SPECIAL
Advances in Drought Monitoring
COLLECTION

Soil Moisture Initialization Error and Subgrid Variability of Precipitation in Seasonal Streamflow Forecasting

RANDAL D. KOSTER

Global Modeling and Assimilation Office, NASA/GSFC, Greenbelt, Maryland

GREGORY K. WALKER AND SARITH P. P. MAHANAMA

SSAI, Lanham, Maryland

ROLF H. REICHLE

Global Modeling and Assimilation Office, NASA/GSFC, Greenbelt, Maryland

(Manuscript received 1 April 2013, in final form 26 July 2013)

ABSTRACT

Offline simulations over the conterminous United States (CONUS) with a land surface model are used to address two issues relevant to the forecasting of large-scale seasonal streamflow: (i) the extent to which errors in soil moisture initialization degrade streamflow forecasts, and (ii) the extent to which a realistic increase in the spatial resolution of forecasted precipitation would improve streamflow forecasts. The addition of error to a soil moisture initialization field is found to lead to a nearly proportional reduction in large-scale seasonal streamflow forecast skill. The linearity of the response allows the determination of a lower bound for the increase in streamflow forecast skill achievable through improved soil moisture estimation, for example, through the assimilation of satellite-based soil moisture measurements. An increase in the resolution of precipitation is found to have an impact on large-scale seasonal streamflow forecasts only when evaporation variance is significant relative to precipitation variance. This condition is met only in the western half of the CONUS domain. Taken together, the two studies demonstrate the utility of a continental-scale land surface–modeling system as a tool for addressing the science of hydrological prediction.

1. Introduction

Because of the importance of accurate streamflow forecasts for water resources planning (e.g., Yao and Georgakakos 2001; Hamlet et al. 2002), the development of approaches for producing useful streamflow forecasts and the evaluation of these approaches over time has a rich history, going back to at least the 1930s (Pagano et al. 2004). Operational streamflow forecasts generally rely on statistical techniques (e.g., Garen 1992). Using various quantities describing the current state of a system (e.g., snow amount, soil moisture, and climate indices), calibrated regressions are applied that transform these quantities into streamflow forecasts.

The historical use of these statistical techniques is arguably a reflection of historical limitations in our ability to model accurately the physical processes that generate streamflow—in particular our ability to provide the high-resolution forcing and boundary condition data needed to support the physical modeling. The advent of improved observational networks in recent decades, however, has supported the growth of the physical-modeling approach. A now common forecast strategy involves the use of spatially distributed land surface modeling: realistic snow and soil moisture fields are used to initialize the models, which are then integrated into the forecast period with atmospheric forcing, producing streamflow forecasts along the way (Day 1985). The atmospheric forcing can take the form of historical time series at the site in question (e.g., Wang et al. 2011), sometimes modified depending on the needs of the study (e.g., Hamlet and Lettenmaier 1999); alternatively, it can be derived from the forcing produced

Corresponding author address: R. D. Koster, Global Modeling and Assimilation Office, Code 610.1, NASA/GSFC, Greenbelt, MD 20771.
E-mail: randal.d.koster@nasa.gov

DOI: 10.1175/JHM-D-13-050.1

© 2014 American Meteorological Society

by a full numerical Earth system model running seasonal forecasts (e.g., Wood et al. 2002). The land surface models used are generally distributed, physically based representations of surface water and energy budget processes that take advantage of the information content of a wide variety of observations. Wood and Lettenmaier (2006) outline strong arguments for the expectation that the land-modeling strategy will, in time, eclipse the regression approach as the preferred means of providing streamflow forecasts.

Recent work has provided important insights into the science of predicting streamflow with the land-modeling strategy. For example, a number of studies have used such systems to examine the relative contributions of state initialization and forecasted meteorological forcing to streamflow forecast skill (e.g., Wood and Lettenmaier 2008; Mahanama et al. 2008, 2012; Bierkens and van Beek 2009; Li et al. 2009). Luo and Wood (2008) demonstrated the effectiveness of a Bayesian approach for producing high-resolution land model forcing for streamflow forecasts, an approach that combines information from multiple coarser-resolution meteorological forecasts and historical data. Yuan and Wood (2012) examined the skill levels achieved through bias correction in the atmospheric forcing prior to its application in the offline (land-only) system versus those achieved with bias-corrected streamflow products from the original seasonal forecast system.

As these examples demonstrate, a given offline-modeling system can serve as a powerful test bed for addressing the science underlying streamflow prediction. Given the potential societal benefits of accurate streamflow predictions, and given the fact that many aspects of the science underlying the predictions still require clarification, the systems have considerable untapped value for basic research. In the present paper, we tap into some of this unmet potential—we use a specific land-modeling system to address two distinct and relatively unexplored facets of the streamflow prediction problem.

In the first exercise, we address the impact of soil moisture initialization error on streamflow forecast skill. Specifically, we add artificial, prescribed levels of error to the realistic soil moisture initializations employed by Mahanama et al. (2012) in their forecast experiments and then quantify the resulting degradation of the streamflow forecasts. The degradation is then interpreted in terms of the *increase* in skill attainable from improvements in soil moisture initialization, improvements that are expected from the assimilation of data from current and upcoming satellite-based soil moisture missions [viz., the Soil Moisture and Ocean Salinity (SMOS; Kerr et al. 2010) and Soil Moisture Active Passive (SMAP; Entekhabi et al. 2010b) missions].

Our second exercise focuses on precipitation downscaling and its relationship to streamflow forecasts. A number of studies have addressed the issue of downscaling, focusing on the statistical or dynamical approaches used to achieve it (e.g., Luo et al. 2007; Yuan et al. 2012) and on the application of the downscaled data to offline hydrologic systems (e.g., Luo and Wood 2008; see Schaake et al. 2010 for a summary of some outstanding issues). Here, we focus on a specific aspect of downscaling, namely, the increase in the spatial resolution of the precipitation data applied to the distributed offline forecast system; we do not address here the additional step of correcting the higher-resolution data for local biases and other errors. Through careful joint analysis of lower resolution ($1° \times 1°$) and higher resolution ($0.125° \times 0.125°$) offline simulations, we examine an essentially unanswered question: under what climatic conditions can increasing the spatial resolution of precipitation add value to forecasts of large-scale streamflow totals? The potential for added value is indeed found to depend strongly on climate regime.

The two applications of the land-modeling system, discussed separately in sections 2 and 3, are fully independent; the second does not build on the first. The two studies nevertheless have a key unifying theme: both use an offline distributed land-modeling system to address an important yet still unresolved issue in streamflow forecasting. By presenting the two studies together, we aim to demonstrate the power and efficiency of such systems for basic hydrological analyses.

2. Impact of soil moisture initialization error on streamflow forecast skill

a. Overview of basic forecast experiment

Our simulation experiments were performed with the catchment land surface model (LSM) (Koster et al. 2000; Ducharne et al. 2000). This LSM is a state-of-the-art model designed for use with global atmospheric models, with detailed treatments of a full range of processes (stomatal conductance, interception, baseflow, snow, etc.) serving to determine the fluxes that make up the surface water and energy budgets. The model's unique feature is its explicit treatment of the impacts of subgrid soil moisture variability on the computed evaporation and runoff fluxes. The subgrid variability is keyed to topography and to the model's internal soil moisture prognostic variables, which respond (through mass conservation) both to variations in precipitation forcing and to variations in the computed evaporation and runoff.

The catchment LSM has been tested in a number of settings (e.g., Boone et al. 2004; Bowling et al. 2003;

Reichle et al. 2011) and recently served as the land model underlying the Modern-Era Retrospective Analysis for Research and Applications (MERRA) reanalysis (Rienecker et al. 2011; Reichle et al. 2011). The particular version of the catchment LSM used here is the version described in section 6 of Koster and Mahanama (2012), a version shown to produce accurate runoff ratios and runoff variances across the conterminous United States when driven offline with observation-based meteorological forcing.

The strategy employed in our simulation experiments can be described briefly as follows. In a ''reference'' set of forecast simulations, the catchment LSM generates offline forecasts of runoff after being initialized with realistic soil moisture contents, and the skill of the forecasts is quantified through comparison of seasonal runoff accumulations over basins (i.e., simulated streamflow) with corresponding streamflow observations. The forecasts are then repeated with degraded soil moisture initial conditions (with different levels of error imposed in different experiments), and the resulting loss of streamflow forecast skill is quantified. We thereby quantify the sensitivity of forecast skill to soil moisture initialization accuracy, the idea being that this sensitivity has implications for forecast *improvement* through the improved estimation of soil moisture.

Naturally, prior to performing any of the forecasts, realistic initialization states for the soil moisture variables must be established. This is accomplished through a long-term (1920–2002) simulation of the catchment LSM across the conterminous United States (CONUS) at a 0.5° resolution using observation-based atmospheric forcing (Andreadis et al. 2005). The use of this approach with this particular LSM does provide soil moisture estimates with significant skill (particularly in terms of their time variations, which are key to producing high correlations between forecasted and observed streamflows), as demonstrated by the analysis of ''open loop'' simulations in numerous data assimilation studies (e.g., Reichle et al. 2007; Liu et al. 2011; Draper et al. 2012). The offline simulation used here is equivalent to that of the ''CTRL'' simulation described by Mahanama et al. (2012) except for the aforementioned use of an updated version of the catchment LSM.

The reference forecasts consist of 3-month model integrations starting on 1 January, 1 April, 1 July, and 1 October of each year in the multidecadal period (1920–2002). The soil moisture states on a given date from the CTRL simulation serve as the initial conditions for the reference forecast with that start date. In contrast, we initialize snow and ground temperatures in the land model to their climatological states, as determined from the multidecadal CTRL simulation. Similarly, the atmospheric forcing used during each forecast is the climatological seasonal cycle of forcing derived from the forcing dataset of the CTRL simulation (though with the year of forecast excluded from the climatology calculation, for proper cross validation)—all July–September forecasts, regardless of year, utilize essentially the same atmospheric forcing. As a result, the differences between, say, the July–September streamflow forecasts for different years stem solely from the different 1 July soil moisture conditions used—any skill obtained in the forecasts reflects the soil moisture initialization alone. While skill would presumably increase if we included snow initialization and forecasted meteorological conditions in the simulations, we are specifically interested here in soil moisture impacts on skill; this experimental design allows us to isolate and analyze these impacts. Note that the reference forecasts are essentially equivalent to the Exp3 simulations of Mahanama et al. (2012), the only difference being that here we rely on the newer version of the catchment LSM alone rather than on an ensemble of four land surface models.

To quantify streamflow forecast skill, the grid-cell-based seasonal runoff values produced in the forecast simulations are averaged over the 20 basins shown in Fig. 1 to generate a time series of seasonal streamflow estimates for each basin. These estimates are in turn compared to contemporaneous observations of streamflow in these basins. [See Table 1, Koster et al. (2010), and Mahanama et al. (2012) for details; like Koster et al. (2010) but unlike Mahanama et al. (2012), all streamflow generated upstream of a gauge site is considered in a given calculation, not just the amount generated between upstream and downstream gauges in the same river network. Also, three basins considered by Mahanama et al. (2012) are not considered here for reasons outlined in Koster and Mahanama (2012).]

The resulting correlation coefficient ($r_{\text{Q-obs}}$) between the simulated and observed time series serves as our skill metric. Given the use of climatological seasonal cycles for the meteorological forcing in our forecasts, this is the most suitable metric for this study; we note in any case that the correlation measure of skill can easily be converted to an RMSE measure of skill through the application of known time series moments (Entekhabi et al. 2010a). Note that the observed streamflows are naturalized, having been modified to remove the impacts of reservoir operations on streamflow totals. The consideration of seasonal totals reduces any errors associated with the neglect, in the model analysis, of the residence time of runoff water in the river network upstream of the gauges.

