Science and Technology Infusion Climate Bulletin
NOAA's National Weather Service
36th NOAA Annual Climate Diagnostics and Prediction Workshop
Fort Worth, TX, 3-6 October 2011

Preliminary Evaluation of Multi-Model Ensemble System for Monthly and Seasonal Prediction

Qin Zhang[1], Huug van den Dool[2], Suru Saha[3], Malaquias Pena Mendez[3],
Emily Becker[1], Peitao Peng[2] and Jin Huang[2]

[1]*Wyle Information System, Climate Prediction Center, NCEP/NWS/NOAA*
[2]*Climate Prediction Center, NCEP/NWS/NOAA*
[3]*Environment Modeling Center, NCEP/NWS/NOAA*

ABSTRACT

A collaborative prediction system, the National Muli-Model Ensemble (NMME), is under development through the NOAA Climate Test Bed (CTB) for experimental monthly and seasonal prediction at Climate Prediction Center (CPC). The CTB NMME project is funded by Climate Program Office MAPP Program (Modeling, Analysis, Prediction and Projection). In the current phase, seven models from different US institutes (NCEP-CFSv1, NCEP-CFSv2, GFDL-CM2.2, NCAR/U.Miami/COLA-CCSM3, NASA-GEOS5, IRI (ECHAM-a and ECHAM-f)) are participating. Three variables (monthly mean precipitation, sea surface temperature, and air temperature at 2 meters on a 1x1 degree grid), all with at least 29 years of hindcasts (1982-2010), are evaluated after removing their systematic errors, and then verified against the observations. Realtime experimental forecasts of the multi-model ensemble were first conducted in August 2011. The bias corrected multi-model ensemble prediction system is designed to contribute to the ongoing monthly and seasonal prediction in CPC.

1. Introduction

Monthly-to-seasonal time scale climate predictions are made at NCEP/CPC routinely by a number of tools, both statistical, *e.g.* canonical correlation analysis (CCA) and optimal climate normal (OCN) (O'Lenic *et al.* 2008), and dynamical models (Climate Forecast System version 1 &2). Although the ratio of signal to noise is low in the dynamical model due to long time integration and growth of the systematic error (Straus and Shukla 2002), the dynamical models have comparable forecast scores to the statistical models (DeWitt 2005; Saha *et al.* 2006). Successful monthly-to-seasonal prediction mostly depends on a revolution in our understanding of the coupled ocean-atmosphere system after the dramatic strong ENSO events in 1982/83 and 1997/98 (Barnston *et al.* 1999; Landsea and Knaff 2000; Shukla *et al.* 2009). This means monthly-to-seasonal predictability relies on the slowly evolving components of the climate system, like the ocean or land surface, that act as boundary conditions for the atmosphere with its shorter intrinsic time scales (Shukla *et al.* 2009; Goddard et al. 2001; Paolino *at al.* 2011). Two types of uncertainties are involved in the monthly-to-seasonal predictability: one is related to the uncertainty of the initial conditions (Keenlyside *et al.* 2005; Luo and Wood 2006), and the other is accounted for model errors in the physics processes related to the sub-grid parameterization (Palmer *at al.* 2004, Kirtman and Min 2009, De Witt 2005).

In recent years, the multimodel ensemble forecast has become a powerful tool for the monthly-to-seasonal time scale prediction to deal with both uncertainties (Krishnamurti *et al.* 2000, Kirtman *et al.* 2003; Peng *et al.* 2002; Hagedorn *et al.* 2005; Doblas-Reyes *et al.* 2005; Palmer *et al.* 2004; Lavers *et al.* 2009). For the monthly-to-seasonal forecast, multimodel prediction has successfully increased the spread by reducing the overconfident forecast of the individual model (Palmer *et al.* 2004; Weisheimer *et al.* 2009). Furthermore, recently research has shown that a multimodel ensemble, even in a simple equal weight combination, has

Correspondence to: Qin Zhang, Wyle Information System, Climate Prediction Center, NCEP/NWS/NOAA, Camp Springs, MD 20746; E-mail: Qin.Zhang@noaa.gov.

higher prediction skill scores than that of any individual model in the prediction of tropical SST anomaly (Kirtman and Min 2009). Several projects, like DEMETER and EUROSIP, have demonstrated the improvement of multimodel seasonal forecast reliability (Hagedorn *et al.* 2005, Mitchell *et al.* 2004, Palmer *et al.* 2004).

This work of reports the latest results from the experimental National Multi-Model Ensemble (NMME) prediction system (Phase1). The models and data involved in Phase 1are briefly described in section 2, and followed by a preliminary evaluation of the multimodel ensemble system for monthly and seasonal prediction. In the last section, we present a summary and discussion.

