University of Nevada, Reno

SEASONAL WATER SUPPLY FORECASTING IN THE WESTERN U.S. UNDER DECLINING SNOWPACK

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science in Hydrology

by Joshua T. Sturtevant

Dr. Adrian A. Harpold/Thesis Advisor

May 2020

THE GRADUATE SCHOOL

We recommend that the thesis
prepared under our supervision by

JOSHUA STURTEVANT

entitled

**Seasonal Water Supply Forecasting in the Western U.S.
Under Declining Snowpack**

be accepted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

Adrian Harpold, Ph.D.
Advisor

Michael Dettinger, Ph.D.
Committee Member

Seshadri Rajagopal, Ph.D.
Committee Member

Stephanie McAfee, Ph.D.
Graduate School Representative

David W. Zeh, Ph.D., Dean
Graduate School

May 2020

Abstract

Seasonal water supply forecasts (i.e. April-July streamflow volume predictions) are a key decision-support tool for water resource managers overseeing reservoir operations, water rights allocation, and ecosystem and wildlife protection, particularly in semi-arid regions such as the western U.S. However, extremely low snowpack and non-stationarity due to climate change challenge the empirical relationships and computer simulations used by operational hydrologic forecasters. In this thesis, we present findings from two studies that use empirical and simulated data to recreate these operational models across two separate study domains: 51 minimally-impacted, snow-dominated basins in the western U.S. and 26 headwater basins in the Sierra Nevada of California and Nevada. In the first study, we use a model benchmarking approach to evaluate and compare the retrospective performance of two of the most widely used water supply forecasting models in the western U.S., the USDA Natural Resource Conservation Service's (NRCS) Principal Component Regression models and the NOAA National Weather Service's (NWS) SNOW17/Sacramento Soil Moisture Accounting models. Results from this retrospective analysis across water years 1981-2014 show that statistical models (NRCS PCR) are generally more skillful than conceptual models (NWS SNOW17/SAC-SMA). However, at longer forecast lead times, are results suggest that statistical models are more disadvantage by low snowpack, making them comparatively less skillful than the conceptual models. In our second study, we extend this analysis by using hydrologic simulations from the Variable Infiltration Capacity model to mimic California Department of Water Resources water supply forecasting procedures in the

Sierra Nevada from 1950-2099 under low and high emissions scenarios. Results show widespread but uneven loss of skill by the second half of the 21st century (average loss 10-20%). Basins at mid-elevations (mean elevation 1000-1700 meters) were simulated to be most vulnerable to loss of snowpack and the forecast information it provides. Within the model environment, we then evaluate mitigation strategies to buffer loss of forecast skill through the introduction of supplemental synthetic observations. These two simulated datasets include 1) basin-wide snowpack measurements (representing remote sensing products) and 2) soil moisture station observations. Our research suggests that remotely-sensed snowpack data may buffer loss of skill by an average of 40% in the most vulnerable basins before 2050. As the century passes and the role of snow in Sierra Nevada hydrology declines, there will eventually be dwindling returns from remotely-sensed snowpack data, but this loss of forecast information may be somewhat ameliorated by the inclusion of soil moisture observations in the regression equations.

Acknowledgements

I would like to thank my advisor, Adrian Harpold, for his mentorship and guidance, as well as my committee members, Mike Dettinger, Stephanie McAfee, and Seshadri Rajagopal, for their dedicated involvement as a project team. Special thanks to researchers at the National Center for Atmospheric Research, Andy Wood and Andy Newman, without whom the model benchmarking efforts of Chapter 2 would not have been possible. This thesis also benefitted from conversations with hydrologic forecasting practitioners including Gus Goodbody and Jolyne Lea at the National Water and Climate Center, as well as others from the California-Nevada and Colorado Basin River Forecast Centers.

Additional thanks to our funding agencies: the Southwest Climate Adaptation Science Center and the Nevada NASA Space Grant.

And last but certainly not least, thank you to my family and friends for their love and support, but especially for their encouragement to return to the sciences.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Seasonal water supply forecasts (i.e. April-July streamflow volume predictions) are a key decision-support tool for water users and managers in semi-arid regions, such as the Western U.S. Seasonal water supply forecasts, also referred to as Water Supply Outlooks, are issued throughout the winter season to predict snowmelt-driven runoff from headwater basins. Responsibility for issuing such forecasts is shared by two federal agencies, the U.S. Department of Agriculture (USDA) Natural Resource Conservation Service (NRCS) (https://www.wcc.nrcs.usda.gov/) and the National Oceanic and Atmospheric Administration (NOAA) National Weather Service (NWS) River Forecast Centers (RFCs) (e.g. https://www.cnrfc.noaa.gov/). Similar seasonal water supply forecasts are independently issued by the California Department of Water Resources for Sierra Nevada headwater basins (https://cdec.water.ca.gov/snow/bulletin120/). For nearly a century, these operational hydrologic forecasts have relied on the high predictive power of winter snowpack to estimate spring runoff (Church, 1935). However, extremely below-average snowpack, or snow drought, and non-stationarity due to climate change challenge the empirical relationships and computer simulations used by operational hydrologic forecasters (Milly et al., 2008), raising questions about the skillfulness of these models in predicting spring streamflow volumes during low snow years.

Mountain snowpack serves as a natural reservoir which bridges the divide between when the majority of precipitation falls (winter) and when water demand is highest (summer). In addition to providing water storage, snowpack is also critical for efficient streamflow generation (Mcnamara et al., 2005), groundwater recharge (Safeeq et

al., 2013), and maintaining summer low flows (Godsey et al., 2013). Interannual to interdecadal climate variability can drive natural occurrences of below-average snowpack from meteorological drought or above-average temperatures (Cayan, 1996), which have been defined as dry snow drought and warm snow drought, respectively (A. Harpold et al., 2017). Yet, an observed 15-30% decline in western U.S. snowpack since the mid 20th century (Mote et al., 2018) has been attributed to increased temperatures and changing precipitation patterns driven by anthropogenic climate change (Pierce et al., 2008). In the western U.S., these smaller snowpacks are melting earlier and slower as a consequence of warming temperatures (A. Harpold et al., 2012; A. A. Harpold & Kohler, 2017; Musselman et al., 2017), which have important implications for streamflow generation (Mcnamara et al., 2005) and late-season water availability (Stewart et al., 2005). Regardless of mechanism, below-average snowpack degrades the predictive power for spring snowmelt-driven runoff, making summer drought conditions more challenging to predict (Livneh & Badger, 2020) and threatening historically accurate spring streamflow forecasting methods.

In this thesis, we evaluate the performance of two of the most commonly used hydrologic models used in operational water supply forecasting: regression-based empirical models and more physically-motivated computer simulated rainfall-runoff models, referred to as conceptual models. In Chapter 2, we present an historical analysis of these two models in which we retrospectively evaluate their relative performance during above and below-average snow years. Specifically, we performed a model benchmark analysis (Newman et al., 2017) in which we compared the performance of the more complex RFCs Sacramento Soil Moisture Accounting (SAC-SMA) model (Burnash

et al., 1973) to the a priori benchmark performance of the longer-running and better established NRCS Principal Component Regression (PCR) models (Garen, 1992). Chapter 2 leverages the recent availability of large meteorological datasets (Addor et al., 2017) and automated U.S. Geological Survey (USGS) stream gauging and NRCS snow telemetry networks to retrospectively recreate and evaluate these two classes of models for a large subset of over fifty basins (forecast points) across the western U.S. for multiple forecast lead times during more than thirty years. Our research questions for Chapter 2 are: 1) Are seasonal water supply forecasts degraded under conditions of below-average snow accumulation (i.e. snow drought)? And, 2) If forecasts are less skillful in low snow years, what are the implications for the skill of statistical and conceptual forecast models faced with future snow water equivalent (SWE) decline? We answer these questions by considering interannual forecast performance of the two models during low snow years and by evaluating longer term changes in forecast skill across the water years 1981-2014 period of record.

In Chapter 3, we turn our focus to the jurisdiction of the California Department of Water Resources (CA DWR) where we take a hybridized approach that runs regression-based models with simulated data as a means to evaluate forecast skill through the 21st century. In this strictly model-based analysis, we use Variable Infiltration Capacity (VIC) model simulations forced by down-scaled climate projections under two emission scenarios to simulate and then evaluate CA DWR seasonal water supply forecasts for water years 1951-2099 for 26 headwater basins in the Sierra Nevada, USA. Within this model environment, we then evaluate two strategies for mitigating the loss of forecast

skill using simulated data to act as surrogates of: 1) remotely-sensed, basin-wide snowpack measurements and 2) soil moisture observations. We use this framework to quantify vulnerability of conventional water supply forecasts to loss of skill under declining snowpack and to evaluate operationally feasible forecast mitigation techniques.

Chapter 2 provides an increased understanding of relative operational hydrologic model capabilities during below-average snowpack during an historical period, while Chapter 3 contextualizes these findings within the longer term threat that declining snowpack presents to conventional water supply forecasting techniques. The latter half of Chapter 3 uses the flexibility of simulated hydrometeorology to develop surrogate methods for buffering loss of skill with the objective of providing recommendations to operational forecasting and government agencies. Finally, Chapter 4 highlights the broader implications of our research to the operational hydrologic modeling community and discusses future research directions.

References

Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: catchment attributes and meteorology for large-sample studies. Hydrol. Earth Syst. Sci, 21, 5293–5313. https://doi.org/10.5194/hess-21-5293-2017

Burnash, R. J. C., Ferral, R. L., & McGuire, R. A. (1973). A Generalized Streamflow Simulation System: Conceptual Modeling for Digital ... - Robert J. C. Burnash, R. Larry Ferral, Robert A. McGuire - Google Books. Retrieved from https://books.google.com/books?hl=en&lr=&id=aQJDAAAAIAAJ&oi=fnd&pg=PR2&dq=A+generalized+streamflow+simulation+system+-+conceptual+modeling+for+digital+computers&ots=4sWc_i8daw&sig=hTA89Z7zHWhqB78bzvBYCmrepYA#v=onepage&q=A generalized streamflow simulation system - conceptual modeling for digital computers&f=false

Cayan, D. R. (1996). Interannual Climate Variability and Snowpack in the Western United States. Journal of Climate, 9(5), 928–948. https://doi.org/10.1175/1520-0442(1996)009<0928:ICVASI>2.0.CO;2

Church, J. E. (1935). Principles of Snow Surveying as Applied to Forecasting Stream Flow. Journal of Agricultural Research, 51(2), 97–130. Retrieved from https://books.google.com/books?hl=en&lr=&id=-EFSAQAAMAAJ&oi=fnd&pg=PA97&dq=james+church+1935&ots=26d2VJVWAT&sig=uV6mZVW8JbxYb_PKMza4p0TAKFg#v=onepage&q=james church 1935&f=false

Garen, D. C. (1992). Improved Techniques in Regression-Based Streamflow Volume Forecasting. Journal of Water Resource Planning Management, 118(6), 654–670.

Godsey, S. E., Kirchner, J. W., & Tague, C. L. (2013). Effects of changes in winter snowpacks on summer low flows: case studies in the Sierra Nevada, California, USA. https://doi.org/10.1002/hyp.9943

Harpold, A., Brooks, P., Rajagopal, S., Heidbuchel, I., Jardine, A., & Stielstra, C. (2012). Changes in snowpack accumulation and ablation in the intermountain west, 48, 11501. https://doi.org/10.1029/2012WR011949

Harpold, A., Dettinger, M., & Rajagopal, S. (2017). Defining Snow Drought and Why It Matters. Eos, 98. https://doi.org/10.1029/2017EO068775

Harpold, A. A., & Kohler, M. (2017). Potential for Changing Extreme Snowmelt and Rainfall Events in the Mountains of the Western United States. Journal of Geophysical Research: Atmospheres, 122(24), 13,219-13,228. https://doi.org/10.1002/2017JD027704

Livneh, B., & Badger, A. M. (2020). Drought less predictable under declining future snowpack. Nature Climate Change, 1–7. https://doi.org/10.1038/s41558-020-0754-8

Mcnamara, J. P., Chandler, D., Seyfried, M., & Achet, S. (2005). Soil moisture states, lateral flow, and streamflow generation in a semi-arid, snowmelt-driven catchment. Process, 19, 4023–4038. https://doi.org/10.1002/hyp.5869

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., et al. (2008). Climate change. Stationarity is dead: whither water management? Science (New York, N.Y.), 319(5863), 573–4. 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

Musselman, K. N., Clark, M. P., Liu, C., Ikeda, K., & Rasmussen, R. (2017). Slower snowmelt in a warmer world. Nature Climate Change, 7(3), 214–219. https://doi.org/10.1038/nclimate3225

Newman, A. J., Mizukami, N., Clark, M. P., Wood, A. W., & Nijssen, B. (2017). Benchmarking of a Physically Based Hydrologic Model. Journal of Hydrometeorology, 18, 2215–2225. https://doi.org/10.1175/JHM-D-16-0284.s1

Pierce, D. W., Barnett, T. P., Hidalgo, H. G., Das, T., Bonfils, C., Santer, B. D., et al. (2008). Attribution of Declining Western U.S. Snowpack to Human Effects. https://doi.org/10.1175/2008JCLI2405.1

Safeeq, M., Grant, G. E., Lewis, S. L., & Tague, C. L. (2013). Coupling snowpack and groundwater dynamics to interpret historical streamflow trends in the western United States, 27, 655–668. https://doi.org/10.1002/hyp.9628

Stewart, I. T., Cayan, D. R., Dettinger, M. D., Stewart, I. T., Cayan, D. R., & Dettinger, M. D. (2005). Changes toward Earlier Streamflow Timing across Western North America. Journal of Climate, 18(8), 1136–1155. https://doi.org/10.1175/JCLI3321.1

Chapter 2: Benchmarking of Statistical and Conceptual Models for Seasonal Water Supply Forecasts: Implications under Declining Western U.S. Snowpacks

Abstract

Seasonal water supply forecasts are a key decision-support tool for semi-arid regions, such as the western U.S. Snow drought and declining snowpack have negative implications for water supplies, including earlier, slower, and smaller snowmelt that impacts runoff efficiency and groundwater recharge. However, the empirical relationships and computer simulations used by operational forecasters remain poorly evaluated under low snowpack conditions. We ask two questions critical to water supply managers: 1) Do lower snowpack years affect forecast skill? and 2) What are the implications for the skill of empirical and physical forecast models as snow declines in the future? We compare the empirical forecasting techniques of the USDA's Natural Resource Conservation Service (i.e. Principal Component Regression, PCR) with NOAA'S National Weather Service (i.e. Sacramento Soil Moisture Accounting model, SAC-SMA) across 51 snow-dominated basins in the western U.S. We analyze 0- to 3-month lead time forecasts for April-July streamflow volumes to understand how a warming climate may degrade forecast skill into the future. April 1st empirical (PCR) forecasts are only 40% as skillful during below-average snow years versus above-average snow years. Physically motivated models (SAC-SMA) regionally offer more skill than PCR forecasts in January and February (3- and 2- month lead, respectively), especially during dry years, but are less skillful than PCR at shorter lead times. Spring precipitation remains a significant source of uncertainty in both forecasts that highlights the value of

improved seasonal to sub-seasonal precipitation forecasts. Benchmarking two operational

models highlight the potential to adaptively use information to best match local

challenges that mitigate the effects of snowpack loss on forecast skill and uncertainty.

2.1 Introduction

Seasonal water supply forecasts are a critical decision-support tool for water

resource managers overseeing reservoir operations, water rights allocation, and

ecosystem and wildlife protection (Mantua et al., 2008). In semi-arid places with high

downstream water demand, like the western U.S. (Bales et al., 2006; Barnett et al., 2005),

operational forecasters use both hydrologic computer model simulations and statistical

regression models based on station observations to produce water supply outlooks (i.e.

seasonal streamflow volumes, usually for April through July) at several months lead

times for the high demand summer months. In snow dominated basins, the robust

relationship between peak snow water equivalent (SWE) and spring streamflow volumes

(Church, 1935) has provided historically accurate forecasting in a water stressed region

(Pagano et al., 2004). However, widespread declines in SWE in the western U.S. (Mote et

al., 2018) are changing the dynamics of snowmelt (Barnhart et al., 2016; A. A. Harpold

& Kohler, 2017; Adrian A Harpold & Brooks, 2018) with consequences for groundwater

recharge (Berghuijs et al., 2014) and streamflow generation (Hammond et al., 2019;

Mcnamara et al., 2005). This shift in snow dynamics challenges the stationarity

assumptions used by operational forecasting models in which the past can predict the

future (Milly et al., 2008, 2015).

Seasonal water supply forecasts in the western U.S. are primarily made by two

agencies: U.S. Department of Agriculture Natural Resource Conservation Service (NRCS) and the National Oceanic and Atmospheric Administration National Weather Service (NWS). Although inter-agency forecast coordination used to occur, the existing forecasts come from independently maintained programs (Pagano et al., 2014). The longer-running and better established NRCS Snow Survey & Water Supply Forecasting Program (https://www.wcc.nrcs.usda.gov/wsf/) leverages an extensive network of snow telemetry (SNOTEL) stations and Principal Component Regression (PCR) techniques to predict April to July streamflow volumes for over 600 forecast points (Garen, 1992). In comparison, the NWS River Forecast Centers (RFCs) have developed and implemented a more complex, physically-motivated rainfall-runoff model (the Sacramento Soil Moisture Accounting, or SAC-SMA, model; also referred to hereafter as a "conceptual model") intended for short lead (i.e. hours to days) flood forecasting but additionally used for seasonal water supply forecasts (e.g. https://www.cnrfc.noaa.gov/) for roughly 900 locations in the western U.S. Previous work benchmarking conceptual (SAC-SMA) and physics-based (Variable Infiltration Capacity) models by Newman et al. (2017) provides a framework for comparing PCR and SAC-SMA water supply models. Despite some of the potential advantages of a more physically-based technique (e.g. explicit parameter representation), the NRCS forecasts have remained an important and highly valued tool for water supply forecasting, particularly for small, snow-dominated headwater basins with nearby observation stations and long periods of record. Despite advances in distributed, physics-based hydrologic models from the research community (Koster et al., 2010), there is little adoption of these models into operational forecasting due largely to unproven potential gains in accuracy when compared to the costs of additional

complexity (Beven, 2002) and decreased model agility (Mendoza et al., 2015). It is important to compare the strengths and weaknesses of operational hydrologic models in light of climate changes that will reduce the forecast skill associated with snow water storage.

Three primary factors are hypothesized to control the future of seasonal-streamflow forecastability in snow-fed regions like the western U.S.: 1) vulnerability to snow loss, 2) sufficiency of model forcing data, and 3) adequacy of model process representation, including parameterization (Figure 2.1). We predict that snow loss has been and will continue to be a primary threat to forecast skill. Since mid 20th century, average declines in April 1 SWE across the western U.S. have been estimated at 15-30% (Mote et al., 2018). This loss of snowpack storage is expected to increase the dependence on other observations (e.g. precipitation) and other stores and fluxes of water by statistical and conceptual models. Long term snowpack decline can arise from a variety of causes (Kapnick & Hall, 2012; Mote, 2003; Pederson et al., 2011). For example, increases in temperature and humidity affect snow accumulation and ablation, particularly in warmer regions and at lower elevations such as the Cascades or northern Sierra Nevada (Hamlet et al., 2005; Mote, 2006; Regonda et al., 2005). In contrast, snowpack in cold, dry winter climates such as the Northern Rockies have been more buffered from these temperature-driven losses (Mote et al., 2018), perhaps translating to a more robust seasonal water supply forecast during that period (Figure 2.1, solid lines). Interannual to interdecadal climate variability can also lead to well-below-average snowpack (Cayan, 1996), also referred to as snow drought (A. Harpold et al., 2017).

Snow drought can arise from both winter precipitation deficits (dry snow droughts) and from increased fraction of winter precipitation falling as rain (Knowles et al., 2006), as well as mid-winter ablation events (A. Harpold et al., 2012). Changes in snowpack from long term declines and snow drought alter the timing and magnitude of streamflow (Stewart et al., 2005) with significant implications for reservoir operations (A. Harpold et al., 2017).
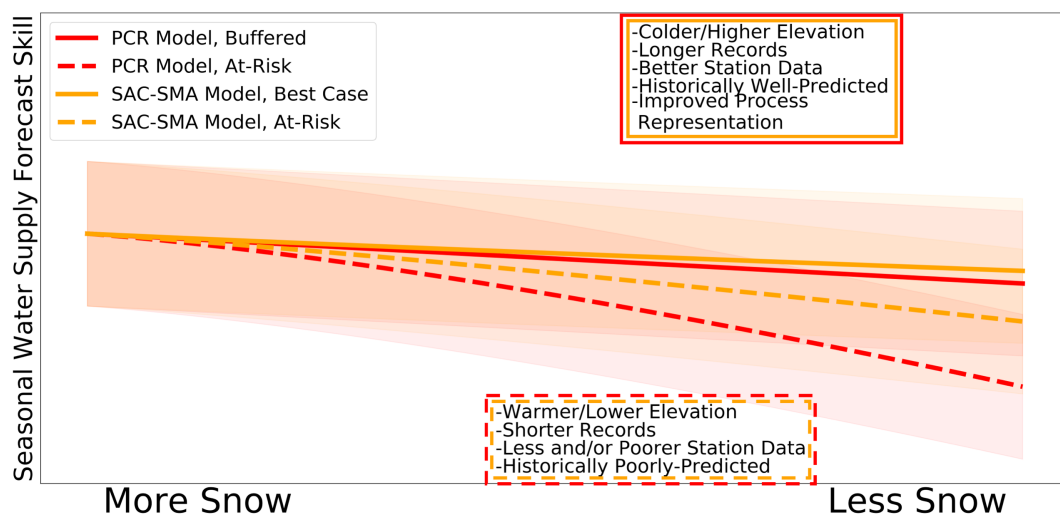
Figure 2.1: A framework for understanding the water supply forecast skill of statistical (PCR, red) and conceptual (SAC-SMA, orange) models under different snow conditions, including factors hypothesized to contribute to model skillfulness. The best case (solid lines) refers to forecast points buffered from skill degradation driven by snow losses; worst case or at-risk (dashed lines) refers to forecast points more sensitive to long-term loss of snowpack and/or snow drought.

It is generally accepted that precipitation received after the forecast issue date, i.e. "climate error", is the primary source (40-80%) of forecast uncertainty; remaining model "structural error" can thus be attributed to forcing data and model process representation (Mantua et al., 2008; Schaake & Peck, 1985). Regions with significant spring and summer precipitation (relative to winter snowpack), such as the American Southwest, are

most affected by post-forecast precipitation error and are therefore challenging for

seasonal forecasting horizons. Seasonal to sub-seasonal precipitation forecasts offer

promise for improved long lead streamflow volume forecasting (Mantua et al., 2008),

with considerable research efforts recently dedicated to this topic (Clark et al., 2004; Cuo

et al., 2011; F. Ralph et al., 2014). However, the relative importance of climate and

structural error varies with forecast lead time and local meteorology; at shorter lead

times, initial condition, forcing data, and model errors may be larger than errors from

meteorological uncertainty.

Sources of model structural error are challenging to disaggregate. For example,

point observations from SNOTEL stations may have biases relative to mean watershed

characteristics (Molotch & Bales, 2006). These biases may not capture declines in low

elevation snow cover or other non-stationary changes. Further, errors in precipitation and

temperature are accentuated in areas with low station density (Gervais et al., 2014) and at

higher elevations (Lundquist et al., 2015). Non-stationary changes to snowmelt and

streamflow generation will likely be captured differently by the implicit and explicit

process representation and parameterization of the PCR and SAC-SMA model structures.

For example, changes in streamflow generation during extremely low snowpacks are

likely to drive a non-linear decrease in streamflow response to precipitation due to lack of

upslope connectivity (Mcnamara et al., 2005), challenging the linear assumptions

inherent to PCR models (Lehner, Wahl, et al., 2017). A model with explicit soil water

storage like SAC-SMA is expected to better represent such conditions (Refsgaard &

Knudsen, 1996). After accounting for the challenges of seasonal precipitation prediction,

basin-scale water supply forecast accuracy is likely to be controlled by declining

snowpack, but also is susceptible to other effects of anthropogenic climate change such as

forest fires, shifts in vegetation, and channel and catchment changes under enhanced

extreme hydrologic events, as well as the degree to which catchment properties buffer

those changes and are adequately represented by forecast model assumptions (Figure

2.1).

Water supply forecasts will become increasingly important as climate change

heightens the risk for drought (Cook et al., 2004) and 'weather whiplash' between dry

and wet periods (Swain et al., 2018). Recent examples like the extreme snow drought

along the West Coast in 2015 (A. Harpold et al., 2017) followed by extreme winter

precipitation in the Sierra Nevada in 2017 demonstrate the need for water supply

forecasts to support reservoir operations (Ralph et al., 2014). While distributed physics-

based models (e.g. the Variable Infiltration Capacity or National Water Model) have the

potential for operational hydrologic forecasting, they are difficult to validate in headwater

basins where many snow-dominated seasonal water supply forecasts are made

(Rasmussen et al., 2011; Yang et al., 2011). Thus, seasonal water supply forecasting in

the western U.S. will continue to rely on the legacy models employed the NRCS and

NWS for a while. Given the need to benchmark these operational models to better

understand skill changes from snowpack decline, we develop a comprehensive dataset for

51 forecast points that are forecasted by both the NRCS and NWS (Figure 2.2).

Fortunately, existing approaches can help reproduce PCR-based water supply forecasts

and validate model runs of SNOW17/SAC-SMA, e.g. Harpold et al. (2017) and Lehner et

al. (2017) recreated modified NRCS PCR forecast equations and Newman et al. (2015; 2017) modified existing SAC-SMA model source code to develop stand-alone versions of the conceptual models used here. This expansive model benchmarking exercise allows us to ask two questions about the reliability of future water supply forecasting:

1. Are seasonal water supply forecasts degraded under conditions of below-average snow accumulation (i.e. snow drought)?
2. If forecasts are less skillful in low snow years, what are the implications for the skill of statistical and conceptual forecast models faced with future SWE decline?

2.2 Methods

2.2.1 Site Selection

Site selection and experimental design were guided by our desire to benchmark existing NRCS forecast points with ensemble streamflow predictions from stand-alone, recalibrated NWS RFC forecasting models.

NRCS forecast points and official regression equations were available from the National Water and Climate Center Air and Water Database (https://wcc.sc.egov.usda.gov/awdbWebService/). All NRCS forecast points included an active USGS streamflow gauge and at least one SNOTEL station measuring snow water equivalent (SWE) and precipitation (P) in or near the basin. All sites were located west of the 104ºW. Basin mean forcing data used to run the NWS's lumped conceptual model, SAC-SMA, were available from the Catchment Attributes and Meteorology for Large-Sample Studies (CAMELS) dataset (Addor et al., 2017) for water years (WY) 1981 to

2014. CAMELS is a research-grade, hydrologic dataset for nearly 700 minimally

impacted basins in the contiguous U.S. CAMELS synthesizes and compiles quality-

controlled streamflow data, basin forcing data (basin lumped and semi-distributed), SAC-

SMA model outputs, and mean basin attributes including topography, climatology,

geology, and hydrology. These features make it an ideal tool for large-scale comparative

basin studies and model evaluation.

Site selection criteria were thus threefold: Each site 1) is an active NRCS forecast

point, 2) is in the CAMELS database, and 3) has a period of record longer than 20 years.

