Characteristics of the water cycle and land–atmosphere interactions from a comprehensive reforecast and reanalysis data set: CFSv2

Paul A. Dirmeyer

Received: 28 January 2013 / Accepted: 1 July 2013 / Published online: 14 July 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract The behavior of the water cycle in the Coupled Forecast System version 2 reforecasts and reanalysis is examined. Attention is focused on the evolution of forecast biases as the lead-time changes, and how the lead-time dependent model climatology differs from the reanalysis. Precipitation biases are evident in both reanalysis and reforecasts, while biases in soil moisture grow throughout the duration of the forecasts. Locally, the soil moisture biases may shrink or reverse sign. These biases are reflected in evaporation and runoff. The Noah land surface scheme shows the necessary relationships between evaporation and soil moisture for land-driven climate predictability. There is evidence that the atmospheric model cannot maintain the link between precipitation and antecedent soil moisture as strongly as in the real atmosphere, potentially hampering prediction skill, although there is better precipitation forecast skill over most locations when initial soil moisture anomalies are large. Bias change with lead-time, measured as the variance across ten monthly forecast leads, is often comparable to or larger than the interannual variance. Skill scores when forecast anomalies are calculated relative to reanalysis are seriously reduced

over most locations when compared to validation against anomalies based on the forecast model climate at the corresponding lead-time. When all anomalies are calculated relative to the 0-month forecast, some skill is recovered over some regions, but the complex manner in which biases evolve indicates that a complete suite of reforecasts would be necessary whenever a new version of a climate model is implemented. The utility of reforecast programs is evident for operational forecast systems.

Keywords Water cycle · Seasonal forecast · GCM · Land–atmosphere interactions · Precipitation · Soil moisture · CFSv2

1 Introduction

Evidence exists from a large number of modeling studies, as well as a more limited number of observational studies, that the state of the land surface can affect the atmosphere on intra-seasonal and longer time scales (e.g., Namias 1960; Charney et al. 1975; Shukla and Mintz 1982; Delworth and Manabe 1989; Koster and Suarez 1995, 2004; Douville and Chauvin 2000; Dirmeyer 2003). As with the impact of ocean surface temperature states on the overlying atmosphere, land surface influences on the atmosphere can be a source of predictability (Shukla 1985; Koster et al. 2004a; Dirmeyer 2006). Predictability is a necessary condition for prediction skill (Shukla 1998). For the purposes of operational forecasts, realistic initialization of land surface states can enhance the skill of sub-seasonal to seasonal climate forecasts in certain regions (Koster et al. 2004b, 2011; Dirmeyer 2005; Jeong et al. 2008) because anomalies in states like soil moisture possess a persistence or memory in many regions that is much longer than the

This paper is a contribution to the Topical Collection on Climate Forecast System Version 2 (CFSv2). CFSv2 is a coupled global climate model and was implemented by National Centers for Environmental Prediction (NCEP) in seasonal forecasting operations in March 2011. This Topical Collection is coordinated by Jin Huang, Arun Kumar, Jim Kinter and Annarita Mariotti.

Electronic supplementary material The online version of this article (doi:10.1007/s00382-013-1866-x) contains supplementary material, which is available to authorized users.

P. A. Dirmeyer (✉)
George Mason University, Fairfax, VA, USA
e-mail: pdirmeyer@gmu.edu

typical deterministic range of weather forecast skill (Schlosser and Milly 2002; Dirmeyer et al. 2009).

The National Centers for Environmental Prediction (NCEP) Coupled Forecast System, version 2 (CFSv2) reanalysis and reforecast data sets (CFSRR; Saha et al. 2010) have several characteristics that are highly useful for exploration of this pathway from terrestrial boundary conditions to predictability and prediction skill. First, the reanalysis is semi-coupled, with an offline land data assimilation system that keeps soil moisture and other hydrologic states constrained by observations. Second, the reforecasts use these realistic land surface states as part of their initial conditions. Third, there are many reforecasts in the CFSRR data set—four per day over a span of more than three decades with durations ranging from 45 days to about 10 months. This allows for statistically sound and thorough investigations of the water cycle and land–atmosphere coupling in the context of short-term climate forecasts. Finally, the existence in parallel of a reanalysis and forecasts at various lead times, all valid at the same “real time” and generated by exactly the same global models, allows for clean comparisons of model behavior when unconstrained (forecast mode) versus constrained by observations (via data assimilation).

There have been some preliminary investigations of aspects of the water cycle in CFSRR. Yuan et al. (2011) found CFSv2 ensemble mean precipitation skill to be poor after the first month of reforecast, but overall skill appeared to be better than for other global forecast models, particularly the older version of CFS. Mo et al. (2012) have shown that CFSv2 precipitation reforecast errors, after bias correction, slightly improve soil moisture simulations over the United States with hydrologic models at longer lead times compared to statistical forcings that emphasize the role of the initial hydrologic state. These studies focused on a very limited subset of seasons, forecast lead times and variables. Kumar et al. (2011) performed a more detailed analysis of the original version of CFS, examining precipitation skill at lead times finer than monthly steps.

In this paper the behavior of the water cycle over global land is investigated within the CFSRR framework. Specifically, two aspects of CFSRR are explored. First, the model climatology is examined, with particular interest in how the model climate drifts in forecast mode over the course of months and seasons. Forecast model drift can be seriously detrimental to hydrologic forecasting (Wood and Schaake 2008). Second, the skill of CFSv2 forecasts of water cycle quantities is analyzed with an eye toward to potential impact of the realistic land surface initialization on the forecasts. Finally, possible mechanisms for the realization (or lack of realization) of potential predictability as prediction skill are explored.

