Evaluation of summer drought ensemble prediction over the Yellow River basin

YAO Meng-Na a,b and YUAN Xing a

aKey Laboratory of Regional Climate-Environment for Temperate East Asia (RCE-TEA), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China; bSchool of Atmospheric Sciences, Chengdu University of Information Technology, Chengdu, China

ABSTRACT

Accurately predicting drought a few months in advance is important for drought mitigation and agricultural and water resources management, especially for a river basin like that of the Yellow River in North China. However, summer drought predictability over the Yellow River basin is limited because of the low influence from ENSO and the large interannual variations of the East Asian summer monsoon. To explore the drought predictability from an ensemble prediction perspective, 29-year seasonal hindcasts of soil moisture drought, taken directly from several North American multimodel ensemble (NMME) models with different ensemble sizes, were compared with those produced by combining bias-corrected NMME model predictions and variable infiltration capacity (VIC) land surface hydrological model simulations. It was found that the NMME/VIC approach reduced the root-mean-square error from the best NMME raw products by 48% for summer soil moisture drought prediction at the lead-1 season, and increased the correlation significantly. Within the NMME/VIC framework, the multimodel ensemble mean further reduced the error from the best single model by 6%. Compared with the NMME raw forecasts, NMME/VIC had a higher probabilistic drought forecasting skill in terms of a higher Brier skill score and better reliability and resolution of the ensemble. However, the performance of the multimodel grand ensemble was not necessarily better than any single model ensemble, suggesting the need to optimize the ensemble for a more skillful probabilistic drought forecast.

1. Introduction

Under the combined influence of global warming and human activity, the frequency and intensity of extreme hydrometeorological events have shown an increasing trend. Among these types of events, drought is considered to be one of the most serious natural disasters, having a huge impact on economies, agricultural production and human livelihoods (Burke, Perry, and Brown 2010; Dai, Trenberth, and Qian 2004). Numerous studies have shown that seasonal drought events in China have also increased significantly (Ma and Fu 2006; Wang et al. 2010, 2015b). Therefore, predicting drought a few months in advance is of great help to various sectors, allowing them enough time for mitigation (Yuan and Wood 2013).

In recent years, dynamical seasonal forecasting systems based on coupled atmosphere–ocean–land general circulation models have been widely used for drought early warning (Luo and Wood 2007; Dutra et al. 2012; Yuan et al. 2013), including those conducted over east Asia (Ding et al. 2004; Ghosh and Mujumdar 2007; Kim and Byun 2009; Tang, Lin, and Luo 2013; Wang et al. 2015a). However, less than 30% of meteorological drought onsets globally can be detected by climate forecast models, especially over midlatitude regions (Yuan and Wood 2013). In addition, soil moisture drought prediction is also very challenging, due to both limited atmospheric predictability and uncertainties in land surface modeling.

A contemporary approach to improving forecast skill is to use an ensemble prediction technique, where the ensemble can either derive from a single model or from...
multiple climate forecast models (Palmer et al. 2004; Doblas-Reyes, Michel, and Jean-Philippe 2010; Becker, Den Dool, and Zhang 2014). For example, Thober et al. (2015) used meteorological forecasts of the North American multimodel ensemble (NMME; Kirtman et al. 2014) to drive a mesoscale hydrological model by following a similar approach to those used by Luo and Wood (2007) and Yuan et al. (2015), and obtained improvement against the ensemble streamflow prediction approach for drought prediction at monthly to seasonal time scales. Yao and Yuan (2018) used a superensemble method to combine soil moisture predictions directly from NMME climate forecast models over China, and reduced the forecast errors, especially over southeastern and northeastern China. However, the soil moisture forecast skill is low over the Yellow River basin in northern China, where drought frequently occurs.