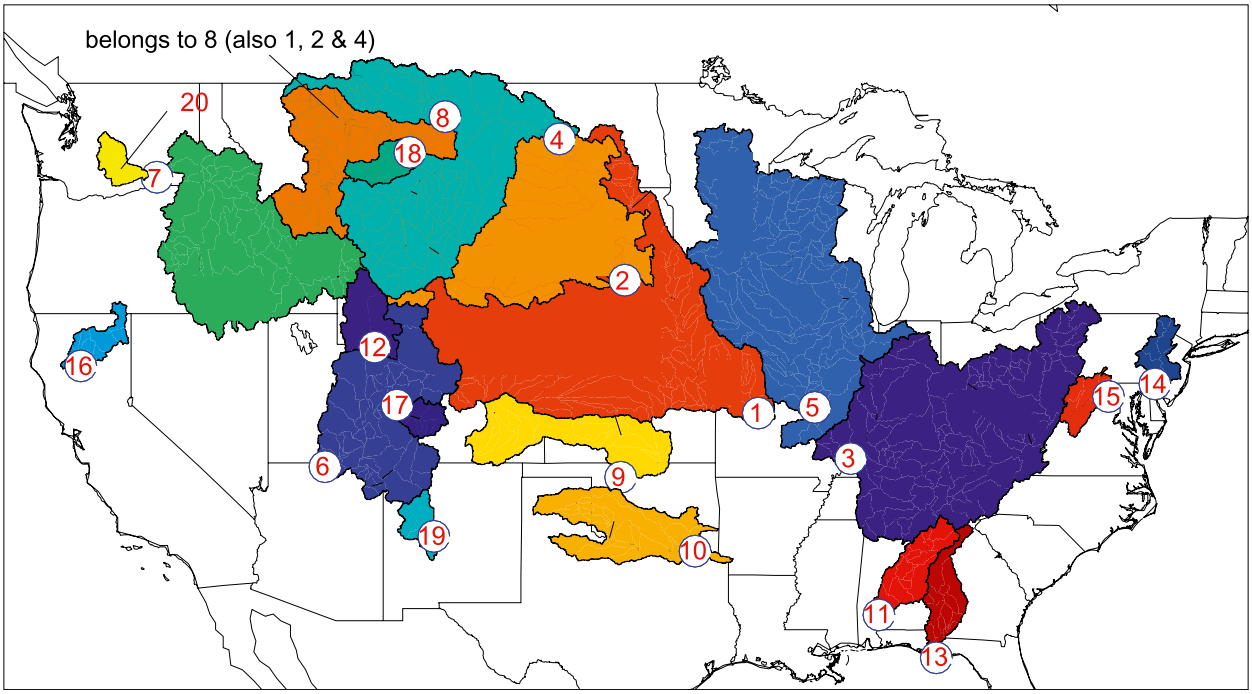

FIG. 1. Basins examined in this analysis. The numbers are located at the stream gauge sites.

b. *Imposition of soil moisture error*

The set of forecast simulations described above serves as the reference—the "unperturbed error" case—for a series of experiments in which we impose specific levels of error in the initial soil moisture fields prior to performing the forecasts. For these latter experiments, we first generate a series of spatially correlated error fields ζ with the error for a given grid cell (i, j) and year t selected from a unit normal distribution. The imposed e-folding decorrelation length scale of the error is taken to be 2° in longitude and latitude, which is consistent with the length scale for errors employed by Reichle and Koster (2003). The error $\zeta(i, j, t)$ is applied to the soil moisture amounts, $W(i, j, t)_{\mathrm{ctrl}}$, in the catchment LSM's subsurface moisture reservoirs (i.e., its root zone excess and catchment deficit variables) at that location in year t from the CTRL simulation; this is done by modifying the standard normal deviate, or Z-score, of each soil moisture quantity:

$$\frac{W(i,j,t)_{\mathrm{pert}} - W(i,j)_{\mathrm{mean}}}{\sigma_{W(i,j)}} = \frac{W(i,j,t)_{\mathrm{ctrl}} - W(i,j)_{\mathrm{mean}}}{\sigma_{W(i,j)}} + \varepsilon\zeta(i,j,t), \tag{1}$$

where $W(i, j, t)_{\mathrm{pert}}$ is the degraded, or perturbed, soil moisture to be used in the experiment forecast, $W(i, j)_{\mathrm{mean}}$ is the mean of the soil moisture (for that time of year) at the grid cell, $\sigma_{W(i,j)}$ is the corresponding standard deviation across the years (again, for that time of year), and ε is the user-imposed scaling factor that determines the average size of the imposed error.

A note of caution is appropriate here. Although the soil moistures represented here are large-scale averages, so that the central limit theorem may impose some Gaussian character to the values (and their errors) encountered in large-scale measurements, the actual distributions will probably be non-Gaussian, if only because soil moisture is bounded from below by zero and from above by the soil porosity. The most ideal strategy for our experiment would be to apply a much more complex, non-Gaussian error field to our soil moistures, with the specified decorrelation length scale. Such a strategy, however, would have two disadvantages: (i) it would be difficult to implement with our current numerical tools, and (ii) the construction of the fields would involve, in any case, some subjective assumptions about the character of the error that might themselves cloud the interpretation of the results. For our experiment, we assume easily managed and easily understood normally distributed errors, applied in the context of (1); our results must be considered in light of this simplifying assumption.

We note that in applying the soil moisture errors, final soil moisture values were naturally constrained to lie within realistic bounds. While this may modify slightly

TABLE 1. Characteristics of the basins examined in this study. Three small basins examined in Mahanama et al. (2012) are not examined here because of apparent inconsistencies in the observation-based precipitation and streamflow datasets; see Koster and Mahanama (2012) for details.

	River name	Station name	Basin area (km^2)	Latitude (°N)	Longitude (°W)	Observation Period
1	Missouri	Hermann (includes basins 2, 4, 8, and 18)	1 353 275	38.71	92.75	1920–97
2	Missouri	Ft. Randall Dam (includes basins 4, 8, and 18)	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 18)	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 17)	289 562	36.87	111.58	1920–2003
7	Snake	Ice Harbor Dam	281 015	46.25	118.88	1927–92
8	Milk	Fort Peck Dam (includes basin 18)	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	Potomac	Point of Rocks	25 000	39.27	77.54	1950–96
16	Sacramento	Bend Bridge	23 051	40.29	122.19	1920–2003
17	Gunnison	Near Grand Junction	20 533	38.98	108.45	1920–2003
18	Musselshel	Moseby	20 321	46.99	107.89	1941–2003
19	Rio Puerco	Bernardo	19 036	34.41	106.85	1940–2003
20	Yakima	Near Parker	9479	46.50	120.44	1925–2003

the effective values of ε used, these slight modifications are implicitly accounted for in our plots, which will show streamflow forecast skill versus actual soil moisture error rather than versus the precise value of the imposed ε.

In analogy to the reference forecast experiment, each "imposed error" forecast experiment consists of a series of 3-month forecasts, one for each season of each year, using a single value of ε. The skill level is again computed as the correlation coefficient $r_{\text{Q-obs}}$ between the simulated streamflows in a basin and the observed streamflows over the multidecadal period. In fact, for thoroughness, we repeat the forecast experiment with a given ε value 10 times, allowing for 10 different sets of realizations of the $\zeta(i, j, t)$ field; the resulting 10 values of $r_{\text{Q-obs}}$ (which are, of course, similar) are averaged together prior to plotting. A total of six experiments are performed, with increasing levels of error: ε is set in turn to 0.1, 0.3, 0.6, 1, 1.5, and 2.

c. Results

Prior to comparing the forecasts with observations, we show in Fig. 2 the temporal correlation coefficient over the multidecadal period, at each grid cell, between model quantities in the reference experiment and those in two of the experiments with imposed soil moisture error. Setting ε to 0.3 produces, for the April–June (AMJ) forecasts, initial soil moisture contents on 1 April that correlate with those of the reference simulation with a largely uniform correlation coefficient (r_W) of about 0.9 (top-left panel of Fig. 2). The correlation coefficient for 1 April soil moisture drops to about 0.5 when ε is set to 1.0 (top right panel of Fig. 2). Again, the spatial distribution of this correlation is fairly uniform, an indication that the error generation technique was applied correctly.

The corresponding correlations between the forecasted AMJ runoff rates in the reference experiment and those in the experiments with imposed error are similar but show some important differences. For the $\varepsilon = 0.3$ case (bottom-left panel of Fig. 2), the correlation for forecasted runoff (r_Q) is about the same as that for soil moisture except in some western mountainous regions, for which it is reduced significantly, and along a north–south swath down the center of the country, for which it is reduced slightly. This pattern of r_Q reduction is also found for the $\varepsilon = 1.0$ case (bottom right panel of Fig. 2). The reductions, which result from nonlinear runoff-generating processes in the model, should act to amplify any streamflow forecast error associated with the inaccurate initialization of soil moisture.

Of course, a high correlation in Fig. 2 does not imply significant skill in streamflow forecasts, since the reference experiment itself may have limited skill. We now turn to a comparison of forecasted streamflows with observations, with an eye toward showing how the

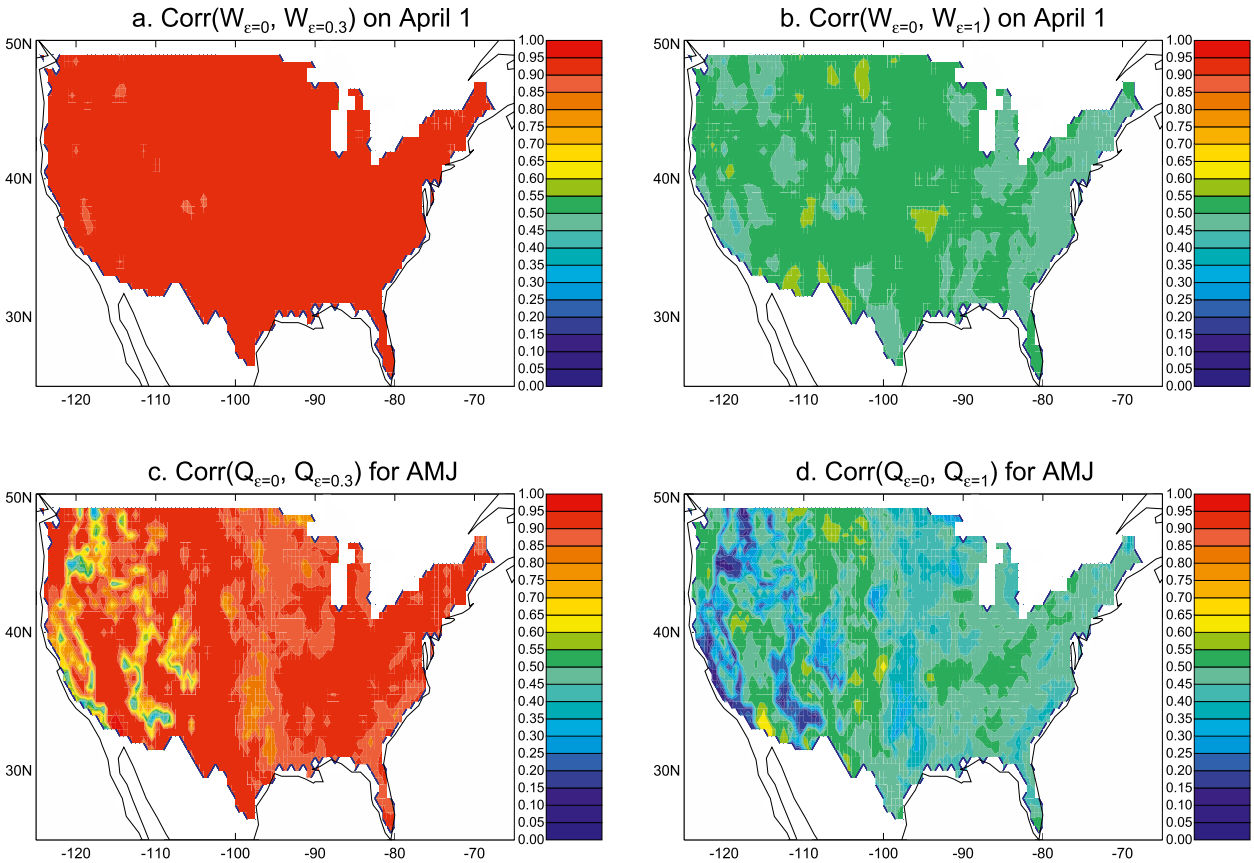

FIG. 2. (a) Time correlation r_W between the initial (1 Apr) soil moisture contents used in the reference forecast experiment ($\varepsilon = 0$) and those used in the experiment with imposed errors corresponding to $\varepsilon = 0.3$. (b) As in (a), but for ε set to 1.0 rather than 0.3. (c) Correlation (r_Q) between the multidecadal time series of April–June runoff totals generated in the reference forecast experiment ($\varepsilon = 0$) and those generated in the experiment with $\varepsilon = 0.3$. (d) As in (c), but for ε set to 1.0 rather than 0.3.

imposition of soil moisture error (beyond that which already exists in the reference experiment) translates to increased error in streamflow forecasts. Again, we focus our analysis on the hydrological basins shown in Fig. 1, which cover a wide range of climatic regimes. We also emphasize again that all of the simulations discussed here represent true forecasts, in that predictions are made without knowledge of conditions during the forecast period and that the forecasts are compared directly with observations. Thus, any skill uncovered represents true skill and is not, for example, an artifact of model parameterization or experimental design.