These criteria identified 51 geographically representative basins and forecast points from

Washington, Oregon, Idaho, California, Nevada, Montana, Wyoming, Utah, and New

Mexico (Figure 2.2). Arizona and Colorado did not have any sites that met our criteria.
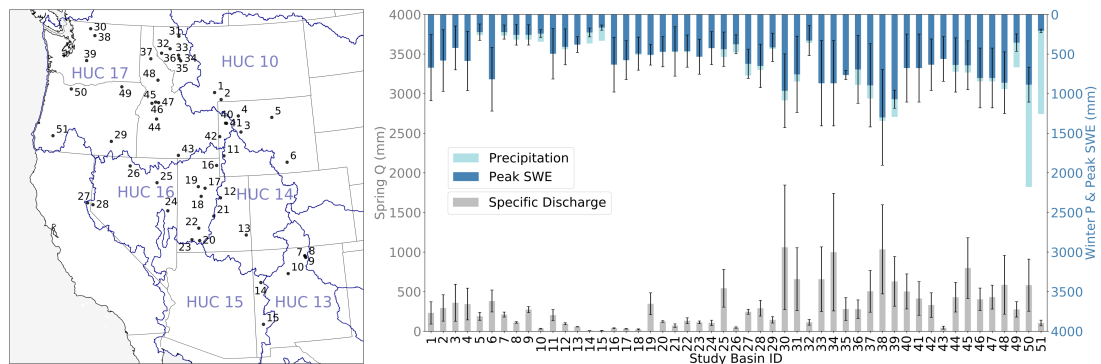
Figure 2.2: (Left panel) Site map with U.S. Geological Survey Regional Hydrologic Unit Codes (HUC; i.e. regional watershed boundaries). (Right panel) Basin mean hydrometeorology including: mean spring (Apr 1 to Jul 31) specific discharge (grey), mean winter (Oct 1 to Apr 1) precipitation (dark blue), and mean peak snow water equivalent (light blue) with max/min values (black bars) for the period of record, i.e. water years 1981 to 2014.

2.2.2 NRCS Principal Component Regression (PCR) Forecasting and Data

NRCS Principal Component Regression (PCR) forecasts were recreated following the National Engineering Handbook Chapter 7, Water Supply Forecasting (U.S. Department of Agriculture, 2011), which standardizes the statistical water supply forecasting methodology originally proposed by Garen (1992). Primary predictor variables for our PCR forecasts are the same as the principle predictors in the official NRCS forecast equations and include 1) end-of-month accumulated P and 2) instantaneous SWE values from one to seven SNOTEL stations in or near each basin (Table 2.1). Supplemental predictor variables used by the NRCS (e.g. climate indices and antecedent streamflow, used at some but not all forecast points) were not included here to allow for more direct inter-basin comparisons. Forecast predictor variables were statistically transformed through a Principal Component Analysis. Only the first Principal Component (PC1) was retained in all cases (Appendix I). The resulting PC1 was linearly regressed against historical April-July streamflow volumes for WY1981-2014. Using this technique, PCR forecasts were recreated for the 51 study basins at four lead times: January 1st (3-month lead), February 1st (2-month), March 1st (1-month), and April 1st (0-month) for WY1981-2014 as data were available as described in Appendix I. Example PCR forecasts for Sagehen Creek in CA (ID# 27, USGS 10343500) are shown in Figure A2.1.

2.2.3 NWS Sacramento Soil Moisture Accounting (SAC-SMA) Forecasting and Data

We employ research-grade versions of the NWS Sacramento Soil Moisture Accounting (SAC-SMA) and SNOW-17 models from Newman et al. (2017)

(https://github.com/anewman89/SACSMA-Snow17) to similarly recreate seasonal water supply forecasts for 51 basins for WYs 1981 to 2014 at four lead times, mirroring the PCR forecast delivery structure. SAC-SMA is a time-stepped lumped conceptual model used for short lead flood forecasting and longer lead seasonal water supply forecasting (National Weather Service, 2016). In snowy areas, SAC-SMA is commonly paired with SNOW-17, a temperature-index based conceptual snow accumulation and ablation model (Anderson, 2006). In all cases, we also paired it with a Unit Hydrograph streamflow routing scheme. Calibrations for the paired models were performed using a Shuffle Complex Evolution (SCE) optimization routine (ftp://ftp.rap.ucar.edu/pub/anewman/sac/) and the daily lumped mean Daymet forcing data from CAMELS for WY1990-1999. The validation period included at least the fifteen years following the calibration period (WY2000-2014) and for the 89.9% of years where data were available, the nine years preceding (WY1981 to 1989). Following model calibration, seasonal water supply forecast ensemble predictions were run in two steps using an Ensemble Streamflow Prediction technique (Day, 1985). Model initialization was first run by applying daily forcing data up to the forecast issue date. The model was then run with ensemble historical forcing data for the period between the forecast issue date (e.g. January 1st) until the end of the forecast period (July 31st). This process was repeated for all four forecast lead times, January 1st through April 1st, as data were available as described within Appendix II. Example SAC-SMA ensemble mean forecasts for Sagehen Creek in CA (ID# 27, USGS 10343500) are shown in hydrograph format in Figure A2.2.

2.2.4 Forecast Evaluation, Error Metrics, and Trend Detection

Standard NRCS forecast evaluations usually rely on percent error, which

measures the magnitude of forecast residuals with respect to observed spring Q. In this study, we also measure prediction bias and skill using percent bias (PBIAS) and a bounded Nash Sutcliffe Efficiency skill score (NSE), as recommended for watershed simulations statistics (Moriasi et al., 2007). We use a slightly modified, bottom-end bounded NSE with possible values ranging from -1 to 1 (Appendix III), allowing for more appropriate model evaluation across a range of study basins, including exceptionally poor-performing sites (Mathevet et al., 2006). When presenting regional results, regions are defined as the Missouri (HUC 10), Upper/Lower Colorado (HUC 14/15) and Rio Grande (HUC 13), Great Basin (HUC 16), and Pacific Northwest (HUC 17) hydrologic units. The Colorado and Rio Grande regions were combined due to small sample size and hydrometeorological similarity.

Lastly, long term trends in forecast skill were quantified by modifying a method from Pagano et al. (2004) that evaluated NSE scores for multi-year rolling windows. For the PCR models developed here, we focus on results from 15-year rolling window spanning WY1981-2014, but also evaluated 5, 10, and 20 year windows. The even shorter validation period of the SAC-SMA forecasts required evaluations across 3, 5, and 10 years windows for WY2000-2014. Rolling NSE scores were evaluated by region. We then used a Regional Mann Kendall Trend Test (Helsel & Frans, 2006) to identify the statistically significant trends in NSE score trends.

Site ID	USGS ID	Station Name and Location	Latitude	Longitude	Area (km2)	Mean Elev. (m)	Mean P (m)	Mean Q (m)	Runoff Ratio	Snow Fraction	P Seasonality	Slope FDC	SNOTELs in PCR
1	06043500	GALLATIN RIVER NEAR GALLATIN GATEWAY MT	45.50	-111.27	2120.4	2,374	2.20	0.94	0.43	0.58	-0.07	0.76	328,385,754
2	06191500	YELLOWSTONE RIVER AT CORWIN SPRINGS MT	45.11	-110.79	6783.6	2,548	2.16	1.14	0.53	0.63	-0.24	1.06	384,635,683,806,822
3	06221400	DINWOODY CREEK ABOVE LAKES, NEAR BURRIS, WYO.	43.35	-109.41	227.9	3,337	1.55	1.46	0.94	0.71	0.01	2.34	822,585,405
4	06280300	SOUTH FORK SHOSHONE RIVER NEAR VALLEY, WY	44.21	-109.55	794.0	3,004	1.95	1.18	0.60	0.67	-0.10	0.95	822,560,350
5	06311000	NORTH FORK POWDER RIVER NEAR HAZELTON, WY	44.03	-107.08	61.2	2,516	1.46	0.61	0.42	0.57	0.27	0.68	325,512,703
6	06632400	ROCK CREEK AB KING CANYON CANAL, NR ARLINGTON, WY	41.59	-106.22	163.0	3,002	2.25	1.10	0.49	0.74	-0.35	0.45	668,731,367
7	08267500	RIO HONDO NEAR VALDEZ, NM	36.54	-105.56	96.3	3,007	1.68	0.88	0.52	0.47	0.29	1.05	715
8	08269000	RIO PUEBLO DE TAOS NEAR TAOS, NM	36.44	-105.50	163.3	2,791	1.55	0.40	0.26	0.40	0.35	0.88	491,715
9	08271000	RIO LUCERO NEAR ARROYO SECO, NM	36.51	-105.53	43.8	2,723	1.69	1.13	0.67	0.48	0.30	1.08	715,491
10	08324000	JEMEZ RIVER NEAR JEMEZ, NM	35.66	-106.74	1208.0	2,459	1.53	0.14	0.09	0.37	0.24	0.70	744,708
11	09223000	HAMS FORK BELOW POLE CREEK, NEAR FRONTIER, WY	42.11	-110.71	333.2	2,449	1.73	0.57	0.33	0.70	-0.52	0.69	509,544,554
12	09312600	WHITE RIVER BL TABBYUNE C NEAR SOLDIER SUMMIT, UT	39.88	-111.04	195.3	2,435	1.58	0.27	0.17	0.60	-0.47	0.65	399,400,864
13	09378170	SOUTH CREEK ABOVE RESERVOIR NEAR MONTICELLO, UT	37.85	-109.37	21.9	2,308	1.62	0.17	0.10	0.50	-0.21	0.32	383
14	09386900	RIO NUTRIA NEAR RAMAH, NM	35.28	-108.55	184.9	2,342	1.21	0.06	0.05	0.31	0.14	0.09	310
15	09430500	GILA RIVER NEAR GILA, NM	33.06	-108.54	4804.9	2,227	1.29	0.09	0.07	0.14	0.47	0.82	757,595,755
16	10023000	BIG CREEK NEAR RANDOLPH, UT	41.61	-111.25	131.5	2,258	1.55	0.21	0.14	0.55	-0.42	1.32	634,533,374
17	10166430	WEST CANYON CREEK NEAR CEDAR FORT, UT	40.41	-112.10	70.2	1,579	1.36	0.11	0.08	0.23	-0.31	1.33	723,330,828,366
18	10172700	VERNON CREEK NEAR VERNON, UT	39.98	-112.38	69.9	1,853	1.39	0.12	0.09	0.41	-0.31	0.90	686,844,723
19	10172800	SOUTH WILLOW CREEK NEAR GRANTSVILLE, UT	40.50	-112.57	11.1	2,086	1.94	1.37	0.70	0.47	-0.35	1.17	631,844,723
20	10173450	MAMMOTH CREEK ABV WEST HATCH DITCH, NEAR HATCH, UT	37.62	-112.52	268.5	2,639	1.68	0.41	0.25	0.64	-0.41	0.87	514,626,853,390,561
21	10205030	SALINA CREEK NEAR EMERY, UT	38.91	-111.53	134.6	2,489	1.65	0.25	0.15	0.58	-0.34	0.87	691,364,475
22	10234500	BEAVER RIVER NEAR BEAVER, UT	38.28	-112.57	236.4	2,499	1.78	0.48	0.27	0.63	-0.31	0.85	621,339
23	10242000	COAL CREEK NEAR CEDAR CITY, UT	37.67	-113.03	208.7	2,035	1.14	0.41	0.35	0.30	-0.41	0.70	626,855,561,390
24	10244950	STEPTOE C NR ELY, NV	39.20	-114.69	28.2	2,404	1.32	0.46	0.35	0.66	-0.34	0.81	849,334
25	10316500	LAMOILLE C NR LAMOILLE, NV	40.69	-115.48	64.8	2,490	2.21	1.69	0.76	0.75	-0.55	0.68	570,453,527,503
26	10329500	MARTIN C NR PARADISE VALLEY, NV	41.53	-117.42	454.5	1,756	1.36	0.18	0.13	0.48	-0.70	0.86	569,373,498,445
27	10343500	SAGEHEN C NR TRUCKEE CA	39.43	-120.24	27.6	2,157	2.57	0.95	0.37	0.71	-1.13	0.93	834,540,539,541
28	10348850	GALENA C AT GALENA C STATE PARK	39.35	-119.86	19.1	2,042	2.51	1.43	0.57	0.83	-1.16	0.96	615
29	10396000	DONNER UND BLITZEN RIVER NR FRENCHGLEN OR	42.79	-118.87	528.9	1,613	1.56	0.55	0.36	0.47	-0.58	1.06	849,759
30	12175500	THUNDER CREEK NEAR NEWHALEM, WA	48.67	-121.07	273.8	1,539	5.46	5.72	1.05	0.69	-0.81	2.07	817,711
31	12358500	MIDDLE FORK FLATHEAD RIVER NR WEST GLACIER MT	48.50	-114.01	2939.2	1,559	3.33	2.34	0.70	0.56	-0.37	1.40	469,307,649,664,693,482,613
32	12374250	MILL CR AB BASSOO CR NR NIARADA MT	47.83	-114.70	50.8	1,380	2.24	0.39	0.18	0.44	-0.24	1.04	510,783,346
33	12375900	SOUTH CROW CREEK NEAR RONAN MT	47.49	-114.03	19.7	1,528	3.43	2.41	0.70	0.61	-0.38	0.86	346,646,783,664
34	12377150	MISSION CR AB RESERVOIR NR ST IGNATIUS MT	47.32	-113.98	32.2	1,512	4.04	3.68	0.91	0.72	-0.51	1.25	346,646,783,664
35	12381400	SOUTH FORK JOCKO RIVER NEAR ARLEE MT	47.20	-113.85	151.0	1,877	3.00	0.96	0.32	0.59	-0.45	1.07	346,646,667,783
36	12390700	PROSPECT CREEK AT THOMPSON FALLS MT	47.59	-115.36	470.2	1,415	3.04	1.13	0.37	0.52	-0.57	1.52	646,594,535,803
37	12414500	ST JOE RIVER AT CALDER ID	47.27	-116.19	2679.0	1,381	3.66	2.07	0.57	0.55	-0.69	1.70	535,623,600,530
38	12451000	STEHEKIN RIVER AT STEHEKIN, WA	48.33	-120.69	830.6	1,510	4.48	4.19	0.94	0.72	-0.88	1.69	681,711,606
39	12488500	AMERICAN RIVER NEAR NILE, WA	46.98	-121.17	205.2	1,453	4.75	2.72	0.57	0.64	-0.90	1.75	375,642,863,692
40	13011500	PACIFIC CREEK AT MORAN WY	43.85	-110.52	404.1	2,401	2.43	1.63	0.67	0.66	-0.47	0.48	837,822,314
41	13011900	BUFFALO FORK AB LAVA CREEK NR MORAN WY	43.84	-110.44	851.8	2,683	2.38	1.47	0.62	0.70	-0.39	0.80	837,822,314
42	13023000	GREYS RIVER AB RESERVOIR NR ALPINE WY	43.14	-110.98	1161.9	2,442	2.36	1.24	0.53	0.65	-0.46	1.13	779,730,419,868,353
43	13083000	TRAPPER CREEK NR OAKLEY ID	42.17	-113.98	133.2	1,863	1.54	0.25	0.16	0.49	-0.53	0.67	534,610,359,698
44	13235000	SF PAYETTE RIVER AT LOWMAN ID	44.09	-115.62	1163.2	2,075	2.57	1.66	0.65	0.68	-0.81	1.06	845,423,312,550,496,637
45	13240000	LAKE FORK PAYETTE RIVER AB JUMBO CR NR MCCALL ID	44.91	-116.00	125.6	1,965	2.58	2.58	0.79	0.73	-0.75	0.92	782,319,370,740,782
46	13310700	SF SALMON RIVER NR KRASSEL RANGER STATION ID	44.99	-115.73	853.1	1,980	3.05	1.40	0.46	0.66	-0.77	1.02	423,338,740,439
47	13313000	JOHNSON CREEK AT YELLOW PINE ID	44.96	-115.50	561.9	2,156	3.63	1.41	0.39	0.74	-0.80	0.76	439,740,338,423
48	13337000	LOCHSA RIVER NR LOWELL ID	46.15	-115.59	3053.4	1,548	3.59	2.16	0.60	0.56	-0.58	1.40	588,425,835,836,735,520
49	14020000	UMATILLA RIVER ABOVE MEACHAM CREEK, NR GIBBON, OR	45.72	-118.32	341.4	1,216	2.64	1.64	0.62	0.62	-0.73	2.14	523,470,362
50	14137000	SANDY RIVER NEAR MARMOT, OR	45.40	-122.14	674.2	1,055	6.30	4.73	0.75	0.32	-0.81	1.73	655,351,712,666
51	14308990	COW CREEK ABV GALESVILLE RES, NR AZALEA, OR.	42.82	-123.13	167.8	987	3.66	1.26	0.34	0.14	-0.94	1.72	558
			42.44	**-113.40**	**717.6**	**2,097**	**2.44**	**1.29**	**0.45**	**0.56**	**-0.42**	**1.06**	

Table 2.1: Study basin IDs, USGS gauge IDs, station names and locations, select CAMELS basin attributes (Addor et al., 2017), and SNOTELs in NRCS forecast equations from the National Water and Climate Center. Runoff ratio is a measurement of streamflow (Q) to precipitation (P) where numbers closer to 1 denote greater streamflow efficiency; P seasonality shows the seasonality and timing of precipitation (Equation 14, from Woods (2009)) where more negative (positive) values indicates that precipitation peaks in winter (summer); and the slope of the flow duration curve (FDC) is a measure of stream variability where steeper slopes indicate flashier streamflow (Equation 3 from Sawicz et al. (2011)). Mean values are listed in bold in the last row of the table.

2.3 Results

2.3.1 Hydrometeorological Variability

Hydrometeorological characteristics vary across and within the study basins, which highlights the need for and challenge of seasonal water supply forecasting (Table 2.1, Figure 2.2). Mean October-April P, for example, varies regionally from 259 mm in the basin of the North Fork of the Powder River WY (ID# 5, USGS 06311000) to 2,177 mm in the Sandy River OR (ID# 50, USGS 14137000). On average, 56% of the annual P within the study basins falls as snow, with mean peak SWE values ranging from 167 mm in the Gila River NM (ID# 15, USGS 09430500) basin to 1,300 mm in the Stehekin River WA (ID# 38, USGS 12451000). Interannually, the coefficient of variation (CV) in winter P and peak SWE is 28% and 31%, respectively. Comparatively, mean April-July Q is more variable (CV 51%), especially for basins with very low volumetric discharge ($<6 \times 10^6$ m3 i.e. <5,000 acre-feet, median CV 86%, n = 10), making them inherently more challenging to predict. Mean spring Q spans several orders of magnitude from 1×10^6 m3 (~1 KAF, thousand acre-feet) in South Creek Monticello UT (ID# 13, USGS 09378170) to 2×10^9 m3 (~1,600 KAF) on the Yellowstone River WY (ID# 2, USGS 06191500).

Streamflow generation is characteristically different between wet and dry years, hereafter defined as years with above and below mean peak SWE, respectively. For instance, aggregating the dry years from all study basins, median annual specific discharge is 294 mm. Wet year median specific discharge, in contrast, is 486 mm; wet year/dry year specific discharge was statistically different in 50 basins. These differences

are not only a result of greater P, but also of more efficient streamflow generation: the median runoff ratio, defined as the ratio of Q to P, increased from 0.31 during dry years to 0.45 during wet years (statistically significant differences in 82% of study basins). Streamflow timing also exhibits significant differences in the fraction of annual Q occurring during winter (DJFM), spring (AMJJ), and fall (ASON) between wet and dry years. During wet years, an average of 68% of annual Q occurred during the spring months, with only 17% in winter and 15% in fall. Dry years, in contrast, saw only 61% of annual Q during spring, with greater proportions of flow during winter (20%) and fall (19%). These differences suggest a shift to an earlier onset of snowmelt during low snow years driving greater streamflow outside of the spring forecasting season compared to high snow years.

2.3.2 Evaluating PCR and SAC-SMA Forecasts

A total of 204 PCR forecast models (51 basins with forecasts each for 3-, 2-, 1-, and 0-month lead times) were developed predicting 5,743 separate April-July Q volumes for WY1981-2014; data gaps, primarily from the years which preceded many SNOTEL station installations, precluded PCR forecasts from being issued for ~17% of the possible 6,936 forecasts (204 PCR forecasts predicting 34 years). In all cases, we selected only the first principal component (PC1) for use in the forecast equations (Appendix I). Using a t-test, 199 of 204 forecasts had statistically significant ($p < 0.05$) regression coefficients for PC1. The regression coefficients of the PC1 for the remaining five forecasts had a maximum p-value of 0.115 but because $t > 1.2$ these PC1s were retained (Appendix I). The 3- and 2-month lead time forecasts for Thunder Creek, WA (ID# 30, USGS 12175500) were the only PCRs where the second PC was also significant. However, for

consistency across forecasts, we only included PC1 in the Thunder Creek regression equations, contrary to Garen (1992). Recall that due to limited training data, separate calibration-validation years were not used for PCR models.

Parallel SAC-SMA model structures were developed for each lead time and forecast point (study basin). However, unlike a statistical model, parameterized conceptual or physics-based models require parameter estimation. Basin-specific parameter values for the SAC-SMA models were determined using a SCE model calibration scheme run over a subset of ten years, WY1990-1999. The calibrated SAC-SMA models were then run with these optimized parameters and an Ensemble Streamflow Prediction technique for the remaining validation years only (WY1981-1989, 2000-2014), predicting a total of 4,692 individual spring Q volumes. With fewer forcing data gaps, SAC-SMA forecasts were issued for ~96% of the total possible 4,896 forecasts during the WY1981-1989 and WY2000-2014 period.

Comparing the aggregated forecast evaluations from the conceptual model with the benchmark performance of the statistical models shows that the statistical models (PCR) generally outperform the conceptual models (SAC-SMA) when evaluated for all model validation years (Figure 2.3, black horizontal bars). Across lead times, model differences in forecast skill (NSE, Figure 2.3a) are largest at short leads (median April 1 NSE: PCR 0. 61, SAC-SMA 0.47) and smallest at long leads (median January 1 NSE PCR 0.24, SAC-SMA 0.27). These values should be considered relative to the forecast skill of the 'perfect prognostic' model runs (i.e. a 'fully-informed' retrospective model; Figure 2.3a, yellow lines), which shows median performance of models forced with data

through the end of July. The lower NSE of the perfect prognostic SAC-SMA models demonstrates a limitation of the models to perform well under optimal conditions, that is, with forcing data from the forecasting period. These findings suggest that the superior performance of PCR models at lead times shorter than 2-months (February 1) may be attributed to higher perfect prognostic and therefore a higher upper-envelope of best-case model performance. Importantly, differences in model calibration techniques and validation years likely introduce some uncertainty.

Evaluating relative model performance during low snow years can be achieved by sub-setting wet years and dry years from one another. Mathematically, evaluating a model trained with 34 years against a subset of data from only 17 years yields a precipitously lower NSE compared to the model evaluated across all years (Appendix III). This mathematical behavior explains why we observe negative NSE scores during some forecasts; unrealistically training the model to a subset of dry years or wet years only and then calculating the NSE demonstrates this idiosyncrasy of the metric (but doesn't simulate a retrospective water supply forecast model). Considering the performance of the wet years with this in mind (Figure 2.3a, blue bars), we find that PCR models consistently outperform the SAC-SMA model at all lead times. These model differences in forecast skill trend similarly across all years and wet years, where forecast skill differences are smaller at long leads and larger at short leads. In contrast, during the subset of dry years (Figure 2.3a, red bars) SAC-SMA models outperform the PCR models at lead times longer than 0-months (April 1). Because snowpack is well-correlated with spring streamflow volumes ($r2 = 0.6$), these findings can also be

interpreted as PCR models are better than SAC-SMA models at predicting higher

flowers, while SAC-SMA models are better than PCR models at predicting lower flows.

Regional differences in relative forecast skill are significant (Figure A2.4).

The experimental design of the calibration-validation datasets between the PCR

and SAC-SMA models manifests in the overall bias of the model predictions (Figure

2.3b). In the PCR model, the least-squares approximation trained on all years of available

data produces an overall unbiased forecast (Figure 2.3b, solid black lines). Sub-setting

wet and dry years shows that expectedly, the PCR model overpredicts streamflow during

dry years (positive bias) and underpredicts streamflow during wet years (negative bias).

In comparison, the SAC-SMA model, which was calibrated on WY1990-1999 and

evaluated for the remaining years, has an overall positive bias (dashed black lines),

largely due to high PBIAS during dry years at lead times greater than 0-months. With

respect to the PCR models, the SAC-SMA models have a tendency to underpredict

during wet years; in comparison, during dry years the model often over predicts, but at

lead times shorter than 3-months, also regularly under predicts, i.e. the inter-quartile

range often bounds zero PBIAS.
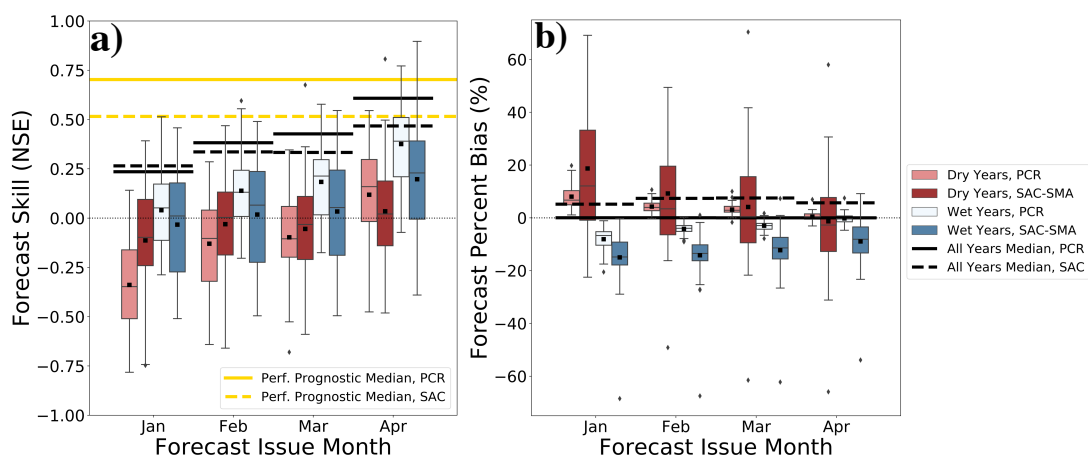
Figure 2.3: Forecast evaluations during above (blue) and below (red) average peak SWE as measured by: **a)** a bounded Nash Sutcliffe Efficiency (NSE) score and **b)** percent bias (PBIAS) across four lead times. Lighter colors are PCR forecasts, darker colors are SAC-SMA forecasts. Horizontal black bars show median forecast performance when evaluated across all years; yellow bars are forecast performance when models are run with forcing data through the end of the forecasting period. Positive PBIAS is a forecast over prediction.

2.3.3 Comparing PCR and SAC-SMA Forecast Skill

Differences in validation years present some limitations to aggregated model comparisons of composite error metrics (Figure 2.3) which can be more evenly compared in a one-to-one plot (Figure 2.4). A direct evaluation between individual spring Q forecasts from PCR and SAC-SMA models shows that the mean difference in the absolute residuals of the forecast error range from 15% (January 1st) to 18% (April 1st), with larger model differences in forecasted values at shorter lead times. However, because regional median spring Q varies by several orders of magnitude (from 1.9x108 m3 in the Pacific Northwest to 5.0x106 m3 in the Great Basin) the site-year differences in forecast performance are lost to the details of Figure 2.4, even on a log-log scale. Regional one-to-one plots of predicted spring Q (Figure A2.5) show more of these

details; median April 1st model differences are 16% in the Missouri, 24% in the

Colorado/Rio Grande, 27% in the Great Basin, and 14% in the Pacific Northwest.

Consistent with earlier findings, the bottom panels of Figure 2.4 demonstrate that

site-year PCR and SAC-SMA forecast performance (percent error) is more similar at long

leads (January 1st, r2 = 0.74), but that by April 1st, larger model differences emerge (r2 =

0.39, Figure 2.4, bottom panels). Differences in model performance are especially evident

in the Great Basin (green colored triangles) and Colorado (red-orange colored triangles).