In order to improve the forecasting of soil moisture drought over the Yellow River basin, Yuan (2016) bias-corrected the NMME meteorological hindcasts and used them to drive the variable infiltration capacity (VIC; Liang et al. 1994) land surface hydrological model to produce soil moisture hindcasts up to six months. Yuan (2016) briefly evaluated the deterministic predictive skill of soil moisture for the NMME grand ensemble mean by comparison with a traditional hydrological forecasting approach. Here, we used the soil moisture hindcast data to carefully assess both the deterministic and probabilistic forecast skill for both the NMME grand ensemble and the individual NMME models, especially for soil moisture droughts during summertime. In addition, we compared the climate–hydrology modeling approach (NMME/VIC) proposed by Yuan (2016) with the NMME raw soil moisture (drought) hindcasts that were released by phase 2 of the NMME project. This study provides a first look at the ensemble (probabilistic) characteristics of seasonal soil moisture prediction from the NMME models, both for the raw hindcasts and the climate–hydrology-model-produced hindcasts.

2. Data and evaluation method

The Yellow River basin is located in North China, and its drainage area is about $7.52 \times 10^5$ km$^2$. With a semi-arid climate, drought is one of the most severe natural disasters in the Yellow River basin. Using eight NMME models, i.e., CanCM3, CanCM4, CM2.2, CFSv2, CCSM4, CM2p5-A06, CM2p5-B01, and GMAO, Yuan (2016) carried out a set of hydrological hindcasts with the VIC land surface hydrological model over the Yellow River basin to produce hindcasts of soil moisture. These soil moisture hindcasts were used to calculate the seasonal means and were then standardized prior to the drought analysis in this study. The lead-0 season forecast represented the seasonal mean soil moisture (or soil drought) forecast started from the beginning of the season, the lead-1 season forecast represented the forecast started from one month before the beginning of the season, and so on.

For convenience, we describe the data and forecasting approach briefly in this paper. The NMME precipitation and temperature hindcasts were downscaled and bias-corrected through the quantile-mapping method, as follows (Yuan 2016):

1. The 1° NMME global meteorological hindcasts during 1982–2010 were first bilinearly interpolated to 0.25°.
2. For each calendar month and each NMME model, all meteorological hindcasts (excluding the target year) with all ensemble members for the target month were used to construct cumulative distribution functions (CDFs) of the forecasts. The CDFs of meteorological observations were constructed similarly (excluding the target year), and the hindcast in the target year was adjusted by matching its rank in the CDFs of the forecasts and observations.
3. The bias-corrected monthly meteorological hindcasts were temporally downscaled to a daily time step by sampling from the observation dataset and rescaling to match the monthly hindcasts.

With the downscaled and bias-corrected meteorological forcings, the well-calibrated VIC land surface hydrological model was used to produce soil moisture hindcasts up to six months.

In this study, we used $-0.8$ as a threshold for drought, and evaluated the ensemble forecast skill for summer drought over the Yellow River basin. The threshold was equivalent to a moderate drought condition, with a probability of about 0.2 (Yuan et al. 2013). The observational dataset was from the VIC offline simulation, driven by observed meteorological forcings. We used it because there was no direct soil moisture observation at a large scale, and the VIC-model-simulated soil moisture was also constrained by observed streamflow through the calibration procedure. Another reason was that our basic aim in this study was to focus on the effect of seasonal climate prediction, where the error from the hydrological model could be removed by using offline simulated soil moisture. Several metrics, including the correlation coefficient, root-mean-
square error (RMSE), and the Brier skill score (BSS; Wilks 2011) were calculated for verification. The Brier score (BS), which is similar to the RMSE, can be used to assess the probabilistic prediction (Wilks 2011):

\[
BS = \frac{1}{n} \sum_{k=1}^{n} (y_k - o_k)^2,
\]

where \( k \) indicates a number of \( n \) forecast-verification pairs, \( y_k \) is the probability of forecast for drought in the \( k \)th pair, and \( o_k \) is the corresponding probability from the observations, with \( o_k = 1 \) if the drought event occurs and \( o_k = 0 \) if it does not.

