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

Enhancing Hydrological Seasonal Forecast by Downscaling CFSv2

Eric F. Wood, Xing Yuan, Joshua K. Roundy
Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ

1. Introduction

Due to the improvements in data assimilation techniques, computing resources, and numerical models such as representing large-scale climate teleconnections, forecasting seasonal climate by using coupled atmosphere-ocean-land general circulation models (CGCMs) has been an emerging area since mid-1990s. Now the major operational weather and climate forecast centers around the world are producing real-time seasonal climate predictions with CGCMs up to nine months. The progress in dynamical seasonal forecasts provides potential opportunity to predict hydrologic variables (*e.g.*, streamflow, soil moisture) at long lead times, which is important for agriculture and water resources management, drought and flood detection and mitigation. In this context, seasonal hydrologic forecast plays an important role in transitioning the scientific advances from the climate research community to the end users of society (Yuan *et al.* 2011b).

Recently, NCEP has updated its operational seasonal forecast system with a new CGCM, the second version of CFS (CFSv2), where 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 have been incorporated (Saha *et al.* 2010, 2011). Therefore, it is necessary to conduct comprehensive seasonal hydrologic reforecasts, and to investigate whether or how much of the improvement from climate forecast models can propagate into the land surface hydrologic forecast. Here, we summarize a deterministic assessment and probabilistic evaluation of CFSv2 by comparing with CFSv1 and European seasonal forecast models, and the CFSv2-based seasonal hydrologic forecast results. Full documentations are available in Yuan *et al.* (2011a, 2011b).

2. Data and method

To have a first look at CFSv2 in a hydrologic forecast perspective, we used the data as follows: 1) the 28-year (1982-2009) ensemble retrospective forecast data set from CFSv2 with 24 members at T126 resolution (Saha *et al.* 2011), and the nine-month reforecast for CFSv1 with 15 members at T62 resolution covering the same period 1982-2009 (Saha *et al.* 2006); 2) the EUROSIP (European Operational Seasonal to Interannual Prediction) model (ECMWF, Météo France (MF), and UK Met Office (UKMO)) reforecasts at 2.5°×2.5° resolution from 1960-2005 (Weisheimer *et al.* 2009); and 3) surface air temperature at 0.5° from the Climate Research Unit (CRU) TS3.1 dataset for the period 1901-2009 (Mitchell and Jones 2005), and precipitation at 0.5° from the Climate Prediction Center (CPC) Unified Gauge-Based Analysis for 1979-2009 (Chen *et al.* 2008).

We downscaled the precipitation and temperature hindcasts from CFSv1 and CFSv2 to 1/8 degree over CONUS by using the Bayesian method described in Luo *et al.* (2007) and Luo and Wood (2008). The downscaled precipitation and temperature fields are used as inputs to the VIC model (in water balance mode) to provide 6-month, 20-member ensemble hydrologic reforecasts starting on the 1st of February, May, August and November of each year during 1982-2008, with initial conditions from a 62-year (1949-2010) offline simulation driven by merged data from 31-year (1949-1979) University of Washington dataset (Maurer *et al.* 2002) and 31-year (1980-2010) dataset from the North-American Land Data Assimilation System Project Phase 2 (NLDAS-2; Xia *et al.* 2011).

3. Results

Figure 1 illustrates the geographic distributions of the month-1 predictive skill in terms of correlation for surface air temperature forecasts over land grids for CFSv1, CFSv2, ECMWF, MF, UKMO and a multi-

---

Correspondence to: Eric F. Wood, Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ; E-mail: efwood@princeton.edu.

model forecast (Yuan *et al.* 2011a). As compared with CFSv1, CFSv2 shows overall improvement, especially in the cold season. 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 equally-weighted multi-model combines the advantages of individual models, and presents generally improved predictability. However, 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 (excluding 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).

