SEASONAL FORECASTING OF GLOBAL HYDROLOGIC EXTREMES
System Development and Evaluation over GEWEX Basins

BY XING YUAN, JOSHUA K. ROUNDY, ERIC F. WOOD, AND JUSTIN SHEFFIELD

A multimodel global hydrologic forecasting system that provides hydroclimate prediction services is evaluated to understand the predictability of seasonal hydrologic extremes over global major river basins.

Persistent hydrologic extreme events such as droughts and wet spells (rainfall anomalies with long durations) have devastating impacts on the human and natural systems and have caused total economic losses of about hundreds of billions of dollars in the United States (Smith and Katz 2013) and around the world (Below et al. 2007). Their occurrences are associated with anomalous atmospheric moisture transport that may be linked to variations of large-scale climate phenomena—for example, El Niño–Southern Oscillation (ENSO), Pacific decadal oscillation (PDO), and Atlantic multidecadal oscillation (AMO)—through ocean–atmosphere teleconnections (Cayan et al. 1999; Hoerling and Kumar 2003; McCabe et al. 2004). Their severities and durations are also influenced by land–atmosphere coupling that can enhance existing extremes (Hong and Kalnay 2000; Schubert et al. 2004). They may be further exacerbated by anthropogenic climate change (Diffenbaugh et al. 2013) and human water consumption (Barnett et al. 2008; Wada et al. 2013).

According to the latest Intergovernmental Panel on Climate Change (IPCC) report, agricultural and hydrological droughts are projected to increase in intensity and duration in presently dry regions by the end of this century under a business-as-usual scenario [representative concentration pathway (RCP) 8.5 scenario], while heavy rainfall events are very likely to increase over most of the midlatitude landmasses and wet tropical regions (IPCC 2013). Such changes impose an increasing risk of hydrologic extremes in the future. Improving understanding of the processes that lead to extremes and establishing operational predictive capabilities that can provide skillful and reliable forecasting of the frequency and intensity of extreme events on regional to global scales are therefore science imperatives of the World

AFFILIATIONS: YUAN—Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, and RCE-TEA, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China; ROUNDY, WOOD, AND SHEFFIELD—Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey
CORRESPONDING AUTHOR: Xing Yuan, RCE-TEA, Institute of Atmospheric Physics, Chinese Academy of Sciences, Huayanli No. 40, Qijiahuozi, Chaoyang District, Beijing 100029, China
E-mail: yuanxing@tea.ac.cn

Climate Research Programme (WCRP) and are being considered as WCRP grand challenges (Karoly 2012; Giorgi et al. 2012).

As a core project of the WCRP, the Global Energy and Water Cycle Experiment (GEWEX; Morel 2001; www.gewex.org) is responsible for facilitating research on quantifying, understanding, and predicting global and regional energy and water variations and extremes through improved observations and modeling of the land, atmosphere, and their interactions, thereby providing the scientific underpinnings of climate services. One of three major research foci of GEWEX is to demonstrate skill in predicting changes in water resources on time scales up to seasonal and annual as an integral part of the climate system [in particular at the regional scale through Regional Hydroclimate Projects (RHP); Coughlan and Avissar 1996; Raschke et al. 1998; Stewart et al. 1998; GEWEX 2012]. Therefore, predicting hydrologic extremes at seasonal scales and investigating their predictability over global major river basins in an integrated hydroclimate forecasting system will help address the WCRP grand challenges and GEWEX RHP research focus.

Seasonal predictability originates primarily from tropical oceans via sea surface temperature (SST) anomalies. The SST anomaly over the tropical Pacific Ocean (i.e., ENSO) has global impacts (Shukla 1998; Goddard et al. 2001), the SST anomaly over the tropical Atlantic Ocean plays a role on the hydroclimate of the Sahel region (Camberlin et al. 2001), and the Indian Ocean dipole (Saji et al. 1999) can contribute to the predictability over Australia, Africa, and southern Asia, somewhat independently from ENSO (Zhao and Hendon 2009; Doblas-Reyes et al. 2013). Other sources of seasonal predictability could come from stratospheric condition (Ineson and Scaife 2009), soil moisture anomaly (Koster et al. 2011), and snow cover (Douville 2010), although these impacts are regional as compared with ENSO. With gradual improvements in observational data assimilation, computing resources that facilitate high-resolution numerical simulations, and the understanding of atmosphere–ocean–land physical processes that account for major seasonal predictability (e.g., ENSO), coupled atmosphere–ocean–land general circulation models (CGCMs) are now widely used for seasonal climate predictions (Weisheimer et al. 2009; Barnston et al. 2012; Kirtman et al. 2014). They have also shown improvement in predictive skill over the past decade, especially for large-scale climate features such as ENSO (Barnston et al. 2012) and the North Atlantic Oscillation (NAO; Scaife et al. 2014).

However, owing to deficiencies in land surface hydrologic parameterizations and/or land surface initializations of CGCMs, the hydrologic forecast products (e.g., soil moisture, runoff) from global seasonal prediction models cannot be directly used for applications. A typical solution is to bias correct the meteorological forecasts from the CGCMs and then drive advanced hydrologic models with refined initial land surface hydrologic conditions to produce seasonal hydrologic forecasts and extreme predictions (Wood et al. 2002; Luo and Wood 2007, 2008; Li et al. 2009; Mo et al. 2012; Yuan and Wood 2012a; Sinha and Sankarasubramanian 2013; Yuan et al. 2013a,b; Mo and Lettenmaier 2014b; Shukla et al. 2014). We refer to this as the CGCM-Hydrology forecasting approach.

Nevertheless, to our knowledge, most CGCM-Hydrology seasonal forecasting studies focus on the soil moisture and/or streamflow prediction over a single basin or continent, usually with a single CGCM (Luo and Wood 2008; Mo et al. 2012; Yuan et al. 2013b; Shukla et al. 2014), while a comprehensive investigation of predictability of global hydrologic extremes in a multi-CGCM framework has not been examined. The multi-CGCM framework not only provides a more reliable assessment of hydrologic predictability but also offers an opportunity to help quantify the uncertainty. Another limitation of previous studies is that most of them assess the forecast skill for extreme indices that blend extreme conditions with normal conditions (Quan et al. 2012; Yoon et al. 2012; Sohn et al. 2013) while ignoring the analysis of hydrologic predictability specifically for individual hydrologic extreme events—for example, the predictive skill for drought onset (Yuan and Wood 2013). Last, theoretical estimates of global hydrologic predictability indicate that initial hydrological conditions provide much of the potential hydrological predictability (Shukla et al. 2013; van Dijk et al. 2013; Yossef et al. 2013), depending on the location and season, but this has yet to be evaluated in terms of actual predictability within the CGCM-Hydrology framework.

This article presents the development and validation of a global seasonal hydrologic forecasting system that is based on multiple CGCMs participating in the North American Multimodel Ensemble (NMME) project (Kirtman et al. 2014) and the Variable Infiltration Capacity (VIC; Liang et al. 1996) land surface hydrologic model. An analysis of droughts and wet spells is carried out over global major river basins, and the CGCM-Hydrology approach is evaluated against the traditional ensemble streamflow prediction (ESP; Twedt et al. 1977) approach, which resamples from the historic record to provide an ensemble of

meteorological forcings and relies on the persistence in initial land conditions to provide forecast skill. The association of the extreme event onset with antecedent oceanic and land conditions is discussed, based on the joint distribution of the forecast and observation.

THE POTENTIAL OF USING NMME FOR HYDROLOGIC APPLICATIONS OVER THE GEWEX HYDROCLIMATE PROJECT BASINS.

During the past four years, the Climate Forecast System, version 2 (CFSv2; Saha et al. 2014), developed by the National Oceanic and Atmospheric Administration (NOAA)'s National Centers for Environmental Prediction (NCEP) has been widely used for hydrologic applications (Yuan et al. 2011; Mo et al. 2012; Quan et al. 2012; Yoon et al. 2012; Yuan and Wood 2012a; Dirmeyer 2013; Yuan et al. 2013b,a; Kumar et al. 2013; Lang et al. 2014; Sheffield et al. 2014; Shukla et al. 2014; Tian et al. 2014). Within the NMME, CFSv2 has been shown to be the most reliable model for seasonal forecasting of global drought onset (Yuan and Wood 2013). But in terms of global gridded analyses, several studies find that combining CFSv2 with other climate forecast models can increase the predictive skill of precipitation (Yuan et al. 2011; Yuan and Wood 2012b; Becker et al. 2014; Kirtman et al. 2014), including extremes (Yuan and Wood 2013).

Here, we target our analysis on major global river basins that are the focus of the GEWEX RHP. Figure 1 shows the continuous ranked probability skill score (CRPSS; Wilks 2011; see appendix for details) for the seasonal mean, basin average precipitation, predicted by CFSv2 and NMME, using all hindcasts started at the beginning of each calendar month during 1982–2009. CFSv2 has 24 ensemble members, while NMME has 71 ensemble members, in total, from six climate models that are producing real-time seasonal climate forecasts. The six models are the National Center for Atmospheric Research (NCAR) Community Climate System Model, version 3 (CCSM3); Geophysical Fluid Dynamics Laboratory Climate Model, version 2.2 (GFDL CM2.2); National Aeronautics and Space Administration (NASA) Goddard Earth Observing System Model, version 5 (GEOS-5); NCEP CFSv2; Canadian Meteorological Centre (CMC) Third Generation Canadian Coupled Global Climate Model (CanCM3); and CMC CanCM4 [see Kirtman et al. (2014) for ensemble information and full references]. Figure 1 shows that NMME has higher probabilistic predictive skill than CFSv2 over 70% of the river basins selected in this study and has comparable skill to CFSv2 for the remaining basins. In particular, obvious improvement can be found over the Amazon and Parana basins in South America, the Colorado
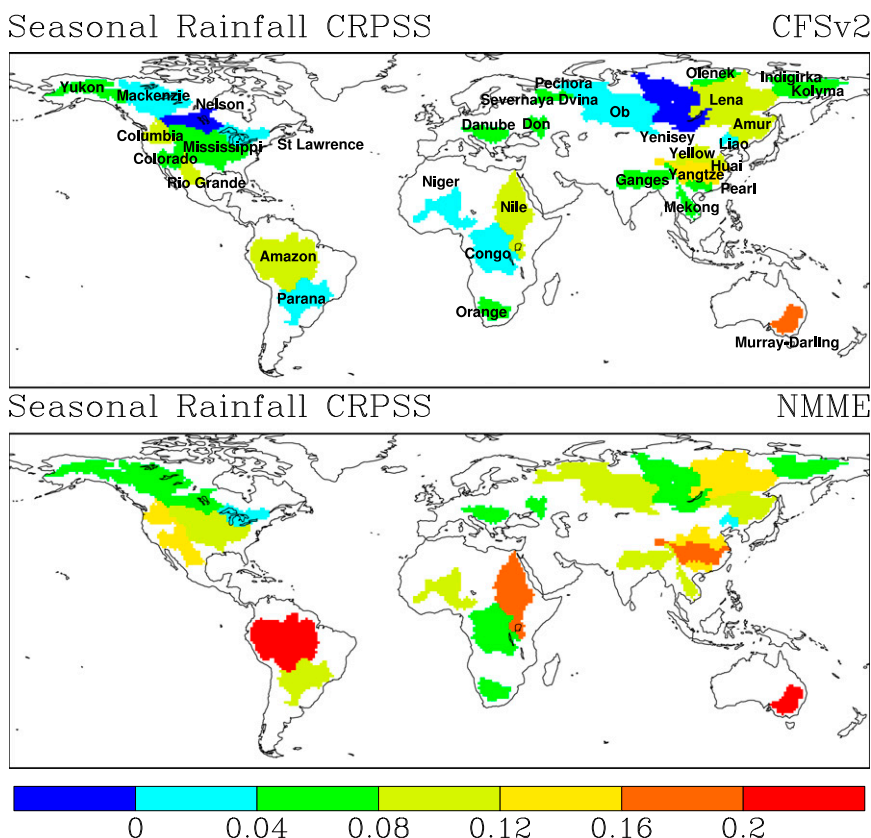

