Sensitivity of Long-Range Forecast Skill Score to Verification Data

Yukiko Naruse, Shuhei Maeda
Climate Prediction Division, Japan Meteorological Agency, Japan
y.naruse@met.kishou.go.jp

1 INTRODUCTION

In the manual (WMO, 2006) on the Standardised Verification System for Long-Range Forecasts (SVSLRF) of WMO, UKMO/CRU (Mitchell and Jones, 2005) and ERA-40 (Uppala et al, 2005) are recommended verification data for surface air temperature anomaly at screen level (2-meter temperature). As an exception, the forecasts issuing centers own analyses may be used. Since the data period of ERA-40 is only up to August 2002, it can not be used as verification data for hindcasts after 2003. Furthermore, the update of UKMO/CRU is not frequent enough to be used as verification data for real-time monitoring forecasts. As a verification system, it is better to use the same dataset for the whole period of hindcasts and real-time monitoring forecasts.

Therefore, useful data is the product of the Japanese 25-year ReAnalysis (JRA-25) (Onogi et al, 2007) and succeeding real time products of JMA Climate Data Assimilation System (JCDAS) which is in real time operation by JMA with the same data assimilation system as JRA-25. The JRA-25 covers the 26-year from Jan.1979 to Dec.2004, and continues from Jan.2005 to the present with JCDAS. Onogi et al (2007) showed that data quality of 2-meter temperature in JRA-25 is similar to that of UKMO/CRU and ERA-40.

In JMA, UKMO/CRU has been used as verification data of hindcasts under the manual on SVSLRF. However, considering the merit of JRA-25 and JCDAS mentioned above, JMA is going to use JRA-25 as verification data of 2-meter temperature prediction instead of UKMO/CRU. Before officially using JRA-25 as verification data, we examine sensibility of forecast skill scores, using JRA-25, UKMO/CRU and ERA-40 as verification data.

The verification methodology is described in Section 2. The forecast skill scores using three verification datasets are compared in Section 3. Conclusions and discussion are drawn in Section 4.

2 Verification methodology

2.1 Data

The forecast skill scores for hindcast experiments are evaluated as deterministic and probabilistic forecasts. The verification element is a 3-month mean 2-meter temperature with about 20 days lead time for four conventional seasons DJF, MAM, JJA and SON. The verification is performed in a cross-validation framework.

The specifications of the hindcasts used in this study are shown in Table 1. Except for the ensemble size, the specifications are the same as those of the operational ensemble prediction system in JMA which was renewed in September 2007.

The predicted 2-meter temperature in the hindcasts is verified with three observational datasets: UKMO /CRU, ERA-40 and JRA-25. The dataset version of UKMO/CRU is CRU TS 2.1 (Mitchell and Jones, 2005), and the original data of UKMO/CRU is converted to the SVSLRF standard 2.5x2.5 degrees grids by the Lead Center.

2.2 Verification scores

The Mean Square Skill Score (MSSS) is used for deterministic forecasts. In a case of perfect forecast, the MSSS equal to 1.0, and tends to 0 for a climatology forecast (no information).

The Relative Operating Characteristics (ROC) area is used for probabilistic forecasts. Under the manual on SVSLRF, the ROC area is used as skill scores of probabilistic prediction. The ROC area is characterized as 1.0 as a perfect forecast and 0.5 as a climatology forecast (no information).

Verification period is 1984-2001 (18years), because the ERA-40 is only available by 2001.

3 Impact of verification data on forecast skill socre

3.1 Deterministic scores

Figure 1 shows the Mean Square Skill Score (MSSS) of 3-month mean 2-meter temperature using each verification data over land of the Northern extra-tropics. The differences between each scores using three verification data are smaller than 10% and are not significant for all seasons. Over the Southern extra-tropics and the Tropics, differences are larger than those over the Northern extra-tropics, however there are no significant differences (not shown).

Figure 2 shows geophysical distribution of the difference between the MSSS using JRA-25 and those using ERA-40 for summer (JJA) and winter (DJF) forecasts in the Northern Hemisphere. It is found that there is larger difference over the Western Pacific and high-latitude of the Southern hemisphere in JJA (Figure 2a), as well as over the Arctic Ocean in DJF (Figure 2b). On the other hand, difference is rather smaller over land.

