Science and Technology Infusion Climate Bulletin
NOAA's National Weather Service
40th NOAA Annual Climate Diagnostics and Prediction Workshop
Denver, CO, 26-29 October 2015

New Measure of Forecast Uncertainty for the North American Multi-Model Ensemble

Qin Zhang[1], Yuejian Zhu[2], Hong Guan[2,5] , Jon Gottschalck[1], Jin Haung[1],
Huug van den Dool[1], Emily Becker[1,3], and Li-Chuan Chen[1,4]

[1]*Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD*
[2]*Environmental Modeling Center, NCEP/NWS/NOAA, College Park, MD*
[3]*Innovim, LLC., Greenbelt, MD*
[4]*Earth System Science Interdisciplinary Center/Cooperative Institute for Climate and Satellites, University of Maryland, College Park, MD*
[5]*System Research Group Inc, Colorado Springs, CO*

ABSTRACT

Since August 2011, realtime monthly and seasonal forecasts from the North American Multi-Model Ensemble (NMME) have been made every month by the NCEP Climate Prediction Center (CPC). Among the most popular NMME products, NMME ensemble mean maps are made from the equally weighted average of the participating models' ensemble means, after removing systematic errors. However, some users are interested in how the models are different – that is, the diversity of the forecasts. In this study, we defined a normalized spread (*SPRnor*) to measure NMME forecast uncertainty, which is calculated from the multi-model predictive variance (including between-model variance and within-model variance) and then normalized by the observed standard deviation. When *SPRnor* is smaller than 1, it indicates the NMME forecast has less uncertainty, since the models are in good agreement over the grid point. When *SPRnor* is greater than 1, it means that the NMME forecast uncertainty is larger than the observed inter-annual variability, as the model forecasts are more dispersed. Generally, the *SPRnor* grows with the forecast leading time, and also varies with season. Therefore, we supply normalized spread maps to complement the NMME ensemble mean forecast and give users additional information of NMME forecast uncertainty in realtime.

1. Introduction

More and more users have gone to the North American Multi-Model Ensemble web page (http://www.cpc.ncep.noaa.gov/products/NMME/) to view NMME products for their operational missions and applications since the first NMME seasonal and monthly prediction was made in August 2011 (Kirtman *et al.* 2014). Among the thousands of uploaded figures of realtime prediction for both North American and global domains, the most popular products are the NMME mean 2m temperature and precipitation anomalies made by the equally weighted average of each NMME model's ensemble mean, after removing systematic errors. However, the information from the NMME ensemble mean anomalies is not enough, since it is akin to a deterministic forecast. Users are also interested in how the forecasts for each model differ and the confidence of the NMME prediction. While the NMME probability forecasts, calculated from all ensemble members with equal weights, have been made each month (Becker *et al.* 2014) since 2012, their weights are not completely consistent with the maps of NMME anomalies. Therefore, the motivation of this work is to define and develop new products to express the prediction uncertainty of NMME and indicate the model forecast diversity.

Correspondence to: Qin Zhang, 5830 University Research Court, Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland; E-mail: qin.zhang@noaa.gov.

2. Definition of the spread for multi-model ensemble

The NMME is a dynamic multi-model ensemble forecast system, initially comprised of 6 US models (CFSv1 & CFSv2/NCEP, ECHAM-a & ECHAM-f/IRI, NCAR-CCSM3/COLAR-UM, GFDL-CM2.1, and GEOS5/NASA). For the past two years, the NMME has included seven models: two Canadian models (Can-CM3&4) (Environment Center of Canada joined in August 2012, when CFSv1 was retired), two models from GFDL, GEOS5/NASA, CFSv2/NCEP, and NCAR-CCSM4 (which replaced NCAR-CCSM3). All of the NMME models are atmosphere-ocean coupled, and the horizontal resolution of the exchanged variables is 1x1 degree, consistent with the retrospective forecasts from 1982 to 2010. The NMME model climatologies are calculated from the 29 years of retrospective forecasts to remove systematic bias in the mean at each leading forecast time for each model before calculating the NMME ensemble mean. The model's prediction skills (as expressed by the anomaly correlation) are also obtained from these retrospective forecasts.

