Science and Technology Infusion Climate Bulletin
NOAA's National Weather Service
41st NOAA Annual Climate Diagnostics and Prediction Workshop
Orono, ME, 3-6 October 2016

Do Statistical Pattern Corrections Improve Seasonal Climate Predictions in NMME Models?

Anthony G. Barnston[1] and Michael K. Tippett[2]

[1]*International Research Institute for Climate and Society (IRI), Columbia University, Palisades, New York*
[2]*Columbia University, New York, New York, and King Abdulaziz University, Jeddah, Saudi Arabia*

1. Introduction

The North American Multimodel Ensemble (NMME) contains global predictions of SST, surface air temperature, precipitation and other variables from 8 or more state-of-the-art coupled general circulation models (Kirtman 2014). In phase one of the NMME project, hindcasts of monthly average climate extending to up to 12 months into the future were created, spanning the 1982-2010 period. Here we seek to determine whether a commonly used multivariate statistical methods—namely, CCA—can improve the temporal anomaly correlation skill of the individual models, with the goal of improving the predictions of the multi-model ensemble. The anomaly correlation is used as the main metric because it measures the ability to reproduce the interannual variability of the climate, regardless of the presence of systematic errors that can be treated locally using simpler statistical methods.

2. Data and methods

The model hindcast data used here are available at the site http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME. The 8 models used include (1) the COLA-RSMAS-CCSM4 model, (2) the NASA-GMAO-062012 model, (3) the GFDL-CM2pl-aer04 model, (4) the GFDL-CM2p5-FLOR-A06, model, (5) the GFDL-CM2p5-FLOR-B01 model, (6) the CMC1-CanCM3 model from Canada, (7) the CMC2-CanCM4 model from Canada, and (8) the NCEP-CFSv2 model. The model global hindcast data are on a 1-degree grid. The 8 models used provide varying numbers of ensemble members, ranging from 10 to 24. Here, the ensemble mean is used to represent the forecast signal, while the ensemble member spread, representing the forecast uncertainty, is not considered.

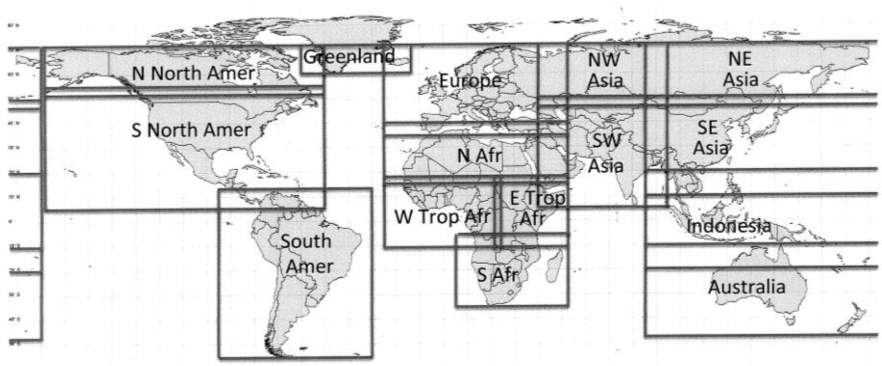

Fig. 1 The 15 slightly overlapping CCA target (predictands) areas, each of which uses a larger predictor area.

The verifying observations, also in a 1-degree grid and available on the above-mentioned web page, are CMAP-URD and GHCN-CAMS for precipitation and temperature, respectively, both created at the NOAA Climate Prediction Center. Most of our attention here is devoted to precipitation prediction.

In the CCA method used here, pre-orthogonalization (EOF analysis) is done separately on the model hindcasts (the X variable) and on the corresponding observations (the Y variable), and a truncated set of the principal component time series from these EOFs are used as input to the CCA. A cross-validation scheme is used, in which 3 consecutive years are withheld from both the pre-EOF and the CCA training sample, and the middle year of the three is predicted. The EOF analysis use the covariance matrix rather than the correlation matrix.

