JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION
Vol. 53, No. 1 AMERICAN WATER RESOURCES ASSOCIATION February 2017

DOES INCLUDING SOIL MOISTURE OBSERVATIONS IMPROVE OPERATIONAL STREAMFLOW FORECASTS IN SNOW-DOMINATED WATERSHEDS?[1]

Adrian A. Harpold, Kent Sutcliffe, Jordan Clayton, Angus Goodbody, and Shareily Vazquez[2]

ABSTRACT: Changing climate and growing water demand are increasing the need for robust streamflow forecasts. Historically, operational streamflow forecasts made by the Natural Resources Conservation Service have relied on precipitation and snow water equivalent observations from Snow Telemetry (SNOTEL) sites. We investigate whether also including SNOTEL soil moisture observations improve April-July streamflow volume forecast accuracy at 0, 1, 2, and 3-month lead times at 12 watersheds in Utah and California. We found statistically significant improvement in 0 and 3-month lead time accuracy in 8 of 12 watersheds and 10 of 12 watersheds for 1 and 2-month lead times. Surprisingly, these improvements were insensitive to soil moisture metrics derived from soil physical properties. Forecasts were made with volumetric water content (VWC) averaged from October 1 to the forecast date. By including VWC at the 0-month lead time the forecasts explained 7.3% more variability and increased the streamflow volume accuracy by 8.4% on average compared to standard forecasts that already explained an average 77% of the variability. At 1 to 3-month lead times, the inclusion of soil moisture explained 12.3-26.3% more variability than the standard forecast on average. Our findings indicate including soil moisture observations increased statistical streamflow forecast accuracy and thus, could potentially improve water supply reliability in regions affected by changing snowpacks.

(KEY TERMS: snow hydrology; streamflow; surface water hydrology; soil moisture; monitoring; statistics; water supply.)

Harpold, Adrian A., Kent Sutcliffe, Jordan Clayton, Angus Goodbody, and Shareily Vazquez, 2017. Does Including Soil Moisture Observations Improve Operational Streamflow Forecasts in Snow-Dominated Watersheds? *Journal of the American Water Resources Association* (JAWRA) 53(1):179-196. DOI: 10.1111/1752-1688.12490

INTRODUCTION

Accurate seasonal streamflow prediction is critical for reservoir operations, drought mitigation efforts, endangered species protection, irrigated agriculture, and power generation (Hamlet and Huppert, 2002). While seasonal streamflow forecasts have seen steady improvement over the last several decades (Pagano *et al.*, 2004), regional warming and altered hydrological processes are increasingly challenging current forecast techniques (Georgakakos *et al.*, 1998). Operational streamflow forecasts in the Western United States (U.S.) rely on three major contributors to skill: meteorological anomalies, snowpack, and soil moisture status. The ability to forecast seasonal precipitation and temperature anomalies is weak and thus, they are typically not employed in operational streamflow

[1]Paper No. JAWRA-16-0059-P of the *Journal of the American Water Resources Association* (JAWRA). Received February 23, 2016; accepted October 24, 2016. **Discussions are open until six months from issue publication**.

[2]Assistant Professor (Harpold), Department of Natural Resources and Environmental Science, University of Nevada, 1664 N. Virginia Street, Reno, Nevada 89557; Soil Scientist (Sutcliffe) and Hydrologist (Clayton), Utah Snow Survey, Natural Resources Conservation Service (NRCS), Salt Lake City, Utah 84116; Hydrologist (Goodbody), National Water and Climate Center, NRCS, Portland, Oregon 97232; and Undergraduate Student (Vazquez), Universidad Metropolitana, San Juan, Puerto Rico 00901 (E-Mail/Harpold: aharpold@cabnr.unr.edu).

forecasting. In contrast, snowpack is measurable with some certainty, and is typically the primary predictor of streamflow volumes across much of the Western U.S. (Pagano *et al.*, 2004). However, altered snowmelt and hydrologic variability from regional warming threatens common hydrological and snowmelt simulation models and the stationarity of statistical models (Pagano and Garen, 2005; Miller *et al.*, 2011; Raleigh and Clark, 2014).

An important and often underutilized source of forecast skill is soil moisture. Soil moisture has a mechanistic basis in streamflow generation: lower soil moisture increases the potential to store precipitation or snowmelt within the soil profile, whereas higher soil moisture can more easily exceed field capacity causing gravity drainage (McNamara *et al.*, 2005; Flint *et al.*, 2008; Seyfried *et al.*, 2009; Clayton, 2016) and controls runoff by promoting hydrologic connectivity between upland and lowland areas (Western *et al.*, 1999). Consequently, effort is being made to utilize new sources of soil moisture information to improve streamflow forecast skill.

Most applications that have included soil moisture information in streamflow forecasts have focused on model-derived soil moisture information rather than employing direct observations. For example, Maurer and Lettenmaier (2004) used a multiple regression approach applying soil moisture from a land surface model to show that soil moisture information dominates streamflow prediction skill in the Mississippi River Basin at 1 to 6-month lead times. Similarly, output from several land surface models has demonstrated that accurate soil moisture initialization can improve streamflow forecast skill in Western U.S. basins (Koster *et al.*, 2010), with the majority of skill provided in the summer and fall. Koster *et al.* (2010) showed that effective initialization of January 1 soil moisture increased streamflow forecast accuracy over forecasts based solely on snowpack by 1-26% depending on the basin. Limited efforts have been undertaken to include modeled soil moisture information in statistical forecasts (*e.g.*, Berg and Mulroy, 2006; Bellingham and Lea, 2014). The lack of inclusion of soil moisture in statistical forecasts is unfortunate because more complex simulation models require substantial experience and computational resources that often do not produce substantial improvements in accuracy (Franz *et al.*, 2003; Pagano *et al.*, 2004). Moreover, simulation models used by the National Weather Service River Forecast Centers (RFC) are lumped models (*i.e.*, the Sacramento Soil Moisture Accounting model) that cannot easily incorporate soil moisture observations due to the model structure and the lack of correspondence between simulated and observed soil moisture values (see http://www.cbrfc.noaa.gov/wsup/sac_sm/sac_sm.php, Accessed June 25, 2016). These limitations in simulation model forecasts are particularly problematic in small watersheds where coarser model output and meteorological forcing information increase uncertainty. Consequently, an opportunity exists to improve streamflow forecasts using soil moisture observations that are more commensurate with the computational complexity of operational statistical models.

Expanding soil moisture observations from *in situ* networks and remote sensing platforms have seldom been integrated into streamflow forecasts. Several remote sensing platforms capable of measuring soil moisture in the top few centimeters are currently operational at a variety of temporal and spatial resolutions: Soil Moisture and Ocean Salinity (SMOS) collecting data every 3 days at 50 km resolution (Kerr *et al.*, 2001) and the Soil Moisture Active Passive (SMAP) collecting data daily at 36 km resolution (Entekhabi *et al.*, 2008). Although remotely sensed soil moisture observations have shown utility for improving runoff predictability (Jacobs and Myers, 2003; Scipal *et al.*, 2005), difficulties arise in the collection of soil moisture in complex, snow-covered topography with forest cover, which is typical of many critical water resource areas in the Western U.S. Expansion of *in situ* soil moisture observations offers an alternative to remote sensing that may better match the data and model complexity of statistical forecasts. New measurement techniques that rely on global positioning information, GPS (Larson *et al.*, 2008) or cosmic nuclides, COsmic-ray Soil Moisture Observing System (COSMOS) (Hunt *et al.*, 2009) have led to a rapid expansion of soil moisture observation networks. Further, standard soil moisture observation techniques, such as dielectric permittivity and time domain reflectometry, have been increasing in spatial coverage as instrumentation costs have decreased. The expansion of soil moisture networks offers a unique opportunity to assess their potential utility in operational streamflow forecasts.

This study focuses on statistical operational streamflow forecasts generated by the Natural Resources Conservation Service (NRCS) to ask the question, "Does including soil moisture observations improve operational streamflow forecasts in snow-dominated watersheds?" The NRCS forecasts are particularly apt for answering this question because of a recent expansion of soil moisture observation by the Snow Telemetry (SNOTEL) network over the last 10 years (Harpold and Molotch, 2015). Further, the NRCS forecasts focus on smaller, mountain watersheds where simulation models and remote sensing soil moisture observations are less reliable. Here, we develop a new method that can incorporate soil moisture observations into Principal Component Regression (PCR) techniques that are used by NRCS forecasters. The study addresses three specific research objectives: (1) quantify the forecast accuracy improvement provided by soil moisture metrics with and without soil property information; (2) determine whether forecast improvements are more valuable across different forecast lead times; and (3) identify the hydrologic and physiographic conditions where forecast accuracy is most likely to be improved. To the best of our knowledge, this study represents one of the first efforts to introduce direct soil moisture observations into statistical streamflow forecasts used by water managers in the Western U.S.

METHODS

Study Areas

Twelve study watersheds were chosen in areas where long-term soil moisture records and soil property information existed at numerous SNOTEL stations (Figure 1) that are currently used in NRCS forecasts (Table 1). At each SNOTEL station daily accumulated precipitation and daily snow water equivalent (SWE) were harvested for the historical period through water year 2014, which resulted in 28 to 34 years (Table 2). Watershed average SWE and precipitation was the mean across all stations used in the forecast with no elevation or area weighting. Daily maximum soil moisture at 5, 20, and 50 cm was collected for the available record of 7 to 12 years (Table 2). Only watersheds with soil moisture available at a minimum of half of the stations with SWE and precipitation observations were used in the analysis. Soil physical property data were available at stations in 6 of the 12 watersheds. All watersheds have U.S. Geological Survey stream gaging station records spanning the historical SNOTEL record (Table 1).