This behavior is expected: for a given residual between the actual and forecasted values,

a lower spring Q will yield a higher percent error. We also find that SAC-SMA models

have a tendency to predict smaller values with respect to PCR models, i.e. percent error

values plot above the one-to-one line. Yet, these trends vary interannually (light grey

crosshairs) as well as regionally and by forecast lead time (Figure A2.6).
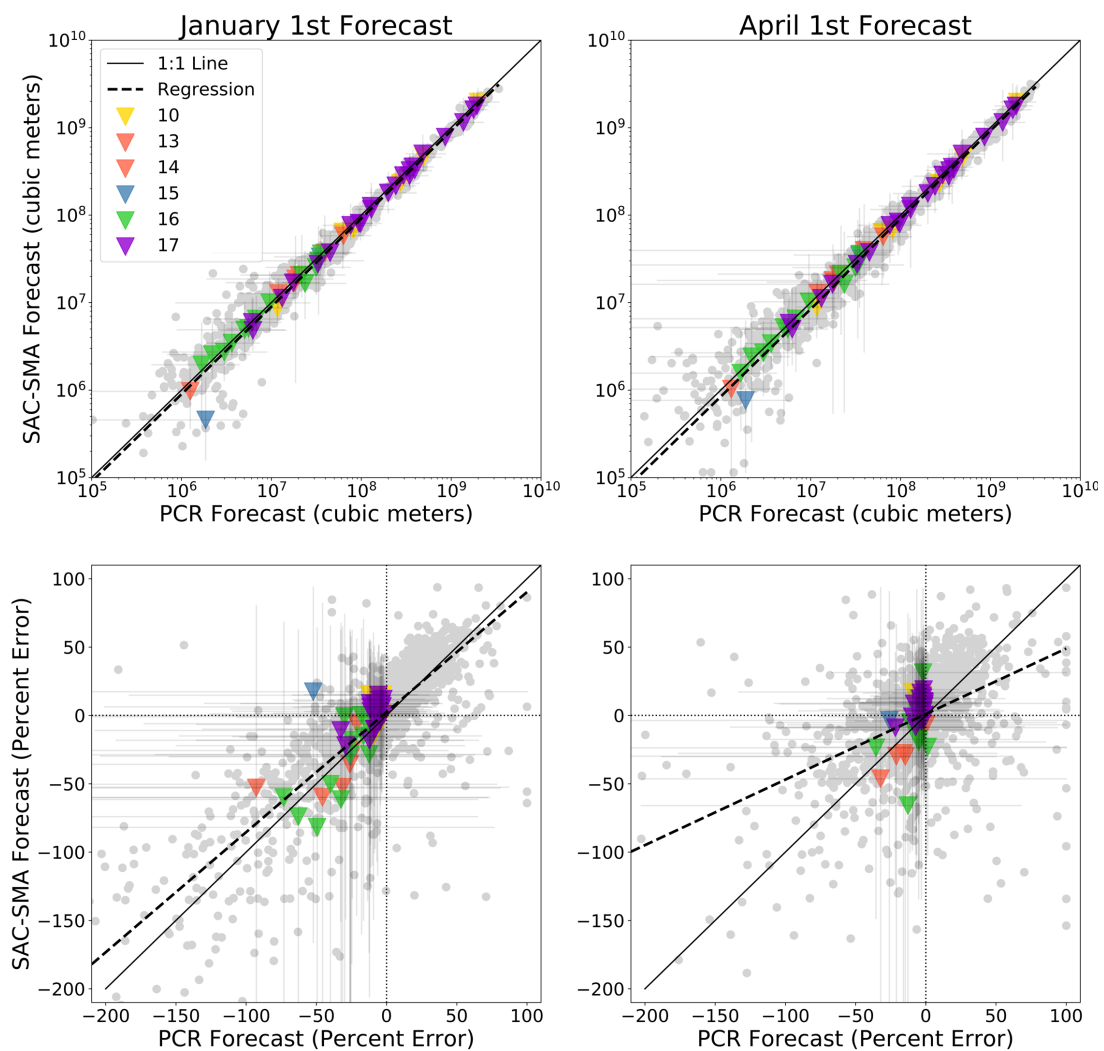
Figure 2.4: SAC-SMA predicted spring Q (m3) as a function of PCR predicted spring Q. Individual site years are in light grey, basin mean values are triangles colored by regional hydrologic unit code, and crosshairs show the interannual variability for each site (minimum/maximum values).

Normalized peak SWE and spring Q are well correlated (r2 = 0.6) across our snow-dominated study basins, where low snow years yield lower spring Q and vice versa. Predicting smaller streamflow volumes during these lower snow years requires greater forecast accuracy (smaller residuals) for equal percent error performance. Because of this, we find a strong relationship between high percent error and low flows. For

example, all forecasts with >60% error occurred when predicting spring Q flows <

5.0x107 m3 (<40 KAF). As a result, comparatively small differences between PCR and

SAC-SMA forecasts result in large absolute percent error during years with below-

average SWE (Figure 2.5). Specifically, we find that years with <75% mean peak SWE

show large (proportional) variability and lack any clear agreement around model

superiority, indicated by 90% confidence intervals bounding absolute differences of zero.

Regional trends are shown in Figure A2.7. In contrast, average to above-average SWE

conditions were modeled similarly by PCR and SAC-SMA models, trending towards

slightly better PCR performance above 125% peak SWE, though 90% confidence

intervals became larger with fewer data points.
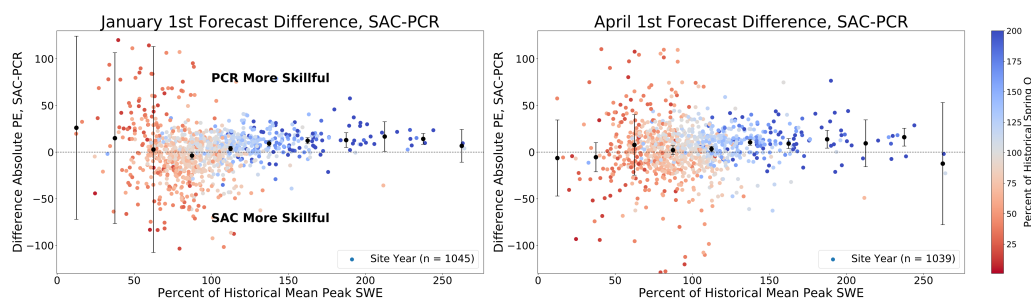
Figure 2.5: Differences between the absolute percent errors of the SAC-SMA and PCR model forecasts for January 1st (left panel) and April 1st (right panel). Median values are bounded by a 90% confidence interval (vertical bars) for 25% SWE bins (e.g. 0-25% mean historical peak SWE). SWE values are SNOTEL observations. Colored by percent of historical mean spring Q.

2.3.4 Regional Variations in Forecast Skill

The challenge of seasonal to sub-seasonal hydrologic forecasting is mostly

attributed to the meteorological uncertainty outside of a two week weather prediction

horizon. The importance of post-forecast precipitation on forecast skill is well-studied

and documented (Mantua et al., 2008; Schaake & Peck, 1985). Across our study domain,

we confirm that forecast percent error is well-correlated with post-forecast precipitation anomalies (Figure 2.6). Specifically, we find that both models (PCR in blue and SAC-SMA in orange) underpredict spring Q volumes when post-forecast precipitation anomalies are positive, and vice versa. A similar relationship between forecast percent error and percent of mean peak SWE (Figure A2.8) is also observed, which is expected in snow-dominated regions where snowmelt is the main driver of spring streamflow. Regional sensitivity of forecast skill to spring P can be estimated by the slope and standard error (SE) of the regression in Figure 2.6. For example, we find that forecast errors in the Colorado/Rio Grande region were most sensitive to spring P (model mean slope = 269.4), but these relationships were weak and highly variable (model mean standard error = 139.3) and dependent on the model and forcing data.

Figure 2.6: Forecast percent error as a function of the April – July fraction of mean annual P. Increases in post-forecast issue date P are positively correlated with larger under predictions of streamflow volume (i.e. positive percent error) across all HUC groups for both the SAC-SMA model (orange) and the PCR model (blue). Mean Spring Q units are in thousands of acre-feet.

We have attempted to identify important differences in regional water supply forecast model performance through the analysis of hydrologic units, while also evaluating site-year forecast performance across snowpacks and spring precipitation regimes. In Figure 2.7, we use a basin-scale approach to benchmark January 1st (top panels) and April 1st (bottom panels) model performance of SAC-SMA to PCR using the NSE. These comparisons highlight the regional differences in forecast skill where, generally, we find better PCR performance, e.g., in the Pacific Northwest and better

SAC-SMA performance in, e.g., the Great Basin and parts of the Upper Colorado. These trends are less distinct, however, during below-average snowpack (dry years). In particular, dry year January 1st forecasts are more skillfully predicted (higher NSE) by SAC-SMA models almost universally: only 12% of basins have higher PCR NSE scores during this time.

There are many basins where model differences are small (i.e. small orange or white inner circles, bottom panels) and/or where model superiority is inconsistent or dependent on wet versus dry years (Figure A2.9). For example, the regional watershed boundary between the Great Basin and Upper Colorado includes several basins with small differences in model performance and no consistent model performance. Similar trends are also observed in other parts of the Colorado/Rio Grande, Pacific Northwest, and Missouri hydrologic units. But, in general, we see a north-south divide where more skillful PCR forecasts are located in northern regions with wetter conditions and higher mean spring Q (Pacific Northwest and Missouri) and more skillful SAC-SMA forecasts are located in southern, more arid regions with lower mean spring Q (Great Basin, Colorado, and Rio Grande).
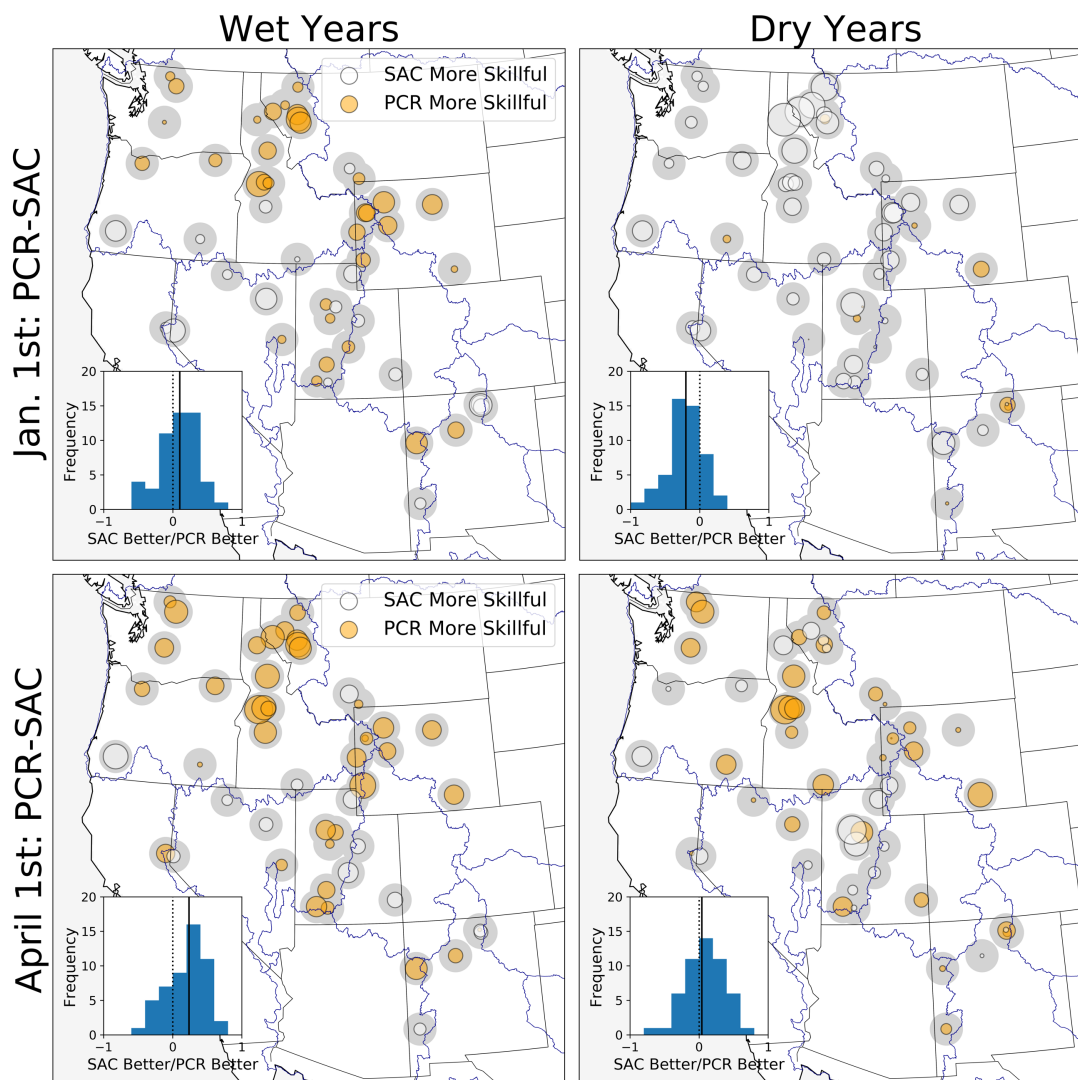
Figure 2.7: Benchmark evaluation of SAC-SMA model NSE to PCR model NSE on January 1st (top panels) and April 1st (bottom panels) across above-average (left column) and below-average (right column) snowpack. Dark grey circles represent a difference in NSE of +/- 1. Orange circles are basins where PCR is more skillful (higher NSE); white circles are basins where SAC-SMA is more skillful. Regional watersheds (hydrologic units) are outlined in blue.

Pagano et al. (2004) demonstrated that official NRCS water supply forecast skill, subject to interannual to interdecadal climate variability, changed through time, but surprisingly, was generally steady (and not increasing), despite a growing period of

record. Extending this analysis, we evaluate basin-scale changes in PCR forecast skill (NSE) within 15-year rolling windows from WY 1981-2014 (Figure 2.8). Mean values for all lead times in each regional hydrologic unit were then evaluated using a Regional Mann Kendall Trend Test, which measures Sen's Slope values and their statistical significance (Helsel & Frans, 2006). This trend test finds statistically significant decreases in forecast skill for 9 of 16 regional-lead time ensembles (Figure 2.8, thick solid lines); the April 1st Colorado/Rio Grande hydrologic unit is the only ensemble mean which is increasing significantly. To evaluate the sensitivity of trend detection to the rolling window size, we evaluated forecasts across 5, 10, 15, and 20-year windows for each lead time and hydrologic unit for the PCR forecasts; the small range of statistically significant (p< 0.05) Sen's Slope values across the four rolling windows suggests that PCR forecast-skill trends in the Missouri (HUC 10), Colorado/Rio Grande (HUCS 13-15), and Pacific Northwest (HUC 17) are mostly insensitive to rolling window size, with the largest decreases in skill in the Pacific Northwest and at longer lead times (Figure 2.9a).

A shorter (WY1981-1989, WY 2000-2014) and more recent continuous period of record complicates a similar analysis for the SAC-SMA model results. If we only evaluate SAC-SMA performance using a 3, 5, and 10-year rolling from WY2000-2014, we do see similar patterns in forecast skill trends between the two models, though SAC-SMA trends are more pronounced in both the positive and negative directions (Figure 2.9b). Direct attribution of changes in forecast skill through time are outside of the scope of this study, though a complimentary trend analysis of SNOTEL-measured SWE and accumulated P was conducted (Figure A2.10). In this analysis, we find statistically

significant declines in SWE and accumulated P in the Pacific Northwest, Colorado/Rio

Grande, and Great Basin, with increases in SWE and P in the Missouri and Great Basin,

though further research is needed to directly attribute factors affecting temporal trends in

water supply forecast skill.
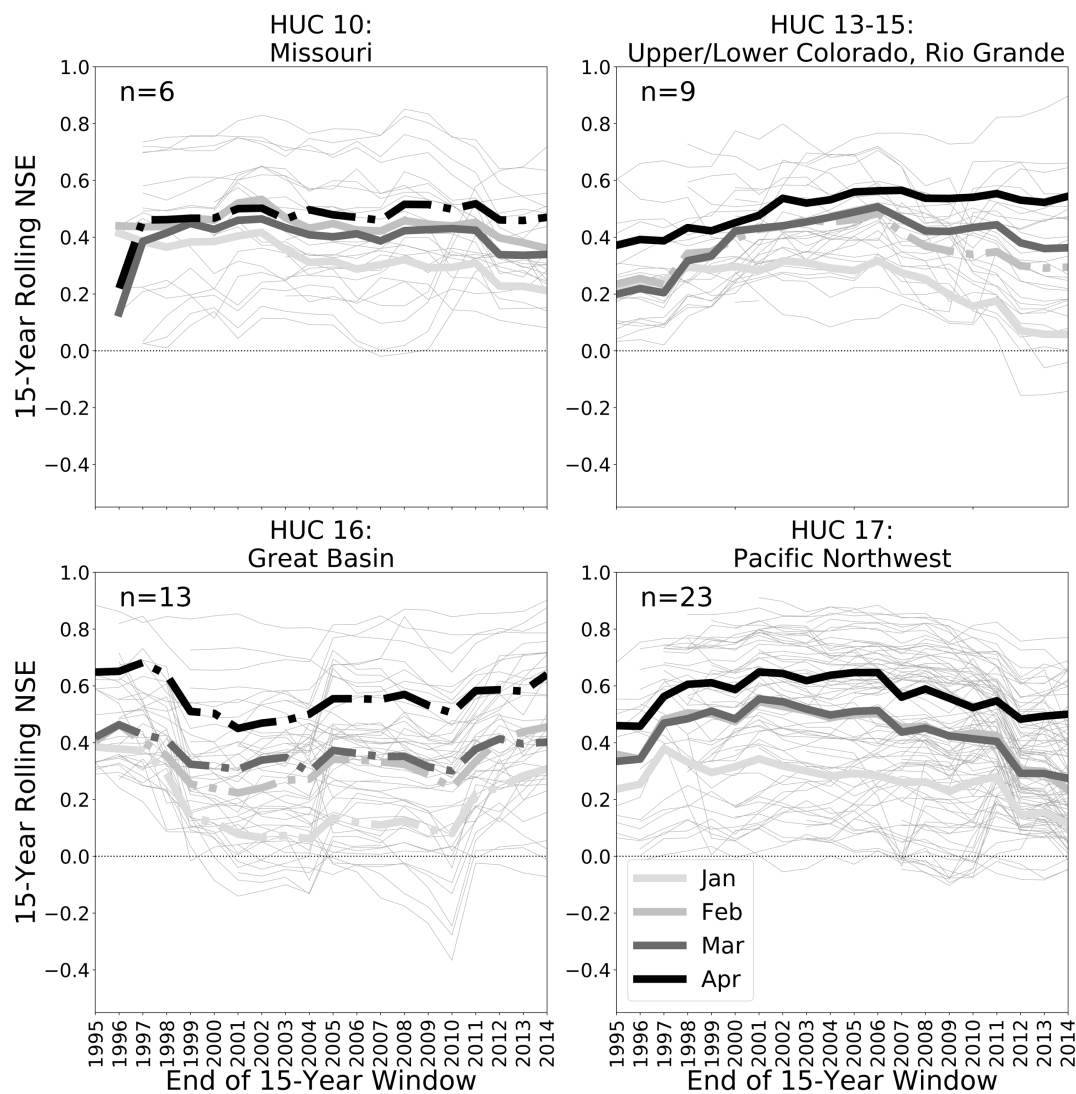
Figure 2.8: 15-year rolling NSE scores across four regional hydrologic units. Thin light grey lines are the individual PCR forecasts for each basin in that region. Heavier weight greyscale lines are the mean of the 15-year rolling NSE for the four lead times; solid thick lines indicate statistical significance (p<0.05) and dashed lines indicate statistical insignificance (p>0.05)
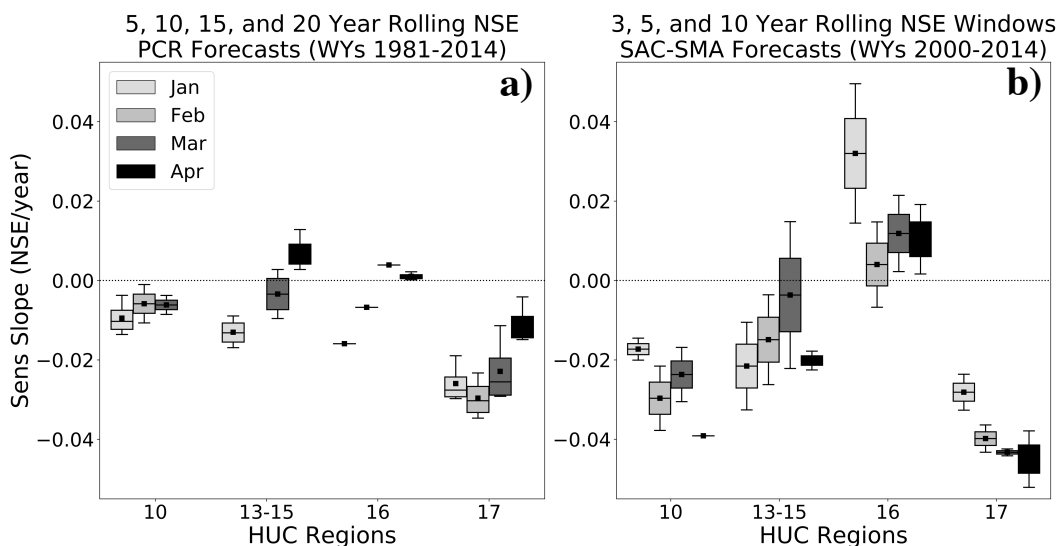
Figure 2.9: Statistically significant Sen's Slope values from a Regional Mann Kendall Trend Test for all forecast lead times in each regional hydrologic unit for a) 5, 10, 15, and 20 year rolling NSE evaluations of PCR forecasts from WY1981-2014 and b) 3, 5, and 10 year rolling NSE evaluations of SAC-SMA forecasts during WY2000-2014.

Our results suggest that statistical water supply forecasts (PCR) generally outperform conceptual models (SAC-SMA) for seasonal streamflow predictions in snow-dominated basins. Yet, these advantages in PCR model performance are most evident during average and above-average snowpack and at shorter lead times (Figure 2.3, 7). In contrast, SAC-SMA model performance is superior at longer lead times regionally, and at longer lead times during below-average snow years almost universally. A caveat to these findings is that model differences are smallest at the longest lead times (Figure 2.3, 4), suggesting that any advantages to using the SAC-SMA model at long lead times is may not be significant. This is complicated by the fact that seasonal to sub-seasonal precipitation forecasting is currently unreliable, making it impossible to confidently determine which model to consider based on snowpack forecasts on January 1st (Figure 2.7). Regionally, there are large spatiotemporal differences in model performance summarized in three broad patterns. First, in the Pacific Northwest and Missouri, PCR

forecasts are reliably favored, particularly during wet years. Second, in the Colorado/Rio Grande and Great Basin, SAC-SMA outperforms PCR at longer lead times. Lastly, SAC-SMA is better during low snow years for all lead times in the Colorado/Rio Grande and Great Basin and at long lead times in the Pacific Northwest and Missouri.

Our results also suggest that forecast skill is generally degrading through time, particularly in regions known to have vulnerable snowpacks such as the Pacific Northwest, though also in colder, interior regions such as the Missouri and Colorado at longer lead times (Figure 2.8, 9). These changes were observed across both models, though limited by different model validation years. A regional analysis of trends in SNOTEL-measured SWE and P showed significant changes (in both directions) for SWE and P. A more robust analysis would be beneficial for attributing forecast skill changes from snow droughts (interannual) versus long term declines in snow or precipitation due to anthropogenic climate change.

2.4 Discussion

The benchmark analysis of two common seasonal water supply forecast models described above shows that increased occurrence of low snow years in the future may threaten the skillfulness of empirical (PCR) models during average to above-average snowpack, whereas conceptual models (SAC-SMA) which are more skillful during lower snowpacks may experience some buffering of skill loss at lead times of 1-month or longer. To be clear, our results show that the forecast skill of historical April to July streamflow volumes is degraded in dry years versus wet years for both SAC-SMA and PCR forecasts, but that PCR models are more sensitive to below-average snow

accumulation. These general patterns highlight how long lead PCR forecasts are more disadvantaged by post-forecast precipitation (Figure 2.3a) and how SAC-SMA under predicts in wetter climates and at higher flows, even at short leads, but is better at lower flows (Figure 2.4). Vulnerability to SAC-SMA model structural error, as estimated by poorer model performance under a prognostic model run (Figure 2.3a, yellow lines), may represent significant potential for model improvement via calibration or otherwise (Gong et al., 2013). In contrast, vulnerability to climate error may disproportionately disadvantage PCR forecasts if there are significant future changes in precipitation timing or amount. However, long term trends in forecast skill show degradation of skill across both models (Figure 2.8), though regional trends are consistent with our hypothesis (Figure 2.1) that forecast skill for some basins will degrade more than others. A longer period of record would enable independent PCR calibration/validation periods and a more complete SAC-SMA forecast skill trend analysis to better understand how relative model forecast skill has changed.

Post-forecast spring P anomalies are a significant control on forecast skill (Day, 1985; Garen, 1992; Schaake & Peck, 1985), however differences in relative forecast performance reflect limitations of model structures, parameterizations, and data quality in streamflow models and forecasts. Disentangling other controls on forecast skill between the two very different model types was challenging in the western U.S. where climate and basin properties vary widely. Returning to Figure 2.1, our results supports our hypotheses about the role of snow vulnerability in forecast skill. Conceptually, with smaller snowpacks, the contribution of snowmelt for springtime streamflow clearly diminishes. In addition to smaller snowmelt contributions elevating the relative importance of

individual spring P events, loss of snow and earlier snowmelt also dilute the importance

of PCR model initial conditions (i.e. forecast-issue date SWE and P) and amplify SAC-

SMA initial conditions, like soil moisture. Explicit soil water storage in the SAC-SMA

model may explain some of the model divergence during low snow years. For example,

Franz et al. (2003) showed that forecasts for Lower Colorado basins were less skillful at

shorter lead times due to early loss of snow and the forecast information it provided.

Relative importance of primary and secondary controls on forecast skill is evident in the

example from the Rio Nutria NM (ID# 14, USGS 09386900), a low-skill (model average

April 1st NSE 0.32) forecast point with large spring water inputs (spring P:peak SWE =

0.84). In comparison to the high skill (model average April 1st NSE 0.74), snow-

dominated, Mediterranean climate (spring P:peak SWE = 0.22) of Sagehen Creek CA

(ID# 27, USGS 10343500), the Rio Nutria forecast is expected to have higher climate

error, since the spring P (unknown to the model) is a larger fraction of the SWE (known

to the model, at least through April 1). At any given forecast point, combinations of

sparse observations; poor forcing data; small, at-risk, and changing snowpack; highly

variable and non-stationary streamflow; and increasing atmospheric water demand

(Lehner, Wahl, et al., 2017) can drive large (and potentially growing) forecast errors.

   Reproducing the actual operational forecasts (in any but limited retrospective

ways) remains elusive and limits benchmarking studies. This study was fortunate to have

descriptions of operational procedures to follow where possible (U.S. Department of

Agriculture, 2011), however the forecasts benchmarked are not identical to those from

full-blown operations. A major difference comes from undocumented but routine forecast

adjustments in operations. Conversations with forecasters at the Portland, Oregon NRCS

and the NWS California-Nevada River Forecast Center all confirm that tuning of initial

conditions and observed driving variables is not only commonplace, but is a critical

element of operational hydrologic forecasting. The purely automated, objective models

used here do not include these adjustments. Notwithstanding bit-reproducibility and

forecast verification (Melsen et al., 2017), we re-created conclusion-reproducible (i.e. the

ability to reproduce other's scientific findings) versions of the forecasts that allows

objective comparisons of sites and years. While precise model reproducibility in

computational hydrology has been criticized (Hutton et al., 2016), Melsen et al. (2017)

argue reproducibility should more broadly be defined as an ability to verify (or falsify)

model conclusions themselves.