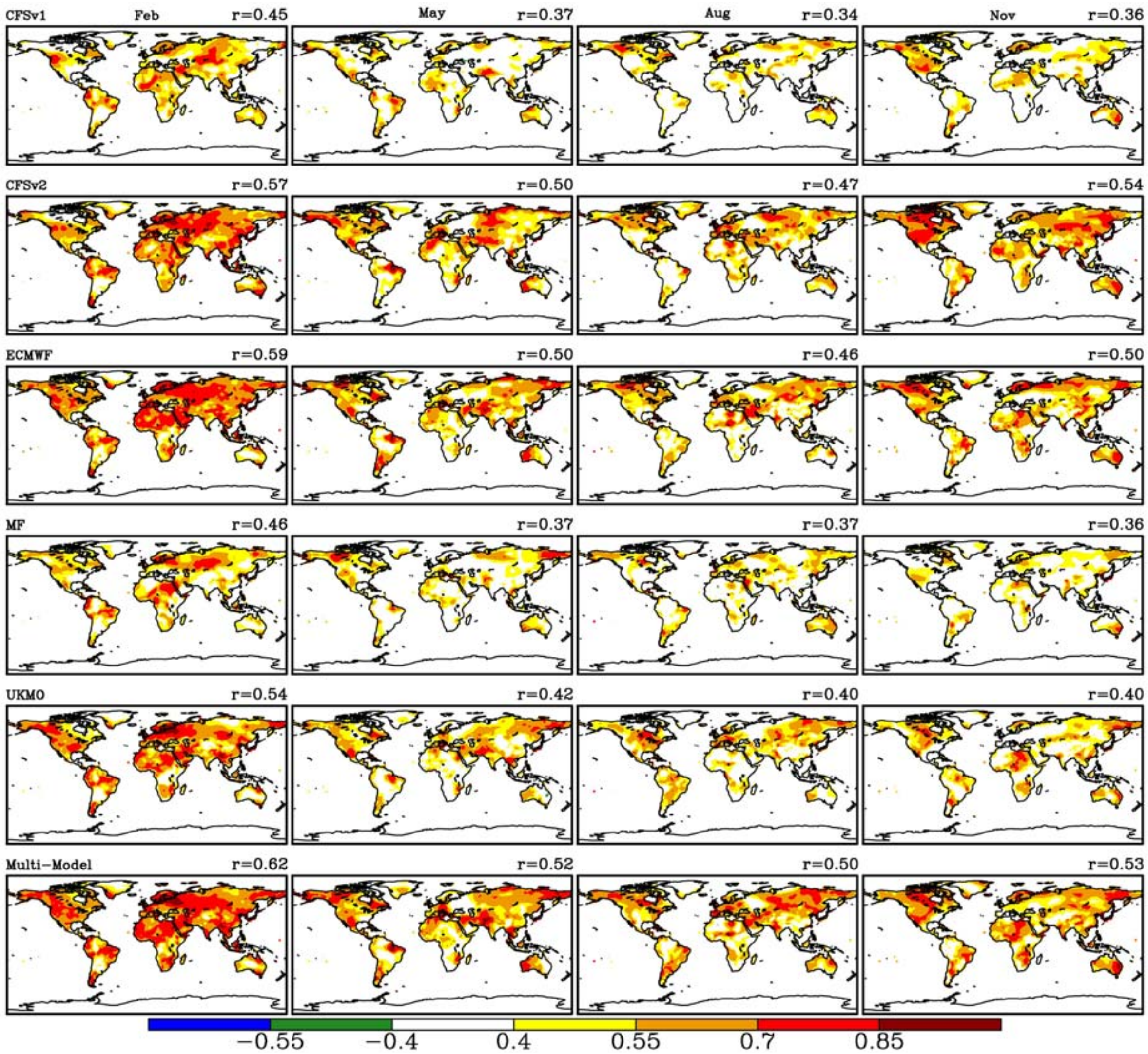

Fig. 1 Month-1 predictive skill of monthly surface air temperature forecasts over land grids in February, May, August and November during 1982-2005. The multi-model is the average among CFSv2, ECMWF, MF and UKMO. Non-white colors represent significant correlation at 0.05 levels. The numbers are the global mean correlations (Yuan *et al.* 2011a).

