Do Climate Forecast System (CFSv2) forecasts improve seasonal soil moisture prediction?

Kingtse C. Mo[1], Shraddhanand Shukla[2], Dennis P. Lettenmaier[2]

and Li-Chuan Chen[3]

1. Climate Prediction Center, NCEP/NWS/NOAA, College Park, MD
2. Department of Civil and Environmental Engineering, University of Washington, Seattle, WA
3. Cooperative Institute for Climate and Satellites/Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD

Geophys. Res. Lett

August 2012

Corresponding author: Kingtse C. Mo, Climate Prediction Center/NWS/NOAA ,5830 University Research Court, College Park, Md. 20740. Email: kingtse.mo@noaa.gov

ABSTRACT

We investigated whether seasonal forecasts from the National Centers for Environmental Prediction (NCEP) Climate Forecast System version 2 (CFSv2) contribute to the skill of seasonal soil moisture (SM) forecasts over conterminous U.S. (CONUS) relative to Ensemble Streamflow Prediction (ESP). The benchmark ESP forecasts were performed using the Variable Infiltration Capacity (VIC) land surface hydrology model (termed ESP_VIC). We compared the ESP_VIC forecasts to SM forecasts performed using VIC with the same initial conditions, but with forcing derived from bias-corrected daily precipitation, temperature, and wind forecasts from CFSv2 (CFSv2_VIC) for the period from 1982 to 2009 initialized on January 1 and July 5. Overall, SM forecast skill is seasonally and regionally dependent. Forecast skill is higher over the western interior of CONUS for both ESP_VIC and CFSv2_VIC relative to the eastern part of the domain. For the western interior of CONUS where soil moisture has strong persistence, ESP_VIC has equal or slightly higher skill than CFSv2_VIC forecasts for all lead times. CFSv2_VIC performs better than ESP_VIC over regions where precipitation (P) is modulated by atmospheric circulation at short lead times. These regions include the Tennessee and Ohio Valleys and the Southwest, where P forecasts from CFSv2 are skillful at one month lead. At leads 2-3 months though, ESP_VIC and CFSv2_VIC have essentially equivalent forecast skill over almost the entire CONUS. We also argue that ESP, rather than persistence (as used in many studies), is a more relevant benchmark for evaluation of seasonal hydroclimate forecasts.

1. **Introduction**

Drought is among the costliest natural disasters in the United States, with average losses exceeding $10 billion [*NCDC*, 2011]. Drought early warning systems based on hydroclimate forecasts can help local and federal governments to reallocate resources for mitigating drought impacts [*Hayes et al.* 2004]. Currently, both the Environmental Modeling Center (EMC) of the National Centers for Environmental Prediction (NCEP) and the University of Washington (UW) routinely produce hydroclimate forecasts of soil moisture and runoff to support the Climate Prediction Center (CPC) operational Seasonal Drought Outlook. The EMC uses the hydrological prediction system developed by the Princeton University group [*Luo et al*. 2007] based on the NCEP Climate Forecast System version 1 (CFSv1). The UW uses the Ensemble Streamflow Prediction (ESP) method to predict soil moisture and runoff on seasonal time scales [*Wood and Lettenmaier* 2006]. Both systems use the VIC model as the core of their hydroclimate forecast systems.

NCEP recently upgraded their CFS system (to CFSv2) with improved model physics and higher spatial resolution [http://cfs.ncep.noaa.gov]. *Yuan et al.* [2011] examined forecast skill of 2m temperature (T_{2m}) and precipitation (P), and found a substantial increase in Lead-1month forecast skill relative to CFSv1 over the CONUS. The question we raise in this paper is whether these improvements in P and T_{2m} forecasts lead to improved ability to forecast soil moisture (SM), a primary variable required for agricultural drought forecasting.