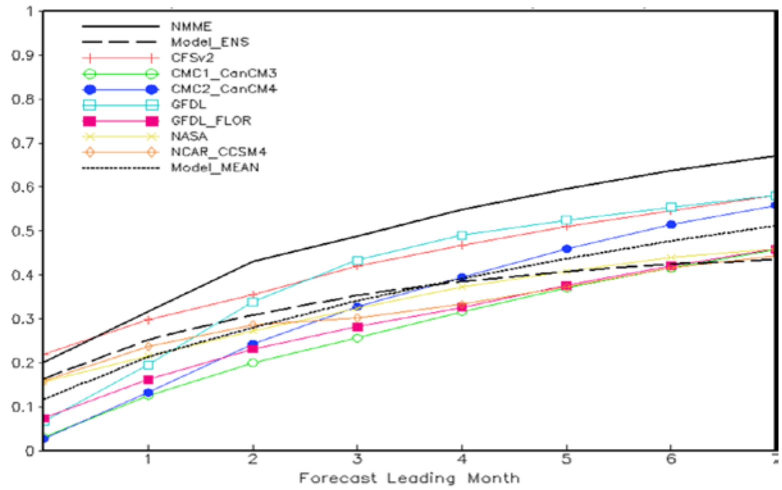

Fig. 1 Spread (black solid line) of Nino3.4 for NMME hindcasts (1982-2010) and the ensemble mean spread (*SPRensm*, dashed line) and the members of spread (*SPRmemb*, dotted line) with the spreads of individual model (color lines).

We define the multi-model ensemble predictive variance in space (s) and time (t), lead (τ) and IC month (m) for an anomalous field, according to Raftery (1993):

$$VAR = \frac{1}{K}\sum_{k=1}^{K}(F_k' - F')^2 + \frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{N}\sum_{n=1}^{N}(f_{nk}' - F_k')^2\right) \quad (1)$$

where F_k' is the k^{th} model ensemble mean anomaly after mean bias correction. F' is the equal weight averaged NMME ensemble mean for (K=7) models and f_{nk}' is the anomaly of each member for each model, as N is the number of the members for each model. *VAR* is a function of space, time, forecast lead and either the start month or the target month.

Here the predictive variance should be the sum of the two terms. One (the first term) is the between-model variance, another one (the second term) is the within-model variance. The between-model variance is the distance of the 7 individual model ensemble means from the multi-model ensemble mean, and the within-model variance is the average distance of each model member from its model's ensemble mean. (Raftery, *et al.* 2005).

The spread of NMME also consists of two terms:

$$SPR^2 = VAR = SPR^2\,ensm + SPR^2\,memb \quad (2)$$

where the first term represents the diversity of the models' ensemble means relative to the forecast signal (we call it ensemble mean spread). The second term is the spread of the individual members relative to their models' ensemble means, which is linked to the forecast noise (hereafter called member spread).

$$SPR^2 ensm = \frac{1}{K}\sum_{k=1}^{K}(F_k' - F')^2 \quad (3)$$

$$SPR^2 memb = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{N}\sum_{n=1}^{N}(f_{nk}' - F_k')^2\right) \quad (4)$$

We also normalized the NMME spread to eliminate spatial and seasonal variation. Normalized multi-model ensemble mean spread indicates the uncertainty of the NMME ensemble mean prediction or the model forecast diversity. It is also known as the "envelope of solutions" for each lead forecast time. We define the normalized spread as a ratio of the root mean square of predictive variance to the observed standard deviation (STD), that is,

$$SPRnor = (VAR)^{1/2} / STD_{obs} \quad (5)$$

When *SPRnor* is smaller than 1, it indicates the NMME forecast has less uncertainty since the models are in good agreement over the grid point. When *SPRnor* is greater than 1, it means that the NMME forecast has more uncertainty than observed inter-annual variability due to the greater dispersion of model forecasts.

3. Results

a. The relationship of spread and the forecast uncertainty

Among the most popular NMME figures are the Nino3.4 plumes (Barnston *et al.* 2015). These show 7-month lead Nino3.4 index forecasts for the individual members and the ensemble mean of each model, as well as the equally weighted NMME ensemble mean (see http://www.cpc.ncep.noaa.gov/products/NMME/current/plume.html). The Nino3.4 plumes also show the forecast uncertainty visually: the denser the member distribution the higher the prediction probability. Here we describe the relationship of the NINO3.4 spread and forecast uncertainty by using 29 years of NMME retrospective forecasts as an example.

Figure 1 shows the *SPR* of NMME Nino3.4 index (black solid line) and its two component terms, the NMME ensemble mean spread (*SPRensm*, labeled Model_ENS, dashed line) and the spread of all members (*SPRmemb*, labeled Model_MEAN, dotted line) with the individual models' spread (colored lines). All of these quantities grow with forecast lead time. However, the NMME ensemble mean spread (*SPRensm*) reaches saturation after 4 lead months and increases slowly after. The spread of NMME is bigger than that of any individual model, indicating that the ensemble mean of NMME covers all members and have a wide PDF for the all kind of predictability from the individual model.

