Soil Moisture, Snow, and Seasonal Streamflow Forecasts in the United States

SARITH MAHANAMA

Science Applications International Corporation, Beltsville, and Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

BEN LIVNEH

Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington

RANDAL KOSTER

Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

DENNIS LETTENMAIER

Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington

ROLF REICHLE

Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 28 April 2011, in final form 26 August 2011)

ABSTRACT

Land surface model experiments are used to quantify, for a number of U.S. river basins, the contributions (isolated and combined) of soil moisture and snowpack initialization to the skill of seasonal streamflow forecasts at multiple leads and for different start dates. Snow initialization has a major impact on skill during the spring melting season. Soil moisture initialization has a smaller but still statistically significant impact during this season, and in other seasons, its contribution to skill dominates. Realistic soil moisture initialization can contribute to skill at long leads (over 6 months) for certain basins and seasons. Skill levels in all seasons are found to be related to the ratio of initial total water storage (soil water plus snow) variance to the forecast period precipitation variance, allowing estimates of the potential for skill in areas outside the verification basins.

1. Introduction

Over the past several decades, great strides have been made in forecasting streamflow at seasonal leads, largely through the incorporation of climate information into hydrologic forecasts. Improved seasonal streamflow forecasts have many societal and economic benefits. Yao and Georgakakos (2001), for example, demonstrated that improved forecasts can increase hydropower revenues. Hamlet et al. (2002) showed that an alternative operating system that exploited climate forecasts could significantly increase nonfirm energy production from a major Columbia River hydropower dam. Improved forecasts have an obvious impact on the ability of reservoir operators to mitigate the destructive capacity of floods and droughts.

Several potential contributors to skill in streamflow forecasts can be identified. Skill can be derived, for example, from the accurate forecasting of meteorological anomalies (particularly precipitation) during the forecast period. A second contributor is the quantification of snowpack; in the western United States, for example, precipitation is winter dominant, and thus a large fraction [by some accounts more than 70% (Christensen and Lettenmaier 2006)] of streamflow there originates from melting snow. It is no surprise, then, that western water resources managers rely heavily on wintertime snow surveys to project water availability in the subsequent spring

Corresponding author address: Sarith Mahanama, Global Modeling and Assimilation Office, Mail Stop 610.1, NASA Goddard Space Flight Center, Greenbelt, MD 20771.
E-mail: sarith.p.mahanama@nasa.gov

DOI: 10.1175/JHM-D-11-046.1

© 2012 American Meteorological Society

(Brumbelow and Georgakakos 2001). A third potential contributor is knowledge of soil moisture. If the soil is dry, incident water at the surface (either snowmelt or precipitation) may infiltrate the soil and later evaporate rather than run off into streams; a wet soil, on the other hand, may encourage greater streamflow and a more efficient filling of reservoirs.

Our ability to take advantage of the first contributor is strongly limited by the minimal skill levels achieved to date in forecasting continental precipitation (and temperature, in snow-dominant areas) at seasonal time scales; indeed, nature, through chaotic atmospheric dynamics, imposes upper limits to the meteorological forecast skill that we can ever hope to achieve. In contrast, we can, at least in principle, estimate water contents in snow and in the soil on the forecast start date with some accuracy, making more tenable the use of the second and third contributors. A number of analyses (e.g., Pagano et al. 2004; Pagano and Garen 2005; Pagano et al. 2009; Wood and Lettenmaier 2008; Li et al. 2009; Bierkens and van Beek 2009) have analyzed the seasonal streamflow forecast skill derived from accurate estimates of snow and other water present in a basin on the forecast issue date. Some recent studies have highlighted in particular the contribution of soil moisture initialization to streamflow forecast skill, both in snow-covered and snow-free areas (Berg and Mulroy 2006; Mahanama et al. 2008). Using a multiple regression approach, Maurer and Lettenmaier (2003) characterized the joint contributions from soil moisture initialization and seasonal climate forecasts in the Mississippi River basin and found that soil moisture dominates runoff predictability for lead times of 1–2 months.

In a recent paper, Koster et al. (2010) used a suite of state-of-the-art land surface models, a multidecadal dataset of meteorological forcing, and time series of streamflow observations in 17 basins (ranging in size from 2000 to 1.4 million km^2) to study the sources of streamflow forecast skill—in particular, to isolate and quantify the contributions of realistic 1 January snow and soil moisture initialization to the forecast skill achieved during the snowmelt season (March–July) in the western United States. We expand here on that study, adding to it and to the overall literature in four important ways: (i) we add six verification basins in the eastern Unites States, giving us 23 basins that span much more of the continent; (ii) we add suites of additional forecasts, allowing us to quantify the variations in the skill contributions as a function of both forecast start date and forecast lead; (iii) we show how the skill levels achieved can be related to the statistical character of initial total water storage and forecast period precipitation; and (iv) we extend the analysis in a ''synthetic truth'' study to the full area covered by the 48 conterminous United States (CONUS). The result is a series of maps showing where and when soil moisture and snow initialization can be expected to contribute to streamflow forecast skill.

2. Experimental design

a. Models

We employ four independent state-of-the-art macroscale land surface models (LSMs) for our numerical experiments: Variable Infiltration Capacity (VIC), Catchment, Noah, and Sacramento (Sac). VIC, Catchment, and Noah were developed specifically for use with atmospheric general circulation models, and their performances have been evaluated in multiple land intercomparison projects (e.g., Bowling et al. 2003; Nijssen et al. 2003; Boone et al. 2004; Dirmeyer et al. 2006). Sac is a lumped, conceptual model generally used for operational hydrology; for this study, it is run in distributed mode (Koren et al. 2003), and to treat cold-season processes, it is coupled to the Snow-17 model (Anderson 1973). The four models differ to varying degrees in their parameterizations schemes, vertical structures, geophysical parameters, and state variables. Details are provided in Table 1, which also includes relevant references. In the present study, we used model-specific default conditions for all parameters. The model versions are essentially the same as those analyzed for other purposes by Wang et al. (2009).

Note that an earlier version of the VIC model was calibrated at a finer spatial resolution (1/8°) to a set of river basins having some overlap with those examined here (Maurer et al. 2002). This should have little impact on its performance in this study; when this model is removed from the analysis and only the remaining three models are used, the results are essentially the same. The other three models were not subject to any direct or indirect calibration. The behaviors of the four models were compared extensively in a recent analysis of drought simulation (Wang et al. 2009).

b. Numerical experiments

A set of four experiments was performed with each LSM. To some extent the design of the experiment mimics that of Bierkens and van Beek (2009), who examined streamflow forecast skill across Europe in the context of land model initialization and seasonal weather forecasts based on the North Atlantic Oscillation. The details and the goal here, however, are different: we aim to quantify, for a different climatic regime (the continental United States), the relative impacts of snow and soil moisture initialization on streamflow forecast skill for a number of different start dates and leads.

All experiments involved the integration of each LSM on a 0.5° × 0.5° array of grid cells encompassing

TABLE 1. The LSMs used in the study with brief descriptions on model structures, snow–soil hydrology schemes, parameters, and relevant references.

Model	Soil hydrology scheme	Soil layers and depth	Snow layers and scheme	Soil/vegetation Parameters	References
VIC	Variable infiltration capacity curve for surface runoff, Arno model for base flow, and drainage given by gravity.	Three soil layers with depths specified differently cell by cell. Total depth ranges from 0.8 to 3 m.	Complete energy balance. 2-layer snowpack with separate canopy snowpack.	From North America Land Data Assimilation (NLDAS) (Mitchell et al. 2004).	Liang et al. (1994), Maurer et al. (2002), and Cherkauer et al. (2003)
Catchment	TOPMODEL-based soil hydrology scheme.	Three soil layers; total soil depth is from 1 to 3.02 m.	Three dynamic snow layers with complete energy balance, liquid retention, densification, and thermal insulation.	Soil depth from the State Soil Geographic Database (STATSGO); soil texture from Food and Agriculture Organization (FAO).	Koster et al. (2000), Ducharne et al. (2000), and Stieglitz et al. (2001)
Noah	Exponential distribution of infiltration capacity for runoff; base flow proportional to storage; drainage driven by gravity.	Four soil layers and fixed depth. Total soil depth is 2 m.	Energy balance. Two-layer snowpack. Snow aging and liquid water retention included.	From NLDAS.	Schaake et al. (1996), Chen et al. (1997), Koren et al. (1999), Ek et al. (2003), Mitchell et al. (2004), and Livneh et al. (2010)
Sac	Runoff from impervious and saturated soils; base flow and percolation between reservoirs based on current storage.	Five soil water storage reservoirs. Total storage capacity ranges from 20 to 600 mm.	Snow-17: one-layer snow model; uses air temperature to index energy exchange and accumulation and ablation processes.	Soil parameters from NLDAS, no vegetation.	Burnash et al. (1973), Anderson (1973), and Mitchell et al. (2004)

CONUS. The atmospheric forcing data used in the integrations include hourly 2-m air temperature, 2-m specific humidity, precipitation, shortwave and longwave radiation, wind speed, and surface pressure covering the 89 yr from 1915 through 2003. The data were extracted from the National Oceanographic and Atmospheric Administration (NOAA) Cooperative Observer (Co-op) station archive, gridded as described by Andreadis et al. (2005). Surface wind was taken from the National Center for Atmospheric Research–National Centers for Environmental Prediction (NCAR–NCEP) reanalysis (Kalnay et al. 1996). As in Koster et al. (2010), prior to 1950, average surface wind values computed for the post-1950 period were used.

