ISSN 1068-3739, Russian Meteorology and Hydrology, 2011, Vol. 36, No. 1, pp. 11–17. © Allerton Press, Inc., 2011.
Original Russian Text © V.M. Khan, V.N. Kryzhov, R.M. Vil'fand, V.A. Tishchenko, A.Yu. Bundel', 2011, published in Meteorologiya i Gidrologiya, 2011, No. 1, pp. 19–29.

Multimodel Approach to Seasonal Prediction

V. M. Khan, V. N. Kryzhov, R. M. Vil'fand, V. A. Tishchenko, and A. Yu. Bundel'

Hydrometeorological Research Center of the Russian Federation, Bolshoi Predtechenskii per. 9–13, Moscow, 123242 Russia

Received January 27, 2010

Abstract—Multimodel forecast fields of temperature at 850 hPa and seasonal precipitation are combined using a procedure of two-step averaging. It is shown that the resulting forecasts averaged over the multimodel ensemble outperform the forecasts of individual models. The verification of forecast production has been carried out on cross-validated hindcasts according to WMO requirements. The simulation of spatiotemporal variability of atmospheric variables is assessed. The results indicate that the combined models are rather skillful in the tropical oceans, while the accuracy in the extratropics is poor.

DOI: 10.3103/S106837391101002X

INTRODUCTION

Atmospheric general circulation models (AGCMs) are being increasingly used as a tool for monthly and seasonal prediction. Although the predictability due to the memory of initial conditions (predictability of the first kind according to Lorenz) is limited to two weeks, forecasts of the average state of the atmosphere can be extended to longer periods (predictability of the second kind). The practical possibility of such forecasts was demonstrated by Shukla [13, 14]. Anomalies of slowly changing boundary conditions, in particular, the underlying surface state, are associated with anomalies of the average state of the atmosphere.

The implementation of ensemble approach was an important step to improving numerical prediction with AGCMs. Instead of one run for the forecast period, multiple (ensemble) runs are performed with the same boundary conditions and a set of perturbed initial conditions of the atmospheric state. The result is an ensemble of forecasts differing to some extent. It is assumed that each individual forecast contains a useful signal and an error due to "weather noise". It follows from the assumption about these errors being random and normally distributed that the useful signal is contained in the ensemble average, while the spread can be interpreted as a forecast uncertainty characteristic. The ensemble approach led to a considerable increase in forecast accuracy. Methods of ensemble generation and examples of using ensemble forecasting systems in leading forecast centers are reviewed in [1]. At present, almost all forecast centers, Russian and foreign, base their forecasts on the ensemble approach. Nevertheless, even these improved forecasts are of modest accuracy. The next step towards increasing forecast quality was the development of multimodel ensembles. A relatively new approach appeared in the last decade and found only limited application until present. This method is based on the assumptions similar to those of the single-model ensemble approach. The ensemble mean of each model in the multimodel ensemble contains a useful signal and an error due to inadequate description of natural processes in the model. If model errors are not interrelated and are normally distributed, the mean of the ensemble of several models will have more useful information than the mean of each model taken separately. Indeed, many studies [3, 7, 9, 11, 16] and experience of a number of forecast centers [8] showed that the multimodel forecast outperforms individual forecasts of all models in the ensemble.

In 2007, the North Eurasian Regional Climate Center (NERCC) was created within the Hydrometeorological Research Center of the Russian Federation. The NERCC carries out works to implement and improve methods of seasonal prediction based on the ensemble of Russian (the model of the Hydrometeorological Research Center of the Russian Federation in cooperation with the Institute of Numerical Mathematics of the Russian Academy of Sciences and the model of the Main Geophysical Observatory) and foreign models collected in the Asia-Pacific Economic Cooperation Climate Center (APCC). As everywhere in the world, the works in NERCC are carried out in two directions. First, much attention is paid to improving the quality of the global hydrodynamic model. The results of experiments aimed at the improvement of parametrizations show a progress. The second direction is better interpretation of model output using statistical

