Title

Mountain snowpacks in the Western U.S: improved estimation and understanding the impact on future water availability

Author

von Kaenel, Manon

Publication Date

2024

UNIVERSITY OF CALIFORNIA

Los Angeles

Mountain snowpacks in the Western U.S: improved estimation and understanding the impact on future water availability

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy in Civil Engineering

by

Manon Line von Kaenel

2024

ABSTRACT OF THE DISSERTATION

Mountain snowpacks in the Western U.S: improved estimation and understanding the impact on future water availability

by

Manon Line von Kaenel

Doctor of Philosophy in Civil Engineering

University of California, Los Angeles, 2024

Professor Steven A. Margulis, Chair

Seasonal snowpack serves as natural reservoir by storing winter precipitation and releasing it as snowmelt. In in the Western U.S., this is a crucial water supply that supports agriculture, hydropower, ecosystems, and millions of users. Rising temperatures are causing reduced snow storage, earlier melt, and increased drought risk. Future climate models predict these trends will continue and intensify, posing challenges for water management. Accurate snow water equivalent (SWE) estimates are essential for water managers in snowmelt-reliant regions, but characterizing the spatial distribution of snow is an ongoing challenge. In situ measurements are not always representative nor widespread and remote sensing of mountain SWE remains elusive. Modeling can fill space-time gaps in the observational record but is impacted by biases in forcings and uncertainties in model physics.

To address the issue of uncertainties in model physics for simulating snow, we evaluate how altering the configurations of a land surface model (Noah-MP) affects its ability to recreate observed SWE across 199 stations in the Western U.S. The base case configuration, which matches that for the National Water Model, overestimates SWE at 90% of sites. Adjustments to model physics for precipitation partitioning, snow albedo formulation, and surface resistance cause significant changes in SWE predictions that vary by season and site climate and geography. No single configuration works best everywhere, but four alternatives outperform the base case at most sites.

To address the challenge of biases in mountain precipitation products, we leverage a historical snow reanalysis dataset to develop, apply, and test a novel precipitation bias correction and downscaling method towards modeling SWE in a real-time context. Over a test domain, this precipitation bias correction is effective in reducing error in April 1st SWE (-58%) and streamflow forecasts (-52%). Assimilating remotely-sensed snow depth observations further reduces errors.

To explore the impact of future shifts in snowpack on water resources, we apply hydrology projections driven by downscaled global climate models (GCMs) and a simple reservoir operations model to 13 major reservoirs in the California Sierra Nevada. Region wide, snowpack reductions (-44%) and earlier snowmelt (11 days) lead to earlier inflow and drops in water deliveries (-19%) and year-end storage (-18%). Reservoir storage and rainfall help offset the impact of snowpack losses, but the extent and mechanisms of this vary on the reservoir's operations, characteristics, and upstream climate and hydrology. Current operating rules are not well-suited to let reservoirs store earlier inflow under future climate conditions.

The dissertation of Manon Line von Kaenel is approved.

Dennis P. Lettenmaier

Timu Gallien

Alvar Escriva-Bou

Marisa Escobar Arias

Steven A. Margulis, Committee Chair

University of California, Los Angeles

2024

Table of Contents

List of Figures

List of Tables

List of Acronyms

AORC	Analysis of Record for Calibration
BATS	Biosphere-Atmosphere Transfer Scheme
CADWR	California Department of Water Resources
CDEC	California Data Exchange Center
DOWY	day of water year
FIRO	forecast-informed reservoir operations
FNR	false negative rate
fSCA	fractional snow-covered area
GCM	global climate model
IQR	interquartile range
KGE	Kling-Gupta efficiency
KS	Kolmogorov-Smirnov
LOCA	Localized Constructed Analogs
LSM	land surface model
MAE	mean absolute error
MD	mean difference
MERRA	Modern-Era Retrospective Analysis for Research and Applications
MNB	mean normalized bias
NOAA	National Oceanic and Atmospheric Administration
Noah-MP	Noah-Multiparameterization
NRCS	National Resources Conservation Service
NRMSD	normalized root mean square difference
NWM	National Water Model
QA/QC	quality assurance and quality control
SLOP	standard linear operating policy
SNOTEL	SNOwpackTELemetry network
SWE	snow water equivalent
TAF	thousand acre-foot
USACE	United States Army Corps of Engineers
VIC	Variable Infiltration Capacity
WRF	Weather Research and Forecasting
WUS	Western United States
WY	water year

Acknowledgements

This work is supported by the NASA FINESST (Future Investigators in Earth and Space Science and Technology) fellowship (80NSSC21K1627), and the NOAA OAR/OWAQ Observations Program (award NA18OAR4590396).

I have many people to acknowledge and thank for their contributions to this dissertation. Thank you first and foremost to my advisor Steve Margulis for setting a strong example and the invaluable guidance that has been instrumental in my growth as a scholar. Thank you to my committee members for the expert insights, patience, and support. Thank you to Alvar Escriva-Bou for providing encouragement, expertise, and guidance in the completion of Chapter 4.

Thank you to my fellow PhD students – past and present – for the hard work, companionship, and inspiration. Specifically, thank you to Sarfaraz Alam for the priceless mentorship, and Yufei Liu and Yiwen Fang for laying the groundwork and the sage advice.

Thank you to my friends for providing me the emotional sustenance to get to the end of this program. Thank you to my family for believing in me and giving me strength and fortitude. Specifically, thank you to my sisters Camille and Noémie for the encouragement and confidence, and my dad Vincent for the steadfast support.

Thank you to all the cheerleaders, teachers, helpers, confidants, role models, and guides who are perhaps unnamed here but not forgotten.

This dissertation is dedicated to the beautiful Sierra Nevada mountains and the Swiss Alps which both have brought me much joy and deserve all the care and attention.

Vita

Education

University of California, Los Angeles

M.S., Civil Engineering | Hydrology and Water Resources, GPA**:** 3.9 *June 2019*

University of California, Berkeley

B.S., Environmental Sciences (honors), GPA: 3.8 *December 2015*

B.A., Geography (high honors), GPA: 4.0

Double minor: Geospatial Information Science and Technology, Spanish

Selected Publications

von Kaenel, M., & Margulis, S. A. (2024). Improved modelling of mountain snowpacks with spatially distributed precipitation bias correction derived from historical reanalysis. *EGUsphere* [preprint]. https://doi.org/10.5194/egusphere-2024-3389

von Kaenel, M., & Margulis, S. A. (2024). Evaluation of Noah-MP snow simulation across site conditions in the Western United States. *Journal of Hydrometeorology, 25*(9), 1389-1406. https://doi.org/10.1175/JHM-D-23-0211.1

Alam, S., **von Kaenel, M.**, Su, L., & Lettenmaier, D. P. (2024). The role of antecedent winter soil moisture carryover on spring runoff predictability in snow-influenced Western US catchments. *Journal of Hydrometeorology, 25*(10), 1561-1575. https://doi.org/10.1175/JHM-D-24-0038.1

Selected Conferences and Presentations

von Kaenel, M., & Margulis, S. A. (2023, December). Leveraging historical snow datasets to predict spatially distributed SWE in near real time. *AGU Fall Meeting 2023*, San Francisco, CA.

Fang, Y., **von Kaenel, M.**, Li, D., Liu, Y., & Margulis, S. A. (2023, January). Characterizing the unusual Western U.S. snow-streamflow drought in 2021 using a novel snow reanalysis dataset and in situ measurements. *103rd AMS Annual Meeting 2023*. American Meteorological Society. [Invited].

von Kaenel, M., Alam, S., Vimal, S., Su, L., Margulis, S. A., & Lettenmaier, D. P. (2019, December). The role of soil moisture memory in spring runoff predictability in Western U.S. river basins. *AGU Fall Meeting 2019*, San Francisco, CA.

Teaching Experience

University of California Los Angeles | Department of Civil and Environmental Engineering

Teaching Assistant, Introduction to Hydrology *Fall 2019-2023*

Relevant Employment

California Department of Water Resources | Snow Surveys & Water Supply Forecasting Unit

Water Resources Engineer *September 2024-present*

Stockholm Environment Institute | U.S. Center Water Group

Scientist *Jan 2017-June 2018*

Research Analyst *Jan 2016-Dec 2016*

Intern *July 2015-Aug 2015*

CHAPTER 1

Introduction

1.1 Background and Motivation

Seasonal snowpack is a natural water tower; by storing winter precipitation and releasing it as snowmelt, it provides an essential resource for downstream ecosystems and society (Viviroli et al., 2007; Sturm et al., 2017; Immerzeel et al., 2020), including an estimated 20% of the Earth's population (Dozier, 2011). In the Western U.S., over half of total runoff originates as snowmelt, and snowmelt-driven runoff makes up two-thirds of the inflow to the region's major reservoirs (Li et al., 2017). This water supplies agricultural, urban, and industrial demands across the region; sustains hydroelectric power production; and supports wetlands, habitat, and other ecosystem services.

Snow water resources in the Western U.S. are both vital and vulnerable to climatic variations. Rising temperatures have already caused changes in snow hydrology that have broad implications for downstream environments and communities. Widespread declining snow storage trends have been observed across the Western U.S. (Mote et al., 2005; Mote et al., 2018). The timing of spring snowmelt-driven streamflow has shifted earlier in the year (Cayan et al., 2001; Stewart et al., 2005) with snowmelt starting earlier (Kapnick and Hall, 2012; Musselman et al., 2021). Climate models project a continued trend towards earlier melt, reduced snow water availability, precipitation extremes, and drought conditions into the future (e.g., Barnett et al., 2005; Pierce and Cayan, 2013; Schwartz et al., 2017; Siirila-Woodburn et al., 2021). These shifts are projected to have important impacts on water supply and management; in mountain ranges of the Western U.S., this includes lower reservoir storage, decreased water supply deliveries and reliability, and reduced hydropower generation (e.g., Vicuna et al., 2007; Rheinheimer et al., 2012; Knowles et al., 2018; Cohen et al., 2020; Ray et al., 2020).

1.2 Challenges for large-scale snow estimation

The water management system in the Western U.S. – including access to information, policy, and physical infrastructure – needs to be resilient to vulnerabilities in regional snow-derived water resources to ensure water security. Access to high-quality water data is key for effective integrated water management (NASEM, 2018). In regions that rely on snowmelt for water supply, water managers need accurate assessments of the distribution and availability of water in snowpack (SWE, snow water equivalent) (e.g., Hamlet et al., 2002; Koster et al., 2010; He et al., 2016).

However, estimating the spatiotemporal distribution and change of SWE remains a significant and important challenge for the snow hydrology community (e.g., Lettenmaier et al., 2015; Dozier et al., 2016; Cho et al., 2022). In situ data are not always representative of the heterogeneity of SWE distribution in topographically complex mountain landscapes (Herbert et al., 2024), and such measurements are sparse globally. Remote sensing can provide observations of snow properties over large areas, but there is currently no reliable way of measuring SWE from spaceborne platforms (Lettenmaier et al., 2015). This implies a continued need for modeling of SWE. Land surface models are commonly used to estimate SWE and other hydrologic variables over large spatial extents and at varying resolutions (e.g., Clark et al., 2011;

Kumar et al., 2013; Cho et al., 2022), but these are susceptible to uncertainties driven by biases in forcing data or model parameterization (Cho et al., 2022). Uncertainty and biases in precipitation products over mountainous terrain are widely acknowledged, and these continue to dominate uncertainty in SWE estimation (e.g., Pan et al., 2003; Raleigh et al., 2015; Liu and Margulis, 2019; Schreiner-McGraw et al., 2020; Cho et al., 2022; Fang et al., 2023).

1.3 Objective and organization of dissertation

The intent of this dissertation is to promote the improvement of snow estimation for regional hydrologic applications to support resilient water management. It addresses two of the most important challenges in modeling SWE: uncertainties in model parametrization and precipitation biases; and explores how and why shifts in seasonal snowpacks affect water supply under climate change. As such, the overarching objective is to demonstrate *how* improved SWE estimation can be achieved and *why* this is important going forward. Specifically, this thesis addresses the following scientific questions:

1) How do changes in land surface model configurations affect SWE simulations across site conditions?
2) How can historically informed mountain precipitation bias correction improve model-based SWE estimates and streamflow forecasts, and how does the assimilation of independent snow observations impact those estimates?
3) How do projected changes in snowpack affect water availability in reservoirs across the California Sierra Nevada under a future climate?

The dissertation is organized into five chapters: Chapter 2 quantifies the performance and sensitivity of different model (Noah-MP) configurations in simulating snow over the Western U.S. to answer the first question. Chapter 3 answers the second question by demonstrating an effective method for mountain precipitation bias correction for SWE estimation by leveraging a historical reanalysis dataset over a test domain. Chapter 4 addresses the third question by quantifying, comparing, and explaining the impact of projected snowpack and hydrology shifts on water deliveries and reservoir storage in the California Sierra Nevada. Chapter 2 is published as a peer-reviewed paper (von Kaenel and Margulis, 2024a), Chapter 3 is published as a preprint currently under review (von Kaenel and Margulis, 2024b), and Chapter 4 provides the initial assessment for a future publication. Chapter 5 summarizes key findings, conclusions, and novel contributions from this research and proposes future work. Appendix A, B, and C provide supplemental information for Chapters 2, 3, and 4, respectively.

1.3 References

Barnett, T. P., Adam, J. C., & Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, *438*(7066), 303–309. https://doi.org/10.1038/nature04141

Cayan, D. R., Dettinger, M. D., Kammerdiener, S. A., Caprio, J. M., & Peterson, D. H. (2001). Changes in the onset of spring in the western United States. *Bulletin of the American Meteorological Society*, *82*(3), 399–415. https://doi.org/10.1175/1520-0477(2001)082<0399:CITOOS>2.3.CO;2

Clark, M. P., Hendrikx, J., Slater, A. G., Kavetski, D., Anderson, B., Cullen, N. J., Kerr, T., Örn Hreinsson, E., & Woods, R. A. (2011). Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review. *Water Resources Research, 47*(7). https://doi.org/10.1029/2011WR010745

Cho, E., Vuyovich, C. M., Kumar, S. V., Wrzesien, M. L., Kim, R. S., & Jacobs, J. M. (2022). Precipitation biases and snow physics limitations drive the uncertainties in macroscale modeled snow water equivalent. *Hydrology and Earth System Sciences*, *26*(22), 5721–5735. https://doi.org/10.5194/hess-26-5721-2022

Cohen, J. S., Zeff, H. B., & Herman, J. D. (2020). Adaptation of multiobjective reservoir operations to snowpack decline in the western United States. *Journal of Water Resources Planning and Management, 146*(12), 04020091. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001300

Dozier, J. (2011). Mountain hydrology, snow color, and the fourth paradigm. *Eos, Transactions American Geophysical Union, 92*(43), 373–374. https://doi.org/10.1029/2011EO430001

Dozier, J., Bair, E. H., & Davis, R. E. (2016). Estimating the spatial distribution of snow water equivalent in the world's mountains. *Wiley Interdisciplinary Reviews: Water, 3*(3), 461–474. https://doi.org/10.1002/wat2.1140

Fang, Y., Liu, Y., Li, D., Sun, H., & Margulis, S. A. (2023). Spatiotemporal snow water storage uncertainty in the midlatitude American Cordillera. *The Cryosphere*, *17*(12), 5175–5195. https://doi.org/10.5194/tc-17-5175-2023

Hamlet, A. F., Huppert, D., & Lettenmaier, D. P. (2002). Economic value of long-lead streamflow forecasts for Columbia River hydropower. *Journal of Water Resources*

Planning and Management, 128(2), 91–101. https://doi.org/10.1061/(ASCE)0733-9496(2002)128:2(91)

He, M., Russo, M., & Anderson, M. (2016). Predictability of seasonal streamflow in a changing climate in the Sierra Nevada. *Climate, 4*(4), 57. https://doi.org/10.3390/cli4040057

Immerzeel, W. W., Lutz, A. F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., Hyde, S., Brumby, S., Davies, B. J., Elmore, A. C., Emmer, A., Feng, M., Fernández, A., Haritashya, U., Kargel, J. S., Koppes, M., Kraaijenbrink, P. D. A., Kulkarni, A. V., Mayewski, P. A., … Baillie, J. E. M. (2020). Importance and vulnerability of the world's water towers. *Nature, 577*(7790), 364–369. https://doi.org/10.1038/s41586-019-1822-y

Kapnick, S., & Hall, A. (2012). Causes of recent changes in western North American snowpack. *Climate Dynamics, 38*(9–10), 1885–1899. https://doi.org/10.1007/s00382-011-1089-y

Knowles, N., Cronkite-Ratcliff, C., Pierce, D. W., & Cayan, D. R. (2018). Responses of unimpaired flows, storage, and managed flows to scenarios of climate change in the San Francisco Bay-Delta watershed. *Water Resources Research, 54*(10), 7631–7650. https://doi.org/10.1029/2018WR022852

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

Kumar, M., Marks, D., Dozier, J., Reba, M., & Winstral, A. (2013). Evaluation of distributed hydrologic impacts of temperature-index and energy-based snow models. *Advances in Water Resources, 56*, 77–89. https://doi.org/10.1016/j.advwatres.2013.03.006

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., & Wood, E. F. (2015). Inroads of remote sensing into hydrologic science during the WRR era: Remote sensing. *Water Resources Research, 51*(9), 7309–7342. https://doi.org/10.1002/2015WR017616

Li, D., Wrzesien, M. L., Durand, M., Adam, J., & Lettenmaier, D. P. (2017). How much runoff originates as snow in the western United States, and how will that change in the future? *Geophysical Research Letters*, *44*(12), 6163–6172. https://doi.org/10.1002/2017GL073551

Liu, Y., & Margulis, S. A. (2019). Deriving bias and uncertainty in MERRA-2 snowfall precipitation over High Mountain Asia. *Frontiers in Earth Science, 7*. https://doi.org/10.3389/feart.2019.00280

Mote, P. W., Hamlet, A. F., Clark, M. P., & Lettenmaier, D. P. (2005). Declining mountain snowpack in western North America. *Bulletin of the American Meteorological Society, 86*(1), 39–50. https://doi.org/10.1175/BAMS-86-1-39

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

Musselman, K. N., Addor, N., Vano, J. A., & Molotch, N. P. (2021). Winter melt trends portend widespread declines in snow water resources. *Nature Climate Change, 11*(5), 418–424. https://doi.org/10.1038/s41558-021-01014-9

National Academies of Sciences, Engineering, and Medicine. (2018). *Thriving on our changing planet: A decadal strategy for Earth observation from space.* Washington, DC: The National Academies Press. https://doi.org/10.17226/24938

Pan, M., Sheffield, J., Wood, E., Mitchell, K., Houser, P., Schaake, J., Robock, A., Lohmann, D., Cosgrove, B., Duan, Q., Luo, L., Higgins, R., Pinker, R., & Tarpley, J. (2003). Snow process modeling in the North American Land Data Assimilation System (NLDAS): 2. Evaluation of model simulated snow water equivalent. *Journal of Geophysical Research: Atmospheres, 108*(D22). https://doi.org/10.1029/2003JD003994

Pierce, D. W., Cayan, D. R., Feldman, D. R., & Risser, M. D. (2023). Future increases in North American extreme precipitation in CMIP6 downscaled with LOCA. *Journal of Hydrometeorology, 24*(5), 951–975. https://doi.org/10.1175/JHM-D-22-0194.1

Raleigh, M. S., Lundquist, J. D., & Clark, M. P. (2015). Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. *Hydrology and Earth System Sciences, 19*(7), 3153–3179. https://doi.org/10.5194/hess-19-3153-2015

Ray, P., Wi, S., Schwarz, A., Correa, M., He, M., & Brown, C. (2020). Vulnerability and risk: Climate change and water supply from California’s Central Valley water system. *Climatic Change, 161*(1), 177–199. https://doi.org/10.1007/s10584-020-02655-z

Rheinheimer, D. R., Ligare, S. T., & Viers, J. H. (2012). Water-energy sector vulnerability to climate warming in the Sierra Nevada: Simulating the regulated rivers of California’s west slope Sierra Nevada. *California Energy Commission* (Publication No. CEC-500-2012-016). Retrieved from https://escholarship.org/uc/item/8h62s4xg

Schreiner-McGraw, A. P., & Ajami, H. (2020). Impact of uncertainty in precipitation forcing data sets on the hydrologic budget of an integrated hydrologic model in mountainous terrain. *Water Resources Research, 56*(12), e2020WR027639. https://doi.org/10.1029/2020WR027639

Schwartz, M., Hall, A., Sun, F., Walton, D., & Berg, N. (2017). Significant and inevitable end-of-twenty-first-century advances in surface runoff timing in California's Sierra Nevada. *Journal of Hydrometeorology, 18*(12), 3181–3197. https://doi.org/10.1175/JHM-D-16-0257.1

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

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

Sturm, M., Goldstein, M. A., & Parr, C. (2017). Water and life from snow: A trillion dollar science question. *Water Resources Research, 53*(5), 3534–3544. https://doi.org/10.1002/2017WR020840

Viviroli, D., Dürr, H. H., Messerli, B., Meybeck, M., & Weingartner, R. (2007). Mountains of the world, water towers for humanity: Typology, mapping, and global significance. *Water Resources Research, 43*(7). https://doi.org/10.1029/2006WR005653

Vicuna, S., Maurer, E. P., Joyce, B., Dracup, J. A., & Purkey, D. (2007). The sensitivity of California water resources to climate change scenarios. *Journal of the American Water Resources Association, 43*(2), 482–498. https://doi.org/10.1111/j.1752-1688.2007.00038.x

von Kaenel, M., & Margulis, S. A. (2024a). Evaluation of Noah-MP snow simulation across site conditions in the western United States. *Journal of Hydrometeorology, 25*(9), 1389–1406. https://doi.org/10.1175/JHM-D-23-0211.1

von Kaenel, M., & Margulis, S. (2024b). Improved modelling of mountain snowpacks with spatially distributed precipitation bias correction derived from historical reanalysis. *EGUsphere* [Preprint]. https://doi.org/10.5194/egusphere-2024-3389

CHAPTER 2

Evaluation of Noah-MP snow simulation across site conditions in the Western US

The text and figures from this chapter are published in:

von Kaenel, M., & Margulis, S. A. (2024). Evaluation of Noah-MP snow simulation across site conditions in the Western United States. *Journal of Hydrometeorology, 25*(9), 1389-1406. https://doi.org/10.1175/JHM-D-23-0211.1 © American Meteorological Society. Used with permission.

2.0 Abstract

Quantifying spatio-temporal variability in snow water resources is a challenge especially relevant in regions that rely on snowmelt for water supply. Model accuracy is often limited by uncertainties in meteorological forcings and/or suboptimal physics representation. In this study, we evaluate the performance and sensitivity of Noah-MP snow simulations from ten model configurations across 199 sites in the Western US. Nine experiments are constrained by observed meteorology to test snow-related physics options, and the tenth tests an alternative source of meteorological forcings. We find that the base case, which aligns with the National Water Model configuration and uses observations-based forcings, overestimates observed accumulated SWE at 90% of stations by a median of 9.6%. The model performs better in the accumulation season at colder, drier sites and in the melt season at wetter, warmer sites. Accumulation metrics are

sensitive to model configuration in two experiments, and melt metrics in six. Alterations to model physics cause changes to median accumulation metrics from -13% to 2.3% with the greatest change due to precipitation partitioning; and to melt metrics from -10% to 3% with the greatest change due to surface resistance configuration. The experiment with alternative forcings causes even greater and wider-ranging changes (medians ranging -29% to 6%). Not all stations share the same best-performing model configuration. At most stations, the base case is outperformed by four alternative physics options which also significantly impact snow simulation. This research provides insights into the performance and sensitivity of snow predictions across site conditions and model configurations.

2.1 Significance statement

The purpose of this work is to evaluate the performance and sensitivity of a land surface model's simulation of snow across site conditions and in response to different model configurations. This is important because estimating snow distribution is a challenge especially relevant for regions that rely on snowmelt for water supply. While land surface models can provide useful large-scale estimates, they are often limited by uncertainties in forcings and/or suboptimal physics representation. The results, which show varying model behavior across geography, climate, vegetation types, and model configurations, highlight inadequacies in model physics representation, emphasize the need for accurate meteorological forcings, and suggest that customizing model configurations to the unique characteristics of the domain could yield more accurate and useful results.

2.2 Introduction

Estimating snow water equivalent (SWE) in remote mountainous areas remains one of the most challenging problems in hydrology (Lettenmaier et al., 2015; Dozier et al., 2016). This challenge is particularly important in areas where snowpack plays a significant role in seasonal water supply and regional hydrology, such as the Western US. To make critical management decisions for flood control, hydropower operations, irrigation, and other competing demands in such regions, water managers need accurate assessments of the space-time distribution and availability of water in snowpack (Vicuña et al., 2011; Tanaka et al., 2006; He et al., 2016). Snowpack observations can come from in-situ stations or remote sensing platforms; however, both data sources are limited in time and space and are prone to substantial errors, especially in topographically complex areas (Dozier et al., 2016).

Land surface models (LSMs) are used to generate spatially-distributed estimates of SWE, snow-derived runoff, and other hydrological variables over large spatial domains and at varying spatiotemporal resolutions by simulating the physics of the water and energy cycles at the land surface. The accuracy of these modeled estimates depends on the reliability of meteorological forcings as well as the fidelity of model physics (Cho et al., 2022). Various studies (e.g., Cho et al., 2022; Kim et al., 2021; Pan et al., 2003; Barlage et al., 2010) have demonstrated that a common weakness in LSMs is that SWE estimates are highly uncertain and often underestimated, which can cascade into uncertainties and errors in SWE-dependent variables like runoff and evapotranspiration. Properly describing and simulating snow processes in a LSM thus helps to not only accurately predict the distribution of snow water resources but also other variables in the land surface water cycle.

Here, we examine the performance and sensitivity of SWE estimates produced by the Noah-Multiparameterization LSM (Noah-MP). Noah-MP is the hydrologic core and LSM for the operational National Water Model (NWM) and associated climate predictions based on the Weather Research and Forecasting (WRF) regional climate model. Of the four LSMs tested by Cho et al. (2022), Noah-MP generated the most accurate maximum SWE values, but it nonetheless significantly underestimated observations, and several limitations in current physics options were noted. Past studies have also identified biases in the Noah-MP representation of snow depth, timing of snow disappearance, and annual SWE (e.g., Li et al., 2022a; Li et al., 2022b; You et al., 2020; Garousi-Nejad et al., 2021; He et al., 2021). The design of Noah-MP allows for user-defined selection of options for physical processes such as precipitation partitioning, snow albedo, and vegetation-snow interactions. Past studies have explored how the incorporation of different schemes for some or most of these physical processes in Noah-MP affects the simulation of snow depth or SWE at specific sites (You et al., 2020; Zhang et al., 2016; Letcher et al., 2022; Wang et al., 2019) and at a global scale (Li et al., 2022a). However, for the purposes of understanding uncertainties in snow simulation in a regional hydrologic model like the NWM, these studies were limited in scope by number of sites or model resolution.

In this study we evaluate the performance and sensitivity of Noah-MP snow simulations across different model configurations at sites spanning the Western U.S. (WUS). The WUS contains extensive in situ data that can force and validate the model and represents a diversity of climate and site conditions. For the NWM, the WUS mountains is also where the snow model plays the largest role in runoff/streamflow prediction. Our intent is to provide insights into model behavior relative to user-selected physics options and promote the improvement of snow simulation in regional climate and hydrologic applications. To this end, we aim to answer: 1) how well can the National Water Model configuration of Noah-MP reproduce observed SWE at sites across the Western U.S.?, 2) how sensitive are the Noah-MP snow simulations to changes in model configurations?, and 3) how do model errors and sensitivities vary by region, climate, and vegetation conditions?

2.3 Data

2.3.1 Study sites

The study domain is comprised of 199 stations from the Snow Telemetry (SNOTEL) network across the WUS that meet the following criteria: a) less than 5% of daily precipitation and hourly temperature observations for Nov-June over water years 2007-2019 are missing; and b) no daily SWE observations over the study record are missing. Daily precipitation and SWE records were taken from the bias-corrected quality-controlled product published by the Pacific Northwest National Laboratory (PNNL; Yan et al., 2018), and hourly temperature data were downloaded from the National Resources Conservation Service (NRCS) web portal.

We classified the 199 sites into groups based on geography, climate, and vegetation type to evaluate how the model performs across site conditions that are relevant to snow accumulation and/or melt (Fig. 2.1). We assigned each station to an eco-region based on the Commission for Environmental Cooperation (CEC) Terrestrial Ecoregions Level III classification (Wilken et al., 2011) (see Text B1). For illustration purposes, certain nearby eco-regions were combined because model behavior was similar. We also developed climate subgroups based on observed temperature and precipitation using the classification scheme outlined in Sun et al. (2022);

stations with mean winter (Nov-Mar) temperature less than -1°C are classified as “cold”, and stations with winter precipitation less than the 25th percentile of all station values are classified as “dry”. Lastly, each station was assigned a vegetation type from the USGS Land Use/Land Cover (LULC) classification system based on site photos posted on the NRCS portal.