The first of the four experiments, labeled CTRL (for ''control''), is essentially the full integration of the LSM with the 89 yr of meteorological forcing data. For spinup, each model was cycled 10 times through the year 1915 and then integrated through 2003. To further minimize spinup effects, we discarded data from 1915 through 1919; thus, we analyzed output for the 84-yr period 1920–2003. The CTRL simulations are identical to those analyzed for other purposes by Wang et al. (2009).

The next three experiments—labeled Exp1, Exp2, and Exp3—are designed to quantify the degree to which streamflow can be forecasted from different sets of initial conditions on a given forecast start date assuming no skill (beyond knowledge of the climatology) in the seasonal forecasting of meteorological forcings. Note that while some skill in forecasting meteorological forcing may be possible with seasonal climate forecast systems, we do not examine this contribution here. Exp1 quantifies the joint contribution of initial snow and soil moisture conditions to streamflow forecast skill, whereas Exp2 and Exp3 quantify, respectively, the isolated contributions of initial snowpack and initial soil moisture to the skill.

Exp1 consists of 12 sets of forecast simulations. The first set consists of 84 separate 1-yr forecast simulations (one for each year of the period 1920–2003), each forecast initialized on 1 January with the 1 January snowpack and soil moisture states (and other, presumably less important states such as soil temperature) produced by CTRL for the year in question. This mirrors the initialization approach used in many seasonal forecasting studies (e.g., Wood et al. 2005). Here, to represent a lack

of knowledge of meteorological forcing during the 1-yr forecast period, we drive the land surface models across CONUS with the geographically varying climatological seasonal cycle of diurnal forcings (at 1-h resolution) determined from the CTRL forcing files, using (for proper cross validation) the 83 yr not including the year of forecast to generate the climatology. The forcing climatology for the year 1932, for example, was computed by averaging the forcings over the periods 1920–31 and 1933–2003. As a result, any skill generated in the forecasted streamflows can be attributed to the initialization alone.

The remaining 11 sets of forecast simulations for Exp1 are analogous to the first but with the forecast initialization on 1 February for the second set, on 1 March for the third set, and so on through December. In this way we can examine the impact of the initial conditions on streamflow forecast skill as a function of both lead (up to a year) and forecast start date.

Exp2 is a repeat of Exp1, except that soil moisture values are initialized on the forecast start date with the geographically varying climatological distribution of soil moisture on that start date, as determined from averaging over the 83 yr (outside the year in question) of soil moisture fields from CTRL. Thus, in Exp2, soil moisture initialization cannot contribute to streamflow forecast skill; skill is derived from snow initialization alone. In direct analogy, Exp3 is also the same as Exp1, but with snow amounts initialized to the geographically varying climatological values of snow variables on the forecast start date as derived from CTRL, again using the 83 years outside the year in question. Thus, in Exp3, no skill is derived from snow initialization; skill is derived from soil moisture initialization alone.

Two notes about the experimental design are warranted here. First, our use of climatological seasonal cycles for such forcing fields as precipitation in Exp1, Exp2, and Exp3 has a potential drawback; rainfall in these experiments is effectively characterized as a drizzle without realistic intermittency. Koster et al. (2010) tested the impact of such drizzle in a supplemental repeat of their version of Exp1 using the Catchment LSM. In this supplemental experiment, the climatological precipitation for every 4-day period was forced to fall at night on the first day, with no precipitation falling during the remaining three days. The impact on the results was essentially negligible.

Second, the experiments ignore the possibility, however unlikely, that soil temperature initialization plays a significant role in streamflow forecasting. To examine this possibility, Koster et al. (2010) performed, for the 1 January forecast start date, an experiment in which both soil moisture and snow were initialized to climatology, so that only the initialization of the remaining variables (viz. soil temperatures) could contribute to skill. None of the skill levels produced in this experiment was significantly different from zero at the 95% confidence level, supporting the idea that the impact of soil temperature initialization can indeed be neglected.

c. *Observational data*

Again, results from running a small subset of the four experiments (the CTRL, Exp1, Exp2, and Exp3 simulations run for 7 months starting on 1 January) have already been documented (Koster et al. 2010) in an analysis of multidecadal March–July forecasts of naturalized streamflow (water management effects removed) in 17 western U.S. basins. The use of these naturalized streamflow observations continues in our analyses below, supplemented by observations in six additional basins covering much of the eastern United States. The locations of the stream gauges and of the upstream basin areas they represent [delineated using HYDRO1k data; Verdin and Verdin (1999)] are shown in Fig. 1. Table 2 provides a description of basin properties and the periods of available data.

d. *Construction of multimodel forecasts; skill metric*

We combine the streamflow simulations and forecasts produced independently by the four land surface models via a two-step process. First, for a given experiment (CTRL, Exp1, Exp2, or Exp3), we convert each model's streamflows into standard normal deviates, or Z scores; this is achieved by subtracting the model's mean streamflow in that experiment for the time of year in question and then dividing by the corresponding standard deviation. Second, for a given year, we average across the four Z scores (one for each model) to compute that year's ensemble mean streamflow. For a given forecast evaluation, the time series of multimodel forecasts (one averaged Z score for each year) is regressed against the corresponding time series of naturalized streamflow observations. Because skill is computed from the Z scores for a specific time of year, no artificial skill is derived from capturing the observed seasonal cycle of streamflow.

The Z scores are used here in part because the design of the forecast experiments strongly limits the variance of the simulated streamflows—by design, the forecasted streamflows do not reflect the variability of meteorological forcings during the forecast period. We do not consider this a limitation to our analysis, as we are interested in measuring the ability of the models to reproduce the time variability and relative magnitudes of year-to-year streamflow anomalies rather than the absolute magnitudes of the variations. We quantify this ability with the square of the correlation coefficient (r^2) between the 84 synthetic truth streamflows and the corresponding

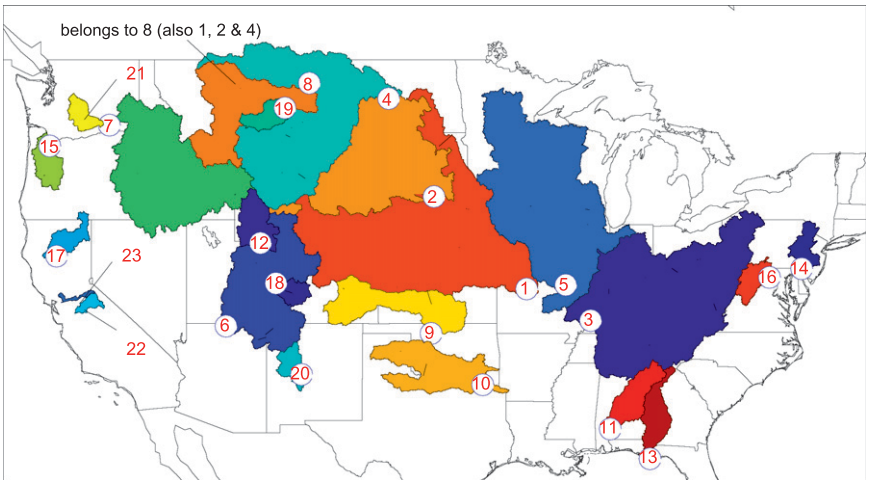

FIG. 1. Locations of basins examined in this study. The identifier numbers (see Table 2) are positioned at the stream gauge measurement sites.

predictions. For the r^2 calculation, the use of an averaged Z score, as described above, is fully appropriate. The r^2 metric of skill is as informative as a root-mean-square error (RMSE) metric, which is of more direct relevance to operations, given that known biases in the model's statistical moments allows the conversion of the first metric to the second (Entekhabi et al. 2010). Indeed, through application of the observed moments of streamflow in a considered basin, a time series of Z-score forecasts, as produced through these experiments, can easily be converted to a time series of forecasts with appropriate magnitude for the basin, with the extremes of interest to forecasters easily distinguished from non-extreme conditions. In any case, our aim is to examine the predictability of streamflow and not the performance of operational streamflow forecast systems, for which the treatment of bias is a particular concern (Shi et al. 2008).