Figure 3 provides a sample set of results, focusing on July–September (JAS) streamflow forecasts (i.e., the 3-month forecasts with 1 July initialization) in the Red River Basin. Each dot in the figure represents a different experiment, that is, it shows the results for a unique value of ε. The dot's ordinate is located according to the average skill in streamflow prediction obtained in the experiment, measured as the correlation coefficient ($r_{\text{Q-obs}}$) between the forecasted JAS streamflow totals and the observed totals. To compute the dot's abscissa, which represents the imposed soil moisture error, we average the grid cell values of r_W for 1 July, as illustrated in Fig. 2 for two of the experiments, across the examined basin. The dot at $r_W = 1$ thus represents the forecast skill obtained with the reference forecasts. As expected, the imposition of greater soil moisture error (i.e., lower values of r_W) leads to a decrease in streamflow forecast skill. The key result of Fig. 3, however, is the near linearity of this reduction in error with respect to r_W. In essence, the forecast skill of each experiment is reduced from that of the reference experiment by a factor of about r_W, and the regression line fitted through the points approximately goes through the origin—marks of a direct linear relationship. This linearity is seen despite the small but significant impacts of nonlinearity indicated for this region in Fig. 2.

Figure 4 provides the results, in the same form, for all 20 basins and all four seasons. A near-linear decrease of

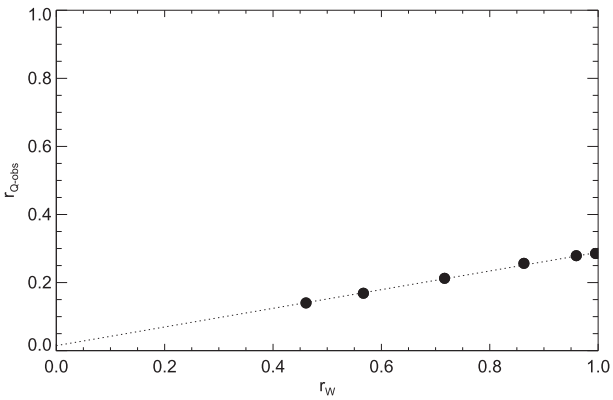

FIG. 3. Plot showing how imposed soil moisture error acts to reduce streamflow forecast skill. Each dot corresponds to a different forecast experiment. A given dot's abscissa is the average correlation (across the Red River basin) between the time series of initial soil moisture contents used by the forecast experiment and those used by the reference forecast experiment ($\varepsilon = 0$). Its ordinate is the correlation between the forecasted streamflow for the basin (i.e., the time series of July–September totals) and the observed streamflow. The dot at $r_W = 1$ (two dots, in fact, fall on top of each other there) thus shows the maximum skill attained with the system for this basin and season.

forecast skill with increasing soil moisture error (decreasing r_W) is seen here for all cases. Most of the plots show direct linear relationships, presumably reflecting the first-order agreement between the r_W and r_Q values plotted in Fig. 2. There are exceptions, however. First, for larger basins (those near the top of the figure), the falloff in the skill for a given decrease in r_W is not as large as would be indicated by a direct relationship; the y intercept of the regression line for these basins tends to be positive. This presumably reflects the e-folding decorrelation length scale (2°) of the error fields imposed. For a large enough basin, the effects of the soil moisture errors imposed in different large-scale (2° scale) subareas may cancel each other out—for example, excessive streamflows induced by positive soil moisture errors in the northern end of a basin may be offset by insufficient streamflows induced by negative soil moisture errors in the southern end. For the smaller basins, which are closer to a 2° scale, the imposed soil moisture errors on a given forecast start date are necessarily more uniform, and such offsets cannot occur. (This is, in fact, verified with additional experiments, not shown here. In these additional experiments, we find that $r_{\text{Q-obs}}$ for a given value of ε decreases with increasing decorrelation length scale only in the larger basins.)

Second, some small basins lie in areas for which nonlinearities amplify the translation of soil moisture error into runoff production error (see Fig. 2 and accompanying discussion), and these basins, as expected, sometimes show a faster reduction of runoff forecast skill with decreasing r_W than would be indicated by a direct relationship (i.e., the regression line has a negative y-intercept). Examples include the Rio Puerco, Sacramento, and Gunnison River basins. Note that while the reduction is faster, it is also still essentially linear.

To illustrate the translation of soil moisture error into streamflow prediction error with a concrete example, we show in Fig. 5 the translation associated with the prediction of a specific hydrological drought. Streamflow during June–August 1984 was anomalously low in the Red River (as measured in Arthur City); the observations show that the naturalized streamflow during this period was one standard deviation below the mean. The forecast model with no imposed soil moisture error ($\varepsilon = 0$) reproduced this anomaly well. (Note that such a strong agreement for a specific event is reasonably frequent but not typical.) As soil moisture error increases, the magnitude of the streamflow deficit, averaged across the 10 ensemble members, is seen to decrease, while the spread of the predictions across the ensemble members is seen to increase. For the maximum imposed error ($\varepsilon = 2$), the predicted streamflow deficit, averaged across the 10 ensemble members, is about 0.3 standard deviations below the mean, with two of the ensemble members for this experiment predicting a streamflow surplus.

d. Implications for improved soil moisture estimates

The initial soil moistures used in the reference experiment, which were derived by driving the land model with antecedent meteorological forcing, are far from perfect given imperfections in the forcing data and in the model's conceptualizations and parameterizations. Important improvements in initialization accuracy are expected from new data sources and new techniques for soil moisture estimation. Both the SMOS and the SMAP satellite missions, for example, have a target RMSE of $0.04\,\mathrm{m^3\,m^{-3}}$ for soil moisture estimation (Kerr et al. 2010; Entekhabi et al. 2010b), with high spatial resolution across the globe (~10-km resolution every 3 days for SMAP). These accuracy levels may or may not be higher than model-based estimates (and require, in any case, information on local soil moisture moments to translate into the correlation metrics examined here), but regardless, the process of data assimilation should produce a soil moisture product that is superior to either the model-based estimates or satellite-based estimates alone (e.g., Reichle et al. 2007; Draper et al. 2012).

To see how the expected increases in accuracy might affect streamflow forecasts, consider the schematic shown in Fig. 6. The abscissa represents $r_{\text{W-truth}}$, the correlation between the estimated time series of initial soil moisture

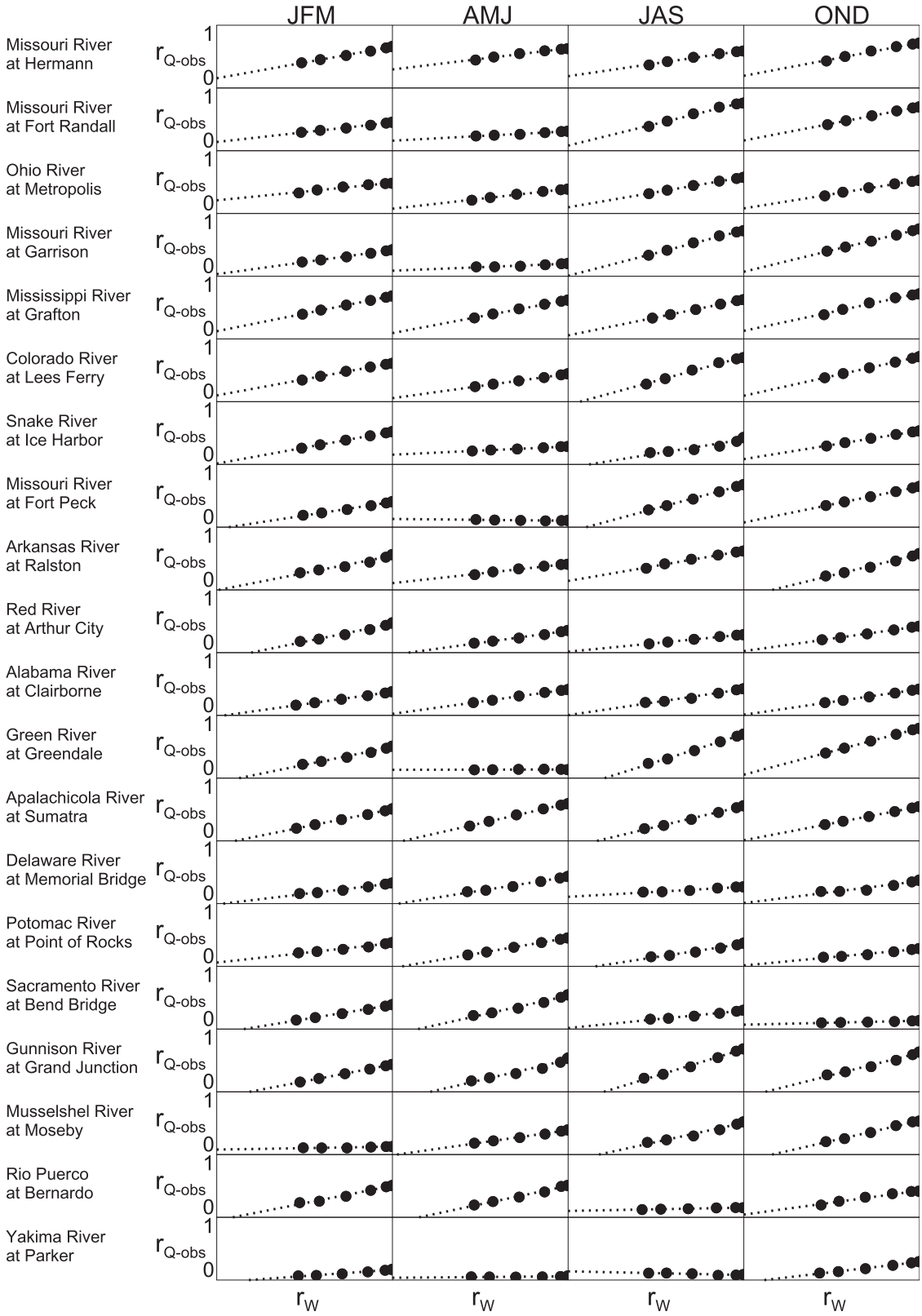

FIG. 4. As in Fig. 3, but for all basins (in order of size) and (left to right) seasons. The *y* axis in each panel goes from 0 to 1.

