Daily Forecast Skill of Multi-Center Grand Ensemble

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

1: Graduate School of Life and Environmental Sciences, University of Tsukuba, Japan
2: Numerical Prediction Division, Japan Meteorological Agency, Japan
3: Center for Computational Sciences, University of Tsukuba, Japan

1 Introduction

An ensemble forecast techniques have drawn more attention in various timescales, such as short-, medium-, and long-ranges for both operational and research purposes. The various ensemble technique except for initial-value ensemble technique, such as multi-model ensemble technique, have been introduced into the ensemble forecast since the ensemble forecast started in numerical weather prediction (NWP) centers. The multi-model ensemble technique combines the independent forecast data from different model in order to represent not only the uncertainties of initial value but also the imperfection of model formulation (e.g., Krishnamurti et al. 1999; Ziehmann 2001; Buizza et al. 2003; Richardson 2001; Mylne et al. 2003). Recently, rapid progress of communication networks enable us to get vast operational ensemble forecast data from some NWP centers. Matsueda et al. (2006) constructed the Multi-Center Grand Ensemble (MCGE) forecast, consisting of three operational ensemble forecast data by JMA, NCEP, and CMC, on a quasi-operational level. They have revealed that MCGE forecasts are more skillful than single-center ensemble forecast without weights among ensemble members and bias corrections using monthly deterministic and probabilistic scores, such as Anomaly Correlation (AC), Root Mean Square Error (RMSE), and Brier Skill Score (BSS) for 500 hPa geopotential height (hereafter Z500) and 850 hPa temperature over the Northern Hemisphere (20°N–90°N) in September 2005. However, their verifications are only for particular month and they didn't discuss the daily forecast skill in detail, excluding the monthly verification of daily forecast.

The purpose of this study is to investigate the daily forecast skill of MCGE in comparison with that of a single-center ensemble using RMSE for Z500 over the Northern Hemisphere (20°N–90°N) from August 2005 to February 2006.

2 Data and Methodology

Medium-Range ensemble forecast data from three operational centers, JMA, NCEP and CMC, are used. The ensemble sizes of JMA, NCEP, and CMC are 25 (25), 44 (11), and 17 (17) per day (run), respectively, as of February 2006. RMSE was calculated for Z500 over the NH (20°N–90°N) from August 2005 to February 2006. Forecast and analysis fields have been interpolated onto a common regular $2.5° \times 2.5°$ grid, and each single-center ensemble has been verified against its own analysis, that is, each control run at initial time is regarded as each analysis. JMA analysis has been adopted as the analysis for verification of MCGE.

Following Matsueda et. al. (2006), we have constructed five ensemble mean forecasts, that is, JMA25, NCEP11, CMC17, J9N8C8, and J25N44C17 using above three single-center ensembles. JMA25, NCEP11, and CMC17 consist of ensemble members of each EPS initialized at 12 UTC, 12 UTC, and 00 UTC, respectively. J9N8C8, whose ensemble size is same as JMA25, contains JMA ensemble control run, 4 perturbation pairs of JMA, 4 perturbation pairs of NCEP starting from 12 UTC, and 4 perturbation pairs of CMC starting from 00 UTC without weights among ensemble members and bias corrections. J25N44C17 has the maximum ensemble size, namely 86. Initial time of MCGE forecasts is set to 12 UTC.

3 Results

3.1 Comparison of single-center ensemble and MCGE

Figure 1 illustrates a scatter diagram of $\mathrm{RMSE_{JMA25}}$ versus MCGE Improvement Rate (MIR) of RMSE for Z500 at 120 hr lead time during verification period from August 2005 to February 2006. The horizontal axis is $\mathrm{RMSE_{JMA25}}$. The vertical axis is MIR defined as ($\mathrm{RMSE_{JMA25}}$-$\mathrm{RMSE_{J9N8C8}}$)/$\mathrm{RMSE_{JMA25}}$. The positive (negative) MIR indicates J9N8C8 (JMA25) is more skillful than JMA25 (J9N8C8). MIR becomes zero (0.0) only when $\mathrm{RMSE_{J9N8C8}}$ is equivalent to $\mathrm{RMSE_{JMA25}}$, whereas MIR becomes one (1.0) only when $\mathrm{RMSE_{J9N8C8}}$ is zero (0.0). The blue and red circles indicate the verification period from August 2005 to October 2005 (ASON) and that from

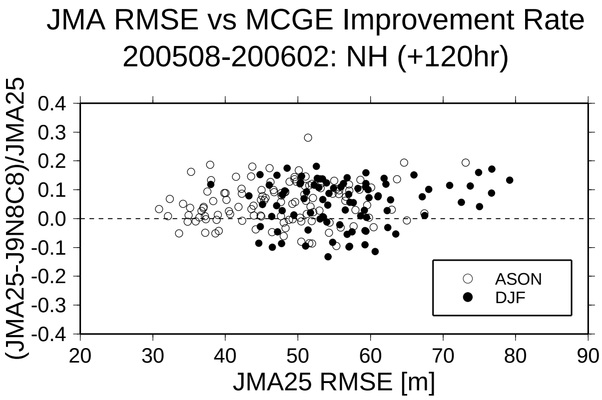

Figure 1. The scatter diagram of the RMSE of JMA25 versus MCGE improvement rate for Z500 RMSE at 120 hr lead time from August 2005 to February 2006. The blue and red circles indicate the period from August 2005 to November 2005 (ASON) and that from December 2005 to February 2006 (DJF), respectively.

