A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction

Xing Yuan,[1] Eric F. Wood,[1] Lifeng Luo,[2] and Ming Pan[1]

Received 14 April 2011; revised 17 May 2011; accepted 17 May 2011; published 1 July 2011.

[1] NOAA's National Centers for Environmental Prediction (NCEP) has transitioned to operationally use the second generation of their coupled ocean-atmosphere-land model, the Climate Forecast System version 2 (CFSv2), with advanced physics, increased resolution and refined initialization to improve the seasonal climate forecasts. We present a first look at the capability of CFSv2 on surface air temperature and precipitation predictions based on analyzing the 28-year (1982–2009) reforecasts. These variables are primary inputs to hydrological seasonal forecast procedures. Averaged globally, the CFSv2 increases the predictive skill for month-1 land surface air temperature and precipitation from the CFSv1 by 37% and 29%, respectively. The CFSv2 has comparable performance to the latest ECMWF model, the best among the current European seasonal forecast models. The soil moisture produced by CFSv2 also provides useful information in identifying several major drought events, especially over tropical regions. Though there is limited skill beyond month-1, the CFSv2 does show promising features for advancing hydrological forecast and application studies.

Citation: Yuan, X., E. F. Wood, L. Luo, and M. Pan (2011), A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction, *Geophys. Res. Lett.*, *38*, L13402, doi:10.1029/2011GL047792.

1. Introduction

[2] Seasonal time-scale climate predictions, though challenging, have been made routinely by using coupled ocean-atmosphere-land models at a number of operational weather and climate centers around the world since the beginning of this century [*Palmer et al.*, 2004; *Saha et al.*, 2006]. The scientific premise for seasonal predictability is that the slowly evolving components of the climate system, like the sea surface temperature (SST) or land surface states, could aid in predicting the atmospheric development [*Palmer and Anderson*, 1994; *Koster et al.*, 2010]. Operational climate forecast centers such as the European Centre for Medium-Range Weather Forecasts (ECMWF) and the NOAA's National Centers for Environmental Prediction (NCEP) are now updating their seasonal prediction systems to the second generation with improved physics and increased resolution [*Weisheimer et al.*, 2009; S. Saha et al., The NCEP Climate Forecast System version 2, manuscript in preparation, 2011]. Assessing the retrospective forecast of the new systems for terrestrial hydrological forecasting, i.e., the predictive skill of surface air temperature, precipitation and soil moisture, can provide valuable baseline information to a wide cross section of application studies, such as flood and drought forecasting, and agriculture and water resources management.

[3] However, the land surface variables are more difficult to predict than the SST especially at long lead times, although the former have more direct impact on society than the latter. For instance, *Weisheimer et al.* [2009] showed that the ensemble of multiple dynamic coupled models have significant anomaly correlation skill for tropical Pacific SSTs up to 14 months ahead, while *Lavers et al.* [2009] found that precipitation and surface air temperature have pronounced drop in skill beyond month-1. Fortunately, the relatively low skill in surface variables can be compensated to some extent by improved representation of physical processes and data assimilation, statistical or dynamical downscaling techniques [*Luo et al.*, 2007; *Alessandri et al.*, 2011; *Yuan and Liang*, 2011]. Recently, the second version of Climate Forecast System (CFSv2) was implemented operationally in NCEP as the real-time seasonal forecast system. The CFSv2 incorporates a number of new physical packages for cloud-aerosol-radiation, land surface, ocean and sea ice processes, as well as a new atmosphere-ocean-land data assimilation system [*Saha et al.*, 2010]. To facilitate real-time forecasts, a set of fully coupled retrospective forecasts covering the 28-year period (1982–2009) were produced with CFSv2 (Saha et al., manuscript in preparation, 2011). A first look at these 24-member ensemble retrospective forecasts for hydrological forecasting needs will provide useful information before updating the hydrological seasonal forecast system [*Luo et al.*, 2007] based on the CFSv2.

[4] In this letter, we assess the predictability of CFSv2 for precipitation and surface air temperature by comparing its forecasts with the reforecast results from CFSv1 [*Saha et al.*, 2006] and three European seasonal forecast models in ENSEMBLES [*Weisheimer et al.*, 2009]. The monthly soil moisture forecast quantiles from CFSv2 are also evaluated against the Climate Forecast System Reanalysis (CFSR) [*Saha et al.*, 2010] for drought prediction.