TABLE 2. Characteristics of the basins examined in this study.

	River name	Station name	Basin area (km^2)	Latitude (°N)	Longitude (°W)	Observation period
1	Missouri	Hermann (includes basins 2, 4, 8, and 19)	1 353 275	38.71°	92.75°	1920–97
2	Missouri	Ft. Randall Dam (includes basins 4, 8, and 19)	682 465	43.07°	98.55°	1950–2009
3	Ohio	Metropolis	525 770	37.15°	88.74°	1928–2010
4	Missouri	Garrison Reservoir (includes basins 8 and 19)	469 826	47.39°	101.39°	1950–2003
5	Upper Mississippi	Grafton	443 660	38.90°	90.30°	1935–2010
6	Colorado	Lees Ferry (includes basins 12 and 18)	289 562	36.87°	111.58°	1920–2003
7	Snake	Ice Harbor Dam	281 015	46.25°	118.88°	1927–92
8	Missouri	Fort Peck Dam (includes basin 19)	149 070	48.04°	106.36°	1950–2009
9	Arkansas	Ralston	141 064	36.50°	98.73°	1940–2008
10	Arkansas–Red	Arthur City	115 335	33.88°	95.50°	1938–2001
11	Alabama	Clairborne	56 900	31.55°	87.51°	1950–93
12	Green	Greendale	50 116	40.91°	109.42°	1920–2003
13	Apalachicola	Sumatra	49 728	29.95°	−85.02°	1950–93
14	Delaware	Memorial Bridge	28 567	39.69°	75.52°	1948–87
15	Willamette	Above falls near Oregon City	25 900	45.34°	122.62°	1930–89
16	Potomac	Point of Rocks	25 000	39.27°	77.54°	1950–96
17	Sacramento	Bend Bridge	23 051	40.29°	122.19°	1920–2003
18	Gunnison	Near Grand Junction	20 533	38.98°	108.45°	1920–2003
19	Musselshel	Moseby	20 321	46.99°	107.89°	1941–2003
20	Rio Puerco	Bernardo	19 036	34.41°	106.85°	1940–2003
21	Yakima	Near Parker	9479	46.50°	120.44°	1925–2003
22	Tuolumne	La Grange Dam	4337	37.67°	120.44°	1920–2003
23	San Joaquin	Mokelunme Hill	1863	38.31°	120.72°	1920–2003

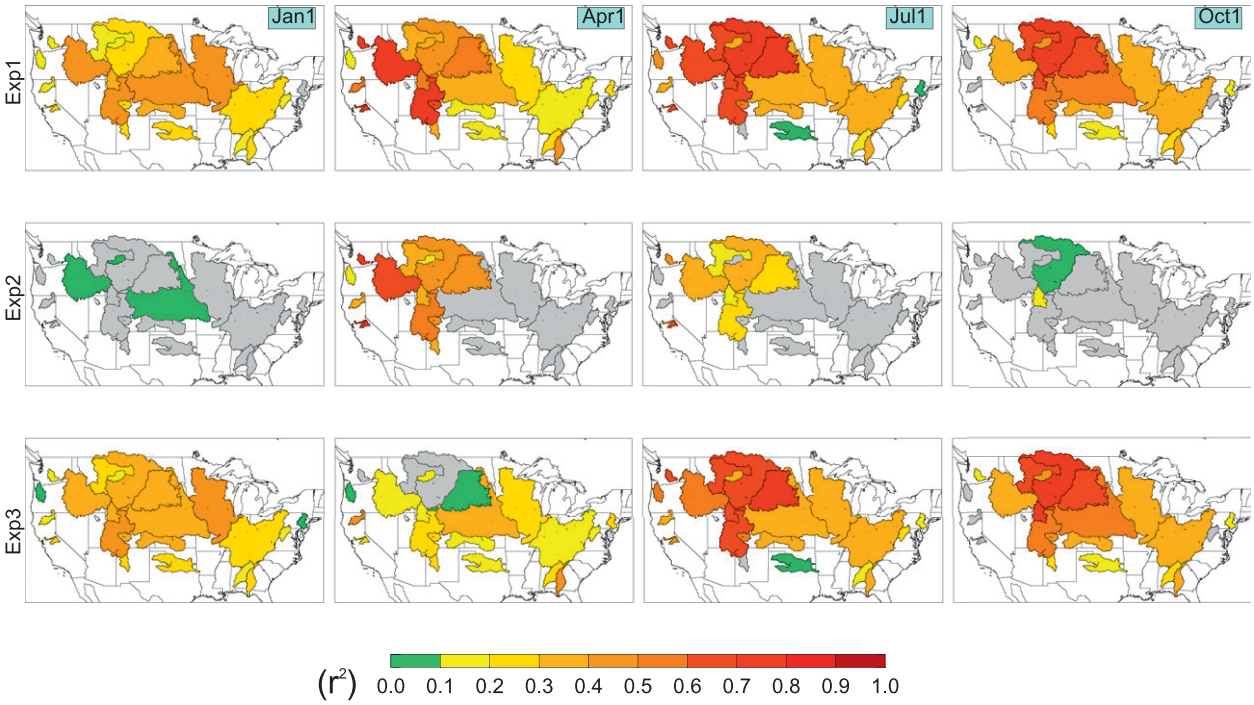

FIG. 2. Skill (measured as r^2 against naturalized observations) of multimodel ensemble 3-month streamflow forecasts at 0-month lead, for four start dates (columns) and the three experiments (rows)—snow and soil moisture initialization (Exp1), snow initialization alone (Exp2), and soil moisture initialization alone (Exp3). Gray shading indicates that skill levels are not significant at the 95% level.

Note that our cross-validation procedure produced an occasional negative correlation between observations and model results, particularly in basins with little or no snow. Any negative correlations were zeroed before computing r^2.

3. Results

a. Skill as a function of start date and lead

For the 17 westernmost basins in Fig. 1, Koster et al. (2010) demonstrated that the models show significant skill in forecasting observed (naturalized) March–July streamflow based on 1 January initial conditions. Both snow and soil moisture initialization contributed independently to skill to different degrees in different basins; the contribution of snow initialization was generally found to dominate, particularly in the more mountainous basins in the northwestern part of the study area, but the contribution of soil moisture initialization was still significant in many basins, and it even dominated in some of those toward the south-central United States.

In this section, through the experimental approach described above, we extend this analysis to 23 basins and to a more comprehensive collection of forecast start dates and lead times. Some key results are summarized in Fig. 2, which presents, for forecasts of 3-month-average streamflow at 0-month lead, the skill levels achieved by the multimodel system for four different start dates: 1 January, 1 April, 1 July, and 1 October. The salient feature of this plot, relative to that in Koster et al. (2010), is the increased relative importance of soil moisture initialization when other seasons are considered. The strong contribution of snow initialization to springtime forecast skill is consistent with that found in the earlier study, although the six newly considered eastern basins show a relatively significant impact of soil moisture initialization. Figure 2, though, shows that in summer, fall, and winter, initializing soil moisture has a dominant impact on skill across the United States, with contributions sometimes exceeding $r^2 = 0.6$ or 0.7. Overall, the summer, fall, and winter skill scores for Exp1 (both soil moisture and snow initialized) rival or exceed those of spring.