Collaborations between the NRCS and NWS forecasters to leverage and build on

their respective (and different) forecasts strategies may allow forecasters to build upon

the strengths of each (Pagano et al. (2014). Wood and Lettenmaier (2006) note that

operational weather predictions have long relied on a suite of models blended to produce

a final weather forecast. Another strategy would be to develop an "ensemble of

ensembles" approach wherein different forecast models are deployed and blended

depending on basins climate conditions and lead times (see Figure 2.7). Among the

basins studied here, a quarter (n=14) of the study basins might benefit from this approach

because one model clearly performed best on January 1st while the other model was

better for April 1st. More than 40% of study basins could improve skill by using

alternative models in dry versus wet years, for their April 1st forecasts (Figure A2.9). For

January 1st forecasts, this number could grow to 65% of basins by explicitly drawing on

SAC-SMA skill in dry years.

Another benefit of multi-model forecasting is improved uncertainty quantifications under different conditions. For example, models that have more climate error may be used to provide an estimate of longer lead time post-forecast precipitation uncertainty, whereas shorter lead time forecast uncertainty may be better described by models with greater structural uncertainty. Benchmarking studies of operational and research-grade models help identify these important model synergies and will be necessary to test new modeling approaches to mitigate skill loss in a future with less snow.

2.5 Conclusion

Improved water supply forecasting could be worth billions of dollars in economic productivity in the United States (Hamlet et al., 2002). Substantial resource allocations will likely be needed to mitigate water supply forecast skill degradations, like those shown in the Pacific Northwest (Figure 2.8), because mitigation will require expanded monitoring networks, improved process understanding and representation, and better model calibration (F. M. Ralph et al., 2014; Wood & Lettenmaier, 2006). Expanded monitoring networks will be needed to improve tracking of shifting rain-snow elevation lines and replace loss of information from snow storage with soil moisture or other observations (Adrian A. Harpold et al., 2017; Koster et al., 2010; Rosenberg et al., 2013). Advancements in distributed, physics-based models (i.e. National Water Model, https://water.noaa.gov/about/nwm) and machine learning through improved automation and computing could improve seasonal water supply forecasting but will be subject to similar benchmarking challenges. This study points to several opportunities to improve water supply forecasts and their impacts on decisions making:

Model reproducibility and benchmarking:

Because of the reliance on professional forecaster judgement, operational streamflow volume forecasts are not reproducible. Without reproducibility, it is difficult to evaluate the potential of new scientific modeling approaches through benchmarking studies. Additionally, moving towards a more automated operational approach will be essential to scaling-up streamflow predictions into a nowcasted domain across a greater number of forecast points. Yet, tradeoffs exist in a strictly-automated forecasting environment. Determining these tradeoffs between a more automated forecasting world and the status quo will be essential to advancing seasonal streamflow forecasting.

Forecast synergy and discussion:

Epistemic uncertainty, or uncertainties that could theoretically be resolved by an adequate model, can be estimated, and theoretically, resolved through a multi-model approach. In addition to multi-model ensembles making the specifics of any one model less important, they also leverage known and unknown model strengths to produce the best forecast. Further, multi-modeling assists in quantifying forecast uncertainty, which is critical in predicting extreme hydrologic events outside of the historical period of record, whether in an automated or forecaster-informed operational hydrologic modeling world.

References

Addor, N., A. J. Newman, N. Mizukami, and M. P. Clark, 2017: The CAMELS data set: catchment attributes and meteorology for large-sample studies. Hydrol. Earth Syst. Sci, 21, 5293–5313, https://doi.org/10.5194/hess-21-5293-2017.

Anderson, E., 1973: National Weather Service River Forecast System - Snow Accumulation and Ablation Model', NOAA Technical Memorandum NWS HYDRO-17. 1–229 pp.

——, 2006: Snow Accumulation and Ablation Model-SNOW-17.

Bales, R. C., N. P. Molotch, T. H. Painter, M. D. Dettinger, R. Rice, and J. Dozier, 2006: Mountain hydrology of the western United States. Water Resour. Res., 42, https://doi.org/10.1029/2005WR004387.

Barnett, T. P., J. C. Adam, and D. P. Lettenmaier, 2005: Potential impacts of a warming climate on water availability in snow-dominated regions. Nature, 438, 303–309, https://doi.org/10.1038/nature04141.

Barnhart, T. B., N. P. Molotch, B. Livneh, A. A. Harpold, J. F. Knowles, and D. Schneider, 2016: Snowmelt rate dictates streamflow. Geophys. Res. Lett., 43, 8006–8016, https://doi.org/10.1002/2016GL069690.

Berghuijs, W. R., R. A. Woods, and M. Hrachowitz, 2014: A precipitation shift from snow towards rain leads to a decrease in streamflow. Nat. Clim. Chang., 4, 583–586, https://doi.org/10.1038/nclimate2246.

Beven, K., 2002: Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. Hydrol. Process., 16, 189–206, https://doi.org/10.1002/hyp.343.

Burnash, R. J. C., R. L. Ferral, and R. A. McGuire, 1973: A Generalized Streamflow Simulation System: Conceptual Modeling for Digital ... - Robert J. C. Burnash, R. Larry Ferral, Robert A. McGuire - Google Books. https://books.google.com/books?hl=en&lr=&id=aQJDAAAAIAAJ&oi=fnd&pg=PR2&dq=A+genera lized+streamflow+simulation+system+-+conceptual+modeling+for+digital+computers&ots=4sWc_i8daw&sig=hTA89Z7zHWhqB78bzvBY CmrepYA#v=onepage&q=A generalized streamflow simulation system - conceptual modeling for digital computers&f=false (Accessed April 14, 2020).

Cayan, D. R., 1996: Interannual Climate Variability and Snowpack in the Western United States. J. Clim., 9, 928–948, https://doi.org/10.1175/1520-0442(1996)009<0928:ICVASI>2.0.CO;2.

Church, J. E., 1935: Principles of Snow Surveying as Applied to Forecasting Stream Flow. J. Agric. Res., 51, 97–130.

Clark, M. P., L. E. Hay, M. P. Clark, and L. E. Hay, 2004: Use of Medium-Range Numerical Weather Prediction Model Output to Produce Forecasts of Streamflow. http://dx.doi.org/10.1175/1525-7541(2004)005<0015:UOMNWP>2.0.CO;2, https://doi.org/10.1175/1525-7541(2004)005<0015:UOMNWP>2.0.CO;2.

Cook, E., C. Woodhouse, C. M. Eakin, D. Meko, and D. Stahle, 2004: Long-Term Aridity Changes in the Western United States. Science (80-.)., 306, 1015–1018, https://doi.org/10.1126/science.1101982.

Cuo, L., T. C. Pagano, and Q. J. Wang, 2011: A Review of Quantitative Precipitation Forecasts and Their Use in Short- to Medium-Range Streamflow Forecasting. J. Hydrometeorol., 12, 713–728, https://doi.org/10.1175/2011JHM1347.1.

Day, G. N., 1985: Extended Streamflow Forecasting Using NWSRFS. J. Water Resour. Plan. Manag., 111, 157–170, https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157).

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.

Garen, D. C., 1992: Improved Techniques in Regression-Based Streamflow Volume Forecasting. J. Water Resour. Plan. Manag., 118, 654–670.

Gervais, M., L. B. Tremblay, J. R. Gyakum, and E. Atallah, 2014: Representing Extremes in a Daily Gridded Precipitation Analysis over the United States: Impacts of Station Density, Resolution, and Gridding Methods. J. Clim., 27, 5201–5218, https://doi.org/10.1175/JCLI-D-13-00319.1.

Gong, W., H. V. Gupta, D. Yang, K. Sricharan, and A. O. Hero, 2013: Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach. Water Resour. Res.,

49, 2253–2273, https://doi.org/10.1002/wrcr.20161.

Gupta, H. V., S. Sorooshian, and P. O. Yapo, 1999: STATUS OF AUTOMATIC CALIBRATION FOR HYDROLOGIC MODELS: COMPARISON WITH MULTILEVEL EXPERT CALIBRATION. J. Hydrol. Eng., 4, 135–143.

Hamlet, A. F., D. Huppert, and D. P. Lettenmaier, 2002: Economic Value of Long-Lead Streamflow Forecasts for Columbia River Hydropower. J. Water Resour. Plan. Manag., 128, 91–101, https://doi.org/10.1061/(ASCE)0733-9496(2002)128:2(91).

Hamlet, A. F., P. W. Mote, M. P. Clark, and D. P. Lettenmaier, 2005: Effects of Temperature and Precipitation Variability on Snowpack Trends in the Western United States*. 4545–4560 pp. https://journals.ametsoc.org/doi/pdf/10.1175/jcli3538.1 (Accessed March 4, 2019).

Hammond, J. C., A. A. Harpold, S. Weiss, S. K. Kampf, and J. C. Hammond, 2019: Partitioning snowmelt and rainfall in the critical zone: effects of climate type and soil properties 2 3. Hydrol. Earth Syst. Sci., https://doi.org/10.5194/hess-2019-98.

Harpold, A., P. Brooks, S. Rajagopal, I. Heidbuchel, A. Jardine, and C. Stielstra, 2012: Changes in snowpack accumulation and ablation in the intermountain west. 48, 11501, https://doi.org/10.1029/2012WR011949.

——, M. Dettinger, and S. Rajagopal, 2017a: Defining Snow Drought and Why It Matters. Eos (Washington. DC)., 98, https://doi.org/10.1029/2017EO068775.

Harpold, A. A., and M. Kohler, 2017: Potential for Changing Extreme Snowmelt and Rainfall Events in the Mountains of the Western United States. J. Geophys. Res. Atmos., 122, 13,219-13,228, https://doi.org/10.1002/2017JD027704.

Harpold, A. A., and P. D. Brooks, 2018: Humidity determines snowpack ablation under a warming climate. PNAS, 115, 1215–1220, https://doi.org/10.1073/pnas.1716789115.

Harpold, A. A., K. Sutcliffe, J. Clayton, A. Goodbody, and S. Vazquez, 2017b: Does Including Soil Moisture Observations Improve Operational Streamflow Forecasts in Snow-Dominated Watersheds? J. Am. Water Resour. Assoc., 53, 179–196, https://doi.org/10.1111/1752-1688.12490.

Helsel, D. R., and L. M. Frans, 2006: Regional Kendall Test for Trend. Environ. Sci. Technol., 40, 4066–4073, https://doi.org/10.1021/ES051650B.

Hutton, C., T. Wagener, J. Freer, D. Han, C. Duffy, and B. Arheimer, 2016: Most computational hydrology is not reproducible, so is it really science? Water Resour. Res., 52, 7548–7555, https://doi.org/10.1002/2016WR019285.

Jolliffe, I. T., 2002: Principal Component Analysis, Second Edition. 2nd Edition. Springer,.

Kapnick, S., and A. Hall, 2012: Causes of recent changes in western North American snowpack. Clim. Dyn., 38, 1885–1899, https://doi.org/10.1007/s00382-011-1089-y.

Knowles, N., M. D. Dettinger, and D. R. Cayan, 2006: Trends in Snowfall versus Rainfall in the Western United States. J. Clim., 19, 4545–4559.

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. Nat. Geosci., 3, 613–616, https://doi.org/10.1038/ngeo944.

Lehner, F., E. R. Wahl, A. W. Wood, D. B. Blatchford, and D. Llewellyn, 2017a: Assessing recent declines in Upper Rio Grande runoff efficiency from a paleoclimate perspective. Geophys. Res. Lett., 44, 4124–4133, https://doi.org/10.1002/2017GL073253.

——, A. W. Wood, D. Llewellyn, D. B. Blatchford, A. G. Goodbody, and F. Pappenberger, 2017b: Mitigating the Impacts of Climate Nonstationarity on Seasonal Streamflow Predictability in the U.S. Southwest. Geophys. Res. Lett., 44, 12,208-12,217, https://doi.org/10.1002/2017GL076043.

Lundquist, J. D., M. Hughes, B. Henn, E. D. Gutmann, B. Livneh, J. Dozier, and P. Neiman, 2015: High-Elevation Precipitation Patterns: Using Snow Measurements to Assess Daily Gridded Datasets across the Sierra Nevada, California. J. Hydrometeorol., 16, 1773–1792, https://doi.org/10.1175/JHM-D-15-0019.1.

Mantua, N., M. Dettinger, T. Pagano, A. Wood, K. Redmond, and P. Restrepo, 2008: A Description and Evaluation of Hydrologic and Climate Forecast and Data Products that Support Decision-Making for Water Resource Managers Decision-Support Experiments and Evaluations using Seasonal to Interannual Forecasts and Observational Data: A Focus. 29–64 pp. https://pdfs.semanticscholar.org/ad74/f7701476a309e366190b246936fe0e150a7d.pdf?_ga=2.113570

47.2146745698.1560960936-1183307250.1560960936 (Accessed June 19, 2019).

Mathevet, T., C. Michel, V. Andréassian, and C. Perrin, 2006: Large Sample Basin Experiments for Hydrological Model Parameterization: Results of the Model Parameter Experiment-MOPEX A bounded version of the Nash-Sutcliffe criterion for better model assessment on large sets of basins. IAHS Publ, 211–219 pp. https://iahs.info/uploads/dms/13614.21--211-219-41-MATHEVET.pdf (Accessed July 5, 2019).

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. Process, 19, 4023–4038, https://doi.org/10.1002/hyp.5869.

Melsen, L. A., P. J. J. F. Torfs, R. Uijlenhoet, and A. J. Teuling, 2017: Comment on "Most computational hydrology is not reproducible, so is it really science?" by Christopher Hutton et al. Water Resour. Res., 53, 2568–2569, https://doi.org/10.1002/2016WR020208.

Mendoza, P. A., M. P. Clark, M. Barlage, B. Rajagopalan, L. Samaniego, G. Abramowitz, and H. Gupta, 2015: Are we unnecessarily constraining the agility of complex process-based models? Water Resour. Res., 51, 716–728, https://doi.org/10.1002/2014WR015820.

Milly, P. C. D., and Coauthors, 2008: Climate change. Stationarity is dead: whither water management? Science, 319, 573–574, https://doi.org/10.1126/science.1151915.

Milly, P. C. D., and Coauthors, 2015: On Critiques of "stationarity is Dead: Whither Water Management?" Water Resour. Res., 51, 7785–7789, https://doi.org/10.1002/2015WR017408.

Molotch, N. P., and R. C. Bales, 2006: SNOTEL representativeness in the Rio Grande headwaters on the basis of physiographics and remotely sensed snow cover persistence. Hydrological Processes, Vol. 20 of, John Wiley & Sons, Ltd, 723–739.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith, 2007: Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Am. Soc. Agric. Biol. Eng., 50, 885–900.

Mote, P. W., 2003: Trends in snow water equivalent in the Pacific Northwest and their climatic causes. Geophys. Res. Lett., 30, https://doi.org/10.1029/2003GL017258.

Mote, P. W., 2006: Climate-Driven Variability and Trends in Mountain Snowpack in Western North America*. www.wcc.nrcs. (Accessed March 7, 2019).

Mote, P. W., S. Li, D. P. Lettenmaier, M. Xiao, and R. Engel, 2018: Dramatic declines in snowpack in the western US. npj Clim. Atmos. Sci., 1, 2, https://doi.org/10.1038/s41612-018-0012-1.

Nash, J. E., and J. V Sutcliffe, 1970: RIVER FLOW FORECASTING THROUGH CONCEPTUAL MODELS PART I-A DISCUSSION OF PRINCIPLES*. © North-Holland Publishing Co, 282–290 pp. https://hydrology.agu.org/wp-content/uploads/sites/19/2016/04/NashSutcillfe1.pdf (Accessed August 13, 2019).

National Weather Service, 2016: Sacramento Soil Moisture Accounting (SAC-SMA) Model. http://www.nws.noaa.gov/oh/hrl/nwsrfs/users_manual/part2/_pdf/23sacsma.pdf (Accessed March 2, 2020).

Newman, A. J., and Coauthors, 2015: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. Hydrol. Earth Syst. Sci., 19, 209–223, https://doi.org/10.5194/hess-19-209-2015.

Newman, A. J., N. Mizukami, M. P. Clark, A. W. Wood, and B. Nijssen, 2017: Benchmarking of a Physically Based Hydrologic Model. J. Hydrometeorol., 18, 2215–2225, https://doi.org/10.1175/JHM-D-16-0284.s1.

Pagano, T., D. Garen, S. Sorooshian, T. Pagano, D. Garen, and S. Sorooshian, 2004: Evaluation of Official Western U.S. Seasonal Water Supply Outlooks, 1922–2002. J. Hydrometeorol., 5, 896–909, https://doi.org/10.1175/1525-7541(2004)005<0896:EOOWUS>2.0.CO;2.

——, A. Wood, K. Werner, and R. Tama-Sweet, 2014: Western U.S. Water Supply Forecasting: A Tradition Evolves. Eos, Trans. Am. Geophys. Union, 95, 28–29, https://doi.org/10.1002/2014EO030007.

Pederson, G. T., and Coauthors, 2011: The unusual nature of recent snowpack declines in the North American Cordillera. Science (80-.)., 333, 332–335, https://doi.org/10.1126/science.1201570.

Ralph, F., and Coauthors, 2014a: A Vision for Future Observations for Western U.S. Extreme Precipitation

and Flooding. 16–32 pp. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1936-704X.2014.03176.x (Accessed May 3, 2019).

Ralph, F. M., and Coauthors, 2014b: A Vision for Future Observations for Western U.S. Extreme Precipitation and Flooding. J. Contemp. Water Res. Educ., 153, 16–32, https://doi.org/10.1111/j.1936-704x.2014.03176.x.

Rasmussen, R., and Coauthors, 2011: High-Resolution Coupled Climate Runoff Simulations of Seasonal Snowfall over Colorado: A Process Study of Current and Warmer Climate. J. Clim., 24, 3015–3048, https://doi.org/10.1175/2010JCLI3985.1.

Refsgaard, J. C., and J. Knudsen, 1996: Operational validation and intercomparison of different types of hydrological models. Water Resour. Res., 32, 2189–2202, https://doi.org/10.1029/96WR00896.

Regonda, S. K., B. Rajagopalan, M. Clark, J. Pitlick, S. K. Regonda, B. Rajagopalan, M. Clark, and J. Pitlick, 2005: Seasonal Cycle Shifts in Hydroclimatology over the Western United States. J. Clim., 18, 372–384, https://doi.org/10.1175/JCLI-3272.1.

Rosenberg, E. A., A. W. Wood, and A. C. Steinemann, 2013: Informing Hydrometric Network Design for Statistical Seasonal Streamflow Forecasts. J. Hydrometeorol., 14, 1587–1604, https://doi.org/10.1175/JHM-D-12-0136.1.

Sawicz, K., T. Wagener, M. Sivapalan, P. A. Troch, and G. Carrillo, 2011: Catchment classification: empirical analysis of hydrologic Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA Catchment classification: empirical analysis of hydrologic. Hydrol. Earth Syst. Sci. Discuss, 8, 4495–4534, https://doi.org/10.5194/hessd-8-4495-2011.

Schaake, J. C., and E. L. Peck, 1985: Analysis of Water Supply Forecast Accuracy. Western Snow Conference Proceedings, Vol. 53 of, 44–53 https://westernsnowconference.org/sites/westernsnowconference.org/PDFs/1985Schaake.pdf (Accessed September 17, 2019).

Stewart, I. T., D. R. Cayan, M. D. Dettinger, I. T. Stewart, D. R. Cayan, and M. D. Dettinger, 2005: Changes toward Earlier Streamflow Timing across Western North America. J. Clim., 18, 1136–1155, https://doi.org/10.1175/JCLI3321.1.

Swain, D. L., B. Langenbrunner, J. D. Neelin, and A. Hall, 2018: Increasing precipitation volatility in twenty-first-century California. Nat. Clim. Chang., 8, 427–433, https://doi.org/10.1038/s41558-018-0140-y.

U.S. Department of Agriculture, N. R. C. S., 2011: Part 622 Snow Survey and Water Supply Forecasting National Engineering Handbook: Chapter 7 Water Supply Forecasting. 1–19 pp. https://directives.sc.egov.usda.gov/OpenNonWebContent.aspx?content=32039.wba (Accessed August 13, 2019).

Wood, A. W., and D. P. Lettenmaier, 2006: A TEST BED FOR NEW SEASONAL HYDROLOGIC FORECASTING APPROACHES IN THE WESTERN UNITED STATES. Am. Meteorlogical Soc., 1699–1712, https://doi.org/10.1175/BAMS-87-I2-I699.

Woods, R. A., 2009: Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks. Adv. Water Resour., 32, 1465–1481, https://doi.org/10.1016/J.ADVWATRES.2009.06.011.

Yang, Z.-L., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. J. Geophys. Res., 116, D12110, https://doi.org/10.1029/2010JD015140.

Appendix 2.1: NRCS Principal Component Regression Forecast Development

Today, the NRCS issues over 600 spring water supply forecasts for the western U.S. and Alaska on a monthly to bi-weekly basis using the PCR technique. These deterministic forecasts begin on January 1st, are updated through the end of the snowmelt season, and are communicated through a variety of products available at: https://www.wcc.nrcs.usda.gov/wsf/. Previous water supply forecast assessment studies (e.g. Pagano et al. (2004)) make direct use of NRCS forecast results for their error assessments. Yet, it is well known that these forecasts are subject to the independent, professional judgement of the forecaster. For example, forecasted streamflow volumes may be adjusted to produce more consistent month-to-month values across a forecast season or to eliminate statistical noise (Garen, 1992). Though admittedly small, such adjustments confound forecast accuracy assessment and are not possible to control for. Accordingly, we follow the approaches of Harpold et al. (2017) and Lehner et al. (2017) to reproduce PCR-based water supply forecasts.

The PCR is the pillar on which modern NRCS water supply forecasting rests. The PCR is a regression technique based on a multivariate statistical process called Principal Component Analysis (PCA). PCA restructures intercorrelated variables into an eigenvalue-eigenvector matrix which provides a new set of uncorrelated variables, or principal components (PCs). The PCA aims to explain the most variance with the fewest PCs, thereby compressing the dimensionality of the original predictor variables. A subset of the PCs can then be used as independent variables in a regression equation (Jolliffe, 2002).

The NRCS water supply forecast prediction variables of P, SWE, antecedent Q- and to a lesser extent, climate indices- exhibit strong multicollinearity. For this reason, the PCR is a statistically more robust streamflow forecasting technique than other regression techniques. In a methods study, Garen (1992) argued that maximum forecast accuracy could be obtained by using monthly basin-specific forecasts which: 1) use PCR instead of multiple linear regression, 2) are informed only by data available at the time of prediction and not the long-term mean values, 3) are backed up through rigorous cross validation, and 4) whose predictor variables are systematically identified by selection algorithms. Chapter 7 of the National Engineering Handbook, Water Supply Forecasting, formalizes this procedure as the standard regression-based water supply forecasting methodology (U.S. Department of Agriculture, 2011).

To reduce the confounding effects of NRCS forecast development and forecaster adjustment, we employ a modified approach to the NRCS PCR method (U.S. Department of Agriculture, 2011) to recreate forecasts for all 51 study basins. The process of developing PCR forecasts begins with the standardization of n predictors, where n = 2 (SWE and P) x the number of SNOTEL stations in the forecast equation (Table 2.1). These predictors, or training data, are selected by the NRCS using an automated search algorithm developed by Garen (1992). We do not replicate this process and instead utilize the primary NRCS-selected predictor variables (SNOTEL-measured P and SWE), forgoing all other NRCS-selected predictors (e.g. antecedent Q) for consistency. Using P and SWE variables only had the added benefit of explicitly considering the sole effects of winter precipitation and snow on forecast performance. The standardized training dataset

comprised of n predictors are then statistically transformed by a PCA, producing n principal components (PCs). To develop a parsimonious model, we aim to select a subset of m<<n PCs to be used as the independent variable(s) in our forecast regression equations, where m represents the number of PCs used in the regression (here, m = 1). The selection process sequentially adds PCs to the regression equation, using a standard t-test to evaluate the statistical significance of the regression coefficient (slope) for each PC. The U.S. Department of Agriculture (2011) specifies a forecaster-selected t-statistic of 1.2 to 2.5, with a reasonable default defined as 1.6.

SNOTEL-measured daily P and SWE data were downloaded from the National Weather and Climate Center's Report Generator (https://wcc.sc.egov.usda.gov/reportGenerator/) for the period October 1st, 1980 to September 30th, 2014, i.e. water years (WYs) 1981 to 2014. SNOTEL stations were identified from the official NRCS-published forecast equations. Daily accumulated P was aggregated to end-of-month totals and daily end-of-month SWE values were extracted. Gap-filling of accumulated P and SWE data was minimal as most data gaps were generally found at the beginning of the period of record and thus inferred to be prior to SNOTEL station installation. Quality-controlled mean daily streamflow (Q) data were downloaded from the CAMELS dataset for all 51 basins for the same period, water years 1981 to 2014. Most (75%) of the study sites had serially complete spring Q records; those missing more than 10% of daily values for a given spring were omitted. In total, 96% (1,670 out of 1,734) of records were retained. As the predictand, no gap filling was performed on the streamflow data. An example PCR forecast for Sagehen Creek CA (ID#

27, USGS 10343500) in shown in Figure A2.1, with principal component loadings (i.e.

eigenvectors) of SWE and P for each forecast issue date.
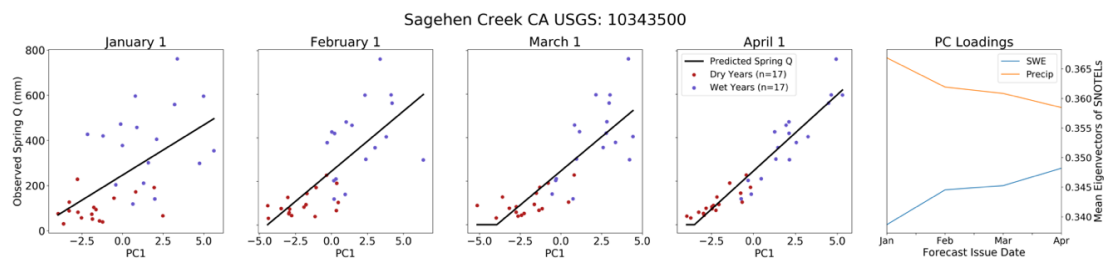
Figure A 2.1: Example PCR forecast for Sagehen Creek CA for four lead time forecasts (first four panels) with PC loadings of SWE and P variables (last panel). Below (above) average peak SWE years are shown in red (blue); forecast regression equations are the black lines. Note that forecasts were corrected to be non-negative (e.g. low flows in March and April forecasts).

Appendix 2.2: NWS Sacramento Soil Moisture Accounting Model Forecast Development

The NWS River Forecast System (NWS RFS) is implemented by 13 regional River Forecast Centers (RFCs) to issue streamflow predictions for short lead time flood forecasts and select longer lead time water supply forecasts for thousands of locations in the United States. Ensemble Streamflow Prediction (ESP) techniques were developed in the 1970s by the NWS and collaborates for use in long-term hydrologic predictions. At the core of the NWS RFS and ESP is the Sacramento Soil Moisture Accounting (SAC-SMA) model, a spatially lumped conceptual rainfall-runoff model (Burnash et al., 1973) coupled to the SNOW17 temperature index snow model (Anderson, 1973).

SAC-SMA is a continuous soil moisture accounting model that uses soil water reservoirs with spatially lumped parameters to simulate runoff, which can be routed to produce streamflow predictions. The reservoir zones represent the upper and lower part of the soil column and simulate processes such as groundwater storage, baseflow, interflow, and surface runoff. The ESP approach uses a deterministic simulation of the initial hydrologic conditions during a model initialization period. Model spin-up then provides hydrologic conditions (i.e. initiate state variables) on the forecast start date. An ensemble of historical meteorology (e.g. precipitation and temperature) is then used to force the model over the forecast period.