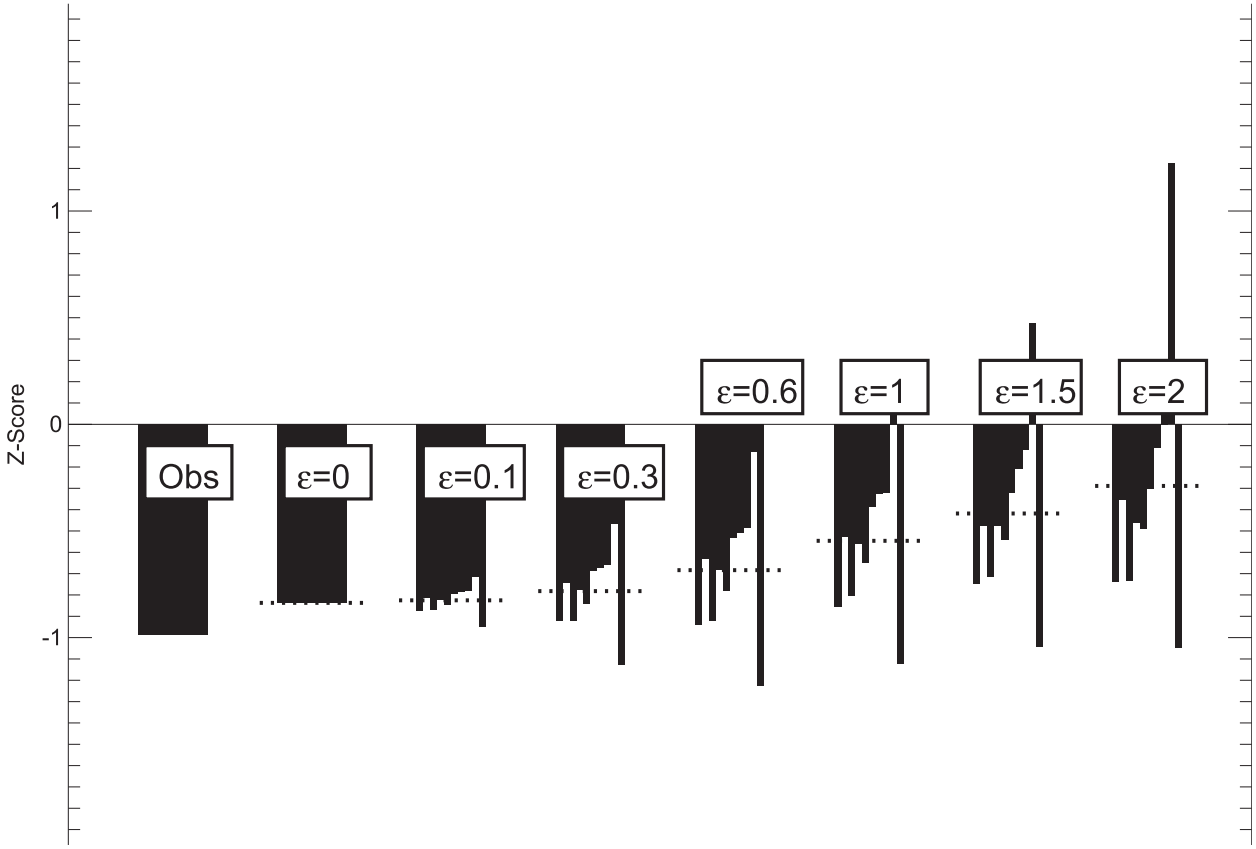

FIG. 5. Total streamflow in the Red River basin (as measured at Arthur City) for July–September 1984 from (left to right) observations, the reference forecast experiment, and the six experiment forecasts with imposed error. All values are shown in terms of dimensionless standard normal deviations. The horizontal dotted lines show the mean values across the ensemble members.

contents and the (unknown) true time series. For the sake of presentation only, we assume in this figure that the reference set of forecasts discussed above has an $r_{\text{W-truth}}$ of 0.8 in the considered basin. (We have, of course, no a priori way of knowing what $r_{\text{W-truth}}$ is.) Given the independence between the actual errors in our reference time series and the errors we impose through (1) in our experiments, it can be shown that the $r_{\text{W-truth}}$ values for the individual "imposed error" experiments are simply their r_W values (i.e., their correlations against the reference time series) multiplied by 0.8. The relationship between streamflow forecast skill and soil moisture estimation skill, where the latter is now assumed measured against an unknown truth, remains linear under this transformation. The plot thereby provides a simple estimate of the increase in streamflow forecast skill that would stem from an increase in soil moisture initialization accuracy:

$$\frac{\partial r_{\text{Q-obs}}}{\partial r_{\text{W-truth}}} = S \approx 0.7(\text{in this example}), \qquad (2)$$

where S is the slope of the line in the plot.

Knowledge of the true value of S would be quite valuable, since it would allow us to translate improvements in soil moisture estimation into quantitative estimates for the associated improvement in streamflow forecast skill. Obtaining S, however, is impossible given our lack of knowledge of $r_{\text{W-truth}}$ for the reference forecasts—we cannot evaluate $r_{\text{W-truth}}$ without the necessary comprehensive observations of historical soil moisture. Nevertheless, we do have an estimate for $S_{\min}$, the *minimum possible* value of S: $S_{\min}$ is the slope obtained when $r_{\text{W-truth}}$ for the reference experiments is assumed to be 1. In essence, $S_{\min}$ provides the *lower limit* of forecast skill increase associated with an improved soil moisture initialization—any skill increase associated with improved soil moisture estimation will be at least as large as that suggested by $S_{\min}$. The $S_{\min}$ values for the different basins and seasons examined above are simply the slopes of the fitted lines in Fig. 4.

The $S_{\min}$ values for the different basins are plotted, for each season, in Fig. 7. The minimum sensitivity of

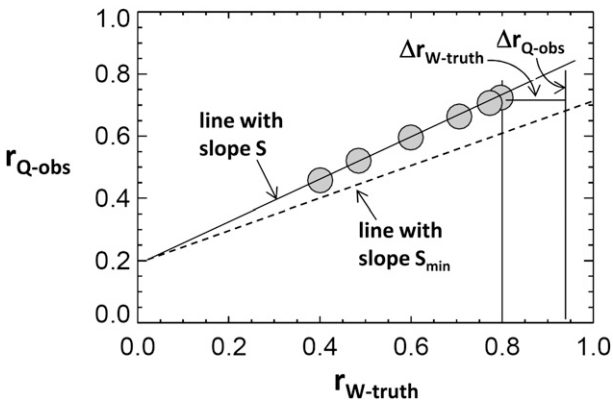

FIG. 6. Diagram highlighting the interpretation of the forecast experiment results in terms of the sensitivity of streamflow forecast skill to improvements in the estimation of initial soil moisture.

streamflow forecast skill to improvements in soil moisture estimation is seen to be largest in the summer and fall, particularly in mountainous areas. Again, the actual improvement in forecast skill for a given improvement in soil moisture estimation should be larger than the values shown. Note that by construction, the plots are similar to the raw streamflow forecast skill plots shown for the Exp3 simulations by Mahanama et al. (2012); the differences stem only from the use of one land model here and, more importantly from a scientific standpoint, from the fact that, for reasons discussed above, not all of the fitted lines through the points in Fig. 4 go through the origin.

Finally, we note the potential for rough estimates of the change in streamflow prediction skill (i.e., the change in $r_{Q\text{-obs}}$) associated with a specific, area-averaged change in the RMSE of soil moisture estimation—a quantity more familiar to many soil moisture scientists. To obtain such estimates, we begin with the relationship between unbiased root-mean-square error (ubRMSE) and correlation coefficient r as provided by Entekhabi et al. (2010a):

$$\mathrm{ubRMSE} = (\sigma_{\mathrm{est}}^2 + \sigma_{\mathrm{true}}^2 - 2r\sigma_{\mathrm{est}}\sigma_{\mathrm{true}})^{0.5}, \qquad (3)$$

where for this discussion σ_{est} is the standard deviation (in time) of the estimated soil moisture, σ_{true} is the true standard deviation of soil moisture, and r is equivalent to $r_{\mathrm{W\text{-}truth}}$ as discussed above. Rearranging (3), and assuming that a reduction in ubRMSE (e.g., through the assimilation of remotely sensed observations) has no impact on σ_{est} or σ_{true}, we compute the derivative of $r_{\mathrm{W\text{-}truth}}$ with respect to ubRMSE:

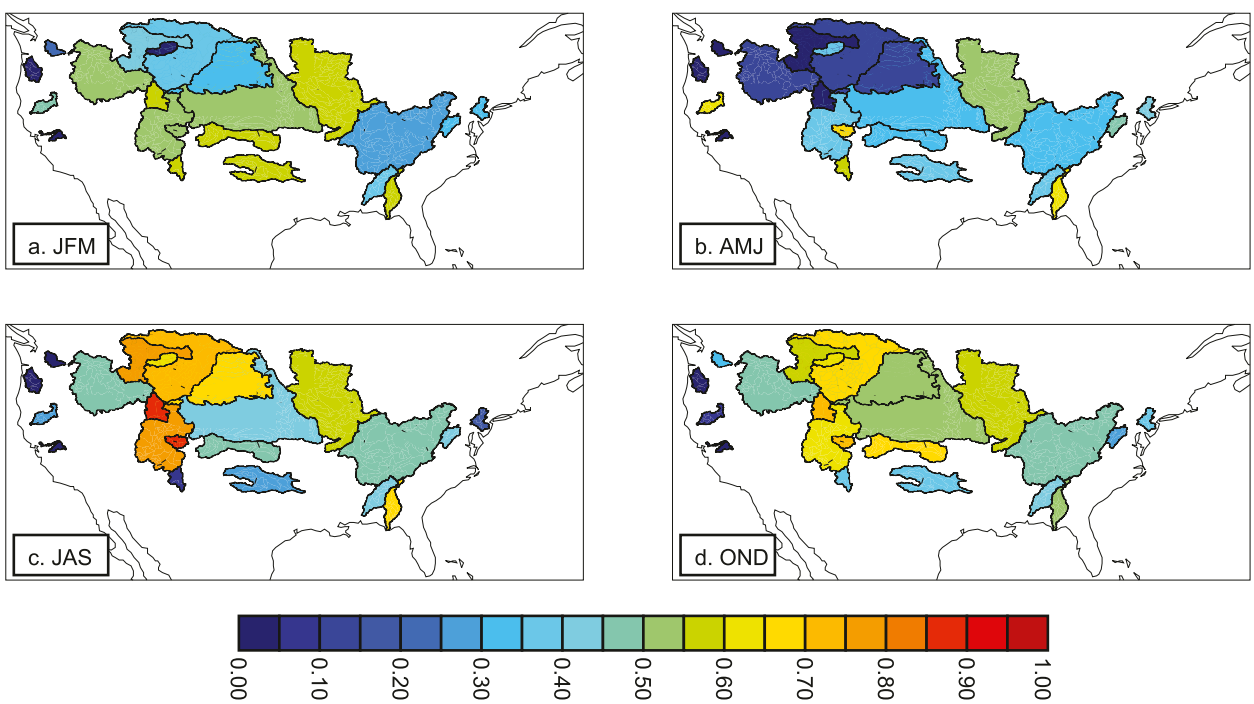

FIG. 7. Estimates of the lower bound of the ratio $\Delta r_{\mathrm{Q\text{-obs}}}/\Delta r_{W\text{-truth}}$ (i.e., the lower bound of the sensitivity of streamflow forecast skill to initial soil moisture accuracy), as derived from the forecast experiments: (a) January–March (JFM), (b) April–June (AMJ), (c) July–September (JAS), and (d) October–December (OND).