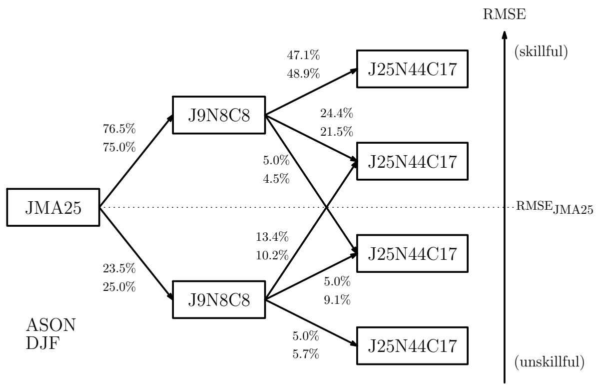

Figure 2. Comparison of RMSE among JMA25, J9N8C8, and J25N44C17. Vertical axis indicates RMSE for Z500 at 120 hr lead time. Note that the top and bottom of vertical axis are reversed, and the vertical distance among RMSEs of each ensemble mean doesn't indicate a realistic difference of RMSE. The percentages in this figure indicate the win rate of MCGE, which is placed at the right-side column of adjoined ensemble means, in each season.

November 2005 to February 2006 (DJF), respectively.

It is found in Fig. 1 that the frequencies that J9N8C8 is more skillful than JMA25 are 76.5% and 75.0% in ASON and DJF, respectively. The maximum (minimum) MIRs are 0.28 (-0.08) and 0.18 (-0.15) in ASON and DJF, respectively. In each season, the range of positive MIR is comparable without relation to the magnitude of RMSE. This indicates that whether the atmospheric field is easily-predictable or not, we can obtain a similar positive MIR. Also, it is noted that when $\mathrm{RMSE}_{\mathrm{JMA25}}$ is large, MIR tends to be positive, especially in DJF. When $\mathrm{RMSE}_{\mathrm{JMA25}}$ is large, RMSEs of another single-center ensemble are not always large. In other word, other single-center ensemble means, at least one single-center ensemble mean, sometimes can reduce the imperfections of model formulation or the uncertainties of initial value, even if JMA ensemble mean cannot reduce the uncertainties of initial value. Therefore, we can reduce them by replacing JMA members by other single-center members. It must be noted here that although of course MCGE does not have the worst forecast skill, MCGE does not always have the best forecast skill. Although the forecast skill is improved by MCGE, the forecast skill is often inferior to the best single-center ensemble. Precisely because we cannot know which ensemble member captures a extreme event correctly in advance, it seems to be appropriate to construct MCGE instead of single-center ensemble.

Although we have compared J9N8C8 with JMA25, it seems to be natural to consider how J25N44C17 with the maximum ensemble size is superior to JMA25 in the operational use of MCGE. The frequencies that J25N44C17 is more skillful than JMA25 are 84.9% and 80.7% in ASON and DJF, respectively (Fig. 2). It is found that 55.5% and 60.6% of them are due to the effects of multi-model and increase of the ensemble size in ASON and DJF, respectively. Furthermore, 28.7% and 26.6% of them are the results of the improved forecast skills by only the effect of multi-model. These results indicate that the effect of multi-model mainly leads to the advantage of J25N44C17 over JMA25. Also, it is rare that neither the effect of multi-model nor that of increase of the ensemble size appears. Even if the effect of multi-model does not appear, we can obtain J25N44C17, which is more skillful than JMA25, at the probabilities of 57.0% and 40.8% in ASON and DJF, respectively, by the effect of increase of the ensemble size.