For seasonal SM forecasting, skill comes from the initial hydrologic conditions (IHCs) and climate forecast (CF) skill. *Shukla and Lettenmairer* [2011] compared the forecast skill of ESP, a method widely used in hydrology which is based solely on knowledge of IHCs (no CF) as represented by a land surface hydrology model, and a method that *Wood and Lettenmaier*

[2008] termed reverse ESP (rESP), which is based on climatology for IHCs but perfect CF. Comparison of SM forecast skill for ESP and rESP isolates the contributions due to IHCs and CF. They found the IHCs dominate SM forecast skill at leads 1 to 2 months, and CF thereafter. For some parts of CONUS, such as the western interior region, IHCs play an important role even at longer leads. It should be emphasized that while ESP is a practical tool that is widely used in hydrology, rESP is not, because it assumes perfect forecasts. In this paper, rather than perfect CFs, we assess SM forecast skill relative to ESP for SM forecasts in which CFSv2 is the CF source.

2. **Methods**

a) **VIC simulation**

We used VIC model version 4.0.6 [*Liang et al.* 1994] to perform the forecast experiments. This is the same version of VIC that is used in the University of Washington (UW) quasi-operational Surface Water Monitor (SWM; http://www.hydro.washington.edu/forecast/monitor). We ran the model in water balance mode (essentially meaning that the effective surface temperature is assumed to be equal to surface air temperature) with a spatial resolution of 0.5 degrees. To spin up the model's SM and snow storages, the VIC model was run from 1 Jan 1979 to 1 Dec 2010 with initial conditions on 31 December 1978 taken from UW's SWM archive. Forcings for the simulation were derived from observations from index stations using the procedure outlined in *Wood and Lettenmaier* [2006]. This long-term simulation is labeled as VIC(SIM). The SM taken from VIC(SIM) was also used for verification and to derive parameters for downscaling and error correction.

b) **Bias correction and spatial downscaling (BCSD) method**

The BCSD method is a quantile mapping approach that is commonly used to correct biases of hydroclimate forecasts [*Wood et al.* 2002; *Wood and Schaake* 2008]. The BCSD method corrects the full probability distribution of the variable in question.

c) ESP_VIC and CFSv2_VIC experiments

All experiments were carried out for the period during which CFSv2 hindcasts are available (1982-2009). We examined forecasts initiated on 1 January and 5 July. Both the ESP and CFSv2_VIC experiments have the same IHCs, obtained from VIC(SIM) on the same forecast date for the target year.

i) ESP_VIC

For a given target year, each member of the ESP_VIC ensemble, was selected randomly from the historical period from 1950-2009 with the target year excluded. VIC Forcings (P, T_{max} and T_{min}) were derived from the time series of observations for that ensemble member starting from the forecast initialization date and proceeding through the end of 3 months. Other variables such as downward solar and longwave radiation, required to force VIC were indexed to the daily mean temperature and temperature range following the approach outlined in *Maurer et al.* [2002], while surface wind was taken from the lowest vertical level of the NCEP/NCAR reanalysis. These forcings were then used to drive the VIC model to obtain daily SM values for that forecast ensemble. The process was repeated for N ensemble members by selecting N different years in the historical period. The ensemble average SM forecast is the equally weighted mean of all members. We tested the ESP forecasts for N=10, 20, 30, 40 and 50 and found that about N=20 produced stable results, and is at least approximately consistent with the 16 ensemble members available for CFSv2 (see below).

ii) CFSv2_VIC

For the CFSv2_VIC forecasts, the VIC forcings were derived from the CFSv2 seasonal hindcast archive from National Climate Data Center (NCDC). Archived CFSv2 seasonal hindcasts were performed every 5 days from 1 Jan 1982 to 27 December 2009 with a frozen model and data assimilation system. On each day, four forecast runs were initialized at 0Z, 6Z, 12Z and 18Z of that day. Each run lasts for 9 months. To obtain a total of 16 ensemble members, we used four ensemble members each initialized on the nominal forecast date (D_{fcst}) as well as D_{fcst} -5 , D_{fcst} -10 and D_{fcst}-15. Time series of 6-hourly P, T_{2m} and 850-hPa winds were obtained from the CFSv2 hindcast archive for each ensemble member. The forecast date D_{fcst} is always the first day of the target month on which a set of (four) ensemble members were initiated. For each ensemble member, the daily T_{max}, T_{min}, P and 850-hPa winds were bi-linearly interpolated to the VIC grids with a spatial resolution of 0.5 degrees from the CFSv2 grid (with approximate resolution of 1 degree). Monthly mean P and T_{2m} at each lead were corrected for bias using the BCSD method. The probability distribution of P or T_{2m} was determined by using all hindcast members in the training period. We chose to correct monthly means instead of daily means because the 28-year record is not long enough to establish a stable daily climatology, and to avoid problems with mis-representing interactions among the three primary variables. The correction was equally distributed to all days within the month. Then, forcings derived from the bias-corrected daily P and T_{max}, T_{min} were used to drive the VIC model to obtain the SM forecasts.