The watersheds used in the analysis span a range of physiographic conditions. The watershed area had a

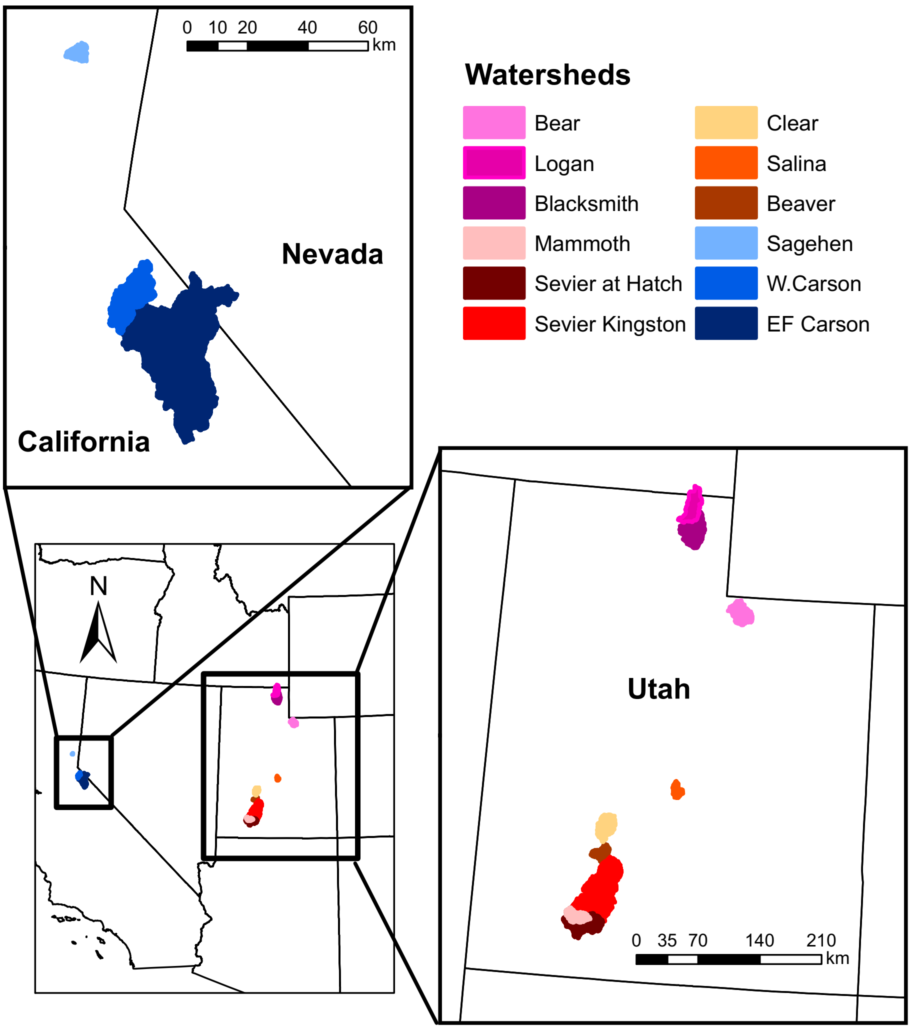

FIGURE 1. The 12 Study Watersheds Used in the Analysis. Nine watersheds were from Utah (bottom right panel) and three watersheds were from California and Nevada (top left panel).

TABLE 1. Watershed USGS Gage Number, Drainage Area, Mean Elevation, Mean Slope, and Percent Forest Cover in the 12 Watersheds.

	USGS Gauge Number	Area (km^2)	Elevation (*m*)	Slope (degrees)	Percent Forest Cover
Bear	10011500	445	2,960	5.7	37
Logan	10109000	236	2,306	6.5	39
Blacksmith	10113500	681	2,178	5.2	34
Mammoth	10173450	272	2,774	3.2	34
Sevier at Hatch	10174500	881	2,505	2.7	31
Sevier at Kingston	10183500	134	2,341	3.9	22
Clear	10194200	425	2,377	6.1	35
Salina	10205030	134	2,669	5.9	23
Beaver	10234500	236	2,801	5.9	46
Sagehen	10343500	27	2,185	4.9	58
EF Carson	10309000	922	2,467	7.4	30
WF Carson	10310000	169	2,325	7.4	29

Note: USGS, U.S. Geological Survey.

TABLE 2. Record Lengths and Number of Stations Used in Each Watershed.

	# w/P and SWE	Start Year	# w/ VWC	Start Year	# w/Soil Properties
Bear	4	1982	4	2007	0
Logan	3	1980	2	2002	0
Blacksmith	3	1982	2	2004	2
Mammoth	5	1981	4	2007	3
Sevier at Hatch	5	1981	4	2007	3
Sevier at Kingston	3	1981	2	2007	2
Clear	4	1986	3	2007	0
Salina	3	1981	2	2007	0
Beaver	2	1981	1	2007	0
Sagehen	3	1981	3	2007	3
EF Carson	5	1981	5	2007	3
WF Carson	3	1980	2	2005	0

Notes: SWE, snow water equivalent; VWC, volumetric water content.
The number of stations used in the standard forecast (# w/P and SWE) and corresponding start year, the number of stations with soil moisture (# w/ VWC) and the start year, and the subset of soil moisture locations with soil properties (# w/ soil properties).

range of 27 to 922 km^2, with the majority between 100 and 450 km^2 (Table 1). The watersheds had mean elevation of 2,178-2,960 m a.s.l. (Table 1). The average slope was more variable across watersheds, ranging from 2.7 to 7.4 degrees (Table 1). The average fraction of forest cover estimated from the 2011 National Land Cover Database (Homer *et al.*, 2015) was between 22 and 58% across the watersheds. We describe the climate during the study period in the Climate Conditions and Hydrological Response section.

Estimating Soil Moisture Metrics

Five soil moisture metrics were derived from a combination of soil water content and soil properties: volumetric water content (VWC), percent saturation (%sat), total storage, available storage, and soil moisture index (SMI). In all cases, soil moisture was measured at SNOTEL sites at 5, 20, and 50 cm depths based on soil dielectric permittivity (Stevens Hydraprobe I and II, Stevens Water Monitoring Systems, Inc., Portland, Oregon), using a standard calibration for all soil types with a measurement uncertainty of 3.4% (Seyfried *et al.*, 2005). Previous quality control and gap filling by the NRCS of precipitation and SWE measurements produced high-quality data that did not require any additional quality control for this study. Conversely, soil water content data required removal of non-realistic values (*i.e.*, > 1 and < 0) and screening and removal of artifacts (*i.e.*, sharp changes in VWC or drift over time) during the first 1-2 years following installation. Missing values were gap filled using cubic convolution spline interpolation in Matlab (Mathworks, 2014). Soil physical properties were downloaded from the National Cooperative Soil Survey (NCSS) Soil Characterization Database (http://ncsslabdatamart.sc.egov.usda.gov, accessed September 2, 2015). All soil analyses were completed by NRCS National Soils Survey Center Kellogg Soil Survey Lab in Lincoln, Nebraska, following standard procedures and reported per physical soil horizon. The specific soil properties used in the analysis were total porosity, water content at 33 kPa (assumed to be field capacity), and water content at 1,500 kPa (assumed to be wilting point) measured at each soil horizon at a given site.

The VWC was depth weighted by assuming the 5 cm soil moisture sensor represented 0-10 cm, the 20 cm represented 10-30 cm, and the 50 cm represented 30-70 cm. The %sat was the ratio of the daily VWC to the maximum daily VWC from the record (*i.e.*, assumed to be saturated) and depth weighted identically to VWC. Total storage was the depth of water stored in the top 70 cm. To estimate total storage each soil horizon was assigned a daily soil moisture value by the soil moisture sensor within that portion of the profile. If no soil moisture sensors intersected that soil horizon, the soil moisture was interpolated between the nearest two horizons or assigned the top or bottom soil moisture sensor if it was above or below the sensors, respectively. The depth of water in each horizon was calculated as the daily maximum VWC multiplied by the total porosity. The total storage was the sum of the depths in each horizon. The available storage was the total porosity minus the total storage, or the air filled porosity. Total porosity was calculated as the sum of porosity in each horizon. The SMI was developed using the method of Hunt *et al.* (2009), that normalizes between −5 at wilting point and +5 at field capacity:

$$\mathrm{SMI} = -5 + 10(\theta - \theta_{\mathrm{WP}})/(\theta_{\mathrm{FC}} - \theta_{\mathrm{WP}}), \qquad (1)$$

where θ was the VWC on any given day, θ_{WP} was the wilting point VWC, and θ_{FC} was the field capacity VWC. Using the methods previously described, an SMI value was assigned to each soil horizon for each day using the maximum daily VWC and subsequently depth weighted.

Standard NRCS Forecasts

We follow a modified version of the standard NRCS PCR forecasts proposed by Garen (1992). The forecasts developed by the NRCS for the 12 watersheds were used to identify the SNOTEL stations of interest for each watershed. However, we did not use the exact forecasts from NRCS because they often include snow course information (4 of 12 watersheds), antecedent streamflow (3 of 12 watersheds), and climate teleconnection indexes (2 of 12 watersheds). Instead, for consistency across watersheds, we modified the NRCS forecasts to only apply water year accumulated precipitation and SWE as predictor variables in the PCR. Following the method of Garen (1992), a PCR was performed, and t-tests evaluated using a p-value of 0.10 to determine which components were significantly different from zero. In all cases, the sign of the regression coefficients was positive (*i.e.*, streamflow increased with increasing precipitation and SWE) and the components were retained. This method resulted in only the first principal component being retained in all cases. It should be noted that p values of < 0.10 were investigated, but resulted in no principal components being retained in some watersheds and thus, the p-value was increased to 0.10. A custom Matlab (Mathworks, 2014) script was used for PCR analysis.

The standard 4-month (April-July) forecast was evaluated and the PCR was developed over the historical record beginning between water year 1980 and 1986 and going through water year 2014 (Table 2). Accuracy metrics were developed for both the historical period and for the period overlapping the soil moisture record. The PCR was run separately for an April-July streamflow volume forecast using four lead times: 0-month lead time (April 1), 1 month (March 1), 2 months (February 1), and 3 months (January 1). These forecasts represent the key water resource planning period in these snow-dominated watersheds and match standard NRCS forecast issue dates.

Two-Step PCR Forecast

The application of a two-step PCR was necessary to address intercorrelation of predictor variables and recognize first-order and second-order controls on streamflow response.

A one-step PCR, similar to that applied in the standard forecast method (see Standard NRCS Forecasts section), was deemed insufficient because it could often result in negative regression coefficients between soil moisture predictor variables and streamflow and thus, those soil moisture predictor variables would not have been retained in the PCR. Further, in a single PCR method, little weight was accorded to soil moisture predictor variables. This effect was not surprising given that P and SWE are clearly first order controls on streamflow, whereas soil moisture is a secondary control (*e.g.*, wet soil moisture conditions have little effect on streamflow during low precipitation years). Consequently, we developed a two-step PCR method that (1) used the standard PCR method based on P and SWE and (2) used a second PCR to explain residuals between predicted and observed streamflow volumes from the standard PCR. The second PCR step used the same two rules for selecting principal components as the standard PCR method (noting that it used soil moisture or antecedent streamflow and the residuals of the first PCR rather than streamflow volume and SWE and P): (1) p value of t-test < 0.10 and (2) regression coefficients were positive. The two-step PCR had several advantages, including reducing the effects of intercorrelation of soil moisture predictor variables, isolating the variability explained by soil moisture, and only allowing for improvements in forecast accuracy from the standard method.

Completing the two-step PCR forecasting method required four tasks. First, the standard forecast was developed from the historical record as described in the Standard NRCS Forecasts section. Second, residuals between observed and predicted streamflow volumes were estimated. Third, the second PCR was applied using soil moisture metrics or antecedent streamflow to predict the residuals between observed and predicted streamflow volumes from the first PCR. Fourth, improvements in forecast accuracy were assessed.