Naturally, such a statement must be tempered by knowledge of the seasonal cycle of streamflow—if skill is higher in seasons for which streamflow is relatively low, the usefulness of this skill for at least some applications is accordingly diminished. The top panels of Fig. 3 show the fraction of the observed annual streamflow that occurs in each season. The western basins in particular are dominated by snowmelt runoff, and thus streamflows there are largest during April–June (AMJ). Even so, a

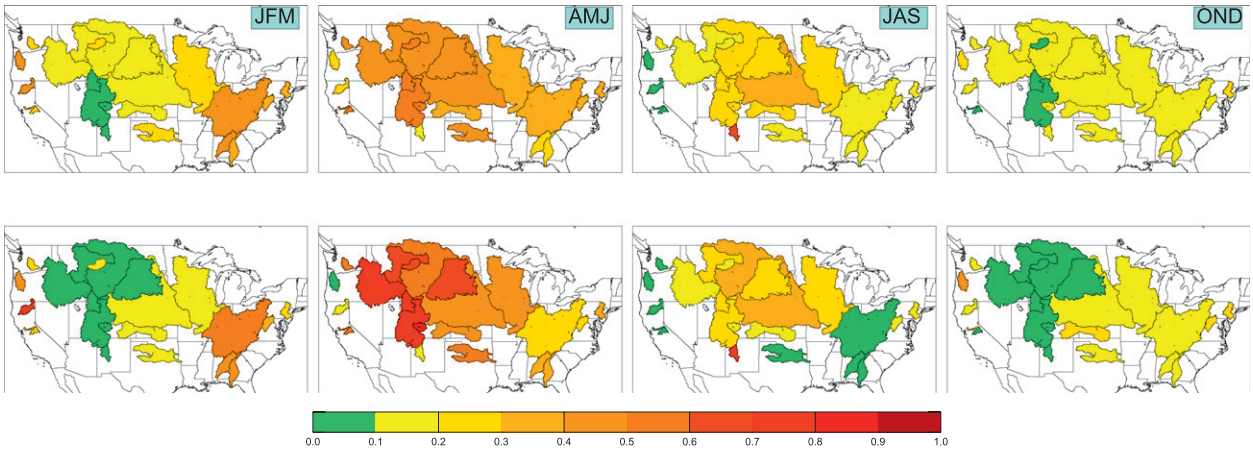

FIG. 3. (top) Fraction of total annual runoff flowing past the stream gauges during each of four 3-month periods, from observations. (bottom) The VF diagnostic for each of the four 3-month periods (see text for details).

large fraction of the streamflow occurs during the other three seasons, suggesting that the dominant impact of soil moisture initialization in these seasons is indeed relevant.

Note, however, that an even stronger consideration is the seasonal cycle of streamflow variance, σ_Q^2. Consider the extreme hypothetical example of a region with high summer streamflow that is almost exactly the same every year and moderate winter streamflow that varies strongly from year to year. Given the very small size of the interannual summer anomalies, a high r^2 skill score is clearly more important for the winter streamflow forecasts than for the summer forecasts. With this in mind, the bottom four panels of Fig. 3 show a diagnostic that we loosely term the ''variance fraction'' (VF), which is a function of season, m:

$$\mathrm{VF}(m) = \frac{\sigma_Q^2(m)}{\sum_{n=1,4} \sigma_Q^2(n)}. \quad (1)$$

This diagnostic is not meant to represent the contribution of each season to the annual streamflow variance, since the annual variance is affected by potential correlations between streamflow in adjacent seasons, which are ignored here. Rather, VF is a simple, ad hoc, first-order representation of how streamflow variance varies with season, made dimensionless so that the value of the diagnostic in different basins can be directly compared. Notice that the seasonal variation in the bottom plots is similar to that in the top plots; in the context of variance, spring (AMJ) is the most important time to predict streamflow accurately. Still, significant variance does exist in other seasons, particularly summer, further supporting the idea that initializing soil moisture accurately in those seasons is important.

The experimental framework allows a look at the skill obtained at longer leads. Figure 4 shows, for Exp1 (both soil moisture and snow initialized), the forecast skill levels obtained in each of the 23 basins for 3-month-average streamflow as a function of lead and start date. A given colored curve corresponds to a 1-yr forecast, with the leftmost point on the curve indicating the skill levels for 0-month lead (i.e., for the first three months of the forecast), the next point on the curve corresponding to the 1-month lead (i.e., for months 2–4 of the forecast), and so on; a point farther to the right on a given curve thus corresponds to a longer forecast lead time, unless of course the curve has wrapped around from the right edge to the left edge of the plot, at the December–February (DJF)–January–March (JFM) boundary.

As should be expected, streamflow forecast skill tends to decrease with increasing lead. (See Pagano et al. (2009) for an operational example of the skill–lead relationship.) The basins clearly show some variability in the lead times allowing skillful forecasts, with some basins having little skill beyond a 0-month lead, regardless of start date (e.g., the Willamette River upstream of Oregon City), and others showing skill for leads exceeding six months for most start dates (e.g., the Green River upstream of Greendale). A number of basins show skillful spring or summer forecasts at multimonth leads, with little or no skill in the fall and winter (e.g., the San Joaquin River upstream of Mokelumne Hill)—a reflection of the dominant contribution of snowmelt to the annual runoff totals in these basins, as indicated by corresponding plots for Exp2 and Exp3 (not shown). Forecasts for start dates early in the year (January–March) in mountainous areas are sometimes characterized by skill that increases with lead and then decreases—a reflection of the fact that skill for these basins does not manifest itself until the snowmelt season begins.

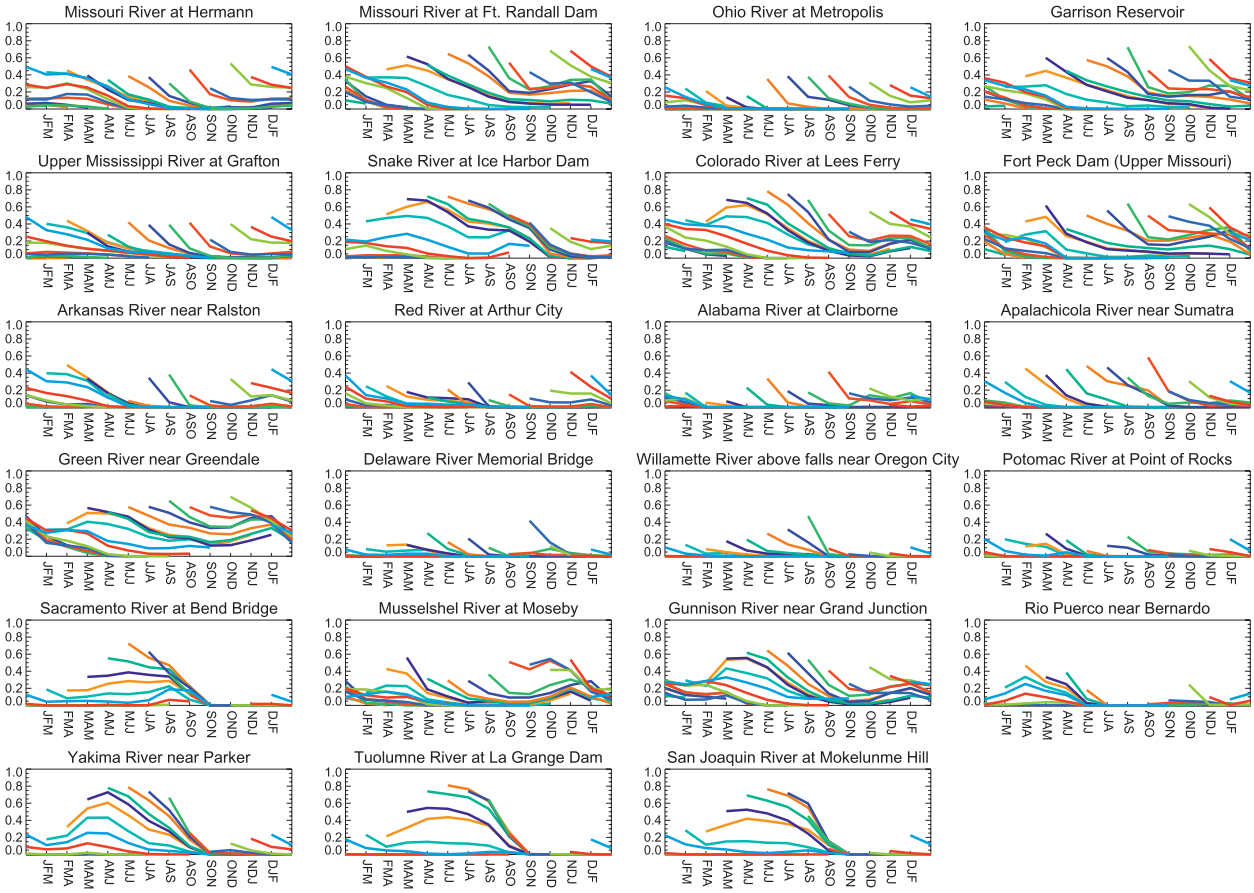

FIG. 4. Skill (r^2) of multimodel ensemble 3-month forecast, as a function of start date and lead, for Exp1. Curves of a single color provide results for forecasts of a specific start date. The leftmost point on a curve indicates a 0-month lead for the forecast; points farther to the right on the curve (except for the wrap around at the end of the year) indicate a longer lead time. (See text for details.)