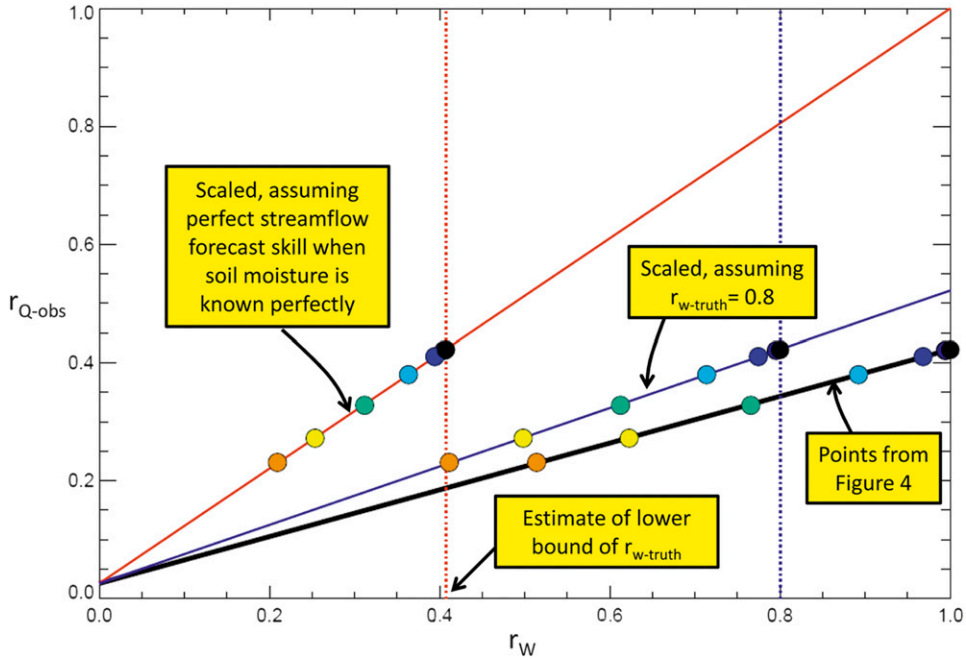

FIG. 8. Demonstration of how the lower bound of $r_{\text{W-truth}}$ (a measure of how well we know the true soil moisture) can be estimated from the experiments addressing the impact of soil moisture initialization error on streamflow forecast skill. See text for details. Dots of the same color represent results from the same experiment, scaled in different ways; these colors are not keyed to the color bars used in any of the maps in the figures above.

$$\frac{\partial r_{\text{W-truth}}}{\partial \text{ubRMSE}} = -(\text{ubRMSE})/(\sigma_{\text{est}}\sigma_{\text{true}}), \qquad (4)$$

from which we derive the derivative of streamflow prediction skill ($r_{\text{Q-obs}}$) with respect to ubRMSE:

$$\frac{\partial r_{\text{Q-obs}}}{\partial(\text{ubRMSE})} = \frac{\partial r_{\text{Q-obs}}}{\partial r_{\text{W-truth}}}(-\text{ubRMSE})/(\sigma_{\text{est}}\sigma_{\text{true}}), \qquad (5)$$

where again, $\partial r_{\text{Q-obs}}/\partial r_{\text{W-truth}}$ has a minimum possible value given by the slope of the $r_{\text{Q-obs}}$– r_W relationship in Fig. 4.

Utilization of (5), of course, is difficult given the need to quantify accurately both σ_{true} and ubRMSE in the general absence of soil moisture observations. To give a flavor, though, for the types of values that might be involved, consider the Ohio basin, the large basin in eastern CONUS in Fig. 1. Assuming (arbitrarily) that ubRMSE for this basin is 0.02 ($\text{m}^3\,\text{m}^{-3}$) prior to using satellite-based soil moisture information, and that σ_{true} is the same as σ_{est} as computed from the model simulation ($0.027\,\text{m}^3\,\text{m}^{-3}$), we can use (5) to estimate that reducing ubRMSE to 0.015 would cause $r_{\text{Q-obs}}$ for JAS streamflow forecasts to increase from 0.58 to at least 0.65.

e. Supplementary result: Lower bound for present-day skill in estimating soil moisture

A curious side benefit of the linearity found in Fig. 4 is worth mentioning. The linearity allows us to infer a minimum value for $r_{\text{W-truth}}$, that is, an estimate for a lower bound for how well we know soil moisture state based on present-day observational networks. The idea is illustrated in Fig. 8. In the plot, the dots along the heavy black line represent pairs of r_W and $r_{\text{Q-obs}}$ from a given panel in Fig. 4. Recall that in Fig. 6, the (unknown) value of $r_{\text{W-truth}}$ was arbitrarily assumed to be 0.8, allowing the abscissas of the points to be scaled by that factor; performing the same operation on the points in Fig. 8 yields the points along the thinner blue line. Notice that according to the blue line, a perfect knowledge of soil moisture ($r_W = 1$, against truth) would lead to a streamflow forecast skill of $r_{\text{Q-obs}} = 0.525$. That is, if with current measuring systems we know soil moisture with an $r_{\text{W-truth}}$ of 0.8, then (making use of the evident linearity) $r_{\text{Q-obs}} = 0.525$ is the best streamflow forecast skill we could ever attain through improvements in soil moisture measurement systems.

Now consider the scaling that brings the points on the heavy black line to the red line in Fig. 8. With this scaling, streamflow is forecasted perfectly when soil

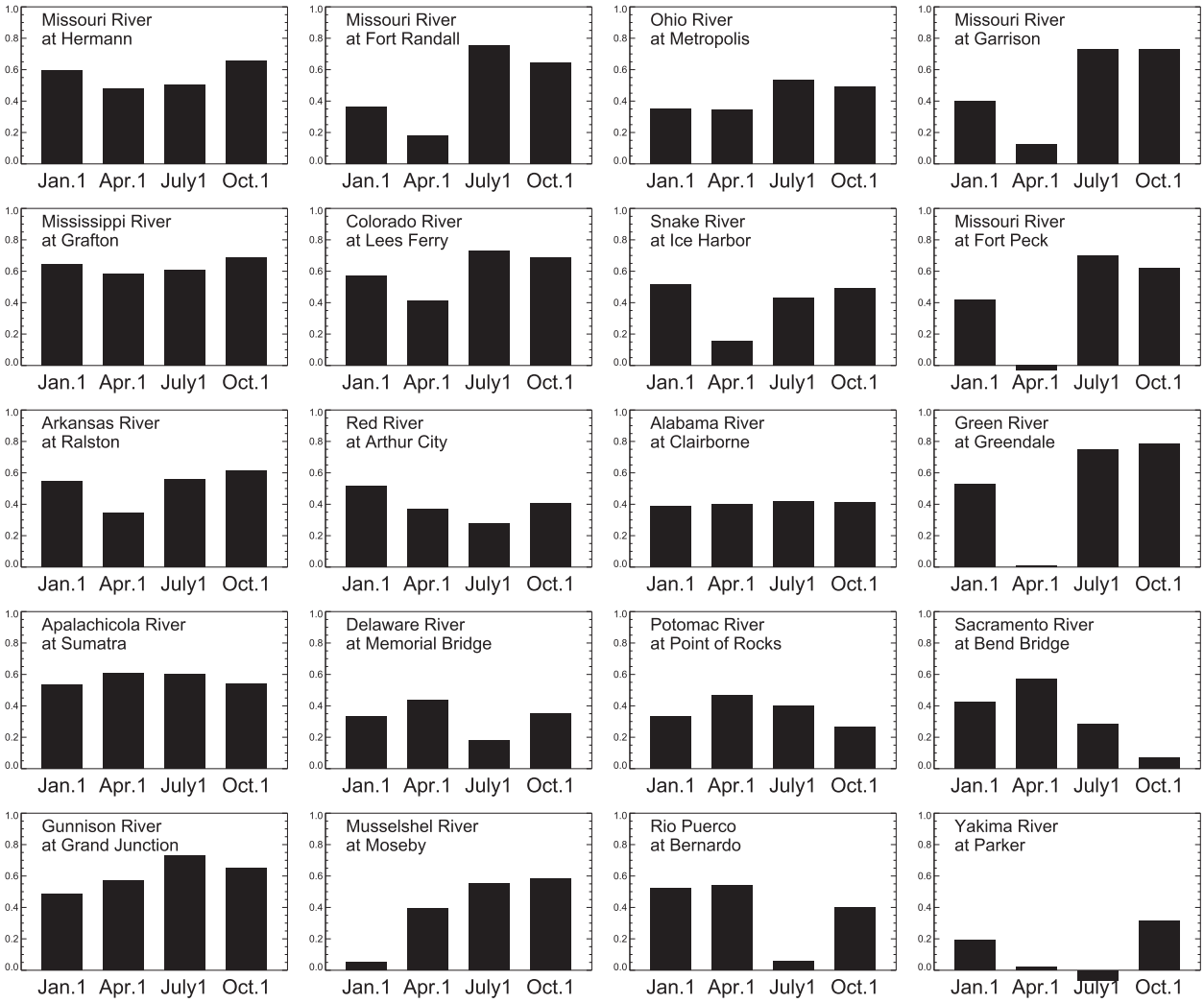

FIG. 9. Estimated lower bounds of $r_{\text{W-truth}}$ for each basin (in order of size) and season considered.

moisture is perfectly known. Any further scaling would lead to the impossible condition of $r_{\text{Q-obs}} > 1$ for perfectly known soil moisture, which has an important implication: although we do not know how well soil moisture is estimated with current measurement systems (i.e., we do not know the value of $r_{\text{W-truth}}$ for the reference forecast experiment), $r_{\text{W-truth}}$ cannot be smaller than the abscissa of the black dot on the red line in Fig. 8 (about 0.42), for a smaller value would imply the aforementioned impossible condition. Because streamflows also depend on precipitation during the forecast period and thus can never be forecasted perfectly with soil moisture information alone (i.e., $r_{\text{Q-obs}}$ will be always be less than 1 even if r_W were 1), the actual value of $r_{\text{W-truth}}$ will presumably be significantly higher than the estimated minimum value.

This estimation procedure requires more extrapolation across the fitted line than the derivative calculation associated with Fig. 6. The degree to which the underlying relationships are not truly linear across all possible values of r_W and are not truly direct (i.e., do not go through the origin) may limit significantly the accuracy of the estimates produced and may even allow an occasional negative value. (Notice, for example, the small but positive, and thus unrealistic, skill seen in Fig. 8 when r_W is 0, an indication that such extrapolation is subject to risk.) With this important caveat, Fig. 9 provides, for each basin and season addressed in Fig. 4, estimates of the lower bound of $r_{\text{W-truth}}$ as produced with the procedure. The calculations imply, for example, that for the Ohio River basin, the model-based soil moistures used in our analyses, which are based on historical time series of meteorological forcing, compare with the true (unknown) soil moistures there with a correlation coefficient higher than 0.35 on 1 January and higher than

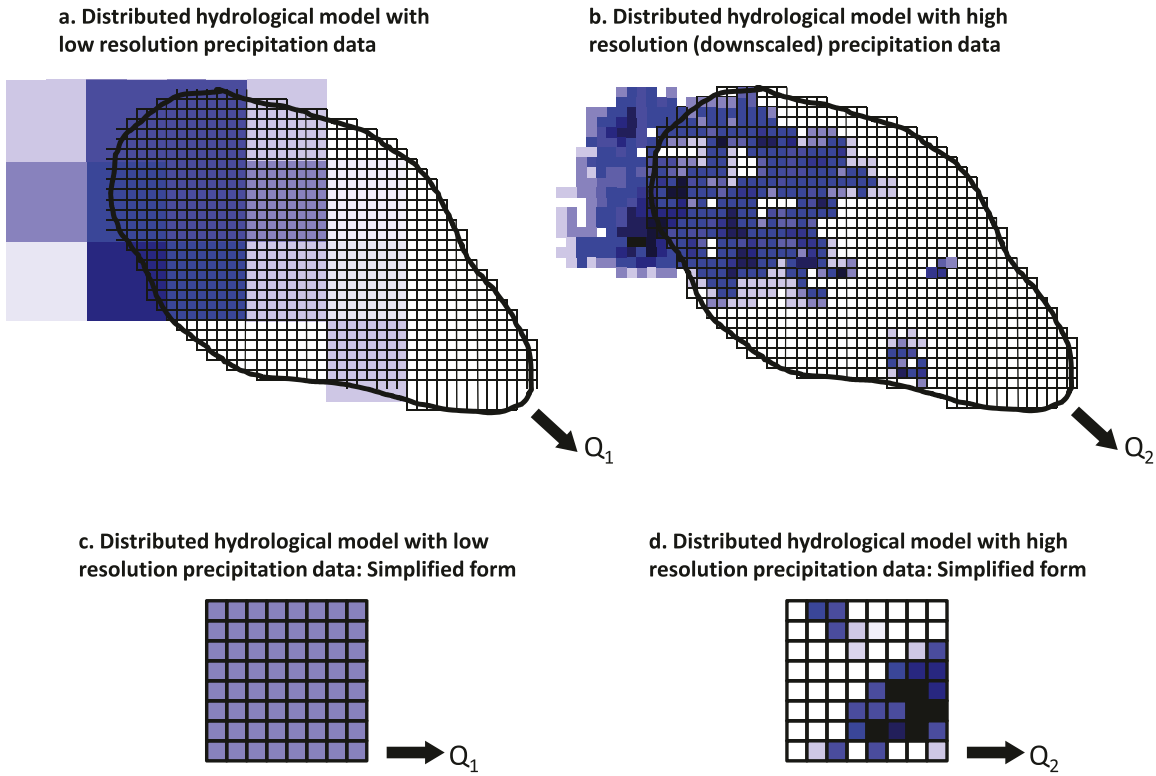

FIG. 10. Schematic illustrating the disaggregation problem addressed in section 3. See text for details.

