Daily Forecast Skill of Multi-Center Grand Ensemble

Mio Matsueda¹, Masayuki Kyouuda¹, H. L. Tanaka²,³,⁴ and Tadashi Tsuyuki²
¹Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan
²Numerical Prediction Division, Japan Meteorological Agency, Tokyo, Japan
³Center for Computational Sciences, University of Tsukuba, Tsukuba, Japan
⁴Frontier Research Center for Global Change/JAMSTEC, Yokohama, Japan

Abstract

In this study, we investigate the daily forecast skill of Multi-Center Grand Ensemble (MCGE), consisting of three operational medium-range ensemble forecast data by the Japan Meteorological Agency (JMA), the National Centers for Environmental Prediction (NCEP), and the Canadian Meteorological Center (CMC). The skill is evaluated by comparison among the daily RMSE of ensemble mean forecasts for 500 hPa geopotential height over the Northern Hemisphere (20°N–90°N) from August 2005 to February 2006.

It is found that MCGE with the same ensemble size as that of the JMA ensemble is more skillful than JMA ensemble for about 75% in frequency both in autumn and winter. Reduction of error with MCGE has little dependence on the atmospheric flow. The RMSE of MCGE can be reduced up to about 20% whether the atmospheric field is easily-predictable or not. Even for the case that MCGE is not more skillful than JMA, the RMSE is increased at most 10%. We argue that the major benefit of MCGE is to avoid the poorest forecast.

1. Introduction

Ensemble forecast techniques have drawn more attention in various timescales, such as short-, medium-, and long-ranges for both operational and research purposes. In 1992, the European Center for Medium-Range Weather Forecasts (ECMWF; Molteni et al. 1996) and the National Centers for Environmental Prediction (NCEP; Toth and Kalnay 1997) employed the ensemble techniques for operational medium-range forecast for the first time followed by other numerical weather prediction (NWP) centers, such as the Japan Meteorological Agency (JMA; JMA 2002), the Canadian Meteorological Center (CMC; Houtekamer et al. 1996), and so on.

Besides the ensemble of the initial value, there are other techniques, such as the multi-model superensemble, poor man’s ensemble, the multi-analysis ensemble, the multi-model multi-analysis ensemble, and so on (e.g., Krishnamurti et al. 1999; Ziehmann 2001; Buizza et al. 2003; Richardson 2001; Mylne et al. 2003). In particular, the multi-model ensemble technique combines the independent forecast data from different models in order to represent not only the uncertainties of initial value but also the imperfection of the model formulation.

Recently, rapid progress of communication networks enables 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 an operational basis. 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 a particular month and they didn’t discuss the daily forecast skill in detail, excluding the monthly verification of daily forecasts.

In general, it is not easy to predict an atmospheric phenomenon correctly, especially an extreme event such as blocking, due to the chaotic nature of the atmosphere, imperfections of model formulation, uncertainties of initial value, and so on. MCGE using some NWP’s ensemble members may suggest the possible occurrence on extreme events, even if a single-center ensemble cannot detect them. 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. Configurations of operational ensembles and MCGE

In this study, ensemble forecast data from three operational centers, JMA, NCEP and CMC, are used. The details of Ensemble Prediction System (EPS) as of February 2006 are summarized in Table 1. The development of ensemble forecast is so much rapid that EPS, thus model resolution or ensemble size, often changes. Since March 2006, there have been apparent changes of EPS in quick succession in each NWP center, so we have chosen the period from August 2005 to February 2006 as a verification period. In our verification period, there were not remarkable changes of EPS, such as increase of the ensemble size, except for the change of the model resolution of NCEP.

Following Matsueda et al. (2006), we have constructed five ensemble mean forecasts, that is, JMA25, NCEP11, CMC17, J9N8C8, and J25N44C17 using above three ensembles. JMA25, NCEP11, and CMC17 consist of ensemble members of each EPS initialized at 12, 12, and 00 UTC, respectively. J9N8C8, which is constructed to compare with 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.

Table 1. Three ensemble configurations at JMA, NCEP, and CMC as of February 2006.

