Skill Assessment of Water Supply Forecasts for Western Sierra Nevada Watersheds

Brent Harrison[1] and Roger Bales, M.ASCE[2]

Abstract: The western slope of the Sierra Nevada contains 13 major river basins with sufficient long-term seasonal forecast and runoff data to assess forecast skill. These seasonal forecasts are issued by the California Department of Water Resources on February 1, March 1, April 1, and May 1 of each year for the major watersheds in California. Annual average precipitation in these river basins goes from 1,500 mm in the Yuba in the north, to 600 mm in the Kern to the south. Average runoff fraction (April to July) for the various watersheds ranged from 0.1 to 0.5 with the lower-elevation watersheds, and the Kern in the south having lowest values. The difference between precipitation and runoff, an index of evapotranspiration, was highest in the lower-elevation Cosumnes and Mokelumne Basins. Approximately half of the April 1 annual forecasts had a percent bias of ±15%. Skill scores for the 13 watersheds showed low scores (0.3) for forecasts in February, increasing through the forecast season to 0.8 for forecasts issued May 1, with 1.0 being a perfect forecast. Correlation skill measures, such as the Nash Sutcliffe scores, also exhibited increases in skill through the season from 0.45 in February to 0.95 in May. A linear regression between Nash Sutcliff scores and watershed elevation yielded a strong relationship with a coefficient of determination of 0.77. This relationship between higher elevation basins and greater forecast skill reflects the stronger statistical relationship between snow accumulations at index sites and seasonal runoff, versus more rainfall dominance in lower elevation watersheds. April through July runoff for each year was classified as the lower 30%, the mid 40%, and upper 30%; categorical skill measures were computed on the three runoff categories. Increases in forecast skill during the forecast season were visible in the low-flow and high-flow years versus the midflow years. Over forecasting of flow in the middle category was especially apparent early in the season, illustrated by high early season false-alarm rate and over forecasting bias. Difficulty in making accurate forecasts for midflow runoff along with the under forecast of high-runoff years and the over forecast of low-runoff years are shown to be common difficulties in runoff forecasting, especially early in the forecast season. Forecast skill is shown to be elevation dependent and can be expected to decrease with increasing temperatures. **DOI: 10.1061/(ASCE)HE.1943-5584.0001327.** *© 2016 American Society of Civil Engineers.*

Author keywords: Forecast; Runoff; Skill.

Introduction

Runoff from the Sierra Nevada is used for agricultural, municipal, industrial, and environmental purposes. To apportion the water to its many uses, the agencies responsible rely on runoff forecasts prepared during the snow accumulation and ablation period to estimate water supplies for the remainder of the water year. These forecasts use information on snowpack, precipitation, and other hydrologic conditions to make runoff projections. In California, the Cooperative Snow Survey Program, part of California Department of Water Resources (DWR), coordinates the measurements and prepares the forecasts. Local agencies sometimes use these forecasts to prepare specialized forecasts for their own use. The DWR forecasts have been prepared on watersheds in California since the 1930s. The forecasts predict April through July runoff and, by extension, estimate runoff for the water year. The forecasts are issued monthly, February through May, of each year.

[1]Lecturer, Sierra Nevada Research Institute, Univ. of California, Merced, 5200 N. Lake Rd., Merced, CA 95343 (corresponding author). E-mail: bharrison3@ucmerced.edu

[2]Director, Sierra Nevada Research Institute, Univ. of California Merced, 5200 N. Lake Rd., Merced CA 95343. E-mail: rbales@ucmerced.edu

Note. This manuscript was submitted on February 25, 2015; approved on October 7, 2015; published online on January 12, 2016. Discussion period open until June 12, 2016; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Hydrologic Engineering***, © ASCE, ISSN 1084-0699.

In preparation of these forecasts, snow water content is measured in snow courses consisting of 7–10 sample points (CDEC 2014a). Runoff forecasts are based on historical relationships defined by regressions between April through July runoff and snow water content of the applicable snow course, precipitation to date, and preceding month's runoff. The forecasts are issued in two parts: (1) the runoff in the April through July period, which is generally considered snowmelt; and (2) runoff for the water year, October through September. Runoff for remaining months in the winter period is estimated using relationships with runoff to date of the forecasts and precipitation accumulations. Runoff amounts for August and September are correlated with April through July runoff for forecasting purposes. Forecasts for one river basin are checked with the forecasts for adjoining basins for quality control purposes. These forecasts are used by water managers and stakeholders to commit deliveries and schedule operations. The forecasts also are used by regulatory agencies for ecosystem protection.

Similar runoff forecasts are made for watersheds in other western states of the United States. The first formal evaluations of the skill of runoff forecasts in other western states were made in the late 1950s on various portions of these forecasts. Additional work to evaluate forecast skill in the western states (excluding the western Sierra Nevada) was done again in the mid 1980s. Shafer and Huddleston (1984) reviewed historical seasonal volume forecasts based on regression techniques and found a small improvement in forecasting skill in recent years, but cautioned that large improvements in skill are not to be expected in the future by refining regression techniques. Starting in 2002, the latest work was initiated to evaluate the skill of water-supply forecasts in the western United States, again excluding the western Sierra Nevada. Franz et al. (2003) evaluated the forecasts at 14 sites in the Colorado River Basin and determined that the Ensemble Streamflow Prediction (ESP) system, developed by the National Weather Service (NWS), performed better than climatology forecasts.

Pagano et al. (2004) evaluated forecasts using Nash-Sutcliffe correlation skill measurement scores and other measures on 29 unregulated rivers in the western United States. They found high skill for forecasts issued on April 1. Forecasts made earlier in the season contained more uncertainty, but were shown to still be skillful. They also found that areas with wet winters and dry springs presented higher skill improvement over the forecast season than for areas with dryer winters and wetter springs. In addition, they also found mixed changes in skill over time when comparing different areas of the study. Pagano and his coauthors found that it is desirable that the measures to evaluate forecast skill be chosen carefully so they are understandable and relevant to forecast users.

Hartmann et al. (2006) performed an assessment of water-supply outlooks in the Colorado River Basin, which established a baseline for identifying improvements in hydrologic forecasts. The following work by Morrill et al. (2007) was an assessment of the strengths and weaknesses of seasonal water-supply outlooks at 54 sites in the Colorado River Basin using an assortment of skill measures. Morrill found that the water-supply outlooks were an improvement over climatology during the historical record for most sites. They also found that most of the forecasts were conservative, with above-average flows under predicted and below-average flows over predicted.

An analysis of percent bias for 28 forecast points within the Colorado Basin indicated no detectable long-term trend (Harrison and Bales 2014). A broader assessment of forecast skill at those same 28 points (Harrison and Bales, unpublished data, 2015) showed increases in skill over the forecast season, but no systematic spatial differences in forecast skill.

