JMA/NCEP/CMC Multi-Center Ensemble Forecast

Mio Matsueda[1], Masayuki Kyouda[2], and H. L. Tanaka[3]

1: Graduate School of Life and Environmental Sciences, University of Tsukuba, Japan

2: Numerical Prediction Division, Japan Meteorological Agency

3: Center for Computational Sciences, University of Tsukuba, Japan
and Frontier Research Center for Global Change

1 Introduction

Ensemble forecast is a collection of a number of forecasts that verify at the same time. In the ensemble forecasts, several model forecasts are performed by introducing perturbations in the initial conditions or in the models themselves. Ensemble forecast has accomplished two main goals: the first one is to provide an ensemble average forecast that beyond the first few days is more accurate than individual forecasts, because the components of the forecast that are most uncertain tend to be averaged out. The second and more important goal is to provide forecasters with an estimation of the reliability of the forecast, which because of changes in atmospheric predictability, varies from day to day and from region to region (Kalnay 2003).

Operational ensemble forecasts for medium-range forecasting have been performed at several numerical weather prediction (NWP) centers. For example, the European Center for Medium-Range Weather Forecasts (ECMWF) has performed an operational Ensemble Prediction System (EPS) daily since May 1994 (Palmer et al. 1993; Molteni et al. 1996), and has the most highest ensemble model resolution with 102 ensemble members per day whose perturbations were generated by singular vector method.

The National Centers for Environmental Prediction (NCEP) produces 11 ensemble members at 0000, 0600, 1200 and 1800 UTC everyday out to 16-days lead time. NCEP has used a bred-vector (BV) perturbation method introduced by Toth and Kalnay (1993). This method is based on the argument that fast-growing perturbations develop naturally in a data assimilation cycle and will continue to grow as short- and medium-range forecast error. Further information for NCEP EPS is described in Toth and Kalnay (1993) and Toth and Kalnay (1997).

The Japan Meteorological Agency (JMA) EPS has been carried out with 25 members at 1200 UTC everyday out to 9-days for medium-range forecasting since March 2001. Perturbed initial fields are obtained using the Breeding of Growing Modes, same as NCEP. For further description of JMA EPS, see JMA (2002).

Also, Canadian Meteorological Center (CMC) has performed an operational EPS at daily 0000 UTC since February 1998, using a multi-model ensemble method. CMC EPS consists of 17 different model runs, mainly with different physical parameterizations. Nine members, containing a control run, are driven by the SEF (Spectral Finite Element) models and other eight members are driven by the Global Environment Multi-scale (GEM) models. Also, CMC EPS adopts perturbed analyses as initial perturbation fields. For further information for CMC EPS, see Houtekamer et al. (1996), Lefaivre et al. (1997) and Pellerin et al. (2003).

Recently, a Multi-Center Ensemble (MCE) has attracted international attention (e.g. Mylne et al. 2003; Richardson 2001; Ziehmann 2000). Creating MCE provides not only increasing ensemble members without further computing resources but also reducing the overall model bias, and thereby the forecast skill of MCE is expected to be superior to that of individual ensemble forecast.

In this study, we have created the MCE forecast, consisting of ensemble forecasts by JMA, NCEP and CMC. We investigated whether the forecast skill of the MCE forecast is improved than that of JMA ensemble forecast. Two variables, 500 hPa geopotential height (hereafter referred to as Z500) and temperature at 850 hPa (hereafter referred to as T850) in the Northern Hemisphere (20°N–90°N) during September 2005, are assessed using Anomaly Correlation (AC), Root Mean Square Error (RMSE), and Brier Skill Score (BSS).

2 Data and Methods

2.1 Data

In this study, three ensemble forecast data, JMA, NCEP, and CMC, are used. The details are summarized in Tabale 1.

Table 1 Three ensemble configurations at JMA, NCEP and CMC.