Table 1. Models participating in the experiment

Country	Institution (model)	Resolution	Ensemble size (forecast/hindcast)
Japan	Japan Meteorological Agency (JMA), http://www.jma.go.jp/jma/indexe.html	T63L40	31/5
China	National Climate Center/CMA (NCC), http://ncc.cma.gov.cn/	T63L16	8/8
Korea	Korea Meteorological Administration (GDAPS), http://web.kma.go.kr/eng/index.jsp	T106L21	20/20
	Seoul National University (GCPS), http://www.useoul.edu/	T63L21	12/12
	Meteorological Research Institute (METRI), http://www.nimr.go.kr/	4 × 5, L17	10/10
Russia	Main Geophysical Observatory (MGO), http://www.mgo.rssi.ru	T42L14	10/6
	Hydrometeorological Research Center of the Russian Federation (HMC), http://www.meteoinfo.ru	1.12 × 1.4, L28	10/10
USA	National Center for Environmental Prediction (NCEP), http://www.cpc.ncep.noaa.gov	T62L64	15/15

methods. In particular, different methods for a predictand (forecast variable) calculation can be developed considering an ensemble of model forecasts as a set of predictors. The simplest method is to represent the predictand as an ensemble mean after removing the model climate bias relative to the observed climate. More complex methods include the calibration according to the skill of models, that is, different weights are assigned to individual models in the ensemble corresponding to their skill [6, 10, 16]. However, the practice shows that changes in the forecast accuracy, both positive and negative, due to applying weights are very inhomogeneous in space and time [8]. It is likely the main reason why most forecast centers do not use calibration according to the skill of models in operational practice.

This paper considers the results of basic method for combining model forecasts, which is not related to the skill of individual models. It is a procedure of two-step averaging: intermodal averaging of means of individual models.

DATA

The monthly mean gridded fields of temperature at 850 hPa (T_{850}) from the NCEP/DOE reanalysis (National Center for Environmental Prediction/Department of Energy) [5] and precipitation fields from CMAP CPC (Merged Analysis of Precipitation, Climate Prediction Center) [15] with 2.5° × 2.5° resolution were used as reference observed data for a period from 1983 to 2004.

Table 1 lists the hindcasts of the models used for multiensemble forecasts and the links to the Internet sites of the institutions developing these models. The output forecast fields with initially different spatial resolutions and periods were interpolated to a unified 2.5° × 2.5° grid format for a period from 1983 to 2004. The focus was on the scores of T_{850} and precipitation in winter and summer.

Eight models were included in the summer multiensemble: GDAPS_F (SST forecast); GDAPS_O (persistent SST); METRI (South Korea); JMA (Japan); HMC, MGO (Russia); NCEP (the USA); and GCPS (South Korea). For the winter season, the same first seven models were taken and the NCC model (China) was used instead of the last one.

All the models are atmospheric general circulation models except for the NCEP model, which is a coupled atmosphere–ocean model. The model of Korea Meteorological Administration produces two kinds of forecasts: with persistence of SST anomalies observed at the beginning of model integration and with SST forecast from a statistical model. Ensembles of forecasts for a season are obtained from four-month model runs starting from the beginning of the last month of the previous season. The seasonal mean forecasts considered here are the forecast fields averaged from the 2nd to 4th months of integration.

METHODS USED IN THE STUDY

Anomalies of forecast variables were considered as deterministic forecast. The model and observed climatologies were calculated for 1983–2004. Then, anomalies were calculated relative to these normals. The combination was performed without any weights: the means were calculated for every individual model, the bias between the model and observed climate was corrected, and the means of all the models were averaged. The forecast not calibrated on the skill of models, that is, the ensemble average over the means of individual forecasts, is considered as a zero forecast skill level. To increase the forecast quality, we plan to apply various regression schemes based on the hindcasts corresponding to the SMIP2 protocol.

