Improved categorical winter precipitation forecasts through multimodel combinations of coupled GCMs

Naresh Devineni[1] and A. Sankarasubramanian[2]

[1]   A new approach to combine precipitation forecasts from multiple models is evaluated by analyzing the skill of the candidate models contingent on the forecasted predictor(s) state. Using five leading coupled GCMs (CGCMs) from the ENSEMBLES project, we develop multimodel precipitation forecasts over the continental United States (U.S) by considering the forecasted Nino3.4 from each CGCM as the conditioning variable. The performance of multimodel forecasts is compared with individual models based on rank probability skill score and reliability diagram. The study clearly shows that multimodel forecasts perform better than individual models and among all multimodels, multimodel combination conditional on Nino3.4 perform better with more grid points having the highest rank probability skill score. The proposed algorithm also depends on the number of years of forecasts available for calibration. The main advantage in using this algorithm for multimodel combination is that it assigns higher weights for climatology and lower weights for CGCM if the skill of a CGCM is poor under ENSO conditions. Thus, combining multiple models based on their skill in predicting under a given predictor state(s) provides an attractive strategy to develop improved climate forecasts. **Citation:** Devineni, N., and A. Sankarasubramanian (2010), Improved categorical winter precipitation forecasts through multimodel combinations of coupled GCMs, *Geophys. Res. Lett.*, *37*, L24704, doi:10.1029/2010GL044989.

1. Introduction

[2]   Seasonal to interannual climate forecasts are regularly issued these days by various national and international agencies using both coupled ocean-atmosphere general circulation models (CGCMs) [*Palmer et al.*, 2004] and atmospheric general circulation models (AGCMs) [*Goddard et al.*, 2003]. Efforts in reducing the errors in climate forecasts have focused on improving the process representation in individual CGCMs/AGCMs as well as by combining multiple models that typically results in error cancellation [*Hagedorn et al.*, 2005]. Recently, using synthetic forecasts, *Weigel et al.* [2008] demonstrated that multimodel forecasts would always result with improved prediction provided individual models' forecasts are overconfident (i.e., ensembles are under-dispersed). Studies from DEMETER

projects show that multimodel combination of CGCMs improves the reliability and skill in predicting summer precipitation in the tropics and winter precipitation in the northern extratropics [*Palmer et al.*, 2004]. Developing multimodel forecasts have been pursued using various techniques that includes simple pooling of the ensembles [*Doblas-Reyes et al.*, 2000] or optimizing weights to maximize the likelihood of the multimodel forecasts [*Rajagopalan et al.*, 2002]. Other methods include various statistical techniques ranging from simple regression [*Krishnamurti et al.*, 1999], dynamic pairwise weighting based on logistic regression [*Chowdhury and Sharma*, 2009], objective Bayesian analysis and multivariate methods.

[3]   Recently, *Devineni and Sankarasubramanian* [2010, hereafter DS10] proposed and evaluated a methodology to combine multiple AGCMs based on their ability to predict under relevant predictor conditions. The motivation behind this combination scheme is to assign higher (lower) weights if a candidate model has good (poor) skill under well-known predictor (e.g., El-Nino Southern Oscillation (ENSO)) conditions that influences the seasonal climate of the region. DS10 demonstrated the utility of the methodology in improving the prediction of winter precipitation and temperature over the U.S conditioned on Nino3.4 state (common index to denote the ENSO conditions). DS10 clearly showed that the weights obtained conditioned on Nino3.4 state were higher for AGCMs that performed well under non-neutral ENSO conditions and weights for climatology (considered as one of the candidate models) were higher since most of the models did not have significant skill under neutral ENSO conditions. The multimodel predictions from DS10 showed improved skill particularly for regions whose winter precipitation and temperature exhibit significant correlation with Nino3.4. However, for this study, DS10 used AGCMs forced with observed SSTs, which typically overestimate the predictive skill of individual models in comparison to the actual forecast skill that could be obtained from CGCMs or from AGCMs being forced with forecasted SSTs. Thus, application of DS10 methodology to retrospective precipitation forecasts would lead to a more objective evaluation of the performance of CGCMs under forecasted Nino3.4 conditions.