The aims of this research were to assess the skill of seasonal water-supply outlooks in a mixed rain-snow mountain system of river basins, the Sierra Nevada, and to analyze the skills as a function of basin-specific topographic, climatic, and hydrologic features. This study also aims to identify future changes in the forecast skill level that will affect managers and other interested stakeholders as they plan and schedule water releases, delivery, and transfers in the region.

Methods and Data

This study assesses the skill of seasonal water-supply outlooks for the 13 main river basins draining the western Sierra Nevada using summary, correlation, and categorical measures of forecast skill. All of the forecast points were at the mountain front, and, in most cases, are at a rim dam on the main river draining the basin (Fig. 1). Records of runoff for these basins extend for multiple decades, with most extending back over 100 years. Forecast records extend back to the 1930s in many cases.

Skill Measures

In this study, summary and correlation (Table 1), and categorical measures (Tables 2 and 3) of the skill of runoff forecasts are introduced. Summary measures indicate the error in forecasts as an arithmetic difference between the forecast and the observation. The mean absolute error (MAE) and mean square error (MSE) have dimensions and depend on the magnitude of the runoff. For the current analysis, a skill score (SS) is used, which is a correlation measure calculated by normalizing by the difference of each observation from the mean (Table 1). A zero skill score indicates no skill more than using the historical average observation as the forecast; a negative value indicates that using the average would be better than using the forecast, and a skill score of 1 indicates perfect skill (no error in the forecast). The SS MSE is mathematically equivalent to the Nash-Sutcliffe score (NS).

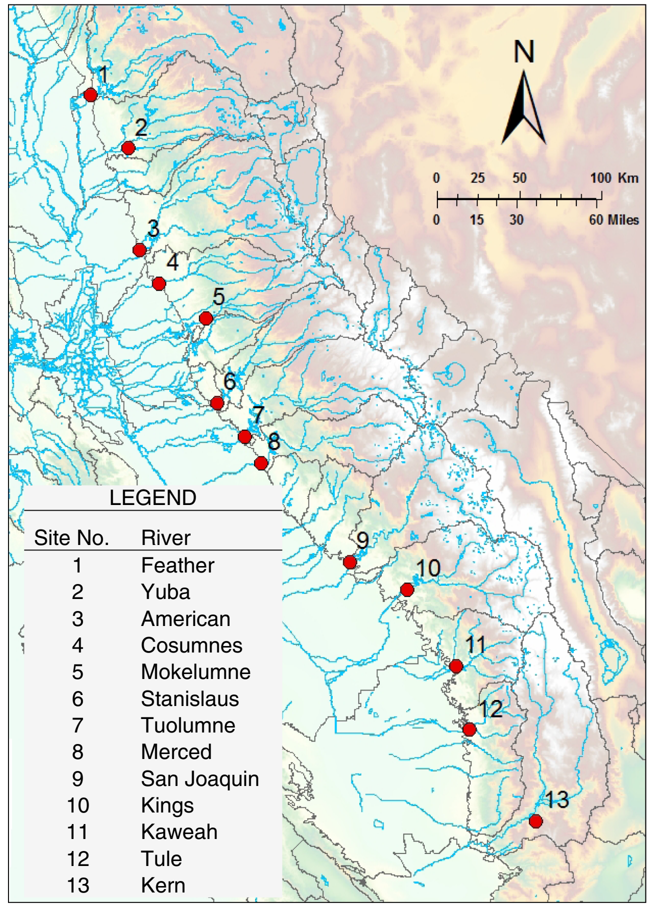

Fig. 1. Map of western Sierra Nevada forecast locations

Table 1. Summary and Correlation Measures of Forecast Skill

Skill measure	Equation[a]
Mean absolute error (MAE)	$\text{MAE} = \sum \lvert f_i - o_i \rvert / n$
Mean square error (MSE)	$\text{MSE} = \sum (f_i - o_i)^2 / n$
MAE skill score	$\text{SS}_{\text{MAE}} = 1 - \text{MAE}/\text{MAE}_{cl}$, where $\text{MAE}_{cl} = \sum \lvert \bar{o} - o_i \rvert / n$
MSE skill score[b]	$\text{SS}_{\text{MAE}} = 1 - \text{MSE}/\text{MSE}_{cl}$, where $\text{MSE}_{cl} = \sum (\bar{o} - o_i)^2 / n$
Percent bias (PBias)	$\text{PBias} = (f_i - o_i/o_i)100\%$

[a]o_i = observation; $\bar{o}$ = mean of the observations; f_i = forecast; n = number of observations.
[b]Equivalent to Nash-Sutcliffe (NS) score.

Table 2. Variables for 2 × 2 Contingency Table

Forecast result	Observed	Not observed
Forecast	a	b
Not forecast	c	d

Table 3. Categorical Measures

Measure	Explanation	Equation	Range
Probability of detection (POD)	Correct forecasts divided by observations	$a/(a+c)$	0–1 (perfect)
False alarm rate (FAR)	Incorrect forecasts divided by forecasts	$b/(a+b)$	1–0 (perfect)
Bias	Correct and noncorrect forecasts divided by observations	$(a+b)/(a+c)$	>1 over; and <1 under forecast
Threat score or critical success index (TS)	Correct forecasts divided by the forecasts plus nonforecast observations	$a/(a+b+c)$	0–1 (perfect)
Hit rate (HR)	Correct forecasts and correct nonforecasts, divided by total forecasts and observations	$(a+d)/(a+b+c+d)$	0–1 (perfect)

The percent bias (PBias) is the direct measure of the error divided by the observation, expressed as a percent. A perfect forecast will have a PBias of zero, a positive value indicates over forecast (forecast exceeds observation), and a negative PBias indicates under forecast. As PBias is normalized by the observation, it is a dimensionless measure. PBias is an easily understood skill measure, and values from a series of annual forecasts can be analyzed readily for the influence of independent climate variables.

Categorical measures indicate the skill of the forecast in predicting the magnitude category of the runoff; in this case, low, middle, and high runoff categories. For example, if the forecast was for flows assigned to the low-flow category, did the low flow actually occur? Historical runoff records for each forecast point were divided into three runoff categories, the lower 30%, the mid 40%, and highest 30% of flows. A 2×2 contingency table was used to count the results of forecasts versus observations in each category (Table 2) (Wilks 2011). Five categorical measures were assessed (Table 3).

The probability of detection (POD) is intuitive, being the proportion of times the event or category was forecast compared to the times it occurred. The false alarm rate (FAR) is the proportion of forecast events in a category that failed to occur. The bias indicates if a category is over forecast (>1) or under forecast (<1). The threat score (TS), also known as the critical success index, normalizes correct forecasts by total forecasts plus observations for that category. Unlike the POD and FAR, the TS takes into account both missed events and false alarms. The hit rate (HR) credits correct forecasts and nonforecasts by dividing by the total forecasts plus observations. It thus reflects both correct detection, or classification, and lack of incorrect classification. The POD, TS, and HR all range from zero (poor) to 1.0 (perfect). FAR has an opposite orientation, ranging from zero (perfect) to 1.0 (poor).