The monthly mean CFSv2_VIC and ESP_VIC SM forecasts were corrected using BCSD to the probability distribution of the historical simulation, VIC(SIM). Even though P , T_{max} and T_{min} are error corrected, the relationship between SM and the forcings is not linear. For example, errors in evapotranspiration feed back to SM forecasts. Therefore, we chose to perform a second stage error correction. For ESP_VIC, we found that the second stage error correction does not result in statistically significant differences in the forecasts, however for consistency, we performed the second stage bias correction to both sets of forecasts.

iii) Verification

The SM hindcasts from both experiments (ESP_VIC and CFSv2_VIC) were cross validated against VIC(SIM) for the target year. The root-mean-square error (RMSE) between hindcasts and VIC(SIM) was used to estimate forecast skill. We normalized the RMSE by the standard deviation of the SM anomalies from VIC(SIM). If RMSE is greater than 1, then there is no skill because the errors are larger than interannual variability.

To measure the relative skill of the two experiments, we calculated the RMSE ratio R between the two experiments. Let RMSE(i) be the RMSE for hindcasts produced by experiment i (i=1, 2),

$$R(exp1/exp2) = \frac{RMSE(1)}{RMSE(2)} \quad \text{Eq. (1)}$$

If R is less than 1, then experiment 1 has higher skill than experiment 2. The reverse is true if R is greater than 1 [*Shukla and Lettenmaier* 2011].

Let the variance S_i be equal to $(RMSE(i))^2$ (assuming bias is small as a result of bias correction). We tested whether the difference in variances (S_1 and S_2) between two experiments is statistically significant at the 5% level using Bartlett's test as applied by *Lettenmaier and Burges* [1978].

3. Forecast skill

Fig. 1 shows the normalized RMSE for forecasts initialized on 1 January and 5 July for the ESP_VIC experiments while the skill for the CFSv2_VIC is given in Fig. 2. Forecasts in areas where the normalized RMSE was greater than 1 were considered unskillful. Both experiments forecasts have strongly seasonally dependent skill. In general, skill is higher in January and lower in July. Skill also is regionally dependent. Skill is higher over the western interior of CONUS where forecasts are skillful even at leads longer than one month and is lower over the eastern U.S. At Lead- 1, forecasts from both ESP_VIC and CFSv2_VIC generally are skillful. At Lead -2 , both forecasts are skillful mostly over the western interior of CONUS. The areas where CFSv2_VIC has low skill are also areas where ESP_VIC has low skill. For July, both forecasts show low skill for Lead -2 east of about 100 ^{o}W except Texas. Figures 1d and 1h show the RMSE ratio for ESP relative to persistence RMSE(ESP)/RMSE(persistence) which is less than 1 essentially everywhere. ESP_VIC forecasts are more skillful than persistence because they have full knowledge of the IHCs and climatologic forcing in comparison to persistence of SM.

The question we seek to answer is whether forcing derived from CFS_VIC forecasts are more skillful than those derived from ESP_VIC forecasts, which resample their forcing from climatology. Figures 3a-3d shows the RMSE ratio (ESP_VIC/CFSV2_VIC) (contoured). The red (black) contour lines indicate where CFSv2_VIC (ESP_VIC) generally has higher skill. The areas that the differences between the variances of CFSv2_VIC and ESP_VIC are statistically significant at the 5% level based on the Bartlett test as applied by Lettenmaier and Burges [1978] are shaded.