[4]   The main goal of this study is to improve the skill of winter (November–February, NDJF) precipitation forecasts over the U.S by combining retrospective forecasts available from multiple CGCMs available from the ENSEMBLES Project [*Weisheimer et al.*, 2009]. We combine precipitation forecasts from multiple CGCMs by evaluating their skill conditioned on the forecasted Nino3.4 state of each model using the multimodel combination algorithm described in DS10. The performance of the multimodel winter precipitation forecasts is compared with the performance of indi-

[1]Columbia Water Center, Earth Institute at Columbia University, New York, New York, USA.
[2]Department of Civil, Construction and Environmental Engineering, North Carolina State University, Raleigh, North Carolina, USA.
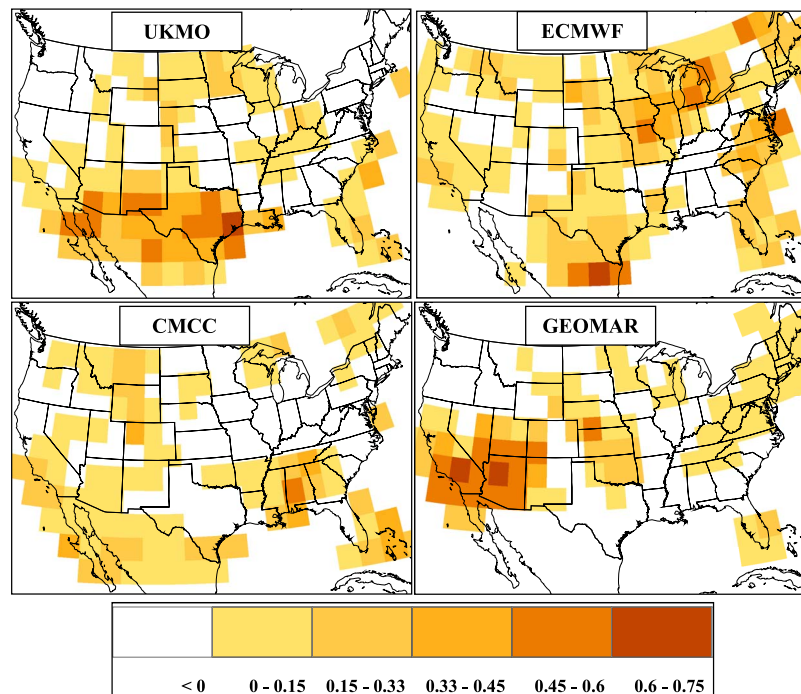
Figure 1. Correlation between squared errors in forecasting DJF Nino3.4 with the squared errors in forecasting DJF precipitation for 37 years over the U.S for four different CGCMs. The significant correlation at 95% confidence interval is 0.33.

vidual models' as well as with the commonly employed techniques for multimodel combination.

2. Data Description

[5] Retrospective precipitation forecasts from the European Union's ENSEMBLES project are used to develop the multimodel forecasts for winter season over the U.S. The ENSEMBLES data is obtained from the public data dissemination system in KNMI Climate Explorer. Forecasts are issued on 1st February, 1st May, 1st August and 1st November and extend up to seven months lead from their respective start times for each year. All the models are represented by nine ensemble members and available for the period of 46 years, 1960–2005. We consider CGCMs' SST forecasts and precipitation forecasts issued on 1st November with a lead time of four months for our analysis. Totally, 200 grid points of forecasts over 123.75W–66.25W and 25N–47.5N are selected from each model across the U.S.