We calibrated the SAC-SMA/SNOW17 models using a Shuffle Complex Evolution (SCE) to optimize model performance as forced by lumped basin mean daily Daymet data for the period October 1 1989 to September 31 1999 (Addor et al., 2017).

Calibrated models were then used to make NWS RFS seasonal water supply forecasts for April 1 – July 31 for WYs 1981-1989 and 2000-2014 following ESP techniques. This process was repeated for each of the four NRCS lead times, including January, February, March, and April 1st. SAC-SMA forecasts were reported as probabilistic values for the mean model runs from the ESP approach. An example April 1st (dark grey) SAC-SMA WY 1900 forecast for Sagehen Creek CA (ID# 27, USGS 10343500) in shown in Figure A2.2, with a 'perfect prognostic' (Figure 2.3a) model run forced with daily data in black. Individual model runs are in lighter grey traces.
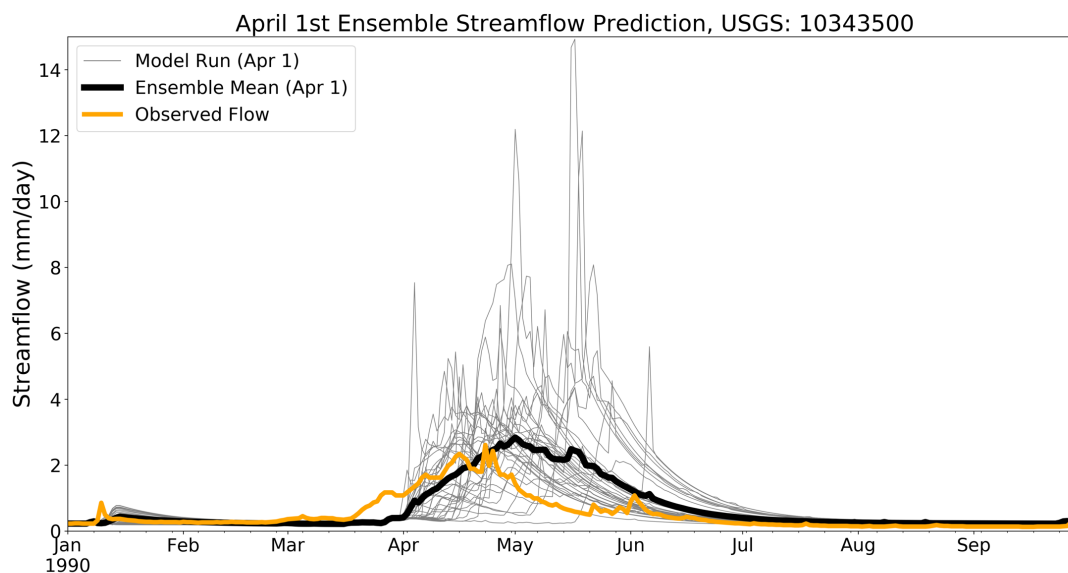
Figure A2.2: Example April 1st SAC-SMA ensemble mean forecast for WY1990 in Sagehen Creek CA (USGS 10343500). Ensemble forcing data were provided to the model from forecast issue date until the end of the forecast period using an ensemble technique.

Appendix 2.3: Statistical Forecast Assessment Metrics

PBIAS measures the average tendency of the forecasted data to be larger or smaller than their observed counterparts; an optimal value is 0 and low magnitude values indicate accurate, non-biased forecast performance (Gupta et al., 1999). Negative values indicate systematic forecast overprediction and positive values indicate systematic forecast underestimation:

$$PBIAS = \frac{\sum_{i=1}^{n}(Q_{obs,i} - Q_{sim,i}) * 100}{\sum_{i=1}^{n}(Q_{obs,i})}$$

The original Nash Sutcliffe Efficiency (Nash & Sutcliffe, 1970), a widely-used goodness of fit measure:

$$NSE \ skill \ score = 1 - \frac{\sum_{i=1}^{n}(Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^{n}(Q_{obs,i} - \overline{Q_{obs}})^2}$$

The Nash Sutcliffe Efficiency (NSE) skill score measures the ratio of the sum of squares of the observed streamflow ($Q_{obs,i}$) and the forecasted streamflow ($Q_{sim,i}$) and the initial variance, a sum of squares between the observed streamflow ($Q_{obs,i}$) and the mean of the observed streamflow ($\overline{Q_{obs}}$). The NSE score is a normalized indicator of model performance in relation to a benchmark, in our case, mean observed streamflow (Gupta et al., 1999). Values vary between the perfect score of 1 and -∞; values less than 0 are less skillful than the mean of the benchmark.

Mathevet et al. (2006) argue that across large sample studies, model assessment will

inevitably be challenged by a wide range of model performances, particularly model failures. When quantifying the average model performance, a metric such as the traditional NSE score with no lower bound may therefore give strong negative values. Hence, Mathevet et al. (2006) propose a bottom-end bounded version of the NSE score (C2M) designed specifically for the assessment of hydrologic models across large sample studies:

$$C_2 M = \frac{NSE}{2 - NSE}$$

The C2M criterion varies between 1 and -1, producing less optimistic values than the traditional metric for positive values and offering overall more meaningful statistics for summarizing forecast performance across our study basins. The C2M was our primary metric of consideration for forecast assessment and improvement. For clarity, we refer to it as the NS score, or simply, NSE. In comparison to metrics such as the relative root mean square error (rRMSE), NSE is normalized by the variance of the observed values (Figure A2.3).

April 1st Forecasts
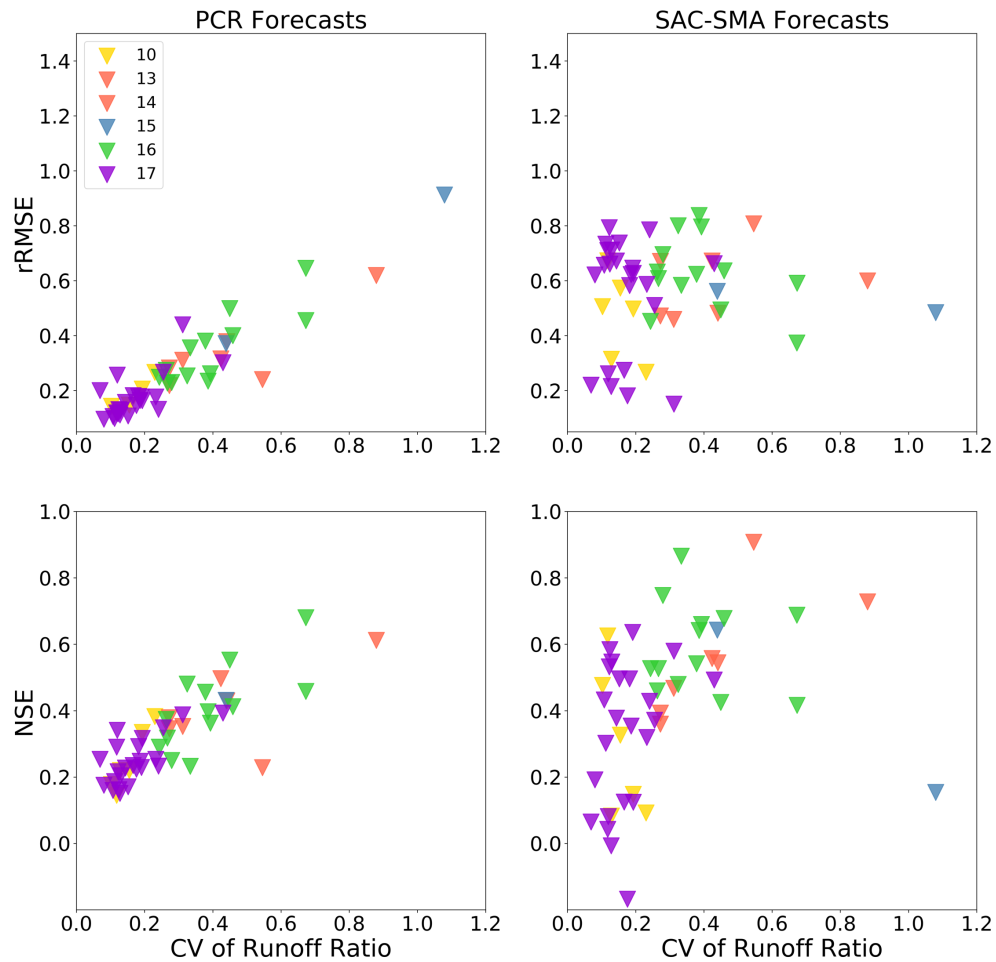
Figure A2.3: Relative RMSE and Nash Sutcliffe Efficiency error metrics as a function of the coefficient of variation of the runoff ratio for April 1st PCR forecasts (left column) and SAC-SMA forecasts (right column). As a metric, NSE scores are normalized by variance. Colored by regional hydrologic unit: Missouri (10), Upper/Lower Colorado (13-14), Rio Grande (15), Great Basin (16), and Pacific Northwest (17).

Appendix 2.4: Supporting Figures
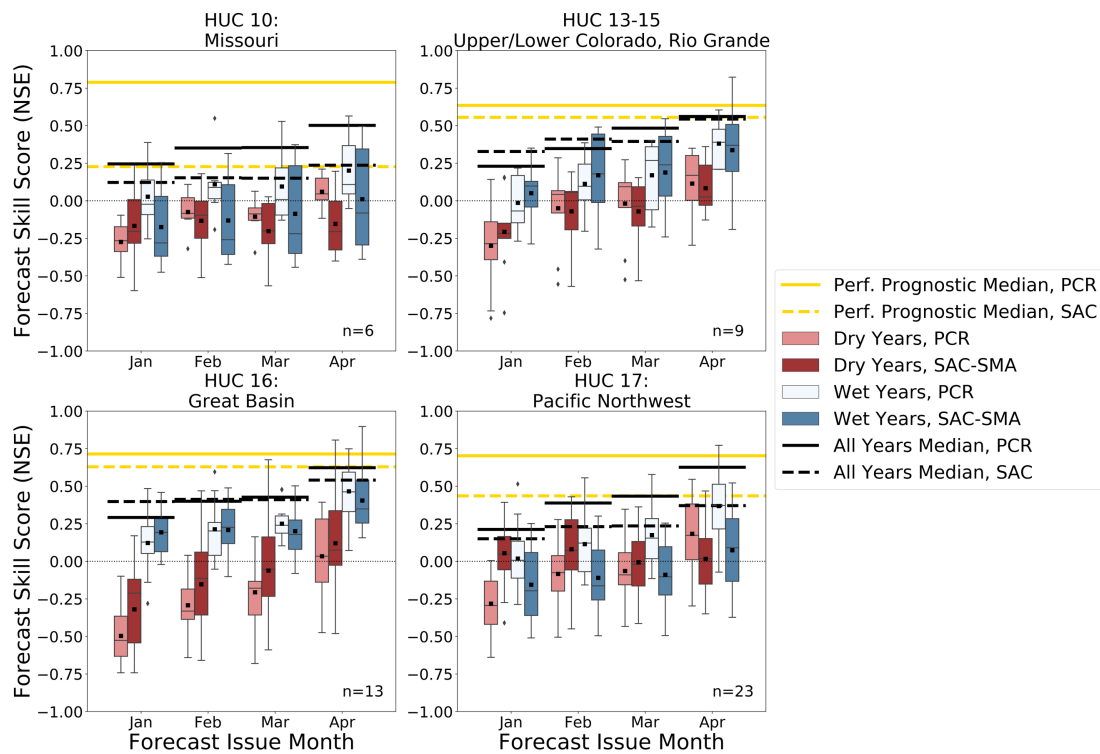
Figure A2.4: Forecast evaluations during above (blue) and below (red) average peak SWE for a bounded Nash Sutcliffe Efficiency (NSE) score across four lead times for each regional HUC group. Lighter colors are PCR forecasts, darker colors are SAC-SMA forecasts. Black bars show median forecast performance when evaluated across all years; yellow bars are forecast performance when models are provided data through the end of the forecasting period.
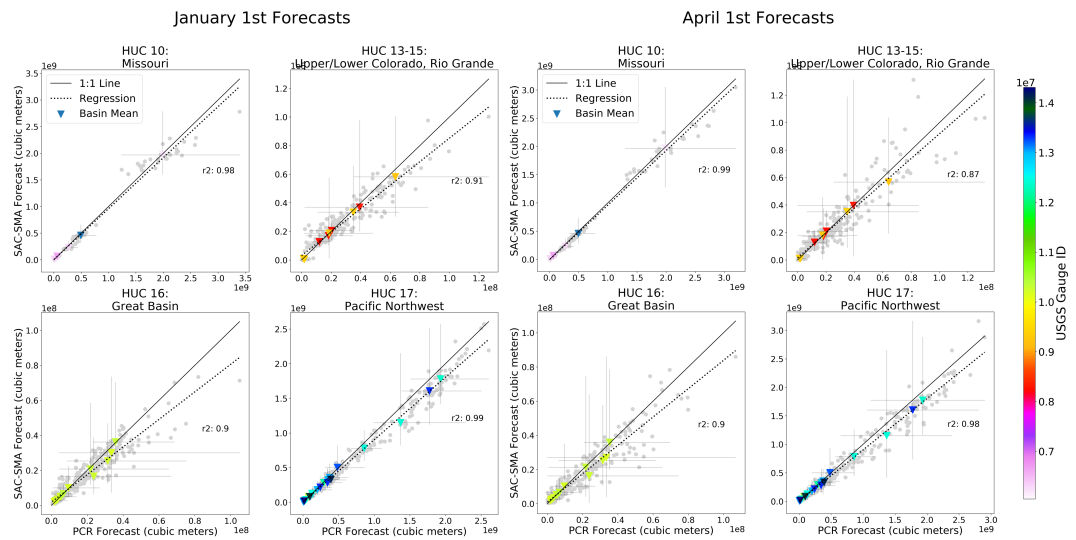
Figure A2.5: January 1st and April 1st volumetric streamflow forecasts from SAC-SMA (y-axis) and PCR (x-axis) for each regional hydrologic unit. Basin mean values are colored by USGS gauge ID. Units are in cubic meters.
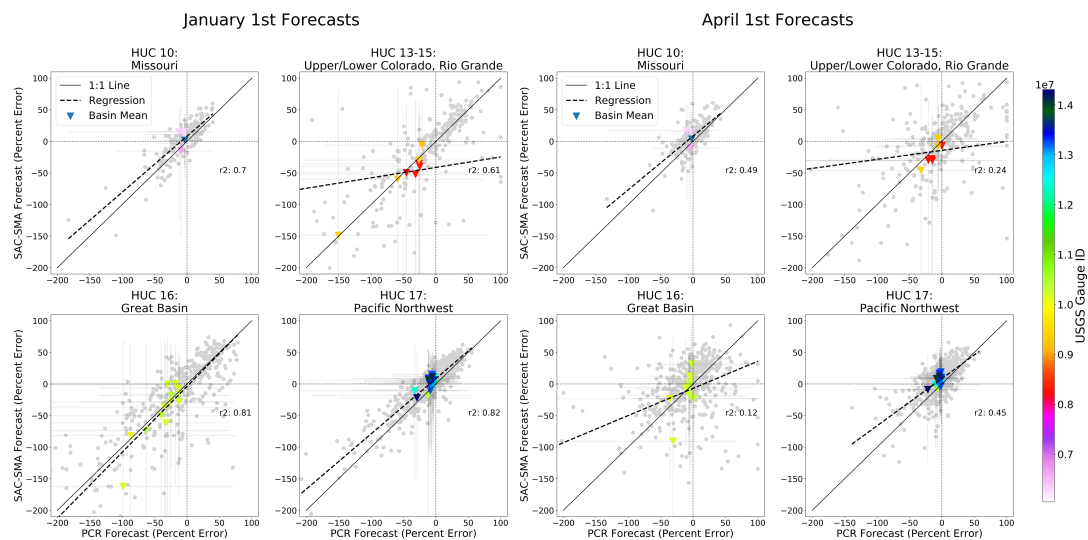
Figure A2.6: January 1st and April 1st volumetric streamflow forecasts evaluations (percent error) from SAC-SMA (y-axis) and PCR (x-axis) for each regional hydrologic unit. Basin mean values are colored by USGS gauge ID. Units are in cubic meters.
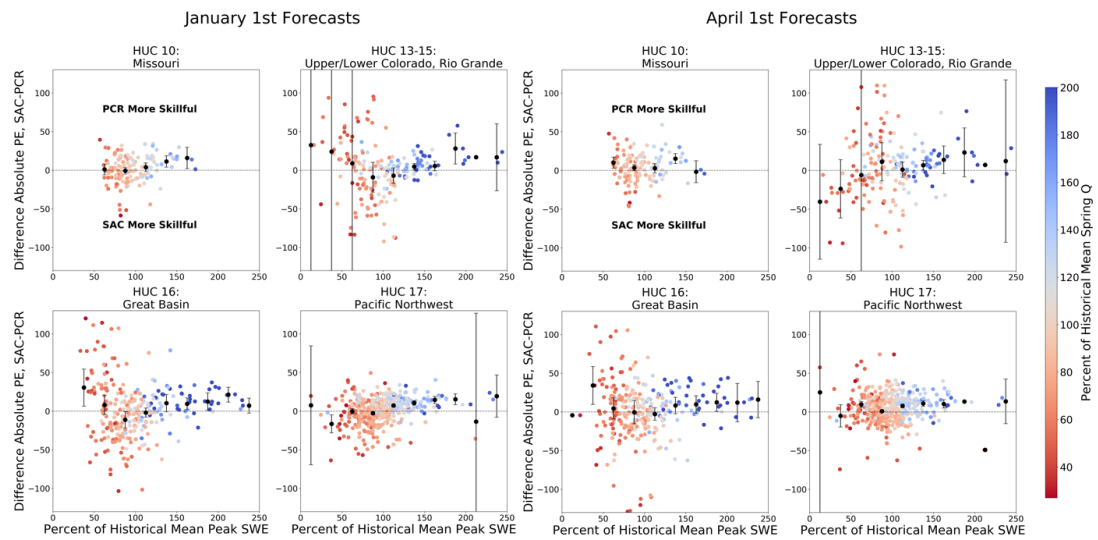
Figure A2.7: Differences between the absolute percent errors of the SAC-SMA and PCR model forecasts from January 1st (left four panels) and April 1st (right four panels) for each regional HUC group. Median values are bounded by a 90% confidence interval (vertical bars) for 25% SWE bins (e.g. 0-25% mean historical peak SWE). SWE values are SNOTEL observations. Colored by percent of historical mean spring Q.
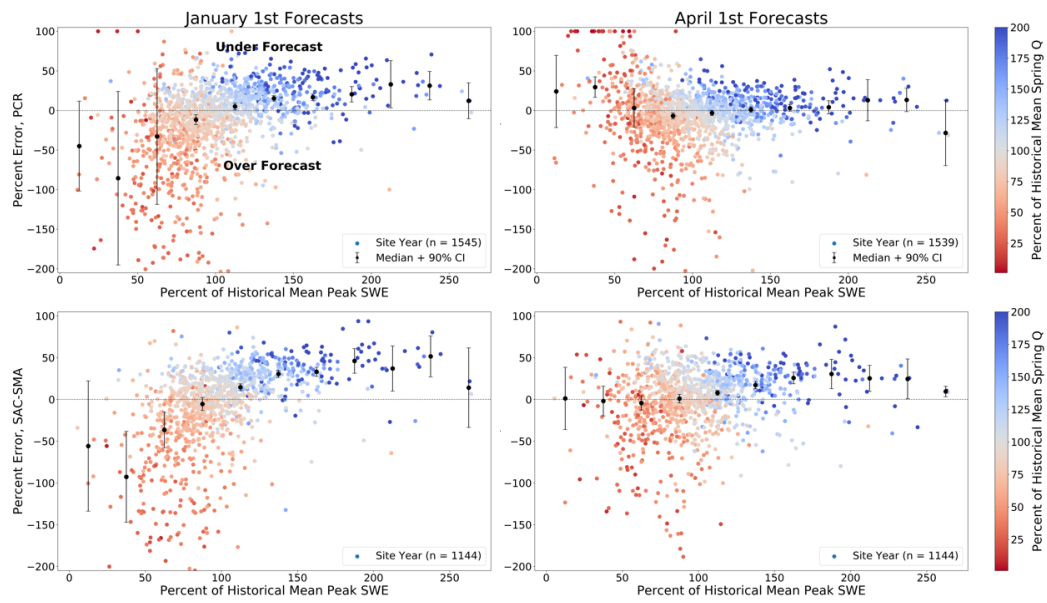
Figure A2.8: Percent error of January 1st forecasts (left column) and April 1st forecasts (right columns) for the PCR (top row) and SAC-SMA (middle row) models as a function of percent of historical peak SWE. Median values are bounded by a 90% confidence interval (vertical bars) for 25% SWE bins (e.g. 0-25% mean historical peak SWE). Colored by percent of mean spring Q.
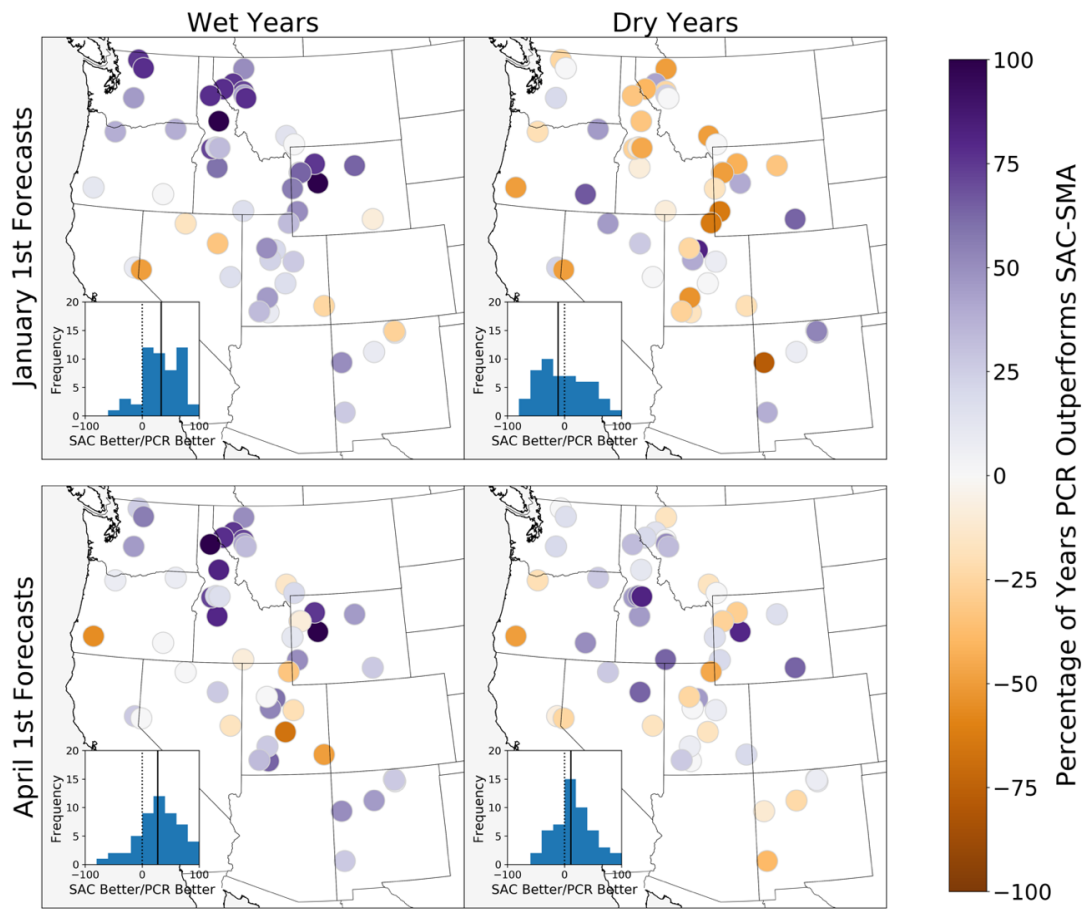
Figure A2.9: Percentage of years PCR forecasts are better than SAC-SMA forecasts for wet and dry years for January 1st and April 1st forecasts. Negative (orange) values are the percentage of years SAC-SMA is better than PCR. Inset histograms show the distribution across all sites; the solid black vertical bar denotes the median.

Figure A2.10: Statistically significant (p-value < 0.05) Sen's Slope values from a Regional Mann Kendall Trend Test for SNOTEL measured beginning of the month SWE values (left) and SNOTEL measured beginning of the month water year accumulated precipitation (right) for January 1st (purple), February 1st (teal), March 1st (light green), and April 1st (red) across the four HUC groups.