	JMA	NCEP	CMC
Model Resol.	T106L40	T126L28	TL149L23-41
Grid	$2.5^\circ \times 2.5^\circ$	$2.5^\circ \times 2.5^\circ$	$1.0^\circ \times 1.0^\circ$
Fore. Leng.	216hr (12hrly)	384hr (6&12hrly)	240hr (12hrly)
Init. Perturb.	BVs	BVs	Anal. cycle
Init. UTC	12	00, 06, 12, 18	00
Mem./run	25	11	17
Mem./day	25	44	17

The data from different ensemble producer is created at different resolutions. CMC ensemble grid is $1.0^\circ \times 1.0^\circ$, whereas JMA and NCEP grids are $2.5^\circ \times 2.5^\circ$. So data is interpolated into 2.5 degree grid spacing as a common grid before the verification of MCEs. MCE forecasts that we have created using above three ensembles are shown in Table 2. As can be expected from Table 2, the alphabet indicates center name, namely J, N and C indicate JMA, NCEP and CMC, respectively, and the number behind each alphabet indicates each ensemble member. So, for example, J25N11C17 consists of 25 ensemble members of JMA, 11 members of NCEP and 17 members of CMC and the total number of the members is 53. J9N8C8, created to compare with JMA25, contains JMA ensemble control run, four perturbation pairs of JMA, 4 perturbation pairs of NCEP and 4 perturbation pairs of CMC. Initial UTC of MCE forecasts is set to 1200 UTC. So, the effect of Lagged Averaged Forecasts (LAF) method is naturally contained in MCE forecasts and J25N44C17 also contains that. In this study, however, the effect of LAF method is not considered.

Table 2 MCE configurations. Left column is abbreviated MCE name.

MCE	JMA mem. (UTC)	NCEP mem. (UTC)	CMC mem. (UTC)
JMA25	25 (12)	-	-
NCEP11	-	11 (12)	-
CMC17	-	-	17 (00)
J9N8C8	9 (12)	8 (12)	8 (00)
J25N11C17	25 (12)	11 (12)	17 (00)
J25N44C17	25 (12)	44 (00, 06, 12, 18)	17(00)

2.2 Methods

2.2.1 Deterministic Verification

We investigate the skill of ensemble mean forecasts of MCE using anomaly correlation (AC) and root mean square error (RMSE). Anomaly correlation (AC) is defined by the following equation:

$$AC = \frac{\sum(x_f - x_c)(x_a - x_c)}{\sqrt{\sum(x_f - x_c)^2}\sqrt{\sum(x_a - x_c)^2}}, \tag{1}$$

where x_f, x_a and x_c indicate ensemble mean forecast, analysis and climate, respectively. In this study, the summation is taken in the Northern Hemisphere (20°N–90°N). AC indicates a patterns correlation between forecast anomaly and analysis anomaly, so AC decreases with time. AC score of one (1.0) demonstrates perfect skill. Based on experience with the anomaly correlation, a score near 0.6 suggests forecast errors are sufficiently large enough to indicate minimal skill while a score below 0.6 signifies a forecast is not useful. In general, the time when AC first becomes 0.6 is called the limitation of predictability (hereafter referred to as LP). We used JMA climate data for calculating ACs of all MCE although AC is sensitive to the choice of the climatological reference. It must be noted that our AC for NCEP11 and CMC17 is not exact value but approximate value. Also, although calculating AC requires the analysis data, we regarded the JMA control run at initial time as the analysis data except for NCEP11 and CMC17.

Root mean square error (RMSE) is defined by the following equation:

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(x_f - x_a)^2}\ , \tag{2}$$

where x_f and x_a indicate the ensemble mean forecast and the analysis, respectively, and the summation is taken in the Northern Hemisphere, as in AC. RMSE indicates a forecast error and RMSE score of zero (0.0) demonstrates perfect skill. RMSE doesn't require the climate data differently from AC, so RMSEs for NCEP11 and CMC17 are accurate.

2.2.2 Probabilistic Verification

The most commonly used verification diagnostic for probabilistic forecasts is the Brier Score, originally introduced by Brier (1950) and described in its modified standard form by Wilks (1995) as:

$$BS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2}\ . \tag{3}$$

The Brier Score is essentially the mean square error for probability forecasts of an event, where p_i and o_i are forecast and observed probabilities, respectively; o_i takes values of unity when the event occurs and zero when it does not occur. N is the number of grid points in the spatio–temporal domain, namely which indicates all grid points in September 2005 in the Northern Hemisphere in this study. BS becomes 0 only when ensemble mean forecast is perfect. However, for example, when BS is equal to 0.6, it's difficult to evaluate whether the ensemble forecast has a good skill or not. Therefore, Brier Skill Score (BSS), defined by the following equation, is often considered:

$$BSS = \frac{BS_{clm} - BS}{BS_{clm}} = 1 - \frac{BS}{BS_{clm}}\ ,$$

where BS_{clm} is a Brier Score for climate forecast. BSS becomes 1.0 only when ensemble mean forecast is perfect. Also BSS becomes 0 when ensemble forecast is equivalent to climate forecast, indicating ensemble mean forecast has no skill. Calculating BSS requires a threshold which is defined for an event. So, we configured 4 thresholds for Z500, whether anomaly is greter (less) than 1 or 2 (−1 or −2) climatoligical standard deviation (hereafter referred to as SD) , and 6 thresholds for T850, whether anomaly is greter (less) than 2, 4 or 8 (−2, −4 or −8)K.