To estimate the deterministic forecasts, the following scores were used: the mean square skill score (*MSSS*) recommended by the WMO [2] showing the skill compared to the climate forecast, the absolute error (*ABSE*), the root-mean-square error (*RMSE*), and the correlation coefficient between forecast and observed anomalies (*ACC*):

$$ABSE = \frac{1}{N}\sum_{i=1}^{N} |(F_i - \overline{F}) - (X_i - \overline{X})|; \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N}\sum [(F_i - \overline{F}) - (X_i - \overline{X})]^2}; \quad (2)$$

$$ACC = \frac{\sum (F_i - \overline{F})(X_i - \overline{X})}{\sqrt{\sum (F_i - \overline{F})^2}\sqrt{\sum (X_i - \overline{X})^2}}, \quad (3)$$

where F stands for forecast; X, for observations; the overbar indicates averaging; and N is the sample size.

The *MSSS* at the jth point is

$$MSSS_j = 1 - \frac{MSE_j}{MSE_{cj}}, \quad (4)$$

where $MSE_j = \frac{1}{N}\sum_{i=1}^{n}(F_{ii} - X_{ij})^2$, $MSE_{cj} = \frac{1}{N}\sum_{i=1}^{n}(X_{ij} - \overline{X}_j)^2$.

The *MSSS* aggregated over a region is

$$MSSS = 1 - \frac{\sum_j \cos(\theta_j) MSE_j}{\sum_j \cos(\theta_j) MSE_{cj}},$$

where θ_j is the geographical latitude of the jth grid point.

The scores for both individual and combined forecasts were calculated through cross-validation.

RESULTS

Figure 1 shows the spatial distribution of the absolute errors of predicted T_{850} anomalies relative to the observed ones for both individual models and the combined forecast with equal weights. The winter maps are given. It is seen from the figure that the forecasts are relatively accurate in the tropics for all the models. The quality is lower in the extratropics, especially over the continents. All the models have the highest absolute errors in the high latitudes of North America (~3.5°C). As the international practice shows, most monthly and seasonal forecasts based on AGCMs forced by underlying surface anomalies influencing heat and mass exchange are rather skillful in the tropics, in the El Niño region in particular, where the inertial ocean influence is most pronounced. However, their quality is rather poor in the mid- and high latitudes with large synoptic variability [4, 9, 12, 14]. The forecast accuracy contrast in the tropics and extratropics is most distinct for the NCEP and MGO models (Fig. 1). It is seen that the combination procedure somewhat increases the forecast quality overall, although the skill in the extratropical continental areas remains low.

Table 2 shows the *MSSS*, *ABSE*, *RMSE*, and *ACC* aggregated for the globe and the Northern Hemisphere extratropics for summer and winter. The table shows how much the multimodel forecast skill exceeds the forecast skill of individual models. The positive *MSSS* indicating the skill compared to the climate forecast is found only for the MGO model for the whole globe. The multimodel forecast has higher *MSSS*, but it is close to zero indicating almost the same value of the combined forecast and the climate forecast. The combination of models decreased the absolute error compared to the errors of individual models by 0.03–