2. Data

[5] The 28-year (1982–2009) ensemble retrospective forecast data set from CFSv2 with 24 members is provided by NCEP (Saha et al., manuscript in preparation, 2011). Beginning at January 1st, 9-month hindcasts were initiated every 5 days with 4 cycles on those days. Currently, only monthly mean values for the ensembles are available, with plans to make 6-hourly resolution available via the National Climatic Data Center (NCDC). NCEP compiled the monthly estimates as follows: for each calendar month, the hindcasts with initial dates after the 7th of that month were used as the

[1]Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, USA.
[2]Department of Geography, Michigan State University, East Lansing, Michigan, USA.

0094-8276/11/2011GL047792

ensemble members of the next month. For instance, the starting dates for the February ensemble members are the January 11th, 16th, 21st, 26th, 31st, and the February 5th. In general, there are at least 16 members having initial dates before the 1st day of the target month. The CFSv2 used in the reforecast consists of the NCEP Global Forecast System at T126 (~0.937°) resolution, the Geophysical Fluid Dynamics Laboratory Modular Ocean Model version 4.0 at 0.25–0.5° grid spacing coupled with a two-layer sea ice model, and the four-layer NOAH land surface model. NCEP also provides the nine-month reforecast for CFSv1 at T62 (~1.875°) resolution covering the same period 1982–2009, where there are 15 ensemble members for each target month [*Saha et al.*, 2006]. The first and second 5 members have initial dates within the 9th–13th and the 19th–23rd of the month before the target month, and the last 5 members have initial dates between the second-to-last day of the month before the target month and the 3rd of the target month.

[6] The ENSEMBLES project, the successor of DEMETER [*Palmer et al.*, 2004], is a seasonal-to-annual multi-model ensemble forecast system containing five models each with nine ensemble members [*Weisheimer et al.*, 2009]. In this study, we selected the EUROSIP (European Operational Seasonal to Interannual Prediction) models, ECMWF, Météo France (MF), and UK Met Office (UKMO), which also have higher predictability among the ENSEMBLES models. The available dataset is at 2.5° × 2.5° resolution. The hindcast period of ENSEMBLES covers the 46 years 1960–2005, with 7-month seasonal forecasts starting on the 1st of February, May, August and November of each year. ENSEMBLES models have similar development strategy as CFSv2: improving the physical parameterizations by including new sea-ice or land surface modules and interannual variability in the greenhouse gas forcing, increasing the resolution, and refining the initialization.

[7] Surface air temperature at 0.5° from the Climate Research Unit (CRU) TS3.1 dataset for the period 1901–2009 [*Mitchell and Jones*, 2005], precipitation at 0.5° from the Climate Prediction Center (CPC) Unified Gauge-Based Analysis for 1979–2009 [*Chen et al.*, 2008], and total column soil moisture at T382 (~0.312°) from CFSR for 1979–2009 [*Saha et al.*, 2010] are used as the reference datasets. The data from model, observation or reanalysis used in this study are over land grids at monthly time scales.

3. Results

[8] To match the ENSEMBLES models' grids, the surface air temperature and/or precipitation from CFSv1, CFSv2, CRU and CPC were regridded to 2.5° × 2.5° resolution. The common period is 1982–2005 (24 years). To assess the deterministic quantities of the forecast models, the product-moment correlation coefficients between the observed and forecast ensemble mean series were calculated in each grid cells [*Wilks*, 2006]. Figure 1 illustrates geographic distributions of the month-1 predictive skill in surface air temperature forecasts over land grids in February, May, August and November for CFSv1, CFSv2, ECMWF, MF, UKMO and a multi-model forecast, where the multi-model forecast is the average of the CFSv2, ECMWF, MF and UKMO. As compared with CFSv1, CFSv2 shows overall improvement, especially in the cold season. CFSv2 has high predictive skill (correlation larger than 0.7) over Amazon, Europe, Middle East, northwestern and eastern Asia, and eastern Australia in February, over Alaska, North America Monsoon region, Amazon, northwestern Africa, middle Asia and western Australia in May, over North America, middle and southern Asia, Siberia and eastern Australia in November. *Wang et al.* [2010] identified the cold bias of CFSv1 over northern hemisphere mid-high latitudes in the real-time seasonal forecast during the warm season, while the CFSv2 reduced the bias for the August forecast by 53%, averaged over the globe, and the reduction was more pronounced over the high latitudes in Eurasia (not shown). As compared with the ENSEMBLES EUROSIP models, CFSv2 has similar performance to ECMWF, where the former has higher skill over North America in November, while the latter has higher skill over northern Africa in February. The multi-model combines the advantages of individual models, and presents generally improved predictability. However, similar to the first generation of ensembles of multiple seasonal forecast models [*Lavers et al.*, 2009]; neither the multi-model nor the individual models in this study have significant predictability over western Russia in May and August, which is an issue that needs further investigation. On average, the global mean (except Antarctica) correlations of surface air temperature for the four months are as follows: CFSv1, 0.38; CFSv2, 0.52; ECMWF, 0.51; MF, 0.39; UKMO, 0.44; and the multi-model, 0.54. Note that the global mean correlation for multi-model in November is slightly smaller than the CFSv2, which may result partly from the low skill of MF and UKMO (Figure 1).