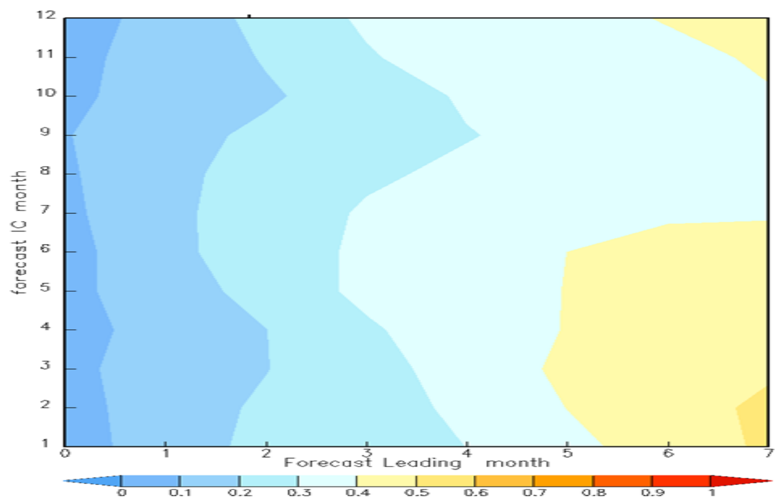

Fig. 2 Spread of NMME calculated for each month from hindcasts (1982-2010).

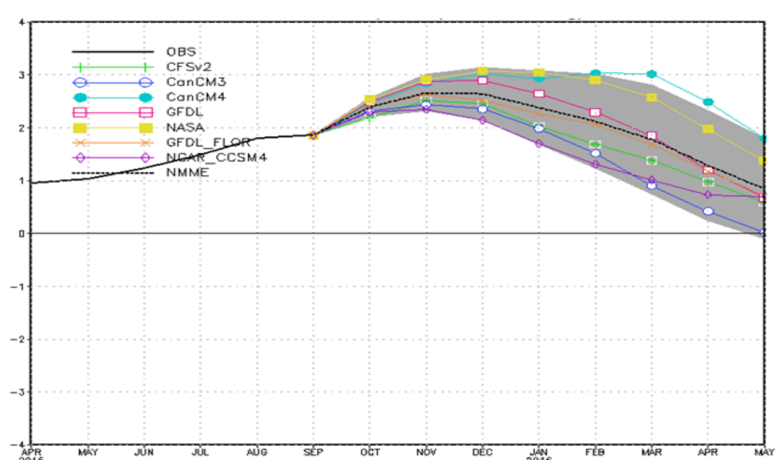

Fig. 3 Realtime forecast of Nino3.4 plumes with the NMME spread in shading for Oct. 2015 IC.

From figure 2, we find that the spread of NMME not only increases with forecast lead time but also varies with the season of initial forecast time (IC). The biggest peak of forecast uncertainty is for spring initial conditions, corresponding to the well- known “spring barrier” of ENSO prediction. The smallest spread (highest forecast confidence) is for forecasts made in the fall (September), when ENSO predictive probability is high, especially within 4 months lead.

Figure 3 shows the realtime forecast of Nino3.4 plumes, with NMME spread in shading, for October 2015 initial conditions. It is easy to see that the spread is consistent with the diversity of the ensemble mean of NMME models on the occasion.

b. Normalized NMME spread

Since the spread varies spatially and temporally, the multi-model spread calculated from formula (2) is hard to compare to the model forecast diversity in a different location or time. For the maps of NMME prediction, we normalize the spread by formula (5) to extract the information of NMME forecast diversity. Figs. 4 and 5 show NMME realtime prediction of 2m temperature and precipitation anomalies (contours) with normalized spread (shading) for North America for October 2015 initial conditions.

The NMME predicts warmer-than-average temperatures over the western half of North America, partially influenced by El Niño developing in the fall of 2015. Forecasts from the NMME models are more consistent in this region than in the south-eastern CONUS, where the forecast has higher uncertainty, shown by the models’ prediction diversity. On the other hand, the forecast for positive precipitation anomalies over the eastern CONUS has less uncertainty than that over the western US (Fig. 5). The normalized spread gives users information about how NMME model forecasts differ, or the diversity in the predictions.