Section 2 of this paper gives a brief description of the CFSRR data sets that are used in this study, as well as independent validation data used to assess systematic errors and forecast skill for precipitation. For other water budget quantities, the reanalysis serves as the validation data set for forecasts. The model climate and its drift in forecast mode are explored in Sect. 3. Section 4 analyzes the forecast skill. Mechanisms of the model’s behavior are probed in Sect. 5, and discussion and conclusions are presented in Sect. 6.

2 Models and data

Saha et al. (2010) describe the CFSv2 reanalysis in detail. The reanalysis covers the period from the beginning of December 1978 through the end of 2009. Relevant to this study, there are data assimilation streams for the atmosphere, ocean and land, which are coupled at 6 or 24 h intervals depending on the model component pairing. The land surface data assimilation stream updates the Noah land surface model using observed precipitation in place of the atmospheric model’s guess forecasts, but otherwise uses the near-surface meteorology from the atmospheric stream. At 0000UTC each day the land surface states from the Noah-only assimilation stream are placed back into the fully coupled stream, to prevent terrestrial states from drifting over time due to systematic errors in precipitation or snowpack. This “semi-coupling” of the assimilation stream for land and atmosphere attempts to strike a compromise, achieving high degrees of both agreement with reality and internal model consistency (cf. Koster et al. 2009).

The reanalysis states are used to initialize a very large suite of coupled retrospective forecasts or “reforecasts”. Every 6 h a new forecast is initialized from the reanalysis and launched, thus ensembles are collected from reforecasts with a range of initial dates and times as illustrated in Fig. 1. Reforecasts are started at 0000UTC, 0600UTC, 1200UTC and 1800UTC on 1 January of each year, and every fifth day thereafter beginning in 1982, and continued ostensibly for 9 months, but in actuality run anywhere from 297 to 329 days. On the other 4 days of this 5-day cycle, 3-month reforecasts are initialized at 0000UTC, and 45 day reforecasts start from the other three 6-h intervals during those days.

Here we use only the 9-month forecasts, which are distributed as a limited set of variables at monthly means accumulated from 24 ensemble members for each month (there are 28 members in the November set, to account for the extra pentad as $365/5 = 73$, which does not divide evenly by 12; we use only the first 24 members of the November forecasts to keep the sample sizes consistent

Fig. 1 Schematic of 1 year of CFS reforecasts illustrating four ensembles initialized every pentad. Large numbers 0–9 indicate the validation month for each initial month’s set of forecasts. See text for details

among months). There are actually 10 monthly values given—the first monthly mean is called the “0-month” forecast, which includes ensemble members that started anywhere from 19 to 24 days before the start of “month zero” through 3–7 days after. Thus, the 0-month forecast includes a mix of weather and sub-seasonal climate forecast time scales, and even the initial conditions for a few of the ensemble members, evident from Fig. 1.

Two observational data sets are used to validate precipitation. Over land, the gauge-based Climate Prediction Center (CPC) unified precipitation analysis (Chen et al. 2008) is used. This data set is gridded at 0.5° horizontal resolution, daily temporal resolution which are averaged to monthly, and covers the entire period of the CFSv2 reforecasts. Over ocean, version 2.2 of the Global Precipitation Climatology Project (GPCP) monthly analysis at 2.5° resolution is used (Adler et al. 2003).

For land surface states, validation of the reforecasts is performed against the CFSv2 reanalysis. This likely produces higher skill scores and correlations than validation against an independent product based directly on observations of quantities like soil moisture, but no such global products exist for the reforecast period. However, validation against the reanalysis ensures consistency in the definition of soil moisture in the forecasts (Koster et al. 2009). There exist other global data sets of model-estimated soil moisture for the period (e.g., CPC, Fan and van den Dool 2004; GLDAS-2, Rodell et al. 2004), but they are also the

products of models driven by gridded meteorology derived from some combination of observations and analyses, like that in the CFSv2 reanalysis. Using the CFSv2 reanalysis has the added benefit of providing an easy assessment of drift in the terrestrial water cycle terms.

In this paper, most of the focus is on results of forecasts that validate during the boreal summer months of June through August, as this is the season when land–atmosphere interactions have the largest potential impact on climate when considered globally. However, figures for the other seasons are included as supplementary material for many of the calculations. The overlap between CFSv2 reforecasts and reanalyses is 28 years from 1982 through 2009, but many of the monthly and seasonal calculations use only 27 years (reforecasts initialized during 1982–2008) as the longer forecasts in that last year validate during 2009.

3 Climate and drift

Climate drift in coupled land–atmosphere models is a significant problem that manifests strongly through the water cycle (Dirmeyer 2001). Even in data assimilation mode, there exist shocks and biases that affect the hydrologic cycle (Betts et al. 2006; Bosilovich et al. 2008). So we first look at errors and drift in the CFSRR products.

Figure 2 shows the mean monthly precipitation errors over the 28 years from 1982 to 2009 relative to GPCP over

ocean and the CPC Unified analysis over land for JJA. Other seasons are shown in the supplementary material (Fig. S1). The top panel is the difference between the CFSv2 Reanalysis precipitation and observations. There is an overall positive bias over both ocean and land, but much regional structure is evident in the errors. We show the reanalysis performance as a baseline of model capabilities when constrained by the best available data in assimilation mode.