3 Results

3.1 Comparison of the JMA, NCEP and CMC ensemble forecasts

First, we make a comparison between the skills of three ensemble mean forecasts, JMA25, NCEP11 and CMC17, using AC and RMSE. When forecast skills among them differ extremely, the effect of MCE might depend largely on the most skillful forecast. AC was calculated for Z500 in the Northern Hemisphere using JMA climate and each analysis data. The monthly mean LP, when AC

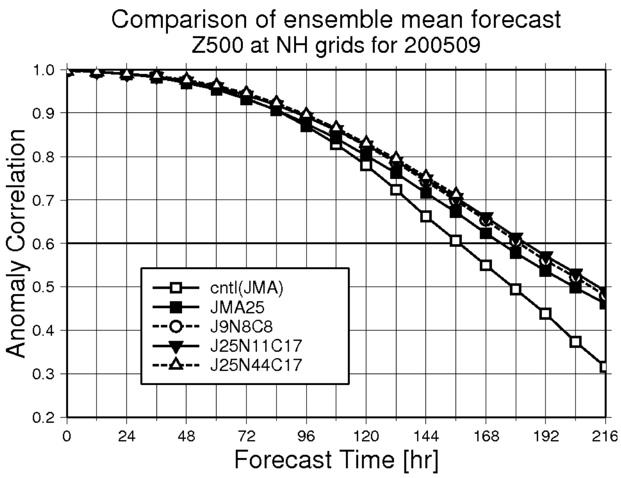

Figure 1. Anomaly correlation skill for JMA control run and ensemble mean forecasts, JMA25, J9N8C8, J25N11C17 and J25N44C17, Z500 in the Northern Hemisphere (20°N–90°N).

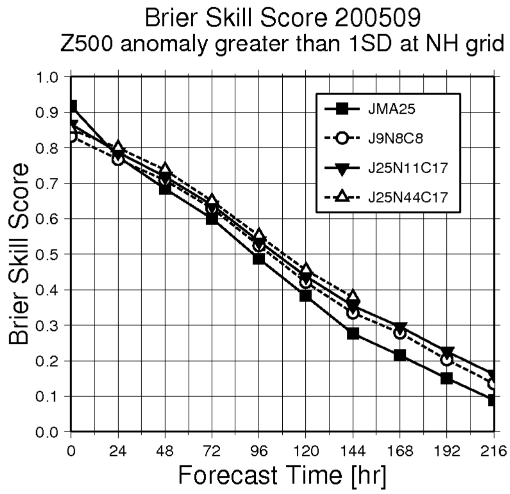

Figure 2. Brier Skill Score of probabilistic predictions, JMA25, J9N8C8, J25N11C17 and J25N44C17, for Z500 in the Northern Hemisphere (20°N–90°N). The threshold is whether the anomaly is greater than 1 standard deviation.

becomes 0.6 first, of JMA25, NCEP11 and CMC17 are 174hr, 180hr, 168hr, respectively. NCEP11 is somewhat superior to JMA25 and CMC17 although LP of NCEP11 and CMC17 may not be accurate because of using the JMA climate. Also, RMSE was calculated for Z500 and T850 in the Northern Hemisphere. RMSE does not require climate data, so RMSE may be more accurate than AC. RMSEs for 500 hPa at 168 hr of JMA25, NCEP11, CMC17 are about 56 m although that of CMC17 is slightly larger than it. Also, RMSEs for 850 hPa at 168 hr of these are all about 2.7 K although that of CMC17 is slightly larger than it. Consequently, we might consider that the skills of three ensemble forecasts are almost equivalent.