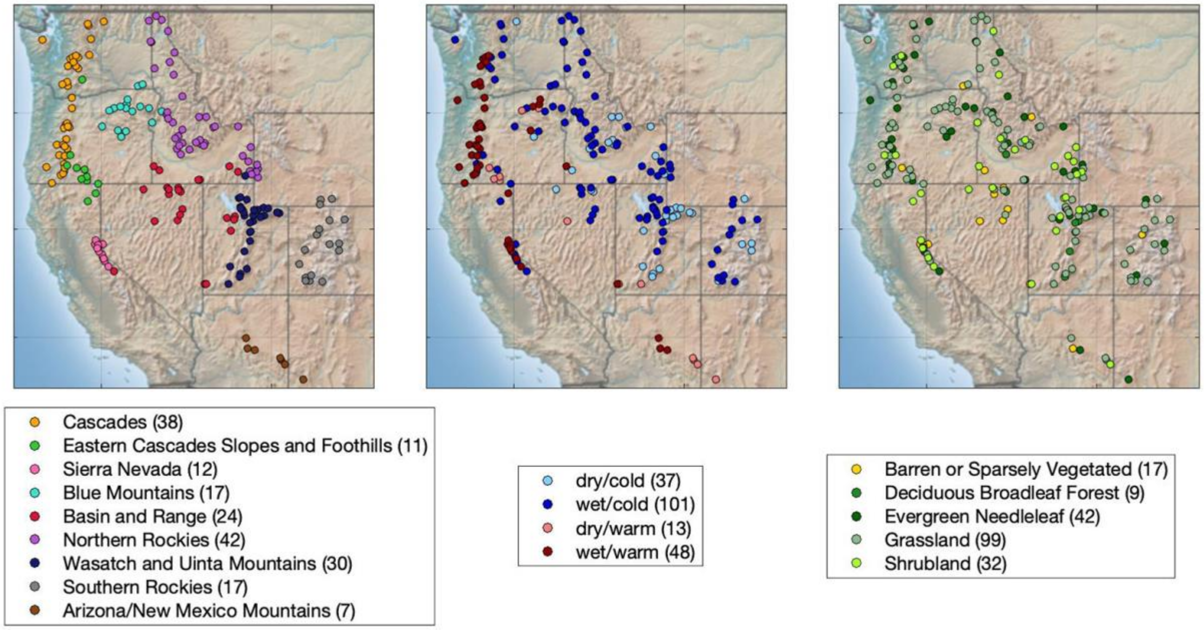


Figure 2.1. Map of 199 SNOTEL stations used, categorized by (left to right panel) geographic region, climate, and vegetation type. The number of stations in each subgroup is noted in the legend entries.

2.3.2 Meteorological forcings and observations

Forcings required by the Noah-MP model at an hourly time step are: precipitation, temperature, specific humidity, terrain-level pressure, downward longwave and shortwave radiation, and surface wind magnitude. For the base case and model experiments in which physics options were altered (see section 2.3.3.2), we assigned hourly temperature and daily precipitation inputs from SNOTEL observations for consistency with the SNOTEL SWE used to evaluate model performance. To generate records that are internally consistent, accurate, unbiased, and complete, extensive quality control, bias correction, and gap-filling procedures were applied (see Text B2). So, to the greatest extent possible, model errors in these experiments can be linked to the model physics rather than forcings biases because the forcings are both constrained by observations and consistent with the validation dataset. Daily precipitation observations were divided into even hourly values to match the model time step; this is a common approach in large-scale hydrologic modeling (for example, as is used by default in the VIC model (Variable Infiltration Capacity; Liang et al., 1994) due to challenges in obtaining reliable sub-daily records. We completed the forcings for these experiments with the Analysis of Record for Calibration dataset (AORC Version 1.1, Kitzmiller et al., 2019), because it was developed by NOAA's Office of Water Prediction (OWP) specifically for the purpose of providing gridded meteorological forcings to calibrate the NWM. It has both high spatial (~800 m) and temporal (hourly for WYs 1979 to 2019) resolutions (see Text B3 for more). We downscaled humidity and pressure from AORC values with SNOTEL temperature and elevation following methods outlined in Liston and Elder (2006) and Cosgrove et al. (2003), respectively. Longwave radiation, shortwave radiation, and wind inputs were extracted from the AORC pixel containing the SNOTEL site. Note here that our decisions to combine AORC hourly radiation forcings with SNOTEL temperature and precipitation, and to divide daily precipitation into even hourly values, will invariably introduce some inconsistencies within the meteorological forcing dataset. However, we believe this was the best approach for the purposes of this study given data limitations and model physics.

To further evaluate model performance and sensitivity relative to changes in meteorological forcings, we set up an additional experiment using solely the gridded AORC dataset, forcing each model point with values from the nearest AORC pixel.

2.3.3 Model

2.3.3.1 Overview of Noah-MP model

Noah-MP is a community model that solves the land surface energy and water balances and simulates land-atmospheric fluxes to generate a suite of hydrological and other land surface variables (Niu et al., 2011). It was developed to improve upon the limitations of the Noah LSM (Gochis et al., 2018). Snowpack is represented by up to three layers. Within each layer, snow density, temperature, and liquid water fraction are computed for each timestep, and snow accumulation and ablation processes are based on mass and energy balance. Snow cover fraction on the ground is a function of snowpack depth and density and tunable parameters (Niu and Yang 2007). Vegetation is represented by a single-layer canopy model with the capability for snow interception and unloading on a canopy snow-cover fraction (Niu et al., 2011). A semitile subgrid scheme evaluates the radiation balance, where shortwave radiation is computed over the entire grid cell using a two-stream approximation and considering canopy gap probabilities, while radiation component fluxes and albedo are computed separately over the vegetated and bare ground areas (Niu et al., 2011). Noah-MP uses multiple user-defined options for key land-atmosphere interaction processes, such that over 4,500 total combinations of physics configurations can be assessed. These configurations relate to 14 physics processes (see Table A1) such as dynamic vegetation, precipitation partitioning, soil temperature, and snow albedo

(Gochis et al., 2018). This setup allows users to customize their modeling scheme. More details about the model representation of these physics processes can be found in Text B4.

2.3.3.2 Application of Noah-MP

We applied Noah-MP as a standalone mode of WRF-Hydro v5.1.1, which is the core physics in NWM v2.1 (Cosgrove et al., 2024), over the 199 SNOTEL sites for the period WYs 2007-2019. This period covers average, dry, and wet years. Each site was treated as a single Noah-MP grid cell with a resolution of 1 km to match that of the AORC forcings and NWM 2.1. We manually adjusted the coded vegetation type at each grid cell to match the corresponding SNOTEL station based on site photos from NRCS. When no photos were available, we used the default LULC class from the WRF preprocessing system (WPS) for the corresponding grid cell. All other surface conditions and parameters, such as default snow albedo parameters, soil classification, vegetation leaf area index (LAI), and other static descriptors were defined by the WPS.

We tested ten model configurations relevant to snow processes (Table 2.1), including one base case and nine alternative experiments. The base case uses SNOTEL-based meteorological forcings, and has physics options set to match the NWM v2.1 recommended configuration. Seven physics processes were examined because of their relevance to snow simulation. You et al. (2020) found that modeled snow depth was sensitive to six of these processes at most sites. The seventh process was included because it offers a specific adjustment for snowy pixels. The tested physics processes (and named experiments) are: precipitation partitioning (Precip2.2, Precip0), snow albedo (Alb), lower soil temperature boundary condition (TempLB), snow/soil temperature time scheme (TempSolv), surface resistance to evaporation/sublimation

(ResisEvap), surface layer drag coefficient (ResisDrag), and dynamic vegetation (DynVeg) (Table 2.1). Additional details on these physics processes and their alternative options are in Text B4. The physics processes tested by the alternative model configurations impact snow accumulation and/or melt processes through various physical mechanisms (Fig. A2). For the physics processes with multiple alternative options, we tested only the recommended Noah-MP alternative (Gochis et al., 2018). Two alternatives were tested for the precipitation partitioning process due to its relevance to snow accumulation. The last experiment (AORC) uses the base case model configuration but AORC forcings. For this experiment, discrepancies between observed SWE and Noah-MP simulations can be attributed to both meteorological input and model errors.

Table 2.1. List of selected Noah-MP physics processes, with the base option listed first and alternative option(s) second. The activated set of physics options for the base case and each of the other nine model experiments are also indicated. AORC is the only experiment that shares the base case configuration of physics options. Note that three versions of the snow albedo BATS model were tested: a case with all default parameter values (Base Case), a case with the snow age parameter $\tau_0 = 3.05e6$ (BATStau_vis), and a case with $\tau_0 = 5.29e5$ (BATStau_NIR) based on optimized values from Abolafia-Rosenweig et al. (2022).

			Model Configurations									
		Options	**Base Case**	**Precip2.2***	**Precip0***	**Alb****	**TempLB****	**TempSolv****	**ResisEvap****	**ResisDrag****	**DynVeg*****	**AORC*****
Noah-MP Physics Processes	Precipitation partitioning	Jordan (1991): prescribed linear snowfall fraction when air temperature > 0.5 °C and < 2.5 °C	x			x	x	x	x	x	x	x
		Air temperature < 2.2 °C		x								
		Air temperature < 0 °C			x							
	Snow albedo	Biosphere-Atmosphere Transfer Scheme (BATS)	x	x	x		x	x	x	x	x	x
		Canadian Land Surface Scheme (CLASS)				x						
	Lower boundary condition of soil temperature	Temperature at soil lower boundary (8m) read from file	x	x	x	x		x	x	x	x	x
		Zero heat flux from bottom					x					
	Snow/soil temperature time scheme	Semi-implicit; flux top boundary condition, fractional snow cover used in calculation	x	x	x	x	x		x	x	x	x
		Semi-implicit; flux top boundary condition						x				
	Surface resistance to evaporation/ sublimation	Sakaguchi and Zeng (2009) for non-snowy pixels; parameter read from file for snowy pixels	x	x	x	x	x	x		x	x	x
		Sakaguchi and Zeng (2009)							x			
	Surface layer drag coefficient	Monin-Obukhov similarity theory	x	x	x	x	x	x	x		x	x
		Original Noah [Chen (1997)]								x		
	Dynamic vegetation	Module turned off (vegetation parameters read from file)	x	x	x	x	x	x	x	x		x
		Module turned on (prognostic vegetation growth)									x	
	Forcings	SNOTEL temperature and precipitation	x	x	x	x	x	x	x	x	x	
		AORC										x

* These model experiments are ones that primarily impact snow accumulation processes.
** These model experiments are ones that primarily impact snowmelt processes.
*** These model experiments are ones expected to impact both snow accumulation and snowmelt processes.

In addition to these user-defined physics options, Noah-MP employs parameters, some of which are designed to be tunable by the model user, related to vegetation and soil classes or model processes like runoff, albedo, and radiation balances. These parameters are not the focus of this study, but it is worth noting that particularly relevant for snow simulation are those parameters related to snow albedo (e.g., Sun et al., 2019; Abolafia-Rosenzweig et al., 2021; He et al., 2021). The Biosphere-Atmosphere Transfer Scheme (BATS), which is the default snow albedo model in Noah-MP and represents both direct and diffuse albedo over visible and NIR spectrums (Dickinson et al., 1986) in a more sophisticated way than the alternative Canadian Land Surface Scheme (CLASS) model, has 12 tunable parameters. Abolafia-Rosenzweig et al. (2022) demonstrate that the BATS model is significantly sensitive to some of these parameters, and provide parameter values that are locally optimized to stations in the Southern Rockies. Letcher et al. (2022) further find that the default parameterization of the BATS model yields albedo values that underestimate the rate of observed snow aging. So, to acknowledge the potential breadth of performance of the BATS model and the caveats of its default parameters, we further test two alternative values for the empirical snow age parameter τ_0, to which Abolafia-Rosenzweig et al. (2022) find BATS is especially sensitive. The experiment hereafter called $\text{BATS}_{\text{tau\_vis}}$ uses τ_0 = 3.05 x 10^6 to match the value optimized by Abolafia-Rosenzweig et al. (2022) for visible albedo, and the experiment $\text{BATS}_{\text{tau\_NIR}}$ sets τ_0 = 5.29 x 10^5, the value optimized for NIR albedo. These two values are respectively greater and less than the default value of $\tau_0 =$ 1 x 10^6. Although these values were optimized locally in the Southern Rockies, we apply them in a global fashion across all sites to evaluate 1) the sensitivity of the BATS albedo to this parameter and 2) how well these parameter values perform at both similar and dissimilar sites. Note that all other parameter values and physics options in these two experiments otherwise match the Base Case (Table 2.1).

2.4 Methods

We computed snow metrics to assess volume, rates, and timing of snow simulation across both the accumulation and melt seasons. These snow metrics, illustrated in Fig. 2.2 and described in more detail in section 2.4.1, are used to evaluate both model performance relative to SWE observations and sensitivity of the different model experiments (Table 2.1).

Table 2.2. Summary of snow, performance, and sensitivity metrics. Annual snow metrics are computed for both observed and modeled results at each station. Performance metrics characterize systematic bias and error magnitude, by comparing modeled (base case) to observed snow metrics at each station. Sensitivity metrics characterize bias relative to the base case and sensitivity level at each station and for each experiment.

Annual snow metrics	**Performance metrics**		**Sensitivity metrics**	
	Systematic bias	*Error magnitude*	*Systematic bias*	*Sensitivity*
Accumulated SWE (mm)*	MNB (%)	MAE (%)	MNB with respect to the base case (%)	KS statistic
Storm rate (mm day^{-1})*				
Daily melt rate (mm day^{-1})**				
Peak SWE DOWY (day)*	Difference (day)	Absolute difference (day)	Difference with respect to the base case (day)	
	FNR for accumulation days (%)*			
	FNR for melt days (%)**			

* These metrics relate to the accumulation season. ** These metrics relate to the melt season.

2.4.1 Derived annual SWE metrics

For each station-year, we computed four annual snow metrics from daily time series of observed and simulated SWE as illustrated in Fig. 2.2: accumulated SWE (mm), storm rate (mm/day), peak SWE day of water year (DOWY, day), and daily melt rate (mm/day).

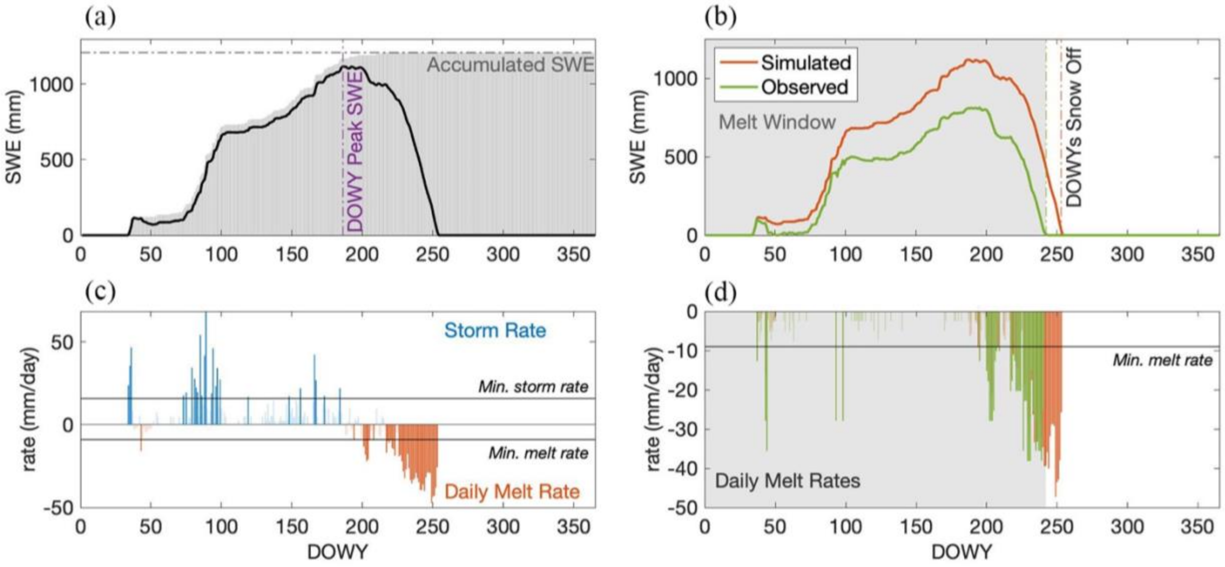


Figure 2.2. Illustration of how the four snow metrics are calculated for a sample SWE time series for one water year. (a) Peak SWE DOWY and accumulated SWE. (b) Illustration of how a melt window is calculated when comparing two SWE time series for a sample water year. Here, snow-off day is first day after peak SWE when daily snowmelt is less than 1 mm. The 1 mm threshold discounts late-season small changes in simulated SWE. (c) Storm rate and daily melt rate (averaged over the year), with a sample minimum melt and storm rates indicated. (d) Simulated and observed daily melt rates within the shaded overlapping melt window.

We calculated accumulated SWE by summing all positive daily changes in SWE (Fig. 2.2a), to represent the total snow available for melt during the water year. This integrated metric is crucial for operational hydrological models like the NWM to correctly predict water supply. We chose this metric over the traditional single-day metric of peak SWE because peak SWE underestimates total snow availability, as it misses winter melt events and post-peak accumulation (for example, Fig. A3). The day of peak SWE (Fig. 2.2a) is included to quantify the timing of spring melt onset.

We computed annual-averaged daily melt rate to independently characterize the snowmelt process. The daily melt rate is the average negative change in SWE over a melt window. The melt window spans from the water year’s first day to the snow-off day (Fig. 2.2b). Note that this window excludes late-season accumulation/melt events that happen after the winter snowpack has melted out. The SNOTEL snow pillow's precision is 2.54 mm, so daily SWE changes below the propagated error of 3.58 mm were set to zero in both simulated and observed records to maintain consistency (see more in Text A5). Only events exceeding a second minimum threshold were included to avoid biasing the average with small rates; this minimum threshold is set for each station as the rate above which 90% of all observed melt occurs. When comparing melt rates from two different time series (i.e., simulated vs. observed for the model performance analysis or simulated vs. simulated for the model sensitivity analysis), the average is applied over a melt window constrained by the earliest snow-off day (Fig. 2.2b), because high late-season melt rates in one time series (for example, Fig. 2.2d) could bias the comparison if the other time series has already melted out. This way, we are comparing melt rates over the period in which the model and measurements agree that snow is present and melting.

The storm rate represents the average rate of snow accumulation on large storm days. Large storm days are defined as days when the daily positive change in SWE exceeds the historical observed 75th percentile. Serreze et al. (2001) found that in the Western US, this top quartile represents over half of total snowfall. Because large storms contribute significantly to snow accumulation and mountain hydrology in general – for example, atmospheric river (AR) events generate 4 times more daily SWE than non-AR storms (Guan et al., 2010), it is important to evaluate the model's ability to recreate those large snowfall events. Note that because the model is run on an hourly time step but generates daily output, the storm rate could be affected by intraday melt events. For example, melt that occurs on a storm day could reduce the accumulation rate if snowmelt leaves the snowpack; this could also occur in the observational time series.

2.4.2 Evaluation of model performance

We assessed model performance across both the accumulation and melt season with six performance metrics (Table 2.2), by comparing simulated to observed SWE over 2,388 station-years of results. For each year at each site, we computed the bias as a normalized percent between simulated and observed snow metrics (M):

$$100\% * (M_{sim} - M_{obs})/M_{obs} \tag{1}$$

We then averaged these biases across all years to derive a mean normalized bias (MNB) for each snow metric and station to evaluate the systematic bias in the model. Years for which the bias magnitude was greater than 1000% (during low to no snow years) were removed as outliers before aggregating. To evaluate error magnitude, we derived a mean absolute error (MAE), expressed as a normalized percentage, by averaging yearly absolute errors for each site and snow metric:

$$100\% * |(M_{sim} - M_{obs})|/M_{obs} \tag{2}$$

The day of peak SWE is compared to its observed value by a difference in days.

Two additional performance metrics measure how accurately the model predicts the timing of melt and accumulation events. The annual false negative rate (FNR) for accumulation is the percent of days when an observed accumulation event occurs but a modeled one does not, such that:

$$FNR_{acc} = 100\% * FN/P \tag{3}$$

where FN is the number of days over the water year that shows observed accumulation without modeled accumulation, and P is the number of days with observed accumulation. A higher FNR_{acc} means that there are more days in which the model fails to simulate an observed accumulation event. FNR_{melt} similarly measures how often the model fails to predict an observed daily melt event. For both FNRs, interannual values are averaged into a single metric for each station. Note that, as opposed to the daily melt and storm rate metrics, these FNRs are not constrained by minimum melt or accumulation thresholds.

2.4.3 Evaluation of model sensitivity

We assessed sensitivity of the modeled snow metrics for each model experiment relative to the base case in two ways (Table 2.2). First, we identified where significant sensitivities exist by applying the Kolmogorov-Smirnov (KS) test (Text B6) to each snow metric at all stations in each experiment. This test has been utilized to assess model sensitivity in numerous hydrologic modelling studies (e.g., He et al., 2011; Sun et al., 2019). A two-sample KS test was performed at each station to test whether the snow metrics from the base case and the experiment come from the same distribution. KS values range from 0 to 1, with higher values indicating greater sensitivities (changes). We used a minimum KS threshold value of 0.5 to identify sensitivity because it yields statistically significant results at p-value < 0.1. Stations with a KS statistic equal to or greater than 0.5 were considered sensitive to that experiment for that snow metric.

Second, we quantified the magnitude (and direction) of changes in snow predictions by comparing each model experiment's simulated snow metrics to those from the base case by applying Equation (1), but with the base case as the reference. Changes in snow metrics for each experiment are thus expressed as the percent change in metric value relative to the base case.

2.5 Results

2.5.1 Model performance relative to SNOTEL observations

2.5.1.1 Base case performance across metrics

Overall, the Noah-MP base case overestimates accumulated SWE observations at 90% of sites, with an overall median overestimate of 9.6% (Fig. 2.3a). This overestimation most likely results from excessive snowfall input due to inaccurate precipitation partitioning. This overestimation is consistent with findings by Letcher et al. (2022) over New York State; they also identify the precipitation partitioning scheme as a source of significant sensitivity in Noah-MP. Stations with underestimated accumulated SWE have a median bias of only -1.7%. The observed storm rate is underestimated by the model in 53% of stations, which have a median bias of -3.2% (relative to the overall median of -0.4%, Fig. 2.3b). The median FNR is 8% for accumulation days, meaning that the model correctly simulates accumulation on 92% of observed days (Fig. 2.3c). Peak SWE occurs later in the base case than observations at 74% of stations, for which the median bias is 9.5 days (relative to an overall median of 5.4 days, Fig. 2.3d). Stations where the model predicts earlier peak SWE have a median bias of only -4 days. In general, model performance during the melt season is worse than in the accumulation season. The daily melt rate is underestimated by the model at 79% of stations, by an overall median bias of -15% (Fig. 2.3e). The median FNR for melt is 38%, which is over 4 times greater than that for accumulation (Fig. 2.3f). This means that the model agreement with the timing of observed melt is much worse than that for accumulation.

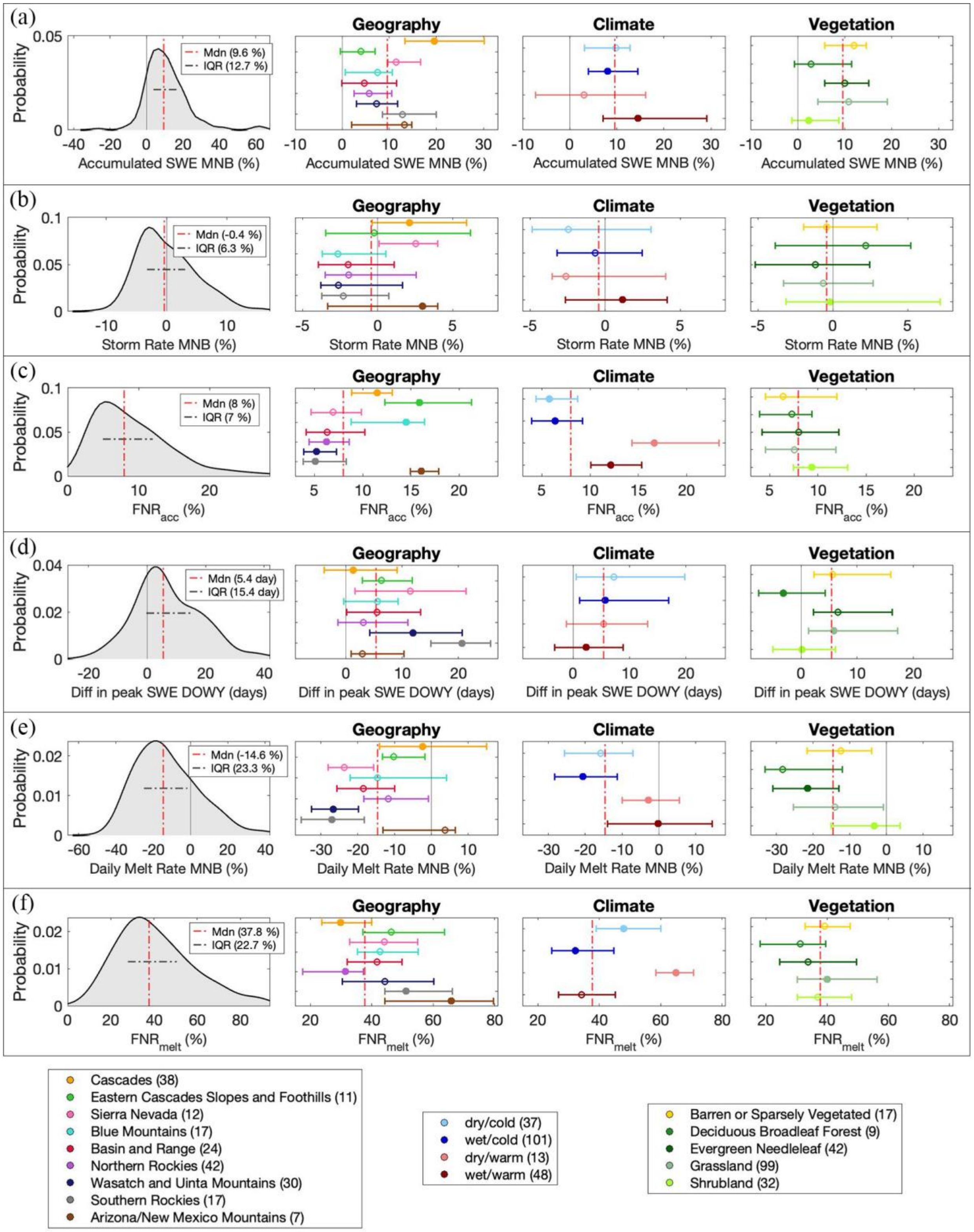

(a)
Probability
0.05
0
Mdn (9.6 %)
IQR (12.7 %)
-40 -20 0 20 40 60
Accumulated SWE MNB (%)
Geography
Climate
Vegetation
-10 0 10 20 30
Accumulated SWE MNB (%)
(b)
0.1
0.05
0
Mdn (-0.4 %)
IQR (6.3 %)
-10 0 10
Storm Rate MNB (%)
-5 0 5
Storm Rate MNB (%)
(c)
Mdn (8 %)
IQR (7 %)
0 10 20
FNR$_{acc}$ (%)
5 10 15 20
(d)
0.04
0.02
0
Mdn (5.4 day)
IQR (15.4 day)
-20 0 20 40
Diff in peak SWE DOWY (days)
0 10 20
(e)
0.02
0.01
0
Mdn (-14.6 %)
IQR (23.3 %)
-60 -40 -20 0 20 40
Daily Melt Rate MNB (%)
-30 -20 -10 0 10
(f)
Mdn (37.8 %)
IQR (22.7 %)
0 20 40 60 80
FNR$_{melt}$ (%)
20 40 60 80
● Cascades (38)
● Eastern Cascades Slopes and Foothills (11)
● Sierra Nevada (12)
● Blue Mountains (17)
● Basin and Range (24)
● Northern Rockies (42)
● Wasatch and Uinta Mountains (30)
● Southern Rockies (17)
● Arizona/New Mexico Mountains (7)
● dry/cold (37)
● wet/cold (101)
● dry/warm (13)
● wet/warm (48)
● Barren or Sparsely Vegetated (17)
● Deciduous Broadleaf Forest (9)
● Evergreen Needleleaf (42)
● Grassland (99)
● Shrubland (32)

Figure 2.3. Model performance across all stations in the base case, as compared to SNOTEL SWE observations. Six performance metrics were used: (a) mean normalized bias (MNB) in accumulated SWE; (b) MNB in storm rate; (c) FNR for accumulation days; (d) difference in timing of peak SWE; (e) MNB in daily melt rate; and (f) FNR for melt days. The leftmost column of panels shows a smoothed histogram of the performance metrics across stations in the base case. A vertical dashed red line indicates the median metric value, and a horizontal dashed gray line indicates the interquartile range (IQR). The performance metrics are separated by geographic region in the second column; by climate subgroup in the third; and by vegetation type in the fourth. Circles mark the median of the subgroup, and the width of the line marks the interquartile range. If the subgroup has a filled-in circle, it is considered significantly different (p-value < 0.05) from the other subgroups. The number of stations in each subgroup is noted in the legend entries.

