MODELING RUNOFF FORECAST UNCERTAINTY USING MAPS OF SNOW WATER EQUIVALENT TO CONDITION ERROR PROPAGATION

Elijah Boardman[1,2], Adrian Harpold[1], Carl Renshaw[2], Evan Dethier[2], Gabriel Lewis[1], Robert Shriver[1], Bruce McGurk[3], Reggie Walters[4], Joseph Boardman[5], Thomas Painter[5], Jeffrey Deems[5], and Kat Bormann[5]

ABSTRACT

Seasonal water supply forecasts are a critical component of water management strategies in Mediterranean climates, where most precipitation falls outside the growing season and the water supply is dependent on sub-annual storage. Successful runoff forecasting not only depends on the skill of the chosen model but also necessitates a critical evaluation of uncertainty. The severity of the 2021 drought in the Sierra Nevada (California, USA) fell outside the historical uncertainty ranges of many statistical water supply forecasts. Here, we develop a flexible framework to separate the impact of hydrological and meteorological uncertainty sources on forecast confidence intervals under a range of antecedent conditions, illuminating observations of heightened forecast uncertainty during 2021 and other drought seasons. Using empirical runoff models of varying complexity to represent a range of forecasting methods, we constrain six water balance equations for the upper Tuolumne and Merced Rivers using full-catchment estimates of snow water equivalent (SWE) from the Airborne Snow Observatory (ASO) and in-situ precipitation measurements. Bayesian sampling of each model's parameter space enables quantification of parameter uncertainty and process variability for each model, and by propagating distributions of post-forecast precipitation through the water balance equations, we demonstrate the separation of uncertainty arising from variations in spring precipitation from uncertainty arising from the hydrological models. Results indicate that the uncertainty of statistical water supply forecasts increases nonlinearly under low-SWE conditions, primarily due to the importance of spring precipitation variability. We suggest that the resiliency of water supply forecasts can be improved by separately propagating uncertainty from random errors and physical variations to more adequately characterize forecast confidence in a warmer future with reduced-SWE winters. (KEYWORDS: forecast uncertainty, Airborne Snow Observatories, spring precipitation, drought, climate)

INTRODUCTION

The synthesis of snow measurements with runoff models to generate water supply forecasts is a crucial component of the water management paradigm in western North America, where snowpack storage supplements artificial water storage capacities (Dettinger & Anderson 2015). However, forecasts always contain some degree of uncertainty, and water management decisions are characteristically hedged against the possibility of worse-than-expected droughts (Stillinger et al. 2021). As such, robustly quantifying uncertainty in runoff forecasts is of the utmost importance for increasing confidence in water supply management decisions.

While the high-precipitation winters and low-precipitation springs typical of California's climate have historically enabled skillful runoff forecasts in the Sierra Nevada (Pagano et al. 2004), significant discrepancies have been noted between forecast values and observed water yields under increasingly variable hydrologic conditions, particularly in the runoff season of 2021, when surprising forecast overpredictions led to calls for an investigation of "missing" streamflow (Abatzoglou et al. 2021). The California Department of Water Resources (DWR) said of 2021 that historical relationships between snowpack and runoff "fell apart this year" (Sean de Guzman, chief of snow surveys and water supply forecasting for DWR, in Rogers 2021). Was 2021 an exceptionally unlucky year, or is the mismatch between modeled uncertainty ranges and observed variation likely to become a persistent problem in the future? To investigate this question, we employ the results of novel snow monitoring techniques to constrain the majority of uncertainty in the water balance, thereby enabling an analysis of residual hydrological and meteorological uncertainty sources.

Paper presented Western Snow Conference 2022
Correspondence: eli.boardman@nevada.unr.edu
[1]University of Nevada, Reno, Department of Natural Resources and Environmental Science, Reno, NV
[2]Dartmouth College, Department of Earth Sciences, Hanover, NH
[3]McGurk Hydrologic, Orinda, CA
[4]Hetch Hetchy Water and Power, Moccasin, CA
[5]Airborne Snow Observatories, Inc., Mammoth, CA

As anthropogenic climate warming proceeds, temperature-related effects are expected to disproportionately affect mountain regions due to elevation-dependent warming (Pepin et al. 2015), and single-degree temperature increases may lead to reductions in mountain SWE storage of 20% to 40% (Huning & AghaKouchak, 2018). With the potential for multi-year periods of significantly reduced snowpack in the Sierra within the next several decades (Siirila-Woodburn et al. 2021), assumptions of climate stationarity are no longer valid (Milly et al. 2008). Climate and snowpack modeling studies have shown an anticipated decline in the California winter snowpack and across much of western North America (Siirila-Woodburn et al. 2021, Mote et al. 2018, Huning & AghaKouchak 2018), with periods of low-to-no snow winters likely emerging in California within 30 years (Siirila-Woodburn et al., 2021). Therefore, water supply forecasts historically based on SWE observations are likely to experience increasing uncertainty. Specifically, the ability to forecast drought occurrence in mountain watersheds, a key decision-making factor for reservoir operations (Stillinger et al. 2021), is expected to decrease due to climate change impacts on the snowpack (Livneh & Badger 2020).

