The Trade-off in Bias Correction between Using the Latest Analysis/Modeling System with a Short, vs. an Older System with a Long Archive

Bo ***Cui***[1]*, Zoltan* ***Toth***[2]*, Yuejian* ***Zhu***[2]*, Dingchen* ***Hou***[1]*, David* ***Unger***[3]*,*
Stéphane ***Beauregard***[4]

[1]SAIC at Environmental Modeling Center, NCEP/NWS, Email Bo.Cui@noaa.gov, Dingchen.Hou@noaa.gov
[2] Environmental Modeling Center, NCEP/NWS, Email Zoltan.Toth@noaa.gov, Yuejian.Zhu@noaa.gov
[3] Climate Prediction Center, NCEP/NWS, Email David.Unger@noaa.gov
[4] Canadian Meteorological Centre, Meteorological Service of Canada, Email Stephane.Beauregard@ec.gc.ca

Abstract

The main tasks of this study is to develop and implement a set of statistical post-processing schemes to reduce the biases in the US National Weather Service (NWS) and the Meteorological Service of Canada (MSC) ensemble forecasts. Three methods are designed to assess and mitigate ensemble biases on model grid with respect to analysis fields in order to adjust the 1st (mean) moment of the ensemble, which are the decaying averaging bias estimate technique, the climate mean bias estimate technique, and the bias estimate using dependent data technique, respectively. These methods provide trade-off studies in bias correction between using the latest analysis/modeling system with a short versus older system with a long archive. Preliminary results show that an adaptive, regime dependent bias correction method works well for the first few days. The reforecast ensembles from the Climate Diagnostics Center (CDC) with and without climate mean bias correction shows that climate mean bias correction can add value, especially for week 2 probability forecasts.

1. Introduction

Within the last decade, ensemble based on global models has been found useful for medium-range probabilistic forecasting. Ensemble forecasting has been embraced as a practical way of estimating the uncertainty of weather forecast and making probabilistic forecast (Toth and Kalnay 1993, 1997; Molteni et al. 1996). However, ensemble forecasts still suffer from model and ensemble formation related shortcoming. As Toth et al. (2003) indicated, these systematic errors would remain and cause bias in the 1st and 2nd moments of the ensemble distribution. In order to make a skillful medium-range forecast it is necessary to run post-processing algorithms to remove these systematic errors before the ensemble forecasts can be used. Thus, the main task of this study is to develop and implement a set of statistical post-processing schemes to reduce the biases in ensemble forecasts against the data on analysis fields.

In this paper, we first investigate a set of statistical post-processing algorithms that are designed to adjust the 1st moment of ensemble forecasts. These methods are being developed jointly by the National Centers for Environmental Prediction (NCEP) of the US NWS and the MSC, and will be implemented operationally for reducing the bias from the NWS and MSC ensemble forecasts before they are merged to form a joint ensemble within the North American Ensemble Forecast System (NAEFS).

Beyond the global ensemble based on the currently available best analysis/modeling system, there is another ensemble run operationally at the NCEP based on a frozen analysis/modeling system, developed by the scientist of the CDC (reforecast). Bias correction of this ensemble is supported by a 25 year ensemble reforecast experiment. Since the operational analysis/modeling system undergoes frequent (once or twice per year) changes, no such long archive is available for the ensemble based on the most recent analysis/modeling system. Bias correction of this ensemble will be base on data for the most recent season. One goal of the this research will be to compare the relative merits of using the current best analysis/modeling system with a small sample, versus an older and frozen analysis/modeling system but with a longer sample of forecasts for the bias correction.

2. Methods

The operational application environment requires that the post-processing algorithms of ensemble forecasts be relatively cheap and flexible for operational real-time running. The 1st moment adjustment method is implemented in two steps. The first step is to estimate the 1st moment bias with respective to the analysis field. The second step is to remove the error from the ensemble forecasts. Three algorithms to asses the 1st moment bias are introduced in the following section.