[6] Since it is well known that the skill of winter climate forecasts over the U.S primarily depends on the anomalous conditions (El Nino/La Nina) in the tropical Pacific [*Quan et al.*, 2006], we consider Nino3.4 as the primary predictor influencing the winter precipitation over the U.S. To assess the skill of each model under ENSO conditions, we use the forecasted Nino3.4 which is obtained by averaging the mean of the forecasted SST anomaly over 5S–5N and 170W–120W from the ocean models of the CGCM. Table S1 of the auxiliary material gives the details of the ocean-atmospheric models used in the study.[1] El Nino (Nino3.4 > 0.5), La Nina (Nino3.4 <− 0.5) and neutral conditions (|Nino 3.4| ≤ 0.5) are identified for each model based on the forecasted

Nino3.4. Thus, the total number of El Nino, La Nina and neutral years are different for each model.

[7] Observed monthly precipitation at 0.5° × 0.5° available from University of East Anglia, Climate Research Unit [*New et al.*, 2000], is used to assess the skill of each model. Grid points (0.5° × 0.5°) of monthly precipitation was spatially averaged to map the grid points of the CGCMs (2.5° × 2.5°). In addition to this, we also use the historical Nino3.4, which is obtained from the Kaplan's SST database to assess the skill in forecasting Nino3.4 over 46 years (1960–2005).

3. Methodology

[8] One source of error in forecasting precipitation arises from the error in forecasting the SSTs. Given that SSTs in the tropical Pacific plays an important role in predicting the precipitation, we associate the error in forecasting Nino3.4 with the error of winter precipitation forecasts available from CGCMs. Figure 1 shows the correlation between the squared error in forecasting Nino3.4 and the squared error in forecasting DJF seasonal average precipitation over 37 (1960–1996) years. The correlation that is statistically significant at 95% confidence interval is 0.33. The correlation is positive at many grid points, but significant primarily over certain grid points in Northeast, South and Southwest regions of the U.S. This clearly indicates the association of precipitation forecasts error to the error in predicting the conditioning state-Nino3.4. Thus, by assessing the skill of CGCMs conditioned on forecasted Nino3.4, we intend to develop a more objective assessment on combining multiple CGCMs under forecasted SSTs.

3.1. Multimodel Combination Algorithm

[9] In this section, we briefly outline the modified multimodel combination methodology of DS10 to combine

---

[1]Auxiliary materials are available in the HTML. doi:10.1029/2010GL044989.
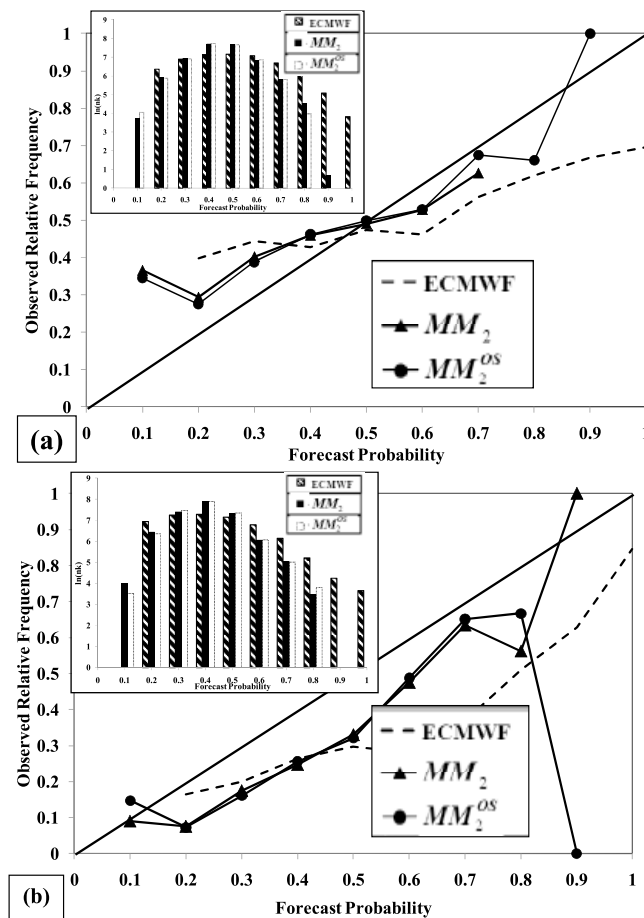
Figure 2. Reliability diagram for ECMWF model and for various multimodel combination schemes based on the seasonal skill in forecasting (a) BN and (b) AN categories of precipitation.