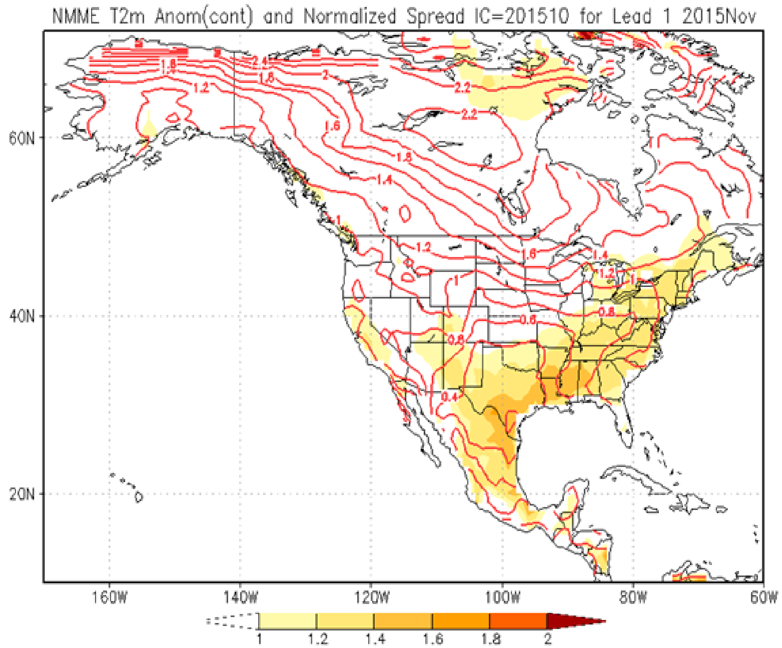

Fig. 4 NMME realtime prediction of 2m temperature anomalies (contours) with normalized spread (shading) of North America for October 2015 initial conditions.

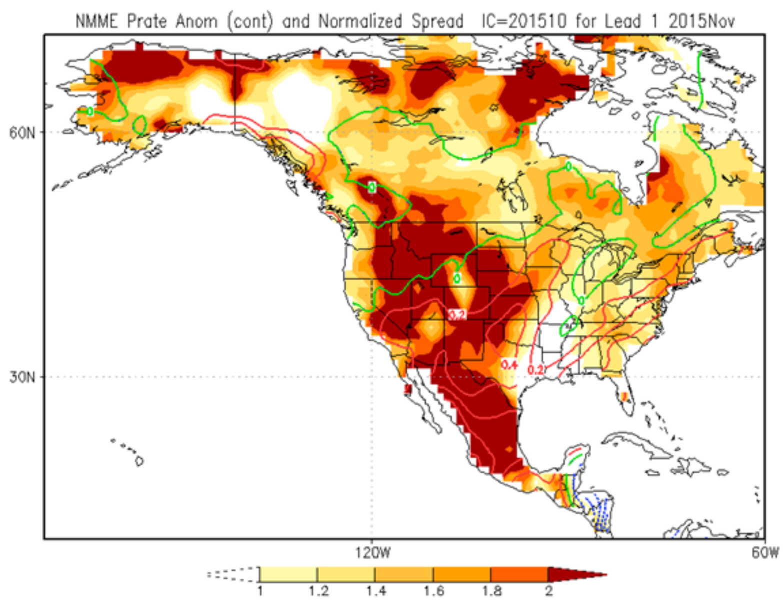

Fig. 5 NMME realtime prediction of precipitation anomalies (contours) with normalized spread (shading) of the North America for October 2015 initial conditions.

4. Summary and discussion

NMME realtime spread is defined as the multi-model ensemble predictive variance, including between-model variance and within-model variance. Normalized ensemble spread is a new measurement for NMME forecast uncertainty, consistent with the forecast of NMME ensemble mean anomalies. *SPRnor* ≤ 1 indicates the model forecasts are in good agreement over the grid points. *SPRnor* > 1 means that the model forecasts are more dispersed, and therefore have more uncertainty, than observed inter-annual variability. In generally, realtime *SPRnor* increases with forecast lead time. However, some variables, such as precipitation, may be influenced by seasonal variance in certain regions.

References

Barnston, A.G., M.K. Tippett, H.M. van den Dool, and D.A. Unger, 2015: Toward an improved multimodel ENSO prediction. *J. Appl. Meteor. Climatol.*, **54**, 1579–1595. doi: http://dx.doi.org/10.1175/JAMC-D-14-0188.1

Becker, E., H.M. van den Dool, Q. Zhang, 2014: Predictability and forecast skill in NMME. *J. Climate*, **27**, 5891–5906. doi: http://dx.doi.org/10.1175/JCLI-D-13-00597.1

Kirtman, B.P., and Coauthors, 2014: The North American Multi-Model Ensemble (NMME): Phase-1 seasonal to interannual prediction, phase-2 toward developing intra-seasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601. http://dx.doi.org/10.1175/BAMS-D-12-00050.1

Raftery, A.E., 1993: Bayesian model selection in structural equation models. *Testing Structural Equation Models,* K.A. Bollen and J.S. Long, Eds., 163–180.

Raftery, A.E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174.