Correspondence to: Anthony G. Barnston, International Research Institute for Climate and Society (IRI), Columbia University, Palisades, New York; E-mail: tonyb@iri.columbia.edu.

The CCA is applied to 15 different regions of the globe (Fig. 1), with the idea that each region would be better treated with respect to the large-scale climate patterns pertinent to it but not necessarily other regions of the globe. The regions are defined so as to capture known leading modes of variability, such as those associated with ENSO. The corrected forecasts of each region will then be merged to form a global forecast.

While the temporal anomaly correlation is used as the main verification measure, the root mean squared error (RMSE) and the uncentered correlation are also computed to detect the presence of calibration issues before and after the CCA.

The regions overlap somewhat so that discontinuities in the forecasts at the boundaries can be smoothed with weightings reflecting the distance from the point in question to the nearest border of the two (or more) regions. The numbers of EOF modes used varies by region to approximately maximize skill, determined by skill sensitivity tests that vary the numbers of X, Y, and CCA modes. The globe as a single region was also used as a 16th "region", allowing for a skill comparison between the merged regional forecasts and the single globe forecast.

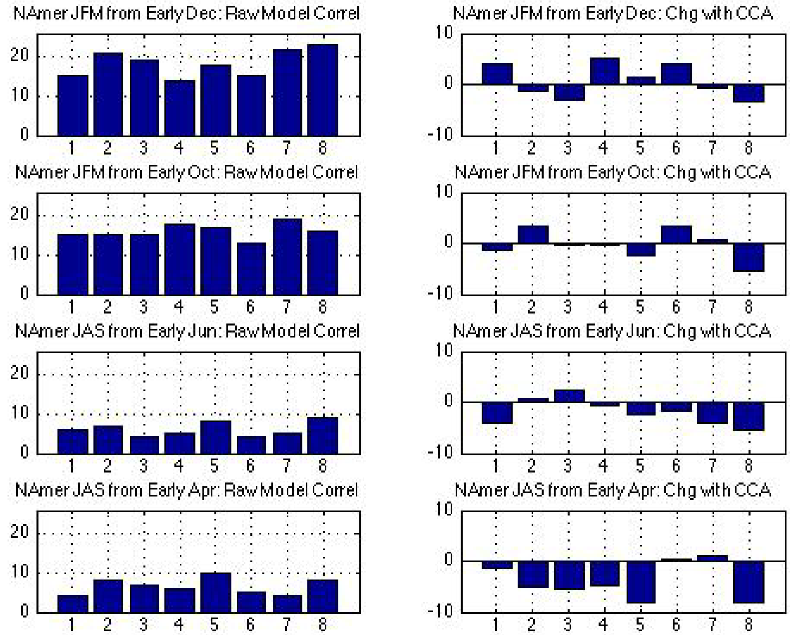

Fig. 2 Original anomaly correlation skill (left), and the change in skill due to the CCA (right). From top to bottom, the results are for (row 1) Jan-Mar precipitation forecasts from early Dec, (row 2) Jan-Mar forecasts from early Oct, (row 3) Jul-Sep forecasts from early Jun, and (row 4) Jul-Sep forecasts from early Apr. The order of the 8 models (horizontal axis) is as listed in the previous section (1:CCSM4, 2:NASA, 3:GFDL, 4:GFDL-FLORA, 5:GFDL-FLORB, 6:CMC1, 7:CMC2, and 8:CFSv2).

3. Results

Figure 2 shows, for the non-northern North America region (including the U.S.) for each of the 8 models, the original anomaly correlation skill, and then the change in skill from the CCA, for the Jan-Mar and Jul-Sep target seasons, each at 1.5-month lead and 3.5-month lead. The CCA corrections generally did not result in substantial skill improvements.