Figure 5 presents this information in a different way; it shows, for each experiment, the number of months into the future for which a forecast of 3-month-average streamflow still has statistically significant skill at the 95% confidence level. (Note that the r^2 value corresponding to this confidence level varies with basin according to the length of its observational record. Note also that the skill levels at longer leads, while determined significant, may still be very small, so that the practical value of this particular diagnostic may be limited.) Soil moisture initialization by itself (Exp3) is, for many basins, most effective in the fall (1 October)—a time when snow initialization (Exp2) has no impact at all. In general, snow initialization is most effective during winter and spring. We explore in the next section some important controls on the overall contribution of initialization to skill and the reasons underlying the geographical and temporal variations of skill. For now, though, we note that certain aspects of the soil–snow distinction in Fig. 5 make intuitive sense. Snow is essentially absent on 1 October, necessitating a low impact of snow initialization for 1 October starts. Soil moisture, on the other hand, does exist (and does vary interannually) on 1 October. At least in relatively cold regions, this soil moisture should remain largely unchanged during winter months while evaporation is low and the ground is, in many cases, covered with snow; the impact of the 1 October soil moisture should thus at least partially manifest itself during the snowmelt season, several months into the future.

b. *Skill as a function of variability in water storage and precipitation*

The spatial and temporal variations in the skill levels illustrated in Fig. 2 can be related, at least conceptually, to two basic water supply quantities: the total water (W_{init}) stored in surface reservoirs (snow plus soil moisture) on the forecast start date and the total water precipitating during the forecast period (P_{fcst}). Simply put, under the assumption that W_{init} and P_{fcst} are the two primary drivers of streamflow variability during the forecast period, the relative degrees to which the two quantities vary from year to year, along with our ability to

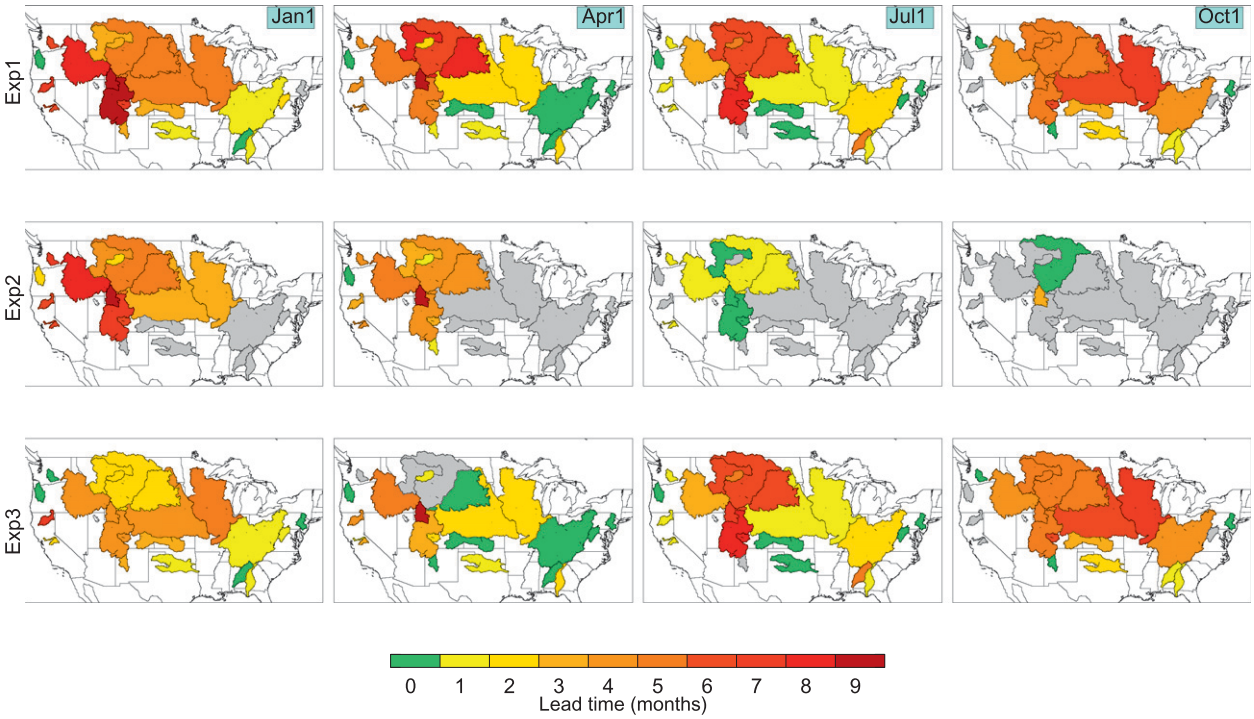

FIG. 5. Number of lead months for which a forecast still has significant skill at the 95% confidence level for (top to bottom) the three experiments and (left to right) Jan, Apr, Jul, and Oct. Gray shading indicates no statistically significant contribution to skill even at a 0-month lead.

estimate each on the forecast start date, determines the degree to which variations in streamflow can be accurately predicted. For example, if the interannual variance of W_{init} (σ_w^2) is large and that of P_{fcst} (σ_P^2) is small, then streamflow forecast skill is mostly predicated on our ability to estimate W_{init}; on the other hand, if σ_P^2 is large and σ_w^2 is small, the skill mostly depends on an accurate prediction of P_{fcst}.

Under the further assumption that W_{init} can be reasonably well estimated on the forecast start date (e.g., through the use of land modeling, as above, or through a data assimilation framework) whereas P_{fcst} is comparatively unknowable given chaotic atmospheric dynamics during the forecast period, a relatively high initial water storage variance (σ_w^2) implies increased streamflow predictability, whereas a relatively high forecast period precipitation variance (σ_P^2) implies a decreased streamflow predictability. More to the point, the stated assumptions imply that forecast skill in the forecast framework presented above should increase with the dimensionless ratio κ, defined here as

$$\kappa = \sigma_w/\sigma_p. \tag{2}$$

Simply put, a higher value of κ corresponds to a higher level of knowledge, on the forecast start date, of the basic underlying controls on streamflow predictability. In essence, κ is the dimensionless ratio of the variabilities of the ''known'' and ''unknown'' water volumes that determine streamflow.

In Fig. 6, the skill achieved for a forecast of 3-month-average streamflow at 0-month lead is plotted against the corresponding value of κ. Each dot corresponds to a specific basin and start date; given 23 different basins and 12 different start dates, there are 276 dots in the plot. The calculation of the κ values required estimates for both σ_w and σ_P. The σ_P values for a given start date were computed directly from the yearly time series of 3-month precipitation totals following that start date in the meteorological forcing data. The corresponding ''multimodel'' value of σ_w was computed from a time series of simulated multimodel total (snow plus soil moisture) water content for the start date in question; this time series was generated by averaging across the models the raw (unstandardized) instantaneous soil moisture and snow water contents generated by each in the CTRL simulation.

The scatterplot shows, to first order, an increase in streamflow forecast skill with κ. This is particularly evident at the extremes; in general, for κ values below about 0.3, the skill achieved is very low ($r^2 < 0.1$), whereas for κ values above about 3, the skill levels are high ($r^2 > 0.7$). The first-order relationship across the full range of κ is

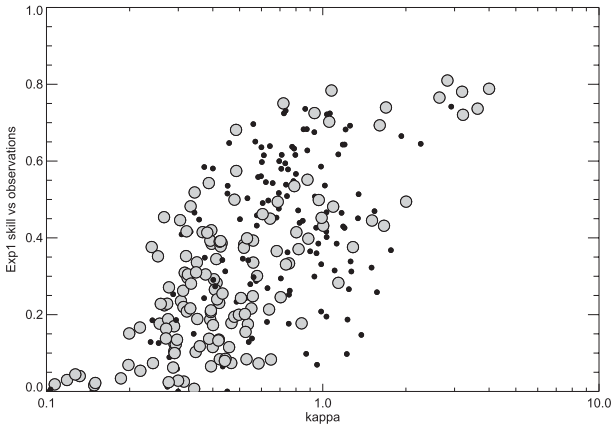

FIG. 6. Skill (r^2) of 3-month streamflow forecast at 0-month lead as a function of κ—the ratio of the standard deviation of initial total water storage (soil water plus snow) to the standard deviation of precipitation during the 3-month forecast period. Each dot represents the skill determined for one of 12 start dates and one of 23 measured basins, for a total of 276 plotted points. The larger dots correspond to locations and start times for which the CTRL simulation (which utilizes observed meteorology during the forecast period) was able to capture the observed streamflow variations with an r^2 of at least 0.75.

made somewhat more evident by considering in isolation the subset of the dots (the larger dots) corresponding to times and locations for which the CTRL streamflows were especially realistic (i.e., $r^2 > 0.75$). A focus on such a subset is sensible because forecast skill for the remaining dots is additionally limited by flaws in the modeling systems, in the forcing data, and/or in the streamflow validation data—for these remaining dots, when the models are run outside of forecast mode, with full knowledge of precipitation during the forecast period, they still have some trouble matching the observed streamflows.

c. *Extension of variability analysis to CONUS*

The intuitive relationship between κ and streamflow forecast skill levels, demonstrated to first order in Fig. 6, has a broader significance. Given our ability to estimate κ outside of the 23 basins through analysis of observational data and land model integrations, we can estimate for these outside locations the streamflow forecast skill that could potentially be achieved with existing systems, even given the absence of suitable verification data.