Overall, the differences in skill between the two experiments are statistically significant only at Lead 1. For January, the areas with statistically significant skill differences cover the

Northern Central and the western interior region. For July, the areas are limited to the area west of 115°W. For these areas, ESP_VIC is more skillful than CFSv2_VIC (ratio <0.9 black contour lines). CFSv2_VIC adds more skill over the eastern CONUS which arguably is more dynamically active (RMSE ratio> 1.1 contoured red) but only if the P forecasts are skillful. These are the regions that show the RMSE of rESP is low relative to ESP, where CF is especially important to overall forecast skill [*Shukla and Lettenmaier* 2011]. The ESP approach is generally most skillful in the areas where SM is persistent and is less skillful over the dynamically active areas where the interannual variability of soil moisture is low compared with that of precipitation during the forecast period. This latter condition leads to low forecast skill along a swath from the Gulf States to the Tennessee and Ohio Valleys in January. Precipitation over these areas depends on the path and strength of moisture transport from the Gulf of Mexico, which is determined by dynamic forcings. For the Southwest, SM increases after the monsoon onset which varies from late June to early August [*Higgins et al.,* 1997]. The timing of monsoon onset and retreat depends on the establishment of monsoon circulation. At that point, ESP_VIC does not have that information, and arguably for this reason ESP_VIC forecast skill is lower than CFSv2_VIC over Arizona and New Mexico for July.

Figures 3e-3h shows the cross validated (normalized) RMSE skill for CFSv2 monthly mean P forecasts after the BCSD bias correction verified against the P analyses. There is a good correspondence between the forecast skill for P and the RMSE R ratio for SM (Figs. 3a-3d). The areas where the P forecasts have high skill are also the region where CFSv2_VIC SM forecasts have higher skill than ESP_VIC. For January, ESP_VIC is more skillful than the CFSv2_VIC forecasts over the interior of the West and the North Central CONUS for Lead 1 (Fig.3a). In these regions, P forecasts have low skill with normalized RMSE>1 (Fig. 3e). On the other hand, CFSv2 P forecasts are generally skillful over the Gulf coast, the Southwest and the Ohio and Tennessee Valleys. These are regions where the CFSv2_VIC SM forecasts are more skillful than ESP_VIC. For July, ESP_VIC has higher skill over most of the West where the P forecast skill is low. The CFSv2_VIC is more skillful than the ESP_VIC over the Southwest monsoon region.

4. Discussion and conclusions

We have evaluated ESP_VIC and CFSv2_VIC SM forecasts over the CONUS for January and July for the period 1982-2009. As pointed out by S*hukla and Lettenmaier* [2011], SM forecast skill is regionally and seasonally dependent. Overall, predictive skill is higher over the western part of CONUS for both ESP_VIC and CFSv2_VIC and lower over the eastern part of CONUS. Over the CONUS, there are two distinct hydroclimate regimes. The Interior of the West is dry and has high water holding capacity. The characteristic time T_o can be considered as a measure of persistence [*Trenberth* 1984]. Fig. 2d shows that T_o computed from SM based on VIC(SIM) is about 2 years over the western U.S. The eastern U.S. is wetter with more frequent precipitation. SM is less persistent but T_o for SM nonetheless is about 6 months, which is much longer than T_o for precipitation. This accounts for the regional differences in SM forecast skill.

Persistence has commonly been used as a baseline for evaluation of forecast skill because of its simplicity and availability [*Schubert et al.* 1992]. For SM, persistence can produce relatively skillful forecasts. However, ESP forecasts generally are more skillful than persistence (Figs.1d and 1h) because they exploit full knowledge of the IHCs and the seasonal cycle of climatologic forcing. Furthermore, it is always possible to obtain reliable IHCs (assuming the existence of consistent long-term model forcing data, which is generally the case over CONUS) from a land surface model such as VIC. Therefore, we argue that for SM (and other hydrologic forecasts) ESP is a more relevant benchmark than persistence. It sets a higher bar for alternative methods, such as CFSv2_VIC.