<table>
<thead>
<tr>
<th>Model Resol.</th>
<th>Init. Perturb.</th>
<th>Init. UTC</th>
<th>Mem./day</th>
<th>Grid size of original data</th>
</tr>
</thead>
<tbody>
<tr>
<td>JMA25</td>
<td>BVs</td>
<td>12, 00, 6, 12, 18</td>
<td>25</td>
<td>2.5°×2.5°</td>
</tr>
<tr>
<td>NCEP11</td>
<td>BVs</td>
<td>00</td>
<td>11×4</td>
<td>2.5°×2.5°</td>
</tr>
<tr>
<td>CMC17</td>
<td>EnKF</td>
<td>25</td>
<td>17</td>
<td>1.0°×1.0°</td>
</tr>
<tr>
<td>J9N8C8</td>
<td></td>
<td>12</td>
<td></td>
<td>1.0°×1.0°</td>
</tr>
<tr>
<td>J25N44C17</td>
<td></td>
<td>25</td>
<td></td>
<td>1.0°×1.0°</td>
</tr>
</tbody>
</table>
3. Verification Score

We investigate the daily forecast skill of JMA25, NCEP11, CMC17, J9N8C8, and J25N44C17 using RMSE of ensemble mean forecast. RMSE is used as a measure of the forecast error. RMSE is calculated for Z500 over the Northern Hemisphere (20°N-90°N). Forecast and analysis fields are interpolated onto a common regular 2.5°x2.5° grid, and each single-center ensemble has been verified against its own analysis, that is, each control run at the initial time is regarded as each analysis. JMA analysis has been adopted as the analysis for verification of MCGE. We denote RMSE of JMA25 as RMSE_{JMA25}. The same manner applies to the other ensemble means.

4. Results

4.1 Daily forecast skills of single-center ensembles

Figure 1a illustrates the daily variabilities of RMSE_{JMA25}, RMSE_{NCEP11}, and RMSE_{CMC17} at the 120-hr lead time (RMSE of each control run is also shown in supplement 1). The day number in the abscissa refers to the initialization date, 0 being 12 UTC 31th July 2005. It must be noted that CMC17 is verified at 00 UTC, whereas JMA25 and NCEP11 are verified at 12 UTC. RMSE_{CMC17} in Fig. 1 is plotted from 0.5 to 211.5 in the abscissa. The solid lines of NCEP11 (green) and CMC17 (yellow) are discontinuous at several places due to data deficiency caused by data transfer.

It is found that each forecast skill varies slowly throughout the verification period irrespective to seasons. The fluctuation of RMSE is large during the winter and small during the summer. Paying attention to daily scores, it is also found that RMSEs vary considerably depending on atmospheric flow of the day, as seen in Kimoto et al. (1992), who have investigated the daily scores with single deterministic forecasts from three operational centers, JMA, ECMWF, and NCEP.

On the other hand, there are some cases that the skill of a certain ensemble mean forecast is extremely unskillful than that of the others, as seen in Fig. 1a. For example, the forecasts for Z500 initialized at 10th December 2005, namely 132 days, are one of the cases in which there is a well-defined difference in RMSE. The difference between RMSE_{NCEP11} and RMSE_{JMA25} is about 30 m. Also, the difference between RMSE_{NCEP11} initialized at 10th December and that initialized at the day before reaches 40 m. These mainly result from the occurrence of blocking at the upstream of the Rocky mountains on 15th December 2005 in Fig. 2. Figure 2a indicates that JMA25 is the best forecast among single-center ensemble mean forecasts, whereas NCEP11 is the worst one in this case. In terms of each member in Fig. 2b, all of NCEP ensemble members initialized at 12 UTC on 10th December predicted the wrong location of the blocking, whereas JMA ensemble members initialized at 12 UTC on 10th December and most of CMC ensemble members initialized at 00 UTC on 10th December predicted the right location of the blocking (We will discuss these predictions of blocking in detail in the next paper).

Interestingly, the peak of RMSE differ among the single-center ensembles, and the poorest forecast alternate among the single-center ensembles. RMSE_{NCEP11} decreases after 10th December but RMSE_{JMA25} and RMSE_{CMC17} increase. Although the forecast error is flow-dependent, the differences in the performance among single-center ensembles on the same day imply that there are significant errors due to the differences in the model formulation, initial value, or initial perturbations.
It is not always true that a certain ensemble mean forecast is most skillful, although ensemble sizes are different from each other. Also, each ensemble mean seems to have strong and weak points of forecast depending on a period, respectively. For example, NCEP11 is almost more skillful than JMA25 and CMC17 during October 2005.