[9] Figure 2 demonstrates the month-1 predictive skill for precipitation. There are fewer grid cells with significant correlation (>0.4) than for temperature, however, CFSv2 does show obvious enhancement compared with CFSv1. CFSv2 increases the ratio of grid cells that have significant predictive skill from CFSv1 by 52%, and the improvement is more pronounced in cold seasons. Similar to the temperature, the CFSv2 has comparable precipitation predictability to ECMWF, and both of them are better than the MF and UKMO. The multi-model forecast produces very promising results over some regions for specific months. Taking the month-1 forecast over United States in May as an example, the CFSv2 has high skill over western U.S. and ECMWF has high skill over southern and eastern U.S.; consequently, the multi-model has significant predictability over almost the whole country. The global mean correlations of precipitation averaged over the four months are as follows: CFSv1, 0.21; CFSv2, 0.27; ECMWF, 0.28; MF, 0.18; UKMO, 0.24 and the multi-model, 0.31. One drawback of CFSv2 as compared with other models is that it under-predicts the inter-annual variability of precipitation by about 30% globally.

[10] For the month-2 forecast results, CFSv2 increases the ratios of grid cells that have significant predictive skill from CFSv1 by 169% for temperature, but only by 1% for precipitation. Even for the multi-model, only 9% of the grid cells have significant correlation for the month-2 precipitation forecast, and most of them are located in the Amazon basin (not shown). Actually the skill of the second half of month-1 is much lower than the first half (not shown), therefore, improvements in precipitation predictability over the mid-latitudes beyond 15 days is still an unresolved issue for dynamical seasonal forecast.

[11] Besides assessing the deterministic quantities above, the probabilistic quality of the forecasts is evaluated by using