3.2 Effects of MCE forecasts and increasing ensemble members

Figure 1 illustrates ACs of JMA control run, JMA25, J9N8C8, J25N11C17, and J25N44C17 for Z500. LP of J9N8C8 is 180 hr, which indicates MCE is more skillful than JMA25 in the latter half of the forecast. J25N11C17 seems to be slightly skillful than J9N8C8. It is interesting that in spite of LPs of JMA25, NCEP11 and CMC17 are 174 hr, 180 hr and 168 hr, respectively, that of J25N11C17 exceeds 180 hr that is maximum within three single model EPS. Also it is hoped that J25N44C17 is slightly skillful than J25N11C17 in the latter half of the forecast although J25N44C17 cannot be created over 156 hr lead time owing to the forecast intervals. Same results are obtained with respect to RMSE for Z500 and T850 (not shown). RMSEs of MCE forecasts are smaller than that of JMA25, and that of JMA25N44C17 is the smallest. The difference among MCEs, J9N8C8, J25N11C17 and J25N44C17, however, is small.

Figure 2 illustrates BSS of JMA25, J9N8C8, J25N11C17, and J25N44C17 for the threshold whether Z500 anomaly is greater than 1 SD. BSS of MCEs after +24h forecast time is superior to that of JMA25, whereas BSSs of MCEs are inferior to that of JMA25 up to +12h forecast time. Also J25N11C17 is more skillful than J9N8C8, and J25N44C17 is more skillful than J25N11C17 in the latter half of the forecast. The same result is obtained with respect to the threshold whether anomaly is greater than 2 SD (not shown). However, for -1 or -2 SD threshold there is not apparent difference among MCE forecasts although all MCE forecasts are superior to JMA25.

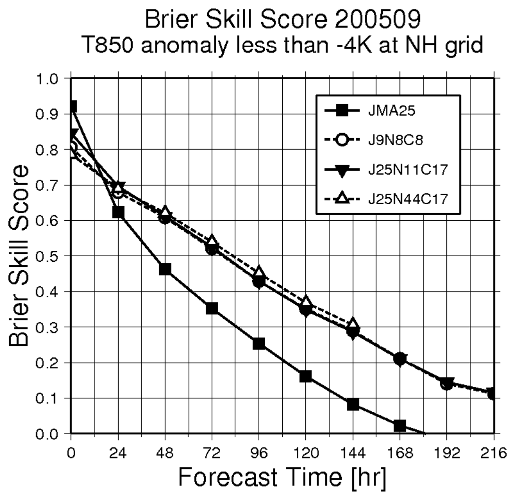

Figure 3. Same as Fig. 2 but for T850. The threshold is whether anomaly is less than -4 K.

Figure 3 illustrates BSS of JMA25, J9N8C8, J25N11C17, and J25N44C17 for the threshold whether T850 anomaly is less than -4 K. BSS of J25N11C17 is almost equal to that of J9N8C8, and J25N44C17 is slightly superior to these, although MCE forecasts are superior to JMA25. The same result is obtained with respect to the threshold whether anomaly is less than -2 or -8 K (not shown) except for that J25N11C17 is superior to J9N8C8 for -8 K threshold. However, BSSs of J9N8C8, J25N11C17 and J25N44C17 are not always superior to that of JMA25 for 2, 4 and 8 K threshold even though in the latter half of the forecast (not shown).

4 Conclusions and discussion

In this study, we have created the multi-center ensemble(MCE) forecast, consisting of ensemble forecasts by JMA, NCEP and CMC. We investigated whether the forecast skill is improved than that of JMA ensemble forecast. Two variables, Z500 and temperature at T850 in the Northern Hemisphere during September 2005, are assessed using AC, RMSE, and BSS.

First, we made a comparison among the skills of three ensemble mean forecasts, JMA25, NCEP11 and CMC17, using AC and RMSE. The monthly mean LPs, when AC becomes 0.6 first, of JMA25, NCEP11 and CMC17 are 174 hr, 180 hr, 168 hr, respectively. NCEP11 is somewhat superior to JMA25 and CMC17. Also, RMSEs for 500 hPa at 168 hr of JMA25, NCEP11, CMC17 are about 56 m although that of CMC17 is slightly large than it. RMSEs for 850 hPa at 168 hr of these are all about 2.7 K although that of CMC17 is slightly large than it. Consequently, we might consider that the skills of three ensemble forecasts are almost equivalent.

Next, we compared JMA25 and J9N8C8 to investigate the effect of MCE. It is noted that both JMA25 and J9N8C8 consist of 25 ensemble members. LP of J9N8C8 is 180 hr, which is longer than that of JMA25. RMSE of J9N8C8 is smaller than that of JMA25. Also, BSS of J9N8C8 for Z500 threshold, whether anomaly is greater (less) than 1 or 2 SD (-1 or -2SD), and T850 threshold, whether anomaly is less than -2, -4 and -8 K, indicates that J9N8C8 is superior to JMA25. BSS of J9N8C8 for T850 threshold, whether anomaly is greater than 2, 4 and 8 K, is equivalent or inferior to that of JMA25. Looking overall, however, MCE forecast is more skillful than single model emsemble forecast.