Such an estimation is presented in Fig. 7, which shows, for each of the four seasons and at each $0.5° \times 0.5°$ grid cell across CONUS, the values of σ_W (calculated following the approach in section 3b above), σ_P (computed from the gridded meteorological forcing data), and the resulting value of κ. The maps show a wide seasonal variation in κ. For much of the United States, κ (and thus expected streamflow forecast skill at 0-month lead) is largest for JFM, with high values across the northern half of the country and south into New Mexico and Texas. For the north-central United States in particular, the high κ values during JFM reflect a seasonal reduction of σ_P rather than a seasonal increase in σ_W. For AMJ, κ values are quite high in the western third of the country and very small elsewhere; this season indeed features particularly low σ_P values and high σ_W values in the southwest and particularly small σ_W values in the east. For July–September (JAS), moderately high κ values are seen mostly in the northwestern quadrant of the country and in the far west, reflecting relatively low rainfall variability in these areas. The October–December (OND) κ distribution looks similar to that for JFM but is weaker, with minimal values (and thus no expected skill) in the west, where σ_W is now small. Notice that most of the southern United States features low values of κ throughout the year, which is a reflection of the consistently high σ_P values there; the corresponding σ_W values there, while also high, are presumably limited by the finite capacities of the soil layers, which provide for efficient runoff as the soil gets very wet.

Again, according to the relationship illustrated in Fig. 6, these patterns can be interpreted in terms of potential streamflow forecast skill. The relationship in Fig. 6, however, while intuitive, is admittedly noisy. We thus supplement the κ calculations here with skill calculations using a synthetic truth dataset. While the computation of true skill levels outside of the basins in Fig. 1 is precluded by a lack of naturalized streamflow observations of sufficient duration, we can utilize model-generated streamflow data (viz. those produced through the CTRL simulations) as verification data, as long as the basic limitation of such synthetic truth is always kept in mind (i.e., the fact that it is based on imperfect land models integrated with imperfect meteorological forcing).

The left column of Fig. 8 is a repeat of the right column of Fig. 7; it shows the computed distribution of κ for each season. The middle column of Fig. 8 shows the corresponding distributions of 3-month runoff forecast skill at 0-month lead for Exp1, using the CTRL results as the synthetic truth. Notice that the plots in these two columns are very similar, particularly for JFM, AMJ, and JAS. In other words, the κ-based estimates of where and when forecast skill might be achieved through land model initialization appear to be well supported by these alternative calculations.

For perspective, the rightmost column of Fig. 8 shows VF, as defined in (1), at each grid cell. A comparison of this column with the others quickly shows that the locations for which skill from initialization is especially attainable (i.e., where values of κ are especially high) tend to correspond to a low value of VF—when runoff is very

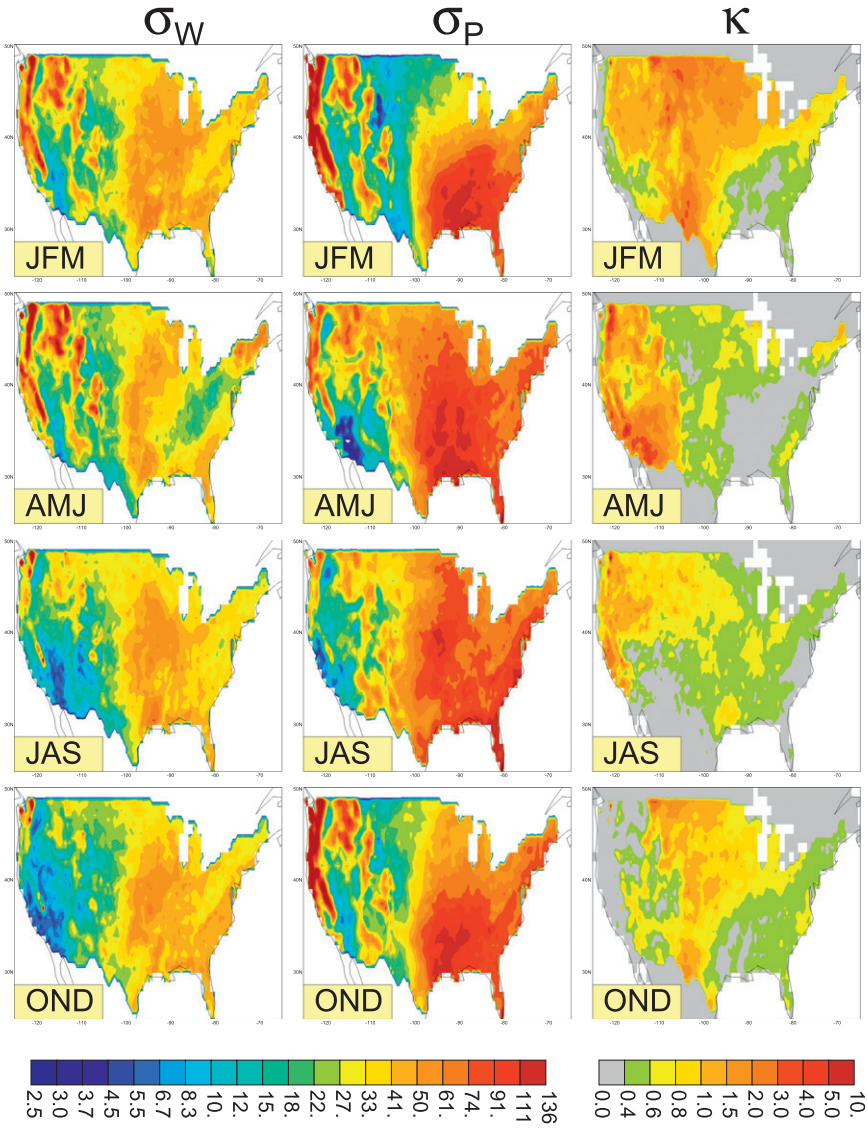

FIG. 7. Distributions, as a function of (top to bottom) season, of (left) σ_W (mm), (middle) σ_P (mm), and (right) the resulting diagnostic κ (dimensionless).

predictable, the variance of the runoff tends to be relatively low. Even so, the fields show substantial overlap between times of significant skill and sizeable runoff variance.

This latter point is demonstrated further in Fig. 9, which shows a rough "annual" estimate of 3-month forecast skill at 0-month lead for Exp1, computed here as a weighted average of the r^2 values over the four seasons:

$$r^2_{\text{ann}} = \sum_{n=1,4} r^2(n)\text{VF}(n), \tag{3}$$

where VF is as defined in (1). The skill levels in Fig. 9 suggest that substantial skill is indeed attained during times of significant runoff variance; if it were not, the values plotted here would be close to zero. The weighted skill levels are particularly large in the west (largely a reflection of snow impacts, as suggested by Figs. 2 and 3) and in the upper Great Plains, toward the Great Lakes.