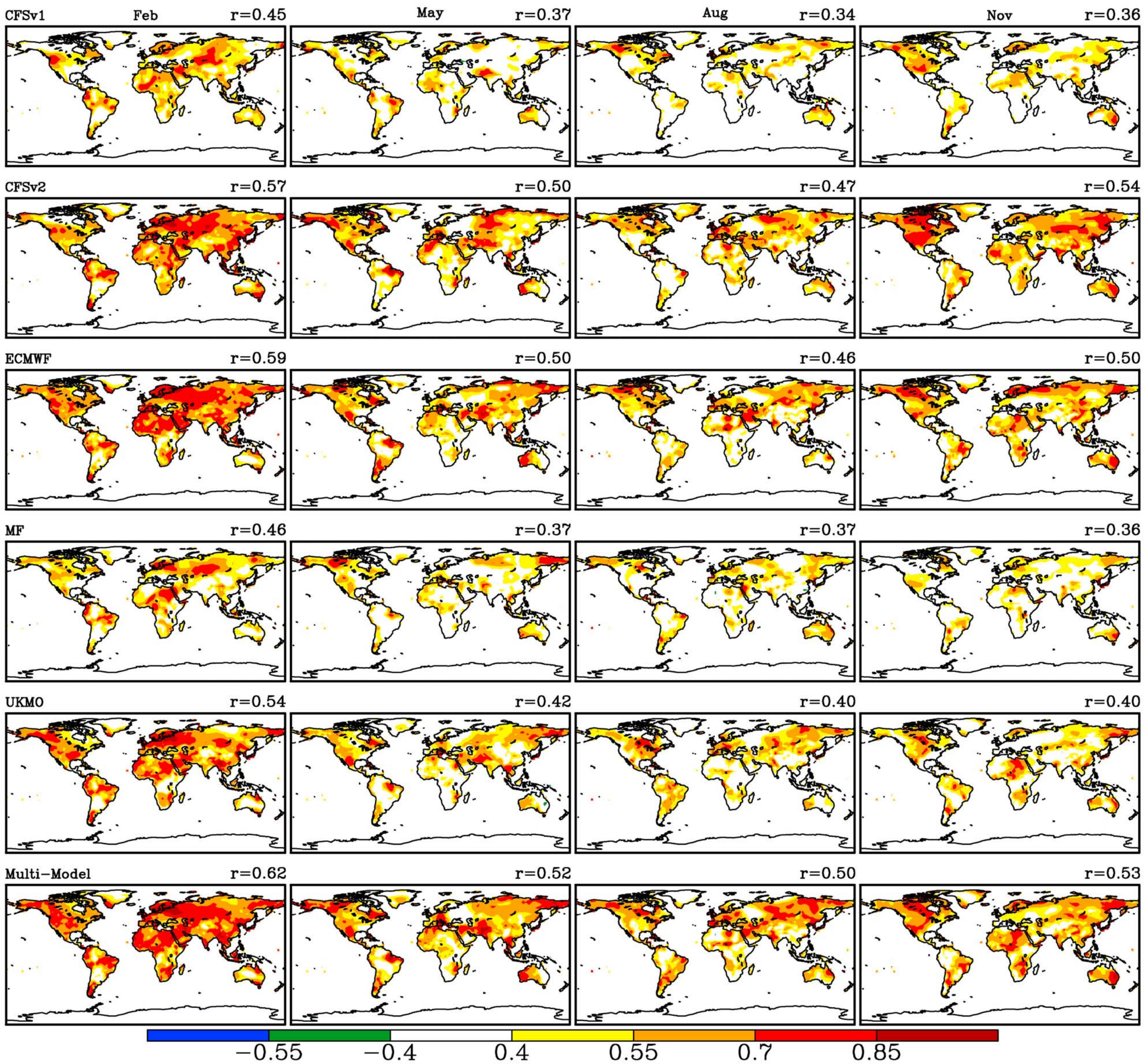

Figure 1. Month-1 predictive skill of surface air temperature forecasts over land grids in February, May, August and November during 1982–2005. The multi-model is the average among CFSv2, ECMWF, MF and UKMO. Non-white colors represent significant correlation at 0.05 levels. The numbers are the global mean correlations.