It is desirable to improve all components of the water supply forecasting system, namely data acquisition, modeling, and uncertainty quantification, to ameliorate the anticipated loss of forecast skill. Considerable effort has been invested in constraining the physical processes driving alpine hydrological responses. The NASA/JPL Airborne Snow Observatory (ASO) project, now Airborne Snow Observatories, Inc., improved the quantification of full-catchment SWE by combining snow depth maps acquired through airborne light detection and ranging (LiDAR) with physical snowpack modeling using iSnobal (Painter et al. 2016). Assimilation of ASO data into snow models resulted in large increases in model skill and suggested the potential for a new paradigm in water management for snowmelt-dominated basins (Hedrick et al. 2018), but the implementation of an ASO-based runoff forecasting paradigm requires the operationalization of hydrological models that can use this novel dataset. Nevertheless, ASO SWE maps influence operational decision-making: for example, an ASO acquisition indicating below-average snowpack in early 2021 was a key factor in the decision of reservoir managers to limit early season water releases.

In light of the availability of novel datasets and the anticipated impacts of climate change, it is important to investigate the physical processes and model assumptions that drive uncertainty in water supply forecasts. In this study, we analyzed uncertainty propagation through a set of statistical runoff models for the upper Tuolumne and Merced River basins, located on the west side of the central California Sierra Nevada mountain range. Flows from the upper Tuolumne River basin are measured at Hetch Hetchy Reservoir, which supplies water and power to the City of San Francisco, and the upper Merced River basin is gauged at the Happy Isles bridge. Both watersheds provide water to agricultural irrigation districts in California's Central Valley, and both have a many-year timeseries of ASO SWE maps, with a combined total of 71 unique SWE maps between the months of January and June. Statistical water supply forecasts for the Tuolumne and Merced Rivers are provided by the California Nevada River Forecast Center (CNRFC), and forecasts are also run by multiple reservoir management groups using a variety of in-house models (e.g., Graham 2018). We use a Bayesian statistical framework to evaluate the likely impacts of drought and climate change on the uncertainty of empirical runoff forecasts for this system across a range of model parameterizations.

Uncertainty in the water balance can arise in two distinct ways: when the magnitude of winter precipitation is large but the fraction remaining as SWE is small, or when the magnitude of winter precipitation is small regardless of the fraction remaining as SWE. In either case, forecasts become more uncertain when the absolute magnitude of the snowpack is small, and the contribution of spring precipitation has the potential to be a large driver of the spring water balance. This result is intuitive and well-established, but our findings present a new opportunity to separately quantify the contribution of different sources of uncertainty to total water supply forecast uncertainty across a range of SWE conditions.

METHODS

To analyze forecast behaviors under a range of hydrologic and climatic conditions, we first built runoff models through which uncertainty could be propagated. The basic structure of most empirical water supply models consists of an expression relating discharge, Q, to precipitation, P, under the influence of one or more additional terms. We generalize this framework to:

$$Q = P - L$$

where L is a placeholder term representing the difference between precipitation and streamflow. The L term includes the lumped effects of evapotranspiration (ET), storage change (ΔS), groundwater fluxes (G), and any other water

cycles active within a catchment. To model cumulative water yield over an extended forecast period of several months, the terms Q, P, and L are further simplified from instantaneous rates to time-integrated volumes, or equivalently, area-normalized specific yields in units of water depth. We define the period of interest as the time interval from October 1 (the beginning of the water year) through July 31 (the end date of most water supply forecasts). In the Sierra Nevada, this time interval includes most of the snowpack accumulation and ablation seasons and typically encompasses the majority of yearly streamflow. We define a water balance on this time interval,

$$L_{Total} \equiv P_{Total} - Q_{Total}$$

such that Q_{Total} is the total water yield measured between October 1 and July 31, P_{Total} is the cumulative precipitation on the same time interval, and L_{Total} is defined as the difference between the former terms. The "loss function" L_{Total} has no explicit physical significance, but is a placeholder for the substitution of other terms representing all hydrological fluxes that are not directly measured. In different cases, the terms constituting L_{Total} may have the same or opposite sign. For instance, a reduction in multi-year storage during a drought could cause L_{Total} to be smaller than ET. We developed six models for L_{Total} to represent several distinct types of empirical runoff model parameterizations (Table 1). For each model, we estimate water balance inputs using the sum of ASO-measured SWE at the start of a given runoff period plus station-measured precipitation during each runoff period.

Model 1 treats L_{Total} as a constant, wherein the difference P-Q is the same each year. This model is assumed to approximate a closed water balance with negligible inter-year storage change and a constant ET demand that can be satisfied even during drought conditions.