4. Summary and discussion

The multimodel ensemble forecasts of 3-month-average streamflow at 0-month lead are shown to have significant skill across the 23 basins for which we have naturalized streamflow observations (Fig. 2). Snow initialization has a positive impact mainly in spring but also in summer, whereas soil moisture initialization contributes to skill in all seasons, with the largest contributions in summer and

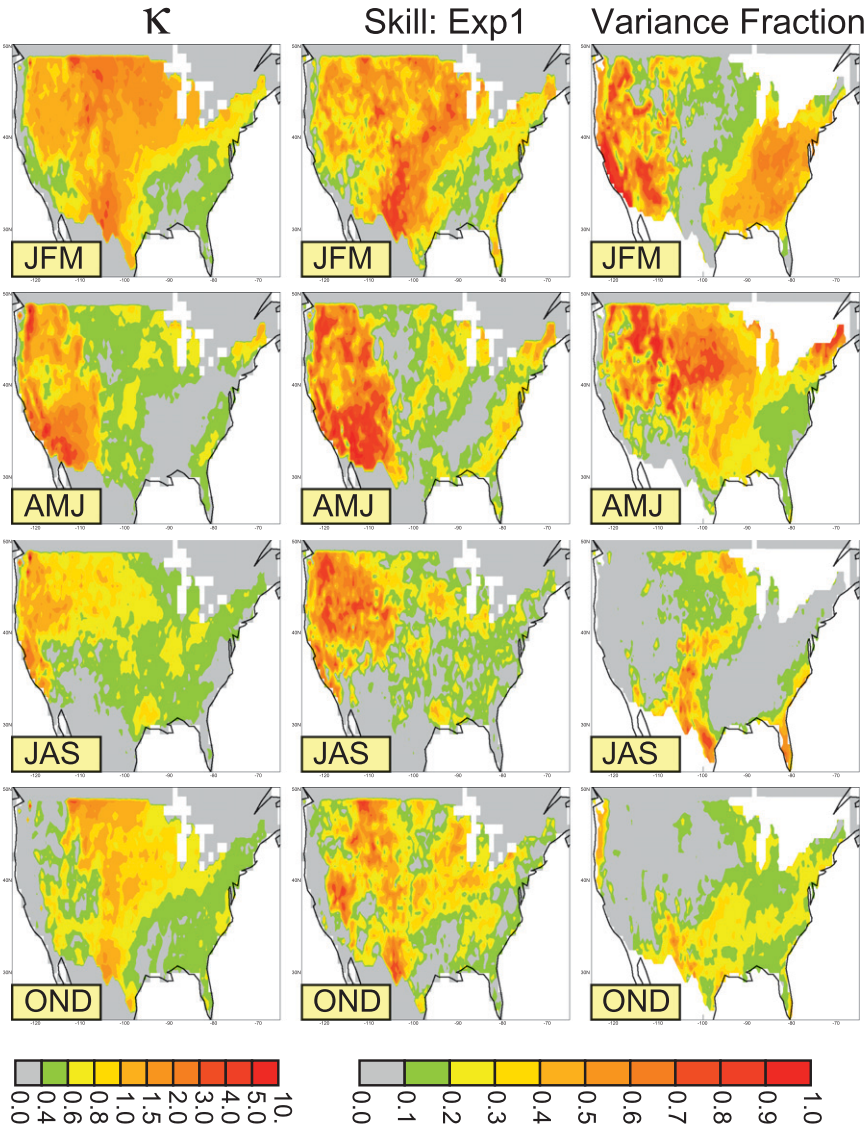

FIG. 8. Comparison between (left) the parameter κ and (middle) the skill (r^2) obtained against synthetic truth in Exp1 for 3-month streamflow forecasts at 0-month lead for (top to bottom) seasons. (right) The VF diagnostic: a rough measure, for a given grid cell, of the relative strength of the variance in that season (see text for details).

fall. In general, outside of spring, the impact of soil moisture initialization dominates over that of snow initialization. 1 October soil moisture initialization contributes, in general, to skill at particularly long leads, presumably because of the reduced activity of soil moisture during the cold season.

As noted in Koster et al. (2010), the skill levels found for Exp1 are similar to those shown in another study for various European stations (Bierkens and van Beek 2009). They are also similar, if sometimes slightly smaller, than those found in two recent studies focusing on the American west (Pagano et al. 2004, 2009). The rough agreement with the latter two studies is interesting because the two Pagano et al. studies examined calibrated forecast systems on (generally) much smaller basins having the benefit, for example, of local snow-course measurements. The models used in our study do not utilize local in situ data or rely on calibration to observed anomalies; they rely instead solely on the integration of antecedent meteorological data at the large scale. This study thus provides evidence that such large-scale, uncalibrated models may prove useful for basin-scale prediction, perhaps in conjunction with existing, proven operational approaches that rely on calibrated, statistics-based models.

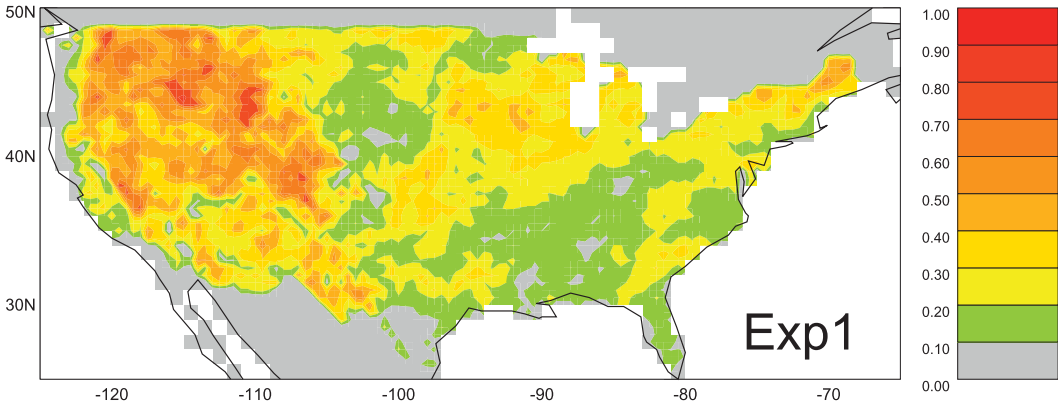

FIG. 9. Estimate of annually averaged skill for Exp1 against synthetic truth (CTRL), computed by weighting the seasonal r^2 values by seasonal runoff variance [Eq. (3) in text].

It is worth mentioning that our experimental design is limited in some ways and that overcoming or mitigating its limitations could boost the skill levels achieved. For example, we neglect here the time scales associated with river routing. To produce the model-generated basin streamflows, the contemporaneous runoffs generated within the different grid cells of a basin are simply added together; no account is taken of the length of time needed to transport a grid cell's river water to a stream gauge site. This limitation is addressed only partially here by the consideration of long-term average (3 month) streamflow totals. Another issue is the necessarily approximate definition of the basin boundaries, given that the models are run on a 0.5° grid. Perhaps the two most important limitations to this study, however, involve the uncertainty in the data used to force and validate the models, particularly early in the study period, and the errors associated with the model parameterizations themselves—land models have oversimplified representations of many critical surface processes. Presumably, improved models driven with more accurate forcing data and validated against more accurate streamflow observations would show greater skill.

Land data assimilation (e.g., Reichle et al. 2007) shows particular promise for providing improved estimates of soil moisture states through the mathematically optimal incorporation of satellite retrievals or radiances into the land modeling environment. Streamflow prediction may thus benefit from data assimilation efforts focused on current and upcoming satellite soil moisture missions, particularly Soil Moisture Ocean Salinity (SMOS; launched in November 2009; http://www.esa.int/esaLP/LPsmos.html) and Soil Moisture Active Passive (SMAP; projected launch in 2014; http://smap.jpl.nasa.gov).

The time and space variability of the skill levels achieved in our experiments is addressed in sections 3b and 3c using a simple framework that relates skill to both the variability of water storage on the forecast start date and the variability of precipitation during the forecast period. The parameter κ captures the joint effects of these two controls; a higher κ value implies that the control quantity that is measurable on the forecast start date (water storage) has an increased impact on skill relative to the immeasurable control quantity (forecast period precipitation), so that a higher κ value implies higher skill. Our skill levels in the 23 examined basins are indeed shown to increase, to first order, with an increase in κ (Fig. 6). Furthermore, the spatial distributions of κ across CONUS agree, again to first order, with the spatial distributions of skill obtained with a synthetic truth dataset (Fig. 8), providing support for the idea that the synthetic skill levels are indicative of what could be achieved across CONUS with an extended streamflow forecast system.

The κ framework is necessarily limited. It does not capture, for example, the simple fact that evaporation is higher during summer than winter, which implies a relatively lower soil moisture memory for summer and thus a reduced potential for summer soil moisture to affect streamflow. The standard deviations used in the κ calculation here assume a stationarity that may not exist in nature. The assumption that precipitation during the forecast period cannot be estimated is not precisely true; some predictive skill for rainfall may be attainable with seasonal forecast systems. The κ framework's neglect of these issues is arguably a weakness; we offer, however, the contrasting view that the framework's simplicity is in fact a strength. The framework is able to capture, in a single parameter, the joint effects of what are probably the two main controls on streamflow prediction skill, allowing a first-order understanding of where and when this skill may be achieved.