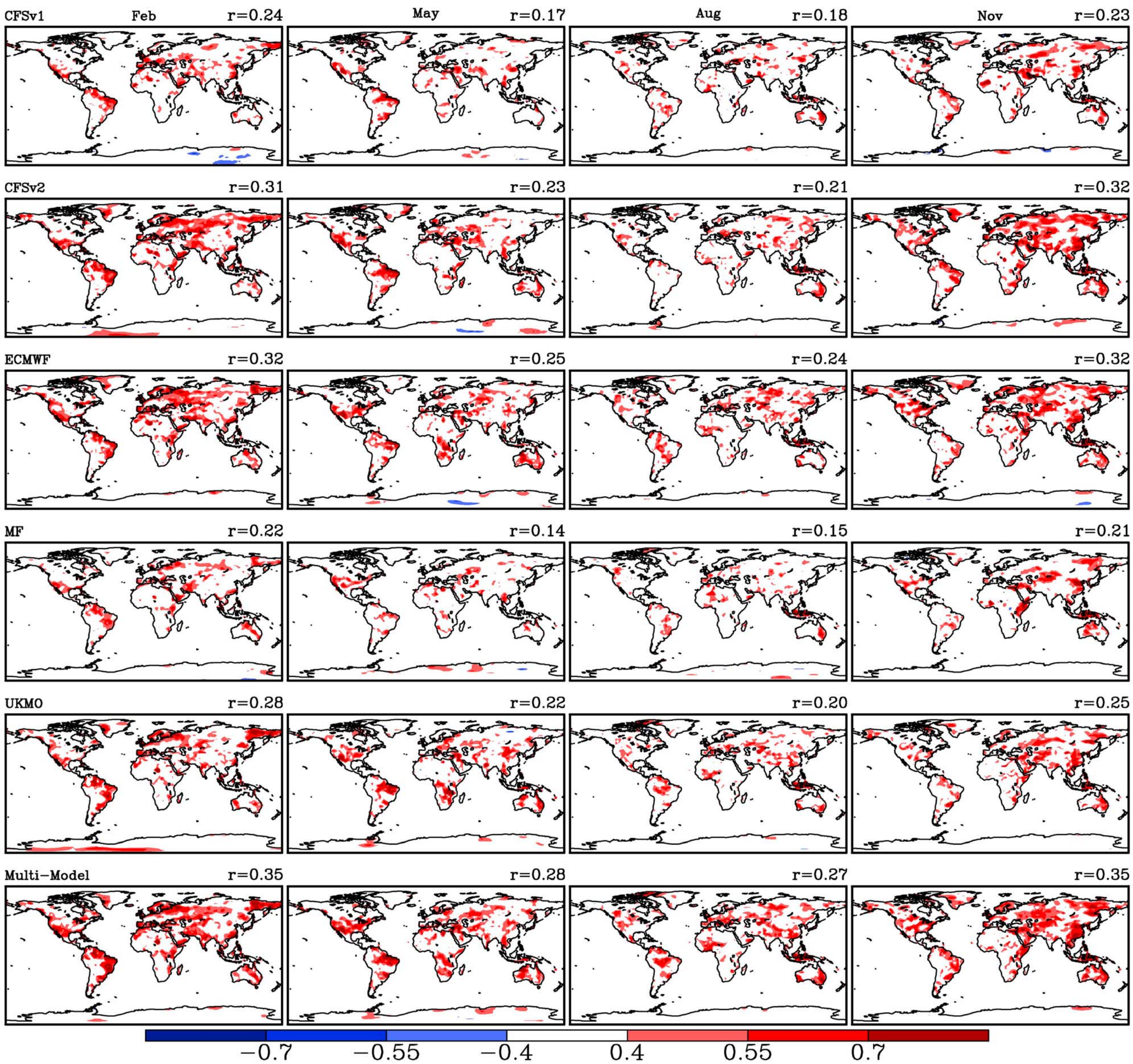

Figure 2. The same as Figure 1, but for monthly precipitation.

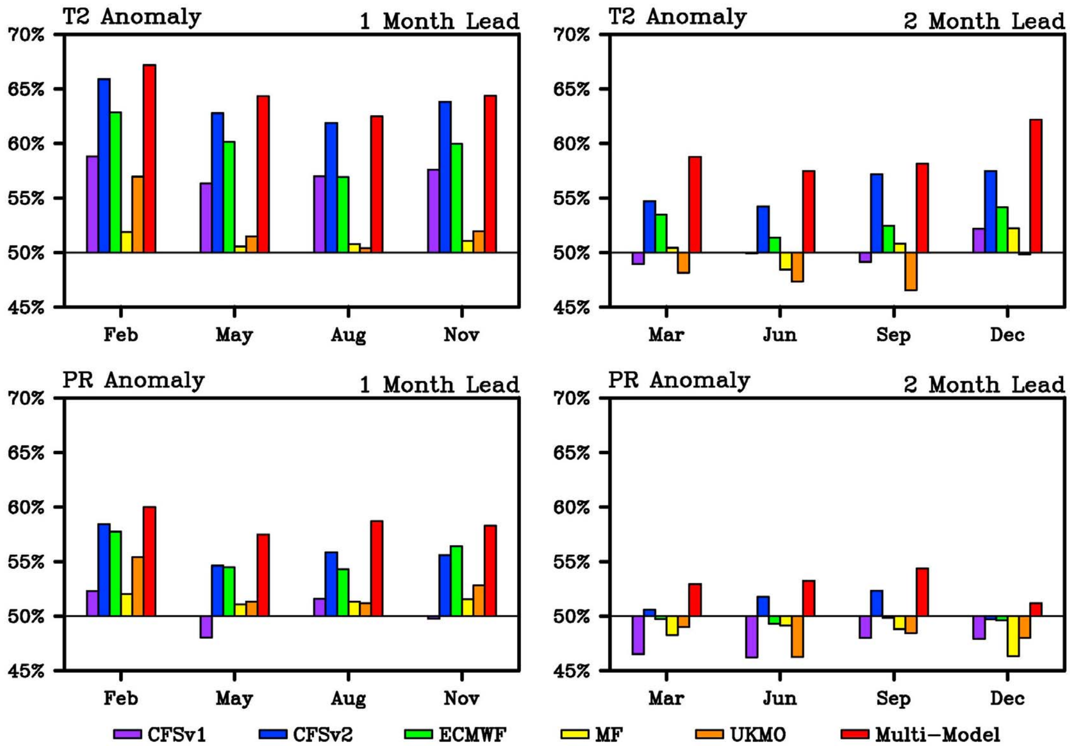

Figure 3. Percentage of positive Ranked Probability Skill Score (RPSS) for monthly surface air temperature and precipitation anomaly over the global land area during 1982–2005.