Figure 2 presents the percentage of the positive RPSS for monthly surface air temperature and precipitation anomaly over the global land area (excluding Antarctic). 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.

Figure 3 shows the correlation between NLDAS-2 observation and downscaled month-1 precipitation forecasts at 1/8 degree over conterminous U.S. in February, May, August and November during 1982-2008 (Yuan *et al.* 2011b). The patterns are similar to the original CFSv1 and CFSv2 forecast results at 2.5 degree (Yuan *et al.* 2011a), where CFSv2 is generally better than CFSv1 especially in cold seasons. For instance, there are obvious improvements over Pacific Northwest, New Mexico-Oklahoma-Texas region and eastern coast for the forecasts in February and November (Figure 3a,d). In May, CFSv2 has higher skill than CFSv1 over southwestern U.S., Ohio basin and Montana, but has lower skill over the southeast (Figure 3b). In the summer time, both CFSv1 and CFSv2 have limited skill, so the CFSv2 has negligible improvement (Figure 3c). On average, CFSv2 increases percentage of grid cells with significant predictive skill (>0.38) from CFSv1 by 61%, 45%, 14% and 60% for the four months, respectively.

Before using the downscaled forecast forcing to provide seasonal hydrologic forecast, a 62-year (1949-2010) offline simulation was performed to generate initial condition and reference data for

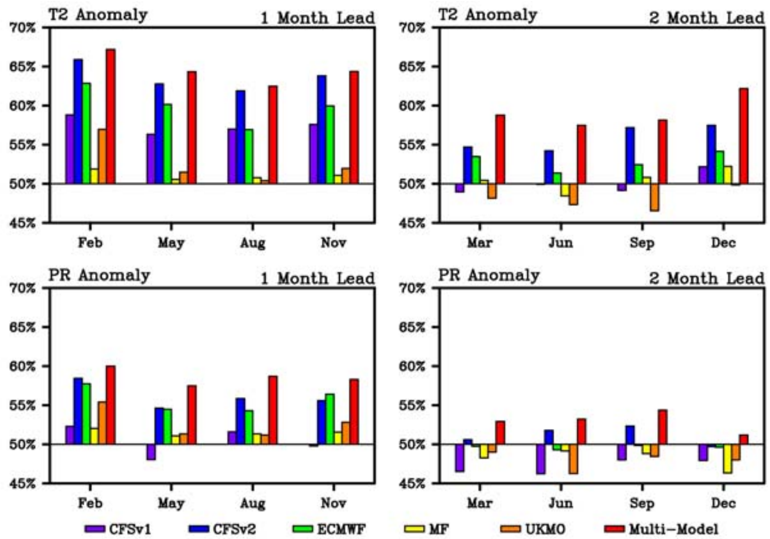

Fig. 2 Percentage of positive Ranked Probability Skill Score (RPSS) for monthly surface air temperature and precipitation anomaly over the global land area during 1982-2005 (Yuan *et al.* 2011a).