Fig. 2 Precipitation error of CFS reanalysis (*top*) and reforecasts and various leads validating during JJA 1982–2009. Errors are relative to GPCP over ocean and the CPC Unified analysis over land. Units are mm d^{-1}

The remaining panels show the errors in the CFSv2 reforecasts at leads of 0, 1, 3 and 7 months, as defined in Sect. 2. The errors for the individual months are assessed for the specified reforecast leads, and then averaged together to give seasonal statistics. Over many regions the errors in the 0-month reforecasts are clearly larger than in the reanalysis. Furthermore, the errors usually continue to grow with longer lead times. However, there are exceptions over both ocean (e.g., the equatorial Indian Ocean) and land (e.g., southern United States) where bias may shrink or reverse sign at longer leads. A more striking feature than regional variations in magnitudes is the consistency in the pattern of errors across all lead times. Similar characteristics are present in the other seasons (supplemental material). Such consistency likely reflects robust model biases in atmospheric circulation or thermodynamic quantities.

The root mean square (RMS) error calculated over only land points is larger in the reanalysis than in the CFSv2 reforecasts at all leads by 2–12 % during DJF (maps in the supplemental material) and over 20 % during JJA. Most of the error in the reanalysis occurs as strong positive biases over seasonal monsoon regions, especially southern Africa in DJF and Southeast Asia in JJA. This apparent discrepancy is a well-known aspect of general circulation model behavior in the simulated hours after initialization. Reanalysis precipitation, like all fluxes from reanalyses, is the product of a very short-term forecast (Saha et al. 2010). The model goes through an adjustment period in the first hours, and sometimes days, while the physical parameterizations spin-up and equilibrate to the model's dynamical state. Thus, the precipitation averaged across the early time steps of a model integration, especially for a model being run in reanalysis mode with frequent application of increments from data assimilation, can be quite different than the climatology of the same model running freely without frequent data assimilation (cf. Betts et al. 2006; Zhang et al. 2012).

For the remainder of this paper we will concentrate on the model behavior over land. The precipitation errors indicated in CFSv2 reforecasts in Fig. 2 accumulate in the land surface reservoir of soil moisture, and affect other components of the water cycle. This has major implications for the interpretation of CFSv2 reforecasts for hydrologic applications. Since soil moisture states are initialized for the reforecasts from the reanalysis, we compare the reforecast soil moisture to the reanalysis in Fig. 3. The averaging period is again 1982–2009, and the same reforecast leads are shown for JJA; other seasons are given in the supplemental section (Fig. S2). The drift in soil moisture between the reanalysis values used to initialize the reforecasts and the forecast values clearly increases as lead-times grow longer. Strong wet biases grow across

Europe, the Sahel, and especially North America, while much of the Amazon basin and parts of South Asia show the strongest dry biases.

The global drift in soil moisture is synthesized in Fig. 4 for all seasons. The divergence with lead-time between the climatologies of the predicted soil moisture and reanalysis is grouped by season for each of the four soil layers of the Noah land surface scheme used in CFSv2. The spatial RMS difference is calculated on the CFSv2 reforecast grid across all land grid points except Antarctica. Generally speaking, the magnitude of the drift is largest in the deepest soil layer, and smallest in the surface layer, although this sorting often takes several months to settle out. For reforecasts of boreal winter months, the drift in the 10–40 cm layer remains slightly larger than the 40–100 cm layer out to 10 months lead-time. The RMS differences in the first month (0-month reforecast) grow by an additional 40–110 % during the

remainder of the forecast period, depending on the season and soil layer.

This drift in soil moisture, driven by systematic errors in precipitation, affects the surface water fluxes of runoff and evapotranspiration. To give an idea of the relative magnitude of the drift, the variance is compared across all reforecast lead-times having the same validation month to the interannual variance for the corresponding month from the reanalysis, which is essentially the climate “signal”. Figure 5 shows the seasonal average of the monthly ratios for reforecasts valid in JJA for precipitation, soil moisture (layer 2; 10–30 cm depth), evapotranspiration and runoff. Again, the other seasons are shown in the supplemental section (Fig. S3). Small values indicate little variance in the reforecast quantities across forecast lead times from 0 to 9 months, and thus little apparent drift. No significance testing is performed as no null hypothesis is tested—the comparison to interannual variability is used as a benchmark for “natural” variations against which drift is compared.

Over many regions soil moisture variance across lead times is from 35 % to more than 100 % the interannual variance. Although no longer evident in precipitation during JJA, systematic errors in snowfall are evident in soil moisture, runoff and even in evapotranspiration in roughly zonal bands across northern Eurasia and northern North America; this snow bias is discussed further in the next section. In middle and low latitudes there is evidence of moderate drift in precipitation over many locations, and strong drift relative to interannual variability over the dry season regions around parts of the Mediterranean, Middle East and southwestern Africa. Many areas have inter-lead variance larger than interannual variance for the other water cycle variables—there is more variability in the growing biases than in the climate signal. Some areas such as northern India and the lower La Plata river basin show much stronger drift in evapotranspiration than the other terms.

It should be recalled that many factors affect soil moisture and surface water flux variability, and thus their drift, besides precipitation. Systematic errors and drifts in surface radiation (clouds), temperature and humidity can also contribute to drift, and frozen soil can prevent soil moisture drift from being realized in evapotranspiration or runoff. Nevertheless, it is clear that for many if not most areas, drift in the model climatology with reforecast lead-time must be taken into account when interpreting the anomalies suggested by the reforecasts.