Does CFSv2_VIC add any values to the ESP_VIC forecasts? Over the western interior of CONUS, ESP_VIC generally is superior to CFSv2_VIC due to the strong persistence of SM and because of the low skill of the P forecasts from the CFSv2 for both winter and summer. Figure 3 shows that there is a good correspondence between areas where P forecasts have skill (Figs. 3e and 3g) and areas where CFSv2_VIC has RMSE ratios greater than 1 – especially at Lead -1 (Figs.3a and 3c, colored red). When and where P forecasts from the CFSv2 are skillful, the CFSv2_VIC does add values to SM forecasts.

One reason that the CFSv2_VIC has low skill is the design of CFSv2 hindcasts. They were performed every 5 days. For 16 member ensemble, the old member is about 15 days old. It does not exploit the skillful weather forecast at the beginning of forecast [*Shukla et al*. 2012]. Better weighting of ensemble members may also improve skill.

Acknowledgements

This project was supported by the NOAA (CTB) grant J8R1RP7-P01 to the Climate Prediction Center, and by NOAA (CTB) grant No. NA10OAR4310245 to the University of Washington.

References

Hayes, M., O. Wilhelmi, and C. Knutson (2004), Reducing drought risk: bridging theory and practice. *Natural Hazards Rev*., **5(2)**, 106-113.

Higgins, R. W., Y. Yao, X. L. Wang (1997), Influence of the North American Monsoon System on the U.S. Summer Precipitation Regime. *J. Climate*, **10**, 2600–2622.

Lettenmaier, D. P., and S. E. Burges (1978), Climate change: detection and its impact on hydrologic design. *Water Resources Res*., **14**, 679-687.

Liang, X., D. P. Lettenmainer, E. F. Wood, and S. J. Burges (1994), A simple hydrologically based model of land surface water and energy fluxes for General Circulation Models. *J. Geophy. Res.,* **99**, 14415-14428.

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

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

National Climatic Data Center (NCDC), 2011: Billion Dollar U.S. Weather Disasters, 1980-2010. Available online at http://www.ncdc.noaa.gov/oa/reports/billionz.html.

Schubert, S. D., M. Suarez and J. Schemm (1992), Persistence and predictability in a perfect model *J. Atmos.Sci.,***49,** 256-269.

Shukla, S. and D. P. Lettenmaier (2011), Seasonal hydrologic prediction in the United States: understanding the role of initial conditions and seasonal climate forecast skill. *Hydro. Earth Syst. Sci.,* 3529-3538, doi:10.5194/hess-15-3529-2011.

Shukla, S., N. Voisin, and D. P. Lettenmaier, (2012) : Value of medium range weather forecasts in the improvement of seasonal hydrologic prediction skill, *Hydrol. Earth Syst. Sci. Discuss.*, 9, 1827-1857, doi:10.5194/hessd-9-1827-2012

Trenberth, K. E. (1984), Some effects of finite sample size and persistence on Meteorological statistics. Part II: potential predictability. *Mon. Wea. Rev*., **112**, 2369-2379.

Wood, A. W., and D. P. Lettenmaier (2006), A test bed for new seasonal hydrologic forecasting approaches in the western United States. *Bull. Amer. Meteor. Soc.* **87**, 1699-1712.

Wood, A. W., and D. P. Lettenmaier (2008), An ensemble approach for attribution of hydrologic prediction uncertainty. *Geophys. Res. Lett*., 35(14), L14401, doi:1029/2008GL034648.

Wood, A. W., and J. Schaake (2008), Correcting errors in streamflow forecast ensemble mean and spread. *J. Hydrometeor.* **9**, 132-148.

Wood, A., E. P. Maurer, A. Kumar, and D. P. Lettenmaier (2002), Long range experimental hydrologic forecasting for the eastern United States. *J. Geophy. Res.*, doi:10.1029/2001JD00659.

Yuan, X, E. F. Wood, L. Luo and M. Pan (2011), A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction, *Geophy. Res. Let.* **38**, doi:10 :1029 /2011GL047792.