Source of Data

The 13 forecast points drain watersheds ranging in size from 1,010 to 9,386 km^2, and range from the Feather River in the north to the Kern River in the south (Table 4). These watersheds are highly developed with water storage and diversion facilities, providing water to extensive areas of cropland and urban development in California. Owing to the many diversions occurring above each main gauging point, this study calculated an April through July runoff total, the full natural flow (FNF). The FNF is the reconstructed flow in the river if there were no diversions or storage, and is calculated on a daily basis from existing flow gauges, adding any increases in reservoir storage and adding estimates of diversions from the river basin. The DWR tabulates the FNF on river basins that have runoff forecasts in Bulletin 120, Water Conditions in California, with the data available over the internet (CDEC 2014b).

Bulletin 120 is published in February, March, April, and May of each year, and contains the snow-survey data and runoff forecasts analyzed in this study. This study used the DWR tabulations of Bulletin 120 runoff forecasts for the April-July runoff period for the years 1930 to 2012, or shorter periods if forecasts started in later years (S. Nemeth, personal communication, 2011). The starting year of forecast for each location varies considerably, especially for forecasts other than the month of April. Most February and March forecasting started by 1953. Most April and May forecasting started by 1939. All forecasts on the Cosumnes River started in 1963, and on the Tule River in 1959. The April to July runoff forecasts were analyzed in this study because the April to July runoff is historically the main period of snowmelt runoff in the Sierra Nevada.

Watershed areas and elevation distributions were extracted from Calwater Basin polygons overlain on 30-m digital elevation data. Precipitation data used in the interpretation were from PRISM, which are spatial datasets incorporating a wide range of climatic

Table 4. Forecast Points and River Basins

Number	Location	Latitude	Longitude	Runoff records	Forecast records	Area (km^2)
1	Feather River at Oroville	39.522	121.547	1906–2012	1938–2012	9,386
2	Yuba River near Smartsville plus Deer Creek	39.235	121.273	1901–2012	1937–2012	3,085
3	American River inflow to Folsom Lake	38.683	121.183	1901–2012	1932–2012	4,921
4	Cosumnes River at Michigan Bar	38.5	121.044	1908–2012	1963–2012	1,373
5	Mokelumne River at Mokelumne Hill (Pardee)	38.313	120.719	1901–2012	1936–2012	1,489
6	Stanislaus River at Goodwin Dam	37.852	120.637	1901–2012	1932–2012	2,422
7	Tuolumne River at La Grange Dam	37.666	120.441	1901–2012	1932–2012	3,963
8	Merced River at Merced Falls	37.522	120.331	1901–2012	1932–2012	2,642
9	San Joaquin River at Friant Dam	36.984	119.723	1901–2012	1932–2012	4,248
10	Kings River at Pine Flat Dam	36.831	119.335	1901–2012	1932–2012	3,989
11	Kaweah River below Terminus Reservoir	36.412	119.003	1901–2012	1932–2012	1,458
12	Tule River below Lake Success	36.061	118.922	1931–2012	1953–2012	1,010
13	Kern River inflow to Lake Isabella	35.556	118.484	1930–2012	1932–2012	5,387

observations. Monthly precipitation data for the years 1896–2013 were downloaded from PRISM and the average precipitation was calculated for each basin (PRISM 2014).

Results

Watershed Features

From a plot of area versus elevation for each watershed (Fig. 2), it is apparent that the southern Sierra basins have more high-elevation, snow-producing area with the highest elevation watershed being the Kings, and the two lowest being the Cosumnes and Tule. Basin-average precipitation declines from north to south (Fig. 3), with precipitation for the three most southern watersheds roughly half of that in the three most northern basins. There is an increase in precipitation when comparing the Yuba to the Feather River, which is opposite of the trend for the remaining 11 river basins. There is also slightly less precipitation for the Cosumnes River than would be expected by the precipitation for neighboring watersheds, which may be related to its lower relative elevation.

FNF records go back to 1900 for many of the forecast points, and cover a range of wet and dry years (Fig. 4). Quite visible in the time series is the extended drought during the 1920s and 1930s, and the historical drought of 1976 and 1977. Also visible in the time series are the heavy runoff years prior to 1920 and the record runoff in 1983. Fig. 4 shows that runoff varies more than does precipitation over the period of record.

As runoff is related to both precipitation and topography, among other factors, it exhibits more variability across the latitudinal gradient than does precipitation [Fig. 5(a)]. There is an uneven trend of decreasing precipitation from north to south. This trend of decreasing precipitation leads directly to decreasing runoff from north to south in the western Sierra Nevada. There is an uncharacteristic increase in precipitation and discharge for the Yuba River, with a decrease for the Cosumnes River. There is also an uncharacteristic decrease in discharge for both the Tule and Kern Rivers. Fig. 5(b) shows the difference between precipitation and discharge, assumed to represent evapotranspiration, overlain on total precipitation.

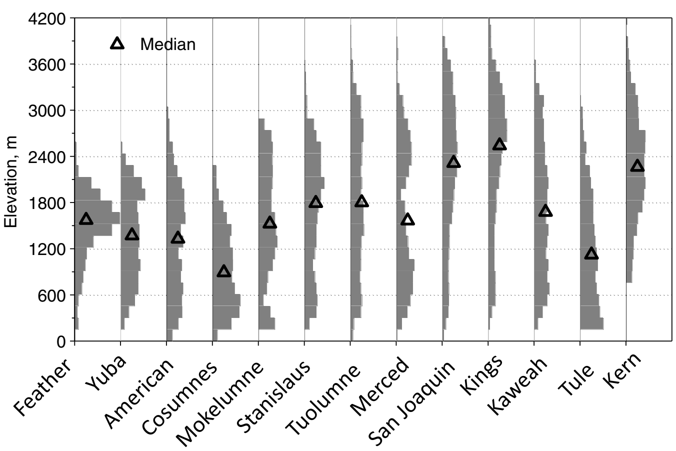

Fig. 2. Western Sierra Nevada watershed elevation histogram with median elevation shown

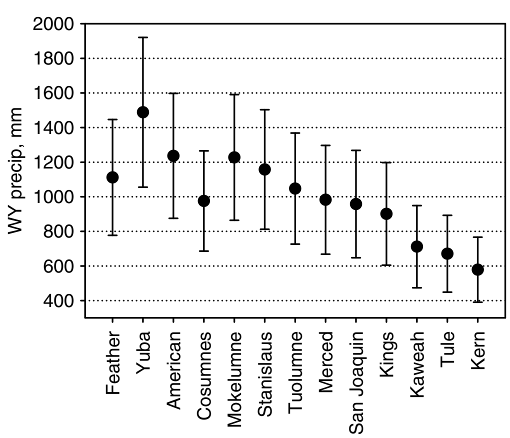

Fig. 3. Mean precipitation from 1896–2013 for 13 western Sierra basins—one standard deviation (PRISM data)