Model	Period	Members	Leads	Arrangement of Members
CFSv1	1981-2009	15	0-8 months	1st 0Z +/-2days, 21st0Z+/-2d, 11th0Z+/-2d
CFSv2	1982-2009	24(28)	0-9	4 members (0,6,12,18Z) every 5th day
GFDL-CM2.2	1982-2010	10	0-11	All 1st of the month 0Z
IRI-Echam4-f	1982-2010	12	0-7	All 1st of the month
IRI-Echam4-a	1982-2010	12	0-7	All 1st of the month
CCSM3.0	1982-2010	6	0-11	All 1st of the month
NASA-GEOS5	1982-2010	6 (8)	0-8	1 member every 5th day Additional 2 members on the beginning of month

Table 1 NMME models information

2. Models and data

Based on two Climate Test Bed (CTB) workshops (February 18 and April 8, 2011), a collaborative and coordinated implementation has been established under the frame work of the CTB project, called National Multi-Model Ensemble (NMME). The experimental realtime experimental forecast system made a first seasonal and monthly multimodel forecast in August 2011 in Climate Prediction Center as Phase 1 of the NMME project. Seven models, from NCEP (CFSv1&2), GFDL (CM2.2), IRI (ECHAM-a and ECHAM-f), NCAR/U.Miami/COLA-CCSM3 (Collins *et al.* 2006), and NASA-GEOS5 are participating. Three variables (monthly mean precipitation, sea surface temperature, and air temperature at 2 meters on a 1X1 degree grid) with 29 years of hindcasts (1982-2010), have been evaluated. The model climatology and prediction skill mask have been calculated after the systemic errors are corrected for each model in every leading month forecasts. More details of the NMME models are given in Table1.

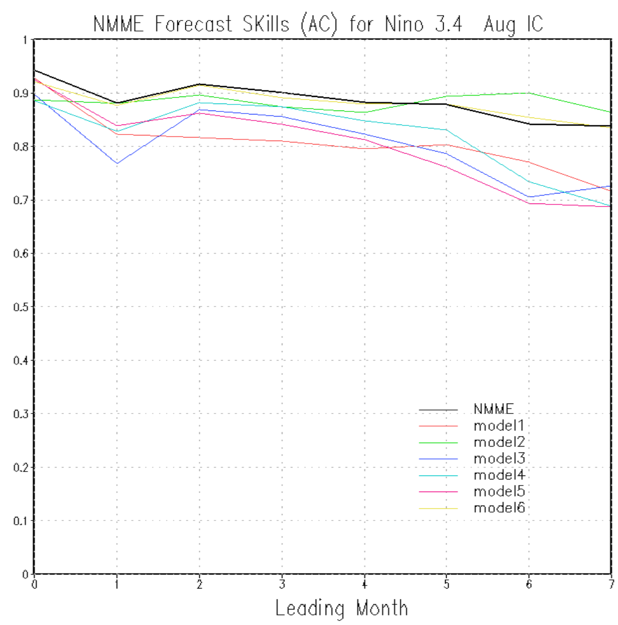

Fig. 1 Multimodel ensemble forecast skills of Nino3.4 (black line) and individual modes (color lines) for 7 lead months with August initial conditions.

The observation data used for verification are the NOAA Optimum Interpolation (OI) Sea Surface Temperature (SST) V2 (SST OISST-QD) (1982-2010) (Reynolds *et al.* 2002) for verification of model SST, CMAP (1982-2010) (Xie and Arkin 1997) for precipitation, and GHCN_CAMS (1982-2010) for temperature at 2m (Fan and van den Dool 2006).

3. Results

The first experimental realtime monthly-to-seasonal forecast of NMME-PHASE1 was made in August 2011. In the month before, hindcasts from 1982-2010 with August ICs for all models, except NASA's, were collected by FTP for all 3 variables: SST, precipitation and temperature at 2m. The evaluation of prediction skills for each model and equal-weight average of multimodel ensemble mean were done after calculating the model climatology for the systematic bias correction. Here, we are reporting some results related to the NMME prediction skill assessment for the subsequent realtime forecast in August.

a. Nino3.4 and SST

Since SSTs in the tropical Pacific are a major source of climate predictability on monthly-to-seasonal time scales, model performance in the tropical Pacific is of particular interest. To demonstrate the typical level of skill in this area, Fig. 1 shows the anomaly correlation coefficient (ACC) of the ensemble mean for the single model ensemble (colored lines) and multimodel ensemble (black line) for the SST anomaly averaged over the Nino3.4 area for each lead in months. The results suggest that all single-model ensembles generally perform as well as El Nino-Southern Oscillation (ENSO) prediction systems. All single models have achieved above 0.7 for 7 months forecast, but the NMME multimodel ensemble system has an ACC above 0.83 for all the 7 months. In addition, note the higher correlation of the multimodel ensemble compared to all single models, except for the last two leads. This indicates the multimodel ensemble indeed has more skill than single model, as pointed out