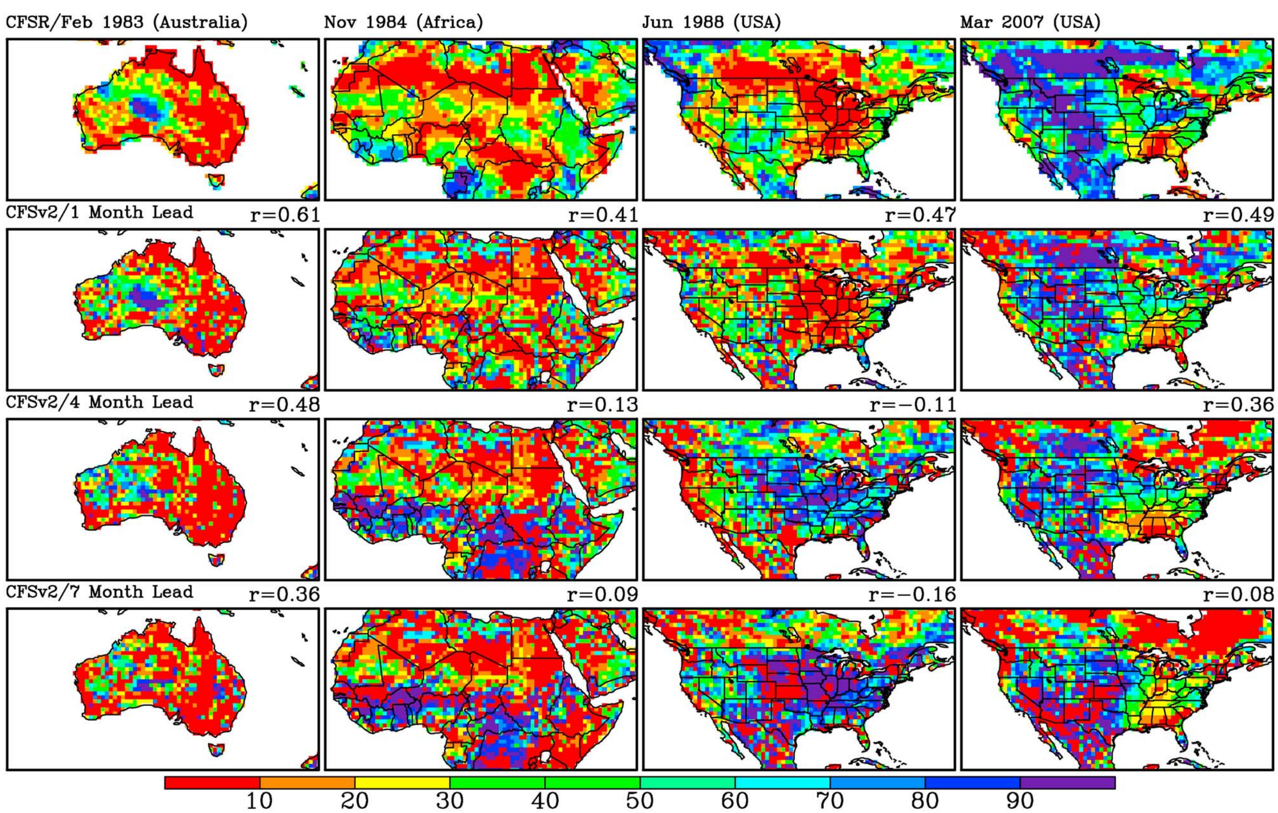

Figure 4. Monthly soil moisture percentiles from CFSR and CFSv2 with different forecast leading times for four regional droughts. The numbers are the pattern correlations.

the Ranked Probability Skill Score (RPSS) defined as 1-RPS/RPS_{Clim}, where the RPS and RPS_{Clim} are the ranked probability scores from the model and climatological forecast respectively [*Wilks*, 2006]. Thus, RPSS = 1 indicates perfect multi-category probability forecast, while RPSS less than zero means the forecast is worse than climatology. Given that the models in this study have at least 9 members, we ignored the effects of ensemble size on the bias of RPSS [*Müller et al.*, 2005]. Figure 3 presents the percentage of the positive RPSS for monthly surface air temperature and precipitation anomaly over the global land area (except Antarctic) for the period 1982–2005. For the surface air temperature anomaly, the CFSv2 has a higher percentage of forecasts with skill beyond climatology than the other models. Even out to two months, more than 54% of the forecasts from the CFSv2 produce useful predictions. For the precipitation anomaly, the performance of the CFSv2 is comparable to the ECMWF, and both are much better than the other models. Similarly to the deterministic evaluation, the skill of probabilistic forecasts for precipitation for each individual model and the multi-model drops greatly beyond one month.

