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

Emily Becker¹, Qin Zhang¹, Huug van den Dool¹, Suranjana Saha², Peitao Peng¹, Malaquias Peña², Patrick Tripp², and Jin Huang¹

¹Climate Prediction Center, NCEP/NWS/NOAA
²Environmental Modeling Center, NCEP/NWS/NOAA

1. Introduction

The National Multi-Model Ensemble (NMME) project supplies routine guidance to users. Phase I (experimental) ran from August 2011 to July 2012. Phase II is in operation now. This assessment focuses on the results of Phase I, and includes diagnostic verification of NMME seasonal and monthly prediction. Evaluation metrics include anomaly correlations (AC) calculated from 29 years of hindcasts (1982-2010) and Heidke skill scores (HSS) for the realtime seasonal and monthly 2 meter surface temperature (T2m) and precipitation rate (Prate) forecasts from August 2011 to July 2012 over the contiguous United States. Prediction of the winter for December-January-February (DJF) 2011/12 is used as a case study. This study is motivated by the desire to provide skill benchmarks for future improvements of the NMME seasonal and monthly prediction system.

2. The National Multi-Model Ensemble project

The NMME is a forecasting system consisting of coupled models from U.S. and Canadian modeling centers. The multi-model ensemble approach has been proven to produce better prediction quality than any single model ensemble, motivating the NMME undertaking. The environmental variables included in Phase I (Aug. 2011 – July 2012) were T2m, sea surface temperature (SST), and Prate; and realtime and archived forecast graphics from Aug. 2011 – present are available at www.cpc.ncep.noaa.gov/products/NMME. Hindcast and forecast data are archived at the International Research Institute for Climate and Society (IRI), accessible from the NMME homepage.

NMME Phase I activities included the following models:

- NCEP CFS version 1: 15 ensemble members
- NCEP CFS version 2: 24 ensemble members

Fig. 1 NMME Prate anomaly forecast (a), probability forecast (b), Climate Prediction Center official forecast (c), and observed Prate anomaly from CPC URD (d) for DJF 2011-2012.
• GFDL CM2.1: 10 ensemble members
• IRI ECHAM4-f: 12 ensemble members
• IRI ECHAM4-a: 12 ensemble members
• NCAR-CCSM3.0: 6 ensemble members
• NASA: 6 ensemble members

All models have 1.0° latitude by 1.0° longitude resolution and forecast leads of 1 – 7 months. 29 years of hindcasts (1982-2010) were available for all models except CFSv1 (28 years: 1982-2009). Model forecasts are produced by the 8th of each month, and graphical forecasts are available on the 9th of each month. Phase I forecasts were all delivered on time.

3. Forecast assessment metrics

Anomaly correlations (AC) were assessed for each model over the 29 years (28 for CFSv1) of hindcasts. Global maps and area averages over the Northern and Southern Hemispheres as well as the tropics were produced for T2m, Prate, and SST. Heidke skill scores (HSS) were used to assess forecast verification of Phase 1 probability forecasts. Verifying data sets used comprise GHCN+CAMS T2m, regridded to 1° x 1° (Fan and van den Dool 2008), CPC global Unified Rain-Gauge Database (URD), regridded to 1° x 1° (Xie et al. 2010), and OI-2 Sea-surface temperature (Reynolds et al. 2002), native resolution of 1° x 1°.

4. Results

Analysis of ACs (not shown) reveal that ACs for DJF seasonal forecasts at lead 1 month are higher for NMME forecasts than for individual models for all three fields. This holds for all area averaged ACs.

DJF 2011-2012 Prate anomaly forecast (Fig 1a), probability forecast (Fig 1b), CPC official forecast (Fig 1c), and observed precipitation (Fig 1d) reveal a low HSS, -0.04, for this season over the contiguous United States. Forecasts for T2m (Fig. 2), with a HSS of 0.47, were reasonably good.

Winter 2012 was a difficult case for forecasting. Looking at the scores for monthly forecasts beginning with September 2011 initial conditions (lead-1 month forecast for October 2012) through May 2012 initial

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>201109</td>
<td>0.03</td>
<td>0.02</td>
<td>0.31</td>
<td>0.27</td>
<td>0.12</td>
<td>0.23</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>201110</td>
<td>0.00</td>
<td>0.16</td>
<td>0.41</td>
<td>0.12</td>
<td>0.17</td>
<td>0.11</td>
<td>0.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>201111</td>
<td>-0.31</td>
<td>0.34</td>
<td>0.06</td>
<td>0.09</td>
<td>0.04</td>
<td>0.58</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>201112</td>
<td>0.18</td>
<td>0.15</td>
<td>0.11</td>
<td>0.05</td>
<td>0.39</td>
<td>0.60</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>201201</td>
<td>0.20</td>
<td>0.10</td>
<td>0.14</td>
<td>0.40</td>
<td>0.54</td>
<td>0.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>201202</td>
<td>0.15</td>
<td>0.15</td>
<td>0.40</td>
<td>0.49</td>
<td>0.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>201203</td>
<td></td>
<td>0.13</td>
<td>0.46</td>
<td>0.53</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>201204</td>
<td></td>
<td></td>
<td>0.46</td>
<td>0.33</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>201205</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.32</td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Precipitation rate Heidke skill scores over the CONUS for monthly forecasts from September 2011 – May 2012 initial conditions. Forecast initial month is in the first column; target months read across.
conditions, higher scores are found for both Prate (Table 1) and T2m (Table 2) for the January – July period, even at leads of 6-7 months. This is encouraging, as the late winter, spring and early summer of 2012 produced unusually hot and dry conditions over the CONUS.

5. Summary

At lead 1 month, NMME anomaly correlations for DJF forecast are higher than those of individual models. DJF 2011-2012 was a difficult case, but lead-1 month T2m forecasts over CONUS were reasonably good; precipitation rate forecast had low skill. The very warm and dry late winter through early summer over CONUS were fairly well forecasted, even at long leads. This is a preliminary examination of the forecast verification. A full verification analysis should help to identify sources of strength and weakness.

References