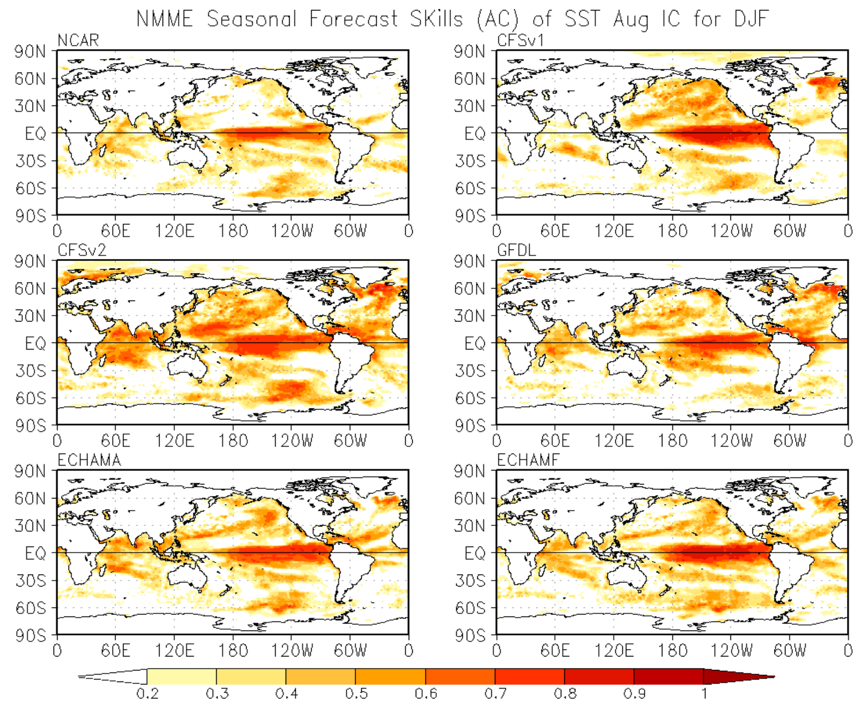

Fig. 2a Maps of anomaly correlation coefficient (ACC) of SST for individual models for prediction DJF with August ICs.

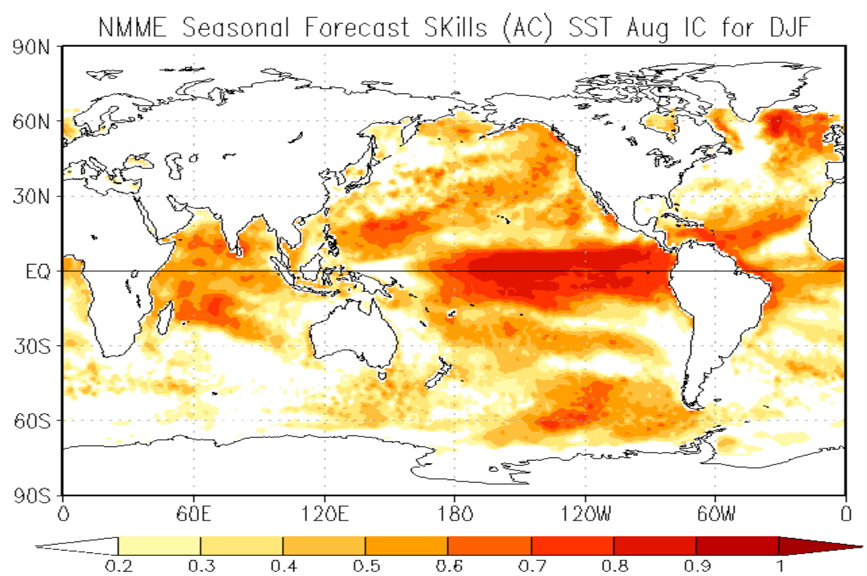

Fig. 2b SST anomaly correlation coefficient (ACC) of multimodel ensemble with observation (1982-2010) for prediction DJF with August ICs.

by the previous studies (Palmer *et al.* 2004, Kertman and Min 2009). This improvement of the Nino3.4 prediction skill is greatly encouraging and has shown the value of the NMME efforts.

To further investigate the multimodel SST hindcast skill, maps of grid point ACC for the target season DJF with August initial conditions are shown for the six individual models (Fig. 2a) and the multimodel ensemble (Fig. 2b). Unsurprisingly, the ACCs of the SST anomaly for both the single models and the multimodel ensemble mean have higher scores over the central eastern Pacific, exceeding 0.8, and then gradually decrease along the equator westward and off the equator in the western tropical Pacific. It is interesting to see that while some models have low skills over the northern west Pacific, the NMME ensemble has skills comparable to the best performing model over the regions. Many areas like this can be found over the tropical Indian Ocean and northern Atlantic. In the extra-tropics, the correlations are generally low, but there are same notable high correlations (*e.g.* greater than 0.6) in the north Atlantic and South Pacific.

Carefully comparing each map in Fig. 2a and Fig. 2b, one may find that the ACC of the multimodel ensemble seems to take over skill scores wherever they are better among the single models. This hypothesis is confirmed by averaging the prediction skill over the global tropical band (from 30ºS-30ºN) shown in Fig. 3. The ACC of the multimodel is the second highest for the first 3 leads and the best one after that for longer lead month forecast. The root square mean error (RSME) is almost the lowest for all the leading months except the first month among all models. Consistent with DEMETER (Hagedorn *et al.* 2005), the NMME multimodel ensemble improvement in SST prediction achieved by the error compensation each other in individual models (Kirtman and Min 2009).