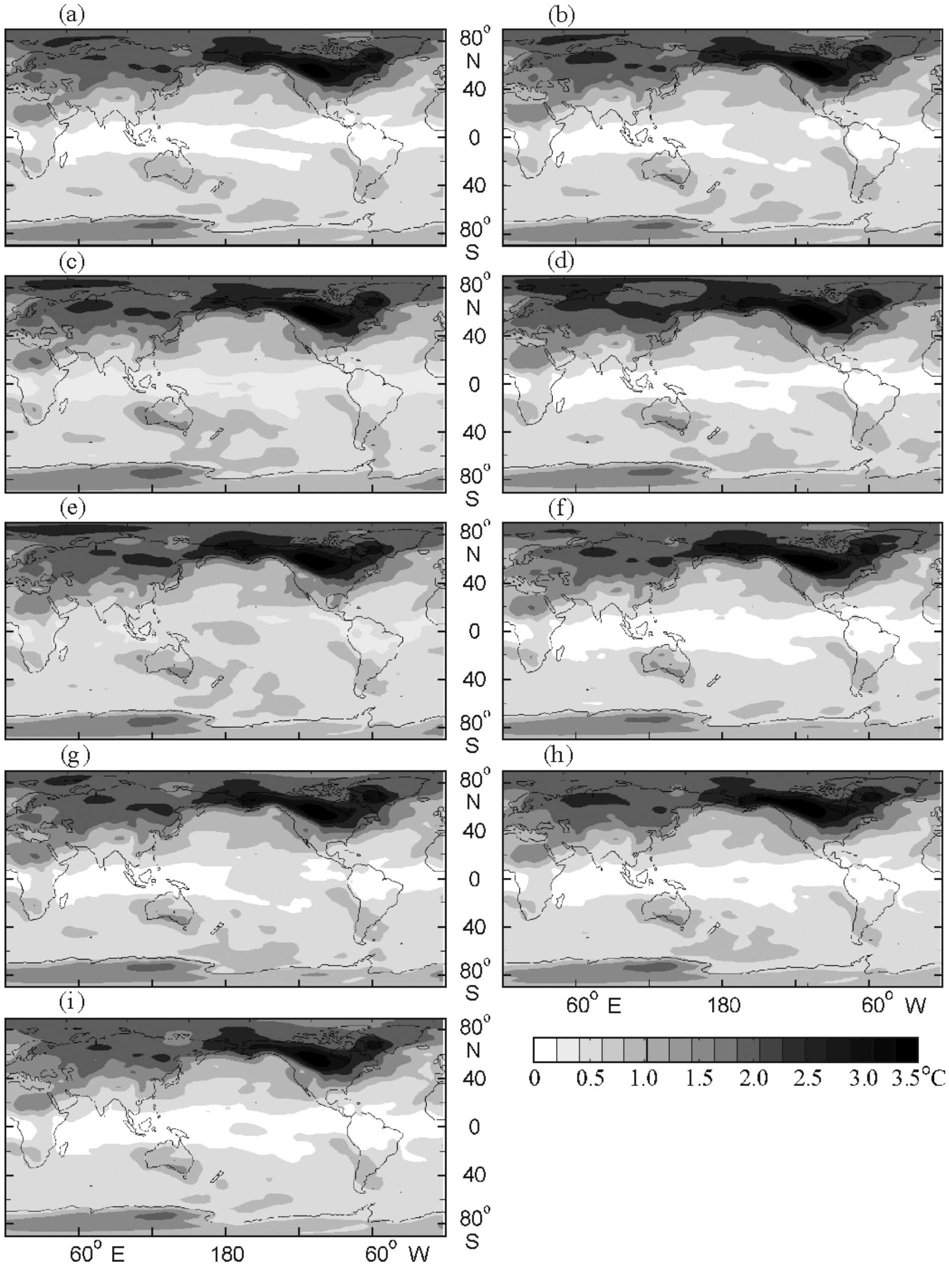

Fig. 1. Spatial distribution of winter absolute errors of the forecasts of temperature at 850 hPa anomalies compared to observed anomalies for (a-h) individual models and (i) multimodel forecast. (a) GDAPS_F; (b) GDAPS_O; (c) HMC; (d) JMA; (e) METRI; (f) MGO; (g) NCC; (h) NCEP.

0.18°C in winter and by 0.01–0.11°C in summer. The *RMSE* decreased by 0.05–0.21°C for winter and by 0.02–0.14°C for summer. Large *RMSE*s and *ABSE*s of seasonal forecasts are typical of the extratropics because of a worse forecast quality and higher temperature variability in the high latitudes. The *RMSE*s and *ABSE*s are naturally higher in winter than in summer because of higher temperature amplitudes in winter. Among the individual models, only the MGO (the globe and extratropics, in summer and in winter) and NCEP (the globe, in winter) models have correlations statistically significant at the 80% level by *F*-test (in bold in the table), the combined forecast having greatest correlations. The useful signal is slightly higher in winter than in summer for all ensemble members and for the combination of models.