The two-step PCR was applied to investigate whether including antecedent streamflow improved 0-month lead time forecasts and whether including soil moisture metrics improved 0, 1, 2, and 3-month lead time forecasts. Antecedent streamflow is sometimes applied in NRCS forecasts to capture the effects of antecedent wetness conditions (*i.e.*, 3 of 12 watersheds investigated in this study used antecedent streamflow in their NRCS forecasts). Antecedent streamflow was calculated for two periods that capture antecedent wetness conditions that might impact snowmelt runoff: the previous summer streamflow volume April 1-July 31, Q_s and the fall volume from September 1-October 31, Q_f. Antecedent streamflow was evaluated, using Q_s and Q_f for forecasts at all 12 watersheds. Soil moisture metrics based on soil properties (*i.e.*, total storage, available storage, and SMI) were investigated in the six watersheds where soil properties were available. Soil moisture metrics based on measured VWC were calculated in all 12 watersheds.

Forecast Accuracy Metrics

Three forecast accuracy metrics were calculated for both standard forecasts and in the two-step PCR for all forecast lead times: coefficient of determination (R^2), the root mean square error (RMSE), and the percent difference in volume (%dV). The R^2 is often utilized in streamflow forecasting to describe the amount of variability in streamflow volumes explained. The RMSE is an oft used measure of accuracy that is sensitive to very large (*e.g.*, high streamflow) values. The %dV is the percent difference between the observed and predicted streamflow volumes and thus, is insensitive to streamflow magnitude. The three accuracy metrics are used to assess differences between forecasts. Correlation between different variables were assessed, using a coefficient of determination with *p*-values computed based on an F-test.

RESULTS

Climate Conditions and Hydrological Response

The 12 watersheds investigated represent a variety of climatic and hydrological conditions. During the recent record coincident with soil moisture observations (Table 2) the average watershed precipitation (P) from October 1-March 31 (October-March) across all years was 540-871 mm, with generally higher P in the California watersheds (Table 3). The average April 1 SWE was more variable, ranging from 273 to 649 mm across the watersheds. Both SWE and P showed substantial inter-annual variability. For example, the 75th percentile of SWE inputs was 54-272% greater than the 25th percentile among watersheds (Table 3). The total streamflow (Q) from April 1 to July 31 (April-July) varied substantially among watersheds, ranging from 23 to 829 mm. The runoff efficiency (Q/P) varied from 0.05 to 1.15 (values > 1 can occur from spring precipitation not included in the forecasts or the release of water previously stored in the catchment as groundwater or soil moisture).

TABLE 3. Climate and Hydrological Variability in the 12 Watersheds Investigated during the Period with Measured Soil Moisture (see Table 2).

	P (mm)	SWE (mm)	Q (mm)	VWC (%)
Bear	653 (523, 761)	420 (307, 527)	748 (522, 796)	11.3 (9.9, 12.6)
Logan	878 (706, 998)	649 (456, 798)	544 (381, 585)	13.4 (11.8, 15.3)
Blacksmith	722 (593, 839)	514 (363, 647)	141 (74, 176)	9.5 (7.8, 11.3)
Mammoth	622 (465, 716)	389 (207, 516)	240 (118, 297)	11 (9.7, 12.5)
Sevier at Hatch	622 (465, 716)	389 (207, 516)	128 (61, 145)	11 (9.7, 12.5)
Sevier at Kingston	540 (396, 632)	273 (108, 401)	25 (6, 26)	10.1 (8.9, 11.4)
Clear	635 (547, 687)	413 (313, 507)	101 (59, 106)	8.9 (7.6, 10.4)
Salina	602 (517, 669)	381 (295, 455)	23 (16, 27)	14.9 (13.4, 16.7)
Beaver	575 (467, 630)	389 (303, 484)	215 (141, 245)	6.7 (4.7, 8.8)
Sagehen	874 (642, 1,030)	563 (346, 691)	523 (182, 749)	5.4 (3.9, 7)
EF Carson	740 (497, 871)	427 (233, 539)	536 (273, 771)	6.5 (4.8, 7.3)
WF Carson	831 (568, 1,117)	581 (322, 901)	829 (454, 1,232)	5.9 (4.8, 6.8)

Note: The mean values of October 1-March 31 total precipitation (P), April 1 SWE (SWE), April 1-July 31 total streamflow (Q), and October 1-March 31 VWC (VWC) are reported for the stations used in the forecast. The corresponding numbers in parentheses represent the 25th and 75th percentiles.

The average volumetric soil water content from October-March (VWC) varied from 5.4 to 14.9%, with generally drier soils in the California sites. The 75th percentile of VWC was 26 to 88% greater than the 25th percentile among watersheds. The presence of inter-annual variations in VWC suggests potential utility for streamflow forecasting.

The runoff efficiency showed a positive relationship with increasing soil moisture across all watersheds (Figure 2). For example, the Z-score of the ratio of October 1 to March 31 Q divided by P *vs*. the Z-score of October-March average VWC had a statistically significant positive relationship ($R^2 = 0.23$, $p < 0.001$; Figure 2a). The relationship indicated that streamflow was more efficiently generated during wetter soil conditions and higher SWE amounts. Similarly, the Z-score of the ratio of Q divided by April 1 SWE ratio also had a statistically significant positive relationship with increasing VWC ($R^2 = 0.12$, $p < 0.001$; Figure 2b). These relationships indicate that including VWC information has the potential to improve estimates of streamflow.

Standard Forecast Accuracy

Streamflow forecasts were made using standard NRCS methods over both the entire historical record and the recent period when soil moisture was collected (Figure 3). Over the historical period, the standard forecast had an R^2 that varied from 0.32 to 0.91 (average of 0.75). The RMSE varied from 16 to 109 mm (average of 87 mm). The percent deviation in streamflow volume (%dV) varied from 15% to 133% (average of 35%). Over the recent period the standard forecast had an R^2 that varied from 0.36 to 0.92 (median of 0.77) (Table 4). The RMSE varied from 19 to 173 mm (median of 63 mm). The percent deviation in streamflow volume (%dV) varied from 16 to 149% (median of 24%). The small difference in streamflow forecast accuracy between the historical *vs*. recent period (Figure 3) indicated that the meteorological and streamflow generation conditions were similar during the two periods, thus the accuracy analysis from the recent period was indicative of the historical period.

Forecast with Antecedent Streamflow

In addition to the standard forecast using P and SWE observations, we investigated the potential improvement from the addition of antecedent streamflow conditions. We examined two antecedent conditions: the total streamflow volume during the previous summer (April 1-July 31) and during fall (September 1-October 31). We found no improvement in streamflow accuracy metrics from including either the previous summer or fall antecedent streamflow (Figure 4) in the second step of the two-step PCR method (see two-step PCR forecast scection).

Forecast with Soil Moisture Metrics

We investigated whether streamflow forecasts using standard methods were improved by the inclusion of five different soil metrics (see Estimating Soil Moisture Metrics section). The soil moisture metrics were applied in a two-step PCR described using average values from October-March. The soil metrics S, S_{avl}, and SMI required soil physical property information, which required investigating a subset of six watersheds. Across those six watersheds, inclusion of the five different soil metrics resulted in nearly identical 4-month streamflow volume accuracy at a 0-month lead time: median R^2 of 0.95, average RMSE of 37 mm, and %dV from 27% (Figure 5).

Including soil property dependent metrics (*i.e.*, S, S_{avl}, and SMI) did not improve forecast accuracy

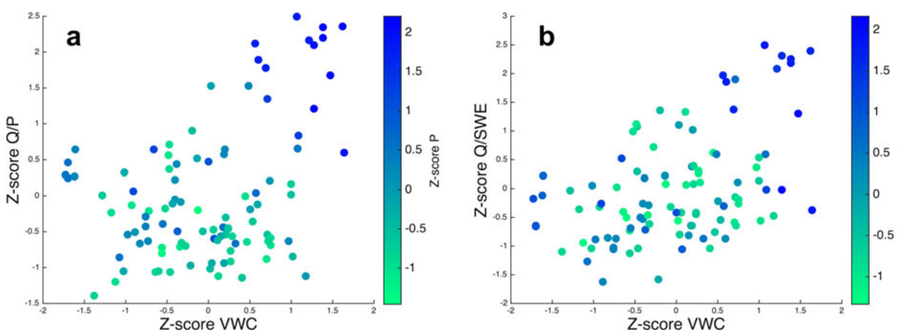

FIGURE 2. Response of the Z-Score of Q/P *vs.* the Z-Score of Average October 1-April 1 Volumetric Water Content (VWC) (a) and Z-Score of Q/Snow Water Equivalent (SWE) Against Z-Score of VWC, with Z-Scores Developed on a Watershed Basis. The symbols are shaded by the Z-score of P (a) and Z-score of SWE (b). The Q/P and Q/SWE runoff efficiency significantly increases with increasing soil moisture ($R^2 = 0.23$, $p < 0.001$ and $R^2 = 0.12$ and $p < 0.01$, respectively).