To show how these drifts evolve, and how they compare with the CFSv2 reanalysis from which the reforecasts are initialized, area means are calculated for the monthly mean layer 2 soil moisture at various lead times, color-coded by the reforecast lead (Fig. 6). Shown in black are the mean values from the CFSv2 reanalysis. The vertical bars show the interannual variability (plus/minus one standard deviation).

Fig. 3 As in Fig. 2 for JJA volumetric soil moisture (10–40 cm layer) forecast errors relative to CFS reanalysis

Fig. 4 Root mean square difference in layer volumetric soil moisture between CFS reanalysis and reforecasts verifying in the same month, averaged over all land grid points excluding Antarctica and averaged for seasons

The annual cycles for three regions are shown. At the top of Fig. 6, the average over land points centered on the Indian subcontinent shows two distinct semi-annual periods of drift. Beginning in July, reforecasts at all leads show drier soils than reanalysis, with the dry bias growing stronger with increasing lead. Toward the end of the year, the shorter leads begin to shift to a wet bias, with all reforecast leads being wetter than reanalysis for validation months February through June. Furthermore, there is more interannual variability in the reforecast values than indicated in reanalysis.

The second panel shows the same evolution of soil moisture over the Great Plains of North America. Here the reforecasts are uniformly biased toward wetter values than reanalysis for all months and lead-times, mirroring biases in precipitation. The biases generally grow with lead-time, and the biggest jump is in the first month of the reforecasts. The bottom panel shows the forecast drift for a region further north and east across much of southern Canada. Here the melting snowpack is a significant source of soil moisture and determines much of the annual cycle. A negative drift is evident during the cold season that reverses to a positive bias in the warm season. There is a heavy snow bias, a cold bias during spring and late snow melt that tend to shift the phase of the annual cycle of soil moisture progressively later with increasing reforecast lead-time.

Other regions of the globe show interesting seasonal cycles of error and drift in the surface water cycle, but the examples in Fig. 6 give some idea of the range of causes and effects. The interannual variability at any month is generally smaller than the annual cycle, and frequently smaller than the spread among the climatologies across different lead-times. The key point is that CFSv2 has a climatological annual cycle that is itself a function of reforecast lead. This extra time dimension should always be

considered when interpreting forecasts, as will be illustrated in the following section.

4 Skill

To quantify reforecast skill, a discrete ranked probability skill score (RPSS_D, Weigel et al. 2007) is used with three categories of equal likelihood based on 27 years of data—above normal, near normal and below normal. Probabilistic forecasts are based on the ensembles of size 24. For this forecast configuration, the ranked probability score of a climatological forecast is exactly 4/9, and bias due to the finite ensemble size is 1/54 (Weigel et al. 2007). The CFS reforecast ranked probability scores are normalized by the expected climatological score, corrected for finite ensemble bias, and the ratio is subtracted from unity so that a positive value indicates skill above and beyond a climatological forecast.

A key determinant of skill scores is how the validation categories are defined. Of course, the proper way to estimate whether a particular model forecast is above, near or below normal, or in other sets of categories if the divisions are not in terciles, is relative to the models' own climatology. This is in fact the primary motivation for the reforecast project, to provide a forecast model climatology for CFSv2. However, for a variety of reasons, there are situations when a complete set of reforecasts are not generated. A forecast model may undergo small incremental changes and improvements on a relatively frequent basis, where the burden of repeatedly generating new datasets of reforecasts is large. Even when model updates are infrequent, the encumbrance of producing a statistically meaningful set of sample reforecasts can be prohibitive.

Fig. 5 Ratio of the variance across all forecast leads validating in June, July and August, averaged across the 3 months, to the interannual variance from the CFS reanalysis. The ratio is expressed as a percentage for each of the variables labeled

The CFSv2 reforecast data set provides an excellent platform to explore the effect of such validation shortcuts to assessing model skill. To do so, we estimate the boundaries between the terciles in several different ways. The most appropriate way is based on the models' own climatology, which we have seen varies as a function of forecast lead-time as well as time of the year. So if the reforecast precipitation, for instance, would fall in the model's own upper tercile for that month and lead-time,

Fig. 6 Evolution of 10–40 cm soil moisture averaged over land points in the boxes indicated on the inset maps for the CFS reanalysis (black) and forecasts initialized in each of the months. Colors indicate the lead-time of the forecast as indicated in the legend, and vertical bars and whiskers show ± 1 interannual standard deviation

that would be the forecast, even if the model rain rate would correspond to the middle or lower tercile among observed values. A shortcut often applied is to use instead the climatology of the reanalysis generated by the same model as the forecasts. Lacking any information about model behavior, the validating observations themselves are often used. Naturally, the observations are used for establishing the terciles in the validation data.

Figure 7 shows $RPSS_D$ of 0-month lead precipitation CFSv2 reforecasts validating during June, July and August estimated three different ways. In the left column at the top, the terciles for estimating the reforecast ranked probability score are calculated from the reforecast climatology at 0-month lead. In the middle panel, the terciles are based on the CPC Unified gridded precipitation observations themselves. Discrepancies are widespread, and reflect the biases not only in mean precipitation but its probability density function in the forecast model. Other seasons are shown in the supplementary material (Figs S4–S6). Regions

where the skill score is comparable across the left-hand panels (e.g., southern Australia) are regions where CFSv2 simulates precipitation means and interannual variability well (compare to Fig. 2). In the bottom panel, the terciles are based on the CFS reanalysis. The apparent model skill is as poor as when observed terciles are used, and in many places worse. As shown in Fig. 2, the reanalysis climate, constrained tightly by data assimilation in many places, can be quite different from the free-running model in forecast mode.