FIG. I. CRPSS for basin-averaged, seasonal mean precipitation (0.5-month lead) from CFSv2 and NMME. The six NMME models are NCAR CCSM3, GFDL CM2.2, NASA GEOS-5, NCEP CFSv2, CMC CanCM3, and CMC CanCM4 [see Kirtman et al. (2014) for ensemble information and full references]. The statistics are based on the hindcasts started in each calendar month during 1982–2009, and the reference is climatological forecast. A value of CRPSS=0.2, for example, indicates that the probabilistic forecast error is 20% less than the climatological forecast error.

basin in North America, the Nile and Niger basins in Africa, the Yangtze and Pearl basins in China, and several high-latitude basins in Eurasia. Figure S1 [find this and more information online (http://dx.doi.org/10.1175/BAMS-D-14-00003.2)] shows the statistics for each season, suggesting that the improvement from CFSv2 to NMME is not necessarily limited to the dry/low flow periods, such as the Yangtze, Pearl, Mekong (East Asia), and Niger (West Africa) basins during June–August (JJA) and the Orange basin (South Africa) and Murray–Darling basin (Australia) during December–February (DJF), where the NMME shows improvement against CFSv2 in their corresponding wet seasons. The improvement over low latitudes originates from better representation of oceanic forcings in other NMME models (Kirtman et al. 2014), while at high latitudes the improvement may be related to the enhancement in reliability from the multimodel ensemble (Yuan and Wood 2013) or a better description of cold season processes. In fact, Fig. 1 in Yuan and Wood (2013) shows that several NMME models (e.g., GFDL CM2.2 and CMC CanCM4) have higher drought onset detectability than the CFSv2 over high latitudes, but the underlying physical reasons for the improvement are unclear. Further diagnosis is needed on the estimation of solid precipitation and the representation of cryospheric processes (e.g., snow and/or frozen soil).

The improvement of NMME against CFSv2 in basin precipitation prediction [more examples can be found in Kirtman et al. (2014)] provides an opportunity to advance the hydrologic forecast over most GEWEX RHP basins. However, the initial condition also has a strong control on the seasonal hydrologic predictability (Shukla et al. 2013; van Dijk et al. 2013), and its spatiotemporal variation can result in quite different hydrologic predictability as compared with the predictability of precipitation. Therefore, this article focuses on a comparison between NMME-based and ESP-based forecasts of hydrologic extremes and explores to what extent the meteorological forecasts from state-of-the-art climate forecast models can improve the hydrologic forecasts relative to the traditional approach that only relies on the information from hydrologic initial conditions.

PRINCETON'S GLOBAL SEASONAL HYDROLOGIC FORECAST SYSTEM.
The global system draws from a legacy of national and continental systems (Luo and Wood 2007; Sheffield et al. 2014) developed by the Terrestrial Hydrology group at Princeton University. Its U.S. Drought Monitoring and Forecast System (Luo and Wood 2007; Yuan et al. 2013b) operates over the conterminous United States (CONUS) at 1/8° resolution and utilizes climate predictions from NCEP's CFSv2 and observations from phase 2 of the North American Land Data Assimilation System (NLDAS-2; Xia et al. 2012a). The system has been transitioned to the NCEP Environmental Modeling Center (EMC) for operational drought prediction (www.emc.ncep.noaa.gov/mmb/nldas/forecast/TSM/perc/). Recently, a continental system for African flood and drought monitoring and
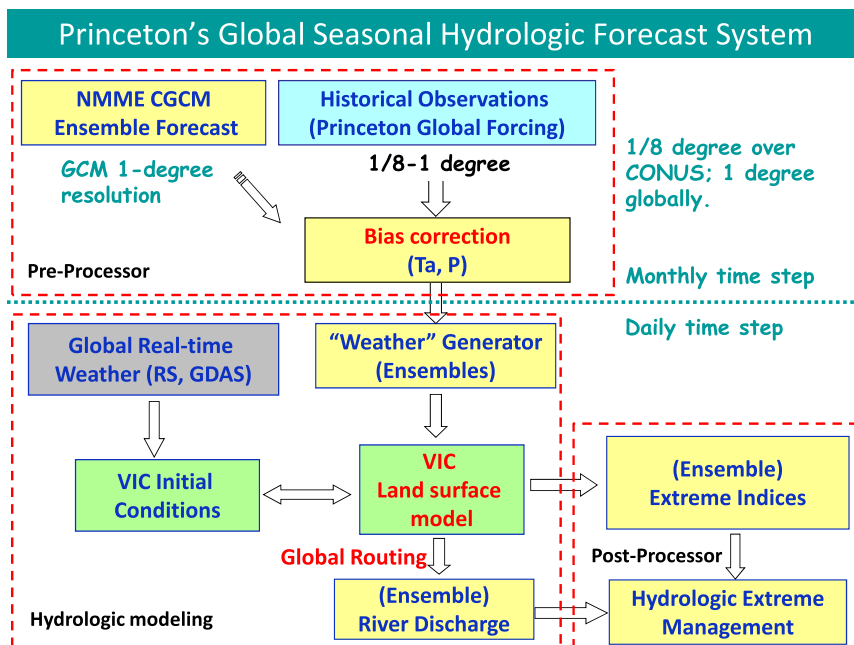

Fig. 2. Schematic view of Princeton's Global Seasonal Hydrologic Forecast System. The system consists of three parts: 1) a preprocessor that bias corrects meteorological ensemble forecasts from global seasonal climate forecast models; 2) VIC land surface hydrologic model calibrated over global major river basins, with a global routing model; and 3) a postprocessor that removes the probabilistic bias of streamflow forecasts and translates hydrologic states to extreme indices. RS = Remote Sensing. GDAS = Global Data Assimilation System.

forecasting (Sheffield et al. 2014; Yuan et al. 2013a) has been developed and installed at regional centers in sub-Saharan Africa.

The central part of a hydrologic forecast system is the land surface hydrologic model. As shown in Fig. 2, the hydrologic modeling part of the global system consists of the VIC land surface model and a global routing model. The VIC model (Liang et al. 1996), version 4.0.5, is used to predict soil moisture and runoff in this study. It is a semidistributed, grid-based hydrologic model with a mosaic representation of land cover and soil water storage capacity. The global routing model utilizes topographic data to derive flow velocity and direction and translates runoff from the VIC grid cells to the river network and routes the flows to the oceans or internal basins, therefore estimating streamflow globally. This first-order approximation of river routing allows for a globally continuous estimate of streamflow that is extremely computationally efficient and therefore can be utilized for both hydrologic monitoring and ensemble forecasting. The VIC model is calibrated over major river basins with 1° global resolution, using monthly streamflow data from Global Runoff Data Centre (GRDC) as compiled by Dai et al. (2009). The stream-flow gauge locations and drainage areas for the basins and length of the records are listed in Table S1 (online supplemental material). The calibration is done using the shuffled complex evolution algorithm (Duan et al. 1994) based on long-term (1952–81) global historical model simulations forced by meteorological data from the Princeton Global Meteorological Forcing dataset (PGF; Sheffield et al. 2006). Except for some basins with heavy water resources management such as the Yellow and Murray–Darling basins, the Nash–Sutcliffe efficiency coefficients for the VIC monthly streamflow simulation vary between 0.6 and 0.9 for the calibration period (1952–81) and 0.5 and 0.8 for the validation period (1982–2006).

A preprocessor component bias corrects the monthly precipitation and temperature hindcasts from each NMME ensemble member using a simple quantile-mapping method (Wood et al. 2002) in a cross-validation mode (leaving the target year out for the climatological distribution). For each calendar month and each NMME model, all hindcasts during 1982–2009 (excluding the target year) with all ensemble members for the target month are used to construct cumulative distribution functions (CDFs) of the forecasts. The 62-yr (1948–2010 excluding the target year) PGF observations in that calendar month are used to construct CDFs of the observations. Both CDFs of forecasts and observations are then transformed to normal probability space through quantile mapping (Wood et al. 2002). Both individual members and NMME multimodel ensemble mean in the normal space are then transformed back to the original space using normal observations as the reference distribution. Finally, the bias-corrected individual members are adjusted according to the bias-corrected NMME multimodel ensemble mean to increase the sharpness (Yuan and Wood 2013). This differs from Princeton's CONUS forecast system, which uses a Bayesian merging method to correct bias (Luo and Wood 2008; Yuan et al. 2013b). The reason is that it is difficult to obtain stable weights for different climate models that can outperform the arithmetic mean from a short sample of hindcast data (Doblas-Reyes et al. 2005; Yuan and Wood 2012b), especially for extremes. The bias-corrected forecasts are temporally downscaled to a daily time step by sampling from the historic PGF dataset and rescaling to match the monthly forecasts ("weather" generator). The daily meteorological forecasts are used to force the VIC model and the river routing model to produce soil moisture and streamflow forecasts, based on initial conditions taken from the historical offline simula-tion. The postprocessor removes the probabilistic bias (Yuan and Wood 2012a) in the soil moisture and streamflow forecasts and calculates extreme indices based on percentile thresholds.

EVALUATION OF HYDROLOGIC HINDCASTS.
The hydrologic hindcasts start from the first day of each calendar month during 1982–2009 and run out to six months, with 71 members for NMME VIC. These are compared with hindcasts based on ESP, which samples random 6-month sequences of daily meteorological data from the PGF dataset that are used to force VIC, to give 20 ensemble members for ESP VIC.

For each calendar month, the monthly soil mois-ture and streamflow are converted into percentiles. Droughts are then defined as monthly soil moisture or streamflow percentiles that are less than 20%, while wet spells are for those larger than 80%. The validation data for soil moisture and streamflow are from the VIC historical simulation.