Model 2 represents a constant runoff efficiency, wherein the ratio Q/P is constant. This model framework approximates a purely supply-limited closed system where ET increases in proportion to water availability.

Model 3 includes both a slope and intercept for the loss term, with the slope of L_{Total} solely dependent on the current year's precipitation. The architecture of this model still assumes a water balance independent of any multi-year effects but allows ET impacts to vary between the endmember scenarios of Model 1 and Model 2.

Model 4 is the same as Model 3 but with the addition of a second slope term, this time predicated on the previous year's total streamflow as a proxy for antecedent deficit or carryover effects. This framework implies an open water balance within any given year, since L_{Total} is recursively dependent on each prior year's water balance.

Model 5 uses the same variable-loss terms as Model 3 but generalizes the functional shape to a logistic equation. This model's sigmoid shape was chosen to simulate increasing ET in wetter conditions up to some asymptotically approached maximum as the landscape approaches saturation. Additionally, the logistic shape is sufficiently flexible that it is representative of the capabilities of simple nonlinear models in general.

Model 6 is also a logistic function with the addition of last year's streamflow to be a proxy for antecedent conditions as in Model 4. As the most complex model, this structure was intended to capture more nuanced nonlinear behaviors in the framework of an open water balance, with the combined effects of ET and storage change depending on the balance of antecedent storage deficits or surpluses.

Since water supply forecasts are typically made on multiple nested time intervals rather than for the entire October-July period, we develop equations to scale L_{Total} over smaller time windows. Runoff efficiency, R,

$$R \equiv \frac{Q_{Total}}{P_{Total}}$$

can be used to establish a relationship between precipitation and streamflow on shorter intervals if L_{Total} is assumed to scale uniformly across temporal subsets of the water balance.

Table 1. Model equations used in this study. L denotes the difference between cumulative precipitation and runoff.

Model 1: constant loss	$L_1 = a$
Model 2: constant runoff efficiency	$L_2 = a * P_{Total}$
Model 3: linearly variable loss (closed water budget)	$L_3 = a + b * P_{Total}$
Model 4: linearly variable loss (open water budget)	$L_4 = a + b * P_{Total} - c * Q_{Last\ Year}$
Model 5: nonlinearly variable loss (closed water budget)	$L_5 = a + \frac{b}{1 + e^{-c*(P_{Total}-d)}}$
Model 6: nonlinearly variable loss (open water budget)	$L_6 = a + \frac{b}{1 + e^{-c*(P_{Total}-d)-f*(Q_{Last\ Year}-g)}}$

While we acknowledge that the assumption of a constant intra-seasonal runoff efficiency, R, is untested over the early season accumulation period, we note that the ablation season time windows used for calibration and forecasting are comparable, and thus the effect of assuming a time-invariant water balance is minimized in our analysis. Given a constant value of R within a single year, we can calculate cumulative runoff during a given forecasting period as R times a water balance input, which consists of the complete ablation of the snowpack plus any precipitation falling during the spring forecast period, P_{Spring}:

$$Q = R * \left(SWE + P_{Spring}\right)$$

We aggregate 50 m resolution ASO SWE maps (Painter et al. 2019) to average SWE depths over the watershed and calculate cumulative post-flight precipitation using the Tuolumne Meadows meteorological station (TUM) gauge (CDEC 2022). We use all ASO flights in the study basins between the months of January and June (inclusive), resulting in a dataset of 54 flights across 9 years (2013-2021) in the Tuolumne and 17 flights across 6 years (2014-2015 and 2018-2021) in the Merced. TUM precipitation values are consistent with ASO basin-wide SWE accumulation values and the station is at a representative elevation and central location within the combined Tuolumne-Merced study area, so we use TUM data as direct estimate of post-flight P for both basins. Cumulative runoff is then calculated from the day of each ASO flight through the end of July. Since the goal of this project is to characterize variations in uncertainty between different years, we calibrate the runoff models such that we can expect random errors arising from imperfect model fit to approximate a constant fraction of each year's water balance. With the effect of random error thus normalized between years, differences in relative uncertainty can be attributed to variations in the system.

A Bayesian modeling framework is ideal for quantifying and propagating uncertainty from random and systematic sources. A set of parameter samples from a Monte Carlo model fitting algorithm constitutes uncertainty in the mathematical shape of the model and quantifies uncertainty in the model's response to systematic differences in inputs, and the standard deviation, σ, of the distribution from which each sample is drawn quantifies the effect of random process variability, accounting for uncertainty from hydrologic processes that might not be captured by any given model. We calibrate each model by optimizing the parameters of L_{Total} to produce a constant seasonal uncertainty $\sigma * Q_{Total}$ in each forecast period, expressed here in sampling notation:

$$Q_{Total,Observed} \sim normal\left(Q_{Total,Modeled}, \sigma\right)$$

We use the No U-Turn Sampler (NUTS) implementation of Hamiltonian Monte Carlo in the Stan Bayesian package (Stan Development Team 2022) to obtain 2,000 parameter samples (not including warmup) for each of the six models. We use flat priors for all parameters except for logistic function parameters, for which we use diffuse normal priors to prevent numerical instability. Effective sample sizes and traceplots of parameter chains are consistent with convergence.