CGCMs conditioned on the forecasted Nino3.4. Figure S1 provides a flow chart of the algorithm that combines tercile forecasts from multiple CGCMs. Tercile categories are computed from the ensembles of CGCMs after removal of systematic bias (i.e., each ensemble is represented as anomaly from the model's seasonal climatology). The squared error in forecasting the observed precipitation from each model is computed for each year over the calibration period (1960–1996). Considering forecasted Nino 3.4 from each model as the predictor that influences the winter precipitation over the U.S we identify 'K' similar conditions to the current Nino3.4 condition. Based on the mean square error (MSE) computed over 'K' neighbors, weights ($w_{t,K}^m$ in Figure S1) for each model are computed for each year. The weights for each model are combined with the models' tercile probabilities to compute the multimodel tercile probabilities. Thus, the weights for each model vary according to the forecasted ENSO conditions. Using the algorithm in Figure S1, we develop five multimodel combinations of winter precipitation forecasts for 200 grid points over the U.S.

3.2. Identification of Similar Predictor Conditions Based on Forecasted Nino3.4

[10] Table S2 provides brief description on the different multimodel schemes considered in the study. Two multi-

model (MM) schemes, MM_1, MM_2, are developed by employing fixed neighbors 'K', which vary conditioned on the forecasted ENSO state for each model (see Table S3 for values K_1, K_2 and K_3, where K_1 = neighbors used if the forecasted Nino3.4 > 0.5, K_2 = neighbors used if the forecasted Nino3.4 < 0.5 and K_3= neighbors used if |forecasted Nino3.4| ≤ 0.5) to obtain the MSE under each scheme. It is important to note that the number of identified neighbors will be different for each model. Under multimodel MM_1, the weights for each month (N, D, J and F) are estimated based on the MSE of each CGCM in predicting precipitation for that month resulting in different weights for each month. Under MM_2, weights are developed based on the MSE in predicting the seasonal (NDJF) average precipitation forecasts. Thus, under this scheme, the weights for (N, D, J and F) for each month in a given year are equal.

[11] Based on the forecasted Nino3.4 by each CGCM over the validation period (1997–2005), values of K provided in Table S3 are used to estimate the MSE of each CGCM in predicting precipitation over the calibration period (1960–1996). MSE obtained for each model are converted into weights for multimodel combination. The weights are then multiplied by the tercile probabilities of the individual models to develop multimodel tercile forecasts. For additional details and a complete discussion on the multimodel combination methodology, see DS10.

[12] We also developed multimodel forecasts by pooling (MM^P) and by overall skill (MM_1^{OS}, MM_2^{OS}) in order to have a baseline comparison with some of the commonly employed techniques in developing multimodel combinations [*Doblas-Reyes et al.*, 2000; *Rajagopalan et al.*, 2002]. Multimodel MM^P obtains multimodel ensembles by pooling all the ensembles from five individual models and climatology. Hence, in this scheme, we have an increased number of ensembles (82). MM_1^{OS} combines individual models based on the overall MSE (unconditional of the ENSO state) for the period 1960–1996 in forecasting N, D, J and F precipitation at a given grid point. Under MM_2^{OS} scheme, the weights are obtained based on the overall MSE in predicting the NDJF precipitation over the entire calibration period. MM_2^{OS} is developed analogous to MM_2 scheme. Both multimodel schemes (MM_1, MM_2) combine individual model forecasts along with climatological ensembles. For climatology, we simply consider observed precipitation at each grid point over the calibration period and mean monthly/seasonal values of Nino3.4 as the conditioning variable for assessing the skill of climatology as a forecasting scheme. The skill of the multimodel forecasts available for each month (9 years of N, D, J and F forecasts) is compared with the skill of individual models' forecasts using standard verification measures such as average Rank Probability Score ($\overline{RPS}$) and average Rank Probability Skill Score ($\overline{RPSS}$). The next section discusses the performance of multimodels in forecasting winter precipitation over the U.S.