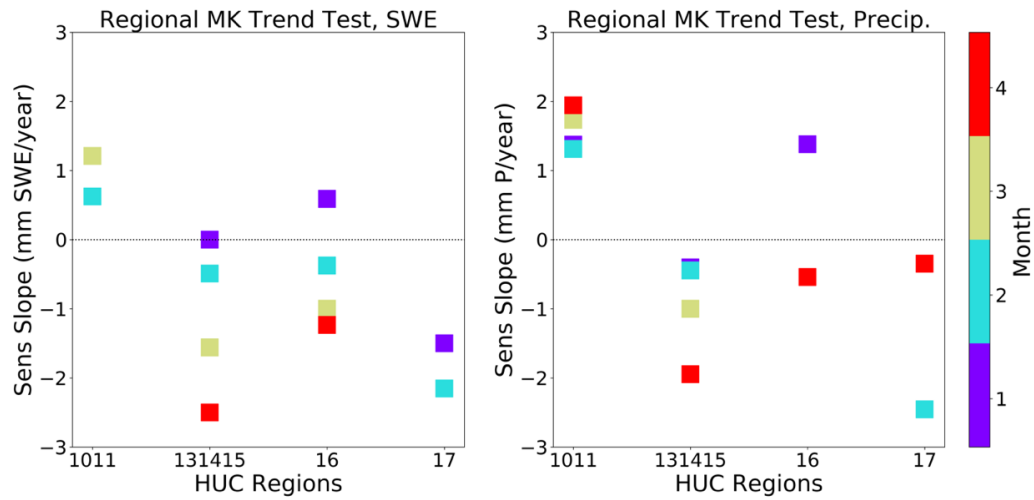| Site ID	USGS ID	Station Name and Location	Mean Spring Q	Mean PCR Forecast	Mean SAC Forecast	Mean Spring Q (Dry)	Mean PCR (Dry)	Mean SAC (Dry)	Mean Spring Q (Wet)	Mean PCR (Wet)	Mean SAC (Wet)
1	06043500	GALLATIN RIVER NEAR GALLATIN GATEWAY MT	397.7	399.9	372.5	329.6	344.8	329.5	480.4	459.0	415.5
2	06191500	YELLOWSTONE RIVER AT CORWIN SPRINGS MT	1616.3	1610.7	1593.3	1325.4	1323.3	1376.5	1907.2	1898.2	1810.1
3	06221400	DINWOODY CREEK ABOVE LAKES, NEAR BURRIS, WYO.	66.4	66.4	55.1	58.4	59.9	50.1	75.9	74.0	61.6
4	06280300	SOUTH FORK SHOSHONE RIVER NEAR VALLEY, WY	221.1	219.9	187.5	177.6	184.9	158.0	264.6	254.9	216.9
5	06311000	NORTH FORK POWDER RIVER NEAR HAZELTON, WY	9.4	9.4	7.2	7.1	7.8	5.7	11.6	11.0	8.7
6	06632400	ROCK CREEK AB KING CANYON CANAL, NR ARLINGTON, WY	50.5	50.4	51.8	36.0	39.5	44.7	65.0	61.4	58.9
7	08267500	RIO HONDO NEAR VALDEZ, NM	16.5	16.5	16.9	11.0	11.3	13.9	21.7	21.6	19.9
8	08269000	RIO PUEBLO DE TAOS NEAR TAOS, NM	14.8	14.8	13.9	7.9	8.4	9.4	21.7	21.2	18.4
9	08271000	RIO LUCERO NEAR ARROYO SECO, NM	9.7	9.7	10.4	6.6	6.6	8.2	12.8	12.8	12.5
10	08324000	JEMEZ RIVER NEAR JEMEZ, NM	31.8	31.9	32.1	16.5	15.9	16.1	47.1	48.0	48.1
11	09223000	HAMS FORK BELOW POLE CREEK, NEAR FRONTIER, WY	54.9	52.0	45.9	31.5	31.9	34.9	78.3	72.2	57.0
12	09312600	WHITE RIVER BL TABBYUNE C NEAR SOLDIER SUMMIT, UT	15.1	14.3	14.5	5.5	5.6	6.0	24.7	23.6	23.1
13	09378170	SOUTH CREEK ABOVE RESERVOIR NEAR MONTICELLO, UT	1.0	1.1	0.8	0.3	0.3	0.3	1.6	1.6	1.4
14	09386900	RIO NUTRIA NEAR RAMAH, NM	1.4	1.5	0.6	0.3	0.4	0.1	3.0	3.2	1.1
15	09430500	GILA RIVER NEAR GILA, NM	28.3	28.3	28.4	15.8	15.2	12.1	40.8	40.8	44.7
16	10023000	BIG CREEK NEAR RANDOLPH, UT	4.1	4.1	4.2	1.8	2.0	2.3	7.1	6.9	6.6
17	10166430	WEST CANYON CREEK NEAR CEDAR FORT, UT	1.8	1.8	2.0	0.8	0.8	1.2	3.3	3.2	2.9
18	10172700	VERNON CREEK NEAR VERNON, UT	1.3	1.4	1.3	0.7	0.6	0.7	2.0	2.1	1.9
19	10172800	SOUTH WILLOW CREEK NEAR GRANTSVILLE, UT	3.1	2.9	2.8	2.0	2.0	2.2	4.2	4.0	3.4
20	10173450	MAMMOTH CREEK ABV WEST HATCH DITCH, NEAR HATCH, UT	26.9	27.1	21.9	15.7	13.5	12.7	38.1	40.0	31.1
21	10205030	SALINA CREEK NEAR EMERY, UT	7.7	7.7	8.3	4.6	4.2	5.5	10.9	11.3	11.1
22	10234500	BEAVER RIVER NEAR BEAVER, UT	25.5	25.5	20.4	15.7	14.7	13.3	35.2	36.2	27.5
23	10242000	COAL CREEK NEAR CEDAR CITY, UT	19.1	19.2	13.2	10.4	9.2	5.7	27.7	28.6	20.7
24	10244950	STEPTOE C NR ELY, NV	2.4	2.4	2.2	1.1	1.1	1.4	3.7	3.7	2.9
25	10316500	LAMOILLE C NR LAMOILLE, NV	28.5	28.7	29.0	20.2	22.0	22.2	36.9	35.3	35.9
26	10329500	MARTIN C NR PARADISE VALLEY, NV	17.3	17.6	17.2	8.0	7.3	8.8	26.7	27.4	25.6
27	10343500	SAGEHEN C NR TRUCKEE CA	5.5	5.5	5.3	2.3	2.3	3.2	8.7	8.8	7.4
28	10348850	GALENA C AT GALENA C STATE PARK	4.5	4.5	4.0	2.6	2.6	2.6	6.3	6.3	5.5
29	10396000	DONNER UND BLITZEN RIVER NR FRENCHGLEN OR	61.7	61.7	62.1	41.4	43.0	48.0	82.1	80.4	76.3
30	12175500	THUNDER CREEK NEAR NEWHALEM, WA	235.3	238.0	233.5	222.8	221.2	224.9	250.0	252.4	242.4
31	12358500	MIDDLE FORK FLATHEAD RIVER NR WEST GLACIER MT	1566.7	1566.7	1437.2	1299.0	1314.7	1211.2	1834.4	1818.6	1663.3
32	12374250	MILL CR AB BASSOO CR NR NIARADA MT	4.7	5.1	4.0	3.2	3.5	3.1	6.3	6.9	5.1
33	12375900	SOUTH CROW CREEK NEAR RONAN MT	10.5	10.7	9.1	9.2	9.6	8.2	12.0	12.0	10.2
34	12377150	MISSION CR AB RESERVOIR NR ST IGNATIUS MT	26.1	26.6	22.3	23.9	25.0	19.9	28.5	28.6	25.1
35	12381400	SOUTH FORK JOCKO RIVER NEAR ARLEE MT	34.3	36.5	30.4	27.6	28.3	23.5	41.4	43.3	37.0
36	12390700	PROSPECT CREEK AT THOMPSON FALLS MT	106.1	105.0	94.1	78.5	80.4	66.0	133.7	131.4	122.3
37	12414500	ST JOE RIVER AT CALDER ID	1095.1	1110.8	932.7	867.6	814.1	739.0	1322.7	1361.9	1126.4
38	12451000	STEHEKIN RIVER AT STEHEKIN, WA	696.2	697.1	634.1	595.8	608.3	548.3	796.5	791.7	720.0
39	12488500	AMERICAN RIVER NEAR NILE, WA	104.7	104.1	102.8	84.3	87.2	83.1	125.0	122.1	122.5
40	13011500	PACIFIC CREEK AT MORAN WY	164.6	163.0	143.6	119.7	118.9	109.7	209.6	207.1	177.5
41	13011900	BUFFALO FORK AB LAVA CREEK NR MORAN WY	285.8	283.3	262.3	224.4	224.9	213.5	347.1	341.8	311.2
42	13023000	GREYS RIVER AB RESERVOIR NR ALPINE WY	311.6	304.2	286.5	227.4	233.5	239.4	395.9	379.5	333.7
43	13083000	TRAPPER CREEK NR OAKLEY ID	4.9	4.7	4.7	3.3	3.5	3.8	6.5	5.8	5.7
44	13235000	SF PAYETTE RIVER AT LOWMAN ID	405.4	391.3	406.1	297.4	288.3	293.9	513.3	501.1	518.3
45	13240000	LAKE FORK PAYETTE RIVER AB JUMBO CR NR MCCALL ID	81.2	79.4	64.6	64.7	64.3	50.2	97.6	99.5	78.9
46	13317700	SF SALMON RIVER NR KRASSEL RANGER STATION ID	278.7	279.0	226.6	216.4	209.4	169.3	341.0	343.9	284.0
47	13313000	JOHNSON CREEK AT YELLOW PINE ID	196.5	196.1	172.9	152.4	146.7	133.1	240.6	242.7	212.6
48	13337000	LOCHSA RIVER NR LOWELL ID	1451.5	1439.9	1298.8	1183.6	1150.6	1054.7	1719.3	1693.0	1542.9
49	14020000	UMATILLA RIVER ABOVE MEACHAM CREEK, NR GIBBON, OR	76.4	76.4	67.9	61.1	63.3	49.3	91.7	89.5	86.5
50	14137000	SANDY RIVER NEAR MARMOT, OR	318.1	317.6	268.9	280.9	282.5	250.2	355.3	350.5	287.7
51	14308990	COW CREEK ABV GALESVILLE RES, NR AZALEA, OR.	14.2	14.2	13.2	9.7	11.8	10.9	18.2	16.8	15.4
			200.3	**199.7**	**183.2**	**161.0**	**160.2**	**151.0**	**239.9**	**238.0**	**215.4**

Table A2.1: Observed and forecasted spring volumetric discharge (thousands of acre-feet) for all years, dry years, and wet years for both PCR and SAC-SMA models. Mean values are in bold on the last

Chapter 3: Potential to Mitigate Skill of Seasonal Water Supply Forecasts From Snowpack Loss in the Sierra Nevada, USA

Abstract

Seasonal water supply outlooks use statistical techniques to relate winter precipitation stored as snowpack with spring (April-July) streamflow volumes. Low snowpack alters streamflow generation which challenges regression-based forecasts like those used in the Sierra Nevada, USA. In this study, we use data from the Variable Infiltration Capacity hydrologic model to simulate seasonal water supply forecasts in the Sierra Nevada from 1950-2099 under low and high emissions scenarios. Our simulations project that forecast skill will decline steadily through the 21st century, with median loss of skill of 12-22% by the second half of the century, most pronounced in mid-elevation basins (1000-1700 m mean elevation). We demonstrated that simulations of remotely-sensed snowpack can replace an average of 40% of lost forecast skill in the most vulnerable basins through midcentury. As snowpack declines, simulations show that the loss of forecast information once provided by snowpack can be partially replaced by soil moisture observations.

3.1 Introduction

In the Sierra Nevada of California and Nevada, winter precipitation stored as snowpack at higher elevations has historically provided reliable streamflow and consistent late season runoff from the headwater basins. Summertime water availability from Sierra Nevada is critical for California's urban centers and $2.7 trillion state GDP, particularly the $50 billion agricultural industry (California Department of Food and Agriculture, 2018). Snowmelt is also essential for in-stream environmental stream flows and temperatures. Appropriation of water resources by the California State Water Resources Control Board to water rights holders, including critical water allocations during drought years, relies on seasonal water supply outlooks issued by the California Department of Water Resources (CA DWR). These streamflow volume forecasts apply regression relations to winter snowpack and precipitation measured in the mountains to predict streamflow discharge volume in the summer, usually for reservoir inflows. Formal forecast guidance from these regression-based models is issued by CA DWR for the state's major watersheds four times a year in their Bulletin-120 (http://cdec.water.ca.gov/snow/bulletin120/), which is similar to other regression-based forecasts issued throughout the western U.S. by the Natural Resource Conservation Service (NRCS) National Water and Climate Center. Unfortunately, declining snowpack (Mote et al., 2018) drives earlier, more intermittent, and slower snowmelt (A. Harpold et al., 2012; A. A. Harpold & Brooks, 2018; Musselman et al., 2017), and loss of snow storage fundamentally alters streamflow generation and groundwater recharge (Berghuijs et al., 2014; Earman & Dettinger, 2011; Hammond et al., 2019; Mcnamara et al., 2005). The consequences of loss of information from snow storage (Livneh & Badger, 2020)

combined with nonstationary changes in hydrologic processes (Hammond et al., 2019)

threatens the reliability of regression-based water supply forecasts (Sturtevant et al.,

2020).

Increasing temperatures from climate change has caused earlier streamflow in the

Sierra Nevada (Stewart et al., 2005) that will continue to shift earlier (Bureau of

Reclamation, 2016), while long-term water yields are typically not projected to change

significantly (Dettinger et al., 2004). New hydrologic measurements, modeling systems,

and analytical techniques (Kirchner, 2006) offer promise for predicting how declining

snowpack alters streamflow generation. Despite providing valuable forecast skill with

gridded, physically based models (Koster et al., 2010; Wood & Lettenmaier, 2006), these

more sophisticated tools are typically too complex (Beven, 2002) or do not show

sufficient skill benefits compared to simpler, statistical models (Sturtevant et al., 2020) to

have motivated their widespread adoption in operations yet. These changes (and

competition from the newer models) are motivating new improvements to the regression-

based streamflow volume forecasts that are widely relied on by water managers. For

example, incorporating climate information (e.g. El Nino-Southern Oscillation climate

indices) is now standard practice for the Natural Resource Conservation Service (NRCS)

following successful research efforts by Yao and Georgakakos (2001) and Hamlet et al.

(2002). Additional water supply forecast predictor variables that have received

consideration to improve regression-based forecasts include soil moisture (A. A. Harpold

et al., 2017; Rosenberg et al., 2013) and satellite-based snow product (Rango, 1980).

Projected declines in snowpack storage (Fyfe et al., 2017) and more frequent snow

drought, or the combination of meteorological drought and reduced snow storage (A.

Harpold et al., 2017), and more intense winter rainfall are expected to stress existing water supply forecasting techniques like those used in California (Sturtevant et al., 2020).

In this study, we project future resilience of statistical seasonal water supply forecasts in the Sierra Nevada. Because testing future skill of the regression forecasts is inherently limited by the lack of future observations and since historical conditions may provide a weak analog for the future (Cooper et al., 2016), we use output from process-based hydrology model (Variable Infiltration Capacity, VIC; Liang et al., 1994) to generate forecasts for the same basins as California DWR's B120 forecast. Long-model records allow us to create forecasts on February 1st and April 1st for water years (WY) 1951-2099 using statistically downscaled projections from 10 global circulation models (GCMs) under RCP 4.5 and 8.5 future climate pathways (Pierce et al., 2018). We then assess the capacity of relevant supplemental observations to improve simulated forecasts: 1) a basin-wide SWE snapshot representing a virtual Airborne Space Observatory (ASO) lidar flight (Painter et al., 2016) and 2) soil moisture measurements co-located with existing CA DWR observation stations. These standard (conventional) and enhanced forecasts are evaluated are compared between historical, mid-21st century, and late 21st century to answer two questions:

1) Do seasonal water supply forecast skills, based on snowpack and precipitation, degrade in the 21st century as snowpacks decline?

2) To what extent can enhanced forecasts using supplemental observations mitigate declines in skill among basins with different characteristics?

3.2 Data and Methods

CA DWR issues April 1 – July 31 streamflow volume forecasts four times annually on February 1st, March 1st, April 1st, and May 1st for twenty headwater basins in the Sierra Nevada and northern California (http://cdec.water.ca.gov/snow/bulletin120/). For the purposes of this study, we represent the Bulletin-120, or B-120, forecast points by simulated flows from twenty-six corresponding Hydrologic Unit Code 8 sub-basins1 (Table 1), as mapped by the U.S. Geological Survey's Watershed Boundary Dataset. The B-120 forecast use a Multiple Linear Regression forecast, but we choose to use a Principal Component Regression (PCR) method that mimics the Natural Resource Conservation Service's own PCR method (Garen, 1992) because it is more repeatable and comparable to other studies (Rosenberg et al., 2011). We develop simulated February 1st (2-month lead time) and April 1st (0-month lead time) forecasts using an adapted version of the NRCS PCR model trained with outputs from the Variable Infiltration Capacity (VIC) model, a large-scale, semi-distributed gridded hydrologic model (Liang et al., 1994) similar to methods by Rosenberg et al. (2011). Forecasts are evaluated using a Nash Sutcliffe Efficiency (NSE) skill score, where values of 1 represent a perfect forecast (Nash & Sutcliffe, 1970).

Simulated hydrologic data from the Localized Constructed Analogue (LOCA)-forced (Pierce et al., 2014) VIC model were downloaded from the Cal-Adapt data repository (https://cal-adapt.org/data/download/) for a subset of 10 GCMs that were previously identified as well-suited for important aspects of California's climate (Pierce

---

[1] As jurisdiction of the City of Los Angeles's Department of Water and Power, the Owens River and Mono Lake basins in the east-draining part of the southern Sierra are not included in our study basins.

et al., 2018). The Cal-Adapt data repository covers California and Nevada at 1/16 °spatial resolution (~6 km). Model simulations are available for historical, RCP 4.5, and RCP 8.5 climate change scenarios for years 1950-2099.

To satisfy objective inter-basin comparisons while balancing representation of B-120 forecast equations, we modify CA DWR forecast predictor variables to include beginning-of-month SWE and beginning-of-month accumulated P for the water year. Antecedent runoff was not considered since it is not included by CA DWR as a predictor variable in many settings. We simulate the observations of predictor variables on the forecast issue date by only extracting VIC data from grid cells with active, corresponding (i.e. SWE or P) observation stations recorded in the California Data Exchange Center (http://cdec.water.ca.gov/cdecstation2/). PCA-transformed February 1st or April 1st SWE and P values were then regressed against VIC-simulated April 1 – July 31 total runoff. No streamflow routing was performed. Five separate PCR forecasts were developed for each of the twenty-six study basins under each GCM to produce ensemble mean predictions: late 20th century (water years 1951-2005) under an historical emission scenario and the early 21st century (WY2006-2052) and late 21st century (WY2053-2099) with climate forcings for low and high emissions scenarios: relative concentration pathways (RCP) 4.5 and 8.5, respectively.

We also hypothesized and evaluated two enhanced forecasts that included synthetic supplemental "observations". The first enhanced forecast system used VIC-simulated soil moisture data to supplement the standard predictor variables. Similar to accumulated P, we extracted simulated top-layer soil moisture values on the forecast-issue dates from grid cells at the CDEC-reported precipitation observation stations,

representing an expansion of California's soil moisture monitoring network (White et al., 2013). The other enhanced forecast system substituted beginning of month SWE at the CDEC stations with basin-wide SWE measurements. This method was modeled after the Airborne Snow Observatory (ASO) objective to remotely estimate SWE through a combination of high resolution lidar snow depth measurements and snow density modeling (Painter et al., 2016). The virtual ASO method is equivalent to having CDEC SWE observation stations in every VIC grid cell for the forecast-issue dates.

ID	USGS ID	River/Basin Name	Elevation (m)	Peak SWE (mm)	WY P (mm)	WY Baseflow (mm)	WY Runoff (mm)	AMJJ Runoff (mm)	Baseflow Index	Runoff Ratio	AMJJ:WY Runoff	NSE Hist. Late 20th C.	NSE RCP85 Early 21st C.	NSE RCP85 Late 21st C.
1	16050101	Lake Tahoe	2138	371	861	221	351	268	0.63	0.41	0.76	0.85	0.77	0.62
2	16050102	Truckee	1800	226	618	125	205	152	0.61	0.33	0.74	0.87	0.76	0.69
3	16050201	Upper Carson	1960	252	684	136	243	182	0.56	0.36	0.75	0.86	0.8	0.69
4	16050301	East Walker	2163	120	384	32	61	40	0.52	0.16	0.65	0.77	0.76	0.78
5	16050302	West Walker	2042	162	485	62	112	77	0.55	0.23	0.69	0.81	0.79	0.8
6	18010208	Scott	1309	239	846	143	290	167	0.49	0.34	0.58	0.81	0.77	0.59
7	18010211	Trinity	1191	425	1563	535	911	429	0.59	0.58	0.47	0.55	0.45	0.36
8	18020003	Lower Pit	1315	188	873	186	321	145	0.58	0.37	0.45	0.66	0.48	0.34
9	18020004	Mccloud	1310	359	1530	493	828	327	0.6	0.54	0.4	0.61	0.54	0.34
10	18020005	Sacramento	1033	276	1742	568	1033	317	0.55	0.59	0.31	0.39	0.32	0.21
11	18020121	North Fork Feather	1517	436	1494	546	910	442	0.6	0.61	0.49	0.63	0.43	0.34
12	18020122	East Branch N. Fork Feather	1601	251	893	235	348	170	0.68	0.39	0.49	0.75	0.59	0.34
13	18020123	Middle Fork Feather	1556	320	1280	429	737	316	0.58	0.58	0.43	0.66	0.48	0.28
14	18020125	Upper Yuba	1276	348	1646	610	1027	401	0.59	0.62	0.39	0.55	0.43	0.33
15	18020128	North Fork American	1386	341	1493	555	900	384	0.62	0.6	0.43	0.56	0.39	0.26
16	18020129	South Fork American	1253	223	1178	371	580	245	0.64	0.49	0.42	0.59	0.48	0.34
17	18030001	Upper Kern	2340	414	812	187	347	251	0.54	0.43	0.72	0.84	0.84	0.84
18	18030006	Upper Tule	578	77	520	92	164	74	0.56	0.32	0.45	0.58	0.51	0.45
19	18030007	Upper Kaweah	739	126	533	89	167	97	0.54	0.31	0.58	0.77	0.77	0.76
20	18030010	Upper King	2264	521	965	239	461	363	0.52	0.48	0.79	0.85	0.86	0.88
21	18040006	Upper San Joaquin	2191	483	1027	284	503	365	0.57	0.49	0.73	0.83	0.83	0.81
22	18040008	Upper Merced	1438	373	982	256	468	308	0.55	0.48	0.66	0.81	0.77	0.68
23	18040009	Upper Tuolumne	1471	331	1016	261	464	274	0.56	0.46	0.59	0.8	0.78	0.71
24	18040010	Upper Stanislaus	1235	302	1069	316	532	292	0.59	0.5	0.55	0.74	0.63	0.47
25	18040012	Upper Mokelumne	764	156	919	250	400	169	0.62	0.44	0.42	0.6	0.44	0.36
26	18040013	Upper Consumnes	540	40	866	201	311	62	0.65	0.36	0.2	0.21	0.16	0.15

Table 3.1: Study basin IDs, USGS gauge IDs, mean basin elevation, mean historical (late 20th century) hydrometeorological characteristics, and select forecast skill scores for historical and RCP 8.5 climate change scenarios for two time periods, early 21st century (WYs 2006 to 2052) and the late 21st century (WYs 2053 to 2099).

3.3 Results

Simulations of historical (WY1951-2005) forecast performance are closely

validated by an analysis of official CA DWR B-120 forecasts for thirteen western Sierra

basins (Harrison & Bales, 2016) from ~1930 to 2012 (Figure A3.1). These model

comparisons provide confidence that our simulations reasonably represent CA DWR B-

120 forecasts; our simulated February 1st forecasts are, however, somewhat more skillful

than the reported results from Harrison & Bales (2016). Projections of seasonal water

supply forecast skill into the 21st century (WY 2006-2099) show substantial but uneven

loss of skill relative to late 20th century historic (1950-2005) baseline conditions (Figures

1, A3.2). By the last half of the 21st century (2053-2099), median April 1st forecast skill

is projected to decline by 12% with lower emissions (delta NSE = -0.1, RCP 4.5),

upwards of 22% with higher emissions (dNSE = -0.16; RCP 8.5). However, these losses

in forecast skill are uneven, ranging from a worse-case loss of 49% (dNSE -0.39, #14,

East Branch North Fork of the Feather, RCP 8.5) to marginal increases of 3% (dNSE

+0.03, #4, East Walker, RCP 4.5). On average, February 1st forecasts are projected to see

even greater relative skill degradation by the last half of the 21st century (median skill

declines of 30%, dNSE = -0.16, RCP 8.5). Smaller declines in skill are projected for the

early 21st century (2006-2052), ranging from median April 1st losses of 6% (dNSE = -

0.05, RCP 4.5) to 8% (dNSE = -0.07, RCP 8.5). We estimate that from the late 20th

century to the late 21st century, a loss of 0.1 NSE under high emission scenarios

increases forecast errors by a Mean Absolute Percent Error (MAPE) of 2.4 (Figure A3.4).

The average RCP 8.5 late 21st century April to July streamflow of 315 thousand acre feet

(KAF); thus, an incremental loss of 0.1 NSE approximates an increase in forecast error of 7.5 KAF, or approximately enough water for 80,000 people for a year2.

Mid- to lower-elevations (1000-1700 m mean elevation) show particular forecast vulnerability by late 21st century (Figure A3.4). Below 1,800 m, median skill loss ranges from 19% (dNSE = -0.13, RCP 4.5) to 33% (dNSE = -0.23, RCP 8.5), while in basins with mean elevations above 2,000 m we simulate virtually no loss of skill. This elevation-mediation of skill degradation is well illustrated by the difference between skill losses in the relatively lower-elevation northern Sierra Nevada and high-elevation southern Sierra Nevada (Figure 3.1). An exception in the well-performing high-elevation band (>2,000m) is the Lake Tahoe basin (#1, mean elevation 2,128m), which has late century skill losses of 8% (dNSE -0.11, RCP 4.5) to 17% (dNSE = -0.23, RCP 8.5). With a high elevation basin outlet (1900m), Tahoe has a basin mean elevation similar to southern Sierra Nevada basins, yet peak elevations that are more than 1000m lower. This suggests that high elevation (>3000m) portions of a catchment are important in buffering loss of snowpack retention. In contrast, differences in forecast skill between east (e.g. #3, Carson) and west-draining (e.g. #25, Upper Mokelumne) basins which share a ridgeline (i.e. the Sierra Nevada crest) may illustrate the vulnerability of lower portions of a basins in future forecast skill, though differences in precipitation on the windward and leeward sides of the mountain range greatly complicate such an west-east comparison.

---

2 Based on an assumed per capita residential usage of 85 gallons per day (California Legislative Analyst's Office, 2016)
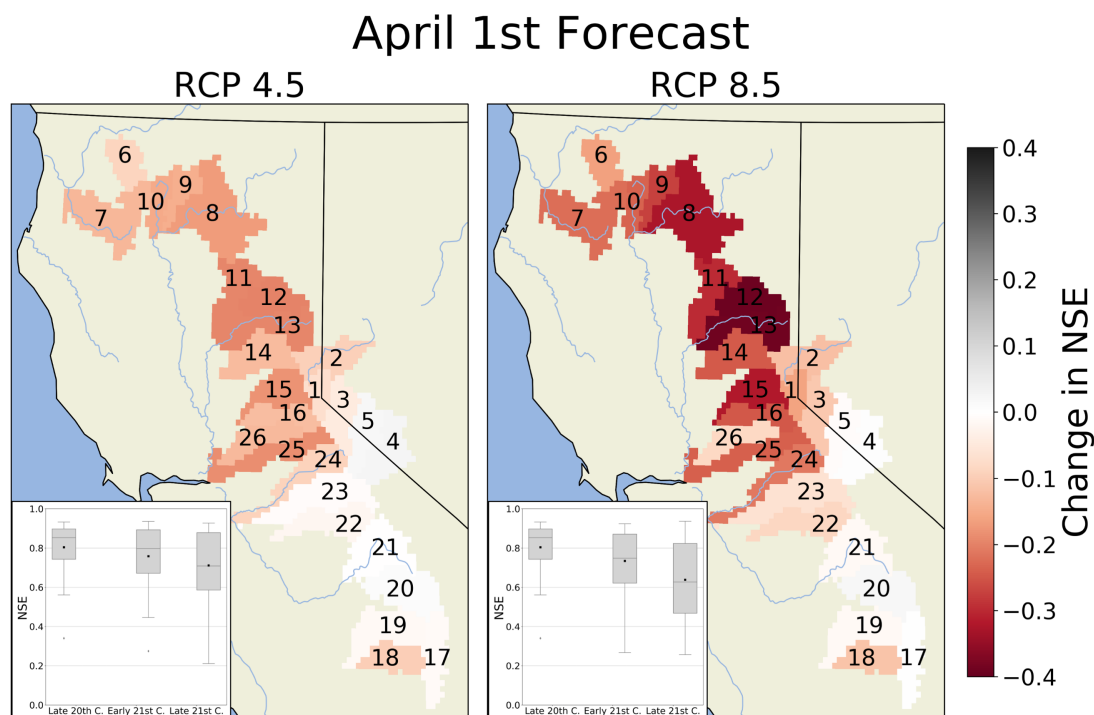
April 1st Forecast

Figure 3.1: April 1st seasonal water supply forecast skill for 26 headwater basins in the Sierra Nevada and Northern California from the late 20th Century (WY1951 to 2005) to the late 21st Century (WY2053 to 2099) for RCP 4.5 (low emissions, left panel) and RCP 8.5 (high emissions, right panel) climate change scenarios from an ensemble mean of 10 global circulation models. Basin ID numbers correspond with Table 3.1. Accompanying February 1st forecast skill maps shown in Figure A3.2.