With model calibration complete, we analyze model behaviors by calculating median values and 90% credible intervals for runoff (Q_{Total}), loss (L_{Total}), and efficiency (R) for each model across the full range of observed P_{Total}. Additionally, we calculate median runoff values for each of the ASO flight-based calibration periods to compare measured and modeled values.

To analyze the runoff models in "forecast mode," that is, only using information that would be available in real-time, we replaced measured values of P_{Spring} with a distribution of likely values that could be propagated through the models to generate forecast runoff exceedance levels. The TUM precipitation record began in water year 1986, so to obtain more than 36 seasonal precipitation samples, it was necessary to build a statistical model of spring precipitation. We calculate the total precipitation occurring from a given day of year through the end of the forecast period for each of the years of record, producing a sparse distribution of spring precipitation patterns. Since precipitation distributions are skewed high by rare large events, we calculate lognormal parameters of the mean (μ_{Spring}) and standard deviation (σ_{Spring}) of the distribution of precipitation occurring after a given date. Sampling from lognormal distributions of P_{Spring} values allows us to iterate the runoff models over many post-forecast precipitation scenarios, thereby propagating uncertainty in P_{Spring} into our final forecast exceedance levels.

To further generalize our forecast analysis, we repeat the same method to characterize lognormal distribution parameters μ_{Winter} and σ_{Winter} for cumulative precipitation between October 1 and the forecast date, giving a distribution of P_{Winter} values. One sample each from the P_{Winter} and P_{Spring} distributions together define a unique seasonal precipitation scenario. Our statistical forecast models only predict runoff and do not have any internal snowpack modeling capabilities, and the amount of P_{Winter} present as SWE on any given forecast date is variable, so the water balance input term SWE + P_{Spring} is undetermined even with known values of P_{Winter} and P_{Spring}. To leverage this innate model flexibility, we defined the SWE fraction X,

$$X \equiv \frac{SWE}{P_{Winter}}$$

which relates the amount of SWE remaining on the day of the forecast to the amount of precipitation up to the day of the forecast. By iterating model runs over a range of X values from 0 (no SWE) to 1 (all P_{Winter} remaining as SWE), we are able to characterize model behaviors across a range of low- to high-SWE proportions within a given precipitation scenario.

To analyze the contribution of spread in the spring precipitation distribution to total forecast uncertainty, we run the runoff models in both "forecast mode" and "backcast mode" for the runoff periods used in model calibration. In contrast to forecast mode, which requires propagation of uncertainty in spring precipitation, a backcast uses only a single value of P_{Spring} which, together with a single value of P_{Winter} and a given value of the SWE fraction X, constitutes a unique sample of the model space.

To generalize these results, we run forecast and backcast modes for synthetic scenarios consisting of precipitation values sampled from P_{Spring} and P_{Winter}. Given a particular model and a single value of X, generalized backcast distributions are created by iterating all 2,000 model parameter sets over 50 spring precipitation samples for each of 50 winter precipitation samples. This results in 2,500 unique backcast distributions, each containing 2,000 samples representing uncertainty in a given hydrological model for a given precipitation scenario. Similarly, we create forecast distributions by iterating the model parameter sets over the same 2,500 winter-spring precipitation combinations representing unique "years," albeit with an additional round of iteration over 50 new spring precipitation samples which are propagated through the model in place of the single P_{Spring} value used by the backcast. Therefore, we obtain 2500 forecast distributions corresponding to the backcast distributions, but each forecast distribution contains 100,000 samples (2,000 parameter samples times 50 P_{Spring} samples), thereby representing uncertainty in the simulated real-time forecast. For both backcast and forecast scenarios, models requiring last year's streamflow as an input variable are given a one-year warm-up period under static precipitation conditions with an initial streamflow value representative of average conditions.

It is impossible to directly validate model results in randomly sampled scenarios since observed runoff values are not available across all possible conditions. However, since the propagation of random and systematic variability in the Bayesian framework can quantify uncertainty even outside the calibration range, the characteristics

of backcast and forecast distributions nevertheless illumine the uncertainty in their respective scenarios. We attained an estimate of expected forecast accuracy by finding the average relative deviation around the median:

$$Expected\ Accuracy \equiv \sum_{i=1}^{n} \left(1 - \left[\frac{|Q_i - Q_{Median}|}{Q_{Median}}\right]\right)\frac{1}{n}$$

Here, Q_{Median} is the median of either a backcast or forecast distribution and represents the central tendency of the model. The spread of the distribution around its median characterizes the uncertainty, with each sample Q_i representing a single possible value that could be the "true" Q. Therefore, the term inside the square brackets represents the relative error between the constant modeled value, Q_{Median}, and a particular possible true value, Q_i. Subtracting the relative error from one yields expected accuracy.