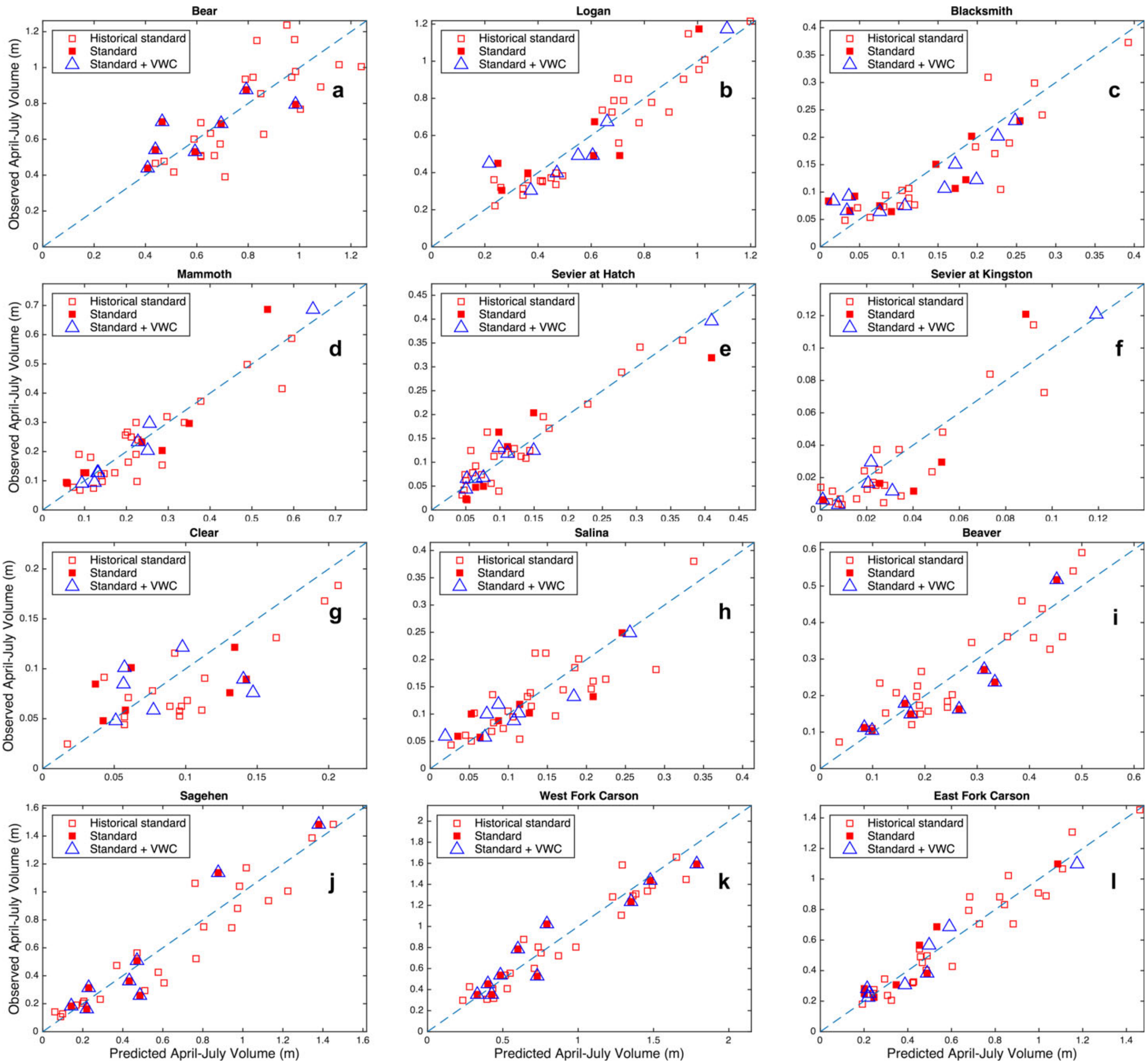

FIGURE 3. Predicted *vs.* Observed 4-Month (April-July) Streamflow Volumes in 12 Watersheds at 0-Month Lead Times. The symbols show the historical standard predictions using Q and P (open squares), recent standard predictions (filled squares), and standard predictions with VWC (open triangles). The letters refer to the 12 different watersheds that are labeled in the title above the panel.

compared to metrics based only on soil moisture (*i.e.*, VWC and %Sat) for reasons we dissect in the Discussion. Consequently, the remaining efforts focus on VWC because it required no soil physical property information and limited assumptions. Moreover, applying VWC in the PCR increased the number of suitable watersheds to 12 (because 6 watersheds did not have available soil property data). Across all 12 watersheds, using the standard forecast at a 0-month lead time with October-March average VWC (Std. + VWC) resulted in median R^2 of 0.77, RMSE of 63 mm, and %dV of 24%, which was a 11.3, 20.6, and 13.5% improvement, respectively, from the standard forecast on average (Table 4).

Effects of VWC Averaging Period

The forecast accuracy improvements were relatively insensitive to the VWC averaging period

TABLE 4. Accuracy Metrics for the Standard 4-Month Forecast with 0-Month Lead Time over the Recent Period Coincident with Soil Moisture Observations.

	Standard April-July Forecast with 0-Month Lead Time			Improvement in 0-Month Lead Time			Improvement in 1-Month Lead Time			Improvement in 2-Month Lead Time			Improvement in 3-Month Lead Time		
	R^2	RMSE (mm)	%dV	R^2	RMSE (mm)	%dV	R^2	RMSE (mm)	%dV	R^2	RMSE (mm)	%dV	R^2	RMSE (mm)	%dV
Bear	0.72	172.99	16	0.000	0.000	0	0.00	0.000	0	0.00	0.000	0	0.00	0.000	0
Logan	0.73	138.92	23	0.098	29.729	4	0.03	14.607	−4	0.10	29.729	4	0.00	0.000	0
Blacksmith	0.72	52.64	35	0.050	4.127	−3	0.15	5.388	−1	0.16	5.750	4	0.15	5.867	6
Mammoth	0.87	66.69	25	0.105	38.353	15	0.27	27.967	5	0.35	28.587	9	0.11	38.019	32
Sevier at Hatch	0.80	48.03	40	0.174	31.241	25	0.38	25.661	12	0.49	25.922	15	0.11	23.137	34
Sevier at Kingston	0.73	18.68	149	0.216	10.110	64	0.47	10.070	46	0.61	10.569	33	0.00	0.000	0
Clear	0.36	46.82	36	0.153	5.717	−2	0.11	4.244	−3	0.24	2.110	0	0.07	6.218	10
Salina	0.77	34.17	23	0.059	5.410	−4	0.20	6.762	8	0.19	8.321	7	0.05	15.108	7
Beaver	0.76	58.42	23	0.000	0.000	0	0.00	0.000	0	0.00	0.000	0	0.00	0.000	0
Sagehen	0.92	131.33	38	0.000	0.000	0	0.18	28.736	−24	0.37	28.527	−35	0.47	28.447	1
EF Carson	0.92	85.75	18	0.020	11.759	2	0.38	42.334	1	0.56	38.014	−3	0.43	22.132	8
WF Carson	0.92	139.56	16	0.000	0.000	0	0.14	18.007	4	0.09	8.950	−3	0.07	6.498	−3
Median	0.77	62.56	24.06	0.050	5.56	0.00	0.16	12.34	0.54	0.22	9.76	1.83	0.07	6.36	3.68
Mean	0.77	82.83	36.74	0.073	11.371	8.40	0.193	15.315	3.68	0.263	15.540	2.40	0.123	12.119	7.88

Note: Improvement in accuracy with the addition of volumetric water content for the 0, 1, 2, and 3-month lead times. Negative improvements indicate a reduction in accuracy.

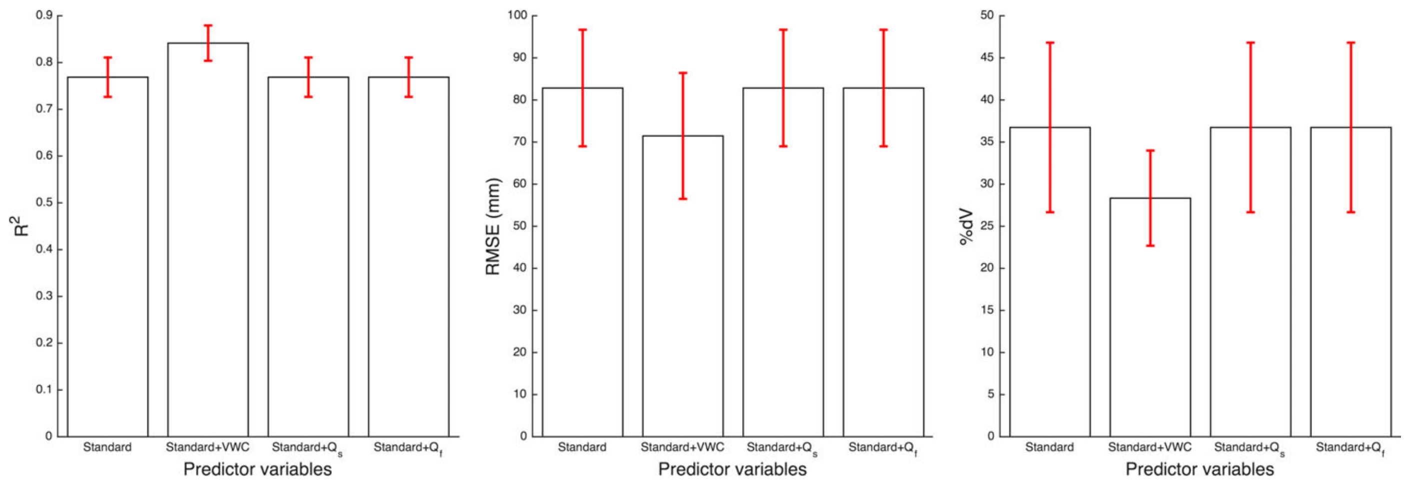

FIGURE 4. Effects of Inclusion of Predictor Variables on the 4-Month Forecast Accuracy of Streamflow Across 12 Watersheds. Four sets of predictor variables were included: standard (Q and P), standard and VWC, standard and summer Q (April-July of previous year), standard and fall Q (September-November of previous year). The accuracy metrics R^2, root mean square error (RMSE), and percent difference in volume are shown on the three different panels. The error bars represent one standard deviation across the 12 watersheds.

overall, but more sensitive within individual watersheds (Figure 6). We compared seven VWC averaging periods in our 0-lead time forecast: day of April 1, March, February-March, January-March, October-March, September-October, and September-November. The R^2 was higher for the October-March period compared to September-November in seven watersheds, lower in two watersheds, and near equal in three watersheds (Figure 6). The median RMSE was relatively insensitive to the averaging period used. The %dV differed substantially at the October-March averaging period, with three watersheds (Mammoth,

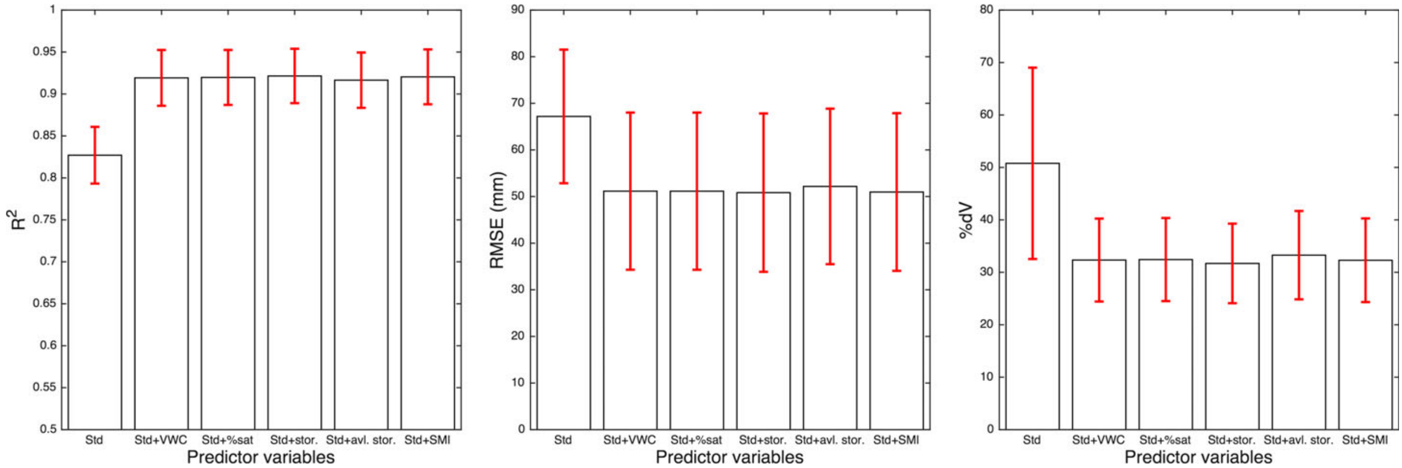

FIGURE 5. Effects of Including Different Soil Water Metrics on the 4-Month Forecast (April-July) Accuracy at 0-Month Lead Times. Five different prediction scenarios were added to the standard (Std) forecast: VWC, percent saturation (%sat), total storage, available storage, and soil moisture index (SMI). The accuracy metrics R^2, RMSE, and percent difference in volume are shown on the three different panels. The error bars represent one standard deviation across the six watersheds.