Model performance varies across geographic regions. For example, stations in the Cascades region exhibit a statistically different model performance ($p < 0.05$) than other regions across all snow metrics. No stations in the Cascades nor the Sierra Nevada underestimate accumulated SWE. The Cascades has the highest median bias in accumulated SWE (19.5%, 2.5 times greater than the median of all other stations) (Fig. 2.3a), indicating lower model accuracy during the accumulation season. Of note, the Sierra Nevada, Cascades, and Arizona/New Mexico Mountains are the only regions where the median bias in storm rate is positive; suggesting that the model tends to overestimate accumulation on large storm days (Fig. 2.3b). Inaccurate precipitation partitioning could cause the model to estimate more SWE accumulation than observed on large storm days. This is notable because large storms like those produced by atmospheric rivers dominate snowpack accumulation in the Pacific mountainous regions, particularly in the Sierra Nevada (Serreze et al., 2001). Generally, stations in the interior ranges show a better performance in the accumulation season than those in the coastal ranges: four out of the five interior regions (Blue Mountains, Basin and Range, Northern Rockies, Wasatch and Unita Mountains, Southern Rockies) have a median bias below the overall median in both accumulated SWE and FNR for accumulation (Fig. 2.3a, c). For the melt season, the Cascades region has the lowest median bias in timing of peak SWE (1.3 days), the smallest median bias in daily melt (-2.4%), and the lowest FNR for melt days (30%) (Fig. 2.3d, e, f). This means that the model performs significantly more accurately over the melt season in the Cascades than in other regions. The Southern Rockies also have a statistically different melt season performance, with the largest bias in timing of peak SWE (median delay of 21 days) (Fig. 2.3d), the most negative median bias in simulated daily melt rate (–27%) (Fig. 2.3e), and the second-highest median FNR for melt days (51%) (Fig. 2.3f). This suggests that melt season model performance is the poorest in the Southern Rockies. It is worth noting that the Cascades stations are the snowiest and second warmest (average observed accumulated SWE of 796 mm and winter temperature of -0.12°C, (Table A2), and the Southern Rockies stations are the coldest (average observed winter temperature of -5.3°C, Table A2). Also relevant here is the role of light absorbing particles (i.e., dust and soot) on snow: albedo decay and shortened snow cover duration caused by dust on snow has been widely observed in the Rocky Mountains and Wasatch and Uinta Range (e.g., Painter et al., 2012, Painter et al., 2007). Noah-MP does not explicitly capture radiative forcing by dust on snow, and so can be expected to underestimate the observed melt rates in regions impacted by dust deposition. This is consistent with the result that melt rate bias is most negative in the Wasatch and Unita Mountains and Southern Rockies (Fig. 2.3e). Note that the alternative BATS experiments use a parameter value optimized for sites with considerable dust on snow effects; median melt rate bias is still negative and largest in the Wasatch and Uinta Mountains and Southern Rockies in these experiments, but the error magnitude in $\mathrm{BATS}_{\mathrm{tau\_NIR}}$ ($\mathrm{BATS}_{\mathrm{tau\_VIS}}$) decreases (increases) at these sites by a median of 14% (16%) relative to the base case (Fig. A7).

Variations in model performance can also be explained by differences in climate. Sites classified as cold (dry/cold or wet/cold) demonstrate similar bias levels in accumulated SWE regardless of precipitation amount, whereas warm sites (dry/warm or wet/warm) show more bias (on median 4.7 times higher) when they are also wet (Fig. 2.3a). This suggests that precipitation at warm sites is a more distinguishing factor for accumulated SWE bias than at cold sites. Dry sites (dry/cold or dry/warm) have a more negative bias (-2.5%) in storm rate than wet sites (-0.2%) (Fig. 2.3b), indicating that the model better predicts the storm rate when precipitation is higher (and storms are typically larger). Temperature instead is the distinguishing factor for FNR: on median, warm sites have an accumulation FNR twice as high as that for cold sites (Fig. 2.3c). So, the model predicts accumulation events better at colder stations, where partitioning thresholds are less influential in determining snowfall. In the melt season, dry/cold stations show the highest bias in timing of peak SWE, and wet/warm stations show the least (Fig. 2.3d). Temperature distinguishes model performance in daily melt rate: cold sites have a median bias 50 times greater (-20%) than warm sites (median of -0.4%) (Fig. 2.3e). FNR for melt days is instead more distinguishable by precipitation: dry stations have a median bias 1.6 times higher than wet stations (Fig. 2.3f). Both wet/warm sites and wet/cold sites perform statistically differently than other climate subgroups in five of the six metrics (Fig. 2.3), suggesting that precipitation amount strongly affects model performance.

Vegetation type also plays a role in model performance, as the amount and type of vegetation can impact both energy and mass balance of snow. Note that vegetation type is manually set in the model to match the SNOTEL station; 50% of stations are located in clearings and labeled as “grassland”. It is worth mentioning that at more vegetated sites, snow simulation becomes more complex. Although the Noah-MP canopy model correctly accounts for shadowing effects on solar radiation, it could be introducing an inconsistency between modeled SWE and observed SWE because it models canopy snow interception whereas the SNOTEL snow pillows are likely less affected by canopy interception because they are not typically placed directly under trees even when the site is forested. This could lead to an underestimation of observed SWE at forested sites in this study by Noah-MP.

Shrubland sites show statistically different model performance at the most metrics (Fig. 2.3a, b, c, d, e), signifying a unique model performance. These sites exhibit the lowest median bias magnitude in accumulated SWE (2.4%) and storm rate (-0.2%), but the highest bias in FNR for accumulation days (9%) (Fig. 2.3a, b, c). This suggests that the model simulates snow volume relatively well at shrubland sites but does so more poorly for the timing of accumulation events, perhaps because of vegetation interception. Deciduous broadleaf sites have a bias in accumulated SWE less than the overall median and are the only ones to have a negative bias in timing of peak SWE (-3 days) indicating earlier simulated snowmelt onset (Fig. 2.3d). In the melt season, the model simulates median melt rate 1.7 times more accurately at non-forested sites than forested sites (Fig. 2.3e), but FNR 1.8 times more accurately at forested sites (Fig. 2.3f). This means that at forested sites, the model simulates daily melt rate less accurately, but those melt events are more likely to occur on the correct (observed) days. It should be noted that patterns in model performance across vegetation types may be cross-correlated with climate, geography, and other factors.

2.5.1.2 Relationship between base case performance and climate variables

Studying the correlations between model error and climate-related variables underscores climate as a crucial factor affecting model performance. Figure 2.4 presents correlations (R) between observed winter temperature, winter precipitation, snow/precipitation ratio (peak

SWE/annual precipitation), and accumulated SWE, and base case model error across all sites. Model error is defined here as mean absolute error, absolute difference, or FNR (Table 2.2). By using the absolute error rather than bias, these correlations reveal the relationship between climate variables and error magnitude irrespective of error direction.

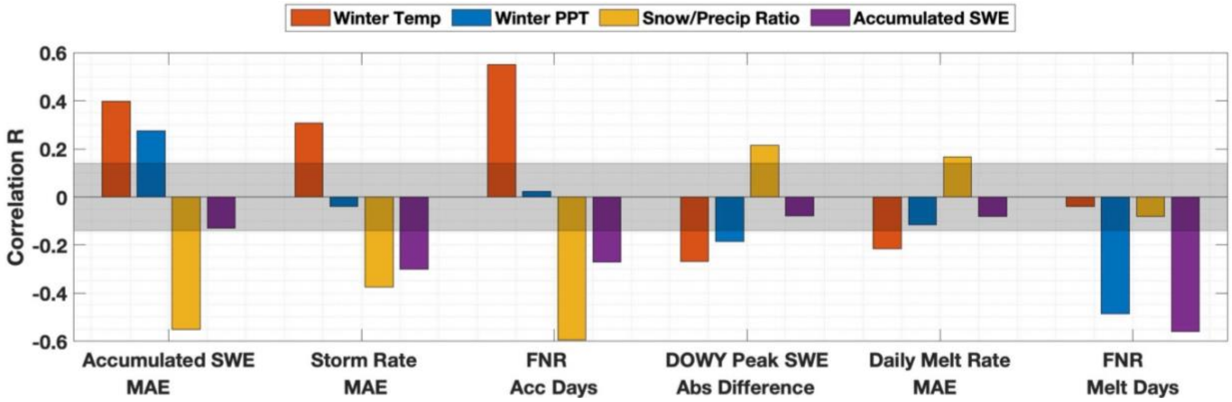


Figure 2.4. Correlations (R) between seasonal climate variables and model error across stations in the base case as represented by six performance metrics: MAE in accumulated SWE, MAE in storm rate, FNR for accumulation days, absolute difference in timing of peak SWE, MAE in daily melt rate, and FNR for melt days (Table 2.2). Climate-related variables are computed for each station-year from observed records over winters (Nov-March) of WYs 2007-2019. The gray area marks R values that are not statistically significant (p-value > 0.05). Positive (negative) bars indicate that a higher value for the climate variable correlates with higher (lower) model error.

Over the accumulation season, warmer winter temperatures correspond to higher error in accumulated SWE (R = 0.39), storm rate (R = 0.31), and FNR (R = 0.55). This supports the finding from Fig. 2.3 that warm stations perform more poorly during the accumulation season than cold stations, likely because these sites are more sensitive to inaccurate precipitation partitioning. Winter precipitation also positively correlates with model error in accumulated SWE (R = 0.28) (Fig. 2.4). This reaffirms the finding in Fig. 2.3 that wet/warm sites have the poorest model performance in accumulated SWE. Snow/precipitation ratio strongly negatively correlates with model error in accumulated SWE (R = -0.55), storm rate (R = -0.38) and FNR (R = -0.59) (Fig. 2.4). This suggests that sites with a lower snow/precipitation ratio (which tend to have warmer temperatures) have higher accumulation errors. This is consistent with the finding that observed accumulated SWE negatively correlates with error in storm rate (R = -0.3) and FNR (R = -0.27); indicating that the model is worse at predicting accumulation events at less snowy sites.

Generally, the trends between climate-related variables and model error are opposite and weaker for melt season metrics. Winter temperature negatively correlates with error in timing of peak SWE (R = -0.26) and daily melt rate (R = -0.21) (Fig. 2.4), indicating that colder stations perform worse than warm stations in the melt season. The strong negative correlation between winter precipitation and FNR for melt days (R = -0.49) (Fig. 2.4) reinforces the finding from Fig. 2.3 that wet sites perform distinctively worse than dry sites for this metric. This is also true, but less significant, for the relationship between winter precipitation and error in timing of peak SWE (R = -0.19) (Fig. 2.4). Observed accumulated SWE has a strong negative correlation with FNR for melt days (R = -0.56) (Fig. 2.4), indicating that the model more often fails to predict observed melt events at sites with less snow. Note that observed accumulated SWE is the only variable which relates negatively to model error across both accumulation and melt seasons; so less snowy sites consistently perform worse than snowier sites. The positive correlations between snow/precipitation ratio and error in timing of peak SWE (R = 0.22) and daily melt rate (R = 0.17) are less significant but demonstrate poorer model performance in the melt season at stations with a high snow/precipitation ratio (which also tend to be colder).

2.5.2 Sensitivity of snow simulations to model configurations

2.5.2.1 Identifying sensitivity to changes in model configuration

To assess whether changes in snow metrics caused by the model experiments amount to a significant model sensitivity, we applied the Kolmogorov-Smirnov (KS) test to four snow metrics (Table 2.2). Snow metrics at a station are considered significantly sensitive to a particular experiment if the KS statistic is equal to or greater than 0.5 (p-value < 0.1). Figure 2.5 summarizes how many stations are considered sensitive to each experiment and snow metric.

	Precip2.2	Precip0	Alb	TempLB	TempSolv	ResisEvap	ResisDrag	DynVeg	AORC
Accumulated SWE	0%	22%	0%	0%	0%	0%	1%	0%	64%
Storm Rate	1%	4%	0%	0%	0%	0%	0%	0%	19%
DOWY Peak SWE	1%	4%	0%	0%	0%	2%	6%	2%	51%
Daily Melt Rate	1%	8%	17%	0%	1%	10%	51%	21%	31%

Figure 2.5. The percent of sensitive stations (KS statistic ≥ 0.5, p-value < 0.1) in each experiment and across four snow metrics. Note that AORC is an experiment that tests forcings input rather than model physics.

Overall, only two experiments show significant sensitivity in accumulation metrics at over 5% of stations: AORC and Precip0 (Fig. 2.5). Precip0 lowers the temperature threshold for snowfall to 0°C, leading to a sensitivity in accumulated SWE and storm rate at 22% and 4% of stations respectively (Fig. 2.5). In contrast, the other precipitation partitioning experiment Precip2.2, which raises the partitioning threshold to 2.2°C, has at most 1% of stations with accumulation metric sensitivity (Fig. 2.5), indicating that the 0°C threshold yields significantly more deviation in results from the base case (which uses a linear increase in partitioning with temperature per Jordan 1991) than 2.2°C. The AORC experiment, which changes the input precipitation and air temperature, demonstrates the highest levels of sensitivity, with 64%, 19%, and 51% of stations sensitive to accumulated SWE, storm rate, and timing of peak SWE, respectively (Fig. 2.5).

Six experiments exhibit significant sensitivity in melt metrics at over 5% of stations: Alb, Precip0, ResisEvap, ResisDrag, DynVeg, and AORC, indicating that melt metrics are sensitive to more changes in model configuration than accumulation metrics. Alb and DynVeg both impact surface albedo, resulting in 17% and 21% of stations showing melt rate sensitivity, respectively. ResisDrag, which impacts the computation of surface heat fluxes, is the physics-related experiment with the most sensitivity in the melt season: 51% of stations have melt rate sensitivity, and 6% have sensitivity to timing of peak SWE. ResisEvap also affects surface heat fluxes, and demonstrates melt rate sensitivity at 10% of stations. The Precip0 experiment exhibits sensitivity to daily melt rate in 8% of stations; this is an indirect effect of changes to accumulation patterns. At most 1% of stations show sensitivity to any metric in TempLB and TempSolv; meaning that these model configurations do not deviate greatly from the base case. AORC shows 31% of stations with sensitivity in melt rate (Fig. 2.5).

2.5.2.2 Changes in snow metrics relative to base case across model configurations

The magnitude and direction of changes to snow predictions caused by alterations to model configuration or forcings input is evaluated by comparing four snow metrics in each experiment to base case results (Table 2.2). Fig. 2.6 illustrates the distributions of these bias metrics for each experiment.

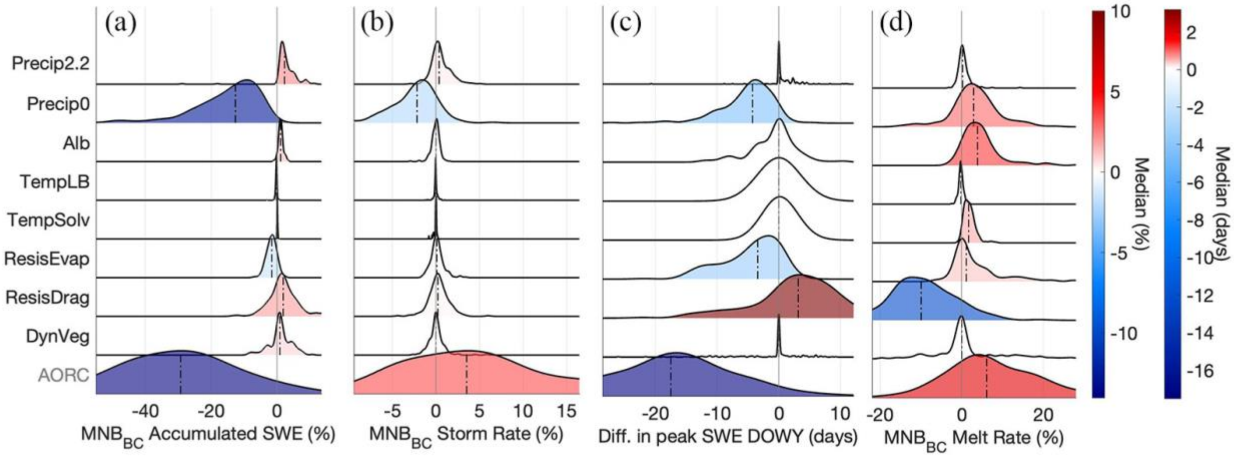


Figure 2.6. Distributions of changes in snow metrics relative to the base case across SNOTEL stations and nine experiments for four snow metrics. From left to right: (a) accumulated SWE, (b) storm rate, (c) timing of peak SWE, and (d) daily melt rate. Bias metrics (MNB_{BC} or difference, Table 2.2) are computed for each station-year with reference to the base case, and then averaged for each station over the time period. The distribution color and the dashed horizontal line correspond to the median bias value for each experiment. A red (blue) color indicates the model configuration produces a higher (lower) median value than the base case.

The precipitation partitioning experiments (Precip2.2, Precip0) are expected to primarily impact snowfall and snow accumulation. Precip2.2 increases the temperature threshold for precipitation partitioning, resulting in a slightly higher accumulated SWE (median of 2.3%) (Fig. 2.6a), but less notable impacts to storm rate (median of 0.2%) and timing of peak SWE (median of 0.2 days) (Fig. 2.6b, c). Precip0 instead lowers the temperature threshold, leading to less snowfall than the base case. The storm rate and total accumulated SWE notably reduce as a result, by a median of 4.3% and 13%, respectively (Fig. 2.6a, b). The day of peak SWE is correspondingly advanced by a median of 4.3 days (Fig. 2.6c). The range of bias for accumulated SWE in Precip0 is the widest for accumulation metrics (IQR of 12.3%), indicating that the effect of this experiment varies greatly across stations (see section 2.5.2.3) (Fig. 2.6a).

The subsequent five experiments relate primarily to melt processes. Alb generates a higher daily melt rate (median of 4%) than the base case (Fig. 2.6d). This is consistent with the observation that the alternative albedo option generates lower snow albedo (Niu et al., 2011, Abolafia-Rosenzweig et al., 2022). ResisEvap, which alters surface resistance to turbulent fluxes, generates a slightly higher melt rate (median of 1.1%) (Fig. 2.6d). This implies that the change in surface resistance was enough to reduce latent heat flux so that the additional energy at the surface initiated higher snowmelt rates. ResisDrag, which alters how the surface drag coefficient is estimated, impacts melt metrics more significantly. The alternative option (Chen 1997) has been noted to produce a lower value for the surface drag coefficient (e.g., Zhang et al., 2014), resulting in higher heat flux values. This is consistent with the experiment reducing daily melt rate, by a median of 9.9% (Fig. 2.6d). This experiment has a wide range of bias values for daily melt rate (IQR of 10%), suggesting that the impact of ResisDrag varies notably across sites (see section 2.5.2.3). TempSolv, which changes the solver approach for the soil temperature equation, increases the daily melt rate by a median of 1.7% (Fig. 2.6d). TempLB, which changes the temperature lower boundary condition of the soil column, generates insignificant to no change (less than 0.2%) to all metrics (Fig. 2.6a, b, c, d).

DynVeg and AORC are expected to impact both accumulation and melt processes. The AORC experiment uses the same physics configuration as the base case, but with different precipitation and temperature forcings. AORC total winter precipitation differs from SNOTEL observations heterogeneously across the WUS (Fig. B1a) but averages out to 10.6% less. AORC winter temperature underestimates observations by an average of 0.2°C (Fig. B1b). As a result of this and/or differences in individual storm events, the storm rate increases in the AORC experiment (median of 3.5%) (Fig. 2.6b). Peak SWE occurs a median of 17.5 days earlier, which, combined with a higher melt rate (median of 6.1%, Fig. 2.6d) and lower input precipitation (Fig.

A1a), is consistent with total accumulated SWE decreasing (median of -29.3%) (Fig. 2.6a). The range of bias values in AORC is wider than in any other experiment (IQRs ranging 11-25% across metrics), suggesting that snow simulation varies greatly across sites in this experiment and that the experiment introduces more uncertainty (Fig. 2.6). DynVeg turns on the dynamic vegetation module. As a result, accumulated SWE changes slightly (median of 1%). There is insignificant to no change in median values (less than 0.2%) for peak SWE timing, storm rate, and melt rates (Fig. 2.6a, b, d). Of note, the IQR for the melt rate biases is 5% (Fig. 2.6d), indicating that DynVeg impacts melt processes at some stations more significantly than the median suggests.

Figure 2.6 also illustrates changes in snow simulation due to indirect effects by the model experiments. For example, Precip0, which directly reduces snowfall relative to the base case, also exhibits a slight increase in daily melt rate (median of 3%) (Fig. 2.6d). This is because a shallower snowpack presumably reaches melt phase earlier. Although ResisDrag primarily impacts melt rate, it also demonstrates a slight increase in accumulated SWE (median of 1.9%) and a corresponding delay in timing of peak SWE (median of 3.2 days) (Fig. 2.6a, c). This is consistent with less winter melt occurring due to lower melt rates. The IQR for biases in timing of peak SWE in this experiment is the second widest (7.5%), indicating a varied impact that can result in either positive or negative biases at individual sites. ResisEvap shows effects similar in magnitude but opposite in direction: a slight decrease in accumulated SWE (median of -1.6%) and advance in timing of peak SWE (median of -3.4 days) (Fig. 2.6a, c). This is consistent with more winter melt occurring due to a higher melt rate.

2.5.2.3 Changes in snow metrics relative to base case across site conditions

Changes in snow metrics caused by different model configurations relative to the base case can also be evaluated across site conditions in order to understand how and why snow simulation responds differently to alterations in physics processes or forcings input. Only the metrics with the highest percent of stations showing a significant KS statistic (Fig. 2.5) are evaluated in this section.

Although the AORC experiment shows high levels of sensitivity (Fig. 2.5), the changes to snow metrics caused by this experiment relate to the localized differences between SNOTEL and AORC temperature/precipitation rather than model physics; and so are more heterogeneous across geography, climate, and vegetation than in other experiments.

Precip0 primarily impacts accumulated SWE as it changes the amount of snowfall input (Fig. 2.5). Fig. 2.7a illustrates how stations closer to the Pacific coast – in the Cascades, Eastern Cascades Slopes and Foothills, and the Sierra Nevada – show a more negative and larger change in accumulated SWE than stations located in the interior ranges – Northern Rockies, Blue Mountains, Basin and Range, Wasatch and Uinta Mountains. This is due to more temperate climate in the more coastal areas, with warmer temperatures that hover more closely to the precipitation partitioning threshold. In fact, stations classified as warm (either dry/warm or wet/warm) demonstrate a higher and more negative median bias in snow accumulation than stations classified as cold (Fig. 2.7a). Precipitation amount also matters: within each temperature grouping (warm or cold), stations classified as wet demonstrate a higher bias than their dry counterparts (Fig. 2.7a). So, Precip0 causes more change in accumulated SWE in wetter and warmer stations, because changing the precipitation partitioning threshold has a bigger impact where winter temperature is closer to that threshold.

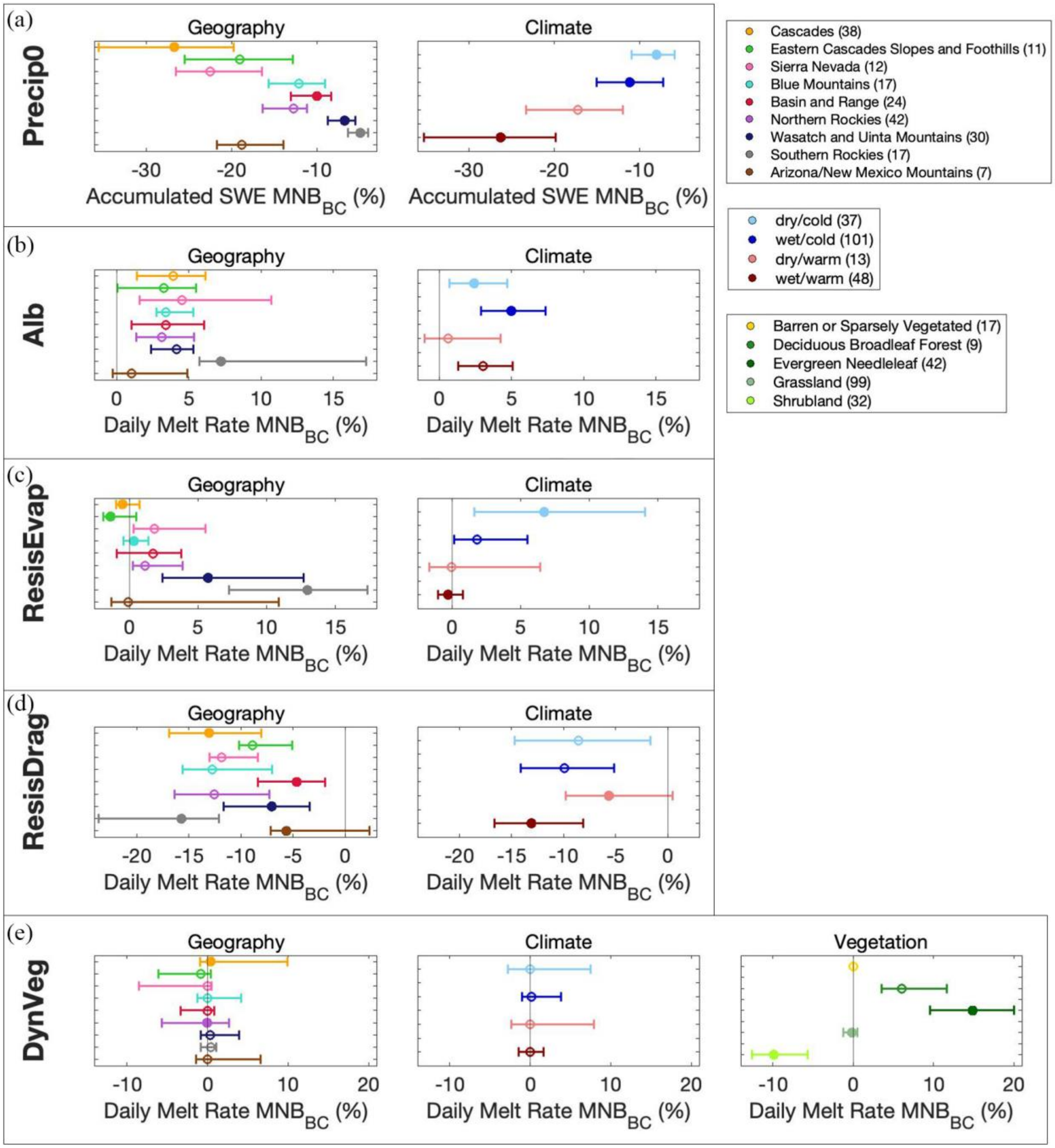


Figure 2.7. Distribution of changes from the base case in key snow metrics across site conditions in (a) Precip0, (b) Alb, (c) ResisEvap, (d) ResisDrag, and (e) DynVeg experiments. The bias metric MNB relative to base case for

accumulated SWE is shown for Precip0, and for daily melt rate for the other experiments. The first column of subpanels presents results across different geographic regions of the WUS, and the second column separates results by climate subgroups. Results are separated by vegetation type for the DynVeg experiment. Circles mark the median of subgroup, and the width of the line marks the interquartile range. If subgroup has a filled-in circle, it is considered significantly different (p-value < 0.05) from the other subgroups. The number of stations in each subgroup is noted in the legend entries.

Alb primarily impacts the daily melt rate (Fig. 2.5), and that impact varies by geography and climate (Fig. 2.7b). The only region to show a negative change in daily melt rate in this experiment is the Arizona/New Mexico mountains (Fig. 2.7b), which is the least snowy region in the study (Table B2). The region with the highest positive change in daily melt rate (and widest IQR) is the Southern Rockies (Fig. 2.7b). Both these regions are statistically different than the others. Precipitation amount rather than temperature is a distinguishing factor for this experiment: stations classified as wet (either wet/cold or wet/warm) have a higher median bias in melt rate than stations classified as dry (Fig. 2.7b). Dry/warm stations have the lowest median bias, whereas wet/cold stations have the highest (Fig. 2.7b).

ResisEvap and ResisDrag also primarily impact daily melt rate (Fig. 2.5). For ResisEvap, stations in the Eastern Cascades and Foothills demonstrate the most negative and highest absolute median bias of all regions (Fig. 2.7c). The only other region with an absolute median bias higher than 2% is the Southern Rockies, which has the highest positive bias (Fig. 2.7c). In this experiment, the IQRs of the bias metrics vary notably across regions – for example, the IQR for the dry Arizona/New Mexico region is much wider than the wet Cascades region (Fig. 2.7c). While the median biases for each climate subgroup do not differ significantly, the IQRs do. Stations classified as dry (either dry/cold or dry/warm) have a wider IQR than stations classified as wet (Fig. 2.7c). So, the changes in daily melt rate caused by ResisEvap range wider across drier stations.

Precipitation amount is also a distinguishing factor for changes to daily melt rate in ResisDrag. Wet stations have a more negative bias than dry stations, with wet/warm stations demonstrating the highest absolute bias Fig. 2.7d). Geographically, the Cascades demonstrate the most negative and highest absolute median bias, whereas the Arizona/New Mexico Mountains region show the lowest median bias but with the widest IQR (Fig. 2.7d).