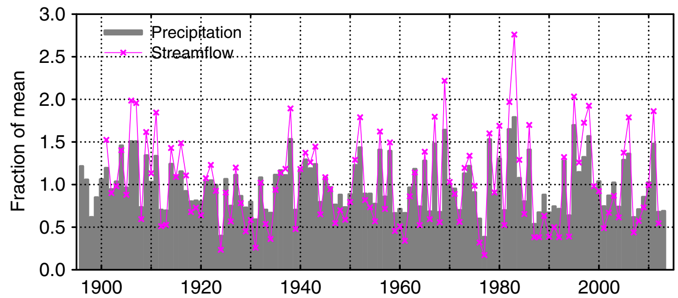

Fig. 4. Mean precipitation and water year runoff for 13 western Sierra basins

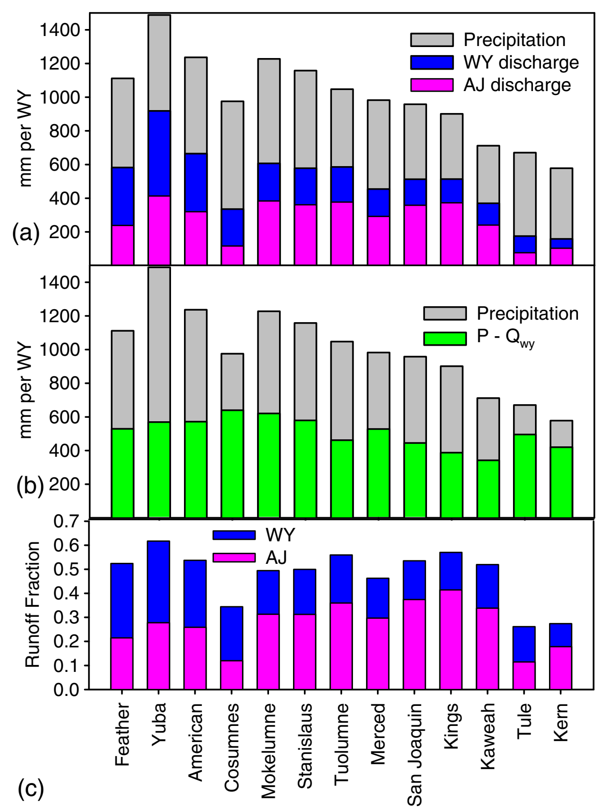

Fig. 5. (a) Precipitation with Water Year (WY) and April through July (AJ) discharge for 13 western Sierra locations; (b) precipitation (P) with P minus WY runoff (Q_{wy}); (c) runoff fraction of precipitation for WY and AJ for 13 western Sierra watersheds

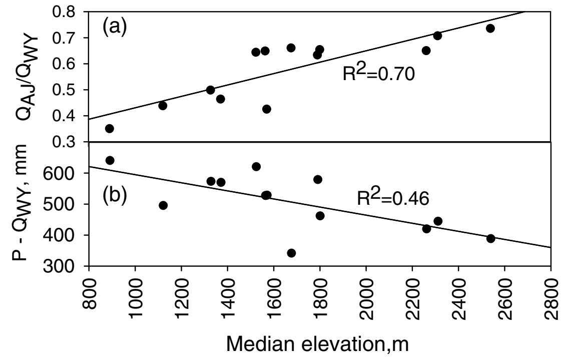

Fig. 6. (a) Runoff ratio (April–July)/water year runoff versus watershed median elevation; (b) precipitation minus WY runoff versus watershed median elevation

The Cosumnes and the Mokelumne have the highest evapotranspiration, with the Tule higher than adjacent watersheds. Fig. 5(c) summarizes the precipitation and discharge relationship by showing the fraction of precipitation that leaves the watershed as runoff. Low-runoff fraction for the Cosumnes is consistent with its lower elevation, with the drop for the Tule and Kern related to the drier conditions of the two southern watersheds.

April-July runoff accounted for an average of 35% (Cosumnes) to 74% (Kings) of the annual runoff, and is correlated with median elevation [Fig. 6(a)]. Precipitation and runoff are not well correlated with elevation (not shown); however, the difference between precipitation and runoff, i.e., evapotranspiration, decreases with increasing elevation [Fig. 6(b)].

Forecast Skill—Summary and Correlation Measures

Annual PBias values for April for one central-Sierra basin, the Tuolumne, range from about 40 to −40% (Fig. 7). The high variability of April-to-July runoff is shown in the bottom panel of Fig. 7. In the top panel, there is little evidence of a trend in PBias over the time series, either in the aggregate or partitioned into low, mid, or high flows. There is evidence of over forecast of low flows, as 15 low flows are over forecast with only eight under forecast. In contrast, 14 of the high flows are under forecast with only five over forecast. The midflow years are nearly evenly split with 15 over forecasts and 16 under forecast. PBias and runoff values for all the 13 river basins are included as supplementary data. The PBias and runoff time series for all the basins exhibit similar patterns.

It is apparent that low-runoff years tend to be somewhat over forecast, with high-runoff years somewhat under forecast for all months (Fig. 8). It also is apparent that forecasts for February and March have a much higher bias than those for April and May. Low-flow years are greatly over forecast early in the forecast season, which would be expected as average climatology is assumed for the remainder of the runoff season. As more information is gathered on snowpack and precipitation, the PBias decreases steadily from 40% with some decrease in width of the distribution, ending near zero for the May forecasts. For the midflow years, there is limited change through the forecast months, with the median staying near zero PBias. The width of the PBias distribution does decrease as information increases. High-flow years behave the opposite of low-flow years. The high-flow years are under forecast early in the forecast season, as limited information is available and average climatology is assumed for the remainder of the forecast period. The PBias increases from −30% to near zero by April, along with a decrease in width of the distribution that extends to May. The range of PBias is similar across most of the 13 forecast points (Fig. 9). Absolute values of PBias are shown as PBias is relatively symmetrical around zero for most basins. Two basins, the Cosumnes and Tule, show a wider distribution of values, with outliers larger than 50% lumped in the highest category of 50%. For most of the basins, 25% of the PBias values exceed 20–25%, with 50% exceeding 10–15%. Again, the Cosumnes and Tule have higher values.