Figure 3 shows the hit rate of 3-month soil mois-ture (agricultural) drought from ESP VIC and NMME VIC hindcasts at different lead times over the hind-cast period (1982–2009). The hit rate is the fraction of actual events that are predicted (the detectability; see appendix for details). The initial hydrologic condi-tions (as shown by the ESP VIC results) have a strong control on agricultural drought over the Northern

0.5 month lead ESP/3m NMME/3m

1.5 month lead

2.5 month lead

3.5 month lead

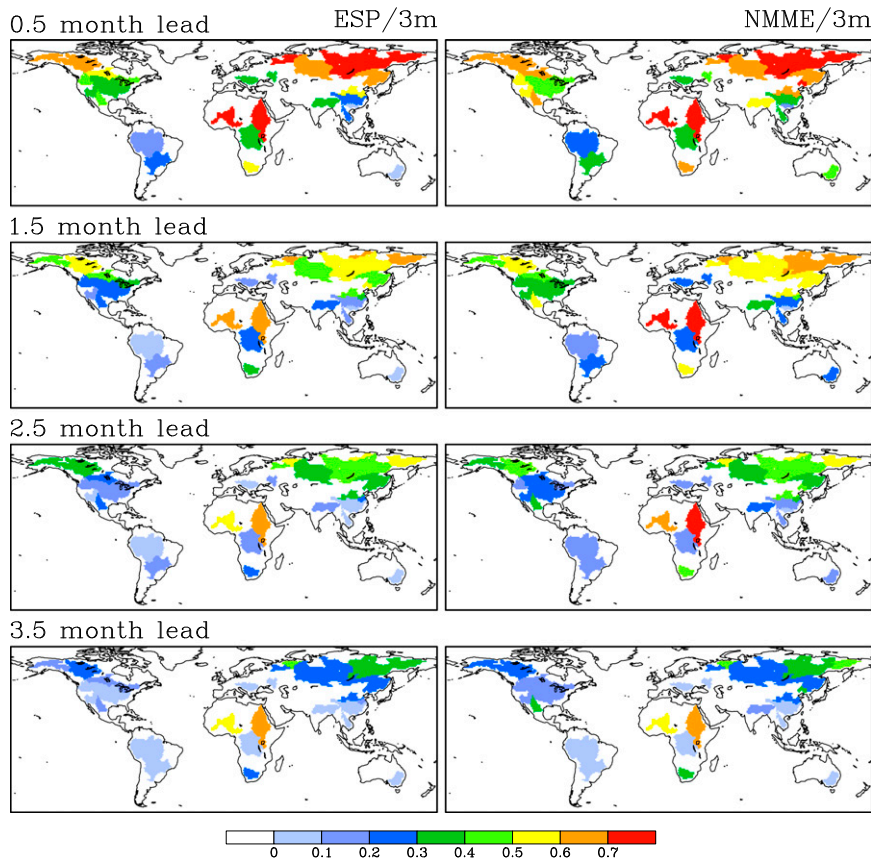0   0.1   0.2   0.3   0.4   0.5   0.6   0.7

FIG. 3. Hit rate for 3-month drought at different forecast leads during the hindcast period (1982–2009). Statistics are calculated using ensemble median soil moisture from all 1° grid cells within each basin. A hit rate of 0.7, for example, indicates that 70% of 3-month drought in the hindcast period can be detected by the forecast system.

Hemisphere, high-latitude basins, and two of the African basins (Fig. 3, left column). The detectability is not negligible (e.g., >0.2) even in some midlatitude basins at short leads. For example, more than 30% of agricultural droughts over the Mississippi basin can be detected by ESP VIC one season ahead (Fig. 3, top-left panel), and the detectability can be as high as 50% during the dry season (Fig. S2, top-left panel). Unlike the precipitation forecast (Fig. 1), NMME VIC's agricultural drought detectability is not necessarily limited to ENSO-affected regions and is actually a compromise between climatic and hydrologic predictability. For instance, NMME has high precipitation predictive skill over the Amazon basin (Fig. 1) because of the ENSO influence, but NMME VIC has low agricultural drought detectability because of lower predictability from the initial hydrologic conditions (Fig. 3). In contrast, NMME VIC has high detectability over high-latitude basins where the influence of initial condition is strong (Fig. 3), regardless of the low precipitation predictive skill

(Fig. 1). As a result, NMME VIC improves drought detectability compared to ESP VIC over basins with moderate control from both remote large-scale oceanic conditions and local initial hydrologic conditions, which are mostly located in midlatitude areas (e.g., Mississippi and Yangtze). For those midlatitude basins, NMME VIC has consistently higher detectability than ESP VIC throughout different seasons, although both methods have higher detectability during dry seasons than during wet seasons (Fig. S2). Similar results can be found for wet spells (not shown), but in general wet spells are less detectable than droughts because of less influence from the initial condition and lower skill in precipitation prediction.

Hit rate (detectability) is one aspect of forecast quality. However, as indicated by Yuan and Wood (2013), the models with high meteorological drought hit rates usually have high false alarm ratios (the fraction of "yes" forecasts that turn out to be wrong; see the appendix for details). Therefore, a balanced index, the equitable threat score (ETS; Wilks 2011; see appendix) that considers both the hit rate and the false alarm ratio is used to quantify the contribution of NMME beyond ESP for the predictive skill of soil moisture extremes. Figure 4 shows the maximum forecast leads for which NMME VIC has significantly ($p < 0.05$) higher ETS than ESP VIC for the prediction of droughts and wet spells. The Student's t test is used here, and samples come from ETS for each grid cell within the basin. For droughts that last for at least 1–2 months, NMME VIC prediction is significantly better than ESP VIC up to 3–6 months over the basins with short soil moisture memory, such as tropical and midlatitude basins (Figs. 4a,b). For the basins with long soil moisture memory, NMME VIC outperforms ESP VIC only if the NMME has high predictive skill for the precipitation (CRPSS > 0.1; Fig. 1), such as the Nile

basin in Africa. For the Niger basin in West Africa with long soil moisture memory, NMME only has moderate precipitation predictive skill (Fig. 1), and so NMME VIC does not have significantly higher ETS than ESP VIC (Figs. 4a,b). For 3-month droughts, NMME VIC also has significantly higher ETS than ESP VIC up to 2–3 months for more than 40% of the basins.

The differences for wet spells (Figs. 4d–f) are much smaller than for droughts, which are expected because of lower skill of the climate models in predicting wet conditions. Similar to drought, the differences between NMME VIC and ESP VIC usually decrease with an increase of the wet spell duration. Some of the exceptions to this are the Columbia and Colorado basins in North America, and the Murray–Darling in Australia, where NMME's advantage emerges as the duration increases.

The results of maximum forecast leads for individual seasons are shown in Fig. S3. Because of the insufficient number of samples for the 2- and 3-month droughts or wet spells, only the results for 1-month duration are plotted. Regardless of the season, NMME VIC drought prediction is significantly better than ESP

VIC up to six months over the Mississippi basin. Even for the wet spells, the NMME VIC prediction has significantly higher skill than ESP VIC up to 4–5 months over the Mississippi during wet seasons [March–May (MAM) and JJA].

Given that several seasonal forecast systems based on CFSv1 or CFSv2 have been developed over the United States (Luo and Wood 2008; Mo et al. 2012; Yuan et al. 2013b), Fig. S4 compares the NNME/VIC system with the CFSv2/VIC for drought prediction. Note that CFSv2/VIC in this study has the same resolution as NMME VIC (i.e., 1°), which is coarser than previous studies. Actually a high-resolution (1/8°) NMME VIC system is currently being developed at Princeton University by following the work of Yuan et al. (2013b). Similar to Fig. 4, Fig. S4 shows the maximum forecast leads when CFSv2/VIC is used as the reference forecast. There is no significant difference over the Columbia basin. The NMME VIC prediction is significantly better than CFSv2/VIC up to two months over the Colorado basin for 1- and 2-month droughts and over the Mississippi basin for 2-month droughts.
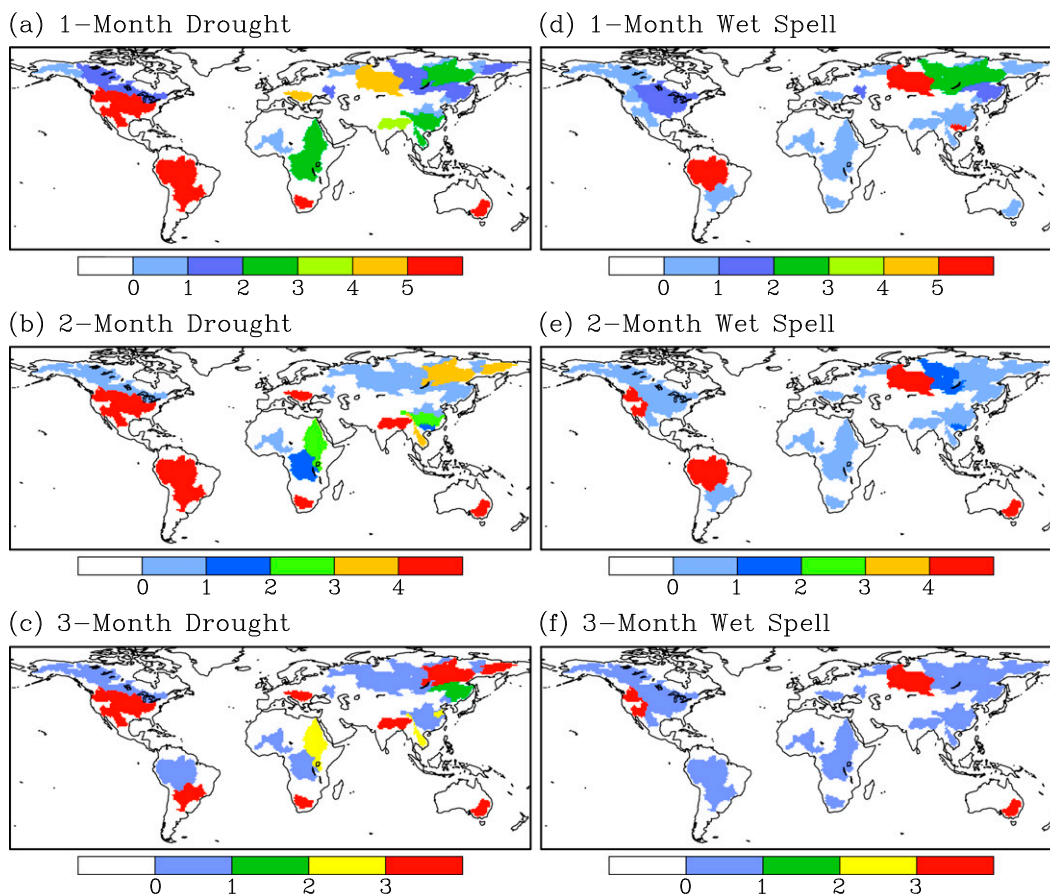

(a) 1−Month Drought (d) 1−Month Wet Spell

(b) 2−Month Drought (e) 2−Month Wet Spell

(c) 3−Month Drought (f) 3−Month Wet Spell

Fig. 4. Maximum forecast lead time (months) that NMME VIC ensemble median has significantly ($p < 0.05$) higher *ETS* than ESP VIC for the prediction of soil moisture droughts and wet spells during the hindcast period (1982–2009).