a. Decaying averaging bias estimate

The first method introduced in this paper is an adaptive (Kalman Filter type) algorithm, which has the following application procedure: (a) Prior estimate to start up the procedure. At a given day T, calculate the time mean forecast errors between day T - 46 and T - 17 to initializes an average. (b) Update step. The average is updated by setting it to the weighted average of the new forecast error at day T –16, with a weight of w, and the previous average, with a weight of 1-w (0<=w<1). (c) Cycling: repeat step (b) every day from day T-15 to T-1. Such a decaying average bias assessment method is a convenient way to consider the most recent behavior of a system. Once initialized, the bias estimate can be updated by only considering the most recent forecast error regarding the storage of the fields (the prior). The weight factor *w* controls how large the influence of the most recent data is. Experiments with different *w* (1%, 2% and 10%, respectively) have been conducted. In general, the 2% work better for most regions and seasons than the 1% and 10% (not shown). Therefore we continue the study of the decaying average approach with a 2% weight and compare it to the other bias correction technique. The decaying average bias assessment method is applied to the NCEP operational ensemble forecast.

b. Climate mean bias estimate

A second method to assess the ensemble bias is by using the climatological mean forecast error, which is gotten from the CDC 25-year reforecast (from 1978 to 2003). Hamill et al (2003) thought that it is not effective to do bias correction with only a short set of prior forecasts because systematic errors may not be well established if only a few cases are tested on but may be more obvious with a larger sample afforded by reforecast. With the Model Output Statistics (MOS) techniques and a frozen forecast model, their results show that dramatic improvements in medium-to extended-range probabilistic forecasts are possible by using retrospective forecasts. Motivated by their success, especially for the probabilistic forecast for week 2, we introduce the climatological mean forecast error into our bias correction study and remove it from the CDC reforecast. The reforecast ensembles with and without the climatological bias correction are then examined and compared to the operational ensemble.

c. Bias estimate using dependent data

A third way to estimate the 1st moment bias of the ensemble is through the calculation of 31-day running mean forecast error centered on day T. The implementation of this method is operationally not feasible but used as an optimal benchmark. The optimal scenarios therefore are compared to the raw and calibrated ensembles to show how large the possibility could be to improve the ensemble forecasting by using the 1st moment adjustment technique. This method is applied on both of the operational and the reforecast ensembles.

3. Experimental data and design

After applying these bias estimate techniques discussed above, each of the operational and the reforest create three different ensembles, respectively, which are the raw, the bias-corrected and the optimal ensemble, respectively. For the operational ensemble, the three ensembles are the OPR_RAW, the OPR_DAV2% (remove the decaying average bias) and the OPR_OPT. For the reforecast, they are the RFC_RAW, the RFC_COR (remove the climitalogical mean bias) and the RFC_OPT.

Bias estimation is carried out separately at each forecast lead time and for individual grid point with respective to ensemble mean. Bias correction is applied on all ensemble member forecasts and for 00Z initial cycle only. The data studied are 500hPa geopotential height and 850 mb temperature (period from 01 March 2004 to 28 February 2005). Other calibrated variables include the 2m temperature, 10m U and V component from the operational ensemble (not shown). The NCEP operational analysis is used for the bias estimation and the verification calculation. All the ensemble forecasts and analysis are on grid point with a resolution 2.5 by 2.5 globally.

After the operational and reforecast ensembles have been calibrated for each day of year 2004 by removing their bias estimates, several probabilistic and traditional evaluation methods are used to evaluate the ensemble forecast accuracy such as ranked probability skill score (RPSS, Toth et al. 2000), excessive outlier (Toth et al. 2003), root mean square error of ensemble mean (RMS), relative operating characteristics (ROC, Zhu et al. 2002) skill score and etc. In this paper, selected results are shown with considerations focused on interpretation of the results. More results can be found from http://www.emc.ncep.noaa.gov/gmb/wx20cb/Bias_Correction_Algorithm/1st_2nd_Moments/Training_1month/Plot_Scores_Second/).