The methods yield an "uncertainty function," which quantifies the uncertainty in statistical runoff forecasts under arbitrary precipitation and SWE conditions. By way of example, we pick April 1 as a forecast date due to the prevalence of April through July water supply forecasts in operational environments. We then calculate the expected accuracy of backcast and forecast distributions for each model across a range of SWE fractions, X, from 0% to 100% SWE remaining on April 1. Finally, we repeat these calculations excluding all but the driest 25% of P_{Winter} and P_{Spring} samples to simulate forecast uncertainty during drought years.

RESULTS

Bayesian sampling of all six models for both the Tuolumne and Merced River basins yields a range of parameter values representing unique individual models. The median, 10th, and 90th percentile runoff values are calculated for the range of SWE + P from 0.4 m to 1.5 m water depth equivalent, representative of the range of variation between drought and abnormally wet conditions observed over the calibration period of 2013-2021. Calculating R over this same range for both the median and the credible interval yields an estimate of runoff efficiency behaviors under various hydrological conditions. In both basins, all models indicate a nonlinear decay of runoff efficiency during drought conditions (Figure 1) with the exception of Model 2, where runoff efficiency is held constant by design.

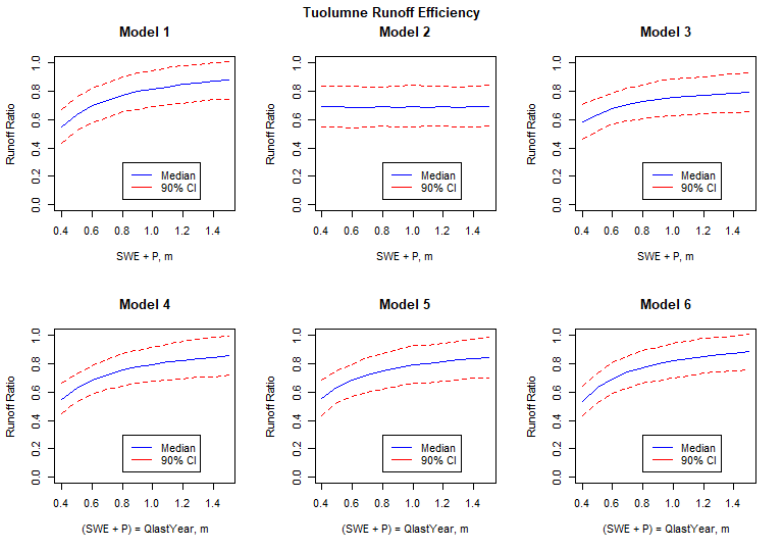

Figure 1. Modeled runoff efficiency behavior in Tuolumne Basin.

We evaluate model performance by comparing measured and modeled values for the runoff periods associated with each ASO flight. Scatter plots of modeled and observed values (Figure 2) indicate minimal bias between high or low magnitudes since the individual points are typically distributed uniformly around the 1:1 line. The exception is again Model 2, which showed a strong bias to underpredicting runoff in wet years. This behavior can be attributed to the model's constant runoff efficiency, which prevents it from capturing the increase in runoff efficiency associated with wetter conditions.

We choose to focus on results pertaining to Model 6 for the present discussion since this model was the most successful at capturing observed variations between drought years. The two water years 2020 and 2021 were particularly useful test cases: the study area received approximately the same amount of precipitation (with spatiotemporally similar ASO SWE distributions) in both years, but 2020 followed an unusually wet year while 2021 was the second year of a sustained drought. As such, this pair of years is useful for testing the importance of antecedent storage or deficit effects. We observe that Model 6 (and to a lesser degree, Model 4) best capture the variation between these years, likely due to the inclusion of the previous year's streamflow as a proxy for antecedent moisture conditions.

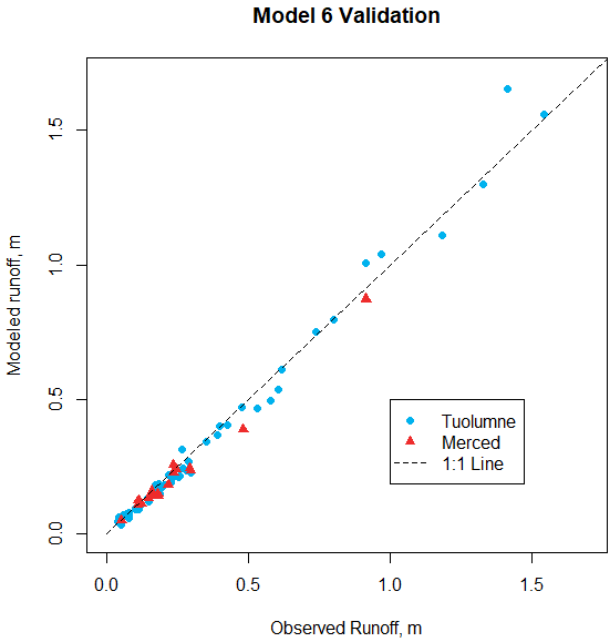

Figure 2. Comparison of measured and modeled runoff values for Model 6.