3.2 Prediction of blocking

Finally, we show a very interesting case in terms of the prediction of blocking. Figure 3 illustrates the spaghetti diagrams (5500m) for (a) JMA, (b) NCEP, and (c) CMC from 00 UTC (CMC) or 12 UTC (JMA and NCEP) on 10th December 2005, and that for (d) MCGE (J9N8C8), valid 12UTC on 15th December. Blocking occurred at the upstream of the Rocky Mountains. All of NCEP ensemble members pre-

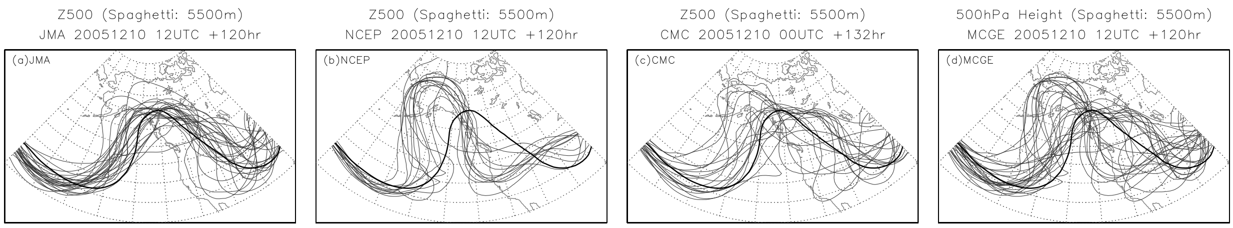

Figure 3. The Z500 spaghetti diagrams (5500 m) for (a) JMA 120 hr forecast, (b) NCEP 120 hr forecast, and (c) CMC 132 hr forecast initialized on 10th December 2005, and that for (d) MCGE (J9N8C8), valid 12 UTC on 15th December (thin solid line for each ensemble member, thick solid line for JMA analysis at the valid time).

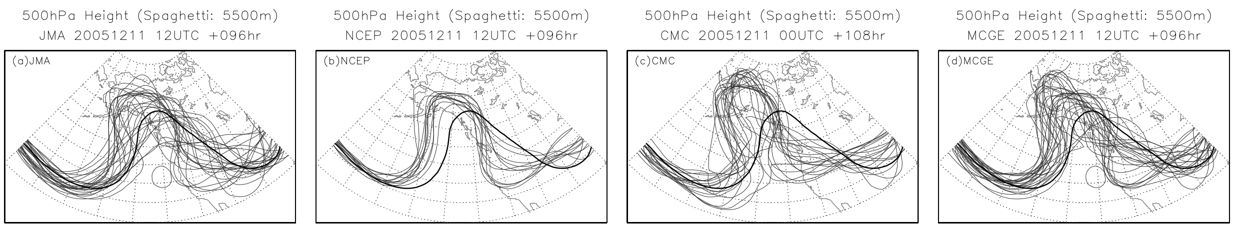

Figure 4. The Z500 spaghetti diagrams (5500 m) for (a) JMA 96 hr forecast, (b) NCEP 96 hr forecast, and (c) CMC 108 hr forecast initialized on 10th December 2005, and that for (d) MCGE (J9N8C8), valid 12 UTC on 15th December (thin solid line for each ensemble member, thick solid line for JMA analysis at the valid time).

dicted the wrong locations of the blocking, whereas JMA ensemble members and most of CMC ensemble members predicted the right locations of the blocking. The difference between RMSE of NCEP11 and that of JMA25 is about 30 m. In the forecast initialized on 11th December, CMC members mis-predict the location of blocking like NCEP members initialized on 10th December, whereas NCEP member seems to predict it better than that initialized on 10th December (Fig. 4). From these results, the mis-prediction seems to be due to not the imperfection of model formulation but the uncertainty of the initial value. In order to verify this hypothesis, we conducted the JMA-GSM (TL159L40) experiment with the initial value of NCEP control run. Figure 5 indicates the JMA analysis field on 12 UTC 15th December and 120 hr forecasts from 12 UTC 10th December 2005 at 500 hPa geopotential height. The JMA model run from the initial value of NCEP control run (Fig. 5d) mis-predicted the location of blocking like NCEP control run (Fig. 5c) . The time evolution of JMA model run from NCEP analysis is same as that of NCEP control run. From these results, we might be able to conclude that this misprediction is due to the initial value. Also, we conducted the JMA-GSM experiment with the initial value of 10 NCEP perturbed run (Fig. 6). As seen in Fig. 6, all perturbed members mis-predicted the location of blocking. They, however, seems to have predicted it better than NCEP original perturbed members shown in Fig. 3b. This might suggest that this mis-prediction was somewhat influenced by the imperfection of the model.

4 Conclusions

We investigate the daily forecast skills of three operational single-center ensembles by JMA, NCEP, and CMC, and MCGE consisting of these single-center ensembles. The forecast skill is evaluated by RMSE for Z500 over the Northern Hemisphere (20°N–90°N) from August 2005 to February 2006.