Figure 1

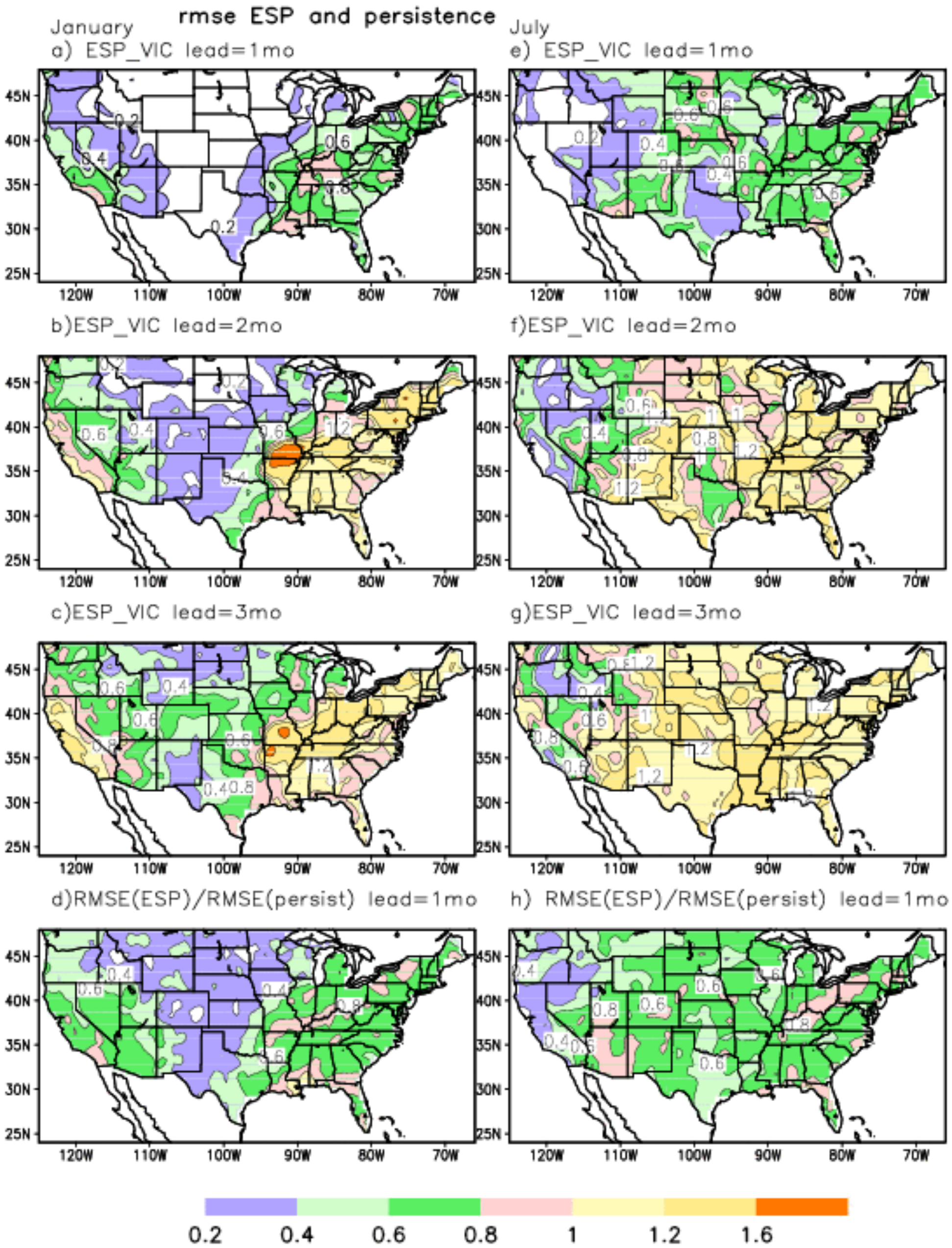

Fig.1: RMSE of ESP_VIC experiments for SM forecasts initialized in January 1 at (a) Lead-1 month, (b) Lead -2 and (c) Lead- 3 , (d) RMSE ratio RMSE(ESP)/RMSE(persistence) for Lead-1 month. (e) -(h) same as (a)-(d) but for July. Contours are indicated by the color bar.