Loss of forecast skill is positively and well-correlated with the loss of February 1st and April 1st SWE from the historical period to early and late 21st century periods for both forecast lead times (Figure 3.2). This relationship is particularly well-defined for February 1st forecasts (RCP 4.5: r2 = 0.6; RCP 8.5: r2 = 0.51; p-values < 0.001); April 1st change in SWE explains less variance in April 1st forecast skill (RCP 4.5: r2 = 0.28; RCP 8.5: r2 = 0.33; p-values < 0.001). Decreases in the baseflow index (BFI), or the ratio of simulated baseflow to total simulated runoff, appear to be associated with larger losses

in forecast skill, particularly for April 1st forecasts (Figure 3.2, bottom panels). For example, from the historic to the late 21st century under RCP 8.5, basins in the top 50th percentile of historical (late 20th C.) BFI have larger projected changes in the BFI (median loss 8.4%, ΔBFI = -0.06) and a greater loss of April 1st forecast skill (median loss 33.6%, dNSE = -0.25), while basins below the 50th percentile had smaller changes in BFI (median loss 3.8%, ΔBFI = -0.02) and smaller losses of forecast skill (median loss 7.3%, dNSE = -0.06). While declining SWE appears be a significant (primary) driver of water supply forecast skill loss into the future, basins with historically larger baseflow (groundwater) contributions, which are more sensitive to decreases in baseflow index (groundwater), accentuate loss of forecast skill.
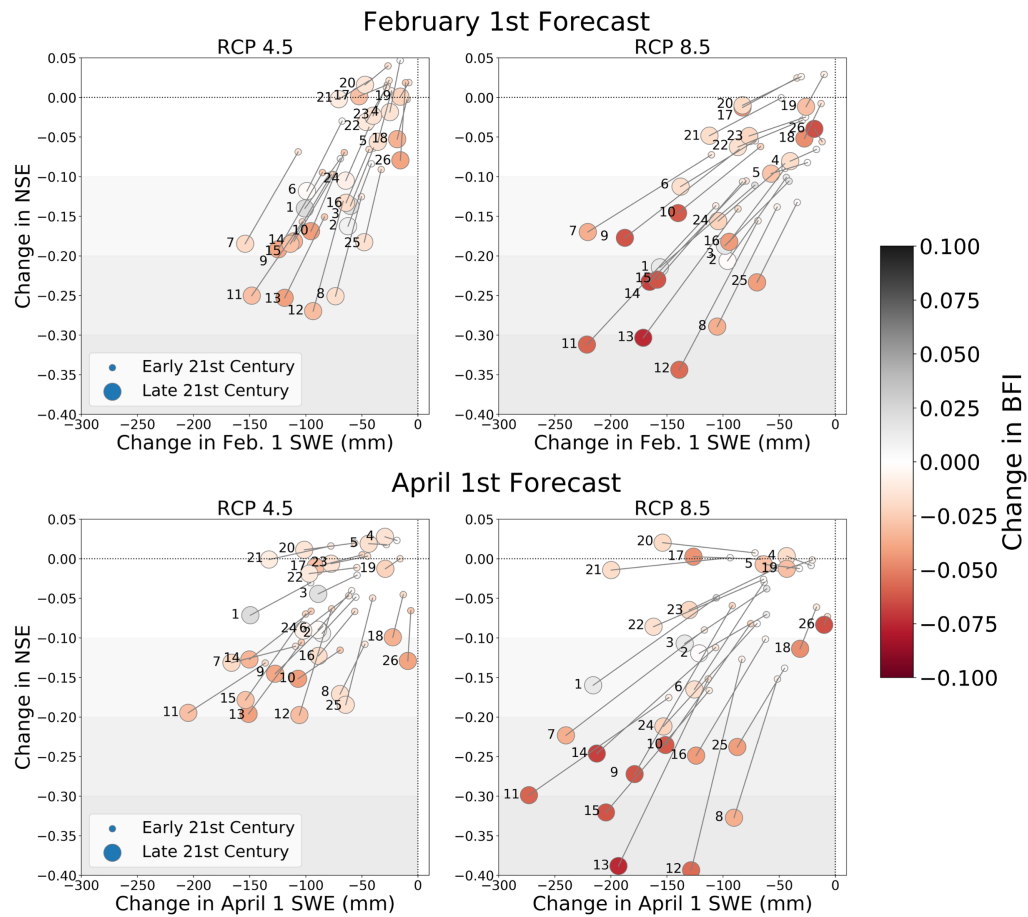
Figure 3.2: Change in forecast skill (Nash Sutcliffe Efficiency, NSE) as a function of change in snow water equivalent from the late 20th century historical baseline to the early 21st century (smaller dots) to the late 21st century (larger dots) for RCP 4.5 (low emissions, left) and RCP 8.5 (high emissions, right) scenarios. Colored by change in the baseflow index, or the ratio of baseflow to total runoff, with respect to the historic baseline.

Introducing VIC-simulated basin-wide SWE (i.e. our simulated version of Airborne Space Observatory SWE estimates) and soil moisture "observations" into the regression models to create 'enhanced' forecasts designed to buffer loss of skill yield modest overall improvements for April 1st forecasts, with better results in isolated instances (Figure 3.3) February 1st forecasts are more challenging to improve upon (Figure A3.5). Similarly, historically very accurate (> ~0.8 April 1st NSE) standard

forecasts often have marginal improvement or actual loss of skill from enhanced forecasting techniques, yet these basins are also simulated to be less susceptible to future forecast skill degradation in most cases. Instead, the greatest improvements in forecast skill are generally simulated in basins with low historical NSE scores and larger loss of future skill, most notably for April 1st forecasts.

ASO enhanced forecasts show some complementary patterns in forecast skill improvement across basins, emissions scenarios, and time (Figures 3, A6). In the early 21st century, basin-wide SWE measurements from the simulated ASO flight could, on average, replace more than 40% of lost forecast skill for basins with dNSE < -0.05, with maximum replacement rates of over 100% (e.g. #18 Upper Tule and #26 Upper Cosumnes). For unclear reasons, some basins, particularly in the east-draining Sierra and in Northern California, have less skillful ASO enhanced forecasts, suggesting that forecast improvements techniques will require a basin-scale approach. However, the simulated widespread skillfulness of ASO-enhanced forecasts does point towards significant operational utility at least through mid-century. With the near eradication of April 1 SWE by the late 21st century under RCP 8.5 projections, the ability of ASO to buffer skill loss for those same mid-elevation basins is diminished (Figure A3.6, A3.7), with isolated success only in the Upper Stanislaus (#24), Upper Tuolumne (#23), Upper Merced (#22), and Upper Tule (#18) basins.

As snowpack declines, our simulations suggest that soil moisture may replace some of the forecast information SWE historically provided (Figure 3.3, A3.8). However, unlike ASO, soil moisture rarely replaces more than 25% of lost forecast skill, with the exception of the Upper Cosumnes (#26, basin mean elevation 540m) and under a late-

century RCP 4.5 scenario. Specifically, our findings suggest that as SWE declines further into the late 21st century, soil moisture observations become increasingly beneficial for April 1st forecasts in most basins below 2,000m, but especially for basins below 1,600m under RCP 8.5. Similar soil moisture forecast improvements are not realized for February 1st forecasts (Figures A3.6, A3.8).
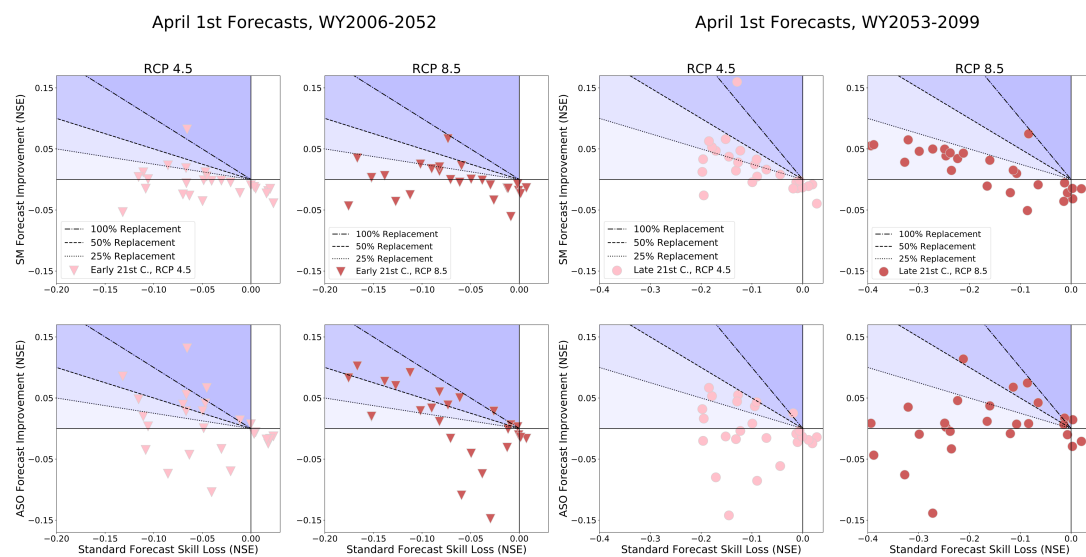
Figure 3.3: Forecast improvement (NSE) from soil moisture (SM, top row) and Airborne Snow Observatory (ASO, bottom row) enhanced forecasts as a function of the loss in forecast skill (NSE) from the late 20th century baseline to the early 21st century (triangles; left panels) and to the late 21st century (circles; right panels) for April 1st forecasts under RCP 4.5 (pink) and RCP 8.5 (red) scenarios. The 25%, 50%, and 100% lines shows the percent of lost forecast skill that is replaced by the enhanced forecast system. Blue regions highlight forecast improvements from the enhanced forecast systems, with greater improvements in darker blue. See Figure A3.6 for accompanying February 1st forecasts figures.

3.4 Discussion

Understanding the mechanisms causing loss of water supply forecast skill is critical. In an operational context, water stored in the mountains as winter snowpack

provides a strong determinant of spring/summer streamflow volumes, along with future

precipitation (Church, 1935). However, as snowfall changes to rainfall (Knowles et al.,

2006) and mid-winter ablation events become more frequent (A. A. Harpold & Brooks,

2018), these relationships decline for a variety of reasons, e.g. making drought less

predictable (Livneh & Badger, 2020). As a first order control, basins with little change in

snowpack have little change in forecast skill unless future post-forecast precipitation

regimes change (Sturtevant et al., 2020). In basins where snow change is substantial,

regression-based forecasts have an implicit assumption that precipitation runoff

efficiency (i.e. streamflow to precipitation) is constant. Changes in runoff efficiency (or

any part of a flow regime), especially outside of the historically observed record, a

statistical method will struggle to capture that non-stationary response. Some of that loss

of skill from snowpack storage loss might be reintroduced using observations of other

storage terms, such as soil moisture (A. A. Harpold et al., 2017; Rosenberg et al., 2013)

or groundwater. Catchment properties, such as baseflow index, appear to be secondary

controls that ameliorate the skill loss due to snowpack loss (Barnhart et al., 2016), at least

in a VIC-model reconstruction of reality. Basins with lower historical baseflow index,

that have slower streamflow responses consistent with significant groundwater inputs,

were more resistant to degradation of forecast skill (Figure 3.2). This is consistent with

the empirical results of Safeeq et al. (2013) who showed that catchments with less storage

are more resistant to changes in late summer flows. However, these inferences rely

strongly on the VIC simulations, and process and parameter assumptions therein (Bennett

et al., 2017), as well as model resolution in space and time. Despite these limitations, our

novel experimental framework allows us to explore future skill degradation and

determine the value of supplemental observations in ways that benefit long-term water supply planning.

Buffering the loss of skill from standard SWE and P-only forecasts was most successful in basins most vulnerable to future skill loss ($>$dNSE = -0.1). During the early 21st century, mid-elevation basins (1000-1700m) benefitted most from ASO-type supplemental information, with widespread improvements in April 1st forecasts, and to a lesser extent, February 1st forecasts (Figure 3.3). This was not as evident during the late 21st century, when soil moisture observations appear to replace the late-season (April 1st) forecast information once offered by SWE. These findings result in basin-specific management recommendations dependent on the degree to which snowpack has declined, and thus, to the vulnerability of the basin to loss of skill (Figure 3.2). The utility of soil moisture information in seasonal water supply forecasting, beyond precipitation and SWE, is consistent with empirical (Harpold et al., 2017) and modeled (Koster et al., 2010; Rosenberg et al., 2013) research. With an intrinsic memory spanning weeks to months (Entin et al., 2000), soil moisture provides terrestrial water storage information outside of snow-covered regions, especially in mid-elevations (Rosenberg et al., 2013).

Consider two examples of basin specific management decisions that arise from our enhance forecast results. The Upper Stanislaus (#24), which flows into California's fourth largest reservoir, New Melones Lake (2.4 million acre-feet of storage), and is a key part of the Central Valley Project (Congressional Research Service, 2020). With projected late century RCP 8.5 skill loss of nearly 25% but higher upper elevations which will retain snowpack, the Upper Stanislaus is a good candidate for ASO mitigation techniques. Our simulations show that remotely measured SWE observations for this time

period would improve skill by 0.11 NSE, buffering half of the RCP8.5 skill loss expected otherwise. In regions with vulnerable snowpacks, soil moisture measurements offer promise for representing soil water storage in once snow-covered areas. The Mokelumne (#25), the primary source of water for the East Bay Municipal Utility District in the San Francisco Bay area, is a good example. Under RCP 8.5, we project late-century skill loss in the Upper Mokelumne of over 30% (dNSE = -0.24), with most (dNSE = -0.14) of the skill loss occurring during the early 21st century. In this case, a targeted installation of a soil moisture sensing network in the Upper Mokelumne could prove helpful; including soil moisture measurements in the regression model buffers loss of skill in the late century, RCP 8.5 Mokelumne by 18%, which may be beneficial for managing small headwater reservoirs, such as Pardee. The examples bound the range of possible management scenarios found in this study (Table 1 and Figure 3.3) that suggest basin-specific management efforts will be critical to buffering water supply forecast skill loss in the 21st century.

3.5 Conclusion

Our research shows that California's statistical water supply forecasts face a considerable threat to skill degradation from a loss of snowpack. It is therefore plausible that seasonal water supply forecast skill has already begun to degrade relative to the 20th century, though this may not yet be discernable within normal interannual climate variability. Buffering of water supply forecast skill degradation should be an important goal for water management agencies tasked with reservoir operations, such as CA DWR. Our model based study showed which basins would be sensitive to forecast skill loss and when, but more work is needed to verify these results and translate them into the

operational models of CA DWR. Further, it is only by assuming the VIC model as truth

that we are able to gain these insights into the nature of future water supply forecast skill.

Despite being well-validated by empirical research (Figure A3.1), the ensembles of

simulated results presented above may or may not come to represent future hydrologic

conditions in the Sierra Nevada.

Numerous paths exist to improve water supply forecast modeling, including

bolstering existing forecast techniques with improved observations. Our research

suggests that remotely-sensed SWE data can play an important role for increasing

forecast skill as snow in Sierra Nevada declines. This result supports building a historical

library of lidar-derived SWE maps that can be used to help train empirical models.

Targeted installation of high-altitude soil moisture sensors (or groundwater wells)

provide another source of data that will prove useful as snowpack decline markedly in the

late 21st century. Soil moisture networks present their own challenges for developing

representative measurements and ensuring data quality (A. A. Harpold, 2016; A. A.

Harpold & Molotch, 2015) Implementing a physically-based, long-range hydrologic

forecast model has been an ultimate goal of the scientific researchers (Vaze et al., 2010),

it has proven challenging to convince operational forecasters that improvements in skill

outweigh costs of model complexity and implementation of new systems into operations

(Clark et al., 2015). The current state of operational and research-grade hydrologic

models suggests that the continued investment in the existing water supply forecasting

infrastructure is a good alternative to adoption of new models at the present time. The

State of California is particularly well positioned to retool their water supply forecasting

program to adapt to declining snowpack through an expanded observational network of

soil moisture and the remote sensing of snow, which could make it a model for other

mountain-based water supply forecasting systems throughout the western U.S. and

globally.

References

Barnhart, T. B., Molotch, N. P., Livneh, B., Harpold, A. A., Knowles, J. F., & Schneider, D. (2016). Snowmelt rate dictates streamflow. Geophysical Research Letters, 43(15), 8006–8016. https://doi.org/10.1002/2016GL069690

Bennett, A., Nijssen, B., Chegwidden, O., Wood, A., Clark, M. P., Bennett, A., et al. (2017). What Makes Hydrologic Models Differ? Using SUMMA to Systematically Explore Model Uncertainty and Error. AGUFM, 2017, H22D-03.

Berghuijs, W. R., Woods, R. A., & Hrachowitz, M. (2014). A precipitation shift from snow towards rain leads to a decrease in streamflow. Nature Climate Change, 4(7), 583–586. https://doi.org/10.1038/nclimate2246

Beven, K. (2002). Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. Hydrological Processes, 16(2), 189–206. https://doi.org/10.1002/hyp.343

Bureau of Reclamation. (2016). Managing Water in the West SECURE Water Act Section 9503(c)-Reclamation Climate Change and Water.

California Department of Food and Agriculture. (2018). California Agricultural Statistics Review State of California. California Agricultural Statistics Review. Retrieved from www.cdfa.ca.gov/statisticsACKNOWLEDGEMENTS

California Legislative Analyst's Office. (2016). Residential Water Use Trends and Implications for Conservation Policy. Retrieved May 5, 2020, from https://lao.ca.gov/Publications/Report/3611

Church, J. E. (1935). Principles of Snow Surveying as Applied to Forecasting Stream Flow. Journal of Agricultural Research, 51(2), 97–130. Retrieved from https://books.google.com/books?hl=en&lr=&id=-EFSAQAAMAAJ&oi=fnd&pg=PA97&dq=james+church+1935&ots=26d2VJVWAT&sig=uV6mZVW8JbxYb_PKMza4p0TAKFg#v=onepage&q=james church 1935&f=false

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., et al. (2015). Improving the representation of hydrologic processes in Earth System Models. Water Resources Research, 51(8), 5929–5956. https://doi.org/10.1002/2015WR017096@10.1002/(ISSN)1944-7973.WRR50

Congressional Research Service. (2020). Central Valley Project: Issues and Legislation. Retrieved from https://crsreports.congress.gov

Cooper, M., Nolin, A., & Safeeq, M. (2016). Testing the recent snow drought as an analog for climate warming sensitivity of Cascades snowpacks - IOPscience. Environmental Research Letters, 11. Retrieved from https://iopscience.iop.org/article/10.1088/1748-9326/11/8/084009

Dettinger, M. D., Cayan, D. R., Meyer, M. K., & Jeton, A. E. (2004). SIMULATED HYDROLOGIC RESPONSES TO CLIMATE VARIATIONS AND CHANGE IN THE MERCED, CARSON, AND AMERICAN RIVER BASINS, SIERRA NEVADA, CALIFORNIA, 1900-2099. Retrieved from https://link.springer.com/content/pdf/10.1023/B:CLIM.0000013683.13346.4f

Earman, S., & Dettinger, M. (2011). Potential impacts of climate change on groundwater resources - A global review. Journal of Water and Climate Change, 2(4), 213–229. https://doi.org/10.2166/wcc.2011.034

Entin, J. K., Robock, A., Vinnikov, K. Y., Hollinger, S. E., Liu, S., & Namkhai, A. (2000). Temporal and spatial scales of observed soil moisture variations in the extratropics. Journal of Geophysical Research: Atmospheres, 105(D9), 11865–11877. https://doi.org/10.1029/2000JD900051

Fyfe, J. C., Derksen, C., Mudryk, L., Flato, G. M., Santer, B. D., Swart, N. C., et al. (2017). Large near-Term projected snowpack loss over the western United States. Nature Communications, 8(1), 1–7. https://doi.org/10.1038/ncomms14996

Garen, D. C. (1992). Improved Techniques in Regression-Based Streamflow Volume Forecasting. Journal of Water Resource Planning Management, 118(6), 654–670.

Hammond, J. C., Harpold, A. A., Weiss, S., Kampf, S. K., & Hammond, J. C. (2019). Partitioning snowmelt and rainfall in the critical zone: effects of climate type and soil properties 2 3. Hydrology and Earth System Sciences. https://doi.org/10.5194/hess-2019-98

Harpold, A., Brooks, P., Rajagopal, S., Heidbuchel, I., Jardine, A., & Stielstra, C. (2012). Changes in snowpack accumulation and ablation in the intermountain west, 48, 11501. https://doi.org/10.1029/2012WR011949

Harpold, A., Dettinger, M., & Rajagopal, S. (2017). Defining Snow Drought and Why It Matters. Eos, 98. https://doi.org/10.1029/2017EO068775

Harpold, A. A. (2016). Diverging sensitivity of soil water stress to changing snowmelt timing in the Western U.S. Advances in Water Resources, 92, 116–129. https://doi.org/10.1016/j.advwatres.2016.03.017

Harpold, A. A., & Brooks, P. D. (2018). Humidity determines snowpack ablation under a warming climate. PNAS, 115(6), 1215–1220. https://doi.org/10.1073/pnas.1716789115

Harpold, A. A., & Molotch, N. P. (2015). Sensitivity of soil water availability to changing snowmelt timing in the western U.S. Geophysical Research Letters, 42(19), 8011–8020. https://doi.org/10.1002/2015GL065855

Harpold, A. A., Sutcliffe, K., Clayton, J., Goodbody, A., & Vazquez, S. (2017). Does Including Soil Moisture Observations Improve Operational Streamflow Forecasts in Snow-Dominated Watersheds? Journal of the American Water Resources Association, 53(1), 179–196. https://doi.org/10.1111/1752-1688.12490

Harrison, B., & Bales, R. (2016). Skill Assessment of Water Supply Forecasts for Western Sierra Nevada Watersheds. Journal of Hydrologic Engineering, 21(4), 04016002. https://doi.org/10.1061/(asce)he.1943-5584.0001327

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, 42(3). https://doi.org/10.1029/2005WR004362

Knowles, N., Dettinger, M. D., & Cayan, D. R. (2006). Trends in Snowfall versus Rainfall in the Western United States. Journal of Climate, 19, 4545–4559. Retrieved from https://journals.ametsoc.org/doi/pdf/10.1175/JCLI3850.1

Koster, R. D., Mahanama, S. P. P., Livneh, B., Lettenmaier, D. P., & Reichle, R. H. (2010). Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow. Nature Geoscience, 3(9), 613–616. https://doi.org/10.1038/ngeo944

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. Journal of Geophysical Research, 99(D7), 14415. https://doi.org/10.1029/94JD00483

Livneh, B., & Badger, A. M. (2020). Drought less predictable under declining future snowpack. Nature Climate Change, 1–7. https://doi.org/10.1038/s41558-020-0754-8

Mcnamara, J. P., Chandler, D., Seyfried, M., & Achet, S. (2005). Soil moisture states, lateral flow, and streamflow generation in a semi-arid, snowmelt-driven catchment. Process, 19, 4023–4038. https://doi.org/10.1002/hyp.5869

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

Musselman, K. N., Clark, M. P., Liu, C., Ikeda, K., & Rasmussen, R. (2017). Slower snowmelt in a warmer

world. Nature Climate Change, 7(3), 214–219. https://doi.org/10.1038/nclimate3225

Nash, J. E., & Sutcliffe, J. V. (1970). RIVER FLOW FORECASTING THROUGH CONCEPTUAL MODELS PART I-A DISCUSSION OF PRINCIPLES*. Journal of Hydrology (Vol. 10). © North-Holland Publishing Co. Retrieved from https://hydrology.agu.org/wp-content/uploads/sites/19/2016/04/NashSutcillfe1.pdf

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., et al. (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

Pierce, D. W., Cayan, D. R., & Thrasher, B. L. (2014). Statistical Downscaling Using Localized Constructed Analogs (LOCA). Journal of Hydrometeorology, 15, 2558–2585. https://doi.org/10.1175/JHM-D-14

Pierce, D. W., Kalansky, J. F., & Cayan, D. R. (2018). CLIMATE, DROUGHT, AND SEA LEVEL RISE SCENARIOS FOR CALIFORNIA'S FOURTH CLIMATE CHANGE ASSESSMENT A Report for: California's Fourth Climate Change Assessment. Retrieved from www.climateassessment.ca.gov.

Rango, A. (1980). OPERATIONAL APPLICATIONS OF SATELLITE SNOW COVER OBSERVATIONS. Journal of the American Water Resources Association, 16(6), 1066–1073. https://doi.org/10.1111/j.1752-1688.1980.tb02549.x

Rosenberg, E. A., Wood, A. W., & Steinemann, A. C. (2011). Statistical applications of physically based hydrologic models to seasonal streamflow forecasts. Water Resources Research, 47(3). https://doi.org/10.1029/2010WR010101@10.1002/(ISSN)1944-7973.SBRS1

Rosenberg, E. A., Wood, A. W., & Steinemann, A. C. (2013). Informing Hydrometric Network Design for Statistical Seasonal Streamflow Forecasts. Journal of Hydrometeorology, 14(5), 1587–1604. https://doi.org/10.1175/JHM-D-12-0136.1

Safeeq, M., Grant, G. E., Lewis, S. L., & Tague, C. L. (2013). Coupling snowpack and groundwater dynamics to interpret historical streamflow trends in the western United States, 27, 655–668. https://doi.org/10.1002/hyp.9628

Stewart, I. T., Cayan, D. R., Dettinger, M. D., Stewart, I. T., Cayan, D. R., & Dettinger, M. D. (2005). Changes toward Earlier Streamflow Timing across Western North America. Journal of Climate, 18(8), 1136–1155. https://doi.org/10.1175/JCLI3321.1

Sturtevant, J., Dettinger, M. D., McAfee, S., Rajagopal, S., Wood, A. W., Newman, A. J., & Harpold, A. A. (2020). Benchmarking Empirical and Physical Models for Water Supply Forecasting: Implications for Declining Western U.S. Snowpack.