4. Results

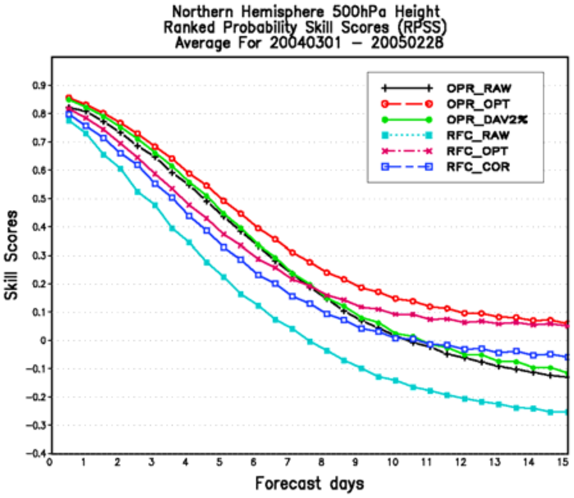

Figure 1. Annual mean of ranked probability skill score (RPSS) of northern hemisphere 500 hPa geopotential height from March 1, 2004 to February 28, 2005.

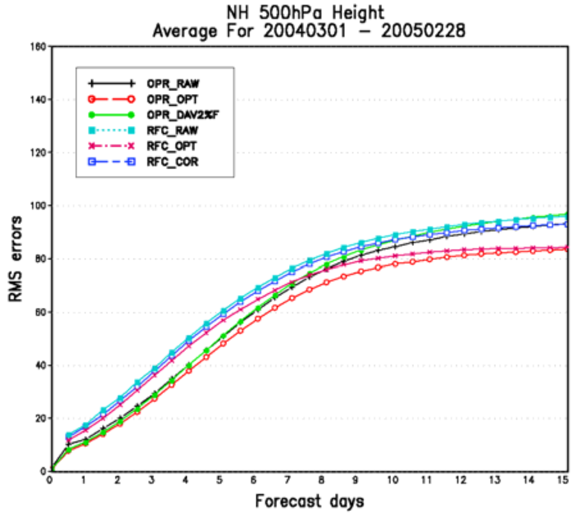

Figure 2. Annual mean of root mean square (RMS) errors for ensemble mean of northern hemisphere 500 hPa geopotential height from March 1, 2004 to February 28, 2005.

Figure 1 shows the annual mean of the RPSS scores of the 500 hPa geopotential heights, verified over the Northern Hemisphere. For the three operational ensembles, the OPR_OPT gets the highest RPSS scores among the six curves. The decaying average bias correction algorithm also works well. The RPSS of the OPR_DAV2% is improved versus the OPR_RAW for all lead time, especially for the short range, judged from the small distance between the two curves of the OPR_DAV2% and OPR_OPT.

For the three reforecast ensembles, it is not surprising to notice that the RFC_OPT shows the best performance compared with the RFC_RAW and the RFC_COR. The RFC_COR gains significant RPSS improvement versus the RFC_RAW for week 2 forecasting. Using the climatological mean bias estimate, it is possible to make probabilistic week 2 forecasts more skillful than the raw reforecast. Though the reforecast use old version model and relative poor quality initial data than the operational ensemble (Figure 2), the RFC_COR has even better performance than the OPR_RAW and the OPR_DAV2% after day 10, indicating the effective of the large data sample for improving week 2 forecasting.

Figure 2 shows the annual mean of the RMS error for 500 mb height ensemble mean forecast. The six curves coming from the six ensembles are divided into two clusters, which belong to the operational and reforecast ensemble forecast groups, respectively. For the three operational ensembles, the OPR_OPT has the lowest RMS error among the six ensembles. The OPR_DAV2% also has reduced RMS errors for the first week compared with the OPR_RAW but its RMS become larger for week 2 forecasting. However, the two close curves of the OPR_OPT and OPR_DAV2% for the first week suggest that there is only a limited opportunity for future improvement in bias correction for the first few days. The big distance between the OPR_OPT and OPR_DAV2% for week 2 indicates that the OPR_DAV2% calibration technique doesn't work very well for extended forecasting.

For the three reforecast ensembles, the RFC_COR has smaller RMS values than the RFC_RAW for all lead times even for week 2. A comparison between the operational and the reforecast ensembles shows that the OPR_RAW has much lower RMS error than the RFC_RAW. The RFC_RAW has around 50% larger short-range error than the OPR_RAW. Though the reforecast starts to run from the relative poor quality initial data than the operational ensemble, the RFC_COR works for short range and extended forecasting and its reduced RMS values becomes close to the OPR_RAW after day 10.