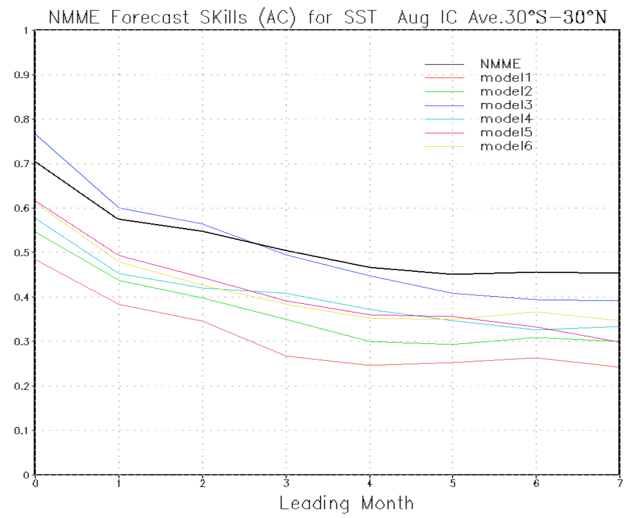

Fig. 3a NMME forecast skills (black line) and individual models (color lines) for SST averaged 30S-30N with August ICs.

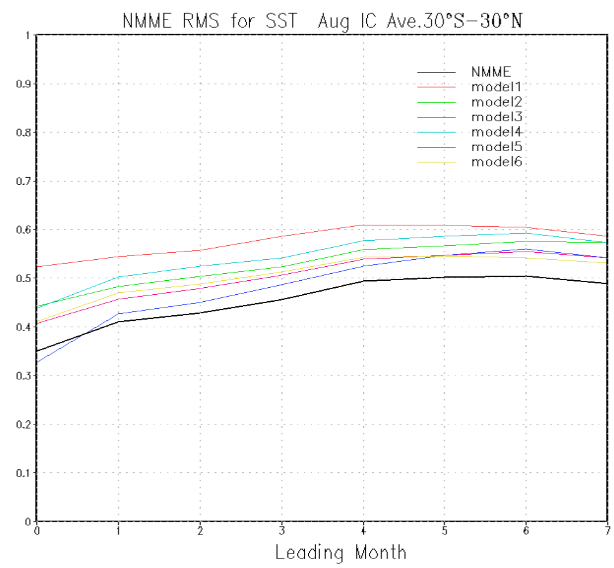

Fig. 3b NMME RMS error (black line) and individual models (color lines) for SST averaged 30S-30N with August ICs.

b. Precipitation

Corresponding to the high scores over the central eastern Pacific in SST anomaly prediction, the ACC of precipitation for each model and multimodel ensemble have the same narrow band of high forecast skills (Fig. 4) for DJF with August initial conditions. Scores quickly decrease westward and off the equator for the individual models and the multimodel ensemble. Only isolated scattered high score areas can be found in the extra-tropics. The ACC over the land is also low for the both individual models and multimodel ensemble. ACC scores averaged over the global tropical band are lower than 0.3 for all the leading months (Fig. 5a). However, the multimodel ensemble has the highest scores, and RSME is much smaller than for any of the single models (Fig. 5b). The superior performance of the multimodel ensemble for the prediction of precipitation indicates the benefit of the NMME approach even with a simple equal-weighting ensemble. This achievement especially helps to improve the CPC's realtime monthly-to-seasonal forecast, since currently precipitation prediction scores are low on land and have large uncertainty.

c. Temperature at 2m

Now we explore the forecast skills of the atmospheric temperature at 2m for the northern hemisphere winter (DJF) in 4 months leading prediction. The strongest signals of the ACC are mainly over the tropical regions related to ENSO impacts on the monthly-to-seasonal time scales. Good scores can be found over South America and Africa in each individual model, indicating that state-of-the-art coupled models have caught the SST forcing and the right response to the forcing of the boundary conditions of ENSO (Fig. 6a). The multimodel ensemble has superior scores for these near equatorial regions, such as the east and west coasts of Australia and Sumatra in the western Indian Ocean and western equatorial Pacific (Fig. 6b). The prediction scores for temperature at 2m averaged over the tropical band (30ºS-30ºN, land only) for each lead show the multimodel ensemble is the second best compared to all models (Fig. 7a), and RMSE is almost as good as the best individual model for the all lead (Fig. 7b). We are very encouraged by these preliminary results for the multimodel ensemble forecast assessment. The prediction scores suggest the NMME will help improve the CPC realtime monthly-to-seasonal prediction, and will provide improved climate forecast to decision makers and downscaling and other user communities.