Sevier at Hatch, and Sevier at Kingston) having substantially higher %dV compared to September-November periods (Figure 6). At the longer lead times the R^2 suggested that water year-scale averaging periods (*i.e.*, October to March 1, February 1, or January 1) were stronger than fall averaging periods. In contrast, %dV showed more mixed differences between water year and fall averaging periods, with roughly half of the watersheds showing greater improvement from fall *vs.* water year periods (Figure 6). We attribute the differences in R^2 *vs.* %dV to improved accuracy at high flows (*i.e.*, R^2) *vs.*

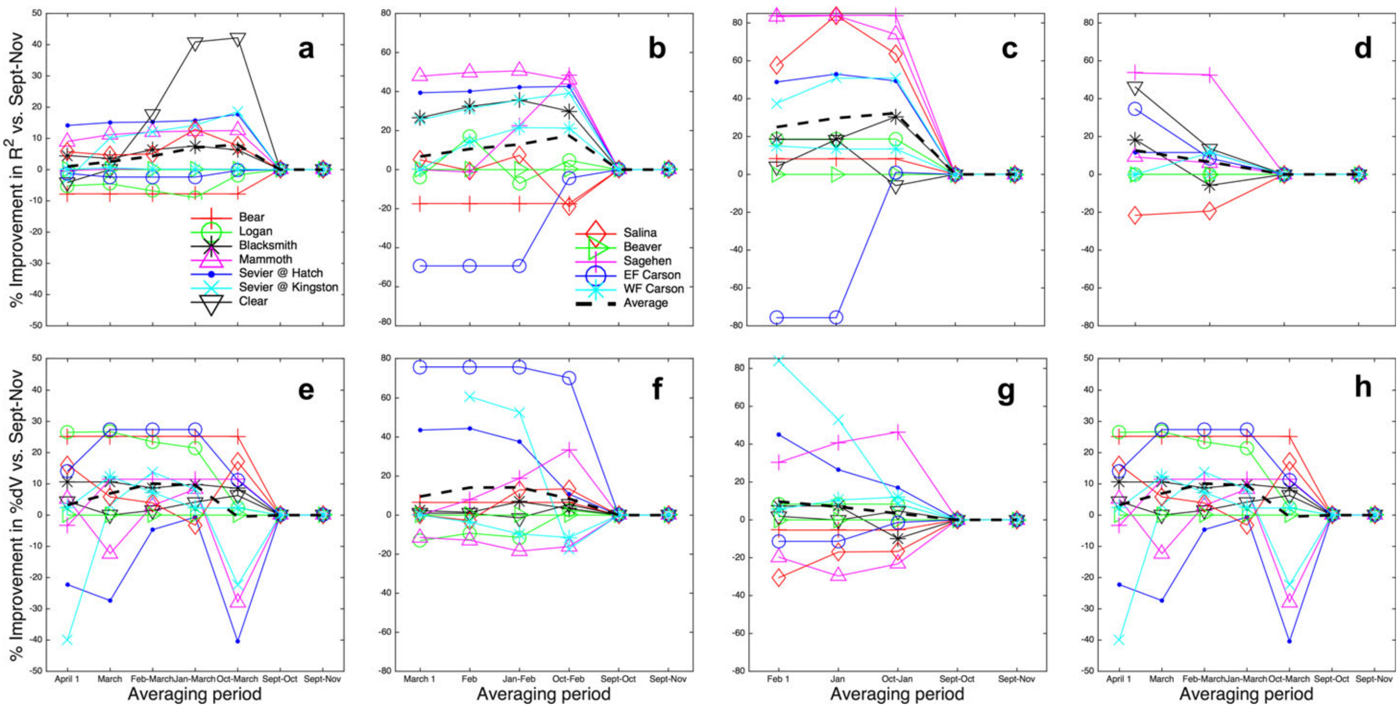

FIGURE 6. Four-Month Forecast Accuracy Using VWC Averaged Over Different Averaging Periods as Compared to the Accuracy from September-November VWC Averaging Periods. The results are shown for each of 12 watersheds at 0-month (left two panels), 1-month (center-left two panels), 2-month (center-right two panels), and 3-month (right two panels) lead times. The thick dashed line is the average of all watersheds.

TABLE 5. The Coefficient of Determination (R^2) between Average VWC over Different Periods Is Shown in the Bottom-Left and Best-Fit Slope between the Columns (dependent variable) and the Rows (independent variables) Is Shown in the Top Right (shaded grey).

	April 1	March	February-March	January-March	October-March	September-October	September-November	October-February	October-January	October-December
April 1	1.00	0.95	0.92	0.90	0.84	0.53	0.64	0.81	0.80	0.77
March	0.91	1.00	0.98	0.97	0.90	0.58	0.69	0.88	0.86	0.83
February-March	0.86	0.98	1.00	0.99	0.93	0.60	0.73	0.92	0.90	0.88
January-March	0.84	0.96	0.99	1.00	0.95	0.61	0.75	0.94	0.93	0.91
October-March	0.74	0.85	0.90	0.93	1.00	0.73	0.89	1.01	1.02	1.02
September-October	0.27	0.32	0.34	0.35	0.48	1.00	0.94	0.68	0.71	0.75
September-November	0.37	0.44	0.48	0.50	0.70	0.86	1.00	0.81	0.84	0.89
October-February	0.68	0.79	0.85	0.89	0.99	0.50	0.73	1.00	1.01	1.02
October-January	0.63	0.74	0.80	0.84	0.98	0.53	0.77	0.99	1.00	1.02
October-December	0.57	0.66	0.72	0.76	0.94	0.56	0.82	0.97	0.99	1.00

Note: Correlation and slope decrease as averaging periods diverge from April 1 VWC values.

improved accuracy at low to medium flows (*i.e.*, %dV). Longer water year averaging periods generally improved forecast accuracy, despite variability across watersheds; thus, the October-March average VWC was used in the subsequent accuracy analyses for consistency.

There was high correlation but variable slope between VWC among different averaging periods. While the correlation indicates the strength of the relationship, the best-fit slope gives information about the relative proportion of VWC between different averaging periods. The VWC between October-March and April 1 alone was strongly correlated ($R^2 = 0.74$, $p < 0.001$), with the correlation and best-fit slope increasing between April 1 and shorter averaging periods (Table 5). The R^2 between April 1 and the water year average remained high, but declined from 0.68, 0.63, to 0.57 for the 1, 2, and 3-month lead times (Table 5). The best-fit slope values suggested that October-March averages were 84% of the VWC value of April 1 on average. Fall VWC values were less correlated to the VWC used in the streamflow forecasts (Table 5). For example, October-March average VWC had R^2 of 0.48 and 0.70 with September-October and September-November averages, respectively (Table 5). The best-fit slope indicated that September-November VWC was 64% of April 1 VWC on average. The R^2 between September-October with water year averages increased from 0.48, 0.50, 0.53, to 0.56 at 0, 1, 2, and 3-month lead times, respectively. At the individual watershed-scale, 7 of 12 watersheds had significant correlation ($p < 0.05$) between September-November and October-March average VWC.

Forecast Accuracy across Different Forecast Lead Times

We evaluated the relative improvement in 4-month forecast accuracy across 0-month (April 1), 1-month (March 1), 2-month (February 1), and 3-month (January 1) lead times, using VWC averaged from October 1 to the first day of the forecast. The median improvement in R^2 from including VWC was 0.05 for 0-month lead times, 0.16 for 1-month lead times, 0.22 for 2-month lead times, and 0.07 for 3-month lead times. The median improvement in RMSE from including VWC was 6 mm for 0-month lead times, 12 mm for 1-month lead times, 10 mm for 2-month lead times, and 6 mm for 3-month lead times. The median improvement in %dV from including VWC was 0% for 0-month lead times, 0.5% for 1-month lead times, 1.8% for 2-month lead times, and 3.7% for 3-month lead times. Including VWC showed improvement in RMSE over standard forecasts in 8 of 12 watersheds for 0-month lead time forecasts, 10 of 12 watersheds for 1 and 2-month lead times, and 8 of 12 watersheds for 3-month lead times (Figure 7). The magnitude of improvement in R^2 was greatest for 0-month lead times in 1 of 12 watersheds, greatest in 1-month lead times in 2 of 12 watersheds, greatest for 2-month lead times in 6 of 12 watersheds, and greatest for 3-month lead times in 1 of 12 watersheds