For DynVeg, more stations show sensitivity to daily melt rate than any other snow metric (Fig. 2.5). Although the median biases across geographic regions and climate subgroups are similar and close to 0, there are notable differences in the IQRs (Fig. 2.7e). For example, stations in the Cascades region have the widest IQR whereas stations in the Southern Rockies have the narrowest (Fig. 2.7e). Across climate subgroups, the IQR is wider in dry stations (either dry/cold or dry/warm) than wet stations (Fig. 2.7e). These differences suggest that station groups with wider IQRs have more uncertainty in how DynVeg affects snow simulation. Changes to melt rate caused by DynVeg are notably discernable by vegetation type. There is no change from the base case daily melt rate at barren or sparsely vegetated stations, suggesting that the dynamic vegetation module (see more in Text A4) does not have any impact on snow simulation at these sites. On the other hand, there is negative change for shrubland stations, and positive change for forested stations (either deciduous broadleaf or evergreen needleleaf) (Fig. 2.7e). Grassland stations show a median bias close to 0, with individual stations showing slight changes (Fig. 2.7e).

2.5.3 Best performing physics options

The variations in model performance and sensitivity across sites speak to inadequacies in physical process representation that manifest differently under different site conditions. This implies that customizing model configurations to the unique characteristics of the model domain, particularly climate type which we have found has strong patterns with model behavior, could yield more accurate and useful results. Figure 2.8 illustrates which option – either the base case or alternative – performs best relative to observations, as measured by the MAE in accumulated SWE and daily melt rate (Table 2.2).

Overall, the Precip0 alternative for precipitation partitioning yields the best accumulation season model performance at most stations across the WUS (Fig. 2.8a). This option reduces snowfall amount relative to the base case and partly compensates for the positive bias in accumulated SWE in the base case (Fig. 2.3a). At most warm stations, which are more sensitive to changes in temperature thresholds for precipitation partitioning (Fig. 2.7a), the base option instead more accurately predicts accumulated SWE.

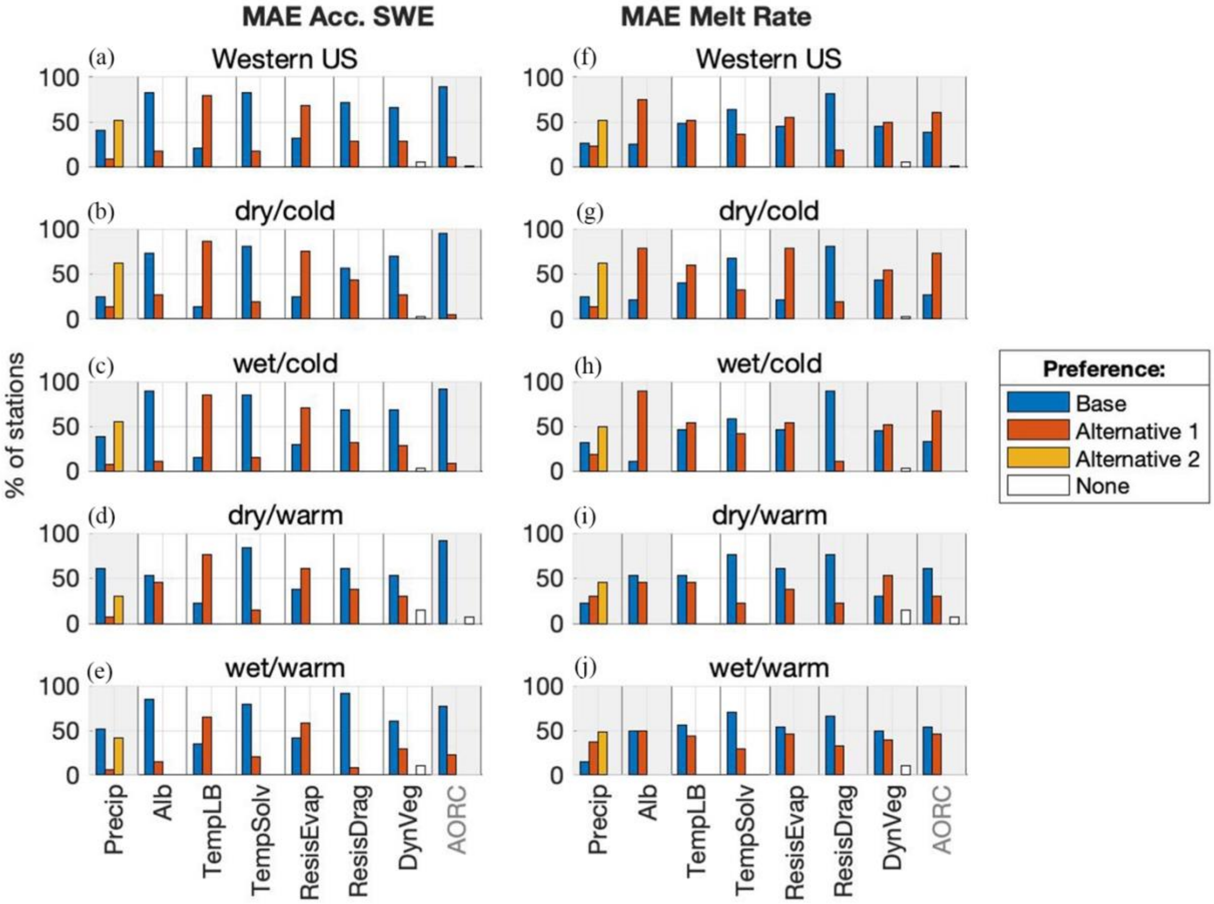


Figure 2.8. Fraction of stations showing better model performance, as represented by the lowest MAE in accumulated SWE (left column, a-e) and MAE in daily melt rate (right column, f-j), for each option under seven physics processes and one pair of forcings input, for all stations and for subsets covering four climate categories. Note that for the precipitation partitioning process, Precip2.2 is "alternative 1" and Precip0 is "alternative 2". There are 37 stations in the dry/cold subgroup; 13 in dry/warm; 101 in wet/cold, and 48 in wet/warm. If a station shows no difference in model performance between the different options, it is added to the white bar. The best performing option has the highest bar. The gray background highlights the most relevant options for that snow metric by indicating which experiment causes sensitivity relative to the base case for that snow metric at over 5% of stations (as shown on Fig. 2.5).

As expected, using observation-based forcings rather than gridded AORC forcings leads to more accurate accumulated SWE at most stations (Fig. 2.8a, b, c, d). At those few stations where AORC instead performs better, the base case model configuration can be considered erroneous because the AORC forcings and validation dataset (observed SWE) are inconsistent and thus the AORC experiment is not expected to yield superior model performance than the experiment which uses observation-based forcings.

At predicting the observed melt rate, the alternative for Alb vastly outperforms the base option at most stations across the WUS and especially at cold stations (Fig. 2.8e, f, g). This is consistent with the finding that observed melt rate is generally underestimated in the base case (Fig. 2.3e), and the alternative snow albedo option yields lower albedo values and thus a higher melt rate (Fig. 2.6d). At warm stations, where melt rates are typically higher and the change relative to the base case is less (Fig. 2.7b), the base option instead outperforms the alternative at most stations. Note that in this case, the base option utilizes all default parameters for the BATS albedo scheme. Figure 2.8 illustrates how tuning one parameter in this scheme to a locally optimized value (Abolafia-Rosenzweig et al., 2022) affects the results.

Similarly, the alternative for ResisEvap, which tends to increase the melt rate relative to the base case (Fig. 2.6d), outperforms the base option at most stations across the WUS (and in particular cold stations). Conversely, the base option outperforms the alternative for ResisDrag, which reduces the melt rate relative to the base case at shrubland and grassland sites (Fig. 2.7e), at most stations across all climate groups except for wet/warm (Fig. 2.8f, g, h, i, j). Note that the impact of the DynVeg alternative relative to the base case is more correlated with vegetation type than climate subgroup (Fig. 2.7e). The second alternative for precipitation partitioning (Precip0), which indirectly impacts melt rate by reducing accumulated snowfall, outperforms the base option across all climate groups (Fig. 2.8f, g, h, i, j). Also, using gridded forcings (AORC) yields more accurate melt rates than using observations-based forcings at most stations except for those in warm climates (Fig. 2.8f, g, h, i, j), suggesting that the base case configuration is not optimized for melt season SWE predictions. Notably, for most stations where AORC outperforms the base option, the alternative snow albedo model also outperforms the base option (Fig. 2.8f, g, h) – this suggests that a key source of error in the base case configuration is the use of an inadequate snow albedo model (in the base option, BATS with default parameter values) especially at cold stations.

Note also that the alternative option sometimes outperforms the base case even amongst the physics processes that do not yield significant changes relative to the base case. This occurs in the accumulation season at most stations for TempLB and ResisEvap, and in the melt season at most cold stations for TempLB.

The BATS snow albedo model, used in the Base Case and the default within Noah-MP, has 12 tunable parameters, some of which have been found to significantly affect the performance of the model (e.g., Abolafia-Rosenzweig et al., 2022). Figure 2.9 illustrates how tuning one of those parameters (τ_0) affects the model's ability to reproduce observed melt rate. Overall and for dry stations, the $\text{BATS}_{\text{tau\_NIR}}$ experiment outperforms the other snow albedo configurations, yielding the lowest melt rate error in 41% of all stations (Fig. 2.9a, b, c). This demonstrates that tuning even one parameter of the BATS model, such as to locally optimized values, can improve the model performance such that it exceeds that of an alternative snow albedo scheme. Specifically, decreasing the τ_0 parameter value (as in $\text{BATS}_{\text{tau\_NIR}}$) improves the ability of the BATS model to reproduce observed melt rate at most stations. See Fig. A6 for distributions of melt rate bias in $\text{BATS}_{\text{tau\_NIR}}$ and $\text{BATS}_{\text{tau\_vis}}$. Conversely, Alb (which uses the CLASS albedo scheme) outperforms the BATS experiments at most wet stations (Fig. 2.9d, e). Note that the adjusted values for the τ_0 parameter were optimized to local observations in the

Southern Rockies (Abolafia-Rosenzweig et al., 2022). The results summarized in Fig. 2.9 suggest that these locally optimized values can effectively improve model performance at other sites with similar climates (i.e., dry/cold, Fig. 2.9b), but less effectively at dissimilar sites (i.e., wet/warm, Fig. 2.9e).

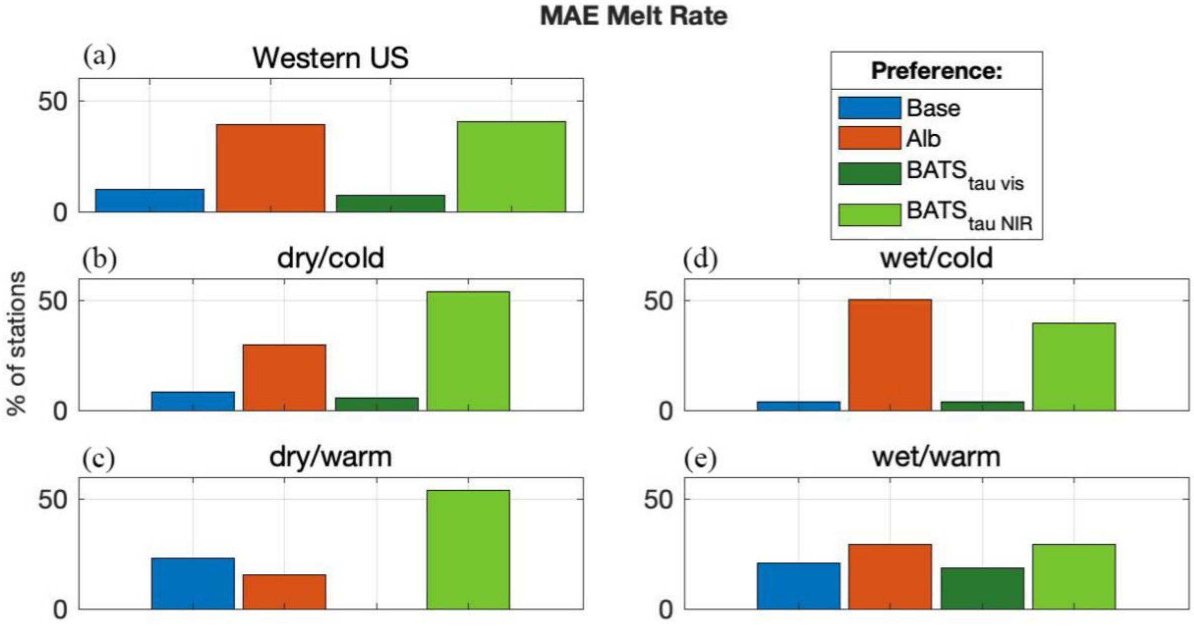


Figure 2.9. Fraction of stations showing better model performance, as represented by the lowest MAE in daily melt rate, for four snow albedo experiments: Base Case (BATS with default parameters), Alb (CLASS), BATStau_vis (BATS with tuned parameter), and BATStau_NIR (BATS with tuned parameter) (Table 2.1), for all stations (a) and for subsets covering four climate categories: (b) dry/cold, (c) dry/warm, (d) wet/cold, and (e) wet/warm. The best performing option has the highest bar.

2.6 Conclusions

This study examines snow water equivalent (SWE) simulation by Noah-MP across 199 sites in the Western US. The base case and eight model configuration experiments test physics options related to precipitation partitioning (Precip2.2, Precip0), snow albedo (Alb), lower soil temperature boundary condition (TempLB), snow/soil temperature time scheme (TempSolv), surface resistance to evaporation/sublimation (ResisEvap), surface layer drag coefficient (ResisDrag), and dynamic vegetation (DynVeg). These experiments are forced by in situ meteorology in order to rigorously identify model deficiencies. The ninth experiment (AORC) tests an alternative source of forcings to provide insight into forcing errors and relative model uncertainties.

With respect to how well Noah-MP can reproduce SWE at sites across the WUS, we find that the base case, which matches the National Water Model (NWM) configuration and uses observed temperature and precipitation forcings, overestimates observed accumulated SWE at 90% of stations by median of 9.6%. Inaccurate precipitation partitioning by the model could explain these errors. At most sites, it also predicts later peak SWE timing (5.4 days) and underestimates daily melt rate (median of -14.6%). The use of globally-applied default parameters in the base case albedo model (BATS) could explain this underestimation. The model more successfully predicts the timing of observed accumulation events than observed melt events.

With respect to the sensitivity of the model to alternative configurations, we find that Precip0 and AORC demonstrate significant sensitivity in the accumulation season, while Alb, Precip0, ResisEvap, ResisDrag, DynVeg, and AORC exhibit significant melt season sensitivity. Precip2.2, TempSolv, and TempLB show little to no significant sensitivity. Of the model configurations that test physics processes, the greatest change to accumulation season predictions occurs in Precip0 (median of -13%). On average, Precip0 reduces accumulated SWE, storm rate, and timing of peak SWE, while Precip2.2 slightly increases accumulated SWE. With regards to melt season predictions, the greatest change occurs in ResisDrag (median of -10%). ResisDrag on average reduces melt rate, increases accumulated SWE, and delays peak SWE timing, while ResisEvap does the opposite. DynVeg demonstrates a wide range of changes to melt rate, which are discernable by vegetation type but average out to a near-zero effect overall. TempLB and TempSolv have minimal effects. AORC causes substantial changes in snow performance metrics that often exceed those caused by alterations in model physics in both magnitude and range (medians ranging -29% to 6%). This suggests that previous assessments of model performance in Noah-MP or other land surface models (LSMs) could be masked by forcings errors or inconsistencies with the validation dataset, and emphasizes that generating accurate meteorological forcings should be prioritized over adjusting model physics representation for promoting more accurate SWE simulation in LSMs.

We also find that the model's performance relative to observations and sensitivity relative to the base case differs across regions, climates, and vegetation types, with the strongest trends related to climate. The model performs better relative to observations in the accumulation season at colder sites with a higher snow/precipitation ratio (such as in the Southern Rockies), and in the melt season at warmer stations with a lower snow/precipitation ratio (such as in the Cascades). The model performs most uniquely in the Cascades region – perhaps because these sites are warmer despite their deep snowpack. We also find that not all stations share the same best-performing model configuration, highlighting inconsistencies in how the model simulates SWE and suggesting the need for models customized to site conditions. Notably, at most stations and especially those classified as cold, the Precip0 alternative outperforms the base option for predicting accumulated SWE; and the Alb, ResisEvap, and DynVeg alternatives outperform the base option for predicting melt rate. We find that tuning just one parameter of the BATS snow albedo model yields lower melt rate errors at most stations. These findings indicate that the current default NWM configuration is not optimized for all sites of the Western US.

Further research should explore how the snow model responds to superposing changes in model configurations, and consider the implications of using varying configurations for different site conditions such as climate when running large-scale models like the National Water Model.

2.7 References

Abolafia-Rosenzweig, R., He, C., Burns, S. P., & Chen, F. (2021). Implementation and evaluation of a unified turbulence parameterization throughout the canopy and roughness sublayer in Noah-MP snow simulations. *Journal of Advances in Modeling Earth Systems, 13*(11), e2021MS002665. https://doi.org/10.1029/2021MS002665

Abolafia-Rosenzweig, R., He, C., McKenzie Skiles, S., Chen, F., & Gochis, D. (2022). Evaluation and optimization of snow albedo scheme in Noah-MP land surface model using in situ spectral observations in the Colorado Rockies. *Journal of Advances in Modeling Earth Systems, 14*(10). https://doi.org/10.1029/2022MS003141

Barlage, M., Chen, F., Tewari, M., Ikeda, K., Gochis, D., Dudhia, J., Rasmussen, R., Livneh, B., Ek, M., & Mitchell, K. (2010). Noah land surface model modifications to improve snowpack prediction in the Colorado Rocky Mountains. *Journal of Geophysical Research, 115*(D22), D22101. https://doi.org/10.1029/2009JD013470

Chen, F., Janjić, Z., & Mitchell, K. (1997). Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP mesoscale Eta model. *Boundary-Layer Meteorology*, *85*(3), 391–421. https://doi.org/10.1023/A:1000531001463

Cho, E., Vuyovich, C. M., Kumar, S. V., Wrzesien, M. L., Kim, R. S., & Jacobs, J. M. (2022). Precipitation biases and snow physics limitations drive the uncertainties in macroscale modeled snow water equivalent. *Hydrology and Earth System Sciences*, *26*(22), 5721–5735. https://doi.org/10.5194/hess-26-5721-2022

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Marshall, C., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., & Meng, J. (2003). Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical Research: Atmospheres*, *108*(D22), 2002JD003118. https://doi.org/10.1029/2002JD003118

Cosgrove, B., Gochis, D., Flowers, T., Dugger, A., Ogden, F., Graziano, T., Clark, E., Cabell, R., Casiday, N., Cui, Z., Eicher, K., Fall, G., Feng, X., Fitzgerald, K., Frazier, N., George, C., Gibbs, R., Hernandez, L., Johnson, D., & Zhang, Y. (2024). NOAA's National Water Model: Advancing operational hydrology through continental-scale modeling. *JAWRA Journal of the American Water Resources Association*. https://doi.org/10.1111/1752-1688.13184

Dickinson, R. E., Henderson-Sellers, A., & Kennedy, P. J. (1986). *Biosphere-atmosphere transfer scheme (BATS) for the NCAR community climate model* (NCAR Tech. Note NCAR/TN-38+STR, Vol. 82). National Center for Atmospheric Research.

Dozier, J., Bair, E. H., & Davis, R. E. (2016). Estimating the spatial distribution of snow water equivalent in the world's mountains: Spatial distribution of snow in the mountains. *Wiley Interdisciplinary Reviews: Water, 3*(3), 461–474. https://doi.org/10.1002/wat2.1140

Garousi-Nejad, I., & Tarboton, D. G. (2022). A comparison of National Water Model retrospective analysis snow outputs at snow telemetry sites across the Western United States. *Hydrological Processes, 36*(1), e14469. https://doi.org/10.1002/hyp.14469

Gochis, D. J., Barlage, M., Dugger, A., FitzGerald, K., Karsten, L., McAllister, M., McCreight, J., Mills, J., RafieeiNasab, A., Read, L., Sampson, K., Yates, D., & Yu, W. (2018). The WRF-Hydro modeling system technical description (Version 5.0). NCAR Technical Note. https://doi.org/10.5065/D6J38RBJ

Guan, B., Molotch, N. P., Waliser, D. E., Fetzer, E. J., & Neiman, P. J. (2010). Extreme snowfall events linked to atmospheric rivers and surface air temperature via satellite measurements. *Geophysical Research Letters, 37*(20). https://doi.org/10.1029/2010GL044696

He, C., Chen, F., Abolafia-Rosenzweig, R., Ikeda, K., Liu, C., & Rasmussen, R. (2021). What causes the unobserved early-spring snowpack ablation in convection-permitting WRF modeling over Utah mountains? *Journal of Geophysical Research: Atmospheres, 126*(22), e2021JD035284. https://doi.org/10.1029/2021JD035284

He, M., Hogue, T. S., Franz, K. J., Margulis, S. A., & Vrugt, J. A. (2011). Characterizing parameter sensitivity and uncertainty for a snow model across hydroclimatic regimes. *Advances in Water Resources, 34*(1), 114–127. https://doi.org/10.1016/j.advwatres.2010.10.002

He, M., Russo, M., & Anderson, M. (2016). Predictability of seasonal streamflow in a changing climate in the Sierra Nevada. *Climate, 4*(4), 57. https://doi.org/10.3390/cli4040057

Kim, R. S., Kumar, S., Vuyovich, C., Houser, P., Lundquist, J., Mudryk, L., Durand, M., Barros, A., Kim, E. J., Forman, B. A., Gutmann, E. D., Wrzesien, M. L., Garnaud, C., Sandells, M., Marshall, H.-P., Cristea, N., Pflug, J. M., Johnston, J., Cao, Y., & Wang, S. (2021). Snow Ensemble Uncertainty Project (SEUP): Quantification of snow water equivalent uncertainty across North America via ensemble land surface modeling. *The Cryosphere, 15*(2), 771–791. https://doi.org/10.5194/tc-15-771-2021

Jordan, R. E. (1991). A one-dimensional temperature model for a snow cover: Technical documentation for SNTHERM, 89. https://erdc-library.erdc.dren.mil/jspui/bitstream/11681/11677/1/SR-91-16.pdf

Kitzmiller, D. H., Wu, W., Zhang, Z., Patrick, N., & Tan, X. (2018). The Analysis of Record for Calibration: A high-resolution precipitation and surface weather dataset for the United States. *AGU Fall Meeting Abstracts*, *2018*, H41H-06.

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., & Wood, E. F. (2015). Inroads of remote sensing into hydrologic science during the WRR era. *Water Resources Research, 51*(9), 7309–7342. https://doi.org/10.1002/2015WR017616

Letcher, T., Minder, J., & Naple, P. (2022). Understanding and improving snow processes in Noah-MP over the Northeast United States via the New York State Mesonet. Engineer Research and Development Center (U.S.). https://doi.org/10.21079/11681/45060

Li, Q., Yang, T., & Li, L. (2022a). Quantitative assessment of the parameterization sensitivity of the WRF/Noah-MP model of snow dynamics in the Tianshan Mountains, Central Asia. *Atmospheric Research*, *277*, 106310. https://doi.org/10.1016/j.atmosres.2022.106310

Li, J., Miao, C., Zhang, G., Fang, Y., Shangguan, W., & Niu, G. (2022b). Global evaluation of the Noah-MP land surface model and suggestions for selecting parameterization schemes. *Journal of Geophysical Research: Atmospheres*, *127*(5). https://doi.org/10.1029/2021JD035753

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

Liston, G. E., & Elder, K. (2006). A meteorological distribution system for high-resolution terrestrial modeling (MicroMet). *Journal of Hydrometeorology*, *7*(2), 217–234. https://doi.org/10.1175/JHM486.1

National Oceanic and Atmospheric Administration (NOAA), National Weather Service (NWS), Office of Water Prediction (OWP), Silver Spring, MD. (2020). *Analysis of Record for Calibration Version 1.1 - Sources, methods, and verification*. Retrieved from https://hydrology.nws.noaa.gov/aorc-historic/Documents/AORC-Version1.1-SourcesMethodsandVerifications.pdf

Niu, G.-Y., & Yang, Z.-L. (2007). An observation-based formulation of snow cover fraction and its evaluation over large North American river basins. *Journal of Geophysical Research*, *112*(D21101). https://doi.org/10.1029/2007JD008674

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., & Xia, Y. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research, 116*(D12). https://doi.org/10.1029/2010JD015139

Painter, T. H., Barrett, A. P., Landry, C. C., Neff, J. C., Cassidy, M. P., Lawrence, C. R., McBride, K. E., & Farmer, G. L. (2007). Impact of disturbed desert soils on duration of mountain snow cover. *Geophysical Research Letters*, *34*(12), 2007GL030284. https://doi.org/10.1029/2007GL030284

Painter, T. H., Skiles, S. M., Deems, J. S., Bryant, A. C., & Landry, C. C. (2012). Dust radiative forcing in snow of the Upper Colorado River Basin: 1. A 6-year record of energy balance, radiation, and dust concentrations. *Water Resources Research*, *48*(7), 2012WR011985. https://doi.org/10.1029/2012WR011985

Pan, M., Sheffield, J., Wood, E., Mitchell, K., Houser, P., Schaake, J., Robock, A., Lohmann, D., Cosgrove, B., Duan, Q., Luo, L., Higgins, R., Pinker, R., & Tarpley, J. (2003). Snow process modeling in the North American Land Data Assimilation System (NLDAS): 2. Evaluation of model simulated snow water equivalent. *Journal of Geophysical Research: Atmospheres*, *108*(D22). https://doi.org/10.1029/2003JD003994

Sakaguchi, K., & Zeng, X. (2009). Effects of soil wetness, plant litter, and under-canopy atmospheric stability on ground evaporation in the Community Land Model (CLM3.5): New schemes for CLM3.5 soil evaporation. *Journal of Geophysical Research: Atmospheres*, *114*(D1). https://doi.org/10.1029/2008JD010834

Serreze, M. C., Clark, M. P., & Frei, A. (2001). Characteristics of large snowfall events in the montane western United States as examined using snowpack telemetry (SNOTEL) data. *Water Resources Research, 37*(3), 675–688. https://doi.org/10.1029/2000WR900307

Sun, N., Yan, H., Wigmosta, M. S., Leung, L. R., Skaggs, R., & Hou, Z. (2019). Regional snow parameters estimation for large-domain hydrological applications in the western United States. *Journal of Geophysical Research: Atmospheres*, *124*(10), 5296–5313. https://doi.org/10.1029/2018JD030140

Sun, N., Yan, H., Wigmosta, M. S., Lundquist, J., Dickerson-Lange, S., & Zhou, T. (2022). Forest canopy density effects on snowpack across the climate gradients of the western United States mountain ranges. *Water Resources Research*, *58*(1). https://doi.org/10.1029/2020WR029194

Tanaka, S. K., Zhu, T., Lund, J. R., Howitt, R. E., Jenkins, M. W., Pulido, M. A., Tauber, M., Ritzema, R. S., & Ferreira, I. C. (2006). Climate warming and water management adaptation for California. *Climatic Change*, *76*(3–4), 361–387. https://doi.org/10.1007/s10584-006-9079-5

Vicuña, S., Garreaud, R. D., & McPhee, J. (2011). Climate change impacts on the hydrology of a snowmelt-driven basin in semiarid Chile. *Climatic Change*, *105*(3–4), 469–488. https://doi.org/10.1007/s10584-010-9888-4

Wang, Y., Broxton, P., Fang, Y., Behrangi, A., Barlage, M., Zeng, X., & Niu, G. (2019). A wet-bulb temperature-based rain-snow partitioning scheme improves snowpack prediction over the drier western United States. *Geophysical Research Letters*, *46*(23), 13825–13835. https://doi.org/10.1029/2019GL085722

Wiken, E., Jiménez Nava, F., & Griffith, G. (2011). *North American terrestrial ecoregions—Level III*. Commission for Environmental Cooperation.