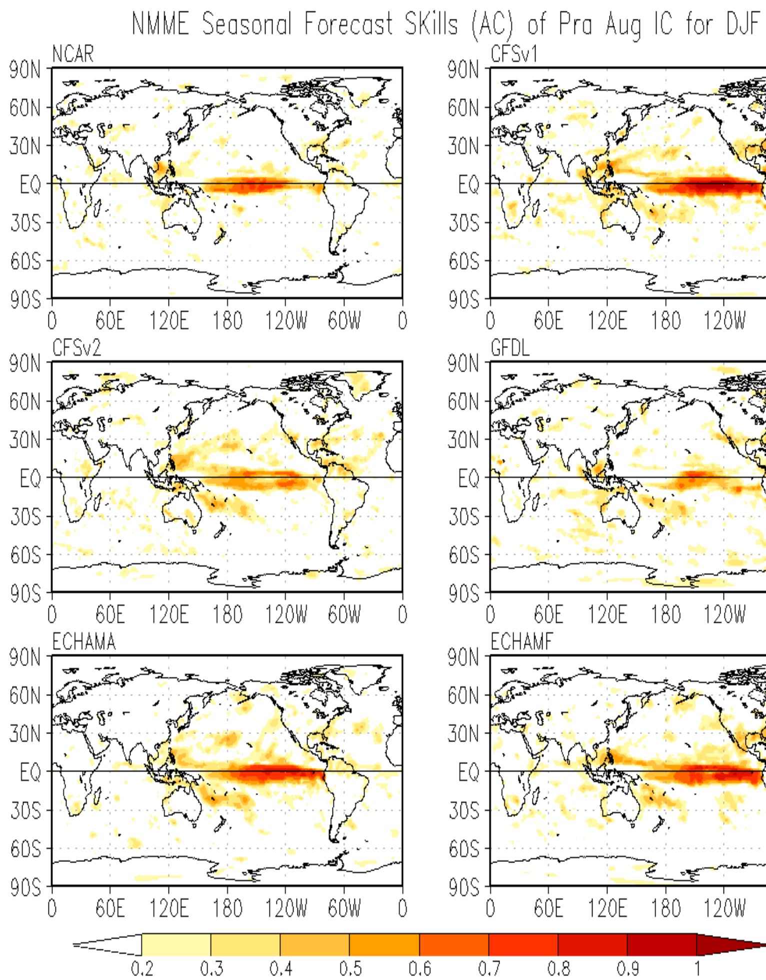

Fig. 4a Maps of anomaly correlation coefficient (ACC) of precipitation for individual models for prediction DJF with August ICs.

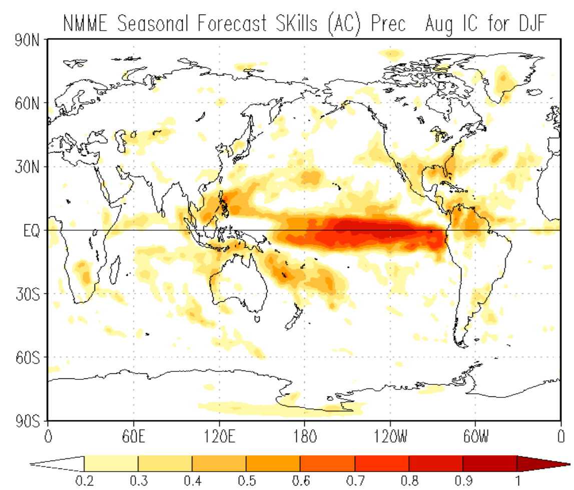

Fig. 4b Maps of anomaly correlation coefficient (ACC) of precipitation for multimodel ensemble for prediction DJF with August ICs.

4. Summary and discussion

Based on existing state-of-the-art US climate prediction models from these institutes (NCEP-CFSv1, NCEP-CFSv2, GFDL-CM2.2, NCAR/U.Miami/COLA-CCSM3, NASA-GEOS5 and IRI-ECHAM4), the Climate Test Bed has launched a phase-1 of National Multi-Model Ensemble project in February, 2011. The monthly-to-seasonal multimodel prediction system is under development and

experimental realtime forecasts have been made on the 8th of each month routinely since August, 2011 to adhere to the schedule of the operational monthly and seasonal forecast in NCEP Climate Prediction Center. Maps of the NMME experimental prediction can be found on the web site:
http://www.cpc.ncep.noaa.gov/products/people/wd51yf/NMME. The graphical forecast guidance includes North America and the global domain of precipitation and temperature at 2m anomalies and SST anomaly. The plots are monthly and seasonal means, with or without skill mask, applied for 7 lead months or 5 lead seasons (3-month averages).

All NMME forecast are bias corrected by using the 29 years of hindcast data for each participating model. The model climatology and skill mask are calculated so as to apply to the realtime forecast in each month. The assessment of prediction scores of the three fields for both individual model and NMME ensemble are also given on the web for information about the confidence and reliability of the prediction. This report only discusses the preliminary evaluation of the NMME prediction system for forecast precipitation, temperature at 2m and SST for 2011/12 winter (DJF) with August initial conditions. More detailed information on forecasts and verifications on other months can be found at the web site mentioned above.