Figure 3 is the maps of the excessive outliers of the 500 hPa geopotential heights. The OPR_DAV2% has smaller values for up to 5-7 days with respective to the OPR_RAW indicating improved performance. The RFC_COR also displays much lower values (significant improvement) for all lead time versus the

RFC_RAW. Note that RFC_COR values become more near zero than the OPR_DAV2% after day 10.

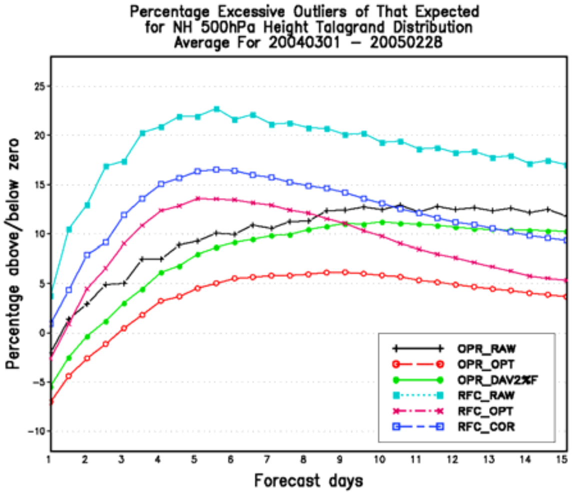

Figure 3. Annual mean of excessive outliers of northern hemisphere 500 hPa geopotential height from March 1, 2004 to February 28, 2005.

In addition to the 500 hPa height, several other variables such as the 850 hPa temperature, 2m temperature, 10m U and V component are also examined (not shown). Some tentative conclusions are obtained. The decaying averaging with 2% weight and 45-day operational training data works very well for short range (almost as good as "optimal), which makes its application for the frequent updates of DA/NWP modeling system possible. On the other hand, the climatological mean bias correction can add value, especially for week 2 probability forecasts. However, the generation of large hind-cast ensemble is expensive but may be helpful. Use of up-to-date data assimilation/NWP techniques is imperative at all ranges.

5. Summary

Statistical post-processing algorithms are being developed jointly by the NCEP and the MSC for eliminating the bias from the NWS and the MSC ensemble forecasts. Preliminary results show that an adaptive, regime dependent bias correction method works well for the first few days. The calibrated NCEP operational ensemble after removing the time mean forecast errors for the most recent period has improved probabilistic performance for all measures and for all lead time. The reforecast ensembles from the CDC with and without climate mean bias correction are also examined. A comparison between the operational and CDC bias-corrected ensemble forecasts shows that climate mean bias correction can add value, especially for week 2 probability forecasts. In the future, the methods developed for bias correction at the NWS and the MSC will be compared, and the best performing methods will be selected for use at both centers within the NAEFS system. The new method will also be tested in the context of the Bayesian Model Averaging algorithm in development at MSC to try to improve upon the associated simple linear bias correction scheme. New bias correction methods developed under the Observing-system Research and predictability experiment (THORPEX) Interactive Grand Global Ensemble (TIGGE) project will also be considered for use in the NAEFS system.

Acknowledgments. We gratefully acknowledge the support of Jeff Whitaker (CDC, NOAA), Tom Hamill (CDC, NOAA), Richard Verret (CMC), Richard Wobus (NCEP). This study would have been much more difficult without their assistance.

References

Hamill, T.M., J.S. Whitaker and X. Wei, 2003: Ensemble re-forecasting: improving medium-range forecast skill using retrospective forecasts. Mon. Wea. Rev., submitted.

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

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

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

Toth, Z., cited 2000: The NCEP global ensemble forecasting system. [Available online at http://www.oneonta.edu/academics/wxclub/JBpages/M361/EnsembleOverview.pdf]

Toth, Z., J. Schaake, Y. Zhu, and J. Du, 2003: Postprocessing ensemble forecasts for hydrological applications, AHPS proposal.

Toth, Z., O. Talagrand, G. Candille, and Y. Zhu, 2003: Probability and ensemble forecasts. Book of: Forecast Verification: A practitioner's guide in atmospheric science. Ed.: I. T. Jolliffe and D. B. Stephenson. Wiley, 137-163.

Zhu, Y., Z. Toth, R. Wobus, D. Richardson, and K. Mylne, 2002: The economic value of ensemble based weather forecasts. Bulletin of the American Meteorological Society, 83, 73-83.