The upper right panel of Fig. 2 shows the CCA did improve the skill of the GFDL-FLOR-A model (model #4). Using this example, Fig. 3 shows the spatial distribution of anomaly correlation skill before and after the CCA correction, and the skill change due to the CCA. Skill was improved in various portions of the U.S., including the Midwest, northern Plains, and other pockets. However, as shown in Fig. 2, in the case of many of the other models for the two seasons and two lead times, average skill was not improved by the

Region	Skill Change	Region	Skill Change
N. North Amer	0.03	South Africa	0.04
S. North Amer	0.01	NW Asia	0.06
South Amer	-0.04	SW Asia	-0.06
Greenland	0.05	NE Asia	-0.01
Europe	-0.05	SE Asia	-0.06
North Africa	-0.04	Indonesia	0.01
W. Trop. Africa	-0.01	Australia	-0.14
E. Trop. Africa	-0.16	**Single Globe**	**-0.001**

Table 1 CCA-related change in anomaly correlation skill for precipitation forecasts for Jan-Mar made in early Dec, averaged over 8 models, for each of 15 individual regions and for the globe as a single region.

CCA.

One can use the CCA to statistically calibrate the forecasts for the entire globe as a single region, rather than merging the corrected forecasts of individual regions. While one might expect the merged skill result to be better than the single globe result (due to the individual attention given to each region), this is not found to be the case. For example, Table 1 shows the change in skill due to the CCA, averaged over all 8 models, for each region and for the globe as a single region, for precipitation forecasts of Jan-Mar made in early December. The average of the skills over the individual regions is slightly lower than the skill of the globe as a single region.

Skill comparisons for other seasons and lead times generally give similar results to that for short-lead forecasts for Jan-Mar, in that the CCA for the single globe does as well as, if not slightly better than, the individually tailored CCAs designed for each region and merged into a global forecast. A summary of these results is shown in Table 2 for the target seasons of Jan-Mar and Jul-Sep, each at 1.5 and 3.5 month lead times. In all cases, the merged CCAs from individual regions does not result as a positive change (or as small a negative change) in skill as the single globe CCA. The original model skills for the merging of individual regions does not exactly equal that of the single globe because the former does not obey area-weighting of skills, while the single globe does.

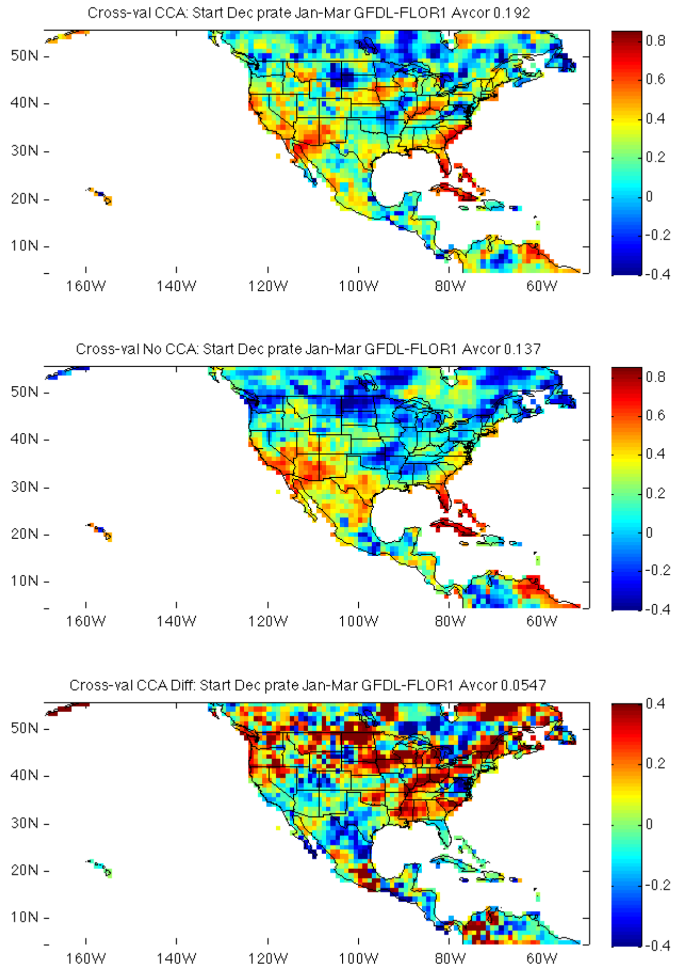

Fig. 3 Geographic distribution of temporal anomaly correlation skill over the non-northern North America region for precipitation forecasts by the GFDL-FLOR-A model for Jan-Mar made in early Dec. The middle panel shows the original skill, top panel the skill after the CCA correction, and bottom panel the skill improvement due to the CCA (note the different scale for the bottom panel).