(a) 0.5 month lead (b) 1.5 month lead
(c) 2.5 month lead (d) 3.5 month lead
(e) 4.5 month lead (f) 5.5 month lead

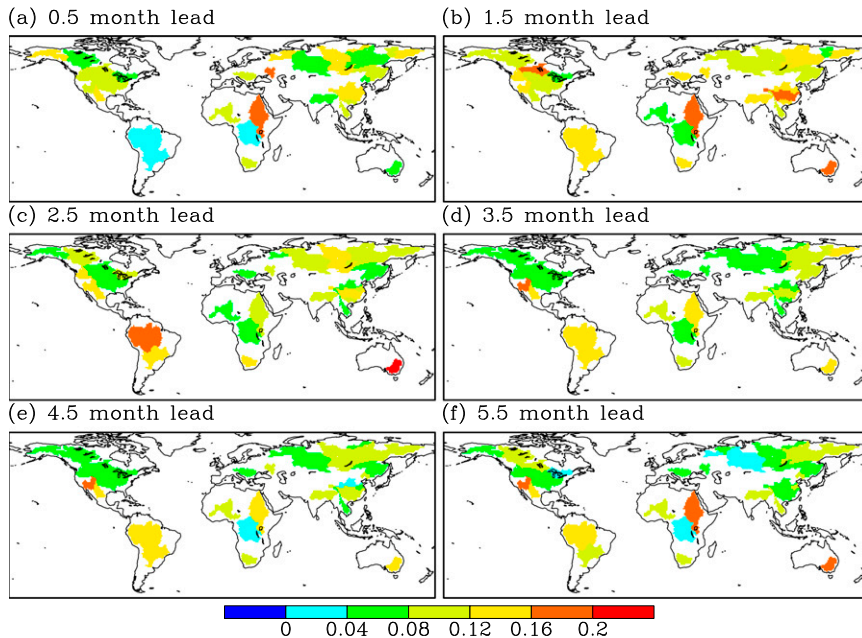0   0.04   0.08   0.12   0.16   0.2

FIG. 5. CRPSS for monthly streamflow predicted by NMME VIC at the outlets of each basin during the hindcast period (1982–2009). The reference forecast is ESP VIC, and VIC offline simulated streamflow is used as validation data. A value of CRPSS=0.2, for example, indicates that the probabilistic streamflow forecast error for NMME VIC is 20% less than the forecast error for ESP VIC.

The streamflow forecasts at the outlet of each basin are also assessed. Figure 5 shows the CRPSS of monthly streamflow predicted by NMME VIC, using ESP VIC as the reference forecast, and Fig. S5 shows the results for different seasons. NMME VIC is generally more skillful than ESP VIC over most regimes. However, the biggest improvement does not necessarily occur in the first month of the forecast (e.g., Amazon, Yangtze, and Murry–Darling). Again, this is another example of the compromise between predictive skill derived from the climate forcing and the initial hydrological conditions. The effects of initial condition on streamflow over these large basins are expected to be larger than the effects on grid-scale runoff because of the memory from upstream areas. This sometimes results in a negligible difference between NMME and ESP skill for the streamflow forecast at the beginning of the forecast, despite that the precipitation predictive skill of NMME is significantly higher than ESP at short leads.

Figure 6 shows the probabilistic forecast quality for low, normal, and high flow conditions, using hindcast samples from all basins. ESP VIC has reliable predictions (results fall along the diagonal lines) for hydrologic droughts and wet spells (red and blue lines, respectively) at 0.5-month lead, but it becomes overconfident as the forecast proceeds. Although the ESP forcings (climatological precipitation and temperature) are very reliable, they are not necessarily

reliable for predicting hydrologic extremes. This is especially true for forecasts that start with anomalously dry or wet initial conditions, for which the ensemble of historical forcings tend to bring the hydrological states to neutral conditions, thus degrading the reliability for extremes. In contrast, NMME VIC maintains the reliability for drought and wet conditions much better, even out to 5.5-month lead. Similar to the 2-m temperature forecasts illustrated in Kirtman et al. (2014), the reliability for neutral streamflow conditions is more difficult to maintain for longer forecast lead times (Fig. 6, gray lines).

REAL-TIME FORE-CASTING OF THE 2012 CENTRAL U.S. DROUGHT. The 2012 summertime drought over the central United States was the most severe seasonal drought in the past 100 years (Hoerling et al. 2014). Most seasonal climate forecast models including CFSv2 failed to predict well the meteorological drought (Kumar et al. 2013; Hoerling et al. 2014). However, some NMME models such as GFDL CM2.2 did capture the 2012 drought (Kam et al. 2014). Kirtman et al. (2014) also showed that a NMME-based 6-month standardized precipitation index that blends antecedent observations with the seasonal forecasts had some skill for the 2012 drought. Here we test the capability of the NMME-based hydrologic forecast system in predicting the 2012 agricultural drought as an extension of the previous results for meteorological drought forecasts (Hoerling et al. 2014; Kirtman et al. 2014).

The approximate "real-time" forecast is done by bias correcting the NMME climate forcings for 2012 using all hindcast data during 1982–2009, which differs from the cross-validation mode of leaving out the target year during the hindcast period. Real-time observational data are used to run the VIC model up to the start of the forecast to produce the initial conditions. The real-time observational data are taken from the Climate Prediction Center (CPC) Unified Gauge-Based Analysis for precipitation (Chen et al. 2008) and the Climate Forecast System Reanalysis (CFSR; Saha

et al. 2010) for other meteorological variables. These are used to extend the PGF data after 2010 and are adjusted to match the monthly climatology (1948–2010) of PGF through quantile mapping. The bias-corrected data are then used to force the VIC model from 2011 to 2012 to generate the initial conditions as well as reference soil moisture data to evaluate the forecasts. Note that an operational real-time forecast would be subject to biases in the real-time meteorological forcings that are likely to be high in regions with sparse gauge networks.

Figure 7 shows the 6-month soil moisture drought area forecasts for 20 ESP VIC members and 71 NMME VIC members initialized on two dates: February 2012 and June 2012. Before the drought onset, ESP VIC has some skill in the first two months (February and March forecast in Fig. 7a). After March, almost all ESP VIC ensemble members underestimate the drought area during 2012, especially during the summer when the drought is quite severe. The NMME VIC grand ensemble encompasses the evolution of the reference drought area (solid black line) quite well (Fig. 7b), and the ensemble mean (blue line) is also much closer to the reference drought area than ESP VIC. Nevertheless, the ensemble mean of NMME VIC shows an earlier drought recovery than the reference data, which indicates the difficulty of predicting drought recovery.

Besides evaluating the 2012 drought forecast with the VIC offline simulation, we also compared the forecast with two satellite-based estimates of the drought: the multidecadal (1979–2013) essential climate variable for soil moisture (ECV_SM) dataset that homogenizes and merges six microwave-based satellite soil moisture retrievals (Liu et al. 2011; Dorigo et al. 2015) and the 11-yr (2003–13) Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage dataset (Wahr et al. 2004). Figure 7 shows that the ECV_SM (dashed black lines) matches the VIC offline simulation (solid black lines) quite well before the drought onset, but they diverge slightly as the drought emerges. This difference can be attributed in part to the representative depth of the satellite soil moisture retrieval, which is for a very thin surface layer (~1 cm) due to the frequency of the sensors, and to the larger errors in more densely vegetated regions for the retrieval. The GRACE data (plus symbols) have larger seasonal variations and show a larger drought area than both the VIC simulation and the ECV_SM, because it represents changes in total water storage that includes surface water bodies (e.g., lakes and reservoirs) and groundwater, which are not represented by other datasets. Additionally, GRACE has a short climatology (11 years) that contributes to a larger uncertainty in its seasonal climatology from which the percentiles

are estimated. Nonetheless, the satellite data provide useful information for validating the hydrologic forecasts, especially if their corresponding time scales and uncertainties are well understood.

OCEANIC AND LAND PRECURSORS FOR HYDROLOGIC EXTREMES.
Because of chaotic nature of the atmosphere, seasonal prediction relies heavily on the memory imparted by both the ocean and land, as do the predictions of hydrologic extremes. For instance, ENSO is recognized as the largest source of
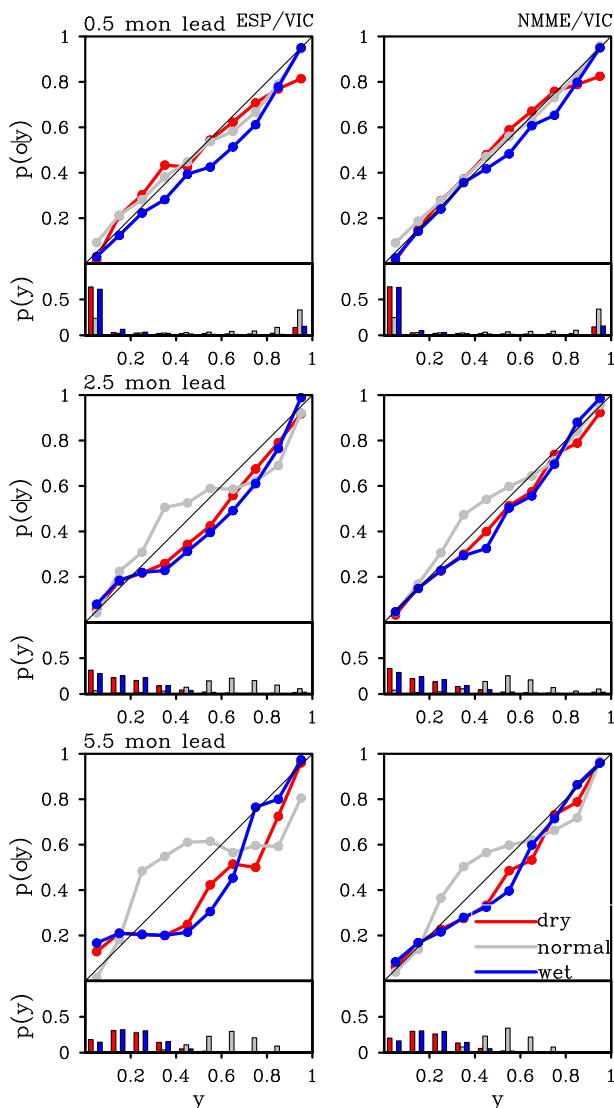

FIG. 6. Reliability diagram for monthly streamflow percentiles predicted by (left) ESP VIC and (right) NMME VIC throughout global major river basins during the hindcast period (1982–2009). Red, gray, and blue are for dry (<20%), normal, and wet (>80%) conditions, respectively. Here p(y) is forecast probability with probabilistic forecast value y, and p(o|y) is the observed probability conditional on the forecast value.

seasonal predictability, and tropical SST anomalies not only alter the Walker circulation and convection in the tropics because of the positive feedbacks between SSTs and wind (Walker and Bliss 1932; Smith et al. 2012) but also affect the climate in midlatitudes through Rossby wave trains (Hoskins and Karoly 1981; Trenberth and Caron 2000). To investigate the impact of ENSO on hydrologic prediction over the GEWEX basins, differences in composite soil moisture percentiles between selected El Niño and La Niña years (i.e., average soil moisture percentiles in El Niño years minus those in La Niña years) during 1982–2009 are shown in Fig. 8. The selected years are according to Smith et al. (2012), which are classified by using a detrended 100-yr SST time series. NMME VIC reproduces the ENSO influence on soil moisture during wintertime very well (Figs. 8a,b) but underestimates its impact over the North American monsoon and East Asian monsoon regions during summertime (Figs. 8e,f). In general, the responses of seasonal soil moisture

to ENSO are roughly captured by the CGCM-Hydrology forecast system. There are moderate differences for the summertime composite among individual models: GFDL CM2.2 and two Canadian models are better over North American basins, while NCEP CFSv2 is better over East Asian basins (not shown).