3.2 Probabilistic scores

Figure 3 shows the ROC area using each verification data over land of the Northern extra-tropics. The event is the probability that the 3-month mean 2-meter temperature is in the upper tercile. The differences between each score using three verification data are smaller than 1% and are not significant over the four seasons. Similar results are obtained for the ROC area in the lower tercile (not shown).

Figure 4 shows geophysical distribution of the difference between the ROC area using JRA-25 and those using ERA-40 for JJA and DJF forecasts. It is found that there is larger difference in the Arctic Ocean in DJF (Figure 4b). On the other hand, the differences in JJA are not so large (Figure 4a).

4 Conclusions and discussion

4.1 Conclusions

This work is focused on the sensitivity of long-range forecast skill score to verification data (JRA-25, ERA-40 and UKMO/CRU).

According to the results of the deterministic scores (MSSS) and probabilistic scores (ROC area) over land using three verification data, the impact of verification data on the forecast skill scores are very small. There are also no significant differences between the scores using UKMO/CRU and ERA-40 recommended in the manual on SVSLRF and those using JRA-25 (Figure 1 and Figure 3). These results suggest that the quality of 2-meter temperature in JRA-25 as verification data is equivalent to that of UKMO/CRU or ERA-40.

Based on the results of this study and the merit of JRA-25 and JCDAS in the verification of hindcasts as real-time availability, JMA is going to use JRA-25 as verification data of 2-meter temperature prediction instead of UKMO/CRU.

4.2 Discussion

95% confidence intervals of the skill scores (Figure 1 and Figure 3) using bootstrap procedures indicate the uncertainty of verification sampling. Considering the confidence interval, results have shown that verification sampling have larger influence on the forecast skill scores than difference of verification data. Figure 5 displays the ROC area using JRA-25 over land of the Northern extra-tropics. The event is same as Figure 3. The hindcast period is 18years (from 1984 to 2001) and 22years (from 1984 to 2005) respectively. The difference between two scores is larger than those of verification data as shown in figure 3.

These results suggest that it is more important than verification data that we have the same verification sampling and hindcast period as other centers when the forecast skill scores are compared to other centers.

Figure 2 and Figure 4 show, the distribution of skill scores using JRA-25 and ERA-40 have larger differences over the high-latitude and sea. Considering difference over sea, the impact of the difference of precipitation between JRA-25 and ERA-40 showed by Onogi et al (2007) appears on the result such as those over Western Pacific in particular. Considering difference over high-latitude, the difference of SST over high-latitude of Sothern Hemisphere between two datasets showed by Onogi et al (2007) appears on the differences between two scores. However, more discussions are necessary, and we like to continue more study.

TABLES and FIGURES

Table 1. Specifications of hindcast

Model type	Two-tiered method is used. The atmospheric model is TL95L40 version of the Global Spectral model used for a short- and medium-range forecast in JMA
Boundary conditions	SST: Combination of persisted anomaly, climate and prediction with the El Nino prediction model (atmosphere-ocean coupled model; CGCM) in JMA
Ensemble size and ensemble method	11 members. Singular vectors are used for atmospheric initial perturbation.
Training period	22 years from 1984 to 2005. Initial date is 10th of every month.
Forecast range	120 days