The right column of Fig. 7 shows the skill of reforecasts validating during JJA at longer lead times; 1 month (top); 2 months (middle) and 3 months (bottom). Here, $RPSS_D$ is calculated in the same way as the top left panel, using the forecast model climatology at the corresponding lead to define the terciles. Except for some tropical regions where precipitation is strongly determined by nearby ocean temperatures, skill drops off quickly after month 0. In fact, month 0 includes the classical weather forecast time scales

Fig. 7 Discrete rank probability skill score for monthly precipitation forecasts, based on terciles of equal population during 1982–2008, averaged for June, July and August. The lead-time for the forecasts is

shown in the lower left of each panel, and the data used to establish the boundaries of the terciles is shown to the right, including lead-time in the case of model forecasts

for many of the ensemble members; the inclusion of these deterministic forecast time scales in these probabilistic forecasts greatly enhances skill scores.

Similar results are shown for 10–40 cm soil moisture (Fig. 8, S7–S9) and runoff (Fig. 9, S10–S12). There are no complete observed global data sets for either soil moisture or runoff, so we validate against the reanalysis values—the land surface state variables and fluxes in the CFS reanalysis are the product of an offline land data assimilation described in Sect. 2. Here we show 1-month lead forecast RPSS_D in the left column, as the autocorrelation time scales for these variables is substantially longer than for precipitation. The right column shows skill scores for 0, 3 and 5 month leads.

Again, the skill for many areas is appreciably lower when the reanalysis is used as the basis for determining forecast terciles (lower left panels in Figs. 8, 9), except over mostly arid regions for soil moisture. This is because over areas with little to no precipitation, the systematic errors in precipitation are not a factor for biasing soil moisture. Instead, soil moisture biases come from other factors (e.g., errors in evapotranspiration formulations, runoff, or the vertical

diffusion of water in the soil) that are the same in the reforecasts and reanalysis, and thus cancel out in terms of formulating terciles and calculating skill scores.

The middle left panels in these figures show the skill scores when the 0-month model forecast climatology is used instead of the 1-month climatology to estimate terciles. This could be thought of as a “shortcut” under the assumption that most of the model drift occurs during the first month. The resulting skill calculated is considerably higher than when reanalysis is used as the basis of estimating terciles, but can still be much worse than using the model climatology from the appropriate forecast lead (e.g., over the northern Amazon Basin).

Overall, reforecast skill for soil moisture is found to be most persistent over semi-arid and arid regions where initial anomalies have the longest autocorrelation time scales. Runoff skill (Fig. 9) is most persistent in many high-latitude and high-altitude regions, where frozen soils and snowpack anomalies can induce persistent anomalies in runoff, as well as over those tropical regions where precipitation forecasts maintain skill well beyond the first month.

Fig. 8 As in Fig. 7 for layer-2 soil moisture. Note the differences in the forecast lead times and validation terciles from Fig. 7

Fig. 9 As in Fig. 8 for runoff

5 Mechanisms

Some clues as to the behavior of CFSv2 skill in predicting components of the water cycle can be found by examining metrics of land–atmosphere interaction. Previous studies have shown that the coupling between land and atmosphere, the ability of land surface states to affect climate on sub-seasonal to seasonal time scales, requires the presence of two feedback “legs”—a connection of surface fluxes to land surface states such as soil moisture (Guo et al. 2006), and a response of the atmosphere to surface fluxes (Guo et al. 2006; Santanello et al. 2009). Zhang et al. (2011) demonstrated that the coupling of the atmosphere and land components of CFSv2 is relatively weak. Wei et al. (2010) showed that the Noah land surface scheme shows weaker coupling than some other land surface schemes when coupled to the same atmospheric model.

The greatest sensitivity of fluxes to soil moisture is expected to occur at intermediate values of soil wetness (e.g., Koster et al. 2000; Dirmeyer et al. 2009). That is indeed the case in Fig. 10, which shows the terrestrial

coupling index defined by Dirmeyer (2011) calculated over all non-forest grid boxes between 25° and 50°N over North America for the first month of forecasts verifying during July. The peak sensitivity of surface daily latent heat flux variations to soil moisture occurs around climatological volumetric soil wetness values of 0.2. Thus, the first leg of the feedback pathway from soil moisture to precipitation appears to be in place in CFSv2.