Running the models in forecast mode requires sampling from precipitation distributions. We observe similarity in the shapes of both P_{Winter} and P_{Spring} distributions with corresponding observed distributions over the past 36 years (Figure 3). We do not suggest that the precipitation distributions derived here should be used for operational forecasting; rather, these distributions are meant to illustrate the range of likely variations in pre- and post-forecast-date precipitation. In the operational environment, short-term and long-term meteorological forecasts are likely to supplement historical distributions to further constrain uncertainty in precipitation.

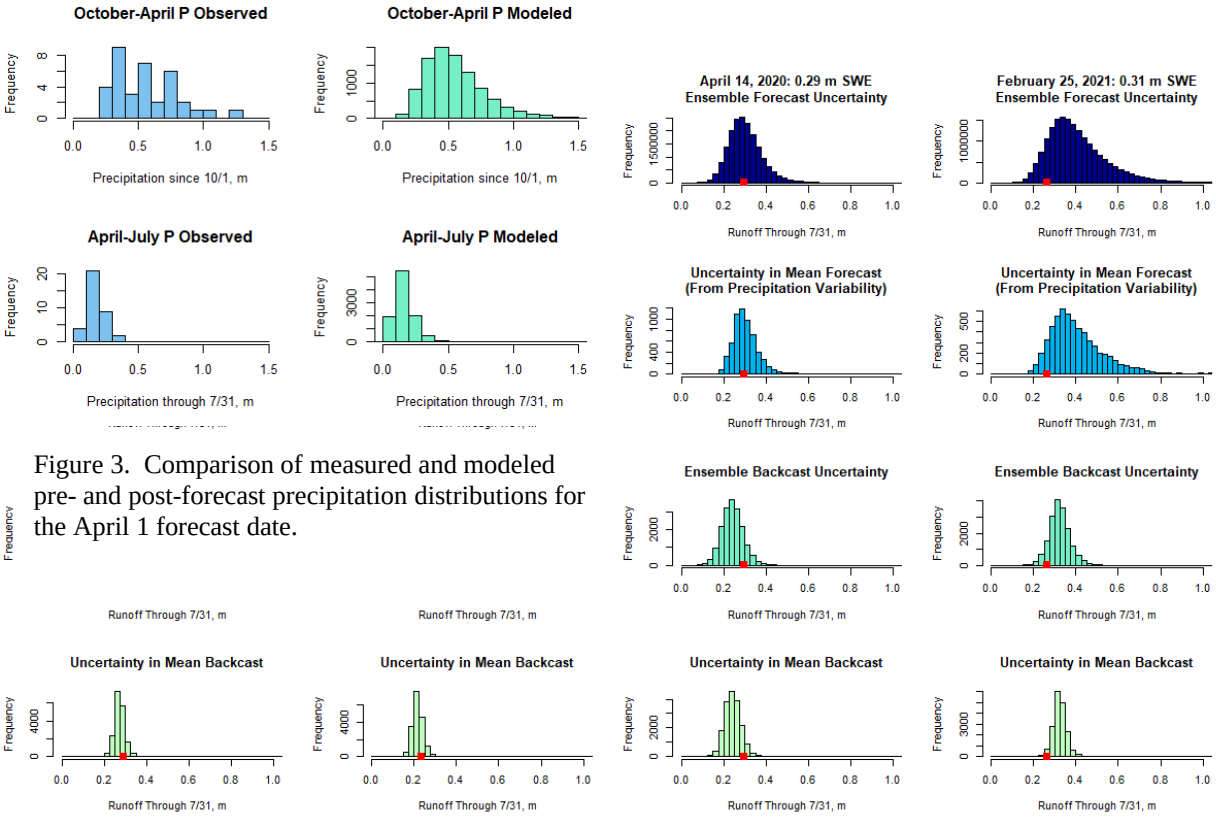

Figure 3. Comparison of measured and modeled pre- and post-forecast precipitation distributions for the April 1 forecast date.