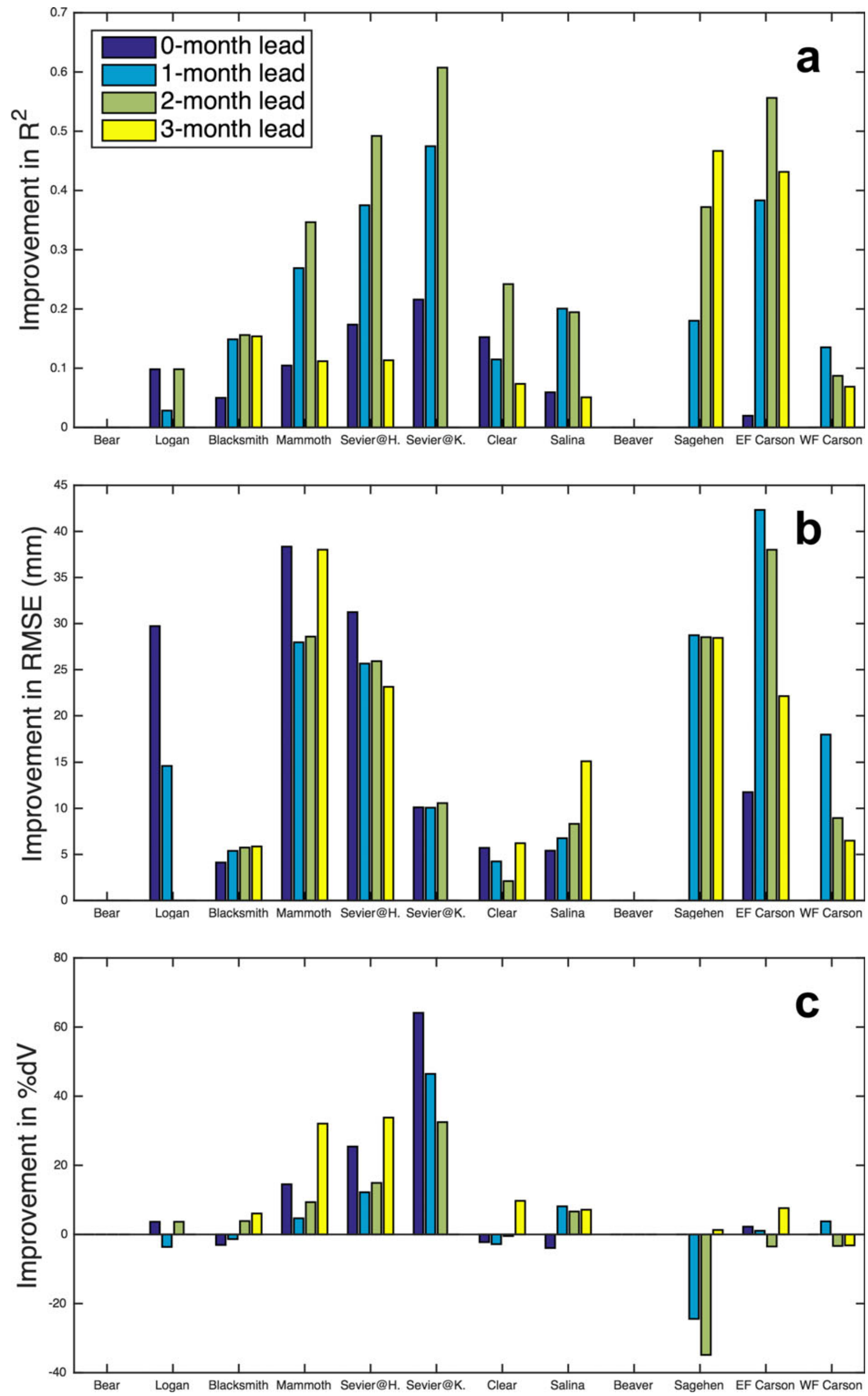

FIGURE 7. Accuracy Improvement by Including VWC for 4-Month (April-July) Forecasts with 0, 1, 2, and 3-Month Lead Times. Improvement is relative to standard forecast. Negative values indicate reduced accuracy.

(Figure 7). Only two watersheds showed no improvement in forecast accuracy across any of the forecast lead times (Bear and Beaver watersheds). These results support the potential to use soil moisture information to improve streamflow forecasts across a variety of lead times.

Forecast Accuracy across Different Hydrological Conditions

The inclusion of VWC into standard streamflow forecasts most consistently improved accuracy during higher Q, with the improvement varying across the different lead times investigated. During the years when the Z-score of Q was above zero (*i.e.*, greater than record mean) the %dV showed improvements of 4.6% for 0-month lead times, 1.3% for 1-month lead times, and 3.6% for 2-month lead times, and −0.4% for 3-month lead times. These forecast improvements during high Q were mainly driven by the three watersheds that had the maximum observed April-July streamflow over the historical record in 2011 (Mammoth, Sevier at Hatch, and Sevier at Kingston watersheds), which showed 15-25% improvement in %dV from including VWC at 0-month lead times. No watersheds had statistically significant relationships between improvement in %dV and Q at any lead time. Improvement in forecast accuracy was generally higher when the watershed average SWE/P ratio was higher, with 5.2, 5.8, 7.9, and 7.8% improvement at 0, 1, 2, and 3-month lead times, respectively, when the Z-score of SWE/P was > 0. Improvement in %dV for 0-month lead time forecasts as a function of VWC were not clear, with generally similar improvement in high and low VWC (Figure 8).

Accuracy Improvement across Watersheds

Watersheds with lower P, more variation in Q/P ratios, and smaller fractions of SWE/P generally saw the greatest forecast improvements from including VWC (Figure 9). At all forecast lead times, the watersheds with the lowest P had the greatest improvement in %dV from including VWC, with R^2 between mean P and improvement in %dV changing 0.20, 0.42, 0.46, and 0.12 for 0, 1, 2, and 3-month lead times, respectively (Figure 9). Watersheds that had greater variation in runoff efficiency, estimated as the difference between the 75th and 25th percentiles of Q/P, had greater improvement in %dV from including VWC; the R^2 was 0.17, 0.34, 0.53, and 0.09 between improvement in %dV and the difference between 75th and 25th percentiles for 0, 1, 2, and 3-month lead times, respectively (Figure 9). Watersheds with lower mean SWE/P fractions (indicating more rainfall or more early season melt) had larger improvements in %dV from including VWC. The improvements in %dV as a function of mean SWE/P generally decreased with longer forecasts, with the R^2 declining from 0.52 to 0.42 to 0.17 to 0.03 for 0, 1, 2, and 3-month lead times, respectively (Figure 9).

DISCUSSION

Our results demonstrate that including soil moisture can improve streamflow volume forecast accuracy over standard 4-month (April-July) statistical streamflow volume forecasts based only on precipitation and SWE across a variety of forecast lead times. The importance of soil moisture was demonstrated by the statistically significant positive relationship between runoff efficiency and average soil VWC from the water year ending on April 1 (Figure 2). Including VWC in the 0-month lead time forecasts explained 11.3% more variability, reduced RMSE by 20.6%, and reduced the %dV by 13.5% on average compared to standard forecasts without VWC (Table 4 and Figure 7). These improvements are substantial given that the standard forecast already explained 77% of the variability in April-July forecasts with 0-month lead times on average (Table 4). The statistical forecast R^2 of 0.84 for 0-month lead times, using soil moisture was comparable to

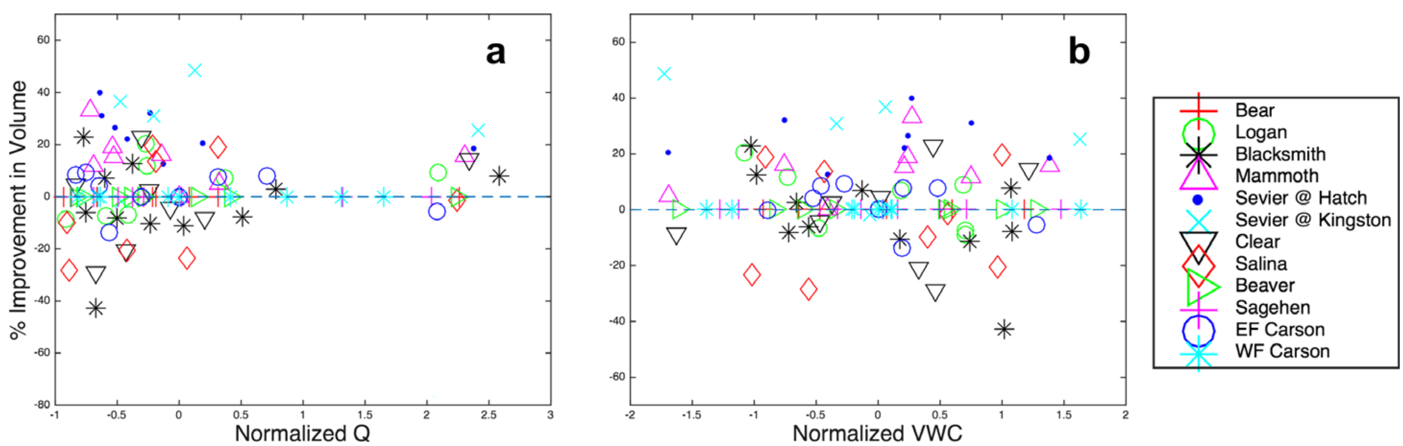

FIGURE 8. Percentage Improvement in Forecast Volume as a Function of the Z-Score of Q and VWC at the 12 Watersheds for 0-Month Lead Time Forecasts.

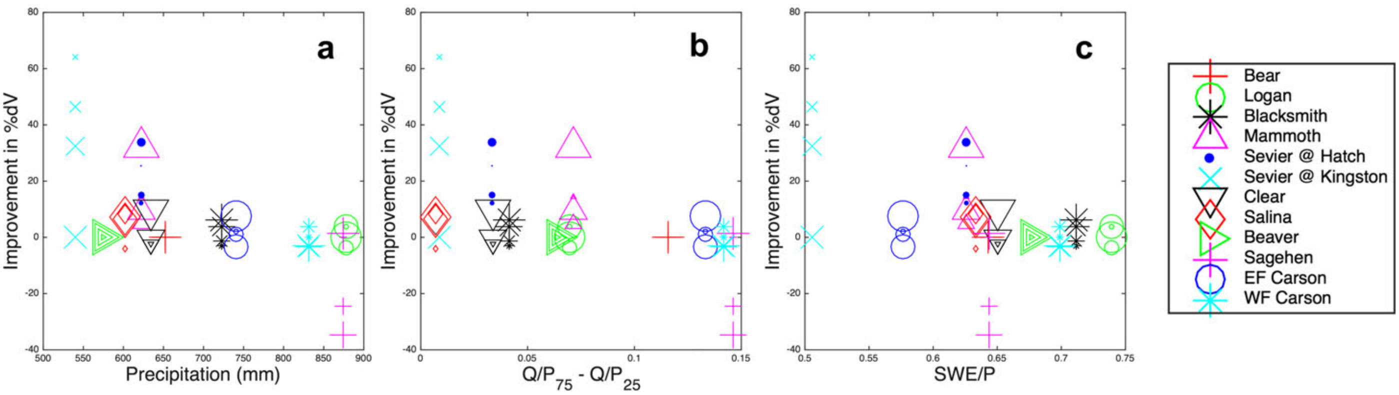

FIGURE 9. Improvement in %dV as a Function of Precipitation, SWE/P Ratio, and the Difference of the 75th and 25th Percentile of Q/P. The symbol type are the 12 watersheds. The symbol sizes increase in size from 0-month to 3-month lead times.

historical verification found for the seven watersheds predicted by the Colorado Basin RFC over the historical period through water year 2015 of 0.73 (Bear = 0.75, Logan = 0.80, Blacksmith = 0.88, Sevier at Hatch = 0.24, Sevier at Kingston = 0.87, Clear = 0.66, and Salina = 0.91; verification data can be found here: http://www.cbrfc.noaa.gov/arc/verif/verif.php, accessed June 25, 2016). At 1, 2, and 3-month lead times the forecasts explained 25.1, 35.2, and 15.0% more variability by including VWC, respectively, than the standard forecast (Table 4). These increases in accuracy have large potential effects on water management; leading to reduced residual between the predicted *vs*. observed streamflow volume of up to ~ 4,327,000 m^3 (~ 3,510 ac-ft) per year on average for a 0-month lead time and ~ 1,555,400 m^3 (~1,260 ac-ft) per year on average for a 1-month lead time. The three watersheds that had their wettest year on record in 2011 (Mammoth, Sevier at Hatch, and Sevier at Kingston watersheds) showed %dV improvements of 15-64%, which was equivalent to ~ 2,142,600-28,747,500 m^3 (~1740 to 23,310 ac-ft) per year, respectively. The large (positive) impact from including soil moisture in streamflow volume prediction in these three watersheds demonstrated the value of soil moisture information in extremely wet years like 2011. The influence of extremely wet years on select sites also explains why the average improvement in R^2 (0.07) was larger than the median improvement in R^2 (0.05) for 0-month lead times. Our results provide the first clear evidence supporting the inclusion of soil moisture information into statistical streamflow forecasts across a variety of lead times.