The BS can be decomposed into reliability and resolution terms as follows (Wilks 2011):

\[
BS = \frac{1}{n} \sum_{i=1}^{I} N_i (y_i - o_i)^2 - \frac{1}{n} \sum_{i=1}^{I} N_i (\bar{o}_i - \bar{o})^2 + (1 - \bar{o})^2
\]

where \( I \) is the discrete number of allowable forecast values (here, \( I \) is the number of member models plus one), \( N_i \) is the number of times for each forecast value \( y_i \) (forecast probability), and \( \bar{o}_i \) is the conditional distribution of an observation, given forecast probability \( y_i \).

The reliability (Rel) measures how close the issued forecast \( y_i \) is to the probability of an observed occurrence conditional on the forecast; the resolution (Res) refers to the differences between the conditional distributions of the observations for different forecast values; and the third term is the uncertainty (Wilks 2011). A lower Rel and higher Res are representative of a better forecasting skill.

The BSS is based on the BS:

\[
BSS = 1 - BS/BS_{clim},
\]

where \( BS_{clim} \) represents a reference forecast (e.g., climatological forecast). The BSS indicates the degree of improvement in climate predictions, and a better forecast has a higher BSS.

Figure 1. The skill (Pearson correlation) for ensemble prediction of June–July–August mean soil moisture over the Yellow River basin during summer time at the lead-1 season: (a–d) NMME raw soil moisture prediction; (e–l) VIC soil moisture prediction driven by bias-corrected NMME meteorological forcings; (m) multimodel ensemble mean from 99 NMME/VIC members. The reference soil moisture is from the VIC offline simulation. All statistics were calculated using standardized soil moisture hindcasts during 1982–2010.
3. Results

3.1. Predictive skill of summer soil moisture in the Yellow River basin

The correlations for summer soil moisture forecasts over the Yellow River basin at the lead-1 season are shown in Figure 1. The numbers in the upper-right corner of each panel indicate the area average of the correlation coefficients. Here, we only show soil moisture raw forecasts from four NMME models because of hindcast data availability for both NMME and NMME/VIC. For the raw soil moisture forecasts from individual NMME climate models, CFSv2 (Figure 1(d)) had the highest skill, especially in the southern part of the Yellow River basin, as compared with the other three models (Figure 1(a–c)). All models showed high forecasting skill in the central part of the Yellow River basin, but had low skill both in the upstream and downstream regions. The skill was significantly improved when predicting soil moisture by combining the NMME climate forecasts and VIC land surface hydrological model with a bias correction method (Yuan 2016). Taking CanCM4 as an example, the mean correlation for CanCM4/VIC was 0.59 (Figure 1(f)), while for CanCM4 it was only 0.11. The improvement may have been due to the reduction in errors both from the climate forecasts and land model simulations, where the VIC model was calibrated by using streamflow data over the Yellow River (Yuan 2016). NMME/VIC was a simple ensemble mean of 99 members from eight models (Figure 1(e–k)). The average correlation for NMME/VIC was 0.6, which was only a marginal improvement compared to the best single model (CanCM4/VIC).

![Figure 2](image-url)  
Figure 2. Frequency distributions of RMSE for the predicted June–July–August mean soil moisture at different lead times. Different dashed color lines represent results for soil moisture raw predictions from the NMME models; solid color lines represent results for soil moisture predictions from the VIC model driven by the NMME models; and the black solid line represents the multimodel ensemble mean prediction (NMME/VIC). All statistics were calculated using standardized soil moisture hindcasts during 1982–2010.
Figure 2 shows the frequency distributions of RMSE at different forecast lead times over the Yellow River basin during June–July–August. Similar to the lead-1 seasons, the NMME models’ raw forecast had much larger errors than those produced by the climate-hydrology approach. The reasons might be threefold:

1. Land surface models (including those in the NMME) are usually less reliable in reproducing the soil moisture dynamics over semiarid regions like the Yellow River basin, while the VIC model’s performance was improved through streamflow calibration.

2. Some of the NMME models did not initialize the land surface component when making the seasonal prediction, while NMME/VIC used a realistic initial condition through offline simulation to the forecast start date.

