics of blockings in Eurasia from 1950 to 2004. *Journal of Nanjing Institute of Meteorology*, **30**(3), 338–344. (in Chinese)
- Stefanova, L., and T. N. Krishnamurti, 2002: Interpretation of seasonal climate forecast using brier skill score, FSU superensemble, and the AMIP-I data set. *J. Climate*, **15**, 537–544.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.
- Warner, B., and M. Misra, 1996: Understanding neural networks as statistical tools. *J. Amer. Stat.*, **50**, 284–293.
- Zhi Xiefei and Shi Xiangjun, 2006: Interannual variation of blockings in Eurasia and its relation to the flood disaster in the Yangtze River valley during boreal summer. Proceedings of the 10th WMO International Symposium on Meteorological Education and Training, 21–26 September 2006, Nanjing, China.
- Zhi Xiefei, Lin Chunze, Bai Yongqing, et al., 2009a: Superensemble forecasts of the surface temperature in Northern Hemisphere middle latitudes. *Scientia Meteorologica Sinica*, **29**(5), 569–574. (in Chinese)
- , —, —, et al., 2009b: Multimodel superensemble forecasts of surface temperature using TIGGE datasets. Preprints of the Third THORPEX International Science Symposium, 14–18 September 2009, Monterey, USA.
- , Wu Qing, Bai Yongqing, et al., 2010: The multimodel superensemble prediction of the surface temperature using the IPCC AR4 scenario runs. *Scientia Meteorologica Sinica*, **30**(5), 708–714. (in Chinese)