Figure 2

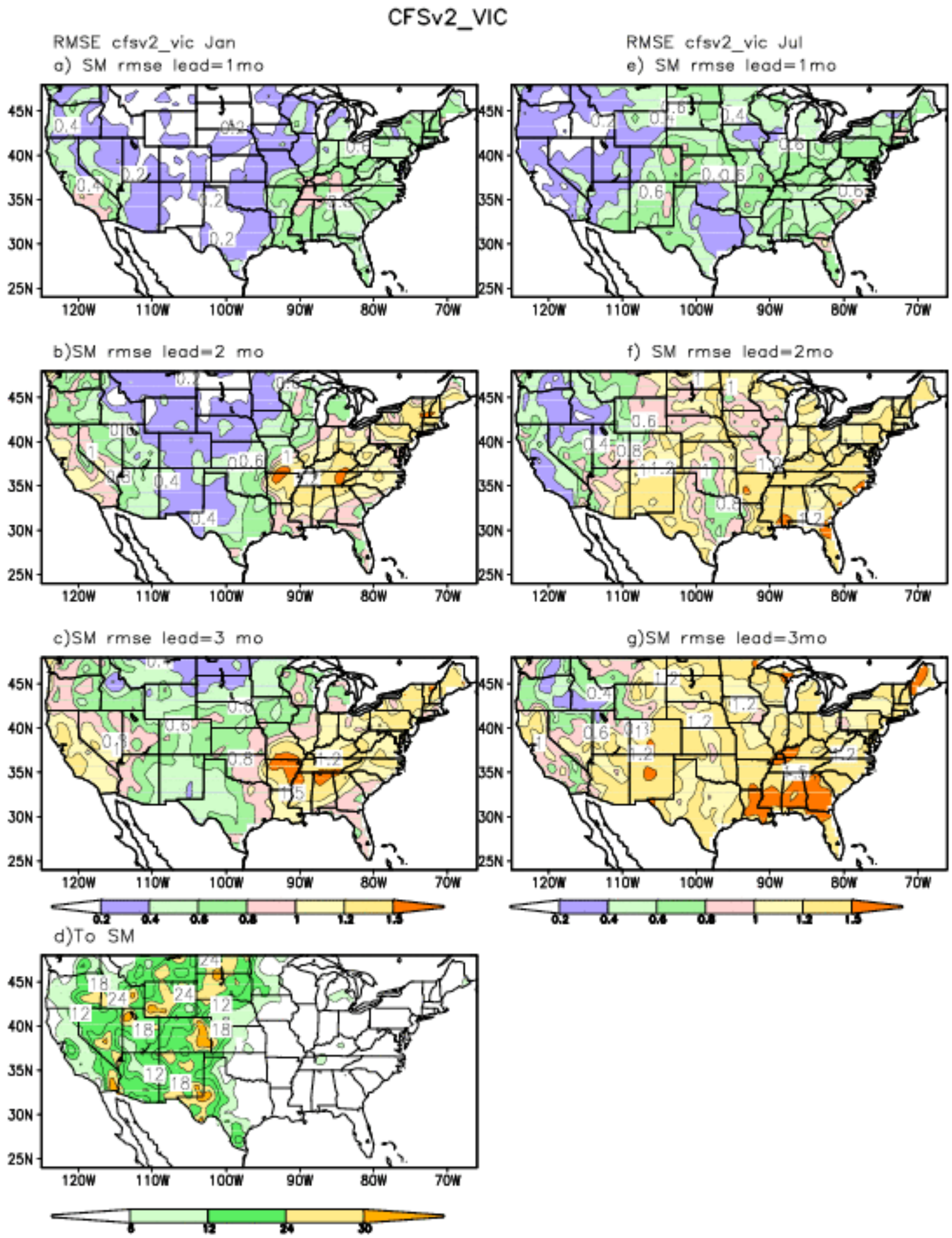

Fig.2: (a)-(c) same as Fig.1 (a)-(c), but for CFSv2_VIC experiment. (d) Characteristic time T_0. Contour interval is 6 months, (e)-(g) same as (a)-(c), but for July.