The connection through to the atmosphere, however, was seen to be weak when measured in terms of forecast skill beyond weather time scales (Zhang et al. 2011). Figure 11 shows one possible reason for this. Correlations among three different quantities verifying during the boreal summer months (JJA) during the 27 years 1982–2008 are shown. Concentrating first on the right column, we see relatively high correlations between the initial soil moisture used in the CFSv2 reforecasts and subsequent monthly mean observed precipitation from the CPC Unified data set at various leads. The 95 % confidence level is at a correlation of 0.11. Basically all but the faintest colors indicate significance, which cover more than half of the land area in

Fig. 10 Terrestrial coupling index (the product of the standard deviation of daily latent heat flux and the correlation between daily latent heat flux and soil moisture; units of Wm^{-2}) as a function of the mean volumetric soil wetness for land grid points over North America between 25° and 50°N during July

the plot through the $\approx 18\text{--}47$ day lead. The gradual breakdown of precipitation skill derived from forecast initial states is evident on intra-seasonal time scales, much like what was found by Kumar et al. (2011) for the previous version of CFS. The large regions with the highest correlations, at least out to the $\approx 13\text{--}42$ day lead, are over Europe, North America, and East Asia, but there are also other regions where strong correlations persist. This pattern corresponds well with rain gauge density, and may reflect, at least in part, how the initial soil moisture is affected by the quality of offline precipitation forcing (cf. Oki et al. 1999; Koster et al. 2011).

The middle column shows the comparable correlation between model forecast monthly precipitation and initial soil moisture. Here, more than half of the area is significant only for the first two leads, and the correlation coefficients are generally much lower. The implication is that there is less persistence, or more noise, in the model precipitation forecasts compared to observations at the monthly time scale. The left column shows the correlation between forecast and observed precipitation, which mirrors the weakness seen between forecast precipitation and initial soil moisture.

We have repeated these correlation calculations only for the forecasts, determined independently at each grid point, where the initial soil moisture lies in either the highest or lowest 20 % (quintile) of the range for those months. The difference of the correlations for the extreme initial soil moisture cases from those shown in Fig. 11 for all forecasts is presented in Fig. 12. The sample sizes are different, so significance for the magnitude of change in the correlation

is not well defined. Rather, consider the proportion of red versus blue in the plots. At all leads, but particularly at shorter leads, red is more widespread than blue in the right column, suggesting that correlations between initial soil moisture and subsequent precipitation are even stronger when the initial soil moisture anomalies are large. Again, the same effect is evident for forecast precipitation, suggesting soil moisture extremes could have a positive impact on forecast skill, but not as strongly as for in observations. After the first three leads, the areas of red and blue colors equilibrate.

There is another strong bias in the water budget in CFSv2. The model systematically over-forecasts the amount of snow. Figure 13 shows the one-month bias over land (Antarctica excluded) in snow water equivalent, in mm of liquid water, for forecasts validating in each of the seasons indicated. Comparisons are to the CFSv2 reanalysis, which assimilates snow data. Nearly everywhere that snow falls, there are positive biases established in the first month. The biases grow steadily with lead-time over most locations. These influence soil moisture states in the spring, contributing to positive biases then (compare to the bottom panel of Fig. 6). The delayed impacts on the surface water cycle can be consequential for climate predictability (Xu and Dirmeyer 2013).

6 Summary and discussion

The behavior of water cycle variables in CFSv2 reforecasts and reanalysis is examined. Model forecast biases evolve as lead-time increases, and may differ substantially from the reanalysis. Precipitation biases arise immediately. For other variables, there are no true global observed data sets, so we use the reanalysis as the basis to calculate biases. For soil moisture, snowpack, evaporation and runoff, biases generally grow throughout the reforecasts. However, they can decrease or change sign during the course of the 9-month forecasts in many regions.

Execution of a large suite of reforecasts is expensive, and it has been common to use reanalysis or observations as the climatology against which forecast model anomalies are reckoned. Skill scores are shown here to be highly dependent on the method of calculating anomalies. Skill scores are much higher when CFSv2 reforecast anomalies are calculated relative to the reforecast climatology at the corresponding lead-time than when they are calculated relative to the reanalysis. A short cut could be to assume that the biases in the first month (0-month forecast) contain most of the model drift. This does help in some situations, but because of the convoluted evolution of biases in many locations and seasons, it often falls far short of using the forecast model climatology at the appropriate lead-time.

Fig. 11 Pair-wise correlations between monthly CFSv2 reforecast precipitation (P_{CFS}), observed precipitation (P_{Obs}) and reforecast initial soil moisture in layer 2 (10–40 cm depth; SM_{IC}), as indicated

above each column, for forecasts validating during JJA, grouped by forecasts leads in days as indicated to the left of each row. Dark colors (beyond ± 0.11) are significant at the 95 % confidence level

Fig. 12 As in the last two columns of Fig. 11, but for the change in correlation when only ensemble members with initial soil moisture in the lowest or highest quintile are considered

The Noah land surface scheme shows sensitivity of evaporation to soil moisture in the transition zone between arid and humid regions, a necessary condition for land-

driven climate predictability and prediction skill (Koster et al. 2004a, 2011). However, the correlation between initial soil moisture and future precipitation drops much

Fig. 13 One-month lead forecast bias in snow water equivalent for all forecasts validating in the months indicated. Units are kg m^{-2}

more quickly over the first 7 weeks in CFSv2 than for observations. This suggests that there may be potential prediction skill that the model is not able to realize. The source of impediment is not diagnosed here. It could be the weakness in lagged correlation is a symptom of a problem apart from land–atmosphere coupling, such as the cloud or convection parameterizations. There is demonstrably more precipitation forecast skill over most locations when initial soil moisture anomalies are large than when they are small. This suggests the feedback of the land state on the atmosphere is not completely shut down in the model.