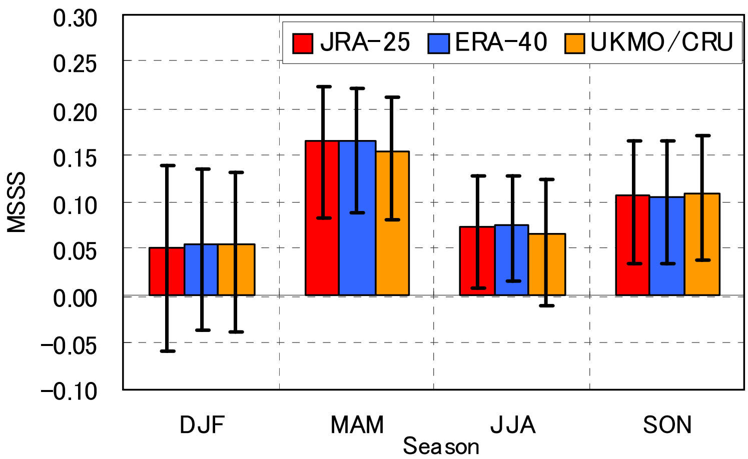

Figure 1 The Mean Square Skill Score (MSSS) of 3-month mean 2-meter temperature over land of the Northern extra-tropics (from 20°N to 90°N) for each season (DJF, MAM, JJA and SON) from 1984 to 2001. Initial dates are 10th Nov. for DJF, 10th Feb. for MAM, 10th May for JJA and 10th Aug. for SON. Red bar represents verification with JRA-25, blue bar with ERA-40 and orange bar with UKMO/CRU. Error bars represent 95% confidence interval using bootstrap procedures.

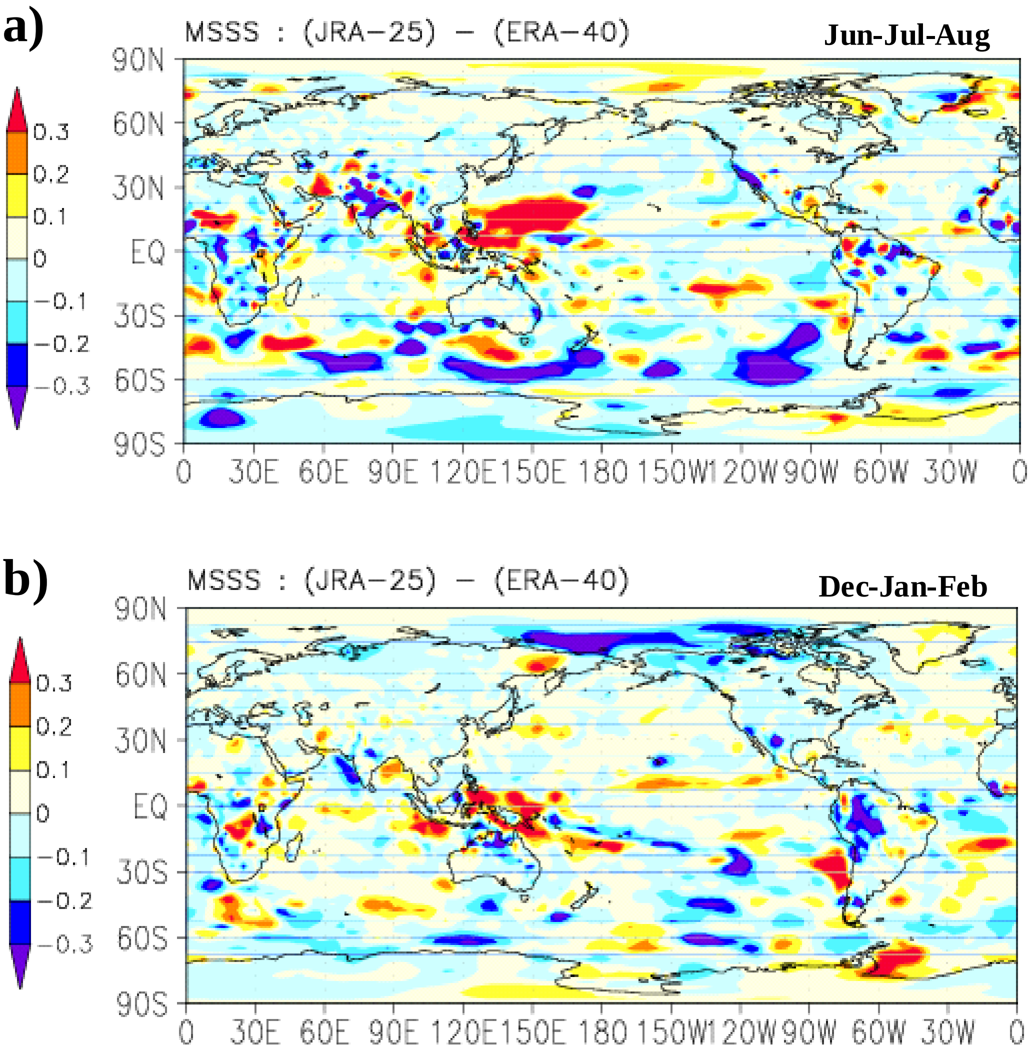

Figure 2 Difference between the MSSS using ERA-40 and the MSSS using JRA-25. Upper panel (a) shows scores in JJA, lower panel (b) shows scores in DJF.