We compare the daily RMSE of MCGE with that of JMA ensemble, where the ensemble size is the same. It is found that MCGE is more skillful than JMA ensemble 76.5% and 75.0% of the time in ASON and DJF, respectively. This indicates that the multi-model ensemble mean can reduce the forecast errors due to the imperfection of model formulation, which the single-center ensemble mean cannot reduce. In each season, it is found that the effect of multi-model has little dependence on the atmospheric flow. This indicates that we can identically reduce the RMSE of MCGE up to about 20% whether the atmospheric field is easily-predictable or not. Also, it is noted that when the RMSE of JMA ensemble is large, MCGE tends to be almost more skillful than JMA ensemble, especially in DJF. This results from the advantage of other single-center en-

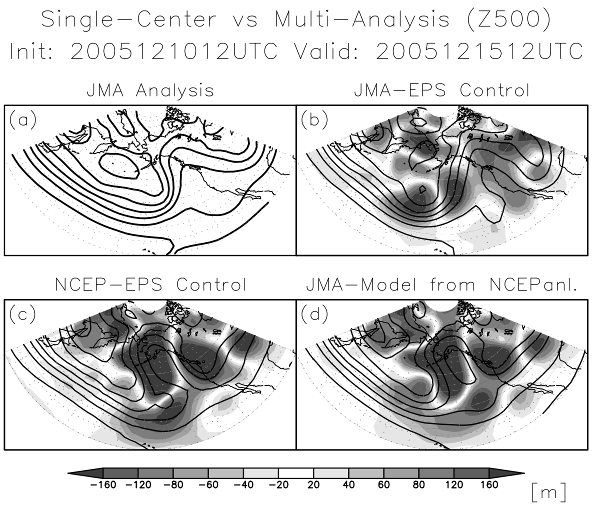

Figure 5. The JMA analysis field on 12 UTC 15th December 2005 and the 120 hr forecasts from 12 UTC 10th December 2005, at 500 hPa geopotential height. (a) JMA analysis filed on 12 UTC 15th december 2005, and 500 hPa height (contour) and the forecast error (shading) of (b) JMA-EPS control run, (c) NCEP-EPS control run, and (d) JMA-GSM run from NCEP analysis.

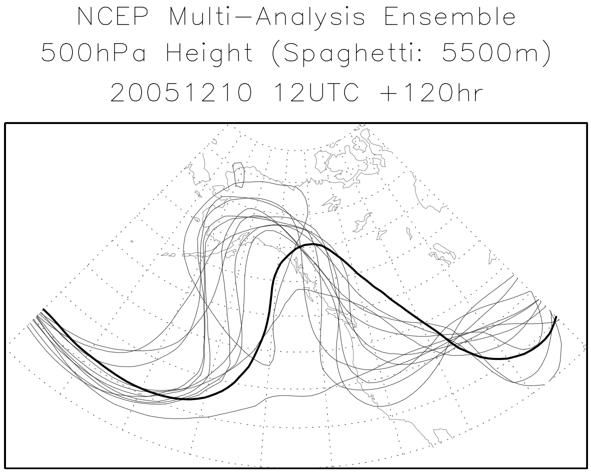

Figure 6. The Z500 spaghetti diagrams (5500 m) for the 120 hr forecast of NCEP multi-analysis ensemble initialized on 10th December 2005 using JMA-GSM (TL159L40).

sembles over JMA ensemble. It must be noted, however, that MCGE is not always the most skillful if not the worst. Although the forecast skill is improved by MCGE, that is often inferior to the best single-center ensemble. We argue that it is a benefit of MCGE to avoid the poorest forecast.

Furthermore, MCGE with the maximum ensemble size of 86 outperforms the JMA ensemble 84.9% and 80.7% of the time in ASON and DJF, respectively. This mainly results from the effect of multi-model, although the increase of the ensemble size improve the forecast skill.

Also, we showed a very interesting case in terms of the prediction of blocking occurred at the upstream of the Rocky Mountains on 15tth December 2005. This is rare case that all of NCEP ensemble members predicted the wrong locations of the blocking, whereas JMA ensemble members and most of CMC ensemble members predicted the right locations of the blocking. From the results of numerical experiment, it is found that the mis-prediction of NCEP members is mainly due to the initial value.

Acknowledgments

This work could not have been done without the operational ensemble data of NCEP and CMC. We are especially thankful to NCEP and CMC.

References

Buizza, R., D.S. Richardson, and T.N. Palmer, 2003: Benefits of increased resolution in the ECMWF ensemble system and comparison with poor-man's ensembles. *Quart. J. Roy. Meteor. Soc.*, **129**, 1269–1288.

Krishnamurti, T.N., C.M. Kishtawal, T.E. LaRow, D.R. Bachiochi, Z. Zhang, C.E. Williford, S. Gadgil, and S. Surendran, 1999: Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, **285**, 1548–1550.

Matsueda M., M. Kyouda, H.L. Tanaka and T. Tsuyuki, 2006: Multi-Center Grand Ensemble using three operational ensemble forecasts. *SOLA*, **2**, 33–36.

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.

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

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