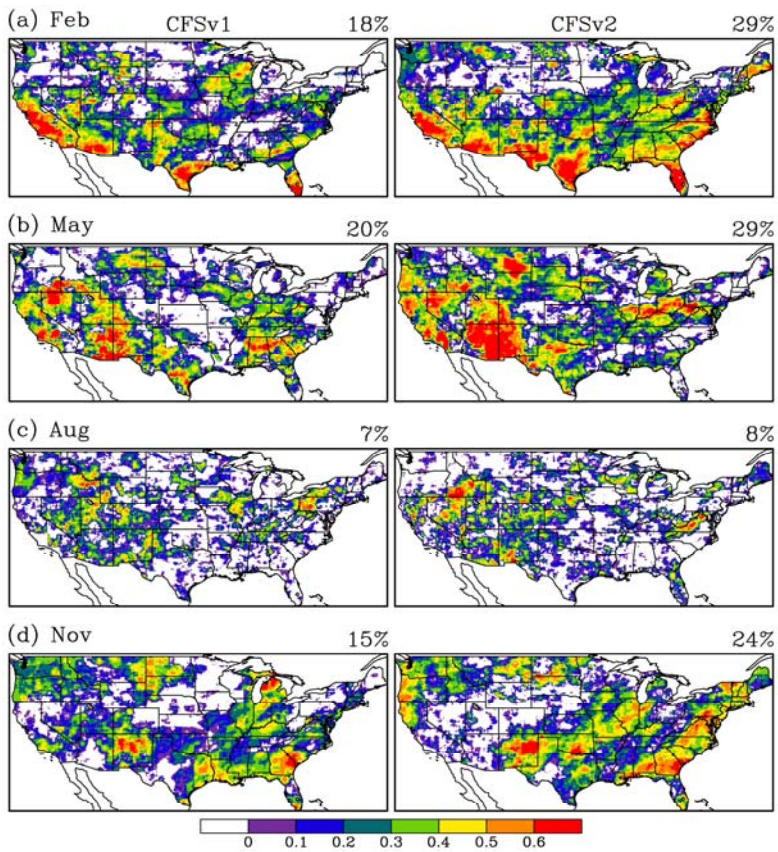

Fig. 3 Correlation between NLDAS-2 observation and Bayesian downscaled CFSv1 and CFSv2 month-1 precipitation forecasts at 1/8 degree over conterminous U.S. in February, May, August and November during 1982-2008. The numbers are percentages of grid cells with significant predictive skill (Yuan *et al.* 2011b).

validation (*e.g.*, soil moisture, runoff), and test the capability of VIC model in capturing the streamflow interannual variations given observed forcing. To match the reforecast period, we calculated the Nash-Sutcliffe efficiency coefficients for monthly streamflow of offline VIC simulations at 416 U.S. Geological Survey (USGS) gauges during 1982-2008. Figure 4 presents locations of the gauges and their corresponding efficiency coefficients (Yuan *et al.* 2011b). Most of the gauges used in this study are in eastern U.S., and the highest coefficients are mainly over Ohio basin, Northeast and lower Mississippi. Among the 416 gauges, there are 315 (75%) and 130 (31%) gauges with coefficients larger than 0.3 and 0.7, respectively. Figure 5 shows interannual variations of streamflow from simulation and observation at four selected gauges with drainage areas ranging from 1027 to 20300 square miles (Yuan *et al.* 2011b). Consistent with high Nash-Sutcliffe coefficients, the VIC model captured seasonal fluctuation of streamflow quite well at the four gauges during 1982-2008, even though we started the model from 1949. Given the limited long-term in situ soil moisture observations over CONUS and the hypothesis that well-calibrated hydrologic model could provide reasonable soil moisture, we used the offline simulated soil moisture as a reference data to validate the hydrologic forecast.

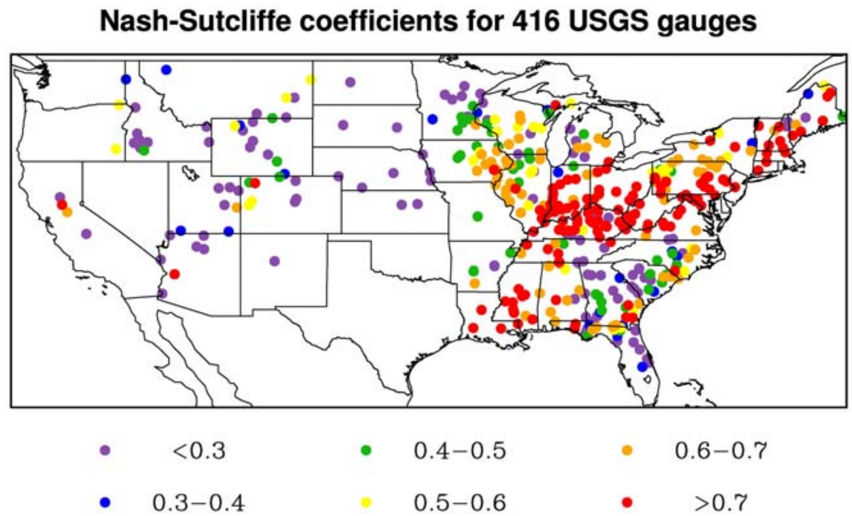

Fig. 4 Efficiency coefficients of monthly streamflow from offline VIC simulations at 416 USGS gauges during 1982-2008 (Yuan *et al.* 2011b).