Yan, H., Sun, N., Wigmosta, M. S., Skaggs, R., Hou, Z., & Leung, R. (2018). Next-generation intensity-duration-frequency curves for hydrologic design in snow-dominated environments. *Water Resources Research*, *54*(2), 1093–1108. https://doi.org/10.1002/2017WR021290

You, Y., Huang, C., Yang, Z., Zhang, Y., Bai, Y., & Gu, J. (2020). Assessing Noah-MP parameterization sensitivity and uncertainty interval across snow climates. *Journal of Geophysical Research: Atmospheres*, *125*(4). https://doi.org/10.1029/2019JD030417

Zhang, G., Zhou, G., Chen, F., Barlage, M., & Xue, L. (2014). A trial to improve surface heat exchange simulation through sensitivity experiments over a desert steppe site. *Journal of Hydrometeorology*, *15*, 664–684. https://doi.org/10.1175/JHM-D-13-0113.1

Zhang, G., Chen, F., & Gan, Y. (2016). Assessing uncertainties in the Noah-MP ensemble simulations of a cropland site during the Tibet Joint International Cooperation program field campaign: Uncertainty in Noah-MP simulations. *Journal of Geophysical Research: Atmospheres*, *121*(16), 9576–9596. https://doi.org/10.1002/2016JD024928

CHAPTER 3

Improved modelling of mountain snowpacks with spatially distributed precipitation bias correction derived from historical reanalysis

The text and figures from this chapter are under peer review:

von Kaenel, M., & Margulis, S. (2024). Improved modelling of mountain snowpacks with spatially distributed precipitation bias correction derived from historical reanalysis. *EGUsphere* [preprint]. https://doi.org/10.5194/egusphere-2024-3389

3.0 Abstract

Accurate estimates of snow water equivalent (SWE) are essential for effective water management in regions dependent on seasonal snowmelt. However, significant biases and high uncertainty in mountain precipitation data products pose significant challenges. This study leverages a SWE reanalysis framework and historical dataset to derive factors that can downscale and bias-correct mountain precipitation in a real-time modelling context. We evaluate through hindcast modelling how different versions of this precipitation bias correction affect errors in 1 April SWE estimates within a representative snow-dominated watershed in the Western U.S. We also evaluate how the additional assimilation of fractional snow-covered area (fSCA) or snow depth observations during the accumulation season impact the 1 April SWE estimates. Results show that spatially distributed historically informed precipitation bias correction significantly improves SWE estimates, reducing the normalized root mean square difference (NRMSD) by 58%, increasing the correlation (R) by 43%, and decreasing mean difference (MD) by 88%. The primary strength of this bias correction method lies in capturing the spatial distribution of precipitation bias rather than its interannual variability. Assimilating snow depth observations further reduces errors both at the watershed scale (NRMSD less by 46%) and pixel level in most years, while accumulation season fSCA assimilation is not generally useful. We demonstrate the value of these methods for streamflow forecasts: bias-corrected precipitation improves the correlation between daily simulated snowmelt and observed streamflow by 31-39% and reduces bias in predicted April-July runoff volumes by 46-52%. This study highlights how historical SWE reanalysis datasets can be leveraged and applied in a real-time context by informing precipitation bias correction.

3.1 Significance statement

Accurate snow water equivalent (SWE) estimates are crucial for water management in snowmelt-dependent regions, but bias and uncertainty in precipitation data make this challenging. Here, we leverage insights from a historical SWE data product to correct these biases and yield more accurate SWE estimates and streamflow predictions. Incorporating snow depth observations further boosts accuracy. This study demonstrates an effective method to downscale and bias-correct global mountain precipitation.

3.2 Introduction

Seasonal snowpack is a natural water tower; by storing winter precipitation and releasing it as snowmelt, it provides an essential resource for downstream ecosystems and an estimated 20% of the Earth’s population (Dozier, 2011). In order to make critical management decisions for flood control, hydropower operations, irrigation, and other competing demands in snow-dependent regions of the world, water managers need accurate assessments of the distribution and availability of water in snowpack (e.g., Hamlet et al., 2002; Koster et al., 2010; He et al., 2016). Estimating the spatiotemporal distribution and change of snow water equivalent (SWE) remains a significant and important challenge for the snow hydrology community (e.g., Cho et al., 2022; Dozier et al., 2016; Lettenmaier et al., 2015).

Large-scale and temporally continuous SWE (snow water equivalent) measurements are generally absent from the real-time observational record. In situ data from networks such as the Western U.S. SNOTEL (snow telemetry) network are not always representative of the heterogeneity of SWE distribution in topographically complex mountain landscapes (e.g., Herbert et al., 2024), and such networks are sparse globally. For example, although seasonal snowpack is crucial to local water availability in High Mountain Asia, the region has almost no in situ data (e.g., Liu et al., 2021). Remote sensing can provide measurements of snow properties like fractional snow-covered area (fSCA, e.g., Selkowitz et al., 2017), snow depth (e.g., Painter et al., 2016), or albedo (e.g., Bair et al., 2019) over large areas, but there is currently no reliable way of measuring SWE from spaceborne platforms (Lettenmaier et al., 2015).

This implies a continued need for modelling of mountain SWE. Land surface models are commonly used to estimate SWE and other hydrologic variables over large spatial extents (Cho et al., 2022; Clark et al., 2011; Kumar et al., 2013), but these are susceptible to uncertainties driven by biases in forcing data or model parameterization (Cho et al., 2022). Uncertainty in precipitation products in mountainous terrain, and its implications for SWE and downstream hydrology modelling, is a widely acknowledged challenge (e.g., Schreiner-McGraw et al., 2020;

Cho et al., 2022; Pan et al., 2003; Raleigh et al., 2015; Liu and Margulis, 2019). Fang et al., 2023 found that the uncertainty of SWE estimates from commonly used global and regional modelling products is primarily explained by precipitation uncertainty. Data assimilation has gained popularity as a way to constrain or correct uncertain model estimates of snow with observations of variables such as fSCA or snow depth, and has demonstrated ability as a method to quantify SWE over both melt and accumulation seasons (Margulis et al., 2016; Cortes et al., 2016; Liu et al., 2021; Fang et al., 2022). This approach is particularly valuable in regions where in situ data are sparse but remotely sensed observations like fSCA are available, such as High Mountain Asia (Liu et al., 2021) or the South American Andes (Cortes et al., 2016). However, such products are typically only generated retrospectively. Recent studies have shown promise in combining historical reanalysis snow estimates with in situ and/or remotely sensed snow observations using statistical methods to specifically develop near real-time SWE estimates (Pflug et al., 2022; Schneider and Molotch, 2016; Bair et al., 2018; Zheng et al., 2018; Yang et al., 2022). While these methods still heavily rely on ground SWE observations, they do demonstrate the value of and potential for historical reanalysis SWE datasets to inform SWE estimation in an operational context.

Real-time spatially distributed SWE estimates have significant potential for application to water management. Climate change impacts in snow-influenced systems, such as earlier runoff of snowmelt and drops in snowpack volume, pose important challenges for water managers (e.g., Berg and Hall, 2017). Accurate and timely seasonal streamflow forecasts help inform management decisions that allocate resources in a way that is resilient to climate variability or drought (e.g., Tanaka et al., 2006). Ensemble streamflow prediction (ESP) uses hydrologic models to forecast future streamflow from current snow, soil moisture, river and reservoir conditions (Wood et al., 2002). The skill of these model-based streamflow forecasts is primarily derived from initial SWE and soil moisture conditions (Koster et al., 2010). This suggests that accurate spatially continuous real-time SWE estimates could be used to reduce uncertainty and error in streamflow forecasts in snow-dominated regions.

In this paper we leverage a SWE reanalysis framework and historical dataset to derive mountain precipitation bias correction estimates, and develop and test spatially continuous SWE estimates on 1 April. The motivating questions are: 1) To what extent can historically informed mountain precipitation bias correction improve model-based spatial SWE estimates? 2) How does the assimilation of accumulation season fSCA and snow depth measurements into this framework affect those estimates? 3) How are snowmelt-driven streamflow predictions affected under these scenarios? We validate these methods over a well-documented study domain, with the potential to extend to areas that have less access to in situ data.

3.3 Methods

3.3.1 Study domain

The study domain comprises the Hetch Hetchy watershed, a headwater catchment for the Tuolumne River in the California Sierra Nevada (Fig. 3.1). Its drainage area (~1,200 km^2) is characterized by complex topography with elevations ranging from 1,150 m to 3,850 m. It is representative of other snow-dominated catchments that provide key water supply in the Sierra Nevada. More broadly, it is a demonstrative basin that represents global mountain watersheds where the tested methods could provide utility for water management purposes; that is, basins with complex terrain at high elevations and seasonal snowpack that plays a significant role in the water budget.

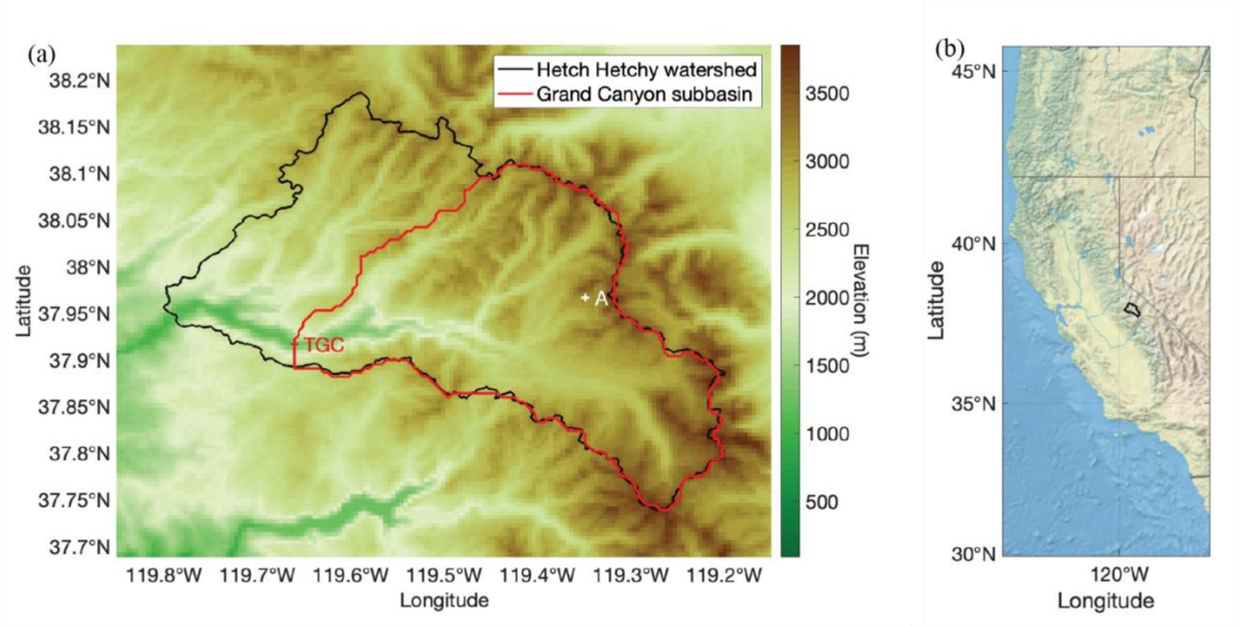


Figure 3.1. (a) A map outlining the Hetch Hetchy and the Grand Canyon subwatersheds of the Tuolumne River. The locations of the TGC streamflow gauge and sample model pixel A. (b) Outline of the Hetch Hetchy watershed illustrating its location in the central California Sierra Nevada.

The Hetch Hetchy reservoir at the watershed's outlet provides water supply for about 2.7 million residents of the San Francisco Bay Area, primarily from snowmelt. This watershed also includes a unique Airborne Snow Observatory (ASO) snow depth dataset (Painter et al., 2016) which provides multitemporal lidar-derived snow depth measurements per year. A subwatershed that drains through the USGS TGC (Tuolumne River at Grand Canyon) gauge located at the inlet of the reservoir was delineated for the streamflow analysis.

3.3.2 Overview of SWE reanalysis framework

A Bayesian reanalysis framework (Margulis et al., 2015) is used in this study for both the development of a historical reference SWE dataset, the derivation of historical precipitation bias correction through retrospective analysis, and testing "real-time" applications using that historical data (along with other data). Note that we use a hindcasting approach to model the real-time applications. Details of the application of this framework in a historical context is in section 3.3.3 and in a real-time context in section 3.3.4; here, we provide a general overview of the method and its previous applications.

Typically applied retrospectively, this reanalysis framework generates spatiotemporally continuous SWE estimates using a particle batch smoother (PBS) data assimilation technique that constrains a prior ensemble of modeled snow estimates with independent observations (most commonly, satellite-based fSCA measurements). The method was developed by Margulis et al. (2015) and has since demonstrated ability to reproduce observed SWE across global mountain regions: Sierra Nevada (Margulis et al., 2016), South American Andes (Cortes et al., 2016), High Mountain Asia (Liu et al., 2021) and the Western U.S. (Fang et al., 2022). It has also demonstrated success in assimilating remotely-sensed snow depth measurements for SWE estimation (Margulis et al., 2019).

First, an ensemble of prior snow estimates is generated using a forward land surface model; here, the modelling core is the SSiB-SAST LSM (Sun and Xue, 2001; Xue et al., 2003) paired with the Liston Snow Depletion Curve (Liston 2004). This LSM is driven by meteorological forcings which each explicitly incorporate some measure of a priori bias and uncertainty. Note that implicit in this SWE estimation is the assumption that precipitation (snowfall) in mountainous regions is the largest source of error in the modelling of SWE. Precipitation fields from raw products like MERRA2 tend be coarse, smooth, and biased (Liu and Margulis, 2019; Fig. 3.2a, b). This is acknowledged by the bias correction and large uncertainty in the postulated prior precipitation distribution, represented by:

$$P_j^-(t) = b_j^- * P_{nom}(t) \tag{1}$$

where $P_j^-(t)$ is the prior precipitation value for ensemble j at time step t, P_{nom} is the nominal precipitation estimate (i.e., interpolated from MERRA2 as in Fig. 3.2a), and b_j^- is a scaling factor where the " $-$ " superscript represents that this is a prior estimate not conditioned on independent observations. The ensemble of scaling factors b (Margulis et al., 2019) are effectively seasonal multiplicative bias correction factors for precipitation. The prior b values are prescribed as a lognormally distributed multiplicative factor that describes first-order bias and uncertainty in the nominal precipitation.

Second, a reanalysis step incorporates independent measurements such as fSCA using a Bayesian PBS update. The a priori (equal) prior weights assigned to each ensemble number are updated to posterior weights that reflect the likelihood that the ensemble member fits the assimilated measurements. These posterior weights are applied to prior ensemble estimates of SWE to derive posterior estimates. Note that fSCA and other potential measurements used for assimilation are connected based on physical processes in the model to other snow variables such as SWE; thus, the whole suite of snow variables is updated both before and after the assimilated measurement time step. So, although this framework has mainly been used to derive posterior SWE estimates, a by-product of this is posterior estimates of all snow states/fluxes and variables like b described above.

3.3.3 Development of historical reference dataset and precipitation bias correction

To generate the historical reference SWE dataset for this study, the SWE reanalysis framework was applied to the study domain in the same way as in Fang et al., 2022, but with an increased ensemble size and initial conditions set to default values to focus on the derivation of posterior *b* values for testing herein. Forcings were sourced from hourly MERRA2 near-surface meteorological forcing data, and the uncertainty models used to perturb input air temperature, precipitation, dew point temperature, and shortwave radiation, as well as model parameters, use the values derived for the Western U.S. domain by Fang et al. (2022) following the methods outlined in Liu and Margulis (2019).

Measurements from Landsat-derived fSCA (raw resolution of ~30 m aggregated to modelling resolution) provide the data assimilated into the historical reference dataset. We apply screening methods consistent with Fang et al. (2022) to exclude Landsat observations with cloud cover fraction greater than 40% and individual cloudy pixels with an internal cloud mask. All remaining fSCA measurements are assimilated into the reanalysis retrospectively and as a batch for each water year. A measurement error standard deviation for retrieved Landsat fSCA is specified as 10% (Fang et al., 2022). A uniform spatial resolution of 16 arcseconds (~500 m) and an ensemble size of 100 members is chosen, with hourly outputs aggregated to a daily timestep for water years (WYs) 1985 to 2021. Initial conditions are set to default values at the start of each water year; for SWE, that value is 0. This assumes that the seasonal snowpack melts out yearly; although this may not happen every year especially at high-elevation shaded areas of the watershed, we argue this is an assumption worth making to avoid accumulating error and to make fair and consistent comparisons between simulations.

In addition to generating a high-resolution reference dataset of SWE estimates for a 37-year historical period to use for validation, this application of the SWE reanalysis framework also yields a rich database of historical precipitation bias correction factors that are conditioned on assimilated fSCA measurements. These values provide insight into the historical distribution of precipitation bias and uncertainty. Hereafter, and unless otherwise indicated, b will refer to the ensemble mean of the posterior b distribution; this is used interchangeably with "bias correction". This database comprises 37 years * 100 ensemble realizations = 3,700 values of b at each pixel. A historical distribution of b can then be derived at each pixel and for each water year, as demonstrated in Liu and Margulis (2019). This study leverages the insights stored in b towards developing real-time SWE estimates by using them to inform the precipitation bias correction of real-time applications (section 3.3.4.1).

Figure 3.2 illustrates how, for sample water year 2016, the b from the historical reference dataset provides valuable built-in downscaling and bias correction information. Note that the posterior precipitation has a much higher resolution than the raw MERRA2 input; for example, ridge and valley features are noticeable in Fig. 3.2c whereas the field is very coarse in Fig. 2a, and smooth and unresolved in the interpolated field of Fig. 3.2b. Furthermore, the raw MERRA2 input fails to capture the expected orographic effect whereas the posterior precipitation clearly shows more precipitation at higher elevations in the north and along the watershed ridgelines. The posterior b's, which are informed by the pixel-wise fSCA ablation time series in the reanalysis (assimilation) step, are the multiplicative (bias correction) factors that bring out these features in the posterior precipitation. They contain both a spatially distributed pattern relating to topography and static physiographic features (Fig. 3.2d) as well as an interannual variation (Fig.

3.2e, f). We observe that, for the study domain, this interannual variation is correlated with the prior MERRA2 precipitation level: a higher watershed-average prior precipitation correlates with a lower watershed-average bias correction (R = -0.6) (Fig. 3.2f).

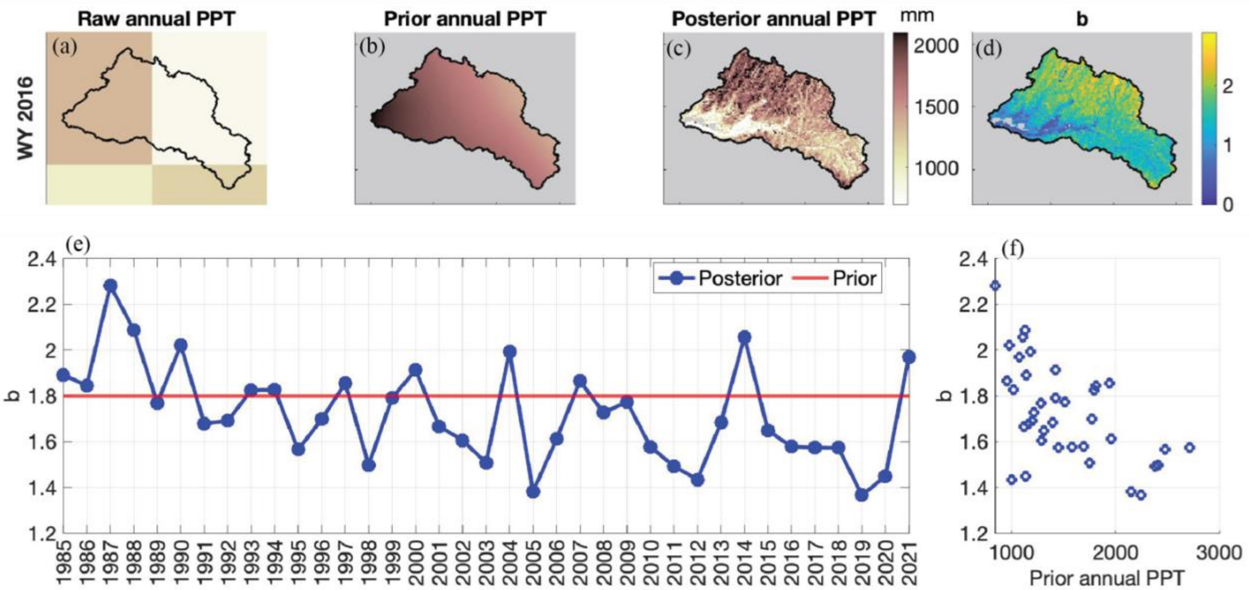


Figure 3.2. For sample WY 2016: (a) Raw MERRA2 annual precipitation at its original resolution. (b) Ensemble-mean annual prior precipitation, which uses the raw MERRA2 precipitation interpolated to the model resolution. (c) Ensemble-mean annual posterior precipitation. (d) Ensemble-mean of the posterior bias correction b. (e) Watershed-average ensemble-mean prior (red) and posterior (blue) b per the reference dataset. (f) Scatter plot showing negative correlation between watershed-average annual prior precipitation and watershed-average posterior b. Note that for (c) and (d), a non-seasonal snow mask screens out model pixels located below 1,500 m, with less than 2 cm of climatological SWE, and/or categorized as glacier.

3.3.4 Design of real-time modelling and data assimilation experiments

We use the *b* database from the historical reference dataset to inform and develop value-added precipitation bias correction for the real-time experiments. The list and characteristics of these experiments are tabulated in Table 3.1. Section 3.3.4.1 describes the bias correction approach for each experiment. Sections 3.3.4.2 and 3.3.4.3 provide further details about the data assimilation experiments (fSCA and snow depth, respectively). Figure 3.3 provides an illustrative schematic of the precipitation bias correction, data assimilation, and resulting SWE estimation for a sample pixel (location A in Fig. 3.1a) and water year (WY 2017).

Table 3.1. Summary of the methods applied to the six real-time experiments. The listed bias correction represents the mean of the prior *b* distribution. The "(x)" notation refers to the bias correction being a spatially distributed field, where x is each pixel in the watershed.

	Experiment name		**Mean bias correction**
Forward modelling experiments	Uncorrected		1
	Uniform		1.8
	Historically informed	Case A	$b_{clim}(x)$
		Case B	$b_{wet}(x)$, $b_{normal}(x)$, $b_{dry}(x)$
Data assimilation experiments	Case B + fSCA*		$b_{wet}(x)$, $b_{normal}(x)$, $b_{dry}(x)$
	Case B + SD*		$b_{wet}(x)$, $b_{normal}(x)$, $b_{dry}(x)$

* In assimilation experiments, only observations up to 1 April is included.

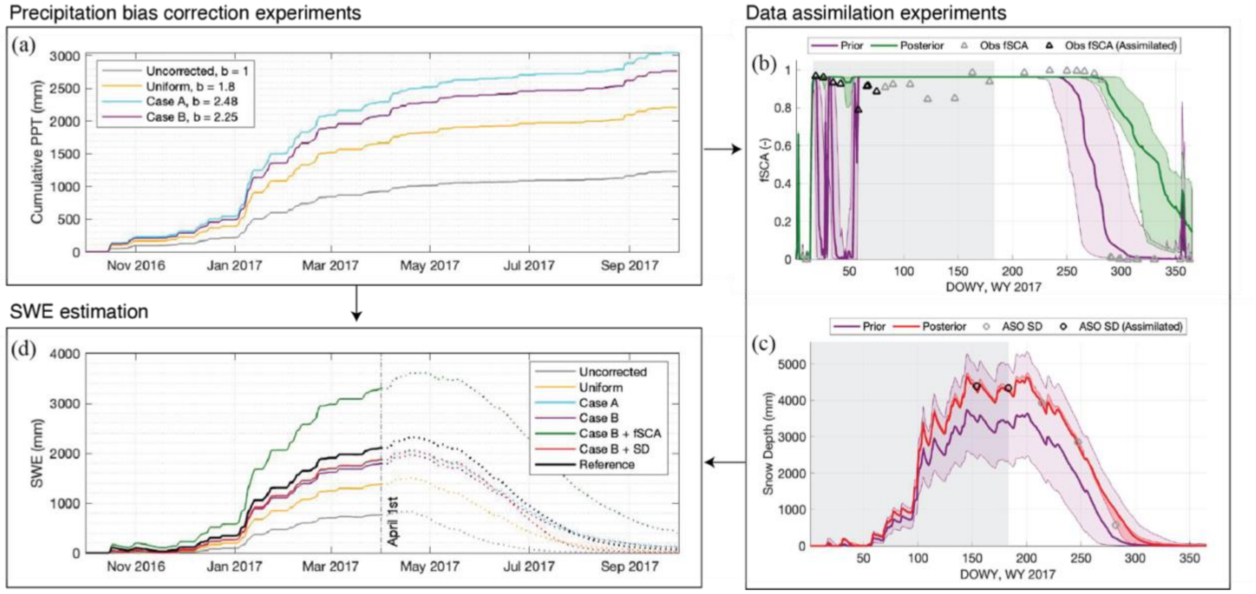


Figure 3.3. For a sample wet year (WY 2017) and sample pixel A: (a) Ensemble-mean cumulative prior precipitation for the four forward modelling experiments. (b) Example of how fSCA assimilation is applied. The prior is the Case B experiment, and the posterior the Case B + fSCA experiment. Note that both the median and interquartile range (IQR) for the prior and posterior ensembles are plotted. The assimilation window, indicated with a grey rectangle, ranges from the snow onset date to 1 April. Observations that fall within the assimilation window and on a day when the prior ensemble is non-zero are labelled "assimilated". (c) Example of how snow depth assimilation is applied. Like (b), the prior is the Case B experiment, and the posterior is the Case B + SD experiment. The assimilation window ranges from the start of the water year to 1 April. (e) Ensemble-mean SWE time series for the four forward modelling experiments, two data assimilation experiments, and the historical reference dataset. The outcome of these experiments is evaluated on 1 April.

Note here that we are applying and evaluating these experiments in hindcast mode and are selecting 1 April as the target representative date. As such, we are testing these methods at the end of the accumulation season, when real-time SWE estimates would provide the most value for water supply forecasts. For true real-time application in an operational context, other factors such as data latency and computation time should be considered.

Each experiment is evaluated by its ability to reproduce SWE spatial fields as compared to the historical reference dataset on 1 April, and ASO-derived SWE on the validation day closest to 1 April. We select 1 April because it has traditionally been used to approximate peak SWE in the Sierra Nevada and is when the key April-July water supply forecasts are made (e.g., He et al., 2016). For both, we compute three metrics: Pearson correlation coefficient R to quantify how closely the reference spatial distribution is captured, normalized root mean square difference (normalized by the observational mean, NRMSD, %) to measure bias and random error, and mean difference (MD, mm) to measure the average bias.

Note that the Uncorrected, Uniform, Case A and Case B experiments (Table 3.1) only use the forward-modelling component of the reanalysis framework; because there is no data assimilation, there is no reanalysis step and therefore no posterior estimates. Instead, validation of these experiments is performed on the modelling-only prior estimates. The Case B + fSCA and Case B + SD experiments include data assimilation and thus yield posterior estimates.

3.3.4.1 Definition of precipitation bias correction factors

We generate a baseline, uninformed case where the prior precipitation is uncorrected (Uncorrected in Table 3.1, Fig. 3.3a). The Uniform experiment adjusts prior precipitation with a uniform (in time and space) mean prior b that matches that used in the historical reference dataset and defined in Fang et al. (2022) (Table 3.1). This represents the case for a simple precipitation bias correction (Fig. 3.3). Note that we maintain the nominal CV and minimum/maximum values from Fang et al., 2022 for the prior *b* ensemble for this and all subsequent experiments; that is, we alter only the ensemble mean.

The more informed experiments leverage the database of historical *b* factors generated as a byproduct of the SWE reanalysis framework. They vary from the less-informed cases (Uncorrected, Uniform) in two key ways: the prescribed precipitation bias corrections are spatially distributed, and historically informed. From the historical reference, we compute a spatially distributed climatological *b* for each water year, withholding the *b* value from a given year in deriving a long-term climatology for that year. These climatological values are used as the mean bias correction for the Case A experiment (Table 3.1, Fig. 3.3a). Because we observe a relationship between precipitation level and watershed-average *b* in the historical reference (Fig. 3.2f), we also derive a bias correction that is conditioned on water year type. For each water year, we determine a type based on the historical prior precipitation (cumulative on 1 April), where <30th percentile is "dry", >70th percentile is "wet", and in between is "normal". We take a spatially distributed average of the historical *b*'s of all the other years classified in that water year type; that average becomes the mean value of bias correction for that year in the Case B *experiment* (Table 3.1, Fig. 3.3a). In Fig. 3.4, we illustrate how the historical *b* factors across the watershed for wet years tends to be less than the climatological values (with the exception of headwater river valley bottoms, as shown in red in Fig. 3.4d); and those for dry years tend to be greater (Fig. 3.4b). In Fig. 3.3a, we see how the prior precipitation in Case B differs from Case A because of the difference in precipitation bias correction.

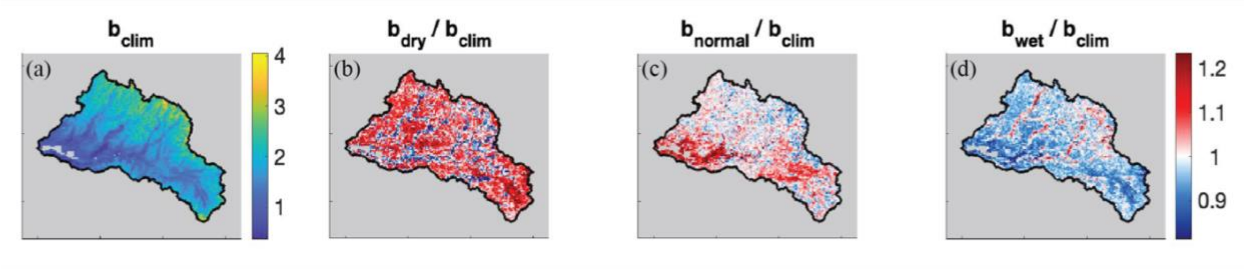


Figure 3.4. Per the historical reference: (a) Climatological (WY 1985-2021) bias correction map. (b) The ratio between the dry year and climatological bias correction. (c) Same as (b) but for normal years. (d) Same, but for wet years.

The last two experiments include the assimilation of either fSCA or snow depth measurements; these experiments are subsequently referred to as Case B + fSCA and Case B + SD, respectively. The data assimilation methods are described further in sections 3.3.4.2 and 3.3.4.3. We choose to use the Case B method of correcting prior precipitation for these experiments because it represents the most informed and sophisticated bias correction.