Furthermore, we investigated the effect of increasing ensemble members in MCE. In the sight of AC, J25N11C17 seems to be slightly skillful than J9N8C8 and JMA25. It is interesting that in spite of LPs of JMA25, NCEP11 and CMC17 are 174 hr, 180 hr and 168 hr, respectively, that of J25N11C17 exceeds 180 hr that is maximum within three single model EPS. Also it is hoped that J25N44C17 is slightly more skillful than J25N11C17 in the latter half of the forecast although J25N44C17 cannot be created over 156 hr lead time owing to the forecast interval. The same results are obtained with respect to RMSE for Z500 and T850. RMSE of MCE forecasts is smaller than that of JMA25, and that of JMA25N44C17 is smallest. The difference among MCEs, J9N8C8, J25N11C17 and J25N44C17, however, is small. In probabilistic verification for two thresholds, whether Z500 anomaly is greater than 1 SD and 2 SD, it is found that J25N11C17 is more skillful than J9N8C8, and J25N44C17 is more skillful than J25N11C17 in the latter half of the forecast. However, for -1 or -2 SD threshold there is no apparent difference among MCE forecasts although all MCE forecasts are superior to JMA25. With respect to threshold, whether T850 anomaly is less than -2, -4 and -8 K, J25N44C17 is most skillful and MCEs are skillful than JMA25. However, BSS of J9N8C8, J25N11CMC17 and J25N44C17 are not always superior to that of JMA25 for 2, 4 and 8 K threshold even though in the latter half of the forecast.

Based on above results, it might be difficult to evaluate the effect of increasing ensemble members in the sight of MCE configuration used in this study. However, it seems to be rare that MCE forecast is always skillful within all verification items. So, it seems that there exist the effect of increasing ensemble members.

References

Brier, G.W., 1950: Verification of Forecasts Expressed in Terms of Probability. *Mon. Wea. Rev.*, **78**, 1–3.

Houtekamer, P.L., L. Lefaivre, J. Derome, H. Ritchie, and H.L. Mitchell, 1996: A System Simulation Approach to Ensemble Prediction. *Mon. Wea. Rev.*, **124**, 1225–1242.

JMA, 2002: OUTLINE OF THE OPERATIONAL NUMERICAL WEATHER PREDICTION AT THE JAPAN METEOROLOGICAL AGENCY. Japan Meteorological Agency. 158pp.

Kalnay, E. 2003: Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press, Cambridge, 341pp.

Lefaivre, L., P.L. Houtekamer, A. Bergeron and R. Verret, 1997: The CMC Ensemble Prediction System. Proc. *ECMWF 6^{th} Workshop on Meteorological Operational Systems*, Reading, U.K., ECMWF, 31–44.

Molteni, F., R. Buizza, T.N. Palmer, and T. Petroliagis., 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.

Mylne, K.R., R.E. Evans, and R.T. Clark, 2002: Multi-model multi-analysis ensembles in quasi-operational medium-range forecasting. *Quart. J. Roy. Meteor. Soc.*, **128**, 361–384.

Palmer, T.N., F. Molteni, R. Mureau and R. Buizza 1993: 'Ensemble prediction'. ECMWF seminar proceedings 'Validation od models over Europe; Vol. 1. ECMWF, Shinfield Park, Reading, UK.

Pellerin, G., L. Lefaivre., P. Houtekamer, and C. Girard, 2003: Increasing the horizontal resolution of ensemble forecasts at CMC.*Non. Proc. Geophys.*, **10**, 463–468.

Richardson, D.S., 2001: Ensembles using multiple models and analyses. *Quart. J. Roy. Meteor. Soc.*, **127**, 1847–1864.

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.

Toth, Z. and E. Kalnay, 1997: Ensemble forecasting at NCEP and the Breeding Method. *Mon. Wea. Rev.*, **125**, 3297–3319.

Wilks, D.S. 1995: Statistical methods in the atmospheric sciences–an introduction. International Geophysics Series Vol 59, Academic Press, New York, USA.

Ziehmann, C., 2001: Comparison of a single–model EPS with a multi–model ensemble consisting of a few operational model. *Tellus*, **52A**, 280–299.