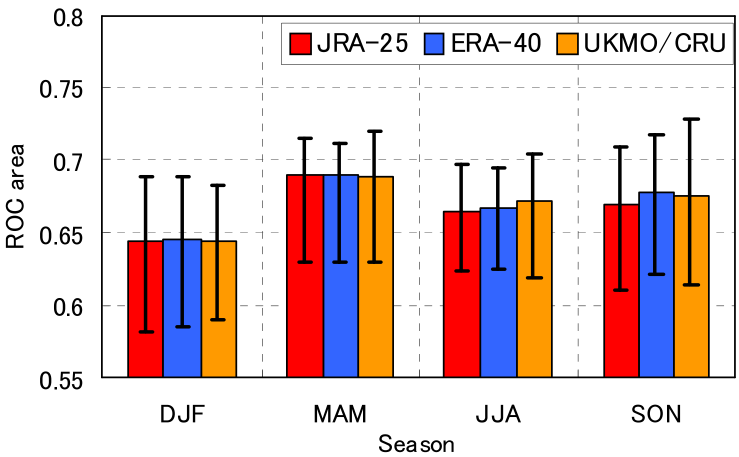

Figure 3 Same as Figure 1, but for the ROC area for the 2-meter temperature in the upper tercile.

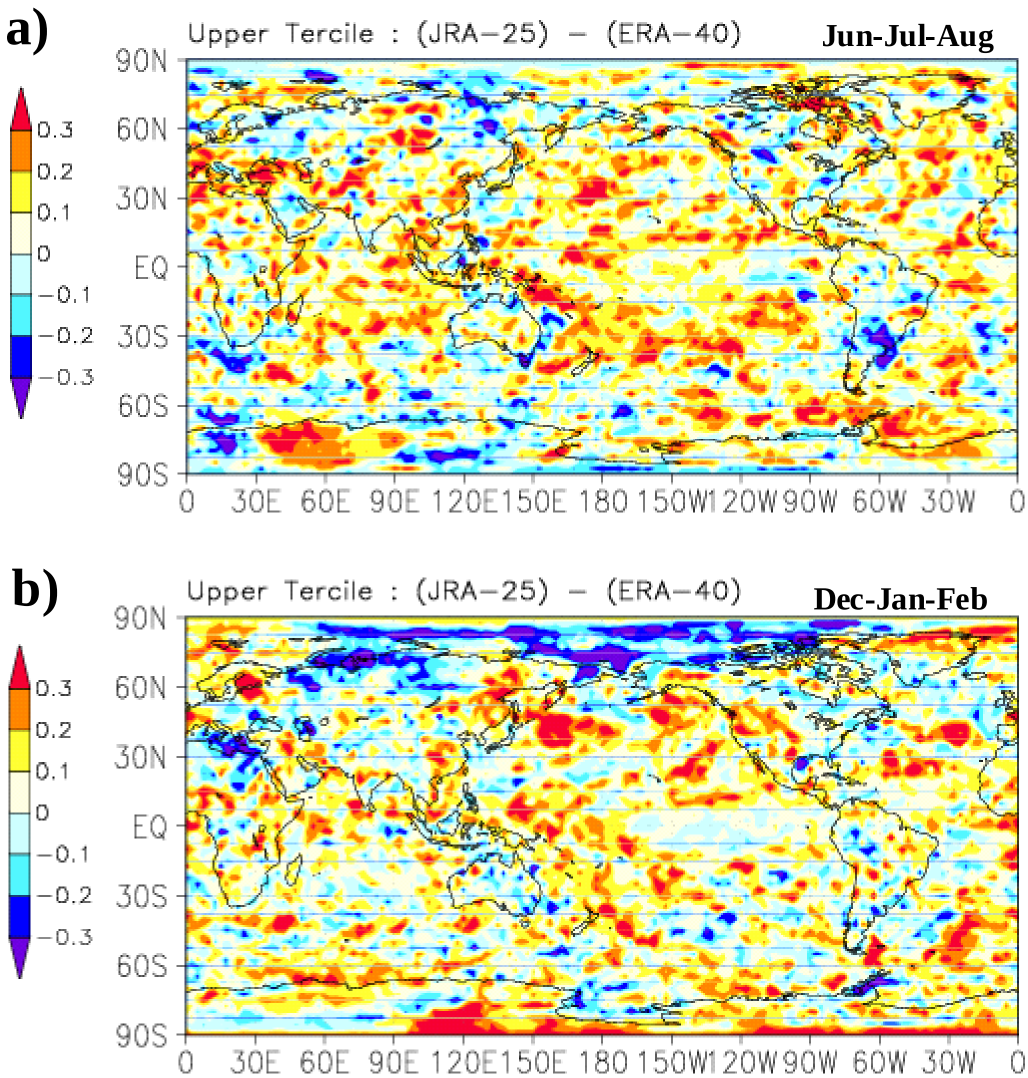

Figure 4 Same as Figure 2, but for the ROC area for upper tercile.

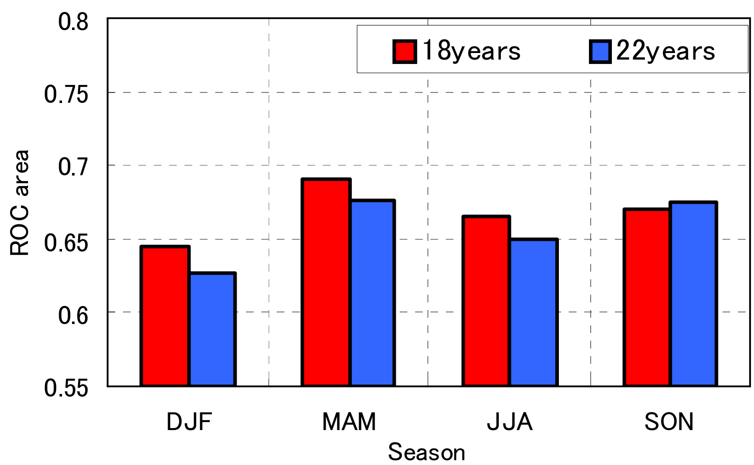

Figure 5 Same as Figure 3, but red bar represents scores with 18years (1984-2001) hindcast, blue bar represents scores with 22years (1984-2005) hindcast. The verification data is JRA-25.

REFERENCES

Mitchell, T.D., P.D. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grid. Int. J. Climatol., **25**, 693-712.

Onogi, K., J. Tsutsui, H. Koide, M. Sakamoto, S. Kobayashi, H. Hatsushika, T. Matsumoto, N. Yamazaki, H. Kamahori, K. Takahashi, S. Kadokura, K. Wada, K. Kato, R. Oyama, T. Ose, N. Mannoji and R. Taira 2007: The JRA-25 Reanalysis. J. Meteor. Soc. Japan, **85**, 369-432.

Uppala, S.M., P.W. Kallberg, A.J. Simmons, U. Andrae, V. da Costa Bechtold, M. Fiorino, J.K Gibson, J. Haseler, A. Hernandez, G.A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R.P.Allan, E. Andersson, K. Arpe, M.A. Balmaseda, A.C.M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Holm, B.J. Hoskins, L. Isaksen, P.A.E.M. Janssen, R. Jenne, A.P. McNally, J.-F. Mahfouf, J.-J. Morcrette, N.A Rayner, R.W. Saunders, P. Simon, A. Sterl, K.E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo and J. Woollen 2005: The ERA-40 re-analysis. Quart. J. R. Meteorol. Soc., **131**, 2961-3012.

WMO, 2006: Standardised Verification System (SVS) for Long-Range Forecasts (LRF). New Attachment II-8 to the Manual on the GDPFS (WMO-No.485), Volume I, 83pp