3.3.4.2 Assimilation of fSCA observations

Previous data assimilation experiments that use fSCA to effectively improve SWE estimates have typically utilized measurements from both the accumulation and melt seasons (e.g., Girotto et al., 2014; Margulis et al., 2016; Fang et al., 2022), which are assimilated retrospectively in a single batch at the end of the water year. This is done with the understanding that it is the fSCA ablation time series combined with estimates of snowmelt that is most directly connected to SWE. The value of fSCA measurements during the accumulation season is expected to be more limited because snow coverage is often complete or near-complete (i.e., fSCA = 1) when snow is accumulating in snow-dominated areas. Past studies have found limited to no improvement from prior modeled SWE estimates when assimilating fSCA observations over the accumulation season, independently from one another, or at sites which experience long periods of near-complete snow cover (Andreadis and Lettenmaier, 2006; De Lannoy et al., 2010).

Here, we test whether there is any additional benefit in SWE estimation with bias-corrected precipitation when assimilating fSCA up to 1 April, and where in the watershed that benefit might be the greatest. We use the same methods described in section 3.3.3 to derive, screen, and assimilate fSCA observations, but only include the subset of fSCA observations that fall between a snow onset date and 1 April (Fig. 3.3b, more details about the assimilation window in Text B1). Note that not all pixels in the watershed assimilate the same number of fSCA observations for a given year because of differences in the snow onset date, cloud cover, and satellite orbital patterns.

3.3.4.3 Assimilation of snow depth observations

The experiment with snow depth assimilation (Case B + SD) incorporates multitemporal lidar-derived snow depth observations taken over the Tuolumne watershed by ASO (Painter et al., 2016) on and before 1 April. The observations cover the entire watershed (Fig. B3). In contrast to fSCA, snow depth observations are expected to provide more insight into accumulation season SWE because of the close relationship between snow depth and SWE. Margulis et al. (2019) demonstrated how the assimilation of even a single day of ASO snow depth observations was able to significantly improve posterior estimates of SWE later in the year. Here, we seek to quantify how much assimilating snow depth observations could improve upon SWE estimates that already incorporate a historically informed precipitation bias correction. We use data for three representative years: WY 2015 (dry), 2016 (near average), and 2017 (wet). The number and dates of the ASO observations used for assimilation and validation purposes are summarized in Table 3.2. Prior to assimilation, the 50 m ASO snow depth product was regridded to the modelling resolution. Following Margulis et al. (2019), we specify a measurement error standard deviation of 5 cm. Figure 3.3c illustrates snow depth assimilation for a sample model pixel: the observations on March 3rd and 1 April of this year fall within the prior ensemble, and so are assimilated and yield a higher posterior mean and narrower ensemble.

Table 3.2. ASO acquisition DOWYs and their corresponding dates. The dates of the observations used for assimilation and for validation (i.e.: the day closest to 1 April) are indicated for each WY.

		Assimilated	**Validation**
WY	2015	140 (Feb 17), 156 (Mar 5), 176 (Mar 25)	185 (Apr 3)
	2016	178 (Mar 26), 184 (Apr 1)	190 (Apr 7)
	2017	154 (Mar 3), 183 (Apr 1)	214 (May 2)

3.3.5 Connection to streamflow

We further evaluate the real-time SWE experiments by their ability to yield snowmelt estimates and streamflow forecasts that match observed streamflow at the TGC (Tuolumne River at the Grand Canyon) USGS gauges. Note that for this comparison, the focus is on a subwatershed that drains through the TGC gauge (Fig. 3.1a). Continuous daily streamflow records are generated from observations at the USGS gauge TGC for WYs 2009-2021 (Text B2).

We measure how well daily watershed-average snowmelt estimates correlate with daily observed streamflow for the key forecasting period April-July with Pearson correlation coefficient R. For each experiment, the SWE reanalysis framework is run forward in time from 1 April with the known meteorological forcings of that year, yielding “perfect” hindcasts for SWE and snowmelt. Snowmelt is estimated as the negative daily changes in SWE, assuming sublimation is negligible.

We quantify the ability of watershed-average estimated 1 April SWE to predict April-July (AJ) streamflow volume with a historical linear regression. For a given year, we build a regression from all other years (i.e., excluding the one being evaluated) using observed AJ streamflow as the predictand (Fig. B8). Observed AJ streamflow from WY 2019 is excluded because of incomplete observations (Fig. B2). We quantify the bias and mean absolute differences between the predicted volumes from the experiments and the reference. Because we are treating the historical reference dataset as the ground truth in this study, the predicted volume from its SWE estimates is treated as the "best case" prediction and so is the target in this comparison. The relationship between estimated 1 April SWE and observed streamflow is affected by other factors like rainfall and soil moisture conditions which make errors between the predicted and observed streamflow volume inevitable, and not the focus of the study. Note that for this analysis, the snow depth experiment is excluded because it only has 3 years of results.

3.4 Results

3.4.1 Value added from historically informed precipitation bias correction

Maps of 1 April SWE estimates and their difference relative to the historical reference (Fig. 3.5) highlight key differences across experiments; in particular, these show how using a historically informed precipitation bias correction (Case A and Case B) yields 1 April SWE that better matches the reference. Here, the reference is the posterior SWE estimates from the historical reference dataset which is constrained by the full set of fSCA observations across the water year. Notably, we can see how Case A and Case B produce 1 April SWE distributions that are much better spatially-resolved than the less-informed Uncorrected and Uniform experiments:

individual ridges and valleys are more prominent and match the historical reference more closely (Fig. 3.5a). This is also notable in the maps illustrating the bias in SWE: in the Uniform experiment, a strong geographical pattern exists in WY 2016 and 2017, where the lower-elevation areas of the watershed (generally, the southern half) consistently show a positive bias in SWE estimation, and the higher elevations show a negative bias (Fig. 3.5b). This geographical distinction is lessened in Case A and Case B in 2017 and 2015, and eliminated in 2016. This illustrates that the historically informed, spatially distributed bias correction is able to correct biases in input precipitation that are related to elevation and topography, whereas the uniform bias correction smooths over these spatial differences. We also see that Case A and Case B yield relatively similar results in terms of the spatial distributions and magnitudes of error.

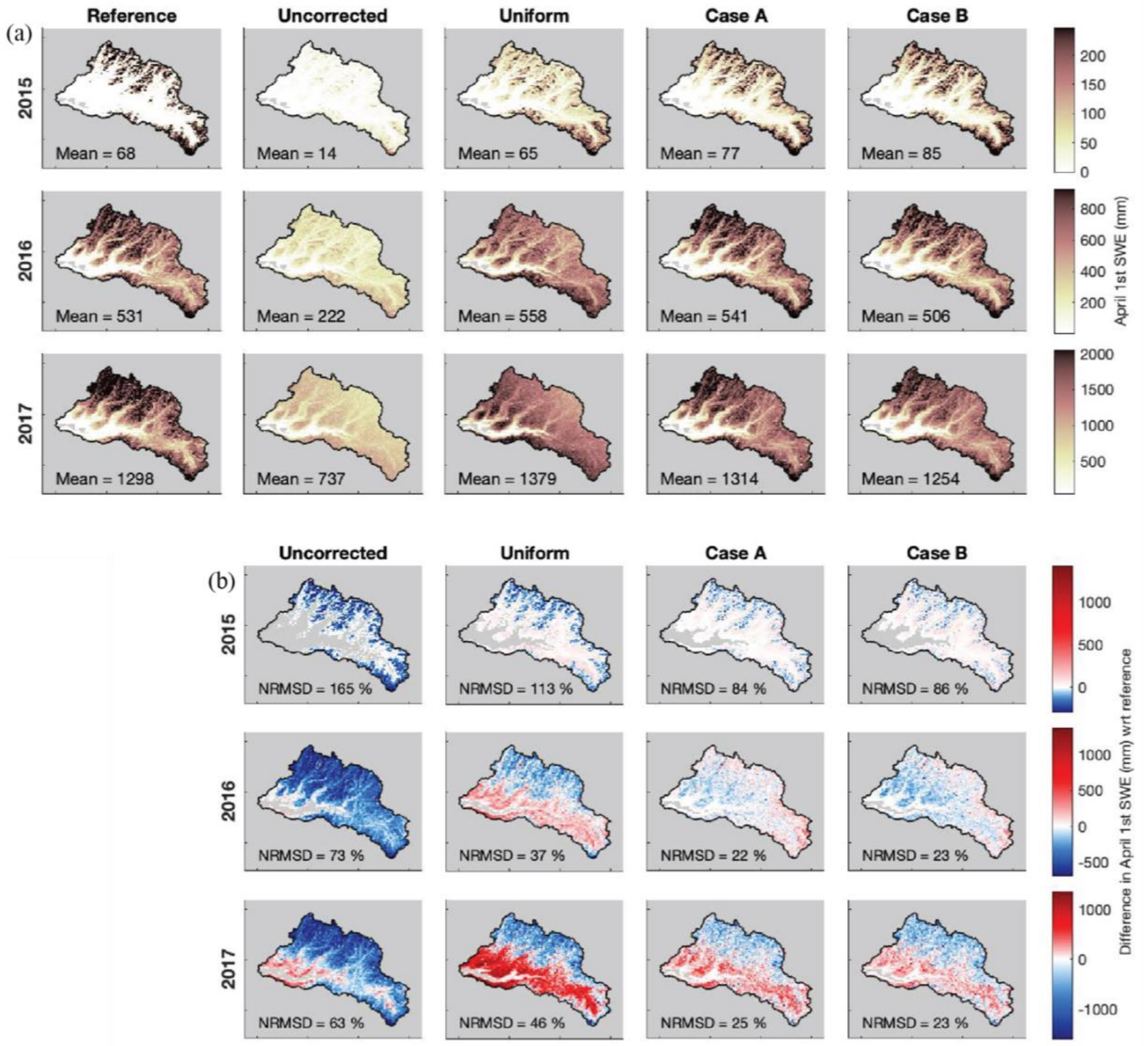


Figure 3.5. (a) Maps of 1 April SWE for the historical reference, Uncorrected baseline, and three experiments for three representative water years: 2015 (dry), 2016 (normal), and 2017 (wet). (b) Maps of the difference in 1 April SWE (experiment - reference) for the same years and experiments. The NRMSD relative to the reference is included. Pixels where both the reference and experiment estimate 0 SWE are greyed out in addition to the mask in these maps.

All modelling-only experiments (Uniform, Case A, Case B) outperform the Uncorrected baseline in at least two 1 April SWE metrics (Fig. 3.6e-g). The greatest reduction in error relative to the Uncorrected baseline occurs when a spatially distributed bias correction is used: on average across all years in the record, Case A and Case B reduce NRMSD by 58% and 57%, improve R by 43%, and reduce bias (MD) by 88% and 85%, respectively (Fig. 3.6e-g). The limited difference in performance between the two cases suggests that the primary value of the historical database of bias correction distribution lies in its description of the (more or less static) *spatial distribution* of precipitation bias, rather than its temporal patterns or uncertainty. A more simple uniform bias correction is also effective at reducing error but to a lesser degree: Uniform reduces average NRMSD by 35% and MD by 97%, and yields an insignificant average impact on R (Fig. 3.6e-g). Note that the MD metric averages values across the watershed and so does not represent the spatial spread of error in these estimates (Fig. 3.5b). For the Uniform experiment in particular, a lower watershed-average MD masks high positive and negative errors across the watershed (Fig. 3.5b).

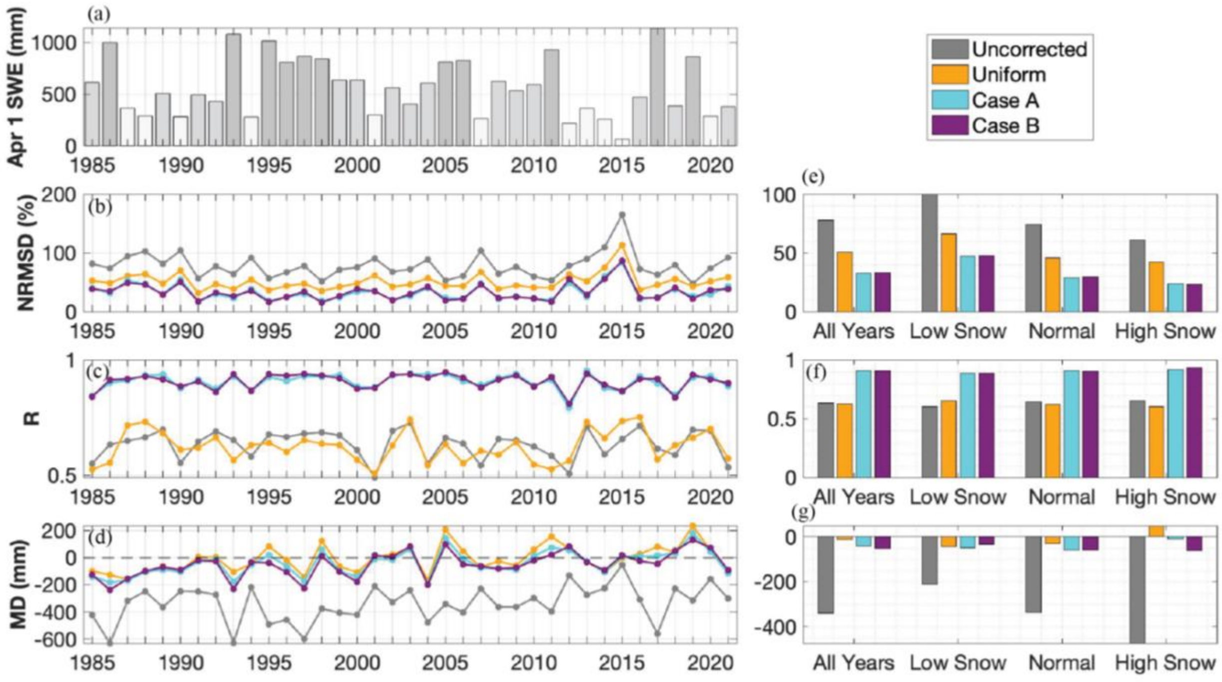


Figure 3.6. (a) Watershed-average 1 April SWE per the historical reference dataset. Low snow (below 30th percentile), normal snow, and high snow (above 70th percentile) years are colored by increasing shades of grey. (b-

d) Yearly performance metrics (NRMSD, R, MD; top to bottom) for 1 April SWE in the four modelling-only experiments as compared to the reference. (e-g) Performance metrics (NRMSD, R, MD; top to bottom) averaged across all years, low snow years, normal snow years, and high snow years.

The performance of these experiments varies by year. For example, in WY 2015, which was historically dry in the Hetch Hetchy watershed, NRMSD is the highest across all experiments (Fig. 3.6b). Although the two spatially distributed precipitation bias corrections yield similar results in most cases, Case B (where the bias correction is also differentiated by water year type) has an 8-30% lower bias than Case A in low and normal snow years (Fig. 3.6g). This indicates that differentiating the bias correction by water year type further reduces bias in years with lower snow accumulation. In those years, the bias correction is generally higher than the climatological mean (Fig. 3.2b), which, when applied to prior precipitation, effectively increases the snowfall input and reduces the negative SWE bias. On the other hand, Case B has a more negative SWE bias in high snow years than Case A (Fig. 3.6g); in these high snow years, the bias correction is generally lower than the climatological value and so, when applied to prior precipitation, would reduce input snowfall. The impact of a uniform bias correction factor also varies slightly by water year type: Uniform improves R from the Uncorrected baseline in low snow years, but reduces R in normal or high snow years, indicating a poorer spatial correlation to the reference in years with higher snow accumulation (Fig. 3.6f). High snow years in Uniform are also the only years to show an average positive bias (Fig. 3.6g), suggesting the uniform bias correction generates higher input precipitation than the reference dataset in higher snow years. This is consistent with the observation in Fig. 3.2f that, historically, wetter years correlate with a lower watershed-average bias correction; in these years, the posterior bias correction is less than the uniform value.

Figure 3.7 illustrates the distribution of error in climatological SWE in these experiments across elevation bands. At low elevations (below 2600 m), the Uncorrected baseline tends to have the least error (lowest RMSD), but also a significant negative bias (MD) that persists across all elevations and indicates an underestimation of SWE (Fig. 3.7b, d). The other three experiments, which incorporate varying levels of precipitation bias correction, instead overestimate SWE at low elevations: Case A and Case B by small magnitudes (~5-10 mm) and Uniform by a greater magnitude (~200 mm) (Fig. 3.7d). Spatial distribution of SWE is best represented by Case A (with Case B a close second) at low elevations, as indicated by the highest R (Fig. 3.7c). At mid elevations (between 2600 m and 3000 m), the lowest error (RMSD) occurs in the Uniform experiment, then Case B and Case A, and then Uncorrected (Fig. 3.7b). The bias in Case A and Case B is consistently negative at mid elevations, with Case A having the slightly lower magnitudes (Fig. 3.7d). The bias in the Uniform experiment switches from positive to negative at around 2760 m (Fig. 3.7d); this trend demonstrates cancellation effects happening in the watershed-averaged bias, where the watershed-average value in Fig. 3.6 may be low despite both high positive and negative biases at different elevations. Error across all three performance metrics increases more steadily and more steeply with increasing elevation in the Uncorrected baseline than the Case A or Case B experiments (Fig. 3.7b-d). This implies that the spatially distributed precipitation bias correction applied to the latter two is effective at reducing error across the watershed, especially at higher elevations where most of the SWE is located. At high elevations (above 3000 m), the lowest error (RMSD) and bias (MD) occurs in Case A, with Case B as a close second and Uncorrected as significantly worse (Fig. 3.7b, d). The spatial distribution of SWE (R) is best portrayed by Case B, with Case A close behind (Fig. 3.7c). Notably, the elevation band (3100-3200 m) with the lowest R across all experiments is also the one with the highest climatological April 1 SWE (Fig. 3.7a, c). This suggests that this elevation band could benefit from ablation season and post-ablation fSCA assimilation, which is included in the historical reference. All experiments underestimate SWE at high elevations (Fig. 3.7d), indicating that the spatially distributed precipitation bias corrections are not enough to fully compensate for the negative bias in SWE estimation at high elevations, although it does reduce that bias.

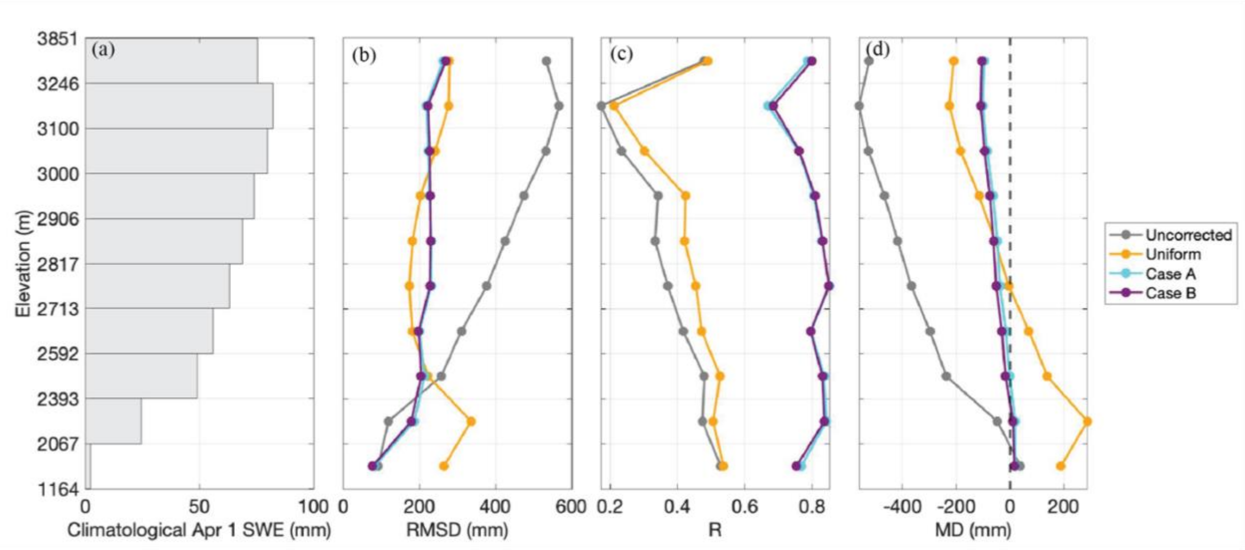


Figure 3.7. (a) Long-term (WYs 1985-2021) average 1 April SWE in each of 10 elevation bands of the watershed. Elevation band bounds were determined by distributing an even number of model pixels into each band. (b-d) Long-term average RMSD, R, and MD (left to right) for each elevation band for the Uncorrected baseline and three modelling-only experiments.

Overall, including a spatially distributed precipitation bias correction significantly improves SWE spatial estimates across all elevation bands, as indicated by consistently higher R values in Case A and Case B than both the Uncorrected and Uniform experiments (Fig. 3.7c). This bias correction also yields error that is more uniform across elevation bands; crucially, it reduces error more significantly at higher elevations where more SWE accumulates (Fig. 3.7b, d). It is worth noting that, to the extent that reanalysis precipitation products such as the MERRA2 input used in these experiments get informed by existing in situ precipitation gages, those data are generally at lower elevations. This emphasizes the need to get accurate spatially distributed bias corrections that adjust uncertain precipitation inputs at higher elevations where most SWE accumulates.

3.4.2 Additional value through data assimilation

3.4.2.1 fSCA assimilation

The value of assimilating fSCA to estimate SWE lies primarily in its ability to track the loss of snow cover during the melt season, so the expectation for additional insight from assimilating fSCA observations only before 1 April (as was done in the Case B + fSCA experiment) is small. This is confirmed by the results: when looking at the overall impact of fSCA assimilation on 1 April SWE estimates, the watershed-scale NRSMD is reduced from the prior in only 2 of 37 years (Fig. 3.7). A further 7 years show a posterior NRMSD within only 10% of the prior, indicating limited difference. In the remaining years (the majority), fSCA assimilation brings the posterior estimates further from the historical reference and increases the NRMSD (Fig. 3.7). Note that here, the experiment uses the spatially distributed precipitation bias correction from Case B. This implies that accumulation season fSCA observations, which comprise most of the observations before 1 April, are more noisy than helpful in data assimilation. In fact, assimilating these observations is detrimental to overall SWE accuracy in most years (Fig. 3.7). Overall, this method is not a useful approach to improving real-time 1 April SWE, especially when the precipitation input is already bias-corrected as is the case here.

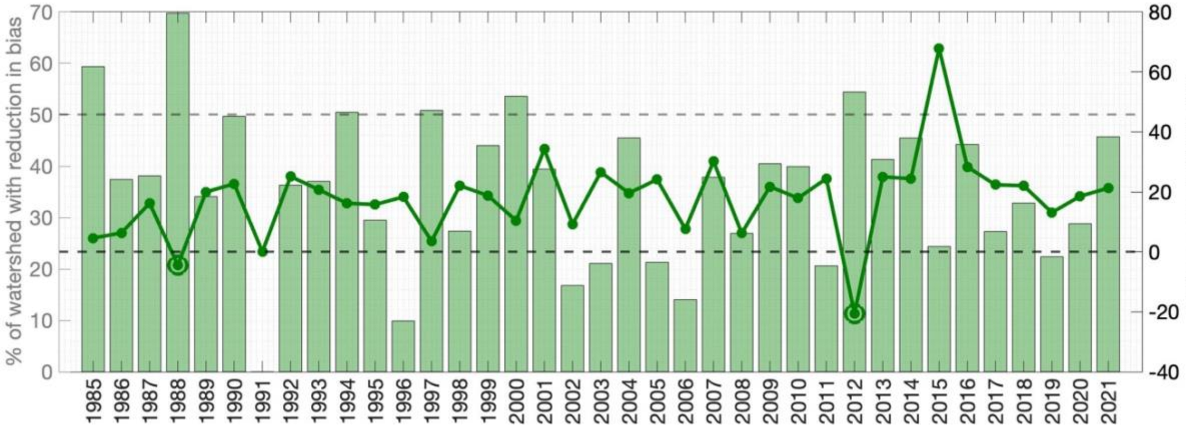


Figure 3.8. The height of the bar indicates the percent of pixels that show a reduction in error (absolute difference) relative to the reference when fSCA assimilation is applied for that WY (left axis). The solid line indicates the differences in watershed-scale NRMSD relative to the reference between the prior and posterior estimates (i.e., NRSMDpost - NRMSDprior) (right axis). The plot points below 0 are highlighted with an extra circle; in these years, the posterior estimate yields a lower NRMSD than the prior. The grey dashed line indicates the point at which 50% of the watershed shows a reduction in bias (left axis); the black dashed line indicates where the posterior NRMSD is less than the prior (right axis). Note that in 1991, no fSCA observations were assimilated so there was no change from the prior to the posterior SWE estimates.

We find that the timing of fSCA observations in the water year is significant to determining whether their assimilation reduces error in SWE estimates. For example, in WY 1988, only 1-2 fSCA observations were assimilated into the fSCA experiment, but all of these occurred after pixel-wise peak SWE and before 1 April (Fig. B4a). In that year, 70% of pixels showed a reduction in posterior error relative to the reference, and the watershed-scale NRMSD was reduced 5% (Fig. 3.7). In WY 2012, pixel-wise peak SWE averaged after 1 April for the watershed, but the NRMSD was still reduced (by 21%) with fSCA assimilation because enough fSCA observations (6-10) were assimilated at low-elevation pixels (Fig. 3.4b). Note that in cases with fewer fSCA observations during the accumulation season, there is often degradation in the posterior SWE estimate.

Individual pixels in the watershed can show improvement with fSCA assimilation even when the total error is increased. The absolute bias relative to the reference is reduced at over

50% of the pixels in the watershed with fSCA assimilation in 7 years (Fig. 3.7). The pixels where fSCA assimilation reduces error tend to have an earlier peak SWE, lower elevation, and higher number of fSCA observations assimilated (Fig. B5). In most years, there is a statistically significant difference in all these metrics between the pixels with an improvement and those without an improvement. This is consistent with the finding in Andreadis and Lettenmaier (2006) that improvements in SWE estimation from fSCA assimilation are more evident at lower elevations and during snowmelt.

3.4.2.2 Snow depth assimilation

In addition to mostly reducing watershed-scale average error (as expected), snow depth assimilation brings more spatial heterogeneity to SWE estimates and reduces pixel-wise bias. Note that here, the reference is SWE from the ASO observations succeeding the last one that was assimilated; that is, day of water year (DOWY) 185 in WY 2015, 190 in 2016, and 214 in 2017. In the maps of Fig. 3.9, we observe that in WY 2016 and 2017, the posterior estimates show a wider range in SWE than the prior, with higher SWE estimates at higher elevation pixels. Note that this prior-posterior pair uses the historically informed spatially distributed bias correction from Case B. In WY 2015, we observe how most of the watershed has already lost its snow cover by the day of comparison according to the posterior estimates (Fig. 3.9a), which contributes to the poor performance of snow depth assimilation in this year.

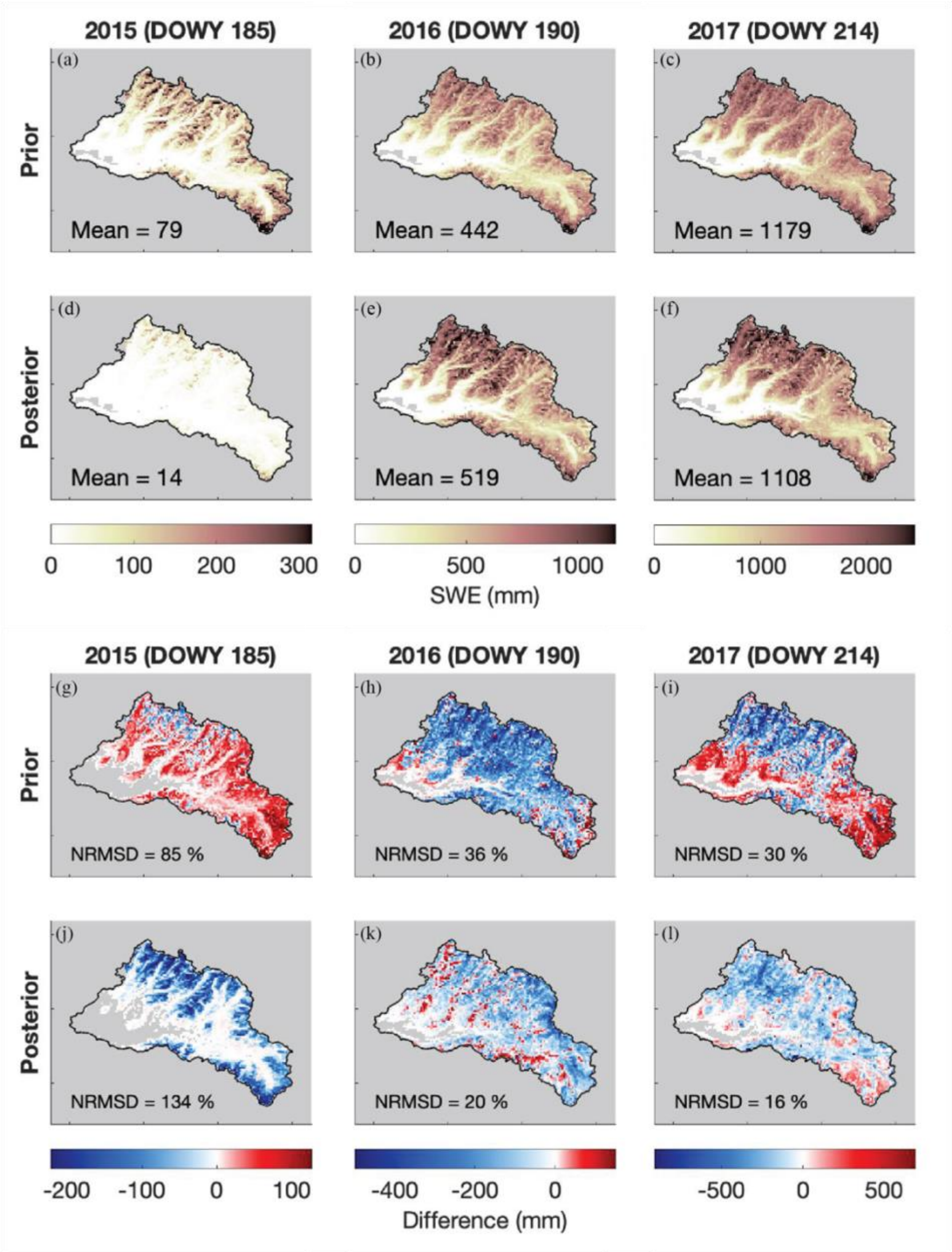


Figure 3.9. (a-c) Maps of prior SWE for the Case B + SD experiment on the validation day in three water years. (d-f) Maps of posterior SWE for the Case B + SD experiment. (g-i) Maps of the difference between prior SWE and ASO-derived SWE. (j-l) Maps of the difference between the posterior SWE and ASO-derived SWE. Values of NRMSD relative to the reference are listed. Pixels where both the experiment and the reference have 0 SWE are greyed out in addition to the mask.