4. Results and Analysis

[13] Five multimodel forecasts of winter precipitation are developed by combining five CGCMs and climatology based on the methodology described in Section 3. Multimodel forecasts are represented as tercile probabilities in 200 grid points over the U.S for the four months (N, D, J and F) during the period 1997–2005.
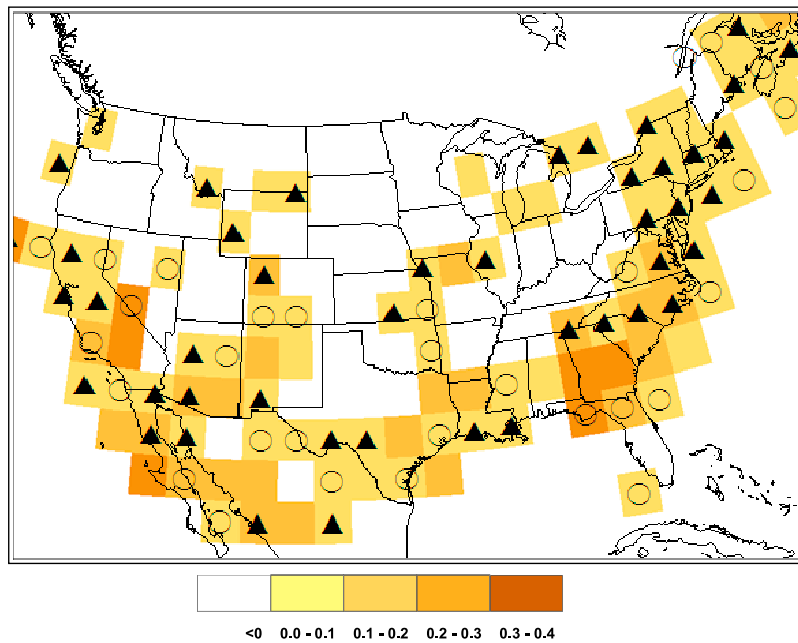
Figure 3. Performance of multimodels and individual models, expressed as $\overline{RPSS}$, in forecasting NDJF winter precipitation. Grid points with no markers, triangle and open circle indicate the best performing model (having the highest $\overline{RPSS}$) being individual CGCMs, multimodels (MM$_1$, MM$_2$) and existing multimodel techniques respectively.

4.1. Comparison Between Individual Models and Multimodels

[14] Figure S2 shows the box plot of $\overline{RPSS}$ for the five individual CGCMs and for five multimodels over the U.S. $\overline{RPSS}$ computes the cumulative squared error between the categorical forecast probabilities and the observed category in relevance to a reference forecast. The reference forecast is usually climatological ensembles that have equal probability of occurrence under each category. A positive (negative) score of $\overline{RPSS}$ indicates that the forecast skill exceeds (less than) that of the climatological probabilities. We computed $\overline{RPSS}$ for both multimodels and individual models using the tercile probabilities for all the four winter months (i.e., a total of 36 forecasts= 4 * 9). Figure S2 also shows the number of grid points that have $\overline{RPSS}$ greater than zero. From Figure S2, we can infer that the individual models' $\overline{RPSS}$ is lesser than zero in most of the grid points which implies that the skill of the CGCMs is poorer than climatology. On the other hand, all the five multimodels perform better than CGCMs with more grid points having positive $\overline{RPSS}$ in forecasting winter precipitation. Among the multimodels, we infer that MM$_1$, MM$_2$, MM_1^{OS} and MM_2^{OS} perform similarly with MM^P having fewer number of grid points indicating better than climatology. Among the individual models, ECMWF perform better than other CGCMs in forecasting winter precipitation. Hence, further analyses in quantifying the improvements resulting from multimodels will focus only on comparing with the performance of ECMWF.