0.5 on 1 July. The lower bounds for soil moisture estimation skill exceed 0.7 for some locations and seasons. Such estimates of soil moisture accuracy at the large scale, of course, are generally very difficult to obtain; the streamflow-based diagnostic provides a lower bound for a highly elusive quantity.

Particularly low values are seen for several larger basins on 1 April. This may reflect the fact that snowpack in these basins on 1 April is more important than soil moisture for determining April–June streamflow. As a result, any streamflow-based diagnostic of how well we know soil moisture on 1 April would presumably be biased low.

3. Impact of an increased spatial resolution of precipitation on streamflow forecast skill

We now turn to our second (and independent) study, which has a different focus. Rather than examining the impact of initialization error on streamflow forecast skill, we examine here the potential impact of high-resolution precipitation information on this skill. As in the first study, an offline-distributed land-modeling system is used to address the problem.

a. Overview of experiment

As discussed in section 1, seasonal precipitation forecasts from global numerical models can be used to drive an offline hydrological model and thereby produce seasonal streamflow forecasts. The mechanisms that control streamflow generation, however, operate at spatial scales much smaller than those captured by the global-modeling systems. While a hydrological model may be calibrated to operate efficiently with low-resolution precipitation products, the subgrid (high resolution) distribution of precipitation in nature does contain information that the low resolution data necessarily lacks, and a distributed hydrological model might, in theory, be able to transform this extra information, if available, into improved streamflow simulation. It is natural to ask whether a realistic increase in the spatial resolution of the precipitation forcing prior to its application in a high-resolution hydrological model would improve the forecast of large-scale streamflow.

Consider, for example, the two hydrological-modeling designs shown in Figs. 10a and 10b. Both designs cover the hydrological basin with a high-resolution array of surface-modeling elements, and both produce, through application of meteorological forcing, streamflow estimates (Q_1 and Q_2) at the basin's outlet. Shown in shading is the nature of the precipitation used in the two cases—in Fig. 10a, the precipitation is applied at low resolution (e.g., at the resolution of the precipitation forecast model), whereas in Fig. 10b, the low resolution precipitation

data are disaggregated spatially in a way that conserves the total precipitation volume but captures its subgrid variability. We can restate our question as follows: assuming the forecasted precipitation and its disaggregation to the finer scale is accurate, is Q_2 inherently more accurate than Q_1?

Of course, under these assumptions, the high-resolution streamflows produced in Fig. 10b have, by definition, useful information not allowed by the setup in Fig. 10a. In this analysis we do not address the impact of the disaggregation on high resolution streamflow data; we only address the impact of the disaggregation on the *large-scale-average* streamflow, as represented in the figures by Q_1 and Q_2.

Note also that we will effectively quantify here the *maximum possible* impact of the disaggregation on accuracy by effectively assuming perfect seasonal precipitation forecasts and a perfect methodology for the disaggregation of the precipitation in both space and time. The idealized design of this experiment is thus in stark contrast to that employed in section 2 above, in which true estimates of forecast skill were derived from a comparison of model-forecasted streamflows to observations. In the present design, we purposefully avoid presenting comparisons of forecasted runoffs with observations; the shortness of the time period considered (as determined by the availability of high-resolution precipitation forcing data), as well as errors in the precipitation data and in the model itself, prevent the observations from providing any definitive conclusions. The justification for our idealized strategy is that observations are not needed to address the broader question posed here, namely, would disaggregation be useful even if all system components were perfect? If no impact is seen even under such idealized conditions, then a disaggregation procedure applied to a real-time large-scale seasonal forecast would have no hope of being beneficial for that forecast.

Two pairs of offline simulations are compared in this analysis. Each simulation covers the CONUS regime, applying the catchment LSM at a resolution of $1/8° \times 1/8°$ over the period 1981–2008 after a multidecadal spinup procedure. In the first pair of simulations [labeled high-resolution precipitation (HRP) and low-resolution precipitation (LRP)], we characterize the land surface with spatially varying fields (at $1/8° \times 1/8°$ resolution) of vegetation, soil, and topography parameters derived from observations. All meteorological forcing except for precipitation is derived from the observations-based dataset of Sheffield et al. (2006); the $1° \times 1°$ (nonprecipitation) forcing in that dataset is applied uniformly across the 64 $1/8° \times 1/8°$ cells contained within. Simulations HRP and LRP differ only in the nature of the precipitation forcing. In simulation HRP, the precipitation is taken from the $1/8° \times 1/8°$ observations-based North American Land Data Assimilation System (NLDAS) dataset (Xia et al. 2012), whereas simulation LRP uses the same dataset, but with a twist: the 64 $1/8° \times 1/8°$ values of a given hour's precipitation in a given $1° \times 1°$ cell are averaged and then applied uniformly across those 64 cells. The design of simulations LRP and HRP thus mimics that illustrated in Figs. 10a and 10b—both simulations apply the same precipitation volumes to the land surface during each time step, but the volumes are disaggregated in a realistic way in simulation HRP.

Results are processed separately for each $1° \times 1°$ cell across CONUS, and thus in effect we examine the simplified view of the problem illustrated in Figs. 10c and 10d. For a given $1° \times 1°$ cell, we compute, for each year in 1981–2006, the runoff produced in simulation LRP spatially averaged over the 64 higher-resolution cells contained within it. The resulting time series of 28 annual totals is then regressed against the corresponding time series generated from simulation HRP. The square of the correlation coefficient between the two time series (r^2) indicates the degree to which a large-scale ($1° \times 1°$) runoff estimate is affected by the disaggregation of precipitation. If r^2 is small, disaggregation has an important impact on the estimate, implying that if precipitation forecasts, disaggregation procedures, and modeling approaches are accurate, disaggregation would indeed contribute to the accuracy of large-scale annual streamflow forecasts. If, however, r^2 is close to 1, we cannot expect precipitation disaggregation to contribute to streamflow forecast skill, regardless of how accurate the disaggregation is. The r^2 diagnostic could, in principle, be similarly computed from monthly or seasonal runoff totals.

The second set of parallel simulations (simulations LRP-h and HRP-h, where h represents homogeneous) is identical to the first set except for the homogenization of the land surface properties—both the LRP-h and HRP-h simulations apply the soil, vegetation, and topographic parameters of a representative grid cell in central Kansas to every grid cell in CONUS. The idea behind this second set of simulations is to demonstrate that the most dominant spatial pattern of r^2 determined from simulations LRP and HRP does not simply reflect spatial patterns of land surface properties; it instead reflects large-scale spatial patterns in the precipitation forcing itself.

A note about the nature of the NLDAS precipitation data is appropriate here. The data are based in large part on gauge measurements, and gauge density tends to be lower in the western half of CONUS. While the dataset contains significant variability at scales below 1° even in sparsely gauged regions, one might wonder if the lower

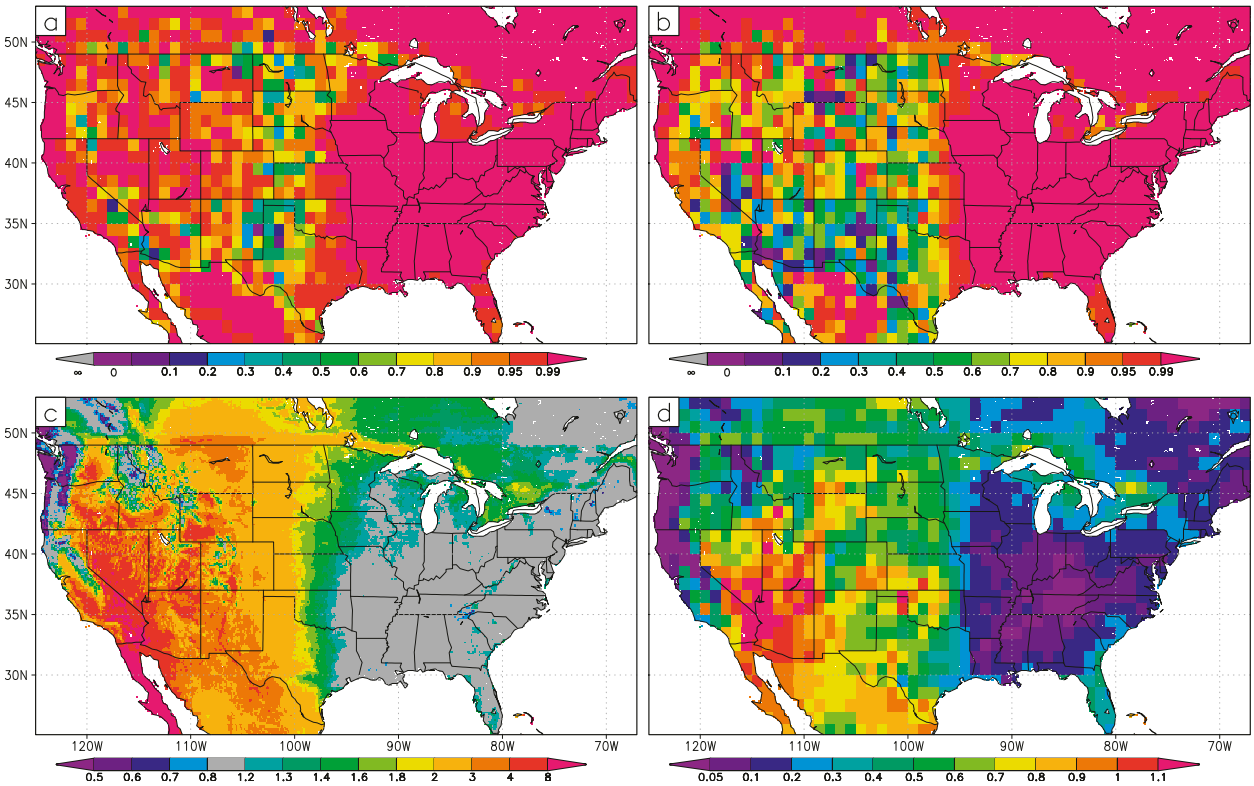

FIG. 11. (a) Correlation between the time series of annual runoffs generated in the experiment with low-resolution precipitation forcing (simulation LRP) and those generated in the simulation with high-resolution precipitation forcing (simulation HRP). (b) The corresponding correlation between simulation LRP-h and simulation HRP-h, which utilized uniform soil, topography, and vegetation characteristics across the United States. (c) Distribution of Budyko's dryness index (the ratio of net radiation to precipitation, made dimensionless with the latent heat of vaporization) across the United States. (d) The ratio of the variance of annual evaporation (from simulation HRP) to the variance of annual precipitation.

gauge density in the west will have an impact on the results there—that is, whether in this experiment the low density in the west will reduce the apparent impact of precipitation disaggregation there. The precise impact of gauge density variations on our results is very difficult to quantify. Nevertheless, as will be shown below, the western half of CONUS is found to be *more* affected by disaggregation than the eastern half despite any disadvantage it has with regard to gauge density. We can safely assume that this distinction, the main result of this experiment, would be retained if precipitation gauge density were uniformly high across CONUS.