[12] Figure 4 shows the monthly soil moisture quantiles from CFSR and CFSv2 with different forecast lead times for four regional droughts identified in previous studies [*Luo and Wood*, 2007; *Sheffield and Wood*, 2007]. The soil moisture percentile is used as a drought index value computed from the current total column soil moisture with respect to its climatological probability distribution. We choose the CFSR soil moisture as the reference because it results from the same land surface model with CFSv2, which may reduce some uncertainty for agricultural drought analysis [*Mo et al.*, 2011]. The strong El Niño of 1982/83 is the major cause of drought in Australia in that period. CFSv2 successfully captured the severe drought conditions in eastern Australia even at the 7 month lead time forecast. Though the drought region is exaggerated at the long lead time, the spatial pattern is highly correlated with the CFSR (Figure 4). The CFSv2 also roughly predicted drought over Sahel region, where the long-term drought conditions greatly affected society and the environment. For the 1988 drought over U.S., which has been considered to be one of the worst natural disasters in recent U.S. history, the CFSv2 totally missed it beyond the month-1 forecast. In fact, the CFSv2 over-predicted the precipitation at long lead times for June 1988, which is somewhat surprising since such a drought is a combination of SST and persistent stationary atmospheric circulation anomalies, and soil moisture-precipitation feedbacks. For this event, CFSv1 has relatively better results (not shown). The underlying reasons for the degradation in forecasting the 1988 drought by CFSv2 needs to be investigated and perhaps can provide insights into seasonal forecasting skill of drought. For the recent drought over southeastern U.S. in early 2007, the CFSv2 performed well at long lead times (as did CFSv1 [*Luo and Wood*, 2007]), although CFSv2 underestimated the soil moisture over the North American Monsoon region.

4. Concluding Remarks

[13] This study provides a first look at the capability of the NCEP's latest operational seasonal forecast model CFSv2 for hydrological seasonal forecasting variables (precipitation, surface air temperature and soil moisture) by comparing its hindcast forecast skill with the CFSv1, ENSEMBLES EUROSIP models and the CFSR data. CFSv2 shows significant skill enhancement for land surface air temperature and precipitation from CFSv1 for month-1 forecasts. CFSv2 has comparable result to ECMWF, where the former and the latter have slightly higher skill in temperature and precipitation respectively. The multi-model ensemble of CFSv2 and three EUROSIP models in ENSEMBLES produces overall better results than the individual models, and 35% of the land areas have significant predictive skill for temperature out to 2 months. The probabilistic evaluation result is consistent with the deterministic assessment, where the CFSv2 has the highest percentage of skillful forecasts beyond climatology. It is anticipated that the increased skill in new seasonal climate models will benefit the seasonal hydrological predictions based on the Bayesian merging framework developed by *Luo et al.* [2007]. CFSv2 also shows some predictability of drought (identified by soil moisture percentiles) at long lead times, especially over the tropical regions. Thus CFSv2 can provide a direct reference for drought prediction at seasonal scales.

[14] However, there are also some concerns in terms of using the new climate forecast results for hydrological predictions. Firstly, even the multi-model has limited predictive skill for precipitation beyond month-1, which is relatively short for a seasonal hydrological forecast. Though efforts have been made by realistically initializing the land surface conditions in climate models [*Koster et al.*, 2010] and in hydrologic ensemble predictions [*Schaake et al.*, 2007], further improvements from seasonal climate models are needed for providing useful long-lead information. In fact, through developing physical parameterizations, increasing model resolution and refining ocean-land-atmosphere initial conditions, the second generation of operational seasonal forecast models does show significant improvement for month-1 forecast. Secondly, CFSv2 underestimates the inter-annual variability for monthly precipitation by about 30%, and has less reliability than the CFSv1 for predicting heavy rainfall events (not shown). The underlying reason and its potential impact on forecasting hydrological extremes need further investigation. Thirdly, besides bias correcting and downscaling meteorological fields to force land surface models (LSMs) that provide hydrological seasonal forecasts, direct downscaling of soil moisture and runoff from the seasonal forecast models may become feasible since the new forecast models contain advanced LSMs. The drought prediction assessment in this study provides an example of the potential hydrological predictive skill from CFSv2. Fourthly, each model result in this study is an ensemble of several equally-weighted members that provide a monthly mean forecast. Since CFSv2 generates four seasonal ensemble members every five days, how to optimally utilize these individual members, especially for months 1 and 2, in our probability hydrological forecast system [*Luo et al.*, 2007] is the focus of future research. Understanding this is an imperative for improved usage of CFSv2 since its 24 ensembles members have different forecast skill due to different forecast lead times. Perhaps a special re-sampling strategy should be adopted. Finally, all of the results in this letter are based on grid point analyses, scrutinizing the model performances in specific climatic zones or river basins are needed for regional studies.