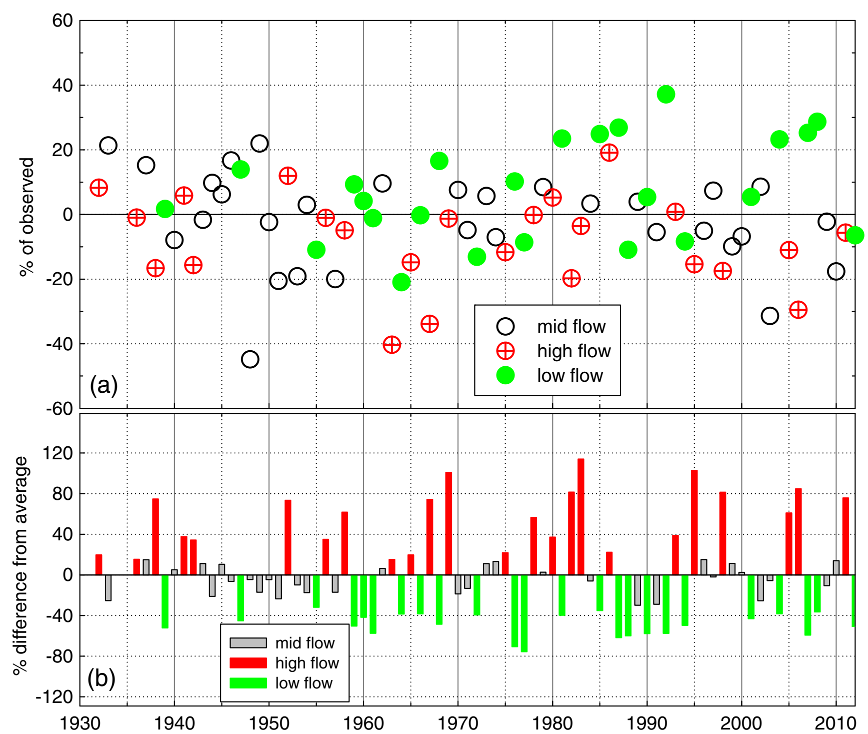

Fig. 7. (a) Tuolumne River (April 1) percent bias by category; (b) AJ runoff by category

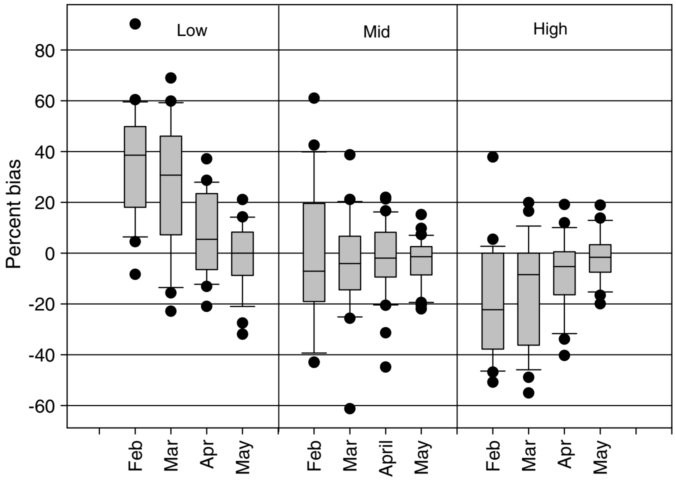

Fig. 8. Tuolumne River (April 1) percent bias categorized by runoff magnitude for each month of forecast period; box is 25 and 75%, with median shown as a bar; tics are 10 and 90% with outliers outside of that range

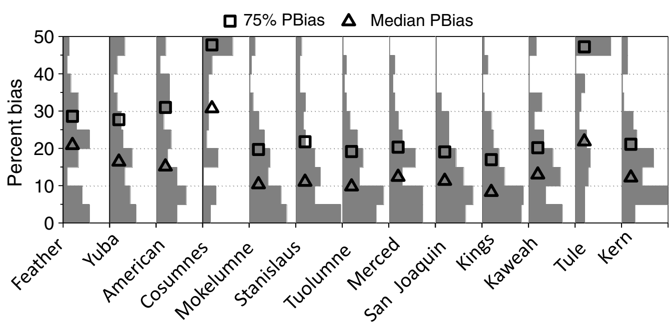

Fig. 9. Histogram of April 1 absolute value of percent bias by river basin; median (50%) and 75%; PBias also shown; histogram truncated with upper range of 45 to 50 containing any PBias 45 and above

The SS for MAE (SSMAE) and the NS coefficient show similar patterns to the PBias, improving from February to May as more information becomes available (Fig. 10). The forecast skill in February is quite low, with the SSMAE centered near 0.3, but with a tight distribution because this early in the forecast season the forecasts are made using average climatology. The forecasts in March contain additional winter-storm information and thus have higher skill, but a much wider distribution. Fig. 10 indicates that the runoff forecast skill increases quite steeply from the start to end of the forecast period, with the SSMAE in March around 0.4, 0.65 in April, and ending the forecasts season in May with a SSMAE of approximately 0.75. The width of the SSMAE distribution decreases from March to April and then again to May, as additional climate information is available.

The distributions of NS for the four forecast months at the 13 forecast locations exhibit a similar pattern [Fig. 10(b)]. The median February NS starts at 0.45 and steadily increases to 0.95 for the median in May. Again, the width of the distribution decreases with increasing information from March to May. For the months of February, March, and April, there is an increase in NS score in going from the Feather River in the north to the Kings River to the south, excepting the smaller Cosumnes Basin (Fig. 11). As previously stated, the drops in scores for the Cosumnes and Tule

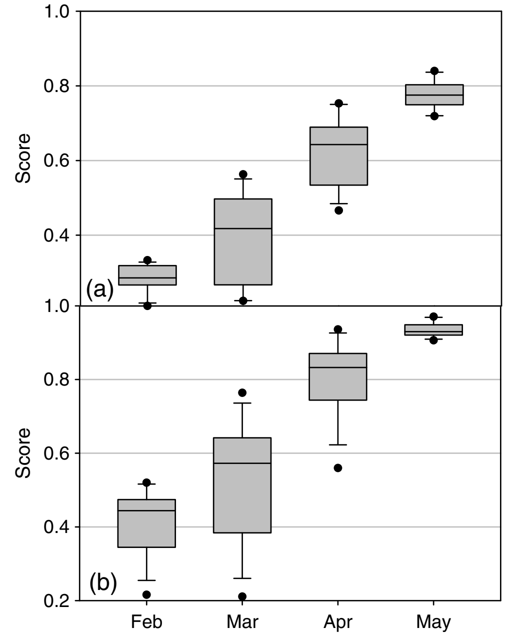

Fig. 10. (a) SSMAE and MAE skill score; (b) NS score by forecast month of February to May

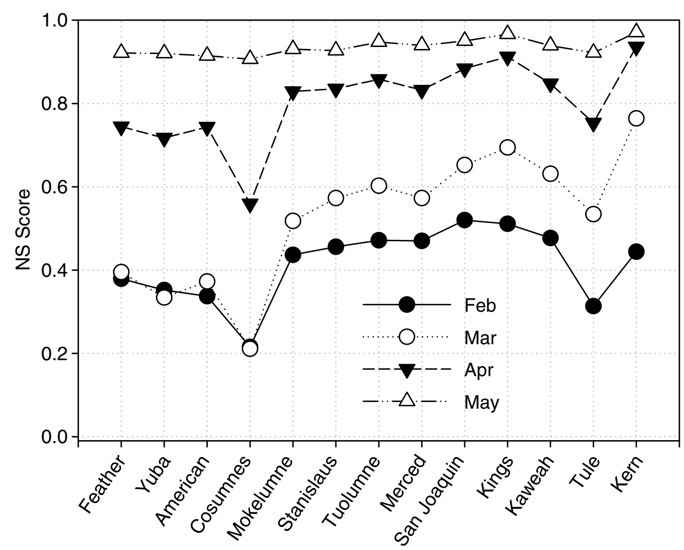

Fig. 11. NS scores for each western Sierra Basin by forecast month

are expected as these watersheds are smaller and lower elevation, with less snow.