While the analyses described so far have been limited to precipitation forecasts, a less extensive examination of temperature forecasts indicates an even less favorable result, with the CCA usually degrading the skill of the original uncorrected forecasts. Exceptions are found for temperature forecasts of Oct-Dec forecasts for eastern tropical Africa and Indonesia, where the CCA increases the anomaly correlation.

As complementary verification measures, the RMSE and the uncentered correlation between uncorrected and corrected and precipitation and temperature forecasts and the corresponding observations were also computed. In contrast to the anomaly correlation, both of these measures show a significant improvement following the CCA for both precipitation and temperature. This result suggests that systematic forecast errors that largely do not involve spatial pattern placement are present in the uncorrected forecasts—errors such as mean bias and amplitude bias.

4. Conclusions and discussion

The CCA corrections did not improve the skills of the precipitation forecasts of the individual models of the NMME as much as had been hoped. In fact, overall, a lack of systematic and substantial improvements is noted, with only slight improvements in about half of the cases.

Start ➔Target	Style	Original Model Skill	Change from CCA
Dec => JFM	Merge	0.149	-0.063
	Single Globe	0.153	-0.001
Oct => JFM	Merge	0.109	-0.009
	Single Globe	0.113	0.023
Jun => JAS	Merge	0.114	-0.014
	Single Globe	0.117	0.012
Apr => JAS	Merge	0.086	-0.009
	Single Globe	0.088	0.024

Table 2 Comparison of the effect on globally averaged anomaly correlation skill of the CCA when performed on individual regions and merged to a global precipitation forecast, and when performed on the globe as a single region. Results are shown for forecasts for Jan-Mar made in early Dec and early Oct, and forecast for Jul-Sep made in early Jun and early Apr.

Although the temporal anomaly correlation is not materially improved by the CCA, the RMSE and the uncentered correlations are notably improved. This indicates the presence of biases in the forecasts, such as mean bias and amplitude bias, which affect the two latter verification measures but not the anomaly correlation. One would expect the CCA to treat both local systematic biases and spatial placement errors together. Even if improving both types of errors, it might require retaining more EOF and CCA modes than if just one type of error were present. Mode truncation sensitivity tests were done, and the truncations resulting in approximately the best cross-validated correlation skills were selected.

A few explanations may be offered to account for the unimpressive ability of CCA to improve the model anomaly correlation skills:

i. A covariance matrix was used in the pre-EOFs, rather than a correlation matrix.
ii. The cross-validation may create a negative skill bias (Barnston and Van den Dool 1993), given the large areas of low skill in many of the regions.
iii. Local systematic biases (mean bias and amplitude bias) may be present to a greater extent in the model hindcasts than pattern placement biases. The CCA did improve local biases (seen in RMSE and uncentered correlation) but these improvements alone do not improve the anomaly correlation. Correction of pattern placement biases would be expected to improve the anomaly correlation.

Possible actions to be taken to address these results are:

i. Test the skill behavior when the correlation matrix rather than covariance matrix is used in the pre-EOFs.
ii. Consider a separate correction of local systematic errors prior to the CCA treatment. It is not clear that the CCA can successfully detect and treat spatial placement errors when there are also larger locally correctible systematic errors.
iii. For temperature, try using an alternative data set in place of the GHCN-CAMS. Although it is unlikely, it is possible there is a problem with GHCN-CAMS temperature data.

References

Barnston, A. G., and H. M. van den Dool, 1993: A degeneracy in cross-validated skill in regression-based forecasts. *J. Climate*, **6**, 963-977.

Kirtman B, 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, DOI 10.1175/BAMS-D-12-00050.1