[15] **Acknowledgments.** The research presented in the paper was supported by the NOAA Climate Program Office through grants NA17RJ2612 and NA10OAR4310246. This support is gratefully acknowledged. The ENSEMBLES data used in this work was funded by the EU FP6 Integrated Project ENSEMBLES (contract 505539). The authors gratefully acknowledge the availability of the forecast information.

[16] The Editor thanks Paul Dirmeyer and an anonymous reviewer for their assistance in evaluating this paper.

References

Alessandri, A., et al. (2011), Evaluation of probabilistic quality and value of the ENSEMBLES multimodel seasonal forecasts: Comparison with DEMETER, *Mon. Weather Rev.*, *139*(2), 581–607, doi:10.1175/2010MWR3417.1.

Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. W. Higgins, and J. E. Janowiak (2008), Assessing objective techniques for gauge-based analyses of global daily precipitation, *J. Geophys. Res.*, *113*, D04110, doi:10.1029/2007JD009132.

Koster, R. D., et al. (2010), Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment, *Geophys. Res. Lett.*, *37*, L02402, doi:10.1029/2009GL041677.

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 (2007), Monitoring and predicting the 2007 U.S. drought, *Geophys. Res. Lett.*, *34*, L22702, doi:10.1029/2007GL031673.

Luo, L., E. F. Wood, and M. Pan (2007), Bayesian merging of multiple climate model forecasts for seasonal hydrological predictions, *J. Geophys. Res.*, *112*, D10102, doi:10.1029/2006JD007655.

Mitchell, T. D., and P. D. Jones (2005), An improved method of constructing a database of monthly climate observations and associated high-resolution grids, *Int. J. Climatol.*, *25*(6), 693–712, doi:10.1002/joc.1181.

Mo, K. C., L. N. Long, Y. Xia, S. K. Yang, J. E. Schemm, and M. Ek (2011), Drought indices based on the Climate Forecast System Reanalysis and ensemble NLDAS, *J. Hydrometeorol.*, *12*(2), 181–205, doi:10.1175/2010JHM1310.1.

Müller, W. A., C. Appenzeller, F. J. Doblas-Reyes, and M. A. Liniger (2005), A debiased ranked probability score to evaluate probabilistic ensemble forecasts with small ensemble sizes, *J. Clim.*, *18*(10), 1513–1523, doi:10.1175/JCLI3361.1.

Palmer, T. N., and D. L. T. Anderson (1994), The prospects for seasonal forecasting—A review paper, *Q. J. R. Meteorol. Soc.*, *120*(518), 755–793.

Palmer, T. N., et al. (2004), Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction (DEMETER), *Bull. Am. Meteorol. Soc.*, *85*(6), 853–872, doi:10.1175/BAMS-85-6-853.

Saha, S., et al. (2006), The NCEP climate forecast system, *J. Clim.*, *19*(15), 3483–3517, doi:10.1175/JCLI3812.1.

Saha, S., et al. (2010), The NCEP Climate Forecast System Reanalysis, *Bull. Am. Meteorol. Soc.*, *91*(8), 1015–1057, doi:10.1175/2010BAMS3001.1.

Schaake, J. C., T. M. Hamill, R. Buizza, and M. Clark (2007), HEPEX: The Hydrological Ensemble Prediction Experiment, *Bull. Am. Meteorol. Soc.*, *88*(10), 1541–1547, doi:10.1175/BAMS-88-10-1541.

Sheffield, J., and E. F. Wood (2007), Characteristics of global and regional drought, 1950–2000: Analysis of soil moisture data from off-line simulation of the terrestrial hydrologic cycle, *J. Geophys. Res.*, *112*, D17115, doi:10.1029/2006JD008288.

Wang, W., M. Chen, and A. Kumar (2010), An assessment of the CFS real-time seasonal forecasts, *Weather Forecast.*, *25*(3), 950–969, doi:10.1175/2010WAF2222345.1.

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.

Wilks, D. S. (2006), *Statistical Methods in the Atmospheric Sciences*, 2nd ed., Academic, Burlington, Mass.

Yuan, X., and X.-Z. Liang (2011), Improving cold season precipitation prediction by the nested CWRF-CFS system, *Geophys. Res. Lett.*, *38*, L02706, doi:10.1029/2010GL046104.

L. Luo, Department of Geography, Michigan State University, 122 Geography Bldg., East Lansing, MI 48824, USA.

M. Pan, E. F. Wood, and X. Yuan, Department of Civil and Environmental Engineering, Princeton University, Engineering Quadrangle, Princeton, NJ 08544, USA. (xingy@princeton.edu)