Forecast Skill—Categorical Measures

The POD shows the previously discussed increase in skill during the forecast months for the low- and high-flow years [Fig. 12(a)]. By April the POD for low- and high-flow years is above 0.6 for most basins and averages about 0.8. For March the average is closer to 0.5. The midflow years exhibit some dispersion around a flat POD during the forecast season. This characteristic may be reflective of the assumption of average climatology for the remainder of the runoff season.

The FAR reinforces the difficulty of forecasting midflow years, as each year looks midflow early in the forecast season [Fig. 12(b)]. For the FAR, it is much higher early in the midflow years, with a steep improvement as the forecast season progresses. The FAR in the low- and high-flow years shows a small decrease over the forecast season.

The bias also illustrates the effect of information on midflow-year forecasts [Fig. 12(c)]. For the midflow years, the graph indicates over forecast early in the season (February and March) with a movement to no bias (1.0) as climate information becomes more available. Interestingly, the bias shows under forecast of both low- and high-flow years early in the forecast season, with a movement to little or no bias by April. This also is related to the assumption of average climatic conditions early in the forecast period.

The TS is shown in Fig. 12(d). The TS values are fairly uniform across the three runoff year types, with the previously observed increase in skill from 0.5 to 0.8 as information increases during the forecast season. The steady increase in TS values over the season for the high and low categories mirrors that of the POD. However, it also improves for the mid category, reflecting the improvement in FAR from February—March to April–May in that category. Still, the median TS is only 0.5 for April for the mid category and 0.6 for the high and low categories.

The HR for February forecasts [Fig. 12(e)] is high for both the low-flow and high-flow years (0.8), whereas lower for midflow years (0.6). The HR increases through the forecast season and for the April forecasts the HR is 0.85 for low and high flows, but for midflow years is approximately 0.75. The slightly lower scores for midflow years may reflect the occurrence of low or high flows for the season even if average flows occur during the early part of the forecast season. More forecasts are in the midflow category because low- and high-flow years are forecast less than they occur. The HR values are somewhat higher than those for TS, reflecting the addition of correct nonforecasts to both the numerator and denominator.

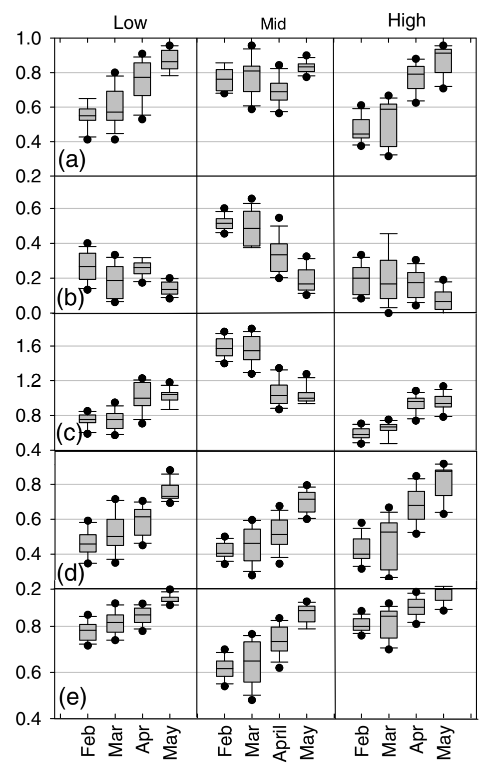

Fig. 12. Categorical measures for western Sierra Nevada for each forecast month; categories are low runoff, mid runoff, and high runoff; forecast months (February to May) are on the abscissa: (a) probability of detection (POD); (b) false alarm rate (FAR); (c) bias; (d) threat score or critical success index (TS); (e) hit rate (HR)

Discussion

The trend of increasing forecast skill by month reflects the occurrence of precipitation and snow accumulation throughout the winter and spring (Fig. 13). Watershed elevation, an index of rain versus snow, also affects the forecast skill. Monthly precipitation and snow accumulation amounts from 1971 to 2013 were obtained from the Central Sierra Snow Laboratory at Soda Springs, California (R. Osterhuber, personal communication, 2014). This site was chosen as an index of precipitation and snow accumulation because of its central location and long period of records. Fig. 13(a) shows the monthly precipitation and cumulative fraction of annual precipitation at that site. Fig. 13(b) shows the snowfall and cumulative fraction of snowfall at the site. There is a strong correlation between the accumulation of the seasonal precipitation and the increase in skill in runoff forecasts. In February, the NS is 0.4, with the cumulative snow at 0.7 and cumulative precipitation at 0.6 (Figs. 10 and 13). The NS, and cumulative precipitation and snow, increases in April to an NS of 0.8 with nearly all of the snow and 0.9 of the yearly precipitation. These measures increase again with the May forecasts. As more of the seasonal precipitation falls, it is measured and incorporated into the runoff forecasts, and the skill of the forecast increases. The relationship is confirmed by the NS to snow correlation coefficient of 0.92, and a NS to precipitation correlation coefficient of 0.93. For the months of February, March, and April, there is an increase in NS score from the Feather River in the north to the Kings River to the south, excepting the smaller Cosumnes Basin (Fig. 11). This is because of the higher elevations and increased snow dominance of the southern watersheds. In Fig. 6(a),