Vaze, J., Post, D. A., Chiew, F. H. S., Perraud, J. M., Viney, N. R., & Teng, J. (2010). Climate non-stationarity - Validity of calibrated rainfall-runoff models for use in climate change studies. Journal of Hydrology, 394(3–4), 447–457. https://doi.org/10.1016/j.jhydrol.2010.09.018

White, A. B., Anderson, M. L., Dettinger, M. D., Ralph, F. M., Hinojosa, A., Cayan, D. R., et al. (2013). A Twenty-First-Century California Observing Network for Monitoring Extreme Weather Events. Journal of Atmospheric and Oceanic Technology, 30(8), 1585–1603. https://doi.org/10.1175/JTECH-D-12-00217.1

Wood, A. W., & Lettenmaier, D. P. (2006). A TEST BED FOR NEW SEASONAL HYDROLOGIC FORECASTING APPROACHES IN THE WESTERN UNITED STATES. American Meteorlogical Society, 1699–1712. https://doi.org/10.1175/BAMS-87-I2-I699
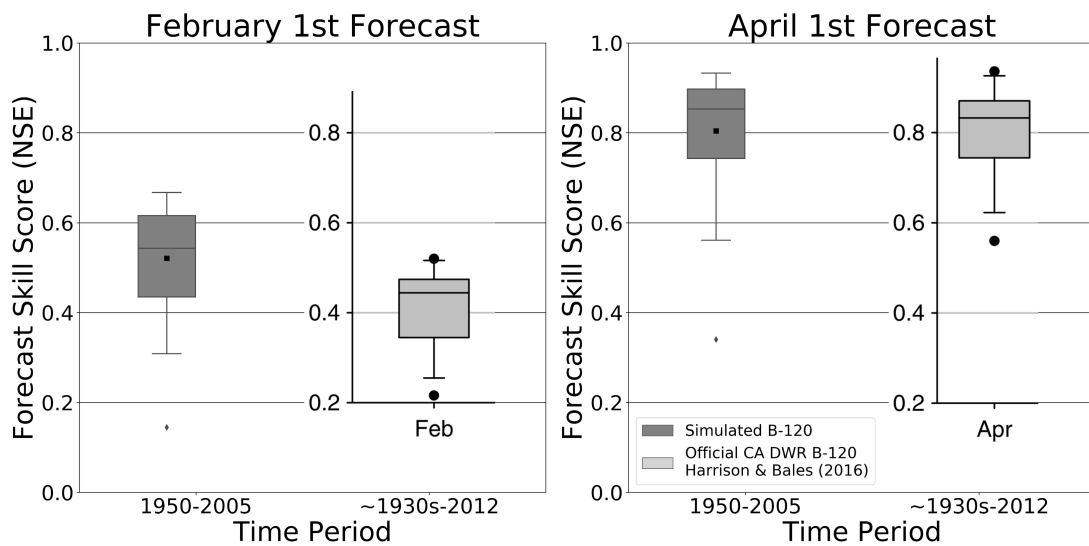
Appendix 3.1: Supporting Figures

Figure A3.1: Validation of simulated historical February 1st and April 1st validation forecasts using VIC data from 1950-2006 for twenty-six HUC8 basins (left box, dark grey) representing twenty CA DWR B-120 forecast points against an elevation of thirteen official CA DWR B-120 forecasts for the western Sierra from ~1930s to 2012 (right box, light grey) from Harrison & Bales (2016) using a Nash Sutcliffe Efficiency (NSE). Boxplots from Harrison & Bales (2016) are overlays of portions of the original figure (Figure 10) from their manuscript.
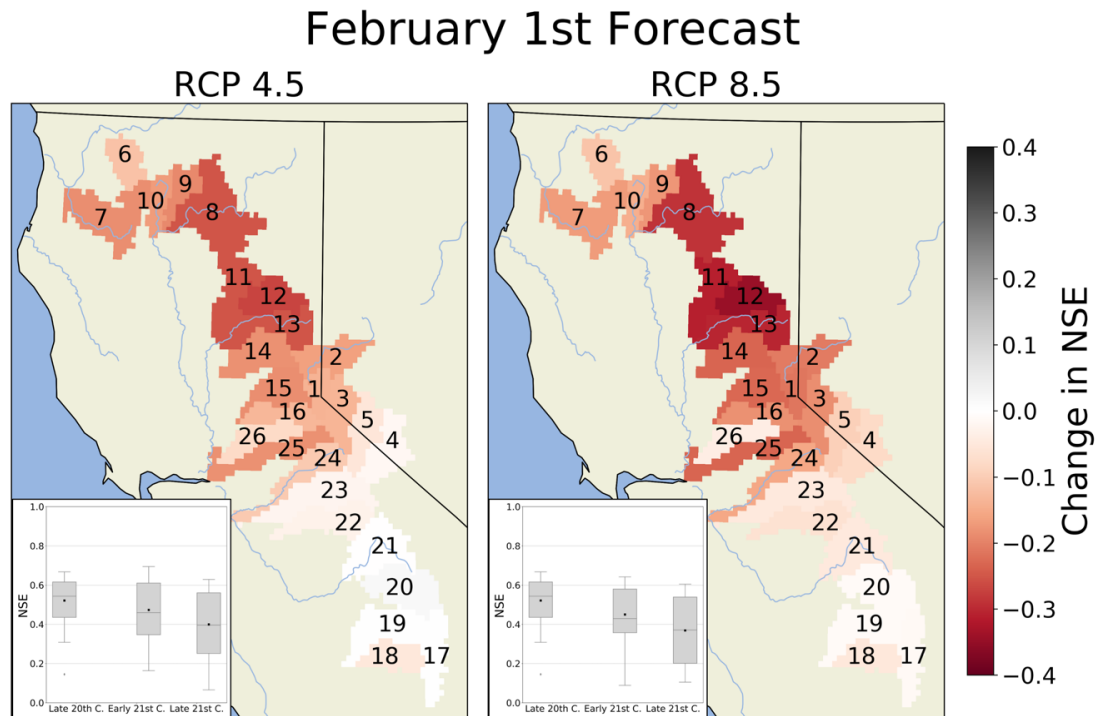
Figure A3.2: February 1st seasonal water supply forecast skill for 26 headwater basins in the Sierra Nevada and Northern California from the late 20th Century (WYs 1951 to 2005) to the late 21st Century (WYs 2053 to 2099) for RCP 4.5 and RCP 8.5 climate change scenarios using an ensemble mean of 10 Global Circulation Models. Basin ID numbers correspond with Table 3.1.
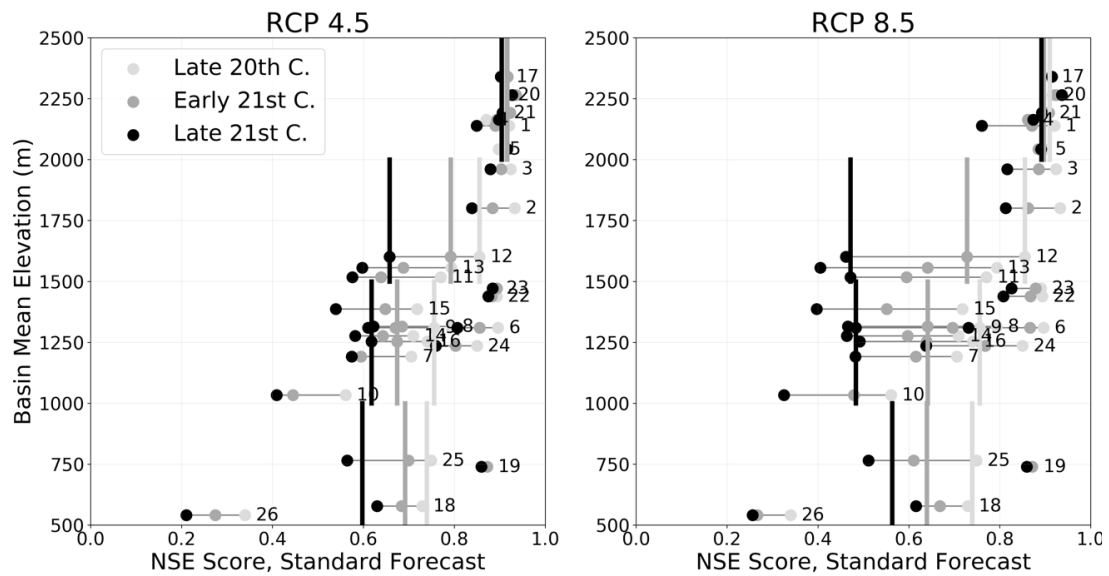
Figure A3.3: Elevation profile (mean elevation, meters) of standard forecast skill score (Nash Sutcliffe Efficiency, NSE) from the late 20th century to the early and late 21st century for RCP 4.5 (left panel) and RCP 8.5 (right panel) climate change scenarios. Vertical bars are median values for 500m elevation bins.

Figure A3.4: Elevation profile (mean elevation, meters) of standard forecast skill score (Nash Sutcliffe Efficiency, NSE) from the late 20th century to the early and late 21st century for RCP 4.5 (left panel) and RCP 8.5 (right panel) climate change scenarios. Vertical bars are median values for 500m elevation bins.
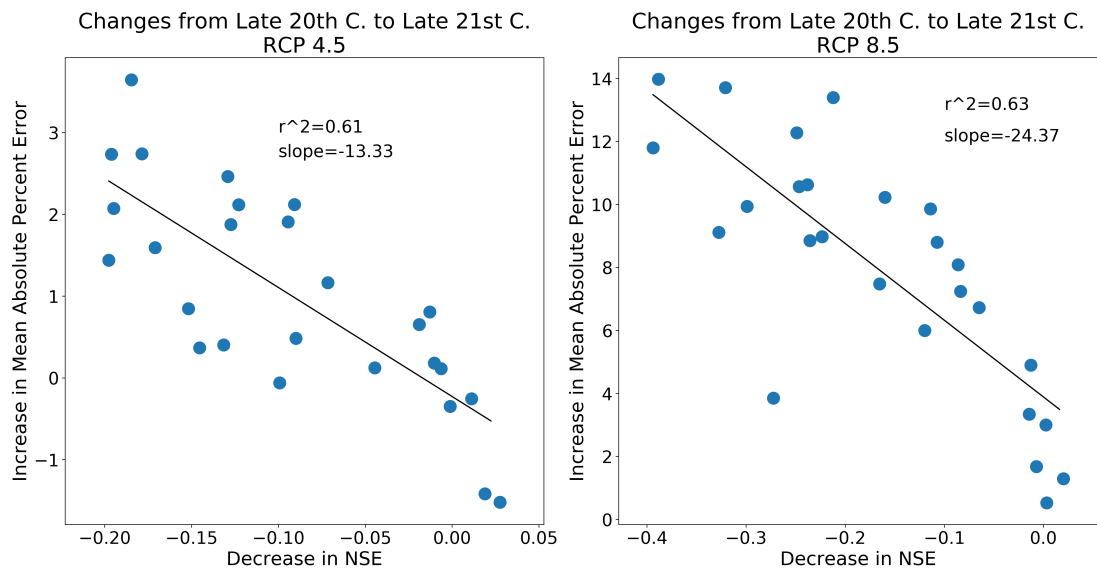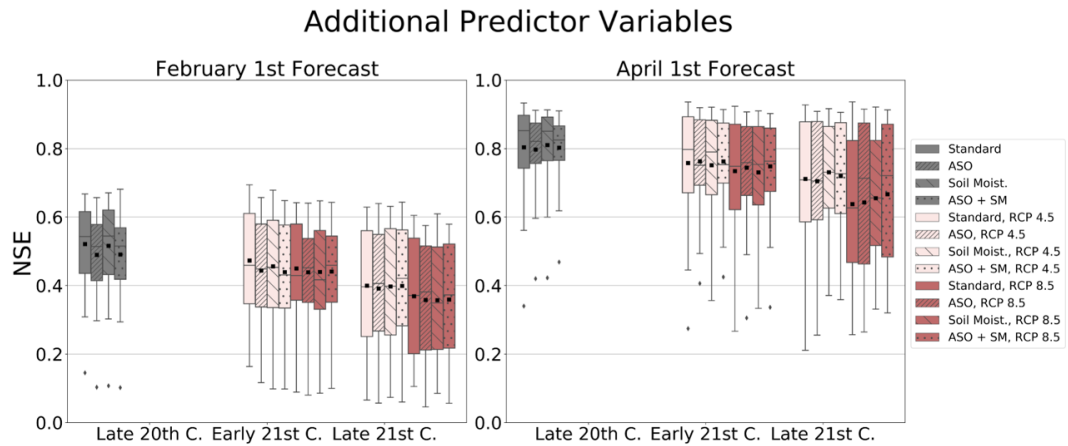
Figure A3.5: Seasonal water supply forecast skill for 26 headwater basins in the Sierra Nevada and Northern California using the CA DWR precipitation and snow water equivalent stations ('Standard', solid colors) and three different scenarios using additional simulated predictor variables including: virtual Airborne Snow Observatory (ASO) flights simulating basin-wide SWE measurements ('ASO', tight backslash), soil moisture measurements ('Soil Moist.', forward slash), and a combination of ASO-measured SWE and soil moisture ('ASO + SM', dots) under historical (grey), RCP 4.5 (pink), and RCP 8.5 (red) emission scenarios.
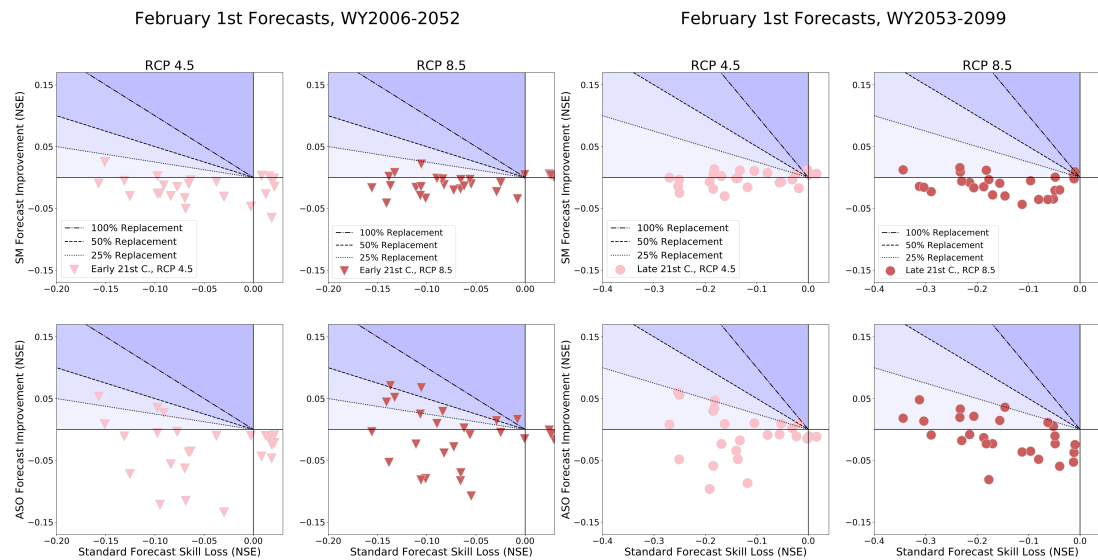
Figure A3.6: Forecast improvement (NSE) from soil moisture (SM, top row) and Airborne Snow Observatory (ASO, bottom row) enhanced forecasts as a function of the loss in forecast skill (NSE) from the late 20th century historical baseline to the early 21st century (triangles; left panels) and late 21st century (circles; right panels) for April 1st forecasts under RCP 4.5 (pink) and RCP 8.5 (red) scenarios. The 25%, 50%, and 100% Replacement lines shows the percent of lost forecast skill that is replaced by the enhanced forecast system. Blue regions highlight forecast improvements from the enhanced forecast systems, with greater improvements in darker blue.
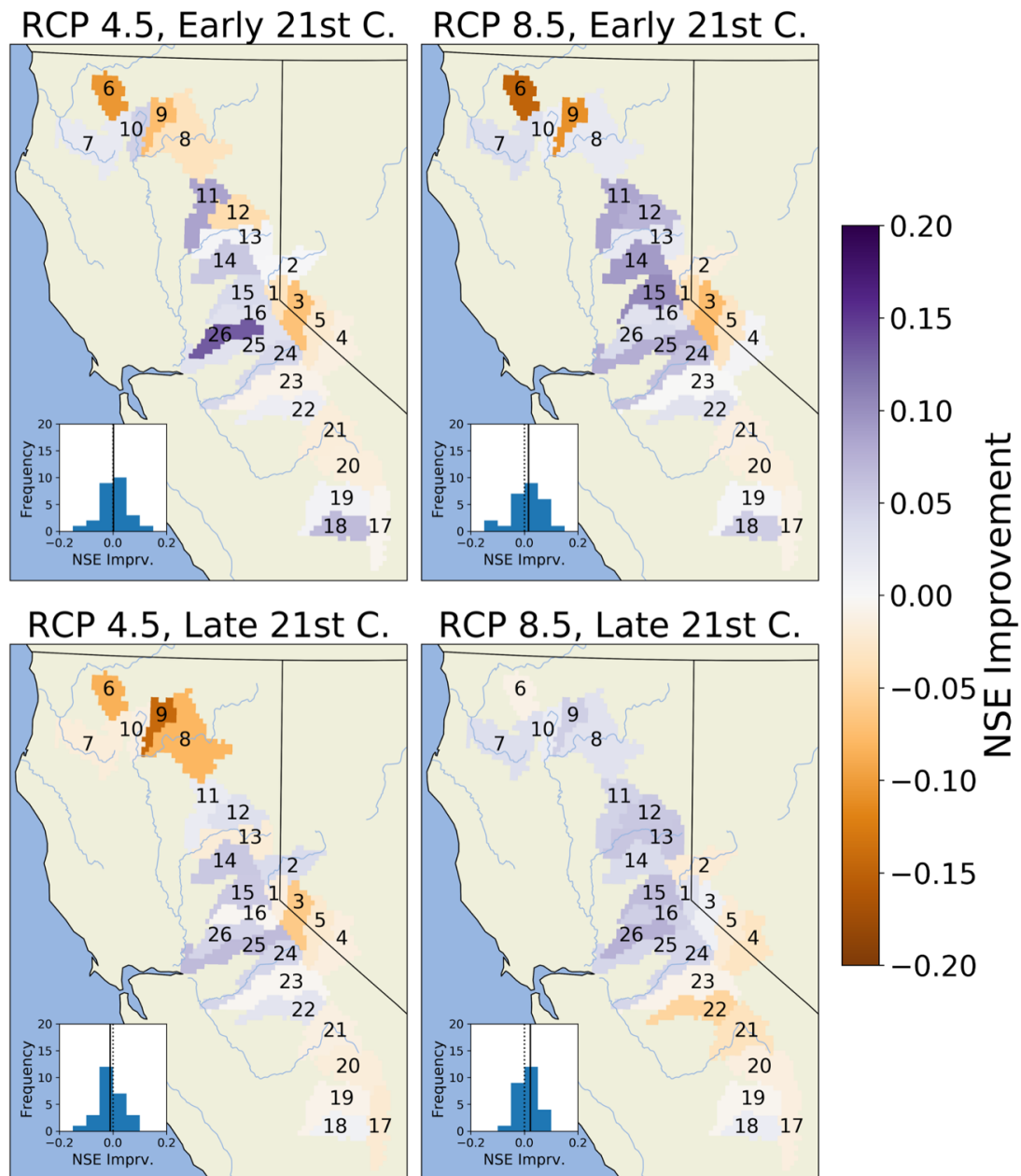
ASO Forecast Improvements

Figure A3.7: Improvement in water supply forecast skill over the April 1st standard forecasts for RCP 4.5 (left panels) and RCP 8.5 (right panels) climate change scenarios using additional predictor variables from a simulated Airborne Space Observatory (ASO) flight representing a remote measurement of SWE across the basin during the early (top panels) and late 21st century (bottom panels). Change in forecast skill measured by Nash Sutcliffe Efficiency (NSE).

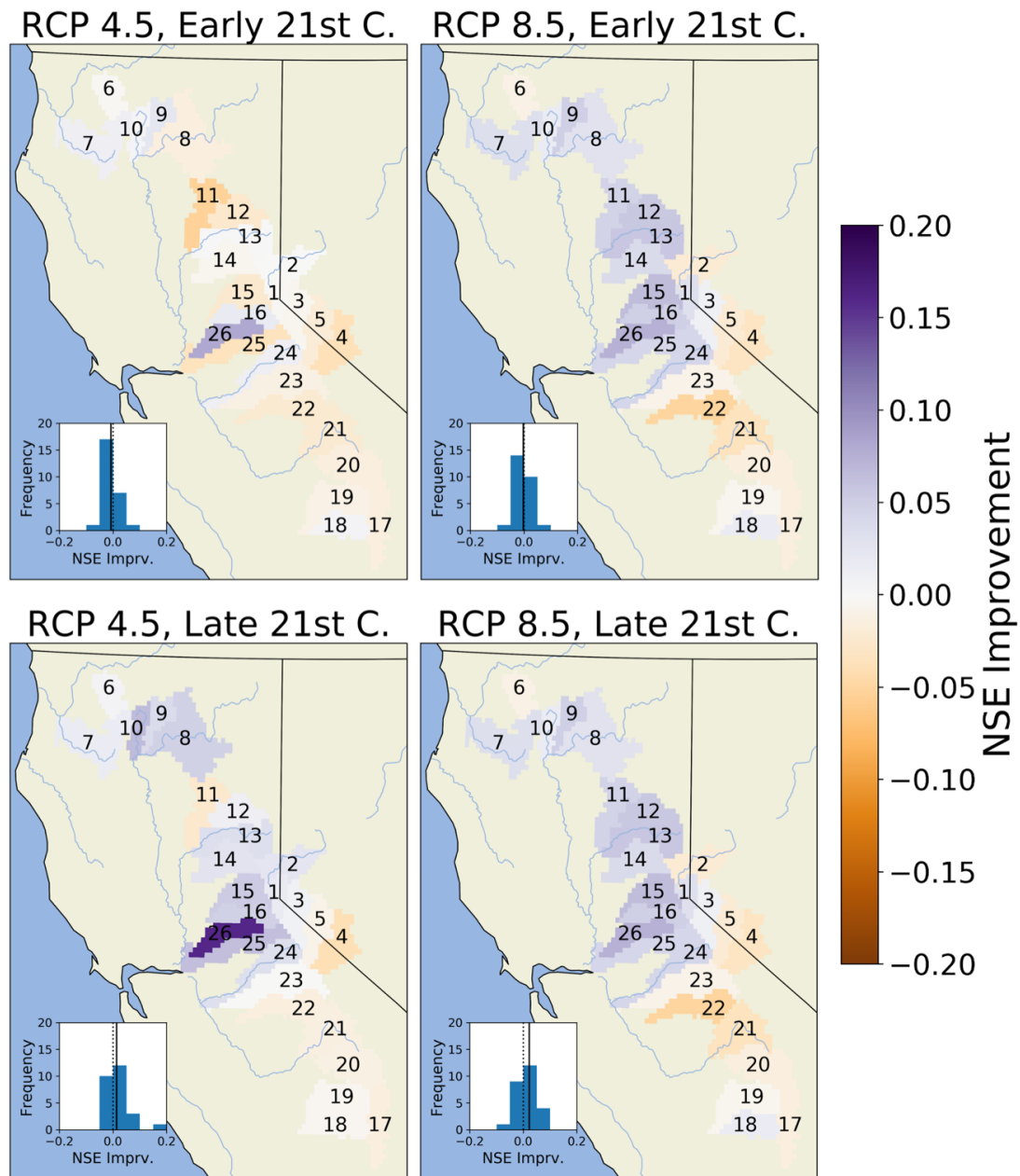**Figure A3.8:** Improvement in water supply forecast skill over the April 1st standard forecasts for RCP 4.5 (left panels) and RCP 8.5 (right panels) climate change scenarios using additional predictor variables from VIC-simulated soil moisture observations at existing precipitation stations during the early (top panels) and late 21st century (bottom panels). Change in forecast skill measured by Nash Sutcliffe Efficiency (NSE).

Chapter 4: Conclusions and Recommendations

Seasonal water supply forecasts will become increasingly important as climate change heightens the risk for drought (Cook et al., 2004) and 'weather whiplash' between dry and wet periods (Swain et al., 2018). Yet, recent research demonstrates the reduced predictability of seasonal drought under declining snowpack (Livneh & Badger, 2020) which this thesis argues similarly threatens the skill of conventional water supply forecasts. We presented research from two studies that directly evaluated the effect of below-average snowpack from interannual snow drought and climate non-stationarity on the historical (Chapter 2) and future (Chapter 3) predictability of seasonal streamflow volumes in historically snow-dominated basins. In this final chapter, we present our final conclusions as well as recommendations for future research into seasonal water supply forecast improvement.

The results of our benchmark analysis in Chapter 2 suggest that water supply forecasts are less skillful during below-average snowpacks and that long term skill decline is evident in several regions across the period of study. Our first key insight into these findings is that the conceptual hydrologic model evaluated in this study (the Sacramento Soil Moisture Accounting model) was generally less skillful than the a priori model benchmark, i.e. the regression-based models used by the U.S. Department of Agriculture's (USDA) Natural Resource Conservation Service (NRCS). Lower 'perfect prognostication' performance of the SAC-SMA model run with data through the forecast period suggests that there remains significant room for improvement of both model structure (e.g. calibration and physical process representation) and forcing data, though disaggregating these sources of error was outside the scope of this research. Specifically,

improvements to SNOW17, the simple temperature-index model used to partition precipitation, may be one fruitful area of future research to improve quantitative streamflow predictions in snow-dominated headwater basins.

Our second key insight is that there were several notable (and consequential) instances during which the conceptual model outperformed the regression-based model. We found that the conceptual model was more skillful at longer lead times, particularly during below-average snowpack. These findings suggest that in a future with less snow, a conceptual or more physically-motivated model may be a better alternative for long-lead seasonal water supply forecasting. Extrapolating this finding into the future may be precarious since previous research has shown that historical snow droughts are not a good analogue for future snowpack decline, at least in the Cascades (Cooper et al., 2016). Regardless, our research does suggest that NRCS and RFC water supply forecast coordination, whether through inter-agency forecast consultation or more formal model ensembling, may produce the best seasonal water supply forecast across snow regimes and lead times. However, regional variability of relative model performance was significant and subject to changes (notably, skill decreases) across the period of record, highlighting the need for additional model benchmarking studies of operational models.

In Chapter 3, we extend our seasonal water supply forecast evaluations from the late 20th century to the end of the 21st century using macro-scale land surface and hydrologic models to more directly quantify long term forecast skill trends. The results of this analysis confirm that future water supply forecast skill in the Sierra Nevada, U.S. is projected to degrade under declining snowpack conditions. These findings were consistent with the Chapter 2 finding forecast performance suffered during below-

average snowpack. Using Variable Infiltration Capacity (VIC) model simulations, we demonstrated that loss of forecast skill is strongly correlated with loss of snow water equivalent (SWE), which is mediated by basin elevation. Because of the importance of elevation, we simulated that northern Sierra Nevada headwater basins in the 1000-1700m mean elevation band were particularly vulnerable, with maximum forecast skill loss of over 40% by the late 21st century. Conversely, high-elevation basins in the southern Sierra Nevada and historically less snow-dominated basins had more resilient forecast performance even under high emission scenarios (i.e. RCP 8.5).

Leveraging the simulated model environment of Chapter 3, we were also able to provide insight into two operationally feasible techniques for mitigating loss of forecast skill. The first technique simulated a remotely-sensed basin-wide SWE measurement modeled after the Airborne Snow Observatory (ASO). We were able to demonstrate that during the first-half of the 21st century, a virtual ASO flight replaced an average of 40% of lost forecast skill in the most at-risk basins. In several instances, enhanced forecast techniques replaced more than 100% of lost skill. However, the ability of enhanced SWE measurements to buffer skill loss dwindled as snow declined over the long term. The second technique simulated the installation of a soil moisture observation network. We were able to show that by the late 21st century, soil moisture observations could replace approximately 10-25% of lost forecast skill, providing a safety net for lost forecast information associated with the loss of snowpack. These two analyses offer further scientific evidence in line with findings from previous empirical (Harpold et al., 2017) and modeled (Koster et al., 2010; Rosenberg et al., 2013) research showing the value of

including supplemental observations in streamflow volume forecasting in snow-dominated basins.

The one-two combination of decreased snow water storage and degrading predictive power of snowpack during low snow years is already or will soon become a significant challenge for reservoir operators and water managers. Improving operational water supply forecasts, however, is a long-term game requiring incremental scientific progress. Use-inspired and actionable scientific research in concert with those who have the authority to implement improvement techniques will accelerate this progress. Promising areas of research for seasonal water supply forecast improvement include the development of advanced statistical and modeling techniques, next-generation distributed hydrologic models, expanded observational networks (including remotely-sensed data assimilation), and model benchmarking studies, each of which may incrementally contribute to the next generation of long-range quantitative streamflow forecasting models ready for predicting the hydrologic response to a non-stationary climate.

References

Cook, E., Woodhouse, C., Eakin, C. M., Meko, D., & Stahle, D. (2004). Long-Term Aridity Changes in the Western United States. Science, 306, 1015–1018. https://doi.org/10.1126/science.1101982

Cooper, M., Nolin, A., & Safeeq, M. (2016). Testing the recent snow drought as an analog for climate warming sensitivity of Cascades snowpacks - IOPscience. Environmental Research Letters, 11. Retrieved from https://iopscience.iop.org/article/10.1088/1748-9326/11/8/084009

Harpold, A. A., Sutcliffe, K., Clayton, J., Goodbody, A., & Vazquez, S. (2017). Does Including Soil Moisture Observations Improve Operational Streamflow Forecasts in Snow-Dominated Watersheds? Journal of the American Water Resources Association, 53(1), 179–196. https://doi.org/10.1111/1752-1688.12490

Koster, R. D., Mahanama, S. P. P., Livneh, B., Lettenmaier, D. P., & Reichle, R. H. (2010). Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow. Nature Geoscience, 3(9), 613–616. https://doi.org/10.1038/ngeo944

Livneh, B., & Badger, A. M. (2020). Drought less predictable under declining future snowpack. Nature Climate Change, 1–7. https://doi.org/10.1038/s41558-020-0754-8

Rosenberg, E. A., Wood, A. W., & Steinemann, A. C. (2013). Informing Hydrometric Network Design for Statistical Seasonal Streamflow Forecasts. Journal of Hydrometeorology, 14(5), 1587–1604. https://doi.org/10.1175/JHM-D-12-0136.1

Swain, D. L., Langenbrunner, B., Neelin, J. D., & Hall, A. (2018). Increasing precipitation volatility in twenty-first-century California. Nature Climate Change, 8(5), 427–433. https://doi.org/10.1038/s41558-018-0140-y