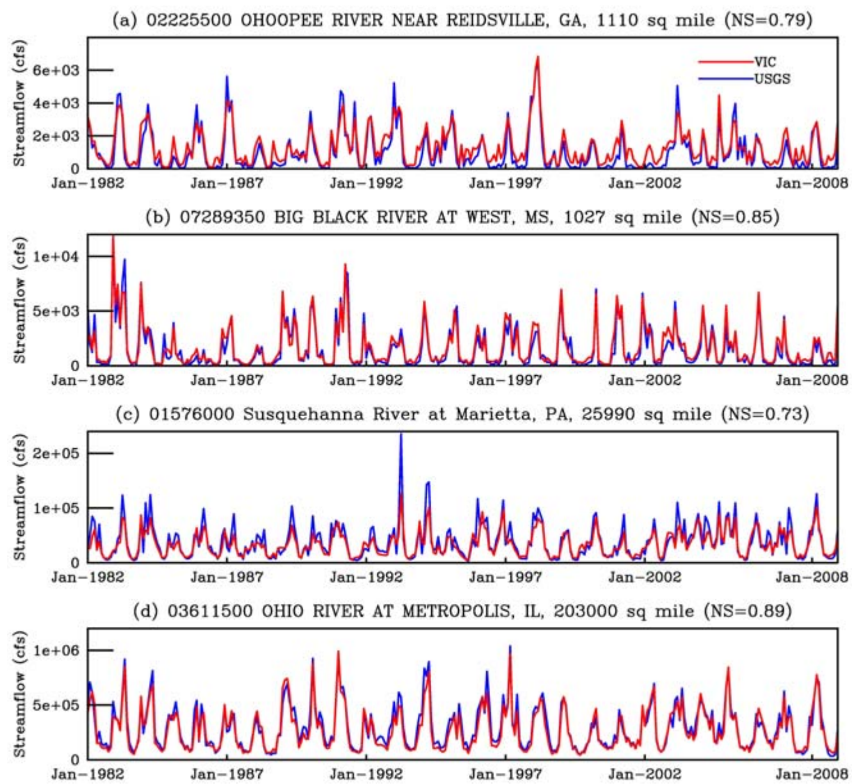

Fig. 5 Interannual variations of VIC offline simulated monthly streamflow compared with USGS observation at four selected gauges (Yuan *et al.* 2011b).

Figure 6 shows the Relative Operating Characteristic (ROC) diagram for low and high flow forecasts (Yuan *et al.* 2011b). The results were calculated for 130 gauges with NS coefficients large than 0.7 (Figure 4). All forecasts were more skillful than random forecast in the first three months. Low flow forecasts were a little better than high flow forecasts in the first month, regardless if the method is ESP or the dynamical climate model; however, their differences diminished beyond month-1. CFSv1 had no advantage in distinguishing low flows over ESP in the first month due to the strong impact on the forecast from initial condition. As the effect of initial conditions decreased and climate model still maintaining some skill, CFSv1 outperformed ESP in the second month;

while their performance became similar again because of the skill decrease in the climate model (Figure 6). With improved skill, CFSv2 was consistently better than ESP in the first three months, and increased the area under ROC curve by 4-7%. For high flows, the climate forecast model-based approach only limited skill beyond ESP in the first month (Figure 6). The moderate advantage manifested the importance of initial hydrologic conditions in identifying low and high flows, and the need of improvement for climate forecasting beyond month-1.