Acknowledgments. We thank James Verdin for discussions at the onset of this study. We also thank the

providers of the naturalized western streamflow data: U.S. Army Corps of Engineers, Omaha NB office, for the Missouri River basin; U.S. Army Corps of Engineers, Tulsa OK office, for the Arkansas–Red River basin; Columbia River Basin Climate Change Scenarios Database for the Columbia River basin; California Data Exchange Commission for the California basins; and U.S. Bureau of Reclamation, Lower Colorado Region, for the Colorado River basin. We further thank Professor Edwin Maurer (Santa Clara University) for his assistance in obtaining naturalized streamflow data for the eastern and central U.S. river basins. Funding for much of this study was provided by NASA's Terrestrial Hydrology Program.

REFERENCES

Anderson, E. A., 1973: National Weather Service River Forecast System: Snow accumulation and ablation model. NOAA Tech. Memo. NWS Hydro-17, 217 pp.

Andreadis, K. M., E. A. Clark, A. W. Wood, A. F. Hamlet, and D. P. Lettenmaier, 2005: Twentieth-century drought in the conterminous United States. *J. Hydrometeor.,* **6,** 985–1001.

Berg, A. A., and K. A. Mulroy, 2006: Streamflow predictability in the Saskatchewan/Nelson River basin given macroscale estimates of the initial soil moisture status. *Hydrol. Sci.,* **51,** 642–654.

Bierkens, M. F. P., and L. P. H. van Beek, 2009: Seasonal predictability of European discharge: NAO and hydrological response time. *J. Hydrometeor.,* **10,** 953–968.

Boone, A., and Coauthors, 2004: The Rhône-Aggregation Land Surface Scheme Intercomparison Project: An overview. *J. Climate,* **17,** 187–208.

Bowling, L. C., and Coauthors, 2003: Simulation of high latitude hydrological processes in the Torne–Kalix basin: PILPS Phase 2e: 1: Experiment description and summary intercomparison. *Global Planet. Change,* **38,** 1–30.

Brumbelow, K., and A. Georgakakos, 2001: Agricultural planning and irrigation management: The need for decision support. *The Climate Report,* Vol. 1, Jones Day Publications, 2–6.

Burnash, R. J. C., R. L. Ferral, and R. A. McGuire, 1973: A generalized streamflow simulation system: Conceptual modeling for digital computers. National Weather Service and California Department of Water Resources Joint Tech. Rep., 204 pp.

Chen, F., Z. Janjic, and K. Mitchell, 1997: Impact of atmospheric surface layer parameterizations in the new land-surface scheme of the NCEP mesoscale Eta model. *Bound.-Layer Meteor.,* **85,** 391–421.

Cherkauer, K. A., L. C. Bowling, and D. P. Lettenmaier, 2003: Variable infiltration capacity cold land process model updates. *Global Planet. Change,* **38** (1–2), 151–159.

Christensen, N., and D. P. Lettenmaier, 2006: A multimodel ensemble approach to assessment of climate change impacts on the hydrology and water resources of the Colorado River basin. *Hydrol. Earth Syst. Sci. Discuss.,* **3,** 3727–3770.

Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki, and N. Hanasaki, 2006: GSWP-2: Multimodel analysis and implications for our perception of the land surface. *Bull. Amer. Meteor. Soc.,* **87,** 1387–1397.

Ducharne, A., R. D. Koster, M. J. Suarez, A. M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general circulation model 2. Parameter estimation and model demonstration. *J. Geophys. Res.,* **105** (D20), 24 823–24 838.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.,* **108,** 8851, doi:10.1029/2002JD003296.

Entekhabi, D., R. H. Reichle, R. D. Koster, and W. T. Crow, 2010: Performance metrics for soil moisture retrievals and application requirements. *J. Hydrometeor.,* **11,** 832–840.

Hamlet, A. F., D. Huppert, and D. P. Lettenmaier, 2002: Economic value of long-lead streamflow forecasts for Columbia River hydropower. *J. Water Resour. Plann. Manage.,* **128,** 91–101.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Koren, V., J. Schaake, K. Mitchell, Q. Duan, F. Chen, and J. Baker, 1999: A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. *J. Geophys. Res.,* **104,** 19 569–19 585.

——, M. Smith, and Q. Duan, 2003: Use of *a priori* parameter estimates in the derivation of spatially consistent parameter sets of rainfall-runoff models. *Calibration of Watershed Models,* Q. Duan et al., Eds., Water Science and Applications, Vol. 6, Amer. Geophys. Union, 239–254.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general circulation model 1. Model structure. *J. Geophys. Res.,* **105** (D20), 24 809–24 822.

——, S. P. P. Mahanama, B. Liveneh, D. P. Lettenmaier, and R. H. Reichle, 2010: Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow. *Nat. Geosci.,* **3,** 613–616.

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

Liang, X., 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. *J. Geophys. Res.,* **99,** 14 415–14 428.

Livneh, B., Y. Xia, K. E. Mitchell, M. B. Ek, and D. P. Lettenmaier, 2010: Noah LSM snow model diagnostics and enhancements. *J. Hydrometeor.,* **11,** 721–738.

Mahanama, S. P. P., R. D. Koster, R. H. Reichle, and L. Zubair, 2008: The role of soil moisture initialization in subseasonal and seasonal streamflow predictability: A case study in Sri Lanka. *Adv. Water Resour.,* **31,** 1333–1343, doi:10.1016/j.advwatres.2008.06.004.

Maurer, E. P., and D. P. Lettenmaier, 2003: Predictability of seasonal runoff in the Mississippi River basin. *J. Geophys. Res.,* **108,** 8607, doi:10.1029/2002JD002555.

——, 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.

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

Nijssen, B., and Coauthors, 2003: Simulation of high latitude hydrological processes in the Torne–Kalix basin: PILPS Phase 2(e): 2: Comparison of model results with observations. *Global Planet. Change,* **38,** 31–53.

Pagano, T., and D. Garen, 2005: A recent increase in western U.S. streamflow variability and persistence. *J. Hydrometeor.,* **6,** 173–179.

——, ——, and S. Sorooshian, 2004: Evaluation of official western U.S. seasonal water supply outlooks, 1922–2002. *J. Hydrometeor.,* **5,** 896–909.

——, ——, T. R. Perkins, and P. A. Pasteris, 2009: Daily updating of operational statistical seasonal water supply forecasts for the western U.S. *J. Amer. Water Resour. Assoc.,* **45,** 767–778.

Reichle, R. H., R. D. Koster, P. Liu, S. P. P. Mahanama, E. G. Njoku, and M. Owe, 2007: Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *J. Geophys. Res.,* **112,** D09108, doi:10.1029/2006JD008033.

Schaake, J. C., V. I. Koren, Q.-Y. Duan, K. E. Mitchell, and F. Chen, 1996: Simple water balance model for estimating runoff at different spatial and temporal scales. *J. Geophys. Res.,* **101,** 7461–7475.

Shi, X., A. W. Wood, and D. P. Lettenmaier, 2008: How essential is hydrologic model calibration to seasonal streamflow forecasting? *J. Hydrometeor.,* **9,** 1350–1363.

Stieglitz, M., A. Ducharne, R. Koster, and M. Suarez, 2001: The impact of detailed snow physics on the simulation of snow cover and subsurface thermodynamics at continental scales. *J. Hydrometeor.,* **2,** 228–242.

Verdin, K. L., and J. P. Verdin, 1999: A topological system for delineation and codification of the Earth's river basins. *J. Hydrol.,* **218,** 1–12.

Wang, A., T. J. Bohn, S. P. Mahanama, R. D. Koster, and D. P. Lettenmaier, 2009: Multimodel ensemble reconstruction of drought over the continental United States. *J. Climate,* **22,** 2694–2722.

Wood, A. W., and D. P. Lettenmaier, 2008: An ensemble approach for attribution of hydrologic prediction uncertainty. *Geophys. Res. Lett.,* **35,** L14401, doi:10.1029/2008GL034648.

——, A. Kumar, and D. P. Lettenmaier, 2005: A retrospective assessment of National Centers for Environmental Prediction climate model–based ensemble hydrologic forecasting in the western United States. *J. Geophys. Res.,* **110,** D04105, doi:10.1029/2004JD004508.

Yao, H., and A. P. Georgakakos, 2001: Assessment of Folsom Lake response to historical and potential future climate scenarios: 2. Reservoir management. *J. Hydrol.,* **249,** 176–196.