The improvements in forecast accuracy using soil moisture observations generally exceeded forecast improvements found by other studies using model-derived soil moisture in large, snow-dominated river basins. For example, the fraction of watersheds with forecast improvements in this study was 66% (8 of 12) for 0 and 3-month lead times and 83% (10 of 12) for 1 and 2-month lead times, which was greater than the fraction of statistically significant correlation between forecast residuals and soil moisture found by Berg and Mulroy (2006) of 32% of watersheds at 1-month lead time and 19% at 2-month lead time, using modeled soil moisture information. Similar to our results, Koster *et al.* (2010) showed statistically significant relationships between soil moisture initialization values from land surface models and March-July streamflow volumes in 11 of 17 Western U.S. basins. Direct comparisons of the magnitude of forecast improvement between this study and others using model-derived soil moisture is difficult however, due to differences in forecast methods, lead times, and duration. For example, Koster *et al.* (2010) found that skill derived from soil moisture initialization explained between 1 and 26% of the 5-month streamflow forecasts (March-July), which was similar with 0-22% increases in explained variability in 0-month lead times and less than the 0-47% increases in explained variability at 1-month lead times found in our study. The decline in improved forecast accuracy provided at 3-month lead time found in this study was consistent with the findings of Maurer and Lettenmaier (2004) who showed that gains in spring and summer forecast skill from soil moisture declined rapidly as lead times exceeded three months. These comparisons in forecast improvement to previous studies highlight the novelty of applying direct soil moisture observations, rather than model products, in smaller, mountain watersheds.

The largest improvements in forecast accuracy were in watersheds that had lower precipitation and were less snow-dominated (Figure 9), but it was difficult to substantiate these trends with the limited number of watersheds. Our finding that forecast improvement was greatest in watersheds with lower average SWE/P ratios (Mammoth, Sevier at Hatch, Sevier at Kingston, and Salina watersheds) was consistent with the findings of Berg and Mulroy (2006) and Maurer and Lettenmaier (2004) who suggested that initial soil moisture conditions were less important in wetter, more snow-dominated watersheds. Presumably, the lower sensitivity to soil moisture arises when snow water inputs are well in excess of soil water storage. The same reasoning can be used to explain why drier watersheds had increasing improvement in forecast accuracy from including soil moisture: the initial soil moisture state becomes a larger determinant of runoff generation as total soil water storage becomes a larger fraction of the net water inputs. The ratio of soil water storage to net water inputs also potentially explains why watersheds with higher and more variable runoff efficiency (Q/P) (Salina, Sevier at Hatch, Sevier at Kingston, and Clear watersheds) showed less improvement from including soil moisture. As an example, we might expect runoff efficiency to be higher but less variable as available soil storage approaches zero. Somewhat contradictory to the watershed-scale averages however, was the finding that improvements within watersheds were generally higher when SWE/P ratios and precipitation were above average (Figure 8), which was most dramatic in Sevier at Hatch, Sevier at Kingston, and Mammoth watersheds. We speculate that these drier and lower elevation watersheds were more sensitive to large precipitation years, which are characterized by high SWE/P ratios. Future efforts and larger datasets with greater quality control are needed to test these potential mechanisms and identify forecast locations and lead times that are most likely to benefit from the inclusion of soil moisture information.

The observed influence of soil moisture on streamflow generation is consistent with previous process-based studies that outline the challenges of upscaling limited field observations in both space and time. Perhaps most widely accepted is that higher soil water storage primes a watershed for runoff by reducing available storage and increasing hydrological connectivity (McNamara *et al.*, 2005). For example, higher antecedent wetness increases runoff, particularly when effective precipitation intensity is below the infiltration capacity (Weiler *et al.*, 2003; Brocca *et al.*, 2008). Thus, if the shallow soil moisture profile can effectively capture inter-annual antecedent wetness variability it helps explain the additional improvement in forecast accuracy found here. However, contrary to our initial expectations, soil water storage metrics that considered soil physical properties did not improve streamflow forecasts beyond VWC (Figure 5). The lack of sensitivity to soil storage metrics could perhaps be due to the statistical normalization done by PCR (*i.e.*, the exact magnitudes are unimportant). The lack of sensitivity to the VWC magnitudes, but rather sensitivity to relative changes in VWC, increases the usability of this approach because it does not require expensive soil property analysis and reduces the need for absolute accuracy in VWC measurements. Moreover, the results suggested that spring soils (averaged during February and March, as well as from October 1 to January 1 or February 1) were wetter than fall soils (averaged September 1 to October 1 or November 1) and those differences were sufficient to improve forecast accuracy in many watersheds (Figure 6). These results are somewhat surprising given that water fluxes are relatively low when snow was present (*i.e.*, low evapotranspiration loss and rainfall/snowmelt gains), but may be indicative of early melt pulses in some watersheds. Longer averaging periods may also reduce the bias from soil frost, which acts to decrease soil moisture values measured with dielectric permittivity. These findings may help explain why 1 and 2-month lead times produced the greatest forecast improvement from the addition of soil moisture: these lead times capture both the fall antecedent wetness as well as early snowmelt that might be occurring at lower elevation stations. At 3-month lead times it is less likely that any melt pulses occurred in the water year. At 0-month lead times the fall signal was overwhelmed and the soil moisture only reflects spring snowmelt inputs. While the averaging period over which VWC was calculated had little effect on forecast accuracy overall, individual watersheds showed different sensitivities to fall *vs*. water year averages that were not accounted for here because of the use of a single averaging period in all watersheds. Overall, it was difficult to assess these watershed-level controls on accuracy improvement without additional information at more locations. Importantly, our results are consistent with a large body of literature indicating a limited number of moisture sensors can provide reliable estimates of antecedent conditions (Thierfelder *et al.*, 2003). The use of limited direct observations has numerous advantages over more sophisticated model and remote sensing estimates for operational forecasting, particularly in small mountainous watersheds where model and remote sensing estimates are more uncertain.

There are three key roadblocks to implementing soil moisture information into operational streamflow forecasting that this study addresses (Pagano *et al.*, 2004): (1) increasing resources necessary to run and maintain forecast models; (2) the reliability and length of soil moisture record; and (3) the lack of evidence of improved forecast accuracy from including soil moisture. Due to a previous lack of direct observations of soil moisture, model-derived soil moisture has seen more integration into streamflow forecasts. While simulation models are in use by the National Weather Service RFC (Franz *et al.*, 2003), they have not been used operationally by the NRCS due to their intensive data and human resource requirements, and a lack of evidence of accuracy improvement compared to statistical forecasts (Pagano *et al.*, 2004). The similar R^2 between historical RFC forecasts and our predictions in a subset of seven watersheds supports this conclusion. Our work poses a tractable solution, using a two-step PCR that includes soil moisture information from direct observations at the stations (or subset thereof) already being used in NRCS forecasts. The issues associated with reliability and record length of *in situ* soil moisture observations are being remedied by expanding spatial and temporal coverage of soil moisture networks. The NRCS SNOTEL network is a particularly good example of datasets with sufficient accuracy to be used in research efforts (*e.g.*, Maurer and Bowling, 2014; Harpold and Molotch, 2015) and the publicly available datasets have recently been quality controlled at stations throughout the Western U.S. However, dielectric sensors have their own sets of issues to be considered, such as decreased moisture during soil frost events, potential sensitivity to poor installations, and an operational life expectancy of around 10-15 years. Novel new measurement techniques are also increasing soil moisture coverage, such as GPS (Larson *et al.*, 2008) and COSMOS (Hunt *et al.*, 2009). Similarly, remote sensing of soil moisture (*e.g.*, SMOS and SMAP) and water storage (*e.g.*, Gravity Recovery and Climate Experiment [GRACE]) is advancing rapidly, but has intrinsic limitations associated with directly observing soil moisture during snow cover, difficulties in complex and forest topography, and coarser spatial resolutions that may not be appropriate for forecasting in headwater basins. Consequently, our results address the major roadblocks to inclusion of soil moisture observations into streamflow forecasts.

Regional warming is expected to alter snowpack dynamics and increase precipitation extremes in ways that will heighten the importance of effective streamflow forecasts for water management while challenging current forecasting techniques and datasets. In particular, shifts to different snowmelt and precipitation regimes (Harpold *et al.*, 2012; Klos *et al.*, 2014; Trujillo and Molotch, 2014) could introduce non-stationarity into statistical forecasts based on historical conditions. Soil moisture offers an additional source of information that could account for shifts from snowpack stores to subsurface stores. In addition, capturing extreme precipitation events that are outside of historical ranges is likely to become increasingly important in the future (Lute and Abatzoglou, 2014; O'Gorman, 2014). The accuracy improvement in watersheds where 2011 was the wettest year on record (Figure 3) bodes well for the potential of soil moisture to help address extreme climate years. Addressing the grand challenge of streamflow forecasting during environmental change will require a concerted effort to simultaneously advance *in situ* observations, simulation models, and remote sensing datasets. Advancing statistical forecasts using existing *in situ* soil moisture observations, as proposed and tested here, provides a potential bridge between statistical techniques currently used by forecasters and more sophisticated distributed simulation models and data assimilation techniques in testing and development. Our study supports the building of a bridge to more sophisticated models, but will require additional testing of the forecasting method across a greater range of study watersheds and hydrological conditions.