In both WY 2016 and 2017, we observe a decrease in pixel-wise error with snow depth assimilation. In 2016, most of the watershed demonstrates an underestimation of reference

(ASO) SWE in the prior; this is lessened in the posterior, with some valleys showing slight overestimation (Fig. 3.9k, l). In 2017, the prior shows an underestimation of ASO SWE in the northern high-elevation region of the watershed and an overestimation in the southeastern headwaters and low river valley area, some of where ASO observes no snow by this time. In the posterior, the magnitude of pixel-wise error is universally decreased, more of the watershed underestimates ASO SWE, and the lack of snow in the low river valley areas is more correctly captured (Fig. 3.9h, i). In 2015, most of the watershed except for higher elevation ridge areas in the north shows an overestimation of prior SWE, including areas where ASO observes no snow cover (Fig. 3.9j). The posterior SWE map shows an almost universal underestimation of observed ASO SWE in this year, except for the areas where the estimates correctly predict no snow cover (Fig. 3.9g). This is consistent with the higher watershed-scale errors observed in the estimations with snow depth assimilation on the validation day in 2015 from Fig. 3.10.

In both WYs 2016 and 2017, the Case B + SD experiment shows the lowest NRMSD (81% less than the Uncorrected baseline) and the highest R (Fig. 3.10a, b). This also holds true for bias (MD) in WY 2016 (Fig. 3.10c). In all three years, Case B + SD shows a negative bias, indicating that SWE estimates are consistently underestimated (Fig. 3.10c). Assimilating snow depth observations reduces the prior NRMSD by 43-46% and increases the R by 6-12% in WYs 2016 and 2017 (Fig. 3.10a). Note that for the Case B + SD experiment, Case B represents its prior (i.e., before assimilation). Fig. 3.10b demonstrates how, for a sample model pixel, the two ASO snow depth observations assimilated in 2016 can bring the prior estimate up to a posterior that better fits the observations both before and after 1 April.

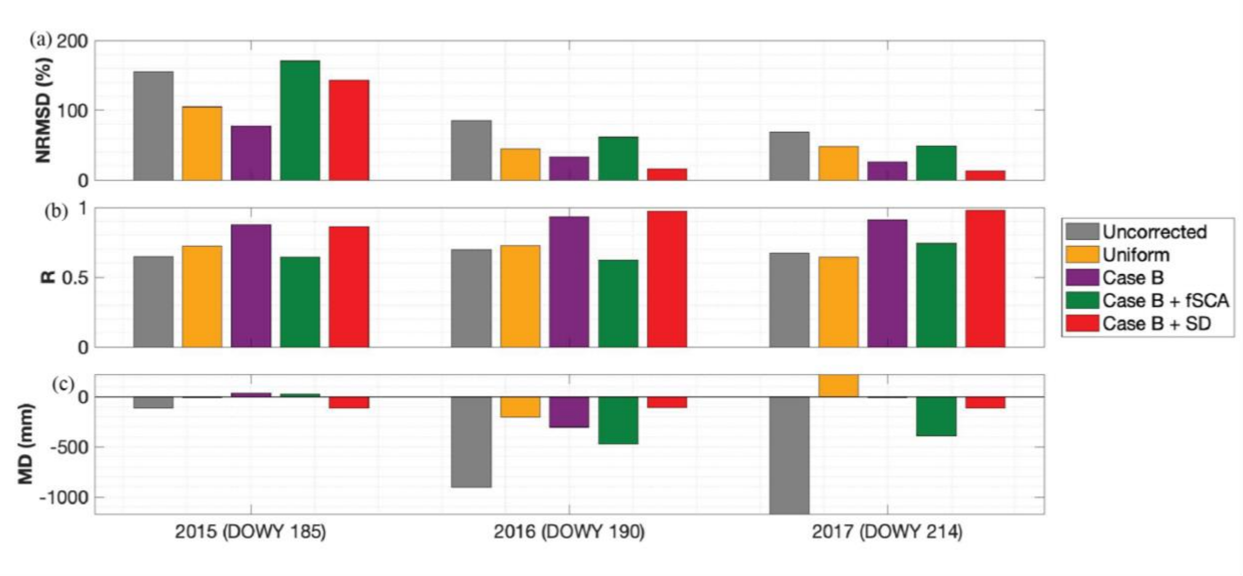


Figure 3.10. Summary of performance metrics: (a) NRMSD, (b) R, (c) MD, relative to ASO-derived SWE on the validation day for WY 2015, 2016, and 2017. Because of the similarities between Case A and Case B, only Case B metrics are included here.

In WY 2015, the outcome of the Case B + SD experiment is different on its validation day. On the days when ASO data is assimilated, the posterior SWE estimates successfully (and as expected) reduce error metrics from the prior (Fig. B6) but by the validation day on DOWY 185, 42% of the basin (mostly higher-elevation higher-snow areas, Fig. B7) has a worse bias in posterior SWE than in the prior. Although overall error is still reduced from the Uncorrected baseline on this day, NRMSD and MD are greater, and R is less, in the posterior than in the prior (Fig. 3.10). We hypothesize that anomalies in precipitation input that occur after assimilation explains the poor performance of snow depth assimilation on this day (Fig. 3.11a). Of the pixels in the watershed that exhibit higher error in the posterior than the prior estimates, 96% experienced an increase in observed snow depth from DOWY 176 (the last day when observations are assimilated) and DOWY 185 (the validation day) (for example, Fig. 3.11a). Note that, in most cases, this increase in observed snow depth corresponds to a decrease in observed SWE (Fig. 3.11b); this implies a decrease in ASO-derived snow density, which is consistent with a fresh snowfall event. The precipitation forcing around those days does not reflect a snowfall event. This inconsistency between precipitation forcing and observations, and the fact that it happened after the last assimilated observations and that WY 2015 was an exceptionally dry year, explains the negative bias at those pixels and in the watershed by the validation day in 2015.

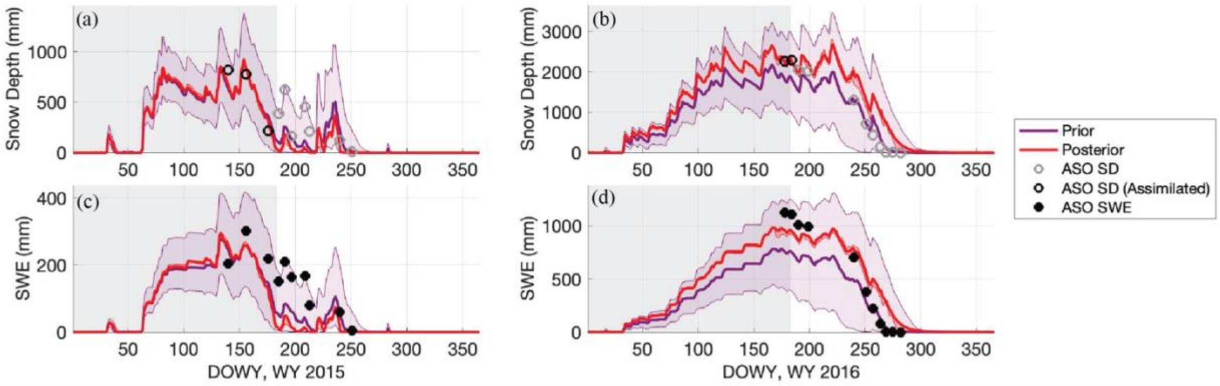


Figure 3.11. Sample time series illustrating snow depth assimilation at a sample model pixel in (a) WY 2015 and (b) WY 2016. The prior ensemble is in purple; the posterior in red. Note that both the median and interquartile range (IQR) for the prior and posterior ensembles are plotted. The assimilation window is indicated with a grey background rectangle. The bottom two plots illustrate prior and posterior ensemble mean and IQR for SWE estimates in (c) WY 2015 and (d) WY 2016.

3.4.3 Value for streamflow forecasting

A key purpose of 1 April SWE estimates is to support streamflow forecasts for spring and summer water supply. We argue that snowmelt is a reasonable proxy for streamflow in this case because Hetch Hetchy is a snow-dominated watershed. We observe that all experiments using bias-corrected precipitation are effective at yielding post-1 April snowmelt estimates that correlate better to observed streamflow than the Uncorrected baseline. Fig. 12a-f illustrates both the cumulative and daily time series of simulated snowmelt and observed streamflow for April-

July in WYs 2015-2017. We observe that the cumulative simulated snowmelt often exceeds the cumulative streamflow (Fig. 3.12a-c); this is expected because some snowmelt will infiltrate into the soil column or evaporate before reaching the river. Regardless, the Uncorrected baseline consistently yields the least amount of cumulative snowmelt (Fig. 3.12a-c). In WY 2015, we see that the experiment with snow depth assimilation also yields cumulative snowmelt that is less than cumulative streamflow (Fig. 3.12a); this is consistent with the high error and significant underestimation of SWE in that year noted in Fig. 3.9 and Fig. 3.10. We observe that the daily simulated snowmelt reasonably captures peaks in observed streamflow: for example, the peak at the end of June 2017 which presumably is driven by a large snowmelt event (Fig. 3.12f).

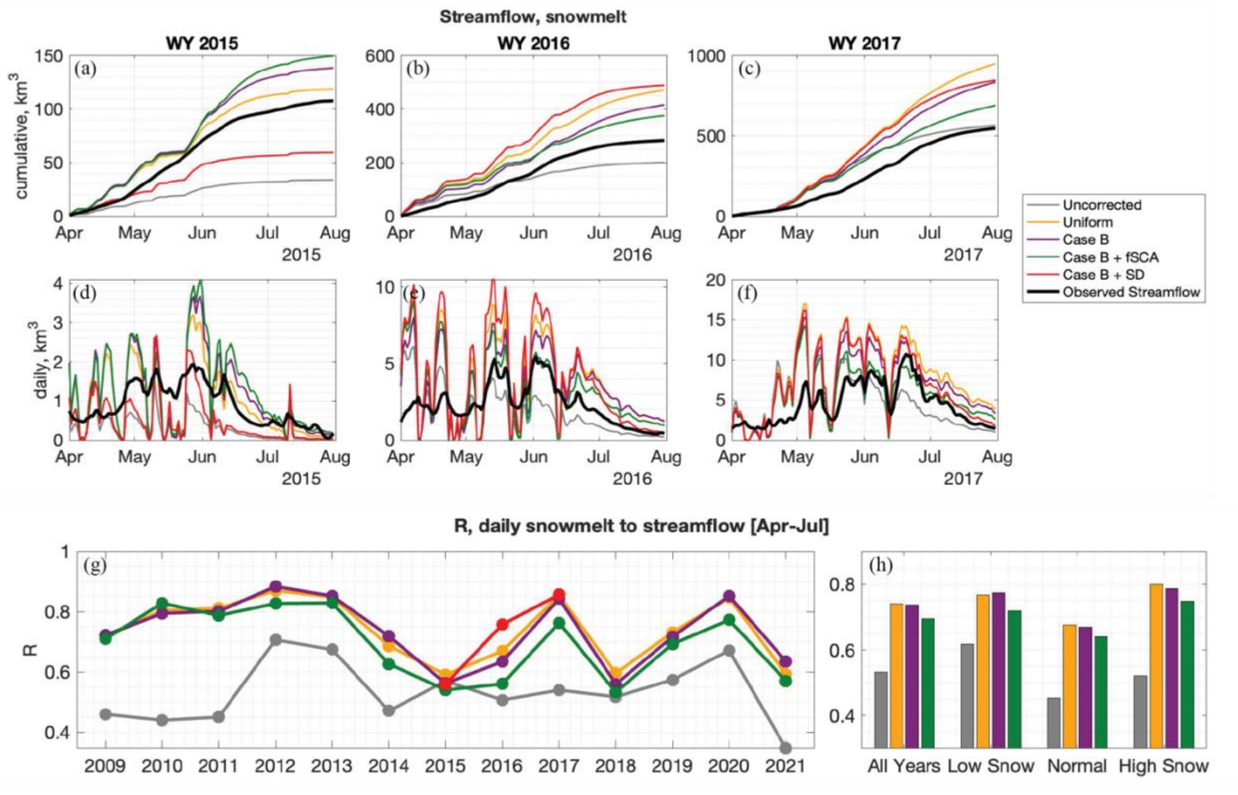


Figure 3.12. (a-c). Cumulative daily snowmelt for the Uncorrected baseline and four experiments, and observed streamflow for April-July in WYs 2015, 2016, 2017. Note that because Case A and Case B had similar results, only Case B is included here. (d-f). Same as (a-c) but showing daily values. (g) Correlation R (lag-1) between daily estimated snowmelt and observed streamflow for WYs 2009-2021. (h) Average correlation for all years, low snow

years, normal years, and high snow years. Note that these averages exclude the Case B + SD experiment because it only includes 3 years of estimates.

How well the simulated daily snowmelt corresponds to observed streamflow is quantified with a correlation coefficient (R). We test different lag times between snowmelt and streamflow and find the highest correlations with 1 day lag. We find that these correlations are significant and higher than that for the Uncorrected baseline in every year for every bias-corrected experiment (except for Case B + SD in 2015); on average, by 31-39% (Fig. 3.12g, h). In WY 2016, the highest correlation occurs in the experiment with snow depth assimilation (Fig. 3.12g). Excluding this experiment because it only yields results for a subset of WYs, the highest average annual correlation coefficient (0.74) is shared amongst Case B and Uniform. It is worth noting that Uniform has cancellation of errors at the watershed scale; that is, high and low within-watershed biases in 1 April SWE are averaged out (as demonstrated in Fig. 3.5b). Here, by aggregating snowmelt to watershed-average values, we are likely similarly averaging out within-watershed biases. The experiment with accumulation season fSCA assimilation consistently has lower correlations than those without in all water year types (Fig. 3.12h). This is consistent with the result that the Case B + fSCA experiment often yields higher errors in 1 April SWE (Fig. 3.7). In low snow and normal years, the highest correlation occurs with Case B snowmelt; in high snow and normal years, the highest correlation is with Uniform (Fig. 3.12h).

Figure 3.13 summarizes results from a linear regression model which predicts April-July (AJ) streamflow volume from watershed-average 1 April estimated SWE. The experiment with snow depth assimilation is excluded because it only has 3 years of results. The multi-year average adjusted R^2 for these regression models, developed and applied separately for each experiment-year, range between 0.94 and 0.96 (Fig. B8); these high values emphasize the strong relationship between 1 April SWE and AJ streamflow. In Fig. 3.13, we compare the predicted streamflow volume from the experiments to the predicted streamflow volume from the historical reference. The predicted AJ streamflow from reference 1 April SWE estimates is considered the "best case" prediction and so is the target in this comparison.

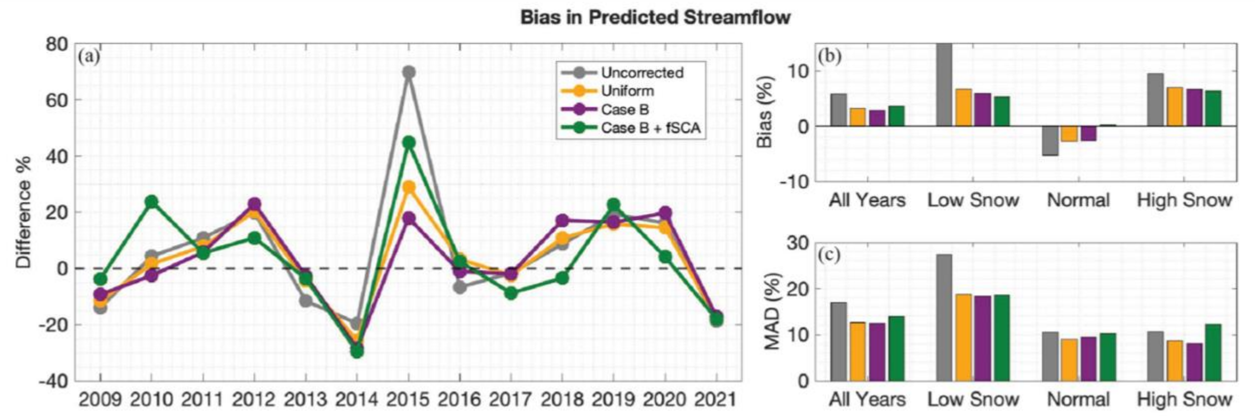


Figure 3.13. (a) Difference (%) computed between the predicted April-July streamflow volume of the experiments and the historical reference. (b) Average biases for all years, low snow years, normal years, and high snow years. (c). Mean absolute differences (MAD).

The yearly differences between the experiment and reference AJ streamflow volume range between -30% and 70% (Fig. 3.13a). The highest differences happen in 2015; this is a historically dry year which also exhibited high errors in 1 April SWE estimates (Fig. 3.6). Because these yearly differences might cancel each other out in a multi-year average, we also look at the mean absolute differences (MAD) across water year types to gauge the magnitude of error (Fig. 3.13c). Generally, the bias and MAD in low snow years is the largest and positive, signifying an overestimation of the reference AJ streamflow volume. The MAD in normal and high snow years are similar in magnitude, but the bias is opposite in direction (positive in high snow and negative in normal) and less in normal years. Overall, the average bias and MAD is reduced from the Uncorrected by the bias-corrected experiments across all water year types (except for Case B + fSCA in high snow years MAD) (Fig. 3.13a). The greatest improvement in streamflow prediction from the Uncorrected occurs with the Case B 1 April SWE estimates, which reduces average bias by 52% and MAD by 26%. The Uniform experiment is a close second (46% and 25% less average bias and MAD). This is consistent with the error reductions in 1 April SWE observed when using bias-corrected precipitation (Fig. 3.6). It is worth noting that averaging over the watershed, as is done to obtain the 1 April SWE predictor for AJ streamflow could mask spatially distributed error (which is high in the Uniform experiment, Fig. 3.5b). A spatially distributed land surface model provides the opportunity to evaluate how improvements in the spatial distribution of SWE, as is observed to happen with spatially distributed historically informed precipitation bias corrections (Fig. 3.6, Fig. 3.7), affects runoff modelling and streamflow forecasts.

3.5 Conclusions

Results demonstrate that spatially distributed historically informed precipitation bias correction significantly enhances SWE estimates. With respect to 1 April SWE fields, it reduces error (NRMSD) by 57-58%, increases spatial correlation (R) by 43%, and decreases bias (MD) by 85-88%. A simpler, spatially uniform bias correction (as used as a first guess prior in the original reanalysis methodology) also reduces error relative to uncorrected precipitation but to a lesser degree. We find that the spatially distributed historically informed bias correction yields SWE error that is not only lower but more homogeneous across elevation bands than the uniform bias correction; crucially, it reduces error more significantly at higher elevations where SWE accumulation is greater. It also significantly improves SWE spatial estimates as indicated by higher R values across all elevation bands. As illustrated by the limited differences in error reduction between the Case A (climatological) and Case B (climatological by water year type) applications, the strength of this approach lies more in its ability to capture the first-order spatial distribution of bias rather than its second-order interannual variability.

Results demonstrate that the assimilation of accumulation season snow depth further improves SWE, whereas fSCA assimilation generally does not. fSCA assimilation prior to 1 April more often degrades than improves posterior SWE estimates, due to the weak relationship between fSCA and SWE outside of the ablation season. In contrast, snow depth assimilation before 1 April leads to a 45% reduction of NRSMD in SWE and a 6% increase in R (excluding the special case of water year 2015). In WY 2015, the precipitation forcing does not capture an observed snowfall event after the last-assimilated snow depth observation; neither the seasonally applied precipitation bias correction nor the assimilation is able to rectify this error. We suggest that assimilation of more observations after single-day anomalies like these could help. This underscores how the assimilation of reliable remote-sensing observations can mitigate forcing anomalies in addition to reducing overall bias and uncertainty. Although remotely-sensed snow depth observations such as those taken by ASO are not readily available everywhere, they are proven to be a good source for improvement in SWE estimation on top of bias correction methods.

The improved SWE estimates provide value for snowmelt-driven streamflow predictions, especially in high snow years. We find that using bias-corrected precipitation reduces average bias in predicted April-July runoff by 46-52% and improves average correlation between daily snowmelt and observed streamflow by 31-39%.

Future work should explore how these methods perform in other global mountain regions. A useful spatially distributed precipitation bias correction in mountain environments can be developed everywhere that historical SWE reanalysis datasets have accurately improved (or are expected to improve) SWE estimates. The power of such an approach lies in the ability to simultaneously downscale and bias-correct globally-available (coarse) precipitation products (e.g. MERRA2 in this work) for use in estimating mountain SWE. Other avenues of investigation could explore more sophisticated methods such as machine learning for bias correction estimation; the assimilation of other sources of real-time snow observations; and the impact of real-time SWE spatial estimates on streamflow forecasts through spatially distributed hydrologic modelling.

3.6 References

Andreadis, K. M., & Lettenmaier, D. P. (2006). Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources, 29*(6), 872–886. https://doi.org/10.1016/j.advwatres.2005.08.004

Bair, E. H., Abreu Calfa, A., Rittger, K., & Dozier, J. (2018). Using machine learning for real-time estimates of snow water equivalent in the watersheds of Afghanistan. *The Cryosphere, 12*(5), 1579–1594. https://doi.org/10.5194/tc-12-1579-2018

Bair, E. H., Rittger, K., Skiles, S. M., & Dozier, J. (2019). An examination of snow albedo estimates from MODIS and their impact on snow water equivalent reconstruction. *Water Resources Research, 55*(9), 7826–7842. https://doi.org/10.1029/2019WR024810

Berg, N., & Hall, A. (2017). Anthropogenic warming impacts on California snowpack during drought. *Geophysical Research Letters, 44*(5), 2511–2518. https://doi.org/10.1002/2016GL072104

Cho, E., Vuyovich, C. M., Kumar, S. V., Wrzesien, M. L., Kim, R. S., & Jacobs, J. M. (2022). Precipitation biases and snow physics limitations drive the uncertainties in macroscale modeled snow water equivalent. *Hydrology and Earth System Sciences, 26*(22), 5721–5735. https://doi.org/10.5194/hess-26-5721-2022

Clark, M. P., Hendrikx, J., Slater, A. G., Kavetski, D., Anderson, B., Cullen, N. J., Kerr, T., Örn Hreinsson, E., & Woods, R. A. (2011). Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review. *Water Resources Research, 47*, W07526. https://doi.org/10.1029/2011WR010745

Cortés, G., Girotto, M., & Margulis, S. (2016). Snow process estimation over the extratropical Andes using a data assimilation framework integrating MERRA data and Landsat imagery. *Water Resources Research, 52*(4), 2582–2600. https://doi.org/10.1002/2015WR018376

De Lannoy, G. J. M., Reichle, R. H., Houser, P. R., Arsenault, K. R., Verhoest, N. E. C., & Pauwels, V. R. N. (2010). Satellite-scale snow water equivalent assimilation into a high-resolution land surface model. *Journal of Hydrometeorology, 11*(2), 352–369. https://doi.org/10.1175/2009JHM1192.1

Dozier, J. (2011). Mountain hydrology, snow color, and the fourth paradigm. *Eos, Transactions American Geophysical Union, 92*(43), 373–374. https://doi.org/10.1029/2011EO430001

Dozier, J., Bair, E. H., & Davis, R. E. (2016). Estimating the spatial distribution of snow water equivalent in the world's mountains. *Wiley Interdisciplinary Reviews: Water, 3*(3), 461–474. https://doi.org/10.1002/wat2.1140

Fang, Y., Liu, Y., & Margulis, S. A. (2022). A western United States snow reanalysis dataset over the Landsat era from water years 1985 to 2021. *Scientific Data, 9*(1), 677. https://doi.org/10.1038/s41597-022-01768-7

Fang, Y., Liu, Y., Li, D., Sun, H., & Margulis, S. A. (2023). Spatiotemporal snow water storage uncertainty in the midlatitude American Cordillera. *The Cryosphere, 17*(12), 5175–5195. https://doi.org/10.5194/tc-17-5175-2023

Girotto, M., Margulis, S. A., & Durand, M. (2014). Probabilistic SWE reanalysis as a generalization of deterministic SWE reconstruction techniques. *Hydrological Processes, 28*(12), 3875–3895. https://doi.org/10.1002/hyp.9887

Hamlet, A. F., Huppert, D., & Lettenmaier, D. P. (2002). Economic value of long-lead streamflow forecasts for Columbia River hydropower. *Journal of Water Resources Planning and Management, 128*(2), 91–101. https://doi.org/10.1061/(ASCE)0733-9496(2002)128:2(91)

He, M., Russo, M., & Anderson, M. (2016). Predictability of seasonal streamflow in a changing climate in the Sierra Nevada. *Climate, 4*(4), 57. https://doi.org/10.3390/cli4040057

Herbert, J. N., Raleigh, M. S., & Small, E. E. (2024). Reanalyzing the spatial representativeness of snow depth at automated monitoring stations using airborne lidar data. *The Cryosphere, 18*, 3495–3512. https://doi.org/10.5194/tc-18-3495-2024

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

Kumar, R., Samaniego, L., & Attinger, S. (2013). Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. *Water Resources Research*, *49*, 360–379. https://doi.org/10.1029/2012WR012195

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., & Wood, E. F. (2015). Inroads of remote sensing into hydrologic science during the WRR era. *Water Resources Research*, *51*(9), 7309–7342. https://doi.org/10.1002/2015WR017616

Liston, G. E. (2004). Representing subgrid snow cover heterogeneities in regional and global models. *Journal of Climate*, *17*(6), 1381–1397. https://doi.org/10.1175/1520-0442(2004)017<1381:RSSCHI>2.0.CO;2

Liu, Y., & Margulis, S. A. (2019). Deriving bias and uncertainty in MERRA-2 snowfall precipitation over High Mountain Asia. *Frontiers in Earth Science*, *7*. https://doi.org/10.3389/feart.2019.00280

Liu, Y., Fang, Y., & Margulis, S. A. (2021). Spatiotemporal distribution of seasonal snow water equivalent in High Mountain Asia from an 18-year Landsat-MODIS era snow reanalysis dataset. *The Cryosphere*, *15*(11), 5261–5280. https://doi.org/10.5194/tc-15-5261-2021

Margulis, S. A., Girotto, M., Cortés, G., & Durand, M. (2015). A particle batch smoother approach to snow water equivalent estimation. *Journal of Hydrometeorology*, *16*(4), 1752–1772. https://doi.org/10.1175/JHM-D-14-0177.1

Margulis, S. A., Cortés, G., Girotto, M., & Durand, M. (2016). A Landsat-era Sierra Nevada snow reanalysis (1985–2015). *Journal of Hydrometeorology*, *17*(4), 1203–1221. https://doi.org/10.1175/JHM-D-15-0177.1

Margulis, S. A., Fang, Y., Li, D., Lettenmaier, D. P., & Andreadis, K. (2019). The utility of infrequent snow depth images for deriving continuous space-time estimates of seasonal snow water equivalent. *Geophysical Research Letters, 46,* 5331–5340. https://doi.org/10.1029/2019GL082507

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

Pan, M. (2003). Snow process modeling in the North American Land Data Assimilation System (NLDAS): 2. Evaluation of model simulated snow water equivalent. *Journal of Geophysical Research, 108*(D22). https://doi.org/10.1029/2003JD003994

Pflug, J. M., Margulis, S. A., & Lundquist, J. D. (2022). Inferring watershed-scale mean snowfall magnitude and distribution using multidecadal snow reanalysis patterns and snow pillow observations. *Hydrological Processes, 36,* e14581. https://doi.org/10.1002/hyp.14581

Raleigh, M. S., Lundquist, J. D., & Clark, M. P. (2015). Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. *Hydrology and Earth System Sciences, 19*(7), 3153–3179. https://doi.org/10.5194/hess-19-3153-2015

Schneider, D., & Molotch, N. P. (2016). Real-time estimation of snow water equivalent in the Upper Colorado River Basin using MODIS-based SWE reconstructions and SNOTEL data. *Water Resources Research, 52*(10), 7892–7910. https://doi.org/10.1002/2016WR019067

Schreiner-McGraw, A. P., & Ajami, H. (2020). Impact of uncertainty in precipitation forcing data sets on the hydrologic budget of an integrated hydrologic model in mountainous terrain. *Water Resources Research, 56*(12), e2020WR027639. https://doi.org/10.1029/2020WR027639

Selkowitz, D., Painter, T., Rittger, K., Schmidt, G., & Forster, R. (2017). The USGS Landsat Snow Covered Area Product—Methods and preliminary validation. In D. Selkowitz, *Automated approaches for snow and ice cover monitoring using optical remote sensing* (pp. 76–119). Salt Lake City, University of Utah. Retrieved from https://www.researchgate.net/publication/331024289_The_USGS_Landsat_Snow_Covered_Area_Products_Methods_and_Preliminary_Validation.