4.2. Reliability and Resolution of Multimodel Forecasts

[15] Reliability diagrams provide information on the correspondence between the forecasted probabilities for a particular category (e.g., above-normal (AN), normal (N) and below-normal (BN) categories) and how frequently that category is observed under the issued forecasted probability. Figures 2a and 2b compare the reliabilities of two multimodels (MM$_2$, and MM_2^{OS}), with the reliabilities of ECMWF in forecasting precipitation for BN and AN categories respectively. We did not consider MM^P since it did not reduce the RPS over many grid points in comparison to the rest of the multimodels in forecasting precipitation.

[16] For developing reliability plots, the tercile probabilities for 36 forecasts under each category are grouped at an interval of 0.1 over all grid points (36 * 200 = 7200 forecasts for a tercile category for each model). The observed category is also recorded using which the observed relative frequency under each forecasted probability is calculated for each tercile category. Inset in each reliability plots show the attribute diagram indicating the logarithm of the number of forecasts that fell under each forecast probability bin for a given model.

[17] From Figure 2, we can see that the individual model forecasts are overconfident and have lower resolution (i.e., the squared error between the observed relative frequency of the forecast and the climatological probability (1/3)). On the other hand, we observe that the multimodels improve the reliability of forecasts showing better correspondence between forecasted probabilities and their observed relative frequencies. We can also observe from Figure 2b (AN category) that both individual model and multimodels are overconfident. But, MM$_2$, performs better than MM_2^{OS} and ECMWF. This can be seen from the attribute diagram which clearly shows reduced number of forecasts from MM$_2$ under high forecast probabilities. (*Reliability diagram for multimodels MM$_1$ and MM_1^{OS} are not shown for better readability and clarity of Figure 2b*). In addition to this, we also computed the average Brier Score ($\overline{BS}$) of multimodel forecasts and ECMWF. $\overline{BS}$ summarizes the mean squared error in categorical forecast probabilities. We observed

(*results not shown*), that all the multimodels (with the exception of MM^p) have smaller reliability score in comparison to the reliability scores of ECMWF under both tercile categories, thereby contributing to the reduction in $\overline{BS}$. In the next section, we compare the performance of five multimodel forecasts under different calibration periods and the improvements resulting from multimodel combination over the U.S. particularly for grid points that exhibit positive $\overline{RPSS}$.

4.3. Performance of Multimodel Forecasts Under Different Calibration Periods

[18] Figure 3 shows the skill, expressed as $\overline{RPSS}$, in forecasting winter (NDJF aggregate, i.e., 36 forecasts) precipitation with each grid point's $\overline{RPSS}$ being indicated by the best performing individual model or the multimodel. From Figure 3, we can clearly understand that multimodels developed in the study perform better than the individual models and over the existing multimodel combination techniques. Among the multimodels, MM_2 seems to be best performing multimodel in forecasting precipitation. Under MM_2 scheme, we basically assign higher weights if a given model performs well under ENSO conditions. From Figure 3, we can also understand that the improvements resulting from multimodel combination in forecasting NDJF precipitation predominantly lies over South and Eastern part of the country as well as over few grid points in the western U.S, particularly over Northern California and Portland. The reason for this improved performance is partly due to strong association between ENSO conditions and the observed winter precipitation (see Figure 1) over these regions. The presented algorithm also assigns higher weights for climatology and lower weights for CGCM if the skill of a CGCM is poor under ENSO conditions (see Figure 10 of DS10).

[19] In addition to this, we also analyzed the performance of multimodels based on different number of years of calibration data. We observed that as the length of calibration period increases, the performance of MM_2 forecasts improved with increased number of grid points having the highest $\overline{RPSS}$. Thus, combining multiple CGCMs conditioned on their skill in predicting under ENSO state offers a new strategy to develop improved multimodel forecasts.