For applications such as hydrologic forecasting, this study touches both issues of initial condition impact and inherent climate model forecast skill (Shukla and Lettenmaier 2011; Mo et al. 2012). The CFSv2 reforecasts are shown to have significant skill in key hydrologic variables such as precipitation in the first month (consistent with Yuan et al. 2011), and in runoff and soil moisture in many locations for several months, but only when the evolving bias climatology is considered and accounted for. A number of studies have considered the effect of model bias on forecast skill and even the interpretation of forecasts (e.g., Wood and Schaake 2008), but this second time dimension of bias has not been directly recognized in most previous studies, even if it is accounted for implicitly in the bias correction. The design of the CFSv2 reforecast suite is ideally suited to expose this issue; knowledge of the time

evolving biases can improve the estimation of forecast anomalies and skill scores. The benefits of executing a complete reforecast suite are clear whenever a model version is changed in an operational climate forecast system.

Acknowledgments I thank Jin Huang for motivating this study though her leadership coordinating the CFSv2 Evaluation Workshop, along with workshop organizers Annarita Mariotti, Wanqiu Wang, Shrinivas Moorthi and James Kinter, held April 30 and May 1, 2012 in Riverdale, Maryland. Credit for the conceptualization of Fig. 1 goes to Jennifer M. Adams. This work was supported by joint funding of the Center for Ocean Land Atmosphere Studies (COLA) from the National Science Foundation (ATM-0830068), the National Oceanic and Atmospheric Administration (NA09OAR4310058), and the National Aeronautics and Space Administration (NNX09AN50G).

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Adler RF, Huffman GJ, Chang A, Ferraro R, Xie P, Janowiak J, Rudolf B, Schneider U, Curtis S, Bolvin D, Gruber A, Susskind J, Arkin P (2003) The version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J Hydrometeorol* 4:1147–1167
- Betts AK, Zhao M, Dirmeyer PA, Beljaars ACM (2006) Comparison of ERA40 and NCEP/DOE near-surface datasets with other