b. Results

Results are shown in Fig. 11a. The r^2 values are seen to be very close to 1 in the eastern half of CONUS, implying that the 1° × 1° annual runoffs generated there with the low-resolution (1° × 1°) and higher-resolution (1/8° × 1/8°) precipitation data are essentially the same. (Indeed, the magnitudes of the runoff as well as their time variability are essentially the same in the east; comparisons of the annual runoffs produced by HRP and LRP, not shown here, show only slightly larger values for Simulation HRP, of the order 0.01 mm day^{-1}.) The r^2 values are particularly low along a longitudinal band near the center of CONUS, and they are higher, though still often significantly below 1, in the western third of CONUS. Figure 11b shows in turn the corresponding results from simulations HRP-h and LRP-h, the simulations for which spatial variations in land surface properties do not play a role. The stark contrast in the r^2 values in the western and eastern halves of CONUS is even more apparent in these latter simulations, demonstrating conclusively that the east–west contrast in r^2 is driven by large-scale spatial variations in the meteorological forcing. Note, however, that in the west, the r^2 values produced with Simulations HRP and LRP do differ in many places from those produced with Simulations HRP-h and LRP-h; in these regions, the imposed land surface properties do have an impact on r^2 and thus on the potential usefulness of precipitation disaggregation with this system. Identifying the particular land

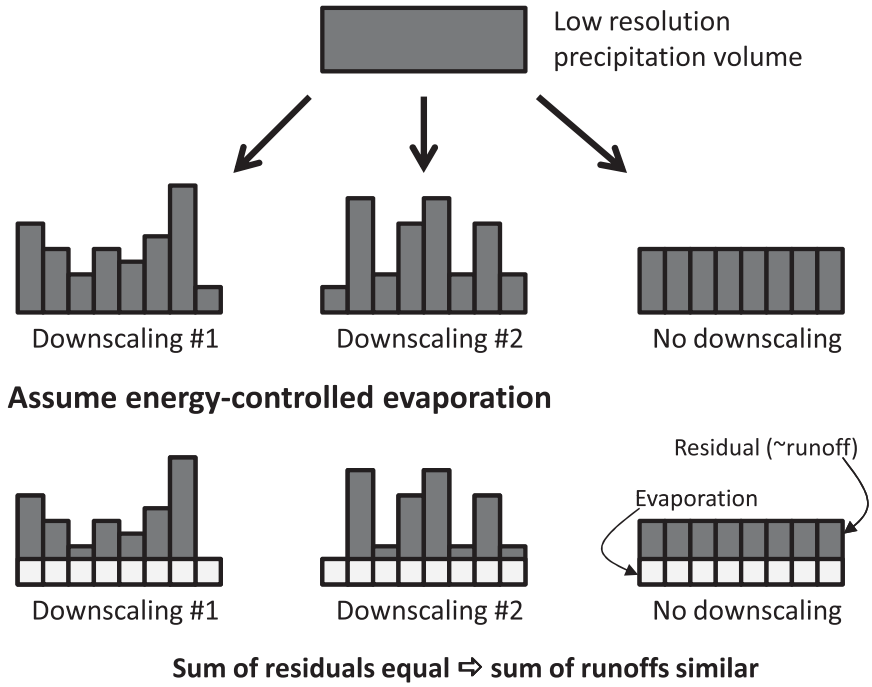

FIG. 12. Illustration of how the runoff generated in an energy-limited evaporation regime may be unaffected by different precipitation disaggregation schemes. See text for details.

surface properties in play here would require additional simulations; we note from the maps, however, that they are not solely associated with topographic variability.

The east–west disparity in r^2, the dominant pattern seen in the maps, has a straightforward explanation. Over long time scales, such as the annual scale considered here, precipitation is roughly balanced by the sum of evaporation and runoff. Evaporation is generally limited by water availability in the west but not in the east, and as a result, for a given precipitation forcing, corresponding impacts are seen for runoff. As argued now, this east–west distinction manifests itself in the r^2 field.

Figure 12 shows how energy-limited evaporation mitigates the ability of precipitation disaggregation to affect the generation of large-scale runoff. The top of the figure shows three different disaggregations of a given, low-resolution precipitation volume to a higher resolution, the third being, in fact, a simple, uniform application of the precipitation. Under the assumption that evaporation is controlled by the availability of energy rather than water at the surface, and under the further assumption that the average energy input across the large-scale area is the same, evaporation is shown in the lower part of the figure to be constant across the high-resolution elements. By construction, then, the sum of the residual water across the high-resolution elements is the same for all three disaggregations. It is this sum of residuals that can contribute to the large-scale runoff, especially when averaged over long time periods, for which changes in storage are insignificant. Because the sum of the residuals is equal, the total runoff is the same regardless of the disaggregation approach (or lack of such an approach) used.

A region with soil moisture–limited evaporation, on the other hand, would not show this behavior. Evaporation in soil moisture–limited regions would vary across the high-resolution elements and from time step to time step (and accordingly from year to year), so that the sum of residuals need not be constant. Different disaggregation approaches (or a lack of disaggregation) would thereby produce different large-scale runoff rates, and the r^2 diagnostic would be less than unity.

Figures 11c and 11d show two measures of evaporation regime. Figure 11c shows Budyko's dryness index (Budyko 1974), defined as the ratio of annual net radiation (converted to water units using the latent heat of vaporization) to annual precipitation. The reflected shortwave and outgoing longwave radiation components of the net radiation were derived from simulation HRP; the remaining radiation components and the precipitation were taken from the prescribed forcing. Clearly seen in this field are the relatively low (roughly unity) dryness indices in the east, implying that the east is not characterized by soil moisture–controlled evaporation rates and thus, according to the argument outlined in

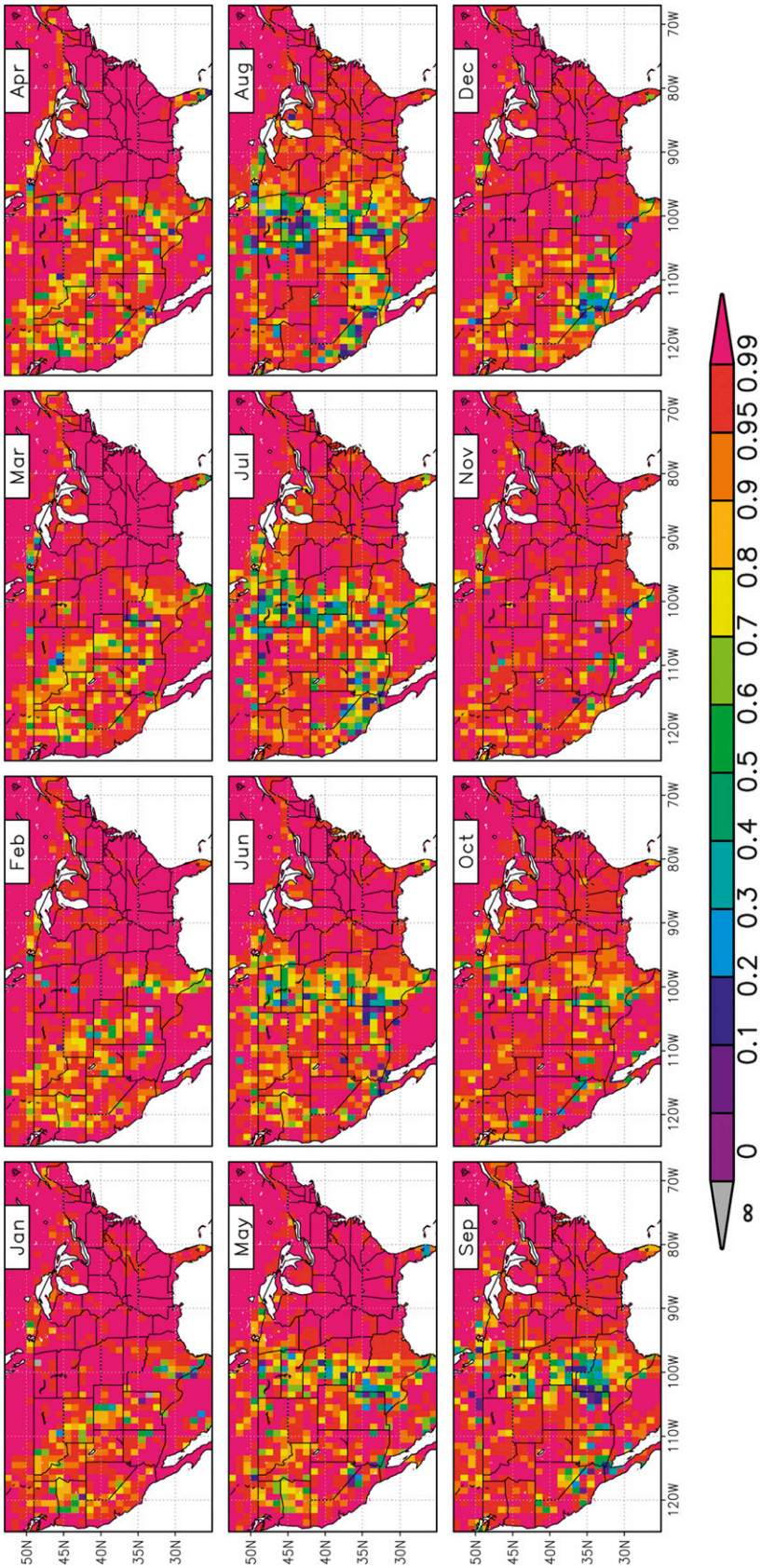

FIG. 13. Correlation between the time series of monthly runoffs generated in the experiment with low-resolution precipitation forcing (simulation LRP) and those generated in the simulation with high resolution precipitation forcing (simulation HRP) for each month of the year.

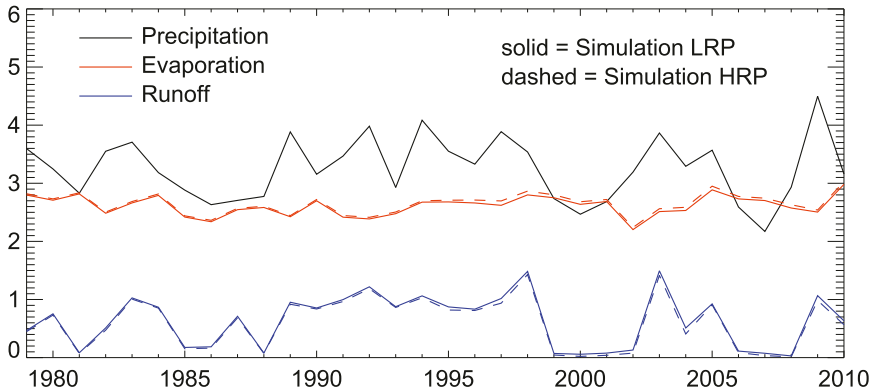

FIG. 14. Time series of annual precipitation (black), evaporation (red), and runoff (blue) at a grid cell in Georgia, as produced in simulation LRP (solid) and simulation HRP (dashed). The y axis has units of mm day^{-1}.

Fig. 12, not as amenable to improved forecasts through precipitation disaggregation. This is underlined further by Fig. 11d, which shows the ratio of the variance of annual evaporation to the variance of annual precipitation, as derived from simulation HRP. In the eastern half of CONUS, the variance ratio is low (and the ratio of mean annual evaporation to mean annual precipitation, not shown, is also relatively low), as expected for dryness index levels characteristic of an energy-limited evaporation regime (Koster and Suarez 1999). The low variance ratios in the east support the idea that the "constant evaporation" assumption used when discussing Fig. 12 is in fact valid there.

Figure 13 provides more evidence that this mechanism is in play and also shows results for a monthly, rather than yearly, averaging period. The 12 panels in Fig. 13 show the r^2 values obtained for simulations LRP and HRP when the runoffs are averaged over each month of the year independently. Clearly seen, especially in the center of CONUS, is the decrease in r^2 during the summer months, when the incident energy is high and evaporation is highest. In the winter months, when the incident energy is lower and evaporation thereby becomes energy limited, the r^2 values are high. The r^2 results for the comparison between simulations LRP-h and HRP-h on the monthly time scale (not shown) are similar, with the lowest r^2 values again appearing in the summer months. We note that the results on this time scale may be of particular relevance to the science of streamflow forecasting, given that monthly and seasonal forecasts are far more common than annual forecasts.