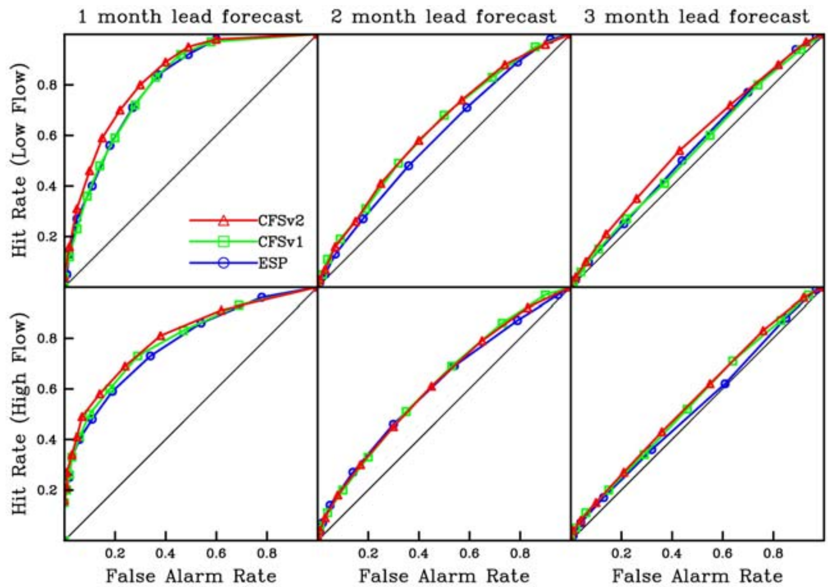

Fig. 6 Relative Operating Characteristic (ROC) diagram for low and high flow forecasts averaged at 130 gauges in the first three months (Yuan *et al.* 2011b).

To investigate the performance for drought severity and duration forecasts, Severity-Area-Duration (SAD) plots were shown in Figure 7 for 3-month duration (Yuan *et al.* 2011b). Severity (S) is defined as $S = (1-\Sigma P/t)*100\%$, where ΣP is the summary of monthly percentile of soil moisture over t months. Given that we were validating drought forecast, the SAD plot was a little different from its traditional way (Andreadis *et al.* 2005). We defined the drought grid cells based on the offline simulation results, no matter whether they were under drought or not in the forecasts. Therefore, the severity was used to quantify the difference between the models' forecasted droughts and the offline simulation conditioning on drought area. The offline simulated severity values were around 0.9 (Figure 7), which was similar to previous studies for 3-month drought duration (Andreadis *et al.*, 2005). Except for 1988, central U.S. drought forecasts had very low skill (Figure 7b) due to the underestimated drought area. CFSv1 and CFSv2 provided more accurate severity values than ESP (Figure 7a,c-d), indicating the added values from climate forecast besides initial conditions. Unlike monthly drought area analysis where the forecast could produce larger drought areas than the offline simulation, the SAD plot demonstrated that the forecasted severity values were generally lower than offline simulation due to the under-prediction of drought areas and/or intensities. For the 1988 drought, all three forecast approaches under-predicted severity compared to the offline simulation by about 40%. Averaged over the other three droughts, ESP underestimated the severity by 31%, CFSv1 by 24%, and CFSv2 by 15%.