Figure 3

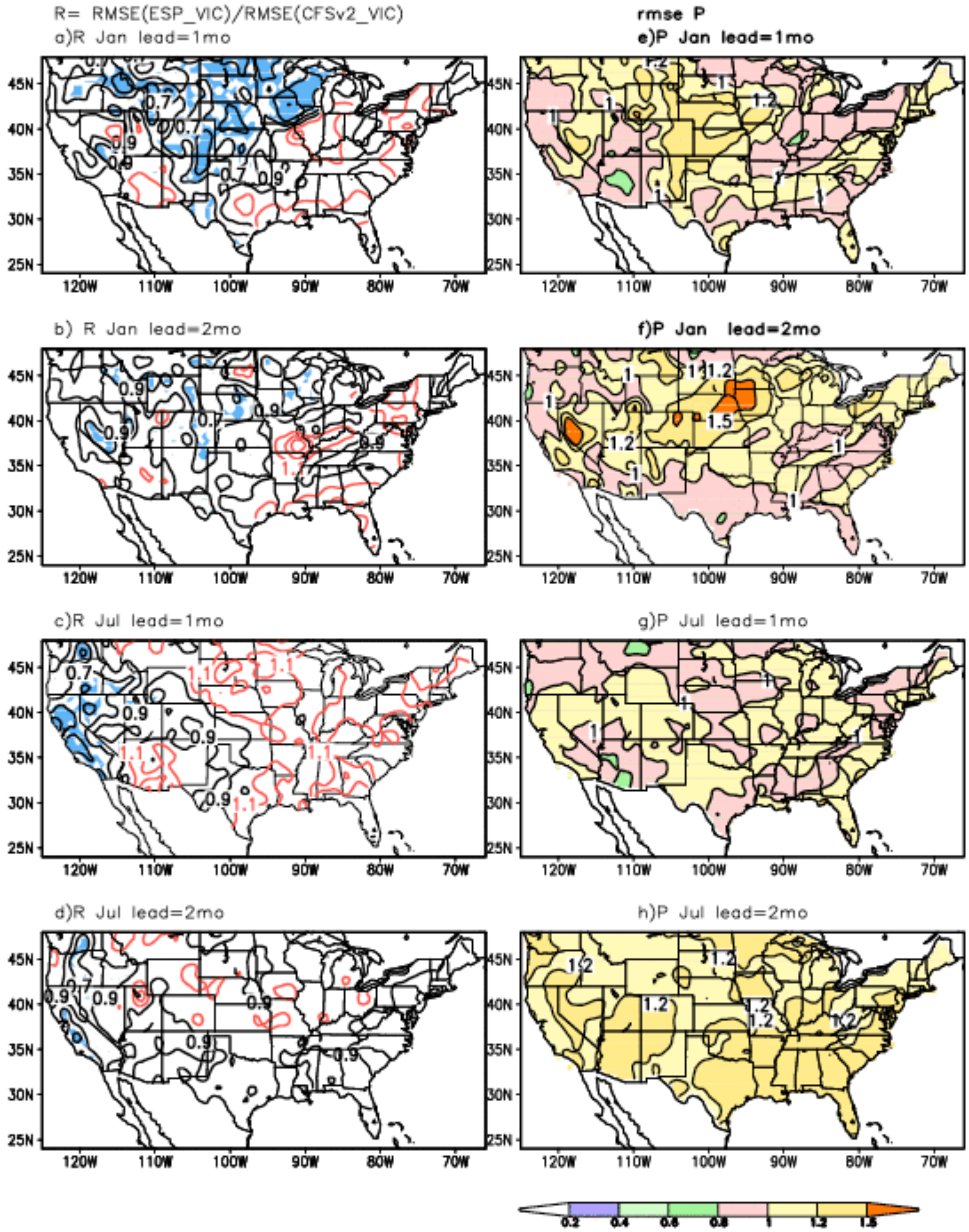

Fig.3: The RMSE ratio RMSE((ESP_VIC)/RMSE(CFS2_VIC) (contoured) for (a) Lead-1 month January forecasts (b) Lead-2 January forecasts. The shading indicates areas that the differences in skill between the ESP_VIC and CFSv2_VIC are statistically significant at the 5% level. (c)-(d) same as (a)-(b), but for July forecasts. (e)-(h)same as (a)-(d), but the RMSE skill for the CFSv2 P forecasts after the BCSD correction. Contour intervals are given by the color bar.