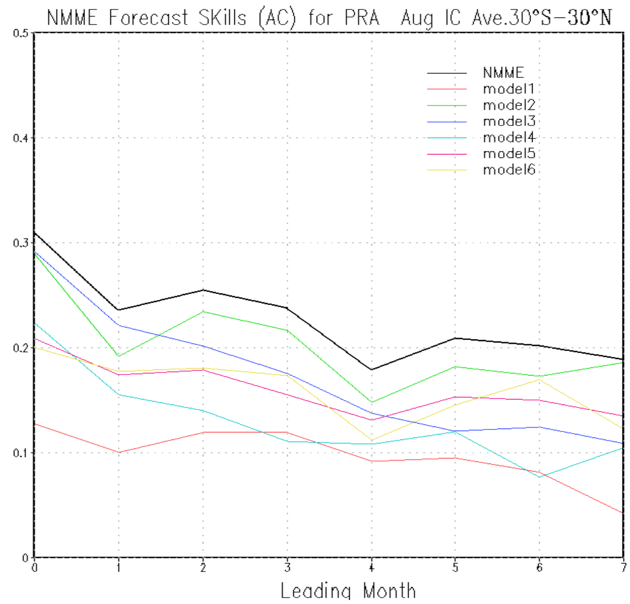

Fig. 5a NMME forecast skills (black line) and individual models (color lines) for precipitation averaged 30S-30N with August ICs.

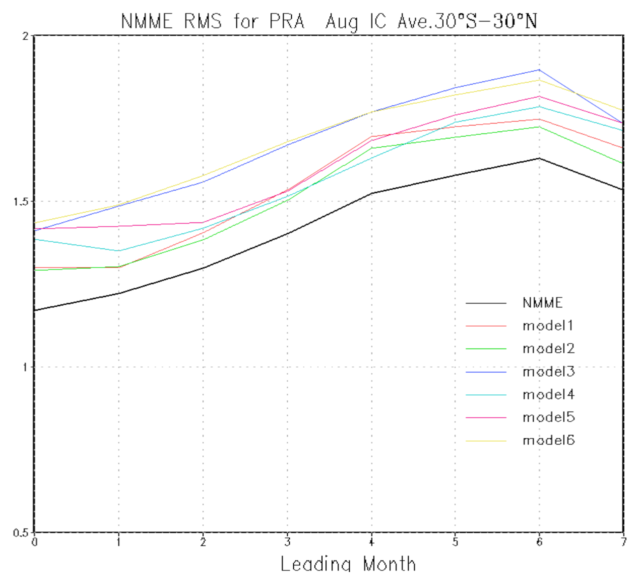

Fig. 5b NMME RMS error (black line) and individual models (color lines) for precipitation averaged 30S-30N with August ICs.

References

Barnston, A. G., M. Glantz, and Y. He, 1999: Predictive skill of statistical and dynamical climate models in SST forecasts during the 1997–98 El Nino and the 1998 La Nina onset. *Bull. Amer. Meteor. Soc.*, **80,** 217–243.

Collins, W. D., and Coauthors, 2006: The Community Climate System Model version 3 (CCSM3). *J. Climate*, **19,** 2122-2143.

DeWitt, D. G., 2005: Retrospective forecasts of interannual sea surface temperature anomalies from 1982 to present using a directly coupled atmosphere-ocean general circulation model. *Mon. Wea. Rev.*, **133**, 2972-2995.

Doblas-Reyes, F. J., R. Hagedorn, T. N. Palmer, 2005: The rationale behind the success of multimodel ensembles in seasonal forecasting - II. Calibration and combination *Tellus A* **57**, 234–252. doi:10.1111/j.1600-0870.2005.00104.x

Hagedorn, R. F. J. Doblas-Reyes, T. N. Palmer, 2005: The rationale behind the success of multimodel ensembles in seasonal forecasting - I. Basic concept *Tellus A*, **57**, 219–233. doi:10.1111/j.1600-0870.2005.00103.x

Goddard, L., and Coauthors, 2001: Current approaches to seasonal-to-interannual climate predictions. *Int. J. Climatol.*, **21,** 1111–1152.

Keenlyside, N.S., M. Latif, M. Botzet, J. Jungclaus and U. Schulzweida, 2005: A coupled method for initializing El Niño Southern Oscillation forecasts using sea surface temperature. Tellus, 57A, 340-356.

Kirtman, B. P., 2003: The COLA anomaly coupled model: Ensemble ENSO prediction. Mon.Wea. Rev., 131, 2324-2341.

Kirtman, B. P., and D. Min, 2009: Multi-model ensemble ENSO prediction with CCSM and CFS. *Mon. Wea. Rev.*, DOI: 10.1175/2009MWR2672.1.

Krishnamurti, T.N., and Coauthors, 2000: Multi-Model ensemble forecasts for weather and seasonal climate. *J. Climate*, **13,** 4196-4216.