4.2 Comparison of single-center ensemble and MCGE

Figure 2b illustrates the daily variations of RMSE$_{JMA25}$ and RMSE$_{J9N8C8}$. The red line shows the same RMSE$_{JMA25}$ as in Fig. 1a. Each ensemble size is 25. It is found that J9N8C8 tends to be more skillful than JMA25 without depending on seasons. When the improvement by MCGE appears clearly in Fig. 1b such as 68, 75, 112, 127, and 141 days, the other single-center ensembles are sometimes more skillful than JMA25 in Fig. 1a. This suggests that 16 members of JMA, which are replaced with 8 members of NCEP and 8 members of CMC, are less skillful than them. On the other hand, JMA25 is sometimes more skillful than J9N8C8, such as around 103, 132, and 160 days. The skills of J9N8C8, however, are not remarkably worse but slightly worse than JMA25. In cases that JMA25 is more skillful than J9N8C8 (red circles in Fig. 1a) or is similar to J9N8C8, RMSE of other single-center ensemble seems to be obviously larger than that of JMA25.

Next, we make a scatter diagram on forecast skill in order to evaluate the above results quantitatively. Figure 3 illustrates a scatter diagram of RMSE$_{JMA25}$ versus RMSE$_{J9N8C8}$ Improvement Rate (MIR) of RMSE for Z500 at the 120-hr lead time during the verification period from August 2005 to February 2006. The horizontal axis is RMSE$_{JMA25}$, whereas 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 November 2005 to February 2006 (DJF), respectively.

It is found in Fig. 3 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 RMSE$_{JMA25}$ is large, MIR tends to be positive, especially in DJF. As seen in Fig. 1, when RMSE$_{JMA25}$ is large, RMSEs of another single-center ensemble are not always large. In other words, other single-center ensemble means, at least one single-center ensemble mean, sometimes can reduce the forecast error due to the uncertainties of initial value, even if JMA ensemble mean cannot reduce it. Therefore, we can reduce the forecast errors due to not only the uncertainties of initial value but also the imperfections of model formulation 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, that is often inferior to the best single-center ensemble. This can be also expected from Fig. 2. As seen in Fig. 2a, JMA25 is obviously the best forecast in all ensemble mean forecasts, whereas NCEP11 is obviously the worst one. Constructing MCGE by using them, MCGE is less skillful than the best single-center ensemble, namely JMA25, but undoubtedly more skillful than the worst single-center ensemble, namely NCEP11. Precisely because we cannot know which ensemble member captures extreme events correctly in advance, it seems to be appropriate to construct MCGE instead of single-center ensemble. It is found from Fig. 2b that MCGE contains various possibilities of detecting extreme events, which a single-center ensemble cannot contain. This leads to avoid the worst ensemble mean forecast. Also, MCGE might enable us to get a helpful measure of the probability of extreme events in the daily use.

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 (see supplement 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 improvement by multi-model mainly leads to a better skill 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.

5. 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.

First, we compare the daily RMSEs of three single-center ensemble mean forecasts. RMSEs vary considerably depending on atmospheric flow of the day throughout the analysis period. On the other hand, RMSEs from different NWP centers vary differently with time. None of the ensemble mean forecasts are always more skillful than the others.

Next, 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 better skills of other single-center ensembles 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 the major benefit of MCGE is 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 improves the forecast skill.

6. Discussions

We showed that the blocking on 15th December 2005 is an example of the case that the forecast skill of a certain single-center ensemble is extremely unskillful than that of the others in Fig. 2. It is generally difficult to predict an extreme event, such as blocking. As seen in Fig. 2b, however, MCGE members represent various possibilities of detecting extreme events. Considering an ensemble mean of them, we can avoid the worst score, and obtain a better score than single-center ensemble if not the best. Also, MCGE might enable us to get a helpful measure of probabilistic information on occurrence of an extreme event in the daily use. Precisely because we cannot know which member captures an extreme event in advance, MCGE instead of single-center ensemble is desirable in terms of the deterministic and the probabilistic forecast.

It is found in comparison with MCGE that a single-center ensemble can sometimes provide misleading forecast information especially on extreme events such as the blocking on 15th December 2005. In these cases, MCGE may represent model-related forecast uncertainty, and may enable us to identify origin of the forecast-error source, that is, model resolution, initial value, or initial perturbations. When the exchange and accumulation of ensemble forecast data provided by various NWP centers start in full swing by TIGGE (THORPEX Interactive Grand Global Ensemble), the research on MCGE or that using MCGE, especially for prediction of extreme events, might be increasingly important.

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. Useful comments made by the editor and anonymous reviewers helped to improve the revisions significantly. A part of the research support comes from Grant-in-Aids (18204043) of the Japan Society for Promotion of Science.

Supplements

1. Time series of daily skills of single-center control runs.
2. Comparison diagram of RMSE.

References


Manuscript received 7 September 2006, accepted 22 January 2007
SOLA: http://www.jstage.jst.go.jp/browse/sola/