CONCLUSIONS

The value of soil moisture to streamflow forecast skill has long been recognized but not properly integrated into operational statistical forecasts like those used by the NRCS. In this study, we showed that limited direct soil moisture observations could improve statistical forecast accuracy. Forecasts were able to explain 7.3, 19.3, 26.3, and 12.3% more variability across 0, 1, 2, and 3-month lead times, respectively. The greater forecast accuracy at longer lead times (particularly 1 and 2 months) is an important strength of including soil moisture information into standard forecasts. Surprisingly, improvements in forecast accuracy were relatively insensitive to soil water storage metrics that included soil properties. These insensitivities bode well for expanding the proposed forecasting technique to other watersheds that lack soil property information. While this study began to identify conditions and watershed physiography that was likely to result in the largest accuracy improvements, the small number of watersheds investigated precluded wide-ranging conclusions. More work is needed to fully explore the potential of soil moisture inclusion into streamflow forecasts: (1) to identify the most sensitive forecasts; (2) identify locations most suited to representing the average watershed soil moisture conditions; (3) and identify the optimum averaging periods on a watershed level. Further, improvements to the two-step PCR developed here might be possible by excluding particular years or optimizing the soil moisture locations and averaging periods for particular watersheds. Despite the need to refine and further test the methods proposed here, early indications suggest that rapid integration of these techniques into NRCS forecasts have the potential to positively affect water management decision-making and build forecast resilience to changing snowpack and climate.

ACKNOWLEDGMENTS

We greatly appreciate the input and facilitation of this research by Troy Brosten and Randall Julander at the NRCS Snow Survey in Salt Lake City, Utah. A. Harpold was supported by USDA NIFA (NEV05293) and NASA EPSCoR (NNX14AN24A) grants. S. Vazquez was supported by an NSF REU summer fellowship (SBE #1263352).

LITERATURE CITED

Bellingham, K. and J. Lea, 2014. Improved Water Supply Forecasts Using Soil Moisture Data. Proceedings of the Fourth International Symposium on Soil Water Measurements Using Capacitance, Impedance and Time Domain, Montreal, Quebec, Canada, July 16-18, 2014.

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. Hydrological Sciences Journal 51(4):642-654, DOI: 10.1175/JHM-D-11-046.1.

Brocca, L., F. Melone, and T. Moramarco, 2008. On the Estimation of Antecedent Wetness Conditions in Rainfall-Runoff Modelling. Hydrological Processes 22:629-642, DOI: 10.1002/hyp.6629.

Clayton, J.A., 2016. Summer Soil Moisture Loss at Utah SNOTEL Sites and Streamflow Recession at Nearby Gauges: Variability in Runoff Generation and the Potential for Flow Forecasting. Hydrological Processes 30:119-134, DOI: 10.1002/hyp.10572.

Entekhabi, D., E. Njoku, P. O'Neill, M. Spencer, T. Jackson, J. Entin, E. Im, and K. Kellogg, 2008. The Soil Moisture Active/Passive Mission (SMAP). IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium 3:1-4, DOI: 10.1109/36.942551.

Flint, A.L., L.E. Flint, and M.D. Dettinger, 2008. Modeling Soil Moisture Processes and Recharge Under a Melting Snowpack. Vadose Zone Journal 7(1):350-357, DOI: 10.2136/vzj2006.0135.

Franz, K.J., H.C. Hartmann, S. Sorooshian, and R. Bales, 2003. Verification of National Weather Service Ensemble Streamflow Predictions for Water Supply Forecasting in the Colorado River Basin. Journal of Cimate 4:1105-1118, DOI: 10.1175/1525-7541(2003)004<1105:VONWSE>2.0.CO;2.

Garen, D.C., 1992. Improved Techniques in Regression-Based Streamflow Volume Forecasting. Journal of Water Resources Planning and Management 118:654-670, DOI: 10.1061/(ASCE)0733-9496.

Georgakakos, A.P., H. Yao, M.G. Mullusky, and K.P. Georgakakos, 1998. Impacts of Climate Variability on the Operational Forecast and Management of the Upper Des Moines River Basin. Water Resources Research 34:799-821, DOI: 10.1029/97WR03135.

Hamlet, A.F. and D. Huppert, 2002. Economic Value of Long-Lead Streamflow Forecasts for Columbia River Hydropower. Journal of Water Planning and Management 128(2):91-101, DOI: 10.1061/(ASCE)0733-9496.

Harpold, A., P. Brooks, S. Rajagopal, I. Heidbuchel, A. Jardine, and C. Stielstra, 2012. Changes in Snowpack Accumulation and Ablation in the Intermountain West. Water Resources Research 48:W11501, DOI: 10.1029/2012WR011949.

Harpold, A.A. and N.P. Molotch, 2015. Sensitivity of Soil Water Availability to Changing Snowmelt Timing in the Western U.S. Geophysical Research Letters 42:8011-8020, DOI: 10.1002/2015GL065855.

Homer, C., J. Dewitz, L. Yang, S. Jin, and P. Danielson, 2015. Completion of the 2011 National Land Cover Database for the Conterminous United States–Representing a Decade of Land Cover Change Information. Photogrammetric Engineering and Remote Sensing 81(5):345-354.

Hunt, E.D., K.G. Hubbard, D.A. Wilhite, T.J. Arkebauer, and A.L. Dutcher, 2009. The Development and Evaluation of a Soil Moisture Index. International Journal of Climatology 29:747-759, DOI: 10.1002/joc.1749.

Jacobs, J.M. and D.A. Myers, 2003. Improved Rainfall/Runoff Estimates Using Remotely Sensed Soil Moisture. Journal of the American Water Resources Association 39:313-324, DOI: 10.1111/j.1752-1688.2003.tb04386.x.

Kerr, Y.H., P. Waldteufel, J.P. Wigneron, J.M. Martinuzzi, J. Font, and M. Berger, 2001. Soil Moisture Retrieval from Space: The Soil Moisture and Ocean Salinity (SMOS) Mission. Geoscience and Remote Sensing, IEEE Transactions 39:1729-1735, DOI: 10.1109/36.942551.

Klos, P.Z., T.E. Link, and J.T. Abatzoglou, 2014. Extent of the Rain-Snow Transition Zone in the Western U.S. Under Historic and Projected Climate. Geophysical Research Letters 41:4560-4568, DOI: 10.1002/2014GL060500.

Koster, R.D., S.P.P. Mahanama, B. Livneh, D.P. Lettenmaier, and R.H. Reichle, 2010. Skill in Streamflow Forecasts Derived from Large-Scale Estimates of Soil Moisture and Snow. Nature Geoscience 3:613-616, DOI: 10.1038/ngeo944.

Larson, K.M., E.E. Small, and E.D. Gutmann, 2008. Use of GPS Receivers as a Soil Moisture Network for Water Cycle Studies. Geophysical Research Letters 35(24):L24405, DOI: 10.1016/S0022-1694(03)00077-5.

Lute, A.C. and J.T. Abatzoglou, 2014. Role of Extreme Snowfall Events in Interannual Variability of Snowfall Accumulation in the Western United States. Water Resources Research 50:2874-2888, DOI: 10.1002/2013WR014465.

Mathworks, 2014. MATLAB 2014a, Natick, Massachusetts.

Maurer, E.P. and D.P. Lettenmaier, 2004. Potential Effects of Long-Lead Hydrologic Predictability on Missouri River Main-Stem Reservoirs. Journal of Climate 17(1):174-186, DOI: 10.1175/1520-0442(2004) 017 < 0174:PEOLHP>2.0.CO;2.

Maurer, G.E. and D.R. Bowling, 2014. Seasonal Snowpack Characteristics Influence Soil Temperature and Water Content at Multiple Scales in Interior Western U.S. Mountain Ecosystems. Water Resources Research 50:5216-5234, DOI: 10.1002/2013WR014452.

McNamara, J.P., D. Chandler, M. Seyfried, and S. Achet, 2005. Soil Moisture States, Lateral Flow, and Streamflow Generation in a Semi-Arid, Snowmelt-Driven Catchment. Hydrological Processes 19:4023-4038, DOI: 10.1002/hyp.5869.

Miller, W.P., T.C. Piechota, S. Gangopadhyay, and T. Pruitt, 2011. Development of Streamflow Projections Under Changing Climate Conditions Over Colorado River Basin Headwaters. Hydrology and Earth System Sciences 15(7):2145-2164, DOI: 10.5194/hess-15-2145-2011.

O'Gorman, P.A., 2014. Contrasting Responses of Mean and Extreme Snowfall to Climate Change. Nature 512:416-418, DOI: 10.1038/nature13625.

Pagano, T. and D. Garen, 2005. A Recent Increase in Western US Streamflow Variability and Persistence. Journal of Hydrometeorology 6(2):173-179, DOI: 10.1175/JHM410.1.

Pagano, T., D. Garen, and S. Sorooshian, 2004. Evaluation of Official Western US Seasonal Water Supply Outlooks, 1922-2002. Journal of Hydrometerology 5:896-909, DOI: 10.1175/1525-7541(2004) 005 < 0896:EOOWUS>2.0.CO;2.

Raleigh, M. S. and M.P. Clark, 2014. Are Temperature-Index Models Appropriate for Assessing Climate Change Impacts On Snowmelt, Proceedings of the Western Snow Conference, Durango, Colorado, April 14-17, 2014.

Scipal, K., C. Scheffler, and W. Wagner, 2005. Soil Moisture-Runoff Relation at the Catchment Scale as Observed with Coarse Resolution Microwave Remote Sensing. Hydrology and Earth System Sciences 9:173-183, DOI: 10.5194/hess-9-173-2005.

Seyfried, M.S., L.E. Grant, and E. Du, 2005. Dielectric Loss and Calibration of the Hydra Probe Soil Water Sensor. Vadose Zone Journal 4:1070-1079, DOI: 10.2136/vzj2004.0148.

Seyfried, M.S., L.E. Grant, D. Marks, A. Winstral, and J. McNamara, 2009. Simulated Soil Water Storage Effects on Streamflow Generation in a Mountainous Snowmelt Environment, Idaho, USA. Hydrological Processes 23:858-873, DOI: 10.1002/hyp.7211.

Thierfelder, T.K., R.B. Grayson, D. von Rosen, and A.W. Western, 2003. Inferring the Location of Catchment Characteristic Soil Moisture Monitoring Sites. Covariance Structures in the Temporal Domain. Journal of Hydrology 280(1):13-32, DOI: 10.1016/S0022-1694(03)00077-5.

Trujillo, E. and N.P. Molotch, 2014. Snowpack Regimes of the Western United States. Water Resources Research 50:5611-5623, DOI: 10.1002/2013WR014753.

Weiler, M., B.L. McGlynn, K.J. McGuire, and J.J. McDonnell, 2003. How Does Rainfall Become Runoff? A Combined Tracer and Runoff Transfer Function Approach. Water Resources Research 39:1315, DOI: 10.1029/2003WR002331.

Western, A.W., R.B. Grayson, G. Blöschl, G.R. Willgoose, and T.A. McMahon, 1999. Observed Spatial Organization of Soil Moisture and Its Relation to Terrain Indices. Water Resources Research 35:797-810, DOI: 10.1029/1998WR900065.