3. Biases of meteorological forcings in NMME would transfer to their land surface modeling, while NMME/VIC removed those biases before producing the soil moisture forecasts.

The NMME/VIC grand ensemble was similar to the individual models in the upper and middle reaches of the Yellow River basin, but it became gradually better than any single model in the lower reaches of the Yellow River basin as the forecast lead time increased.

3.2. Predictive skill of summer drought in the Yellow River basin

Soil moisture drought refers to the phenomenon of a water shortage due to the imbalance between soil moisture supply and demand. The performances of the models in predicting summer soil moisture droughts at the lead-1 season are shown in Figure 3. Similar to Figure 1, CFSv2 (Figure 3(d)) had the lowest error (1.34) among the four NMME models for the soil drought raw forecasts. The error was smaller in the middle and lower reaches of the Yellow River, and larger in the upstream areas. CFSv2/VIC (Figure 3(h)) had the smallest regional mean error (0.7), as compared with the seven other models.

Figure 3. RMSE for June-July-August soil moisture drought predictions over the Yellow River basin at the lead-1 season: (a–d) NMME raw predictions of soil drought; (e–l) VIC soil moisture drought prediction driven by bias-corrected NMME meteorological forcings; (m) multimodel ensemble drought prediction from 99 NMME/VIC members. The reference soil moisture drought was from the VIC offline simulation. All statistics were calculated using standardized soil moisture hindcasts during 1982–2010. Drought months were identified when standardized soil moisture was lower than −0.8 based on VIC offline simulation.
climate–hydrology forecasting models. NMME/VIC, the multi-model grand ensemble mean, performed better than the best individual model for soil moisture drought forecasting, with errors decreased by 6%. For the eight climate–hydrology forecasting models, they showed high drought forecasting skill in the central part of the Yellow River basin, which was similar to the soil moisture forecast skill (Figure 1). Moreover, most models had their worst drought forecasting skill in the upstream regions, except for CCSM4/VIC (Figure 3(i)). CM2.2 (Figure 1(g)) and CM2p5-A06/VIC (Figure 1(j)) had the lowest soil moisture forecasts skill, as compared with other models, with an average correlation of 0.49; meanwhile, they also had the lowest drought forecasting skill, with a regional mean error as large as 0.83.

In order to evaluate the probabilistic drought prediction skill, we calculated the BSS, the results of which, at the lead-0 season, are shown in Figure 4. The climate–hydrology approach produced a much better probabilistic soil moisture drought forecast than the NMME raw forecasts. Similar to the deterministic forecasts, there was also a higher probabilistic drought forecast skill over the middle reaches of the Yellow River. The area-average BSS of GMAO/VIC (Figure 4(l)) was the highest (0.43), which was even better than the grand ensemble mean of NMME/VIC (0.39). This was a little surprising because given that the deterministic drought forecasting skill from NMME/VIC was higher than any individual model (Figure 3). One of the reasons was perhaps the skill spatial patterns from individual models were too similar to one another, where a simple combination did not help without spatial complementary skill. Another reason was perhaps related to the bias correction approach (Yuan 2016), where each NMME model was mapped to the observed climatology, without considering the skill of the climate forecasts. A Bayesian approach for the bias correction can account for the model hindcast skill, where the ensemble forecast distribution will match reality if the hindcast skill is high, and the forecast distribution will be close to climatology if the hindcast skill is low (Yuan et al. 2013). However, whether to use full samples including drought, neutral and wet conditions to calibrate the Bayesian model, or just those samples in drought conditions, is still under debate. For the

![Figure 4. The Brier skill score for probabilistic forecasts of summer (June-July-August) droughts at the lead-0 season.](image-url)
latter, three decades of hindcasts may not be enough for calibrating extreme forecasts.