Landsea, C. W., and J. A. Knaff, 2000: How much skill was there in forecasting the very strong 1997–98 El Nino? *Bull. Amer. Meteor. Soc.*, **81,** 2107–2120.

Lavers, D., L. Luo, and E. F. Wood, 2009: A multiple model assessment of seasonal climate forecast skill for applications. *Geophys. Res. Lett.*, **36**, L23711, doi:10.1029/2009GL041365.

Luo, L., and E. F. Wood, 2006: Assessing the idealized predictability of precipitation and temperature in the NCEP Climate Forecast System. *Geophys. Res. Lett.*, **33**, L04708, doi:10.1029/2005GL025292.

Mitchell, K.E., and Coauthors, 2004: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *J. Geophys. Res.* **109**(D7). doi:10.1029/2003JD003823.

O'Lenic, Edward A., David A. Unger, Michael S. Halpert, Kenneth S. Pelman, 2008: Developments in Operational

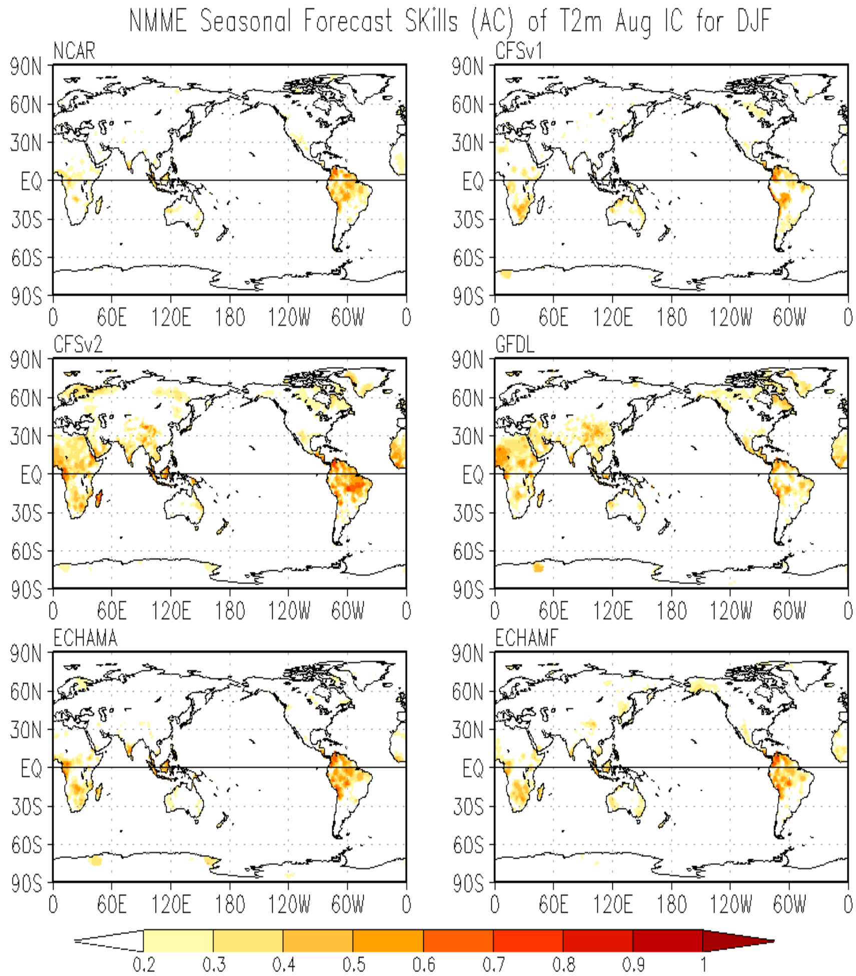

Fig. 6a Maps of anomaly correlation coefficient (ACC) of temperature at 2m for individual models for prediction DJF with August ICs.

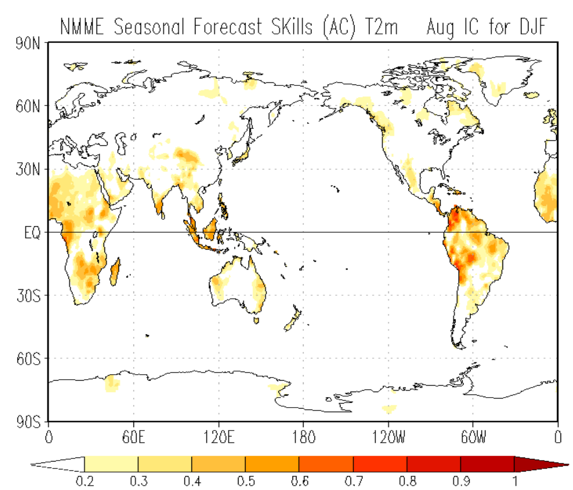

Fig. 6b Maps of anomaly correlation coefficient (ACC) of temperature at 2m for multimodel ensemble for prediction DJF with August ICs.