Besides ENSO, the initial soil moisture is also thought to influence both the potential and actual subseasonal to seasonal climate predictability via land–atmosphere coupling (Koster et al. 2004, 2006, 2010). However, the association of oceanic and land precursors with model performance for individual hydrologic extreme events at the global scale is still unclear. Here we investigate the ENSO and soil moisture associations based on the joint distribution of the forecast and observation for the onsets of droughts and wet spells. We would like to answer the following question: What are the probability distributions of antecedent oceanic and land conditions for hit cases (observed extreme events that are captured by the models) and for false

alarms (forecasted extreme events that do not occur in the observation)? The onset events of droughts or wet spells are defined as three continuous months when the soil moisture percentile is consistently below 20% or above 80%, respectively.

Figure 9 shows the spatial frequency of conditional mean, antecedent Niño-3.4 SST absolute anomaly and the initial soil moisture percentiles calculated over the GEWEX basins. For example, the green lines (from six NMME models) in Fig. 9a represent the frequency distributions of seasonal mean Niño-3.4 SST (three months before the onset of extreme events) averaged over those forecast events where the models issue a soil moisture drought onset forecast (fcst = T) but drought does not occur in the observation (obs = F), where T and F represent that drought occurs or does not occur, respectively, either for the
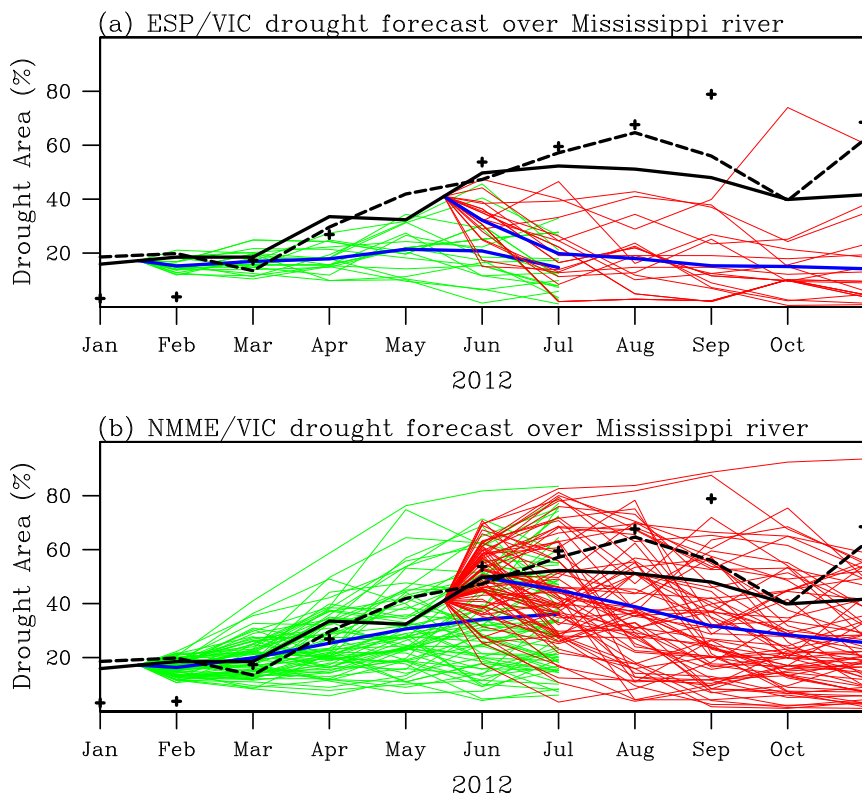

FIG. 7. Real-time seasonal forecasting of the 2012 drought over the Mississippi River basin with ESP VIC and NMME VIC. The drought area is quantified using soil moisture percentiles. Solid black lines represent the VIC offline simulations, dashed black lines are from the ECV-merged satellite soil moisture retrievals, plus symbols are for the GRACE terrestrial water storage anomalies (data for May and Oct are missing), blue lines are the ensemble means of the forecasts, and the green and red lines are for individual ensemble members.

forecast (fcst) or observation (obs). Red and blue curves are for detected and missed events, respectively. The higher peaks of the blue curves (fcst = T) compared to those of the green and red curves (fcst = F) around small SST anomaly values indicate that climate models have a higher chance of missing the agricultural drought onset when the antecedent SST anomaly is smaller (Fig. 9a). As the SST anomaly increases, climate models have a higher chance of issuing a drought forecast than missing a drought (i.e., the red and green curves show higher frequency than the blue curves), but there is also a higher chance of a false alarm (green curves). This is similar to the meteorological drought analysis in Yuan and Wood (2013). The association for the wet spell onset forecast (Fig. 9b) is similar to the drought onset, but different models have moderate differences for the false alarms (green curves).
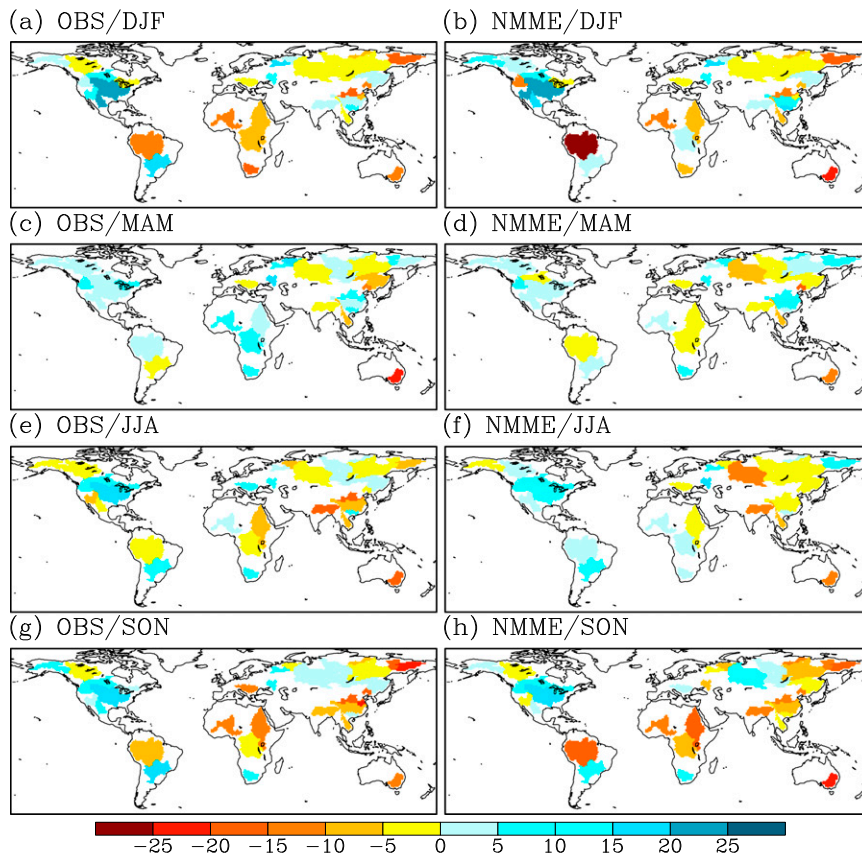

Fig. 8. Composite differences in offline simulated (OBS) and model-predicted (NMME, 0.5-month lead) soil moisture percentiles between positive and negative phases of ENSO for different seasons. Gridpoint soil moisture percentiles are averaged to basin scale before composite analysis. During the hindcast period (1982–2009), positive ENSO years are 1982, 1986, 1991, 1997, and 2009 and negative ENSO years are 1984, 1988, 1999, and 2007.

This asymmetric performance for predicting soil moisture droughts and wet spells is more obvious in the analysis of land precursors (Figs. 9c,d). The spread of initial soil moisture percentiles for model-predicted drought events (red and green curves, Fig. 9c) is smaller than that for wet spell events (Fig. 9d), suggesting that there is less dependence of wet spell onset forecast on the initial land conditions than for the drought onset. In fact, as the drought occurs, initial soil moisture memory (anomaly) could persist for a period of time, while when a wet spell occurs, an individual rainfall event can sometimes erase all soil moisture memory. This interacts with the atmospheric asymmetry mentioned above and amplifies the difference between dry and wet conditions. Figures 9c,d also demonstrate that the missed drought (wet spell) events (blue curves) are associated with higher (lower) initial soil moisture (i.e., less information from the land precursor). Therefore, some droughts and wet spells occur without clear SST and soil moisture precursors (e.g., the 2012 central U.S.

drought), and they are the most difficult to predict at seasonal time scales. The differences in the red and green curves in Figs. 9c,d are larger than those in Figs. 9a,b, suggesting that the oceanic precursor facilitates a higher probability that the climate models issue an extreme forecast (although it is sometimes difficult to determine whether the forecast is correct or a false alarm), while the land precursor will reduce the probability of false alarms.

BEYOND THE NORTH AMERICAN MULTI-MODEL ENSEMBLE. Phase 2 of the NMME project (Kirtman et al. 2014) will provide higher-temporal-resolution datasets (e.g., three hourly) with more variables besides precipitation, 2-m surface air temperature, and SST. This will enable a more comprehensive diagnostic study that can provide feedback to model development. Nevertheless, in terms of applications (e.g., hydrologic forecasts), there are other concerns. As pointed out by Yuan and Wood (2012b), six of the seven

original NMME models (without the two Canadian models) use the ocean model developed at GFDL, which to some extent may result in similarity or overconfidence in the seasonal climate forecasts. While combining those seven models does not gain much predictability in terms of a deterministic forecast, skill can be increased by including European models, which are considered to be more independently developed (Yuan and Wood 2012b).