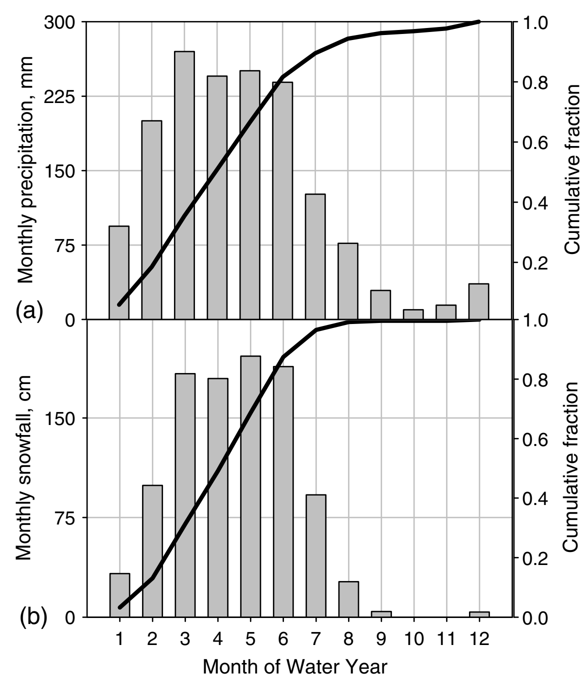

Fig. 13. (a) Mean precipitation and cumulative mean fraction precipitation; (b) mean snowfall by month and cumulative mean fraction snowfall; location is the Central Sierra Snow Laboratory

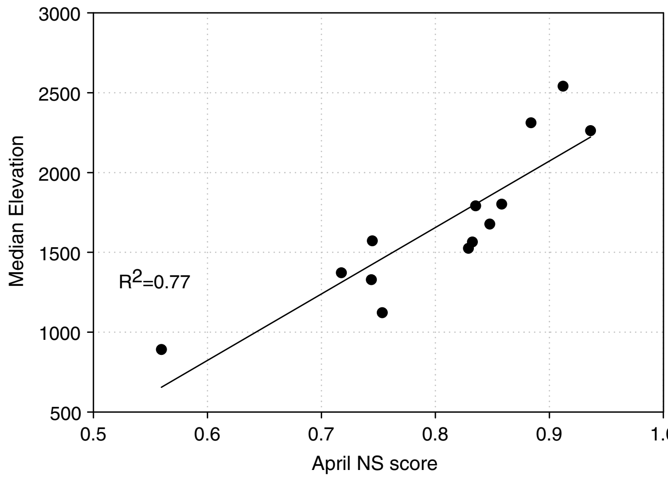

Fig. 14. April 1 NS score for each watershed versus median elevation of each watershed

the ratio of April through July runoff is plotted as a function of median watershed elevation. The associated linear regression indicates a solid relationship with an R^2 of 0.70. These results reflect increasing snow dominance of the central and southern Sierra watersheds.

The uneven skill of the forecasts south of the Kings River can be attributed to the lower elevation of the watersheds and to the lower levels of precipitation to the south and the north-south aspect of the Kern River watershed, in contrast to the east-west aspect of most of the other watersheds. As seen previously, the drops in scores for the Cosumnes and Tule are expected as these watersheds are smaller and lower elevation. This increase in forecast skill in the southern direction is attributable to the higher elevations of the Sierra to the south. The increase in forecast skill as elevation increases is illustrated further in Fig. 14, which shows a linear regression between April NS and median basin elevation with an R^2 of 0.77. The relationship would be steeper if the 0.56 NS score for the Cosumnes were omitted from the regression.

Of the categorical measures, POD and FAR are the simplest mathematically and conceptually straightforward, though they are poorly correlated (not shown). Bias is correlated with POD for the low-flow category, suggesting that noncorrect low-flow forecasts are not emphasized (Fig. 15). However, the lack of correlation between Bias and POD for mid flows reflects the high number of incorrect midflow forecasts for April. This point is reinforced by the correlation between Bias and FAR. The lower importance of incorrect low-flow forecasts also is reflected in the similar correlation between TS and POD, as between Bias and POD. TS differs from POD by including incorrect forecasts in the denominator. The higher correlations between TS and POD for high and mid flows, versus the correlations for Bias and POD, also reflect the addition of incorrect forecasts. But in the case of TS, the incorrect forecasts are in the denominator, resulting in lower TS versus POD values. Slopes of the TS and POD correlations are near 1.0 for mid and high flows. This also is reflected in the high correlation between TS and FAR for mid and high flows, versus lower correlation for low flows, and the relatively high FAR values for mid flows, versus lower FAR values for high flows. HR differs from TS by including correct nonforecasts in both the numerator and denominator. The HR and TS skill measures are thus very highly correlated, with R^2 values of 0.92–0.98 (not shown). The HR values are higher than POD values, especially for low and high flows, reflecting the importance of correct nonforecasts for those categories.

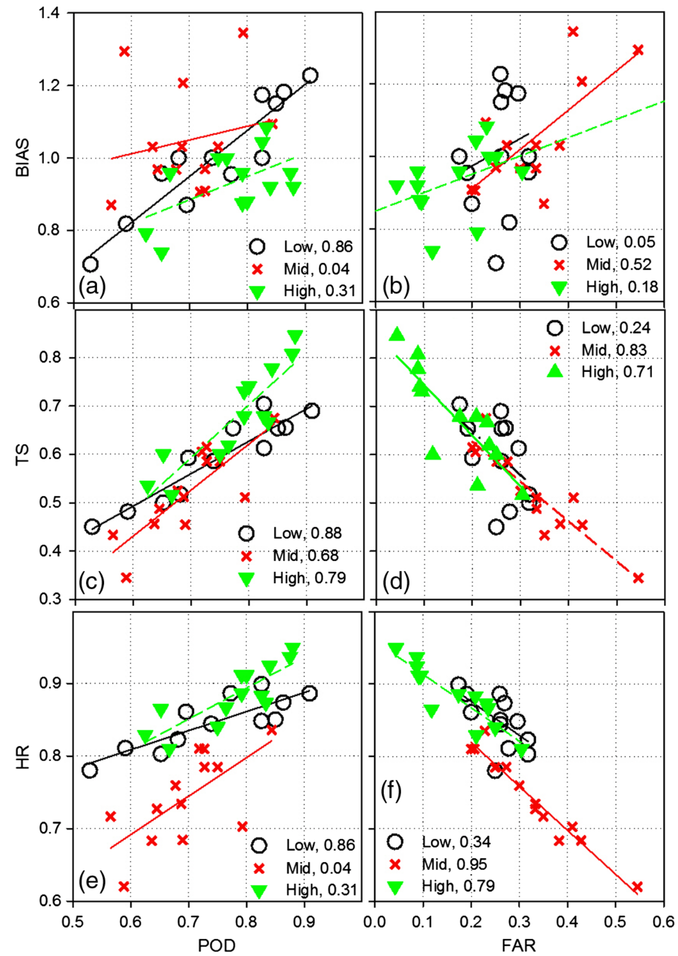

Fig. 15. Correlation between categorical measures; values in legend of each panel are R^2

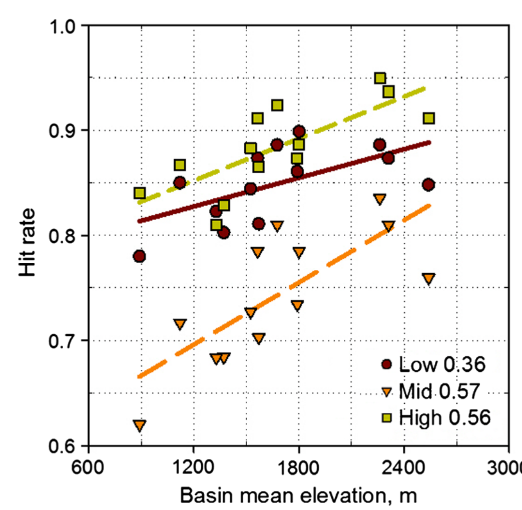

Fig. 16. Hit rate versus elevation; values in legend are R^2

Mid flows have HR values much closer to their POD values, and also a very high correlation between HR and FAR. The HR is more highly correlated to elevation in mid and high-flow years (Fig. 16), thus illustrating the influence of snow dominated runoff in forecasting skill.