The above discussion brings up a question: could the accurate downscaling of temperature and net radiation (rather than precipitation) in the east, particularly in the summer, have a positive impact on large-scale streamflow forecast skill? Though one might expect that the (nonstatic) spatial variability of temperature or net radiation would be substantially less than that of precipitation, so that the impacts of their downscaling would be reduced, additional analyses (not performed here) would be needed to address this question adequately.

One might also wonder if the noted behavior in the east in Fig. 12 breaks down during drought periods, when water availability rather than energy availability has the potential to limit evaporation. Figure 14 suggests that this is not the case, at least according to this particular land model's simulation of Georgia drought periods in the 2000s. The figure shows, for a representative 1° × 1° grid cell (at 33.5°N, 83.5°W), the simulated time series of annual precipitation (black curve), evaporation (red curves), and runoff (blue curves) averaged across the 64 subgrid cells, with results from simulations LRP and HRP shown as dashed curves and solid curves, respectively. The years 1999–2002 and 2006–08 have particularly low precipitation at this cell, and this low precipitation is manifested almost completely in reduced runoff rates—evaporation rates are not reduced during these periods, and thus the arguments outlined in Fig. 12 suggest that simulations LRP and HRP should continue to show similar runoff estimates during this time. They indeed do, as indicated by the comparisons of the dashed and solid blue curves. Of course, this result might be different for a more severe drought or with a different land surface model. Note, however, that the land surface does have a propensity to convert a precipitation anomaly into a runoff anomaly rather than an evaporation anomaly in wetter regions, as demonstrated, for example, by Koster et al. (2006; their Fig. 7) with a purely observational dataset.

4. Summary and discussion

The two analyses above address some unanswered questions in the science of seasonal streamflow prediction,

a science that is key to addressing societal concerns regarding water supply and drought. In the first analysis, streamflow forecast skill was found, for the most part, to decrease linearly with increasing soil moisture initialization error. (The applied error was assumed to be normally distributed, a simplifying assumption.) This allows the inference of the *increase* in skill that should be achieved with improvements in soil moisture monitoring, as made possible, for example, with the advent of spaceborne L-band soil moisture sensors. In the second analysis, we examined how streamflow forecast skill is connected to the information content of subgrid precipitation distributions, as might be established with a downscaling algorithm applied to the output of global seasonal forecast systems. Here we found that this information content would have little impact on streamflow prediction in the eastern half of CONUS, apparently because of the fact that this region is not characterized by soil moisture–limited evaporation. In contrast, in the western half of CONUS, where evaporation is soil moisture–limited, accurate downscaling applied to accurate low-resolution precipitation forecasts may indeed lead to improved forecast skill.

We should mention, of course, the fact that our results are potentially model dependent. Mahanama et al. (2012) examined streamflow forecasts with four different land-modeling systems and found that the different systems varied in the skill levels they produced; the precise values of $r_{\text{Q-obs}}$ in Figs. 3 and 4 and of the derivatives plotted in Fig. 7 would thus presumably differ if a different land model were used. Also, the values of the subunity correlations plotted in Figs. 11a and 11b might differ if computed with a land model with a different treatment, for example, of topographic impacts on hydrology. We expect, however, that the linearity shown in Figs. 3 and 4 and the impact of evaporative regime illustrated in Fig. 11 are robust results. This robustness can be verified through a repeat of our experiments with alternative models.

We should also emphasize again that the second study does not purport to be a full analysis of precipitation downscaling, which would involve more than disaggregation—it would also involve careful calibration and bias correction. Here we examine only a subset of the downscaling problem: how the information content of high-resolution versus low-resolution precipitation data affects the simulation of large-scale streamflow. Yuan and Wood (2012) found that the main effect of downscaling in the Ohio basin is to correct the errors of the atmospheric model providing the forecasted meteorology; any advantage from precipitation disaggregation there is likely not as important, a result consistent with our findings in Fig. 11 (X. Yuan, personal communication 2012).

While the two studies discussed here are largely independent, they are presented here in a single manuscript because, together, they show that an offline-distributed land-modeling system can be a very useful tool for tackling basic questions in the science of streamflow forecasting. The lower bounds of soil moisture estimation skill provided in Fig. 9 illustrate yet another example of the usefulness of such systems. We expect that these systems have great untapped potential as test beds for basic hydrological research.

Acknowledgments. The authors thank Lifeng Luo and Xing Yuan for their helpful comments on the precipitation disaggregation study. Providers of basin streamflow measurements include the U.S. Army Corps of Engineers (Omaha and Tulsa offices), the Columbia River Basin Climate Change Scenarios Database, the California Data Exchange Commission, and the U.S. Bureau of Reclamation; Edwin Maurer of Santa Clara University provided additional streamflow data. This work was supported by the Modeling, Analysis, Predictions, and Projections Program of the NOAA/Climate Program Office, the NASA Terrestrial Hydrology Program, and the SMAP Science Definition Team.

REFERENCES

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, doi:10.1175/JHM450.1.

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, doi:10.1175/2009JHM1034.1.

Boone, A., F. Habets, J. Noilhan, and Coauthors, 2004: The Rhone-aggregation land surface scheme intercomparison project: An overview. *J. Climate,* **17,** 187–208, doi:10.1175/1520-0442(2004)017<0187:TRLSSI>2.0.CO;2.

Bowling, L. C., D. P. Lettenmaier, B. Nijssen, 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, doi:10.1016/S0921-8181(03)00003-1.

Budyko, M. I., 1974: *Climate and Life.* Academic Press, 508 pp.

Day, G. N., 1985: Extended streamflow forecasting using NWSRFS. *J. Water Resour. Plann. Manage.,* **111,** 157–170, doi:10.1061/(ASCE)0733-9496(1985)111:2(157).

Draper, C. S., R. H. Reichle, G. J. M. De Lannoy, and Q. Liu, 2012: Assimilation of passive and active microwave soil moisture retrievals. *Geophys. Res. Lett.,* **39,** L04401, doi:10.1029/2011GL050655.

Ducharne, A., R. D. Koster, M. J. Suarez, 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.

Entekhabi, D., R. H. Reichle, R. D. Koster, and W. T. Crow, 2010a: Performance metrics for soil moisture retrievals and applications requirements. *J. Hydrometeor.,* **11,** 832–840, doi:10.1175/2010JHM1223.1.

——, and Coauthors, 2010b: The Soil Moisture Active Passive (SMAP) mission. *Proc. IEEE,* **98,** 704–716, doi:10.1109/JPROC.2010.2043918.

Garen, D. C., 1992: Improved techniques in regression-based streamflow volume forecasting. *J. Water Resour. Plann. Manage.,* **118,** 654–670, doi:10.1061/(ASCE)0733-9496(1992)118:6(654).

Hamlet, A. F., and D. P. Lettenmaier, 1999: Effects of climate change on hydrology and water resources in the Columbia River basin. *J. Amer. Water Resour. Assoc.,* **35,** 1597–1623, doi:10.1111/j.1752-1688.1999.tb04240.x.

——, 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, doi:10.1061/(ASCE)0733-9496(2002)128:2(91).

Kerr, Y. H., and Coauthors, 2010: The SMOS mission: New tool for monitoring key elements of the global water cycle. *Proc. IEEE,* **98,** 666–687, doi:10.1109/JPROC.2010.2043032.

Koster, R. D., and M. J. Suarez, 1999: A simple framework for examining the interannual variability of land surface moisture fluxes. *J. Climate,* **12,** 1911–1917, doi:10.1175/1520-0442(1999)012<1911:ASFFET>2.0.CO;2.

——, and S. P. P. Mahanama, 2012: Land surface controls on hydroclimatic means and variability. *J. Hydrometeor.,* **13,** 1604–1620, doi:10.1175/JHM-D-12-050.1.

——, 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,** 24 809–24 822.

——, B. M. Fekete, G. J. Huffman, and P. W. Stackhouse Jr., 2006: Revisiting a hydrological analysis framework with International Satellite Land Surface Climatology Project Initiative 2 rainfall, net radiation, and runoff fields. *J. Geophys. Res.,* **111,** D22S05, doi:10.1029/2006JD007182.

——, S. Mahanama, B. Livneh, D. Lettenmaier, and R. 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.

Liu, Q., and Coauthors, 2011: The contributions of precipitation and soil moisture observations to the skill of soil moisture estimates in a land data assimilation system. *J. Hydrometeor.,* **12,** 750–765, doi:10.1175/JHM-D-10-05000.1.

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

——, ——, and M. Pan, 2007: Bayesian merging of multiple climate model forecasts for seasonal hydrological predictions. *J. Geophys. Res.,* **112,** D10102, doi:10.1029/2006JD007655.

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 prediction A case study in Sri Lanka. *Adv. Water Resour.,* **31,** 1333–1343, doi:10.1016/j.advwatres.2008.06.004.

——, B. Livneh, R. D. Koster, D. Lettenmaier, and R. Reichle, 2012: Soil moisture, snow, and seasonal streamflow forecasts in the United States. *J. Hydrometeor.,* **13,** 189–203, doi:10.1175/JHM-D-11-046.1.

Pagano, T., D. Garen, and S. Sorooshian, 2004: Evaluation of official western U.S. seasonal water supply outlooks, 1922–2002. *J. Hydrometeor.,* **5,** 896–909, doi:10.1175/1525-7541(2004)005<0896:EOOWUS>2.0.CO;2.

Reichle, R. H., and R. D. Koster, 2003: Assessing the impact of horizontal error correlations in background fields on soil moisture estimation. *J. Hydrometeor.,* **4,** 1229–1242, doi:10.1175/1525-7541(2003)004<1229:ATIOHE>2.0.CO;2.

——, ——, P. Liu, S. P. P. Mahanama, E. G. Njoku, and M. Owe, 2007: Comparison and assimilation of global soil moisture retrievals from AMSR-E and SMMR. *J. Geophys. Res.,* **112,** D09108, doi:10.1029/2006JD008033.

——, and Coauthors, 2011: Assessment and enhancement of MERRA land surface hydrology estimates. *J. Climate,* **24,** 6322–6338, doi:10.1175/JCLI-D-10-05033.1.

Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate,* **24,** 3624–3648, doi:10.1175/JCLI-D-11-00015.1.

Schaake, J., J. Pailleux, J. Thielen, R. Arritt, T. Hamill, L. Luo, E. Martin, D. McCollor, and F. Pappenberger, 2010: Summary of recommendations of the first workshop on postprocessing and downscaling atmospheric forecasts for hydrologic applications held at Meteo-France, Toulouse, France, 15–18 June 2009. *Atmos. Sci. Lett.,* **11,** 59–63, doi:10.1002/asl.267.

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

Wang, E., Y. Zhang, J. Luo, F. H. S. Chiew, and Q. J. Wang, 2011: Monthly and seasonal streamflow forecasts using rainfall-runoff modeling and historical weather data. *Water Resour. Res.,* **47,** W05516, doi:10.1029/2010WR009922.

Wood, A. W., and D. P. Lettenmaier, 2006: A test bed for new seasonal hydrologic forecasting approaches in the western United States. *Bull. Amer. Meteor. Soc.,* **87,** 1699–1712, doi:10.1175/BAMS-87-12-1699.

——, and ——, 2008: An ensemble approach for attribution of hydrologic prediction uncertainty. *Geophys. Res. Lett.,* **35,** L14401, doi:10.1029/2008GL034648.

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

Xia, Y., M. Ek, H. Wei, and J. Meng, 2012: Comparative analysis of relationships between NLDAS-2 forcings and model outputs. *Hydrol. Processes,* **26,** 467–474, doi:10.1002/hyp.8240.

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, doi:10.1016/S0022-1694(01)00418-8.

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

——, X.-Z. Liang, and E. F. Wood, 2012: WRF ensemble downscaling seasonal forecasts of China winter precipitation during 1982-2008. *Climate Dyn.,* **39,** 2041–2058, doi:10.1007/s00382-011-1241-8.