Long-Range Climate Prediction at CPC. *Wea. Forecasting*, **23**, 496–515. doi: 10.1175/2007WAF2007042.1

Palmer, T.N., and Coauthors, 2004: Development of a European multi-model ensemble system for seasonal-to-interannual prediction (DEMETER), *Bull. Amer. Meteor. Soc*., **85,** 853-872.

Paolino, D.A., and Coauthors, 2011: The Impact of Land Surface and Atmospheric Initialization on Seasonal Forecasts with CCSM. Journal of Climate, in press. doi:10.1175/2011JCLI3934.1

Peng, P., A. Kumar, and H. van den Dool, 2002: Ananalysis of multimodel ensemble prediction for seasonal climate anomalies. J. Geophys. Res., 107, 4710, doi:10.10129/2002JD002712.

Reynolds, R.W., N.A. Rayner, T.M. Smith, D.C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. J. Climate, 15, 1609-1625.

Saha, S., and Coauthors, 2006: The NCEP Climate Forecast System. *J. Climate*, **19**, 3483–3517. doi: 10.1175/JCLI3812.1

Shukla, J., R. Hagedorn, M. Miller, T. N. Palmer, B. Hoskins, J. Kinter, J. Marotzke, J. Slingo, 2009: Strategies: Revolution in Climate Prediction is Both Necessary and Possible: A Declaration at the World Modelling Summit for Climate Prediction. *Bull. Amer. Meteor. Soc*., **90**, 175–178. doi: 10.1175/2008BAMS2759.1

Straus, D. M. and J. Shukla, 2002: Does ENSO Force the PNA? *J. Climate*, **15**, 2340–2358.

Weisheimer, A., F. Doblas-Reyes, T. Palmer, A. Alessandri, A. Arribas, M. D´equ´e, N. Keenlyside, M. MacVean, A. Navarra, and P. Rogel, 2009: ENSEMBLES: A new multi-model ensemble for seasonal-to-annual predictions—skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophys Res Lett*. **36**, L21, 711, DOI 10.1029/2009GL040896.

Xie, P., and P. A. Arkin, 1997: Global Precipitation: A 17-Year Monthly Analysis Based on Gauge Observations, Satellite Estimates, and Numerical Model Outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539--2558.

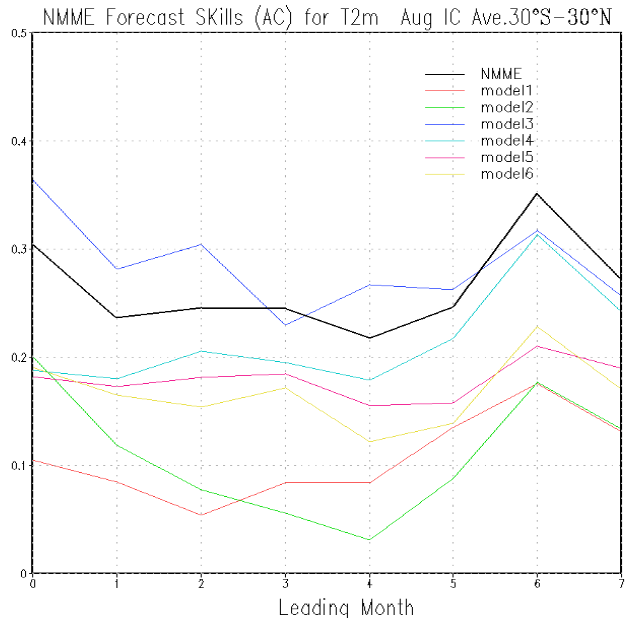

Fig. 7a NMME forecast skills (black line) and individual models (color lines) for temperature at 2m averaged 30S-30N with August ICs.

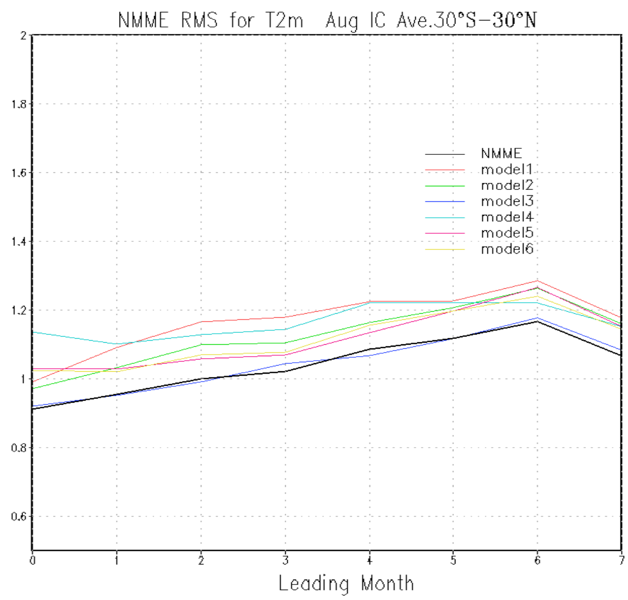

Fig. 7b NMME RMS error (black line) and individual models (color lines) for temperature at 2m averaged 30S-30N with August ICs.