When deciding to spill water, such as releasing water over a dam to provide storage for expected high flows, over forecasts can result in lower hydropower production if the system is already producing at a maximum, result in flashiness in streamflow that affects habitat, and result in potential loss of water supplies that would otherwise be captured by seasonal storage behind dams or in downstream groundwater. Underforecasts are also problematic in wet years, when storage is limited. Bias and POD is of interest in both cases. FAR also should be considered when storage is limited and spilling water would have important economic and environmental consequences. HR and TS are more balanced indicators; which one is more sensitive for a given river basin should be assessed on a case-by-case basis. HR may be more applicable where overall forecast skill is higher.

Increasing average temperatures in the Sierra, as the Earth's climate warms, will mean that all basins will receive a greater fraction of precipitation as rain rather than as snow. Using an approximate temperature lapse rate of 2°C per 300 m elevation, a 2°C increase could result in a drop in the NS score by approximately 0.1 (Fig. 14). Statistical forecasts that are based on indices of snow accumulation would not be expected to yield as much skill in forecasts if the snow is no longer accumulating. Similarly, the skill of categorical forecasts, also strongly elevation dependent, can be expected to decrease as temperature increases and higher elevations tomorrow become more like lower elevations are today.

One other projected effect of changing climate is an increase in the amount of total precipitation falling as rain, emphasizing the potential of distributed, representative rainfall and snowfall measurements to enhance forecasts. This shift to rain will decrease the skill level of forecasts in the northern Sierra Nevada and lower-elevation basins more than the higher-elevation southern basins.

Conclusions

The summary and correlation skill measures such as SSMAE and NS can appraise forecast skill over a period of years. In contrast, PBias will generate a time series for each location that can be reviewed for changes in skill over time. Currently, there is no indication of a long- term trend in forecast skill as shown in the PBias record.

This analysis suggests that use of POD, FAR, and bias together can be a good diagnostic for the Sierra Nevada overall. However, examining individual basins, these measures individually show only moderate correlations with elevation (supplementary data, Appendix H). HR shows a stronger correlation, reflecting the combined effects of correct and incorrect forecasts plus nonforecasts. This correlation with elevation also suggests that better measurements of winter precipitation and snowpack, which are generally better in the higher-elevation basins with more snow than rain, contributes to improved forecasts.

Any changes in forecast skill, especially decreases in skill, would have direct effects on users of these forecasts. Skill changes will impact water allocation between competing uses, hydroelectric system operation, and spills to provide flood storage. Preparers and users of these forecasts should be aware of the potential for skill changes and should incorporate processes to recognize and track forecast skill.

Various changing climate scenarios present the possibility of an increase of 2 or 4°C in temperature (CEC 2015). This will result in the loss of snow cover along with a rise in elevation of the snow line. Both results will decrease forecast skill. As a result, users of water-supply forecasts may consider use of forecasting tools and data that are based on principles of mass balance and on the spatially distributed data needed to drive the models. Although current modeling tools are sufficiently flexible to incorporate immediate and larger future changes in climate that are outside the current stationary assumption, data for those models are largely lacking.

Future increases in skill level could be enabled by incorporating snow-cover data estimated by remote sensing blended with representative ground measurements into the forecasting process. Increased forecast skill earlier in the water year, if new measurements can facilitate that, can provide significant economic benefits to water users, and introduce flexibility and resiliency into water-management decisions.

Acknowledgments

The authors wish to thank the staff of the California Cooperative Snow Survey Program and the staff of the Central California Snow Laboratory for assembling the data used in this study. University of California Merced's support of the first author's graduate studies is gratefully acknowledged.

Supplemental Data

Appendices A through H are available online in the ASCE Library (http://www.ascelibrary.org).

References

CDEC. (2014a). "California's cooperative snow surveys program." ⟨Cdec.water.ca.gov/snow/meeting/CCSS.2003.Meeting/Calif_Coop_Snow_Program.pdf⟩ (Dec. 10, 2014).

CDEC. (2014b). "Welcome to California data exchange center." ⟨cdec.water.ca.gov⟩ (Apr. 22, 2014).

CEC. (2015). "Our changing climate 2012 vulnerability and adaptation to the increasing risks from climate change in California." ⟨http://www.energy.ca.gov/2012publications/CEC-500-2012-007/CEC-500-2012-007.pdf⟩ (Jul. 10, 2015).

Franz, K. J., Hartmann, H. C., Sorooshian, S., and Bales, R. (2003). "Verification of national weather service ensemble streamflow predictions for water supply forecasting in the Colorado River basin." *J. Hydrometeorol.*, 4(6), 1105–1118.

Harrison, B., and Bales, R. (2014). "Percent bias assessment of water supply outlooks in the Colorado River basin." *Proc., Western Snow Conf.*, Brush Prairie, WA, 91–100.

Hartmann, H. C., Morrill, J. C., and Bales, R. (2006). "Baseline for identifying improvements in hydrologic forecasts: Assessment of water supply outlooks for the Colorado river basin." *Eos. Trans. AGU*, 87(52), 649.

Morrill, J. C., Hartmann, H. C., and Bales, R. C. (2007). "An assessment of seasonal water supply outlooks in the Colorado River basin." Dept. of Hydrology and Water Resources, Univ. of Arizona, Tucson, AZ.

Pagano, T., Garen, D., and Sorooshian, S. (2004). "Evaluation of official western US seasonal water supply outlooks, 1922–2002." *J. Hydrometeorol.*, 5(5), 896–909.

PRISM Climate Group Oregon State University. (2014). "PRISM climate data." ⟨http://prism.oregonstate.edu⟩ (Sep. 15, 2014).

Shafer, B. A., and Huddleston, J. M. (1984). "Analysis of seasonal volume streamflow forecast errors in the western United States." *Proc., Critical Assessment of Forecasting in Water Quality Goals in Western Water Resource Management*, American Water Resource Association, Bethesda, MD, 117–126.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*, Academic press, Oxford, U.K., 308–313.