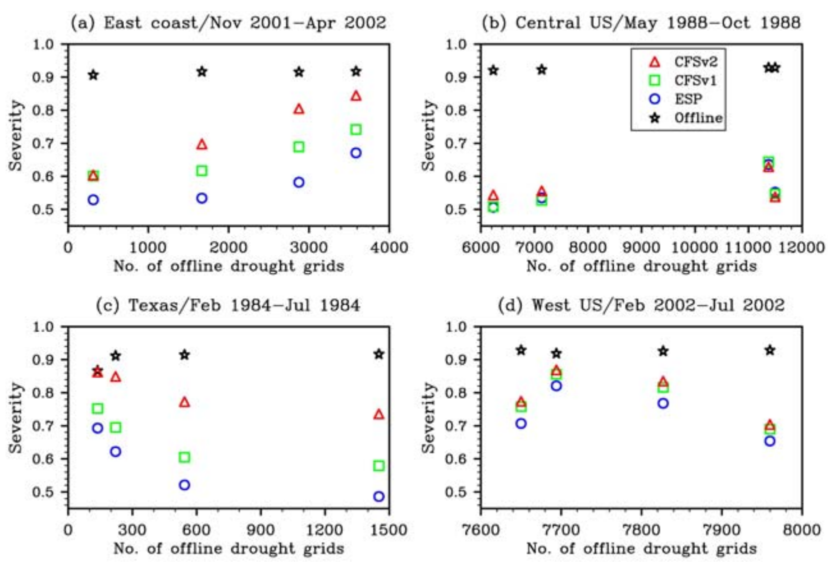

Fig. 7 Severity-Area-Duration (SAD) plots for 3-month duration. The 3-month drought grid cells are identified from offline simulation (Yuan *et al.* 2011b).

4. Conclusion

We provided a first look at the capability of the NCEP's latest operational seasonal forecast model CFSv2 by comparing its hindcast forecast skill with the CFSv1, ENSEMBLES EUROSIP models. CFSv2 shows significant skill enhancement for land surface air temperature and precipitation from CFSv1 for month-1 forecasts, and has comparable result to ECMWF, where the former and the latter have slightly higher skill in temperature and precipitation respectively (Yuan *et al.* 2011a).

CFSv1 and CFSv2-based seasonal hydrologic forecasts were generally more skillful than ESP in the first three months, and CFSv2 increased the skillful forecast percentage from CFSv1 by 10% on average. The climate model-based approach could outperform ESP out to three months in identifying low flow, but not for high flow beyond one month. CFSv2 had better discrimination than CFSv1 for the month-1 streamflow forecast. The offline simulated soil moisture was used to validate short-term drought forecasts. The SAD plots for 3-month duration illustrated the underestimation of drought severity, but the CFSv2 had the least error (Yuan *et al.* 2011b).

Acknowledgements. The research was supported by the NOAA Climate Program Office through Grants NA17RJ2612 and NA10OAR4310246.

References

Andreadis, K. M., E. A. Clark, A. W. Wood, A. F. Hamlet, and D. P. Lettenmaier, 2005: 20th Century drought in the conterminous United States, *J. Hydrometeorol.*, **6**, 985– 1001.

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.

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.

Luo, L., and E. F. Wood, 2008: Use of Bayesian merging techniques in a multimodel seasonal hydrologic ensemble prediction system for the eastern United States. *J. Hydrometeorol.*, **9**, 866-884.

Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen, 2002: A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States, *J. Climate*, **15**, 3237– 3251.

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**, 693-712.

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

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Am. Meteorol. Soc.*, **91**, 1015-1057.

Saha, S. and Coauthors, 2011: The NCEP Climate Forecast System Version 2, to be submitted to *J. Climate* (http://cfs.ncep.noaa.gov/)

Wang, W., M. Chen, and A. Kumar, 2010: An assessment of the CFS real-time seasonal forecasts. *Weather and Forecasting,* **25**(3), 950-969.

Weisheimer, A., and Coauthors, 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.

Xia, Y., K. Mitchell, M. Ek, J. Sheffield, B. Cosgrove, E. F. Wood, L. Luo, C. Alonge, H. Wei, J. Meng, B. Livneh, D. Lettenmaier, V. Koren, Q. Duan, K. Mo, Y. Fan, and D. Mocko, 2011: Continental-scale water and energy flux analysis and validation for the North-American Land Data Assimilation System Project Phase 2 (NLDAS-2), Part I: Intercomparison and application of model products. *J. Geophys. Res.* (submitted).

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

Yuan, X., E. F. Wood, J. K. Roundy, M. Pan, and L. Luo, 2011b: CFSv2-based seasonal hydrologic forecasts over conterminous United States. *J. Geophys. Res.* (to be submitted)