To test the ability of backcast and forecast models to capture observed runoff variability during drought conditions, we also generate post-forecast precipitation distributions for the day-of-year corresponding to a selection of three ASO flights selected from the drought years of 2015, 2020, and 2021. Histograms of the forecast and backcast distributions (Figure 4) for an ensemble of all six models indicate that the models successfully capture the measured value and illustrate the tightening of the distribution in backcast mode when the value of P_{Spring} is known. In addition to the ensemble forecast and backcast distributions, we estimate ensemble distributions in forecast mode and backcast mode propagating only parameter uncertainty and neglecting the addition of random process variability, essentially modeling the uncertainty in the mean forecast and the mean backcast. We note that in forecast mode, the contribution of process variability is small compared with the large uncertainty introduced by propagating the spring precipitation distribution, while in backcast mode, substantial uncertainty is contributed by both random error and parameter uncertainty.

Generalizing our forecast and backcast analysis to sample from both P_{Winter} and P_{Spring} and sampling across the full range of X (from 0 to 100% SWE remaining), we observe that the hydrological uncertainty of Model 6, that is, the uncertainty remaining in the runoff model even when P_{Spring} is known, increases from ± 10% (90% accuracy) to ± 20% (80% accuracy) as the SWE fraction decreases from 100% to 0%. When P_{Spring} is sampled from the precipitation distribution in forecast mode, the overall uncertainty increases, and the reduction in accuracy under low-SWE conditions is substantially more pronounced. With 100% SWE remaining on April 1, the forecast model indicates ± 20% uncertainty, including the ± 10% contributed by uncertainty in the runoff model itself, suggesting that hydrological and meteorological uncertainty sources are equally important. However, with reduced SWE on April 1, the contribution of meteorological uncertainty increases nonlinearly, and in a hypothetical zero-SWE scenario, the total forecast uncertainty is greater than ± 60%. This indicates that meteorological uncertainty is about twice as important as hydrological uncertainty in this extreme, but not impossible in the 21st century, scenario.

We observe a similar pattern of increasing runoff uncertainty under reduced-SWE conditions when we repeat the analysis using only the 25% driest precipitation scenarios. Hydrological uncertainty increases slightly, but the most notable change is the increase in total forecast uncertainty at the high-SWE end when considering only drought scenarios. Our results indicate that expected forecast accuracy decreases by ~10-15% during droughts even when most winter precipitation remains as SWE on the forecast date.

DISCUSSION AND CONCLUSION

Of the six tested models, those that include a constant or near-constant loss term perform significantly better at capturing observed variations compared with the constant runoff efficiency model, consistent with Hedrick et al. (2020), who found that evapotranspiration losses are relatively constant in the Tuolumne across a sample of dry, average, and wet years. The decoupling of plant water consumption from precipitation variability has been observed in other regions of California (Hahm et al. 2019), and we suggest that our similar finding in the Tuolumne and Merced watersheds could be indicative of drought-limited ecosystem adaptation, which results in relatively constant water loss across a wide range of wet and dry conditions. Our results indicate the relationship between precipitation and runoff in the upper Tuolumne and Merced Basins is better modeled as a constant or nearly constant difference than as a constant or nearly constant ratio.

With regard to the propagation of hydrological and meteorological uncertainties through runoff forecasts, we observe the expected result that runoff forecasts experience greater uncertainty when the seasonal water balance is poorly constrained. In particular, we draw attention to the comparison of forecast uncertainty between all sampled precipitation scenarios (Figure 5a) and the subset of drought scenarios (Figure 5b). The average forecast uncertainty in a drought year with 100% of winter precipitation remaining as SWE on April 1 is comparable to the expected forecast uncertainty for all precipitation conditions when only 30% of winter precipitation remains as SWE. We suggest that the behavior of forecast uncertainty during droughts is an analogue for the behavior of forecast uncertainty in a reduced-SWE climate.

Our results support the findings of Livneh and Badger (2020) by quantifying increased forecast uncertainty in reduced-SWE conditions at the watershed scale. Moreover, we argue that high water supply uncertainty during current droughts foreshadows high uncertainty in future low-SWE climate scenarios. Climate change is likely to exacerbate the challenges associated with adequately modeling runoff uncertainty. We suggest that the development

of forecasting approaches that rigorously quantify and hopefully reduce uncertainty in out-of-sample conditions should be an area of future investigation.

REFERENCES

Abatzoglou, J., Rallings, A., Bernacchi, L., Viers, J., & Medellin-Azuara, J. (2021). California's Missing Forecast Flows in Spring 2021 – Challenges for seasonal flow forecasting. California WaterBlog. https://californiawaterblog.com/2021/07/18/californias-missing-forecast-flows-in-spring-2021-challenges-for-seasonal-flow-forecasting/

California Data Exchange Center (CDEC). (2022). Tuolumne Meadows historical data. California Department of Water Resources/DFM-Hydro-SMN. https://cdec.water.ca.gov/dynamicapp/staMeta?station_id=tum

Dettinger, M. D., & Anderson, M. L. (2015). Storage in california's reservois and snowpack in this time of drought. *San Francisco Estuary and Watershed Science*, *13*(2). https://doi.org/10.15447/sfews.2015v13iss2art1