Figure 2 shows the spatial distribution of the correlations between T_{850} forecast and observed anomalies. The regions with coefficients of correlation less than the 80% level of statistical insignificance are in white. There are high correlations in the tropics, while the extratropics in both hemispheres are mostly characte-

Table 2. Aggregated scores of T_{850} forecasts over the globe and in the Northern Hemisphere extratropics

Model	*MSSS*		*ABSE*, °C		*RMSE*, °C		*ACC*	
	globe	20°–90° N	globe	20°–90° N	globe	20°–90° N	globe	20°–90° N
				Winter				
GDAPS_F	–0.00	–0.02	1.23	1.84	1.53	2.28	0.15	0.11
GDAPS_O	–0.03	–0.03	1.25	1.86	1.55	2.29	0.03	0.05
HMC	–0.20	–0.13	1.29	1.90	1.64	2.37	0.14	0.10
JMA	–0.10	–0.17	1.30	1.97	1.61	2.44	0.15	0.08
METRI	–0.40	–0.14	1.31	1.91	1.64	2.37	0.12	0.13
MGO	0.04	–0.02	1.22	1.85	1.52	2.30	**0.24**	**0.16**
NCC	–0.08	–0.07	1.27	1.88	1.58	2.33	0.01	0.02
NCEP	–0.01	–0.08	1.24	1.89	1.55	2.34	**0.19**	0.09
MULTI	0.08	0.04	1.19	1.79	1.48	2.21	**0.25**	**0.20**
				Summer				
GCPS	–0.05	–0.10	1.08	1.16	1.34	1.45	0.15	0.05
GDAPS_F	–0.02	–0.02	1.06	1.12	1.33	1.39	0.14	0.13
GDAPS_O	–0.04	–0.04	1.08	1.13	1.34	1.41	0.01	0.04
HMC	–0.19	–0.21	1.10	1.17	1.37	1.45	0.15	0.10
JMA	–0.14	–0.19	1.12	1.20	1.40	1.49	0.14	0.09
METRI	–0.39	–0.25	1.14	1.19	1.42	1.49	0.08	0.07
MGO	0.03	–0.03	1.04	1.12	1.30	1.40	**0.26**	**0.17**
NCEP	–0.05	–0.08	1.09	1.15	1.36	1.43	0.14	0.07
MULTI	0.07	0.03	1.03	1.10	1.28	1.36	**0.23**	**0.18**

Note: Explanations are given in the text.

rized by values below the threshold of statistical significance. In Eurasia, the anomalies are most difficult to predict in European Russia and in vast regions of North Asia. Useful forecast information can be obtained for some regions in Southern Siberia and the Far East, the west of Europe, and several regions in Central Asia. The scores are slightly better in winter than in summer.

Almost the same features of geographical distributions are found for the *MSSS*. The large *MSSS*s (the perfect *MSSS* value is unity) are located in the tropics only. As for the *ACC*s, the *MSSS*s are higher in winter.

Precipitation is one of the meteorological variables most difficult to predict. The precipitation scores were found lower than for temperature. Figure 3 shows the maps of correlations between forecast and observed seasonal precipitation anomalies. The areas with correlations below the 80% level of statistical significance are in white. Significant correlations are located mainly in the tropical Pacific. Over the continents, the precipitation forecast is almost useless, except for small areas. In winter, precipitation is slightly more predictable than in summer.

Table 3 reports aggregated scores of seasonal precipitation forecasts. Overall, the globally averaged scores demonstrate that the accuracy is not high. The *MSSS* is negative even for the resulting forecast indicating that the climate forecast is better. The highest correlations do not exceed the 56% level of statistical significance by *F*-test. As with temperature, the scores for the Northern Hemisphere extratropics are worse than the globally averaged scores.

CONCLUSION

For the first time in Russia, the results of multimodel seasonal forecasts are given. Two Russian models, of the Hydrometeorological Research Center of the Russian Federation and of the Main Geophysical Observatory, and six foreign models used in the Asia-Pacific Economic Cooperation Climate Center entered the multimodel ensemble. A comprehensive assessment of the accuracy of deterministic forecasts is made for the individual models and for the multimodel ensemble average based on the hindcasts for 1983–2004. Overall, the scores of forecasts combined using equal weights are better than the scores of individual fore-