To further assess the probabilistic drought forecasting skill, we plotted the frequency distribution of the BSS, Rel, and Res terms for summer drought at the lead-0 season over the Yellow River basin (Figure 5). The climate model raw forecasts were the worst for both the BSS and its components. The NMME/VIC multimodel probabilistic forecast was not much different from the single model probabilistic forecasts, laying between the best and the worst models. As compared with the results shown in Figure 2(a), where NMME/VIC only improved the soil moisture forecasts with large errors, Figure 5 shows that a simple multimodel ensemble did not improve the probabilistic prediction of extremes (droughts).

4. Concluding remarks

After the bias correction of the meteorological forecast and the implementation of a well-calibrated land surface hydrological model, summer soil moisture drought forecasting was improved significantly, especially over the middle and lower reaches of the Yellow River basin. The simple multimodel ensemble improved the forecasting skill for soil moisture, but not for extreme events, such as drought, where the performance of the multimodel ensemble lay between the best and worst individual models. It is necessary to find appropriate methods to improve the ensemble prediction of extreme events, either for deterministic or probabilistic forecasting.

Acknowledgments

We would like to thank the IRI (http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/) for making the NMME forecast information available.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the China Special Fund for Meteorological Research in the Public Interest (Major projects) (Grant No. GYHY201506001) and the National Natural Science Foundation of China (Grant No. 91547103).

References

Becker, E., H. V. Den Dool, and Q. Zhang. 2014. “Predictability and Forecast Skill in NMME.” Journal of Climate 27 (15): 5891–5906. doi:10.1175/JCLI-D-13-00597.1.
Burke, E. J., R. H. J. Perry, and S. J. Brown. 2010. “An Extreme Value Analysis of UK Drought and Projections of Change in the Future.” *Journal of Hydrology* 388 (1): 131–143. doi:10.1016/j.jhydrol.2010.04.035.

Dai, A., K. E. Trenberth, and T. Qian. 2004. “A Global Dataset of Palmer Drought Severity Index for 1870-2002: Relationship with Soil Moisture and Effects of Surface Warming.” *Journal of Hydrometeorology* 5 (6): 1117–1130. doi:10.1175/JHM-386.1.

Ding, Y. H., Q. Q. Li, W. J. Li, Y. Luo, P. Q. Zhang, Z. Q. Zhang, X. L. Shi, et al. 2004. “Advance in Seasonal Dynamical Prediction Operation in China.” *Acta Meteorologica Sinica* 62 (5): 598–612. doi:10.11676/qxxb2004.059.

Doblas-Reyes, F. J., D. Michel, and P. Jean-Philippe. 2010. “Multi-Model Spread and Probabilistic Forecasts in PROVOST.” *Quarterly Journal of the Royal Meteorological Society* 126: 2069–2087. doi:10.1002/qj.49712656705.

Dutra, E., L. Magnusson, F. Wetterhall, H. L. Cloke, G. Balsamo, S. Boussetta, and F. Pappenberger. 2012. “The 2010–2011 Drought in the Horn of Africa in ECMWF Reanalysis and Seasonal Forecast Products.” *International Journal of Climatology* 33: 1720–1729. doi:10.1002/joc.3545.

Ghosh, S., and P. P. Mujumdar. 2007. “Nonparametric Methods for Modeling GCM and Scenario Uncertainty in Drought Assessment.” *Water Resources Research* 43 (7): W07405. doi:10.1029/2006WR005351.

Kim, D. W., and H. R. Byun. 2009. “Future Pattern of Asian Drought under Global Warming Scenario.” *Theoretical & Applied Climatology* 98 (12): 137–150. doi:10.1007/s00704-008-0100-y.

Kirtman, B. P., D. Min, J. M. Infanti, J. L. I. Kinter, D. A. Paolino, Q. Zhang, H. Van Den Dool, et al. 2014. “The North American Multimodel Ensemble: Phase-1 Seasonal-To-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction.” *Bulletin of the American Meteorological Society* 95: 585–601. doi:10.1175/BAMS-D-12-00050.1.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges. 1994. “A Simple Hydrologically Based Model of Land Surface Water and Energy Fluxes for General Circulation Models.” *Journal of Geophysical Research: Atmospheres* 99: 14415–14428. doi:10.1029/94JD00483.