To explore this, we briefly evaluate the benefit of combining the NMME models with those in the Climate-System Historical Forecast Project (CHFP; Kirtman and Pirani 2009). Figure 10 shows the CRPSS for basin-averaged, May–July (MJJ) mean precipitation predicted by NMME and NMME+CHFP. Because of data availability during 1982–2009, the CHFP models used here only include the European Centre for Medium-Range Weather Forecasts (ECMWF) Seasonal Forecast System 4 (S4; Molteni et al. 2011; Dutra et al. 2013), two models from Japan [Meteorological Research Institute Coupled Atmosphere–Ocean General Circulation Model, version 3 (MRI-CGCM3) and Model for Interdisciplinary Research on Climate, version 5 (MIROC5)], one model from Germany [Max Planck Institute Earth System Model (MPI-ESM)], and one from Australia [Predictive Ocean Atmosphere Model for Australia (POAMA-2)]. Figure 10 shows

that including CHFP does not improve the MJJ precipitation prediction over the basins in North America (except for Columbia, but MJJ is a transition season between the snow-dominated winter and mostly dry summer), suggesting that the NMME is the best multimodel ensemble in predicting hydroclimate over North America. Nevertheless, improvement over the basins in East Asia and Australia is not negligible (CRPSS difference larger than 0.05) in basins such as the Yangtze, Mekong, Ganges, and Murray–Darling. Therefore, increasing the number of international models (and presumably the level of international collaboration) may be necessary to advance hydrological forecasting at the global scale.

An alternative multimodel ensemble is the ensemble of multiple land surface hydrologic models. In fact, work over the past decade has shown that hydrological models, even when forced with identical atmospheric boundary conditions, can produce results that are substantially different (Dirmeyer et al. 2004; Mitchell et al. 2004; Duan et al. 2007; Wang et al. 2009; Xia et al. 2012b). Recently, Mo and Lettenmaier (2014a) have reported the challenge of applying hydrologic models in monitoring the droughts with different severity categories, and Nijssen et al. (2014) have developed a prototype global drought information system based on multiple land surface models. These studies suggest that augmenting the NMME-based multimodel hydrologic forecasting system with multiple land surface models would enhance its capability in handling the hydrologic extremes with different severity levels.

CONCLUDING RE-MARKS. A global seasonal hydrologic forecasting system based on the NMME climate forecast models and VIC land surface hydrologic model has been established, and its performance against the traditional ESP forecast approach in predicting droughts and wet spells is assessed over the GEWEX RHP basins for a 28-yr hydrologic hindcast experiment and a "real time" case study. The ESP forecast skill relies on the information
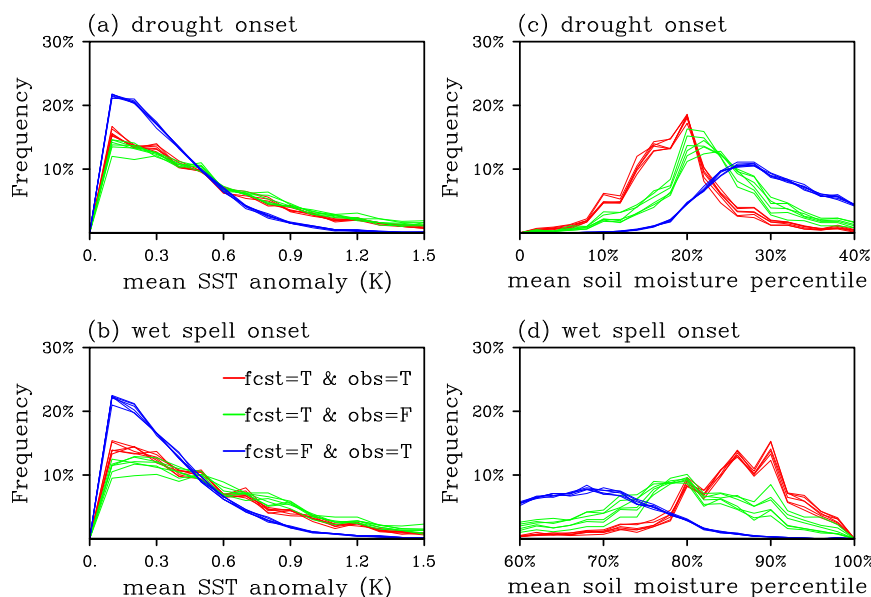
FIG. 9. Frequency distributions of (a),(b) absolute anomaly of antecedent seasonal mean Niño-3.4 SST and (c),(d) initial soil moisture percentile conditional on the joint distribution of forecast and observation for soil moisture drought and wet spell onset events over GEWEX basins. SST absolute anomaly and soil moisture percentile are averaged among all drought/wet events for each grid cell, and spatial frequency is then calculated using averaged values from all grid cells within GEWEX basins.

from the initial hydrological conditions, and so the comparison between the output of the NMME VIC and ESP VIC provides an opportunity to quantify the origin of hydrological predictability from the ocean and land states.
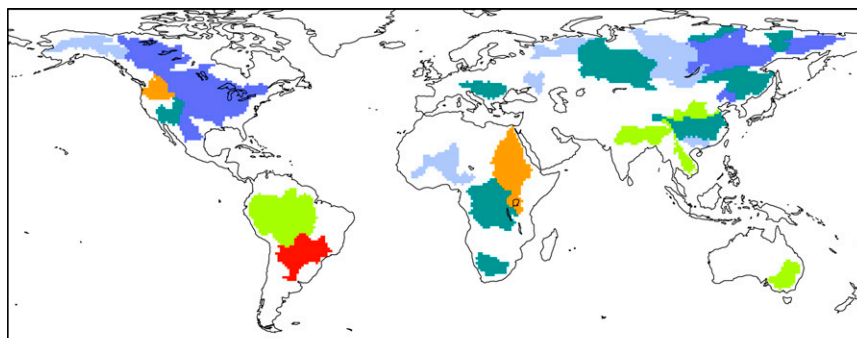
NMME VIC improves drought detectability against ESP VIC mostly over midlatitude basins where the controls of both remote large-scale oceanic states and local initial hydrological conditions are moderate. It is found that NMME VIC has significantly ($p < 0.05$) higher ETS values than ESP VIC up to 3–6 months over basins with short soil moisture memory. In terms of the streamflow forecasts, the NMME VIC is superior to ESP VIC for accuracy and reliability, and the biggest improvement does not necessarily occur in the first
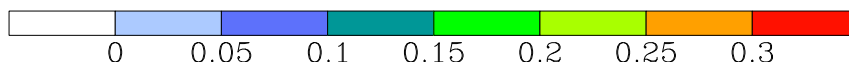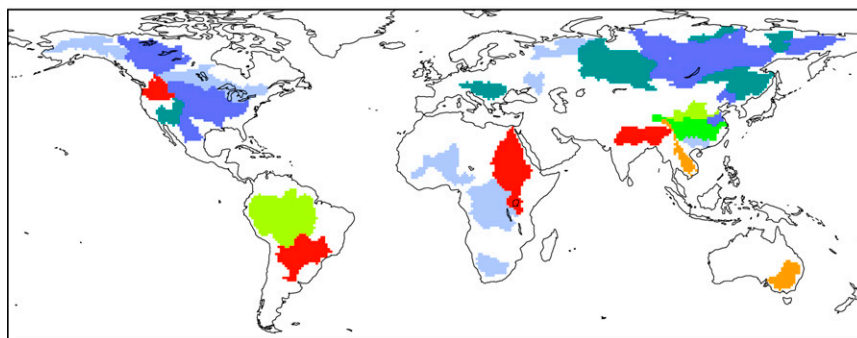

Fig. 10. CRPSS for basin-averaged MJJ mean precipitation (0.5-month lead) from NMME and those combined with CHFP models. The statistics are based on the hindcasts started in each May during 1982–2009, and the reference is climatological forecast.

month of the forecast. A real-time forecasting of the 2012 central U.S. drought shows that none of the ESP VIC ensemble members is able to forecast the drought onset; however, the NMME VIC grand ensemble covers the evolution of drought area quite well, with an ensemble mean closer to the reference data. The association of the onsets of extreme hydrologic events with oceanic and land precursors is also investigated on the basis of the joint distribution of the forecast and observation. Climate models have a higher probability of missing the onset of hydrologic extremes when the antecedent SST anomaly is smaller. Larger SST anomalies offer a higher probability for the models to issue a forecast for extremes but also bring higher probability of issuing a false alarm. The probability of such a false alarm can be reduced if there is a large anomaly in land surface conditions.

Overall, the global hydrologic forecast system established in this study shows encouraging performance when compared with the ESP approach for predicting hydrologic extremes, such as higher detectability for historical soil moisture droughts, more reliable streamflow ensemble forecasts for low or high flow

conditions, and better prediction for the 2012 North American extreme drought in a real-time forecast mode. The system also shows the potential for successfully utilizing climate models to advance GEWEX RHP. A website is being established to make real-time hydrological forecasts available, drawing from the existing Princeton CONUS and African monitoring and seasonal forecast websites (http://hydrology.princeton .edu/forecast; http://hydrology.princeton.edu/adfm). Linking the real-time forecasting of soil moisture and streamflow with impact models for predicting reservoir inflow, crop yield, and wild fire, etc. will amplify the usefulness of the system. Therefore, the NMME VIC system can serve as a prototype system for the Global Framework for Climate Services (GFCS), both in providing hydroclimate information services and in contributing to the science underpinning the prediction and predictability of the terrestrial hydrologic systems, including droughts and wet spells.

However, initial tests also indicate that the superiority of climate model-based streamflow forecasts tend to diminish when using the observed streamflow data for validation. There are a number of reasons for this,

which include the often poor representation of water resources management (e.g., reservoir operation, irrigation) in land surface models and inadequate process parameterization in the hydrological models. Examples of the latter include surface–subsurface interactions; regional to continental surface water transportation (river routing); insufficient parameters for soil and vegetation properties; and inadequate hydrologic model, which includes calibration that requires either statistical postprocessing (e.g., Yuan and Wood 2012a; Ye et al. 2014), or an ensemble of multiple land surface hydrologic models, and/or model improvements, especially for simulating water resources managements (Jaranilla-Sanchez et al. 2011; Wang et al. 2012). Uncertainties in the observed forcings (e.g., precipitation) can also be significant over basins with sparse in situ observations, which may be a nontrivial problem when implementing a hydrologic forecast system globally.

With the planned release of higher-temporal-resolution NMME datasets (phase 2) with more variables besides monthly precipitation, 2-m temperature, and SST that will also include land surface conditions as well as pressure level atmospheric variables, there will be opportunities to diagnose more completely individual model performance that influence seasonal extreme predictability (e.g., stationary waves, land–atmosphere coupling) and provide feedback to model development. Furthermore, the benefit of incorporating climate models from international centers for improving hydrological predictability globally calls for an international ensemble seasonal prediction system.