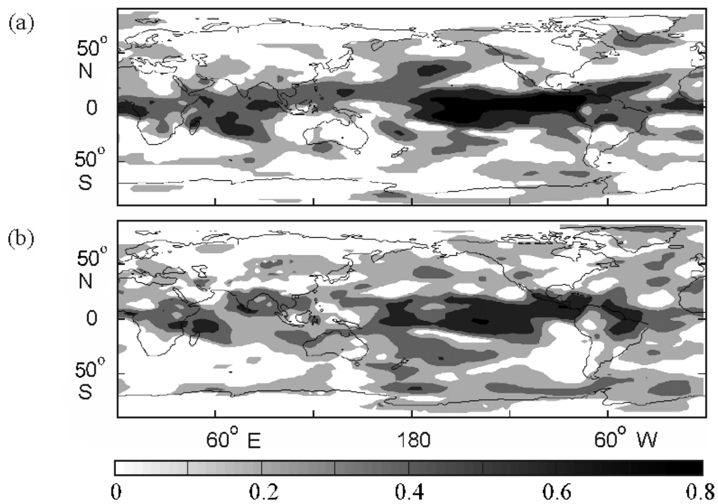

Fig. 2. Spatial distribution of correlation coefficients between forecast and observed anomalies of air temperature at 850 hPa for (a) winter and (b) summer.

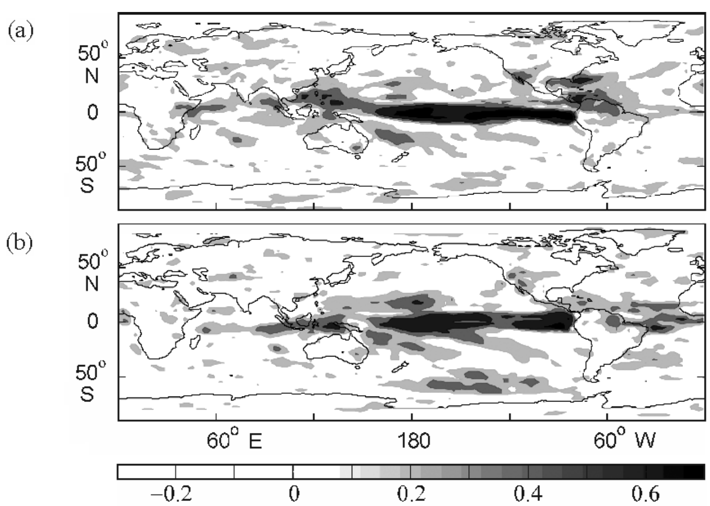

Fig 3. Same as in Fig. 2 but for precipitation anomalies.

casts. However, the forecast accuracy in the extratropics is far from being satisfactory even within this approach. The forecast quality is better in the tropics than in the mid- and high latitudes. The accuracy over the oceans is higher than over the continents due to larger inertia of the oceans. As it was anticipated, the air temperature is better predicted than precipitation.

The development of combination methods lies in the implementation of regional combination based on the forecast calibration on the past skill of the models.

ACKNOWLEDGMENTS

The work is supported by the Russian Foundation for Basic Research (grants 09-05-13570-ofi_ts and 10-05-00784a).

REFERENCES

1. A. V. Muraviev and I. A. Kulikova, “Ensembles of Forecasts: Methods, Problems, and Perspectives,” Meteorol. Gidrol., No. 3, 2005 [Russ. Meteorol. Hydrol., No. 3, 2005].

Table 3. Aggregated scores of precipitation over the globe and in the Northern Hemisphere extratropics