5. Summary

[20] A multimodel combination methodology is presented to combine multiple models conditioned on its performance under ENSO conditions and evaluated for combining retrospective winter precipitation forecasts from CGCMs over the U.S. Considering forecasted Nino3.4 from each model as the primary predictor influencing the winter precipitation, the study combines five CGCMs with climatological ensembles to develop multimodel forecasts over the U.S. Five different multimodel schemes are developed for the period, 1997–2005, by training the retrospective forecasts available during 1960–1996. The skill of the multimodel forecasts are evaluated using $\overline{RPSS}$ and reliability plots based on the monthly forecasts available during winter months. The study clearly shows that all multimodels perform better than

individual models and the proposed multimodel combination algorithm conditioned on the ENSO state performs better than multimodel combinations based on pooling and long-term skill. We also understand that the proposed multimodel methodology is dependent on the length of the calibration period. The main advantage in using this algorithm for multimodel combination is that it assigns higher weights for climatology and lower weights for CGCM if the skill of a CGCM is poor under ENSO conditions. Thus, combining multiple models based on their skill in predicting under a given predictor state(s) provides an attractive strategy to develop improved climate forecasts.

References

Chowdhury, S., and A. Sharma (2009), Long-range Niño-3.4 predictions using pairwise dynamic combinations of multiple models, *J. Clim.*, *22*, 793–805, doi:10.1175/2008JCLI2210.1.

Devineni, N., and A. Sankarasubramanian (2010), Improving the prediction of winter precipitation and temperature over the continental United States: Role of ENSO state in developing multimodel combinations, *Mon. Weather Rev.*, *138*, 2447–2468, doi:10.1175/2009MWR3112.1.

Doblas-Reyes, F. J., M. Deque, and J. P. Piedelievre (2000), Multimodel spread and probabilistic seasonal forecasts in PROVOST, *Q. J. R. Meteorol. Soc.*, *126*, 2069–2088, doi:10.1256/smsqj.56704.

Goddard, L., A. G. Barnston, and S. J. Mason (2003), Evaluation of the IRI's "net assessment" seasonal climate forecasts: 1997–2001, *Bull. Am. Meteorol. Soc.*, *84*, 1761–1781, doi:10.1175/BAMS-84-12-1761.

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

Krishnamurti, T. N., et al. (1999), Improved weather and seasonal climate forecasts from a multimodel superensemble, *Science*, *285*, 1548–1550, doi:10.1126/science.285.5433.1548.

New, M., M. Hulme, and P. D. Jones (2000), Representing twentieth-century space–time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate, *J. Clim.*, *13*, 2217–2238, doi:10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2.

Palmer, T. N., et al. (2004), Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER), *Bull. Am. Meteorol. Soc.*, *85*, 853–872, doi:10.1175/BAMS-85-6-853.

Quan, X., et al. (2006), Diagnosing sources of US seasonal forecast skill, *J. Clim.*, *19*, 3279–3293, doi:10.1175/JCLI3789.1.

Rajagopalan, B., U. Lall, and S. E. Zebiak (2002), Categorical climate forecasts through regularization and optimal combination of multiple GCM ensembles, *Mon. Weather Rev.*, *130*, 1792–1811, doi:10.1175/1520-0493(2002)130<1792:CCFTRA>2.0.CO;2.

Weigel, A. P., M. A. Liniger, and C. Appenzeller (2008), Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts?, *Q. J. R. Meteorol. Soc.*, *134*, 241–260, doi:10.1002/qj.210.

Weisheimer, A., F. J. Doblas-Reyes, T. N. Palmer, A. Alessandri, A. Arribas, M. Déqué, 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*, L21711, doi:10.1029/2009GL040896.

N. Devineni, Columbia Water Center, Earth Institute at Columbia University, 500 West 120th St., New York, NY 10027, USA. (nd2339@columbia.edu)

A. Sankarasubramanian, Department of Civil, Construction and Environmental Engineering, North Carolina State University, Campus Box 7908, 314 Mann Hall, Raleigh, NC 27695, USA.