APPENDIX: HIT RATE, FALSE ALARM RATIO, ETS, AND CRPSS. The nonprobabilistic forecasts for discrete predictands (e.g., a drought event) can be verified by several measures that are based on a 2×2 contingency table. Taking the drought event forecast as an example, define a as the number of events when drought occurs in both the forecast and observation, b for when drought occurs in the forecast but not in the observation, c for when drought occurs in the observation but not in the forecast, and d for when drought does not occur in either the forecast or observation. Then, the hit rate is

$$\frac{a}{a+c}, \tag{A1}$$

where it is also called the probability of detection. The false alarm ratio is

$$\frac{b}{a+b}, \tag{A2}$$

where it is fraction of "drought" forecasts that turn out to be false. The ETS is

$$\frac{a - a_{ref}}{a - a_{ref} + b + c}, \tag{A3}$$

where $a_{ref} = (a+b)(a+c)/(a+b+c+d)$.

The probabilistic forecasts for continuous predictands (e.g., precipitation) can be verified through the CRPSS. First, the CRPS is defined as

$$CRPS = \int_{-\infty}^{+\infty} [F(y) - F_o(y)]^2 \, dy, \tag{A4}$$

where $F(y)$ is the CDF of the forecast with a predictand value of y and $F_o(y)$ is the CDF of the observation and

$$F_o(y) = \begin{cases} 0, & y < observed \quad value \\ 1, & y \geq observed \quad value \end{cases}. \tag{A5}$$

Then, the CRPSS is defined as

$$CRPSS = 1 - \frac{CRPS}{CRPS_{ref}}, \tag{A6}$$

where $CRPS_{ref}$ is the CRPS from the reference forecast (e.g., the climatological forecast used in this study). So, a value of CRPSS = 0.2, for example, indicates that the probabilistic forecast error is 20% less than the climatological forecast error.

REFERENCES

Barnett, T. P., and Coauthors, 2008: Human-induced changes in the hydrology of the western United States. *Science,* **319,** 1080–1083, doi:10.1126/science.1152538.

Barnston, A. G., M. K. Tippett, M. L. L'Heureux, S. Li, and D. G. DeWitt, 2012: Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bull. Amer. Meteor. Soc.,* **93,** 631–651, doi:10.1175/BAMS-D-11-00111.1.

Becker, E., H. Van den Dool, and Q. Zhang, 2014: Predictability and forecast skill in NMME. *J. Climate,* **27,** 5891–5906, doi:10.1175/JCLI-D-13-00597.1.

Below, R., E. Grover-Kopec, and M. Dilley, 2007: Documenting drought-related disasters: A global reassessment. *J. Environ. Dev.,* **16,** 328–344, doi:10.1177/1070496507306222.

Camberlin, P., S. Janicot, and I. Poccard, 2001: Seasonality and atmospheric dynamics of the tele-connection between African rainfall and tropical sea-surface temperature: Atlantic vs. ENSO. *Int. J. Climatol.,* **21,** 973–1005, doi:10.1002/joc.673.

Cayan, D. R., K. T. Redmond, and L. G. Riddle, 1999: ENSO and hydrologic extremes in the western United States. *J. Climate,* **12,** 2881–2893, doi:10.1175/1520-0442(1999)012<2881:EAHEIT>2.0.CO;2.

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.

Coughlan, M., and R. Avissar, 1996: The Global Energy and Water Cycle Experiment (GEWEX) Continental-Scale International Project (GCIP): An overview. *J. Geophys. Res.,* **101,** 7139–7147, doi:10.1029/96JD00125.

Dai, A., T. Qian, K. E. Trenberth, and J. D. Milliman, 2009: Changes in continental freshwater discharge from 1948 to 2004. *J. Climate,* **22,** 2773–2791, doi:10.1175/2008JCLI2592.1.

Diffenbaugh, N. S., M. Scherer, and M. Ashfaq, 2013: Response of snow-dependent hydrologic extremes to continued global warming. *Nat. Climate Change,* **3,** 379–384, doi:10.1038/nclimate1732.

Dirmeyer, P. A., 2013: Characteristics of the water cycle and land–atmosphere interactions from a comprehensive reforecast and reanalysis data set: CFSv2. *Climate Dyn.,* **41,** 1083–1097, doi:10.1007/s00382-013-1866-x.

——, Z. Guo, and X. Gao, 2004: Comparison, validation, and transferability of eight multiyear global soil wetness products. *J. Hydrometeor.,* **5,** 1011–1033, doi:10.1175/JHM-388.1.

Doblas-Reyes, F. J., R. Hagedorn, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting—II. Calibration and combination. *Tellus,* **57A,** 234–252, doi:10.1111/j.1600-0870.2005.00104.x.

——, J. García-Serrano, F. Lienert, A. P. Biescas, and L. R. L. Rodrigues, 2013: Seasonal climate predictability and forecasting: Status and prospects. *Wiley Interdiscip. Rev.: Climate Change,* **4,** 245–268, doi:10.1002/wcc.217.

Dorigo, W. A., and Coauthors, 2015: Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote Sens. Environ.*, **162,** 380–395, doi:10.1016/j.rse.2014.07.023.

Douville, H., 2010: Relative contributions of soil and snow hydrology to seasonal climate predictability: A pilot study. *Climate Dyn.,* **34,** 797–818, doi:10.1007/s00382-008-0508-1.

Duan, Q., S. Sorooshian, and V. K. Gupta, 1994: Optimal use of SCE-UA global optimization method for calibrating watershed models. *J. Hydrol.,* **158,** 265–284, doi:10.1016/0022-1694(94)90057-4.

——, N. K. Ajami, X. Gao, and S. Sorooshian, 2007: Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Resour.,* **30,** 1371–1386, doi:10.1016/j.advwatres.2006.11.014.

Dutra, E., F. D. Giuseppe, F. Wetterhall, and F. Pappenberger, 2013: Seasonal forecasts of droughts in African basins using the Standardized Precipitation Index. *Hydrol. Earth Syst. Sci.,* **17,** 2359–2373, doi:10.5194/hess-17-2359-2013.

Giorgi, F., C. Vera, and F. Semazzi, cited 2012: WCRP Grand Challenges: Regional climate information. [Available online at www.wcrp-climate.org/gc-regionalclimate.]

Goddard, L., S. J. Mason, S. E. Zebiak, C. F. Ropelewski, R. Basher, and M. A. Cane, 2001: Current approaches to seasonal to interannual climate predictions. *Int. J. Climatol.,* **21,** 1111–1152, doi:10.1002/joc.636.

Hoerling, M., and A. Kumar, 2003: The perfect ocean for drought. *Science,* **299,** 691–694, doi:10.1126/science.1079053.

——, J. Eischeid, A. Kumar, R. Leung, A. Mariotti, K. Mo, S. Schubert, and R. Seager, 2014: Causes and predictability of the 2012 Great Plains drought. *Bull. Amer. Meteor. Soc.,* **95,** 269–282, doi:10.1175/BAMS-D-13-00055.1.

Hong, S. Y., and E. Kalnay, 2000: Role of sea surface temperature and soil-moisture feedback in the 1998 Oklahoma–Texas drought. *Nature,* **408,** 842–844, doi:10.1038/35048548.

Hoskins, B. J., and D. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.,* **38,** 1179–1196, doi:10.1175/1520-0469(1981)038<1179:TSLROA>2.0.CO;2.

GEWEX, 2012: GEWEX plans for 2013 and beyond. GEWEX Doc. Series 2012-1, 45 pp. [Available online at www.gewex.org/pdfs/GEWEX_IMPERATIVES_final.pdf.]

Ineson, S., and A. A. Scaife, 2009: The role of the stratosphere in the European climate response to El Niño. *Nat. Geosci.,* **2,** 32–36, doi:10.1038/ngeo381.

IPCC, 2013: Summary for policymakers. *Climate Change 2013: The Physical Science Basis,* T. F. Stocker et al., Eds., Cambridge University Press, 1–29.

Jaranilla-Sanchez, P. A., L. Wang, and T. Koike, 2011: Modeling the hydrologic responses of the Pampanga River basin, Philippines: A quantitative approach for identifying droughts. *Water Resour. Res.,* **47,** W03514, doi:10.1029/2010WR009702.

Kam, J., J. Sheffield, X. Yuan, and E. F. Wood, 2014: Did a skillful prediction of sea surface temperatures help or hinder forecasting of the 2012 Midwestern US drought? *Environ. Res. Lett.,* **9,** 034005, doi:10.1088/1748-9326/9/3/034005.

Karoly, D., cited 2012: WCRP Grand Challenges: Science underpinning the prediction and attribution of extreme events. [Available online at www.wcrp-climate.org/index.php/gc-extreme-events.]

Kirtman, B. P., and A. Pirani, 2009: The state of the art of seasonal prediction outcomes and recommendations from the First World Climate Research Program (WCRP) Workshop on Seasonal Prediction. *Bull. Amer. Meteor. Soc.,* **90,** 455–458, doi:10.1175/2008BAMS2707.1.

——, and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.,* **95,** 585–601, doi:10.1175/BAMS-D-12-00050.1.

Koster, R. D., and Coauthors, 2004: Realistic initialization of land surface states: Impacts on subseasonal forecast skill. *J. Hydrometeor.,* **5,** 1049–1063, doi:10.1175/JHM-387.1.

——, and Coauthors, 2006: GLACE: The Global Land–Atmosphere Coupling Experiment. Part I: Overview. *J. Hydrometeor.,* **7,** 590–610, doi:10.1175/JHM510.1.

——, and Coauthors, 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.

——, and Coauthors, 2011: GLACE2: The second phase of the Global Land–Atmosphere Coupling Experiment: Soil moisture contribution to subseasonal forecast skill. *J. Hydrometeor.,* **12,** 805–822, doi:10.1175/2011JHM1365.1.

Kumar, A., M. Chen, M. Hoerling, and J. Eischeid, 2013: Do extreme climate events require extreme forcings? *Geophys. Res. Lett.,* **40,** 3440–3445, doi:10.1002/grl.50657.

Lang, Y., and Coauthors, 2014: Evaluating skill of seasonal precipitation and temperature predictions of NCEP CFSv2 forecasts over 17 hydroclimatic regions in China. *J. Hydrometeor.,* **15,** 1546–1559, doi:10.1175/JHM-D-13-0208.1.

Li, H., L. Luo, E. F. Wood, and J. Schaake, 2009: The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. *J. Geophys. Res.,* **114,** D04114, doi:10.1029/2008JD010969.

Liang, X., E. F. Wood, and D. P. Lettenmaier, 1996: Surface soil moisture parameterization of the VIC-2L model: Evaluation and modifications. *Global Planet. Change,* **13,** 195–206, doi:10.1016/0921-8181(95)00046-1.

Liu, Y. Y., R. M. Parinussa, W. A. Dorigo, R. A. M. De Jeu, W. Wagner, A. I. J. M. van Dijk, M. F. McCabe, and J. P. Evans, 2011: Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydrol. Earth Syst. Sci.,* **15,** 425–436, doi:10.5194/hess-15-425-2011.

Luo, L., and E. F. Wood, 2007: Monitoring and predicting the 2007 U.S. drought. *Geophys. Res. Lett.,* **34,** L22702, doi:10.1029/2007GL031673.