Model	*MSSS*		*ABSE*, mm/day		*RMSE*, mm/day		*ACC*	
	globe	20°–90° N	globe	20°–90° N	globe	20°–90° N	globe	20°–90° N
Winter								
GDAPS_F	–0.27	–0.15	0.62	0.37	0.81	0.47	0.07	0.07
GDAPS_O	–0.20	–0.13	0.61	0.37	0.80	0.47	0.03	0.03
HMC	–1.33	–0.57	0.76	0.42	0.99	0.54	0.04	0.04
JMA	–0.88	–0.72	0.70	0.45	0.90	0.57	0.07	0.03
METRI	–0.71	–0.42	0.63	0.38	0.81	0.49	0.05	0.01
MGO	–0.64	–0.59	0.66	0.41	0.86	0.53	0.04	0.01
NCC	–0.58	–0.42	0.64	0.39	0.84	0.50	0.01	0.01
NCEP	–0.69	–0.56	0.67	0.42	0.87	0.53	0.08	0.04
MULTI	–0.08	–0.04	0.57	0.35	0.74	0.45	0.10	0.06
Summer								
GCPS	–0.45	–0.59	0.61	0.52	0.79	0.66	0.06	0.01
GDAPS_F	–0.46	–0.37	0.61	0.50	0.79	0.63	0.02	0.02
GDAPS_O	–0.25	–0.31	0.57	0.49	0.74	0.62	0.00	0.01
HMC	–1.90	–1.36	0.76	0.60	0.99	0.77	0.04	0.03
JMA	–1.22	–0.85	0.69	0.56	0.90	0.72	0.05	0.02
METRI	–0.89	–0.73	0.60	0.51	0.78	0.64	0.04	0.02
MGO	–0.87	–0.87	0.65	0.55	0.84	0.71	0.06	0.01
NCEP	–0.81	–0.57	0.64	0.52	0.83	0.66	0.09	0.03
MULTI	–0.18	–0.20	0.54	0.47	0.70	0.59	0.09	0.04

2. *Manual on the Global Data-processing and Forecasting System*, Vol. I (Annex IV to WMO Technical Regulations). Global Aspects, 1992, Attachment II-8. Standardized Verification System (SVS) for Long-range Forecasts (LRF), WMO, Geneva, 2005.
3. F. Doblas-Reyes, M. Déqué, and J.-P. Piedelievre, “Multi-model Spread and Probabilistic Seasonal Forecast in PROVOST,” Quart. J. Roy. Meteorol. Soc., **126** (2000).
4. R. J. Graham, A. D. L. Evans, K. R. Mylne, et al., “An Assessment of Seasonal Predictability Using Atmospheric General Circulation Models,” Quart. J. Roy. Meteorol. Soc., **126** (2000).
5. M. Kanamitsu, W. Ebisuzaki, J. Woollen, et al., “NCEP-DOE AMIP-II Reanalysis,” Bull. Amer. Meteorol. Soc., 83 (2002).
6. V. V. Kharin and F. W. Zwiers, “Improved Seasonal Probability Forecast,” J. Climate, **16** (2003).
7. T. N. Krishnamurti, C. M. Kishtawal, T. E. LaRow, et al., “Improved Weather and Seasonal Climate Forecasts from Multimodel Superensemble,” Science, **285 (5433)** (1999).
8. V. N. Kryjov, H.-W. Kang, D. Nohara, et al., “Assessment of the Climate Forecasts Produced by Individual Models and MME Methods,” APCC Technical Report, No. 1, **1** (2006).
9. T. N. Palmer, “A Nonlinear Dynamical Perspective on Climate Prediction,” J. Climate, **12** (1999).
10. V. Pavan and F. J. Doblas-Reyes, “Multi-model Seasonal Hindcasts over the Euro-Atlantic: Skill Scores and Dynamic Features,” Climate Dynamics, **16** (2000).
11. P. Peng, A. Kumar, and H. van den Dool, “Analysis of Multimodel Ensemble Predictions for Seasonal Climate Anomalies,” J. Geophys. Res., **107** (2002).
12. T. Reichler and J. O. Roads, “Time-space Distribution of Long-range Atmospheric Predictability,” J. Atmos. Sci., **61** (2004).
13. J. Shukla, “Dynamical Predictability of Monthly Means,” J. Atmos. Sci., **38** (1981).
14. J. Shukla, J. Anderson, D. Baumhefner, et al., “Dynamical Seasonal Prediction,” Bull. Amer. Meteorol. Soc., **81** (2000).
15. P. Xie and P. A. Arkin, “Global Precipitation: A 17-Year Monthly Analysis Based on Gauge Observation, Satellite Estimates, and Numerical Model Outputs,” Bull. Amer. Meteorol. Soc., **78** (1997).
16. W. T. Yun, L. Stefanova, and T. N. Krishnamurti, “Improvement of the Multimodel Superensemble Technique for Seasonal,” J. Climate, **16** (2003).