Luo, L., and E. F. Wood. 2007. “Monitoring and Predicting the 2007 U.S. Drought.” *Geophysical Research Letters* 34 (22): 315–324. doi:10.1029/2007GL031673.

Ma, Z. G., and C. B. Fu. 2006. “The Basic Fact of Drying Tendency in North China from 1951 to 2004.” *Chinese Science Bulletin* 51 (20): 2429–2439. doi:10.1007/s11434-006-2159-0.

Palmer, T. N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Déécluse, M. Déqué, et al. 2004. “Development of a European Multimodel Ensemble System for Seasonal-To Prediction (DEMETER).” *Bulletin of the American Meteorological Society* 85: 853–872. doi:10.1175/BAMS-85-6-853.

Tang, W., Z. H. Lin, and L. F. Luo. 2013. “Assessing the Seasonal Predictability of Summer Precipitation over the Huaihe River Basin with Multiple APCC Models.” *Atmospheric and Oceanic Science Letters* 6 (4): 185–190. doi:10.3878/j.issn.1674-2834.13.0025.

Thober, S., R. Kumar, J. Sheffield, J. Mai, D. Schäfer, and L. Samaniego. 2015. “Seasonal Soil Moisture Drought Prediction over Europe Using the North American Multi-Model Ensemble (NMME).” *Journal of Hydrometeorology* 16 (6): 2329–2344. doi:10.1175/JHM-D-15-0053.1.

Wang, H. J., K. Fan, J. Q. Sun, S. L. Li, Z. H. Lin, G. Q. Zhou, L. J. Chen, et al. 2015a. “A Review of Seasonal Climate Prediction Research in China.” *Acta Meteorologica Sinica* 32 (2): 149–168. doi:10.1007/s00376-014-0016-7.

Wang, J. S., F. H. Chen, L. Jin, and H. Z. Bai. 2010. “Characteristics of the Dry/Wet Trend over Arid Central Asia over the past 100 Years.” *Climate Research* 41 (1): 51–59. doi:10.3354/cr00837.

Wang, L. W. Chen, W. Zhou, and G. Huang. 2015b. “Drought in Southwest China: A Review.” *Atmospheric and Oceanic Science Letters* 8 (6): 339–344.

Wilks, D. S. 2011. “Statistical Methods in the Atmospheric Sciences.” In *International Geophysics Series*, 676. 3rd ed. 100 vols. San Diego: Academic Press.

Yao, M., and X. Yuan. 2018. “Superensemble Seasonal Forecasting of Soil Moisture by NMME.” *International Journal of Climatology* 38: 2565–2574. in press. doi:10.1002/joc.5436.

Yuan, X. 2016. “An Experimental Seasonal Hydrological Forecasting System over the Yellow River Basin – Part 2: The Added Value from Climate Forecast Models.” *Hydrology & Earth System Sciences* 20 (6): 2453–2466. doi:10.5194/hess-20-2453-2016.

Yuan, X., J. K. Roundy, E. F. Wood, and J. Sheffield. 2015. “Seasonal Forecasting of Global Hydrologic Extremes: System Development and Evaluation over GEWEX Basins.” *Bulletin of the American Meteorological Society* 96: 1895–1912. doi:10.1175/BAMS-D-14-00003.1.

Yuan, X., and E. F. Wood. 2013. “Multimodel Seasonal Forecasting of Global Drought Onset.” *Geophysical Research Letters* 40 (18): 4900–4905. doi:10.1002/grl.50949.

Yuan, X., E. F. Wood, N. W. Chaney, J. Sheffield, J. Kam, M. Liang, and K. Guan. 2013. “Probabilistic Seasonal Forecasting of African Drought by Dynamical Models.” *Journal of Hydrometeorology* 14 (6): 1706–1720. doi:10.1175/JHM-D-13-054.1.