——, and ——, 2008: Use of Bayesian merging techniques in a multimodel seasonal hydrologic ensemble prediction system for the eastern United States. *J. Hydrometeor.,* **9,** 866–884, doi:10.1175/2008JHM980.1.

McCabe, G. J., M. A. Palecki, and J. L. Betancourt, 2004: Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States. *Proc. Natl. Acad. Sci. USA,* **101,** 4136–4141, doi:10.1073/pnas.0306738101.

Mitchell, K. E., and Coauthors, 2004: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *J. Geophys. Res.,* **109,** D07S90, doi:10.1029/2003JD003823.

Mo, K., and D. P. Lettenmaier, 2014a: Objective drought classification using multiple land surface models. *J. Hydrometeor.,* **15,** 990–1010, doi:10.1175/JHM-D-13-071.1.

——, and ——, 2014b: Hydrologic prediction over the conterminous United States using the National Multi-Model Ensemble. *J. Hydrometeor.,* **15,** 1457–1472, doi:10.1175/JHM-D-13-0197.1.

——, S. Shukla, D. P. Lettenmaier, and L. Chen, 2012: Do Climate Forecast System (CFSv2) forecasts improve seasonal soil moisture prediction? *Geophys. Res. Lett.,* **39,** L23703, doi:10.1029/2012GL053598.

Molteni, F., and Coauthors, 2011: The new ECMWF seasonal forecast system (System 4). ECMWF Tech. Memo. 656, 49 pp. [Available online at www.researchgate.net/profile/Frederic_Vitart/publication/230688826_The_new_ECMWF_seasonal_forecast_system_%28System_4%29/links/0912f51028aa651240000000.pdf.]

Morel, P., 2001: Why GEWEX? The agenda for a global energy and water cycle program. *GEWEX News,* Vol. 11, No. 1, International GEWEX Project Office, Silver Spring, MD, 1–11.

Nijssen, B., and Coauthors, 2014: A prototype global drought information system based on multiple land surface models. *J. Hydrometeor.,* **15,** 1661–1676, doi:10.1175/JHM-D-13-090.1.

Quan, X.-W., M. Hoerling, B. Lyon, A. Kumar, M. Bell, M. Tippett, and H. Wang, 2012: Prospects for dynamical prediction of meteorological drought. *J. Appl. Meteor. Climatol.,* **51,** 1238–1252, doi:10.1175/JAMC-D-11-0194.1.

Raschke, E., and Coauthors, 1998: The Baltic Sea Experiment BALTEX: A brief overview and some selected results of the authors. *Surv. Geophys.,* **19,** 1–22, doi:10.1023/A:1006567924820.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.,* **91,** 1015–1057, doi:10.1175/2010BAMS3001.1.

——, and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate,* **27,** 2185–2208, doi:10.1175/JCLI-D-12-00823.1.

Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, 1999: A dipole mode in the tropical Indian Ocean. *Nature,* **401,** 360–363.

Scaife, A. A., and Coauthors, 2014: Skillful long-range prediction of European and North American winters. *Geophys. Res. Lett.,* **41,** 2514–2519, doi:10.1002/2014GL059637.

Schubert, S., M. Suarez, P. Pegion, R. Koster, and J. Bacmeister, 2004: On the cause of 1930s Dust Bowl. *Science,* **303,** 1855–1859, doi:10.1126/science.1095048.

Sheffield, J., G. Goteti, and E. F. Wood, 2006: Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling. *J. Climate,* **19,** 3088–3111, doi:10.1175/JCLI3790.1.

——, and Coauthors, 2014: A drought monitoring and forecasting system for sub-Sahara African water resources and food security. *Bull. Amer. Meteor. Soc.,* **95,** 861–882, doi:10.1175/BAMS-D-12-00124.1.

Shukla, J., 1998: Predictability in the midst of chaos: A scientific basis for climate forecasting. *Science,* **282,** 728–731, doi:10.1126/science.282.5389.728.

Shukla, S., J. Sheffield, E. F. Wood, and D. P. Lettenmaier, 2013: On the sources of global land surface hydrologic predictability. *Hydrol. Earth Syst. Sci.,* **17,** 2781–2796, doi:10.5194/hess-17-2781-2013.

——, A. McNally, G. Husak, and C. Funk, 2014: A seasonal agricultural drought forecast system for food-insecure regions of East Africa. *Hydrol. Earth Syst. Sci. Discuss.,* **11,** 3049–3081, doi:10.5194/hessd-11-3049-2014.

Sinha, T., and A. Sankarasubramanian, 2013: Role of climate forecasts and initial conditions in developing streamflow and soil moisture forecasts in a rainfall-runoff regime. *Hydrol. Earth Syst. Sci.,* **17,** 721–733, doi:10.5194/hess-17-721-2013.

Smith, A. B., and R. W. Katz, 2013: US billion-dollar weather and climate disasters: Data sources, trends, accuracy and biases. *Nat. Hazards,* **67,** 387–410, doi:10.1007/s11069-013-0566-5.

Smith, D. M., A. A. Scaife, and B. P. Kirtman, 2012: What is the current state of scientific knowledge with regard to seasonal to decadal forecasting. *Environ. Res. Lett.,* **7,** 015602, doi:10.1088/1748-9326/7/1/015602.

Sohn, S.-J., J.-B. Ahn, and C.-Y. Tam, 2013: Six month–lead downscaling prediction of winter to spring drought in South Korea based on a multimodel ensemble. *Geophys. Res. Lett.,* **40,** 579–583, doi:10.1002/grl.50133.

Stewart, R. E., and Coauthors, 1998: The Mackenzie GEWEX Study: The water and energy cycles of a major North American river basin. *Bull. Amer. Meteor. Soc.,* **79,** 2665–2683, doi:10.1175/1520-0477(1998)079<2665:TMGSTW>2.0.CO;2.

Tian, D., C. J. Martinez, and W. D. Graham, 2014: Seasonal prediction of regional reference evapotranspiration based on Climate Forecast System version 2. *J. Hydrometeor.,* **15,** 1166–1188, doi:10.1175/JHM-D-13-087.1.

Trenberth, K. E., and J. M. Caron, 2000: The Southern Oscillation revisited: Sea level pressures, surface temperatures, and precipitation. *J. Climate,* **13,** 4358–4365, doi:10.1175/1520-0442(2000)013<4358:TSORSL>2.0.CO;2.

Twedt, T. M., J. Schaake, and E. L. Peck, 1977: National Weather Service extended streamflow prediction. *Proc. 45th Western Snow Conf.,* Albuquerque, NM, Colorado State University, 52–57.

van Dijk, A. I. J. M., J. L. Peña-Arancibia, E. F. Wood, J. Sheffield, and H. E. Beck, 2013: Global analysis of seasonal streamflow predictability using an ensemble prediction system and observations from 6192 small catchments worldwide. *Water Resour. Res.,* **49,** 2729–2746, doi:10.1002/wrcr.20251.

Wada, Y., L. P. H. van Beek, N. Wanders, and M. F. P. Bierkens, 2013: Human water consumption intensifies hydrological drought worldwide. *Environ. Res. Lett.,* **8,** 034036, doi:10.1088/1748-9326/8/3/034036.

Wahr, J., S. Swenson, V. Zlotnicki, and I. Velicogna, 2004: Time-variable gravity from GRACE: First results. *Geophys. Res. Lett.,* **31,** L11501, doi:10.1029/2004GL019779.

Walker, G. T., and E. W. Bliss, 1932: World weather V. *Mem. Roy. Meteor. Soc.,* **4,** 53–84.

Wang, A., T. J. Bohn, S. P. Mahanama, R. D. Koster, and D. P. Lettenmaier, 2009: Multimodel ensemble

reconstruction of drought over the continental United States. *J. Climate,* **22,** 2694–2712, doi:10.1175/2008JCLI2586.1.

Wang, F., L. Wang, H. Zhou, O. C. Saavedra Valeriano, T. Koike, and W. Li, 2012: Ensemble hydrological prediction-based real-time optimization of a multiobjective reservoir during flood season in a semiarid basin with global numerical weather predictions. *Water Resour. Res.,* **48,** W07520, doi:10.1029/2011WR011366.

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.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences.* 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 676 pp.

Wood, A. W., E. P. Mauer, A. Kumar, and D. P. Lettenmaier, 2002: Long-range experimental hydrologic forecasting for the eastern United States. *J. Geophys. Res.,* **107,** 4429, doi:10.1029/2001JD000659.

Xia, Y., and Coauthors, 2012a: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res.,* **117,** D03109, doi:10.1029/2011JD016048.

——, and Coauthors, 2012b: Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *J. Geophys. Res.,* **117,** D03110, doi:10.1029/2011JD016051.

Ye, A., Q. Duan, X. Yuan, E. F. Wood, and J. Schaake, 2014: Hydrologic post-processing of MOPEX streamflow simulations. *J. Hydrol.,* **508,** 147–156, doi:10.1016/j.jhydrol.2013.10.055.

Yoon, J. H., K. Mo, and E. F. Wood, 2012: Dynamic-model-based seasonal prediction of meteorological drought over the contiguous United States. *J. Hydrometeor.,* **13,** 463–481, doi:10.1175/JHM-D-11-038.1.

Yossef, N. C., H. Winsemius, A. Weerts, R. van Beek, and M. F. P. Bierkens, 2013: Skill of a global seasonal streamflow forecasting system, relative roles of initial conditions and meteorological forcing. *Water Resour. Res.,* **49,** 4687–4699, doi:10.1002/wrcr.20350.

Yuan, X., and E. F. Wood, 2012a: Downscaling precipitation or bias-correcting streamflow? Some implications for coupled general circulation model (CGCM)-based ensemble seasonal hydrologic forecast. *Water Resour. Res.,* **48,** W12519, doi:10.1029/2012WR012256.

——, and ——, 2012b: On the clustering of climate models in ensemble seasonal forecasting. *Geophys. Res. Lett.,* **39,** L18701, doi:10.1029/2012GL052735.

——, and ——, 2013: Multimodel seasonal forecasting of global drought onset. *Geophys. Res. Lett.,* **40,** 4900–4905, doi:10.1002/grl.50949.

——, ——, 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.

——, ——, N. W. Chaney, J. Sheffield, J. Kam, M. Liang, and K. Guan, 2013a: Probabilistic seasonal forecasting of African drought by dynamical models. *J. Hydrometeor.,* **14,** 1706–1720, doi:10.1175/JHM-D-13-054.1.

——, ——, J. K. Roundy, and M. Pan, 2013b: CFSv2-based seasonal hydroclimatic forecasts over conterminous United States. *J. Climate,* **26,** 4828–4847, doi:10.1175/JCLI-D-12-00683.1.

Zhao, M., and H. H. Hendon, 2009: Representation and prediction of the Indian Ocean dipole in the POAMA seasonal forecast model. *Quart. J. Roy. Meteor. Soc.,* **135,** 337–352, doi:10.1002/qj.370.