Graham, C. (2018). *Airborne Snow Observatory in the Tuolumne River Basin: Forecasting streamflow using basin wide SWE estimates*. Yosemite Hydroclimate Meeting. https://www.cafiresci.org/events-webinars-source/category/yosemite-hydroclimate-meeting-2018

Hahm, W. J., Dralle, D. N., Rempe, D. M., Bryk, A. B., Thompson, S. E., Dawson, T. E., & Dietrich, W. E. (2019). Low subsurface water storage capacity relative to annual rainfall decouples mediterranean plant productivity and water use from rainfall variability. *Geophysical Research Letters*, *46*(12), 6544–6553. https://doi.org/10.1029/2019GL083294

Hedrick, A. R., Marks, D., Havens, S., Robertson, M., Johnson, M., Sandusky, M., Marshall, H., Kormos, P. R., Bormann, K. J., & Painter, T. H. (2018). Direct insertion of nasa airborne snow observatory-derived snow depth time series into the *isnobal* energy balance snow model. *Water Resources Research*, *54*(10), 8045–8063. https://doi.org/10.1029/2018WR023190

Hedrick, A. R., Marks, D., Marshall, H., McNamara, J., Havens, S., Trujillo, E., Sandusky, M., Robertson, M., Johnson, M., Bormann, K. J., & Painter, T. H. (2020). From drought to flood: A water balance analysis of the tuolumne river basin during extreme conditions(2015 – 2017). *Hydrological Processes*, hyp.13749. https://doi.org/10.1002/hyp.13749

Huning, L. S., & AghaKouchak, A. (2018). Mountain snowpack response to different levels of warming. *Proceedings of the National Academy of Sciences*, *115*(43), 10932–10937. https://doi.org/10.1073/pnas.1805953115

Livneh, B., & Badger, A. M. (2020). Drought less predictable under declining future snowpack. *Nature Climate Change*, *10*(5), 452–458. https://doi.org/10.1038/s41558-020-0754-8

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity is dead: Whither water management? *Science*, *319*(5863), 573–574. https://doi.org/10.1126/science.1151915

Mote, P. W., Li, S., Lettenmaier, D. P., Xiao, M., & Engel, R. (2018). Dramatic declines in snowpack in the western US. *Npj Climate and Atmospheric Science*, *1*(1), 2. https://doi.org/10.1038/s41612-018-0012-1

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

Painter, Thomas H., Berisford, Daniel F., Boardman, Joseph W., Bormann, Kathryn J., Deems, Jeffrey S., Gehrke, Frank, Hedrick, Andrew, Joyce, Michael, Laidlaw, Ross, Marks, Danny, Mattmann, Chris, McGurk, Bruce, Ramirez, Paul, Richardson, Megan, & Skiles, S. McKenzie. (2018). *ASO lidar snow*

water equivalent 50m UTM grid, version 1. NASA National Snow and Ice Data Center DAAC. https://doi.org/10.5067/M4TUH28NHL4Z

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., Hedrick, A., Joyce, M., Laidlaw, R., Marks, D., Mattmann, C., McGurk, B., Ramirez, P., Richardson, M., Skiles, S. M., Seidel, F. C., & Winstral, A. (2016). The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. *Remote Sensing of Environment*, *184*, 139–152. https://doi.org/10.1016/j.rse.2016.06.018

Pepin, N., Bradley, R. S., Diaz, H. F., Baraer, M., Caceres, E. B., Forsythe, N., Fowler, H., Greenwood, G., Hashmi, M. Z., Liu, X. D., Miller, J. R., Ning, L., Ohmura, A., Palazzi, E., Rangwala, I., Schöner, W., Severskiy, I., Shahgedanova, M., Wang, M. B., … Mountain Research Initiative EDW Working Group. (2015). Elevation-dependent warming in mountain regions of the world. *Nature Climate Change*, *5*(5), 424–430. https://doi.org/10.1038/nclimate2563

Stan Development Team. (2022). Stan Modeling Language Users Guide and Reference Manual, 2.29. https://mc-stan.org

Rogers, P. (2021). Where did Sierra snow go this spring? Not into California rivers and water supplies. Phys.Org; Mercury News. https://phys.org/news/2021-06-sierra-california-rivers.html

Siirila-Woodburn, E. R., Rhoades, A. M., Hatchett, B. J., Huning, L. S., Szinai, J., Tague, C., Nico, P. S., Feldman, D. R., Jones, A. D., Collins, W. D., & Kaatz, L. (2021). A low-to-no snow future and its impacts on water resources in the western United States. *Nature Reviews Earth & Environment*, *2*(11), 800–819. https://doi.org/10.1038/s43017-021-00219-y

Stillinger, T., Costello, C., Bales, R. C., & Dozier, J. (2021). Reservoir operators react to uncertainty in snowmelt runoff forecasts. *Journal of Water Resources Planning and Management*, *147*(10), 06021010. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001437