- ISLSCP-II datasets. *J Geophys Res* 111:D22S04. doi:[10.1029/2006JD007174](https://doi.org/10.1029/2006JD007174)
- Bosilovich MG, Chen J, Robertson FR, Adler RF (2008) Evaluation of global precipitation in reanalyses. *J Appl Meteor Clim* 47:2279–2299
- Charney J, Stone PH, Quirk WJ (1975) Drought in the Sahara: a biogeophysical feedback mechanism. *Science* 187:434–435
- Chen M, Shi W, Xie P, Silva VBS, Kousky VE, Higgins RW, Janowiak JE (2008) Assessing objective techniques for gauge-based analyses of global daily precipitation. *J Geophys Res* 113:D04110. doi:[10.1029/2007JD009132](https://doi.org/10.1029/2007JD009132)
- Delworth T, Manabe S (1989) The influence of soil wetness on near-surface atmospheric variability. *J Clim* 2:1447–1462
- Dirmeyer PA (2001) Climate drift in a coupled land-atmosphere model. *J Hydrometeor* 2:89–100
- Dirmeyer PA (2003) The role of the land surface background state in climate predictability. *J Hydrometeor* 4:599–610
- Dirmeyer PA (2005) The land surface contribution to boreal summer season predictability. *J Hydrometeor* 6:618–632
- Dirmeyer PA (2006) The hydrologic feedback pathway for land-climate coupling. *J Hydrometeor* 7:857–867
- Dirmeyer PA, Schlosser CA, Brubaker KL (2009) Precipitation, recycling and land memory: an integrated analysis. *J Hydrometeor* 10:278–288
- Dirmeyer PA (2011) The terrestrial segment of soil moisture-climate coupling. *Geophys Res Lett* 38:L16702. doi:[10.1029/2011GL048268](https://doi.org/10.1029/2011GL048268)
- Douville H, Chauvin F (2000) Relevance of soil moisture for seasonal climate predictions: a preliminary study. *Clim Dyn* 16:719–736
- Fan Y, van den Dool H (2004) The CPC global monthly soil moisture data set at 1/2 degree resolution for 1948-present. *J Geophys Res* 109. doi:[10.1029/2003JD004345](https://doi.org/10.1029/2003JD004345)
- Guo Z, Dirmeyer PA, Koster RD, Bonan G, Chan E, Cox P, Davies H, Gordon T, Kanae S, Kowalczyk E, Lawrence D, Liu P, Lu S, Malyshev S, McAvaney B, Mitchell K, Oki T, Oleson K, Pitman A, Sud Y, Taylor C, Verseghy D, Vasic R, Xue Y, Yamada T (2006) GLACE: the global land-atmosphere coupling experiment. Part II: analysis. *J Hydrometeor* 7:611–625
- Jeong J-H, Ho C-H, Chen D, Park T-W (2008) Land surface initialization using an offline CLM3 simulation with the GSWP-2 forcing dataset and its impact on CAM3 simulations of the boreal summer climate. *J Hydrometeor* 9:1231–1248
- Koster RD, Suarez MJ (1995) Relative contributions of land and ocean processes to precipitation variability. *J Geophys Res* 100:13775–13790
- Koster RD, Suarez MJ (2004) Suggestions in the observational record of land-atmosphere feedback operating at seasonal timescales. *J Hydrometeor* 5:567–572
- Koster RD, Suarez MJ, Heiser M (2000) Variance and predictability of precipitation at seasonal-to-interannual timescales. *J Hydrometeor* 1:26–46
- Koster RD, Dirmeyer PA, Guo Z, Bonan G, Chan E, Cox P, Davies H, Gordon T, Kanae S, Kowalczyk E, Lawrence D, Liu P, Lu S, Malyshev S, McAvaney B, Mitchell K, Oki T, Oleson K, Pitman A, Sud Y, Taylor C, Verseghy D, Vasic R, Xue Y, Yamada T (2004a) Regions of strong coupling between soil moisture and precipitation. *Science* 305:1138–1140
- Koster RD, Suarez MJ, Lu P, Jambor U, Berg A, Kistler M, Reichle R, Rodell M, Famiglietti J (2004b) Realistic initialization of land surface states: impacts on subseasonal forecast skill. *J Hydrometeor* 5:1049–1063
- Koster RD, Guo Z, Dirmeyer PA, Yang R, Mitchell K, Puma MJ (2009) On the nature of soil moisture in land surface models. *J Clim* 22:4322–4335
- Koster RD, Mahanama SPP, Yamada TJ, Balsamo G, Berg AA, Boissiere M, Dirmeyer PA, Doblas-Reyes FJ, Drewitt G, Gordon CT, Guo Z, Jeong J-H, Lee W-S, Li Z, Luo L, Malyshev S, Merryfield WJ, Seneviratne SI, Stanelle T, van den Hurk BJJM, Vitart F, Wood EF (2011) The second phase of the Global Land-Atmosphere Coupling Experiment: soil moisture contributions to subseasonal forecast skill. *J Hydrometeor* 12:805–822. doi:[10.1175/2011JHM1365.1](https://doi.org/10.1175/2011JHM1365.1)
- Kumar A, Chen M, Wang W (2011) An analysis of prediction skill of monthly mean climate variability. *Clim Dyn* 37:1119–1131
- Mo K, Shukla S, Lettenmaier DP, Chen L-C (2012) Do Climate Forecast System (CFSv2) forecasts improve seasonal soil moisture prediction? *Geophys Res Lett* 39:L23703. doi:[10.1029/2012GL053598](https://doi.org/10.1029/2012GL053598)
- Namias J (1960) Factors in the initiation, perpetuation and termination of drought. [Extract of publication No. 51, I.A.S.H. Commission of Surface Waters], pp 81–94
- Oki T, Nishimura T, Dirmeyer P (1999) Assessment of annual runoff from land surface models using Total Runoff Integrating Pathways (TRIP). *J Meteor Soc Japan* 77:235–255
- Rodell M, Houser PR, Jambor U, Gottschalk J, Mitchell K, Meng C-J, Arsenault K, Cosgrove B, Radakovich J, Bosilovich M, Entin JK, Walker JP, Lohmann C, Toll D (2004) The global land data assimilation system. *Bull Am Meteor Soc* 85:381–394
- Saha S et al (2010) The NCEP climate forecast system reanalysis. *Bull Amer Meteor Soc* 91:1015–1057. doi:[10.1175/2010BAMS3001.1](https://doi.org/10.1175/2010BAMS3001.1)
- Santanello JA, Peters-Lidard CD, Kumar SV, Alonge C, Tao W-K (2009) A modeling and observational framework for diagnosing local land-atmosphere coupling on diurnal time scales. *J Hydrometeor* 10:577–599
- Schlosser CA, Milly PCD (2002) A model-based investigation of soil moisture predictability and associated climate predictability. *J Hydrometeor* 3:483–501
- Shukla J (1985) Predictability. *Adv Geophys* 28B:87–122
- Shukla J (1998) Predictability in the midst of chaos: a scientific basis for climate forecasting. *Science* 282:728–731
- Shukla J, Mintz Y (1982) Influence of land-surface evapotranspiration on the earth's climate. *Science* 215:1498–1501
- Shukla S, Lettenmaier DP (2011) Seasonal hydrologic prediction in the United States: understanding the role of initial hydrologic conditions and seasonal climate forecast skill. *Hydrol Earth Syst Sci* 15:3529–3538
- Wei J, Dirmeyer PA, Guo Z, Zhang L, Misra V (2010) How much do different land models matter for climate simulation? Part I: climatology and variability. *J Clim* 23:3120–3134
- Weigel AP, Liniger MA, Appenzeller C (2007) The discrete Brier and ranked probability skill scores. *Mon Wea Rev* 135:118–124
- Wood AW, Schaake JC (2008) Correcting errors in streamflow forecast ensemble mean and spread. *J Hydrometeor* 9:132–148. doi:[10.1175/2007JHM862.1](https://doi.org/10.1175/2007JHM862.1)
- Xu L, Dirmeyer P (2013) Snow-atmosphere coupling strength. Part II: Albedo effect versus hydrological effect. *J Hydrometeor* 14:404–418. doi:[10.1175/JHM-D-11-0103.1](https://doi.org/10.1175/JHM-D-11-0103.1)
- Yuan X, Wood EF, Luo L, Pan M (2011) A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction. *Geophys Res Lett* 38:L13402. doi:[10.1029/2011GL047792](https://doi.org/10.1029/2011GL047792)
- Zhang L, Dirmeyer PA, Wei J, Guo Z, Lu C-H (2011) Land-atmosphere coupling strength in the Global Forecast System. *J Hydrometeor* 12:147–156. doi:[10.1175/2010JHM1319.1](https://doi.org/10.1175/2010JHM1319.1)
- Zhang L, Kumar A, Wang W (2012) Influence of changes in observations on reprecipitation: a case study for the Climate Forecast System Reanalysis (CFSR). *J Geophys Res* 117:D08105. doi:[10.1029/2011JD017347](https://doi.org/10.1029/2011JD017347)