Sun, S., & Xue, Y. (2001). Implementing a new snow scheme in the Simplified Simple Biosphere Model. *Advances in Atmospheric Sciences*, *18*(3), 335–354. https://doi.org/10.1007/BF02919314

Tanaka, S. K., Zhu, T., Lund, J. R., Howitt, R. E., Jenkins, M. W., Pulido, M. A., Tauber, M., Ritzema, R. S., & Ferreira, I. C. (2006). Climate warming and water management adaptation for California. *Climatic Change*, *76*(3-4), 361–387. https://doi.org/10.1007/s10584-006-9079-5

Wood, A. W., Lettenmaier, D. P., & Vail, L. W. (2002). Long-range experimental hydrologic forecasting for the eastern United States. *Journal of Geophysical Research*, *107*(D20), 4429. https://doi.org/10.1029/2001JD000659

Xue, Y., Sun, S., Kahan, D. S., & Jiao, Y. (2003). Impact of parameterizations in snow physics and interface processes on the simulation of snow cover and runoff at several cold region sites. *Journal of Geophysical Research: Atmospheres*, *108*(D22), 2002JD003174. https://doi.org/10.1029/2002JD003174

Yang, K., Musselman, K. N., Rittger, K., Margulis, S. A., Painter, T. H., & Molotch, N. P. (2022). Combining ground-based and remotely sensed snow data in a linear regression model for real-time estimation of snow water equivalent. *Advances in Water Resources*, *160*, 104075. https://doi.org/10.1016/j.advwatres.2021.104075

Zheng, Z., Molotch, N. P., Oroza, C. A., Conklin, M. H., & Bales, R. C. (2018). Spatial snow water equivalent estimation for mountainous areas using wireless-sensor networks and remote-sensing products. *Remote Sensing of Environment*, *215*, 44–56. https://doi.org/10.1016/j.rse.2018.05.029

CHAPTER 4

Impact of climate warming and snowpack declines on California Sierra Nevada reservoir storage and water deliveries

4.0 Abstract

Climate warming is already and will continue to influence snow-derived water supply due to snowpack declines and earlier snowmelt. Here, we quantify and explain the impact of projected hydrometeorological shifts on reservoir storage and water deliveries for 13 major California Sierra Nevada reservoirs using hydrology projections from seven downscaled GCMs and a reservoir operations model. We find that by the end of the 21st century and under current operating rules, domain-wide average summer water deliveries are projected to decline by 19% and end of year reservoir storage by 18%, despite a 3% increase in annual inflow. This occurs due to shifts in seasonal inflow patterns driven by earlier snowmelt (11 days), decreased snowpack (-44%), and increased winter rainfall. The severity of these changes varies geographically from north to south, with reservoirs in the Sacramento region experiencing more pronounced snowpack, timing, and reservoir storage shifts, while Tulare reservoirs see greater increases in winter precipitation and inflow. The extent to which projected snowpack declines impact water availability is attenuated by reservoir storage and rainfall, varying based on characteristics of the reservoir, its operations, and its upstream catchment. However, current reservoir operating rules, specifically around seasonal flood pool constraints, are not optimal for maintaining water supply under the future climate.

4.1 Significance statement

This study highlights the significant impact of climate warming on California's water supply, particularly through the loss of snowpack and earlier snowmelt, and the role of reservoirs in moderating that impact. By the end of the century, decreases in snowpack, earlier snowmelt and inflow centroid, and increased winter rainfall are expected to lead to declines in summer water deliveries and end of year reservoir storage across 13 major reservoirs in the California Sierra Nevada**.** The severity of these shifts varies along a north-south gradient. The extent to which reservoir storage and rainfall help mitigate the effects of snowpack declines depends on geography, climate, and reservoir operating rules. The findings emphasize the need to reassess current reservoir management practices, as existing operating rules are not well-suited to capture earlier inflow under future climate conditions.

4.2 Introduction

Climate warming has already impacted the California Sierra Nevada region by causing reduced snowpack accumulation and earlier snowmelt and streamflow timing (Cayan et al., 2001; Stewart et al., 2005; Mote et al., 2018; Huang et al., 2018). As temperatures go up – projected by 6-9 °F in the Sierra Nevada by the end of the 21st century (Dettinger et al., 2018) – more precipitation will fall as rain rather than snow (Klos et al., 2014). As a result, the volume of Sierra Nevada snowpack is expected to decline significantly by the end of the century (e.g., Siirila-Woodburn et al. 2021), ranging from 64% less than the historical under a business-as-usual scenario to 30% less under a mitigation scenario (Sun et al., 2018). Associated changes in snowmelt timing and amount mean that water will flow out of mountain watersheds earlier in the year (Rauscher et al., 2008; Rhoades et al., 2018). By the end of the 21st century, the runoff midpoint in the Sierra Nevada is expected to be an average of 30 days earlier than it is now under a mitigation scenario, and up to 80 days earlier in some locations under a business-as-usual scenario (Schwartz et al., 2017). Average future annual precipitation in California is uncertain and projected to vary -5%-10% over the Sierra Nevada (Dettinger et al., 2018), it is widely projected that the state will experience more frequent and more extreme storm events (Dettinger et al., 2018; Swain et al., 2018). The impact of climate change in the Sierra Nevada is expected to be heterogeneous across watersheds, with northern watersheds most susceptible to decreased mean annual flow, southern watersheds most susceptible to runoff timing changes, and central watersheds most affected by longer periods with low flow conditions (Null et al., 2010).

Changes in hydrometeorology in the Sierra Nevada are important for water management across the state, as the mountain snowpack typically provides up to a third of California’s total water supply according to the California Department of Water Resources. This includes crucial water supply for agriculture in the California Central Valley and millions of downstream users. High elevation hydropower plants, many of which are fed by snowmelt, generate over 70% of the state’s hydroelectric power and are also susceptible to climate change-induced shifts in streamflow (Madani et al., 2014). Impacts of projected climate warming on snow-derived water resources from the Sierra Nevada include lower reservoir storage, decreased water supply deliveries and reliability, increased flood risk, and reduced hydropower generation (VanRheenen et al., 2004; Vicuna et al., 2007; Purkey et al., 2008; Willis et al., 2011; Rheinheimer et al., 2012; Madani et al., 2014; Knowles et al., 2018; Ray et al., 2020; Gupta et al., 2024).

The impacts of climate change in the Sierra Nevada for downstream water users is moderated by reservoirs and water management infrastructure. Reservoirs in the California Sierra Nevada serve multiple objectives including flood protection, water supply, hydropower generation, recreation, and environmental flows. These objectives can conflict at times, requiring important decision-making by reservoir operators. For example, operators in the California Sierra Nevada maintain an unused flood pool in the reservoir that can accommodate high inflows to reduce downstream flood risk; releasing water to keep this flood pool empty in the wet months sometimes occurs at the expense of water supply storage. These flood pools are based on fixed seasonal rule curves designed by the U.S. Army Corps of Engineers and informed by historical streamflow trends (USACE, 2017a). When reservoirs fill earlier due to early snowmelt and winter rainfall, they run the risk of both summer water shortages and flooding during the winter (Christensen et al., 2004; Cohen et al., 2020). As such, climate change will amplify the challenge for reservoirs to balance flood protection with maintaining water supply. Reservoirs and river basins historically reliant on snowmelt for water supply may require additional storage and/or adaptation in reservoir operating policies to mitigate these vulnerabilities (Tanaka et al., 2006; Medellín-Azuara et al., 2008; Georgakakos et al., 2011; Cohen et al., 2020; Taylor et al., 2024). However, no study has comprehensively assessed the role of reservoirs in moderating the effect of snowpack declines across California's Sierra Nevada on downstream water availability.

This study fills that research gap by quantifying, comparing, and explaining projected changes in reservoir storage and water deliveries across the entire North-South gradient of the California Sierra Nevada with the latest future climate modeling. We leverage hydrology model output produced for the California Energy Commission's Fifth Climate Change Assessment to characterize future conditions. The hydrology model uses as input downscaled global climate model (GCM) meteorology from the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Eyring et al., 2016). As such, this study uses the latest and most comprehensive future climate and hydrology modeling. We use the CMIP6-driven hydrology model output as input to a reservoir operations model, which allows us to evaluate: a) how will climate change affect seasonal storage and releases from the major reservoirs in the California Sierra Nevada?, b) how will snowpack volume and timing impact those changes?, and c) how will these changes differ across reservoirs?

4.3 Methods

4.3.1 Study area

We include 13 reservoirs in watersheds on the western-draining rivers of the Sierra Nevada, which combine to form the Sacramento, San Joaquin, and Tulare river basins (Table 4.1, Fig. 4.1). These watersheds all are strongly affected by seasonal snowpack that contribute significantly to streamflow. The southern watersheds are mainly high mountains with peak elevations above 4000 m and snowmelt dominated runoff, whereas the watersheds in the north mostly are lower elevation (peak elevations less than 3000 m) and have mixed seasonal runoff patterns driven by a combination of winter rainfall and winter and spring snowmelt. From north to south, these watersheds and their largest reservoirs and key characteristics are listed in Table 4.1. We use available historical daily records of observed storage, outflow, inflow, evaporation, and top of conservation for these reservoirs from the California Data Exchange Center (CDEC) for WYs 1985-2023 to inform, calibrate, and validate the reservoir operations model (4.3.3). We perform QA/QC checks on these data to remove obvious outliers.

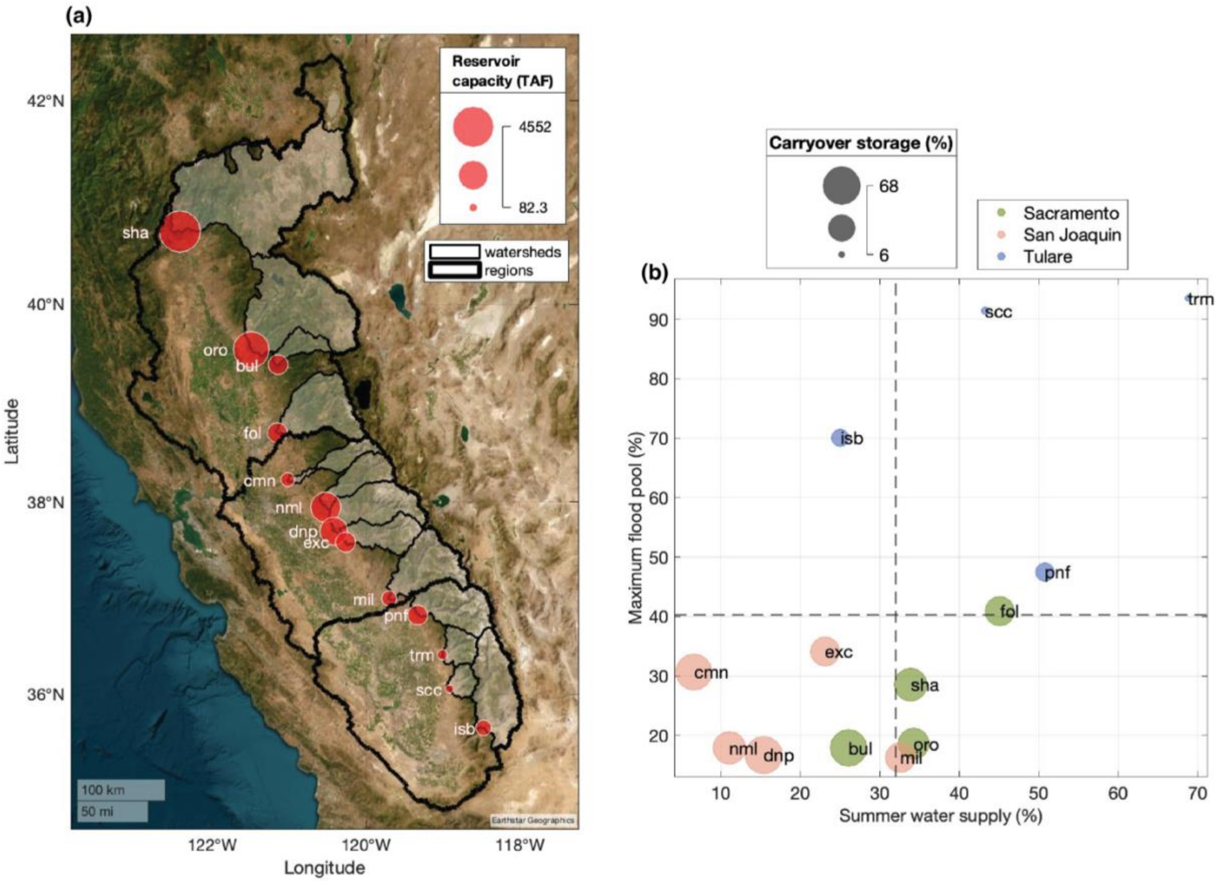


Figure 4.1. (a) Map of 13 reservoirs and their contributing watersheds. From top to bottom, the delineated HUC4 hydrologic regions are the Sacramento, San Joaquin, and Tulare. The three-letter labels refer to the reservoir ID, which is defined in Table 4.1. (b) Comparison of the historical trends in the primary uses of the reservoirs: flood control, storage, and water supply (all expressed as % of total reservoir capacity). Maximum flood pool is defined as the difference between reservoir capacity and the minimum top of conservation (TOC) generated for the reservoir operations model. Summer water supply is defined as the difference between median historical observed reservoir storage on June 1st and September 30th. Carryover storage is the median historical observed reservoir storage on September 30th. Horizontal and vertical lines indicate the averages.

These reservoirs are managed by state, local, or federal agencies for multiple purposes including irrigation and water supply, hydropower, recreation, and flood control (Table 4.1). We quantify and compare the reservoir storage dedicated to flood control, summer water supply, and carryover (end of year) uses from the historical observed record in Fig. 4.1b. The maximum (winter) flood pools range from 16% to 94% of the total reservoir capacity (Table 4.1, Fig. 4.1b); this indicates how much of the reservoir is used for flood control purposes. The reservoirs in the Tulare region have the largest flood pools. The observed changes in reservoir storage over the summer months (June to September) range from 7% to 69% of total reservoir capacity (Fig. 4.1b); this indicates how much of the reservoir is used for summer water supply. In general, reservoirs in the Sacramento and Tulare regions have higher summer water supply storage than reservoirs in the San Joaquin region. Reservoirs in the San Joaquin region have below-average flood pools and below-average summer water supply, and so are used more for long-term storage. Carryover storages range from 6 to 64% of total reservoir capacity; this indicates how much water reservoirs typically store from one year to the next. The reservoirs in the Tulare region have much less carryover storage, suggesting higher inter-annual variability in reservoir storage.

Table 4.1. List of reservoirs and their key characteristics. Information is taken from CDEC unless otherwise noted. TAF = thousand acre feet; k ac = thousand acres.

ID	Reservoir	Watershed	Region	Capacity (TAF)	Drainage area (k ac)	Use*	Agency**
SHA	Shasta	Sacramento	Sacramento	4552	4270	MULTI, IRR, REC, POW	USBR
ORO	Oroville	Feather	Sacramento	3424.75	2310	MULTI, IRR, MUN, POW	CADWR
BUL	New Bullards Bar	Yuba	Sacramento	969.6	308	DOM, IRR, REC, MUN, POW	YCWA
FOL	Folsom	American	Sacramento	977	1210	MULTI, IRR, REC, POW	USBR
CMN	Camanche	Mokelumne	San Joaquin	417.12	396	STO, MUN, FC	EMBUD
NML	New Melones	Stanislaus	San Joaquin	2400	576	IRR, MUN, POW, REC	USBR
DNP	Don Pedro	Tuolumne	San Joaquin	2030	987	STO, FC, DIV, DOM, IRR	TID
EXC	McClure (Exchequer)	Merced	San Joaquin	1024.6	666	STO, FC, IRR, REC, POW	MID
MIL	Millerton (Friant)	San Joaquin	San Joaquin	520.5	1070	STO, FC, IRR, REC	USBR
PNF	Pine Flat	Kings	Tulare	1000	989	MULTI, IRR, REC	USACE
TRM	Terminus (Kaweah)	Kaweah	Tulare	185.6	359	MULTI, IRR, REC	USACE
SCC	Success	Tule	Tulare	82.3	252	FC	USACE
ISB	Isabella	Kern	Tulare	568	1330	MULTI, IRR, REC	USACE

* Use descriptions as defined on CDEC: MULTI = multi-purpose; IRR = irrigation; REC = recreation; POW = power generation; MUN = municipal; STO = storage; DIV = diversion; DOM = domestic; FC = flood control

** Agencies: USBR = U.S. Bureau of Reclamation; DWR = California Department of Water Resources; YCWA = Yuba County Water Agency; EBMUD = East Bay Municipal Utility District; TID = Tuolumne Irrigation District; MID = Merced Irrigation District; USACE = U.S. Army Corps of Engineers.

4.3.2 Climate and hydrology projections

The flow chart in Figure 4.2 outlines the key models, inputs, and outputs used in this study. We use hydrology simulations that were produced as part of the California Energy Commission Group 1 Data Production for the Fifth Climate Change Assessment and include meteorology from statistically-downscaled (Localized Constructed Analog version 2 - LOCA)

GCMs. LOCA is a statistical method that downscales coarse scale GCM projections to finer scale regional projections (Pierce et al., 2014). It uses observationally-based training data for bias correction and to generate a pattern library. In LOCA version 2, an updated hybrid approach couples the statistical downscaling of LOCA with a pattern library obtained from dynamically-downscaled GCM-Weather Research Forecasting (WRF) runs (Pierce et al., 2023). We select a subset of seven GCMs that best match California climate as determined by high rankings in local climate and process-based metrics per Krantz et al. (2021) (see Text S1 for more). These are: ACCESS-CM2, CNRM-ESM2-1, EC-EARTH3-Veg, FGOALS-g3, MIROC6, MPI-ESM1-2-HR, and MIR-ESM2-0. Here, we treat these GCMs as equally likely realizations of future climate.

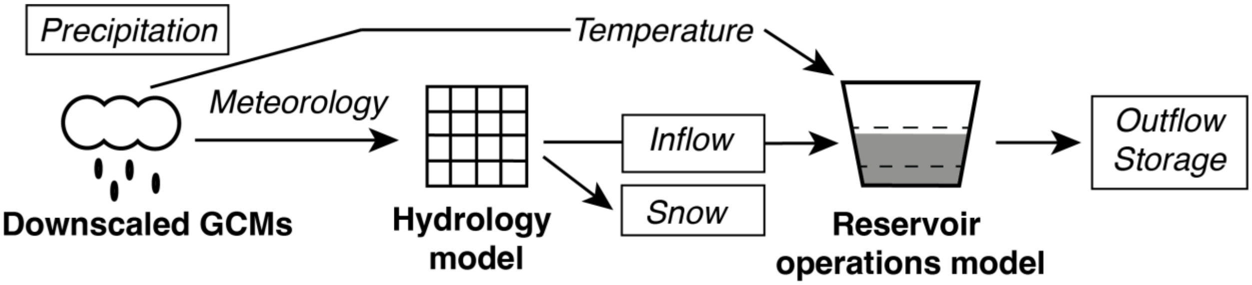


Figure 4.2. Flow chart representation of the methods to generate GCM-driven projections of reservoir outflow and storage, with key variables highlighted in boxes.

A single ensemble member of the downscaled and bias-corrected GCMs are run through historically-calibrated land surface models to generate hydrology projections on a ~ 3 km grid (Bass et al, 2023a). Although both VIC (Variable Infiltration Capacity, Liang et al., 1994) and Noah-MP (Niu et al., 2011) hydrologic simulations were produced by Su et al. (2024) and Bass et al. (2023), respectively, we chose to use the VIC model results here because of the superior performance during validation with observed streamflow over Sierra Nevada watersheds (Su et al., 2024). We compare model results from the historical time period (water years [WYs] 1951-

2014) and under the future climate scenario SSP370, which is a mid-range to high emissions scenario with a radiative forcing of 7.0 W/m2 by 2100 (O'Neill et al., 2016). We partition the future time horizon into three 25-year periods following Siirila-Woodburn et al. (2021): near future (2026-2050), mid century (2051-2075), and end of century (2076-2100).

To run the reservoir operations model (section 4.3.3) with future and historical climate conditions, we compute daily reservoir inflow from VIC modeled runoff. The inflow volume, which represents the unimpaired natural runoff from upstream hydrologic processes, is the product of the daily watershed-average modeled runoff rate and the drainage area (Table 4.1). We do not apply streamflow routing because the short lag times in these watersheds do not impact results at the monthly or seasonal scale. We also extract temperature at the grid cell closest to each reservoir location from VIC meteorology inputs (downscaled GCMs) for input to the evaporation model (section 4.3.3.2). Hereafter, historical conditions are described by model results from the GCMs over the historical time period, except for the calibration and validation of the reservoir operations model which uses observed records.

4.3.3 Reservoir operations model

We develop a reservoir operations model to simulate reservoir storage and outflow on a daily time step in the 13 reservoirs in this study under both historical and future climate conditions. We implement basic historical operating rules related to flood control and water supply. We choose to do this over using established water resources models like CalSim (Draper et al., 2004) because the complexity and computational time of operational models, and the reprogramming required to adapt them to run future climate scenarios, exceeds the needs and purposes of the study. Steinschneider et al. (2023) demonstrate that reduced-complexity systems models are sufficiently accurate for use in the context of climate change vulnerability assessments because the contribution of systems-based model error is small relative to climate-based uncertainties. The reservoir operations model developed for this study applies mass balance with inputs including inflow and temperature (to estimate evaporation, section 4.3.3.2), and simulated outputs including storage and outflow (Fig. 4.3), such as:

$$S_t = S_{t-1} + I_t - E_t - O_t \tag{1}$$

where for day *t*, S is the simulated storage, I is the inflow, E is evaporation, and O is the outflow.

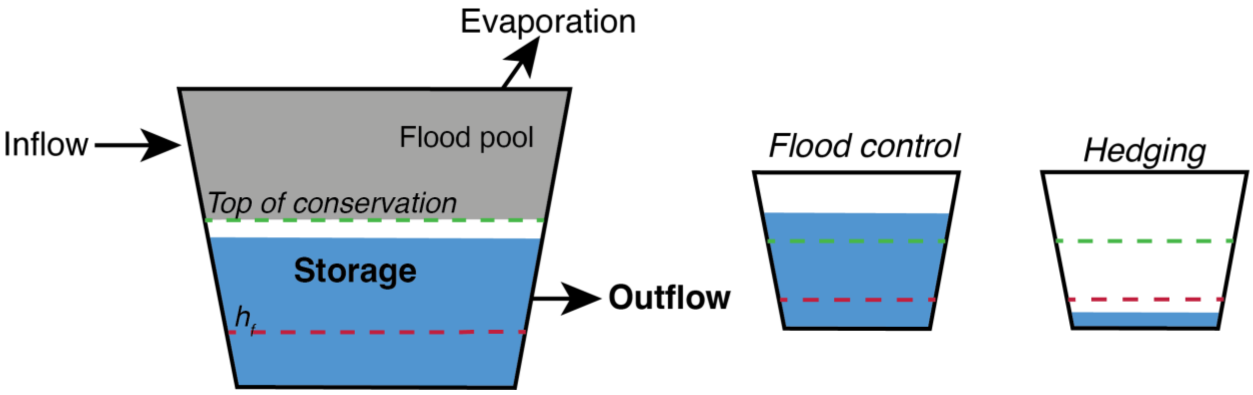


Figure 4.3. Schematic representing the reservoir operations model. Mass balance is applied with daily inflow, evaporation, and outflow volumes to simulate storage. Outflow is determined by reservoir operating rules. The flood pool, top of conservation and hedging threshold h_f for a sample day is indicated. If the simulated storage is above top of conservation, flood control rules are activated. If the simulated storage is below the hedging threshold, hedging rules are activated.

Simulated outflows are determined by reservoir operating rules which prescribe releases based on storage and a target demand. The three rules are: standard linear operating policy (SLOP), flood control, and hedging (described in section 4.3.3.1) (Lund et al., 1996). We determine a target demand (i.e., release) for each day of the water year in each reservoir based on median outflows from the observed record. We suggest that these median outflows are a good approximation for typical historical downstream demands that include irrigation, municipal, and environmental needs (Table 4.1).

We implement a constraint on the maximum daily release allowed from each reservoir and the maximum storage capacity based on maximum values from the observed record. This approximates historical operational constraints. We also include a minimum storage constraint based on the minimum observed value for each reservoir to approximate an operational deadpool storage.

We note that this reservoir operations model does not explicitly include rules related to environmental regulations, recharge, or forecast-informed reservoir operations (FIRO). Instead, we simulate reservoir storage and outflow based on basic water supply and flood control rules. Also, the operating rules and constraints we are using represent a historical strategy; so, we are evaluating how future water supply in and downstream of the reservoirs is affected if water delivery and flood control operating rules remain as they have been historically. In reality, these rules are likely to change in the future as operators adapt to changes in climate and inflows. In this study, we apply the current operating rules and constraints into the future to evaluate vulnerabilities and assess how and where adaptation might be needed the most.

4.3.3.1 Reservoir operating rules

The SLOP rule determines a target release based on demands; if there is enough water in the reservoir to make a release that meets the target demand, the reservoir will release that amount. If the water available in storage is less than the demand, whatever water is available is released.

Hedging restricts these releases when the reservoir storage is below a certain threshold (Fig. 4.3). This reflects a reservoir operator’s choice to strategically reduce releases in favor of keeping water in the reservoir for use later in the water year. The storage threshold at which hedging is triggered and the rate (factor) at which releases are restricted are controlled by thresholds h_1 and h_2. For more details and equations on hedging, see Text C2.

The flood control rules determine how much of the reservoir capacity is available to store water on a given day of the year, and force flood control releases when storage is too high. A seasonal flood pool maintains an unused space in the reservoir that can capture high inflows during the wet season and mitigate downstream flood risk (Fig. 4.3). These requirements are designed and set by the U.S. Army Corps of Engineers (USACE) for the reservoirs in this study based on historical hydrometeorological and storage conditions. We define flood pool curves for each reservoir from USACE Manuals and the literature where available (USACE, 1970; USACE, 1977; USACE, 1980; USACE, 1981; USACE, 1987; USACE, 2004; USACE, 2017b; Zeff et al., 2021). For Camanche, we instead estimate a seasonal flood pool curve from the median daily top of conservation values from the observed record. For Shasta, Oroville, and Folsom reservoirs, we formulate three flood pool curves based on water year type. This is meant to approximate the real flood pool curves which differ based on precipitation-based "wetness" indices. In the model, we define water year type by total inflow, where <30th historical percentile is dry, >70th percentile is wet, and in between is normal. We look at the observed median top of conservation patterns for each water year type and find the best-matching flood pool curve from the manuals. Note that for these water-year dependent cases, we define the day of the year when the flood pool constraint ends from the observed median pattern rather than the published flood pool curves because we find this better corresponds to the observed storage and outflow.

When and if the simulated reservoir storage encroaches on the designated flood pool, flood control releases are prescribed to bring the storage back down (Fig. 4.3). In the model, a flood control factor *fc* determines how much of the flood pool encroachment (i.e., the reservoir storage above the flood pool curve) is released every day until it is cleared, such that:

$$O_t = fc * (S_t - TOC_t) + T_t \tag{2}$$

where for day t, O is the predicted outflow, fc is the flood control factor, S is the simulated reservoir storage (with that day's inflow gain and evaporation loss already incorporated), TOC is the top of conservation (flood pool curve), and T is the target demand. This follows the modeling convention used in Zeff et al. (2021), which sets a flood control factor of 20%.

4.3.3.2 Evaporation losses from the reservoir model

Evaporation losses in the model are estimated from a linear regression that predicts daily evaporation rate from temperature in the grid cell closest to the reservoir location. The predicted evaporation rate is multiplied by the simulated reservoir area to get a volume which is subtracted from the daily mass balance. Reservoir area is predicted from storage by a linear regression based on historical reservoir storage and area records from the Global Reservoir Storage dataset (Li et al., 2023). Note that historical evaporation on average across all the reservoirs is less than 3% of annual reservoir losses from the reservoir (outflow + evaporation). So, although our statistical evaporation model does not consider energetics or meteorological factors other than temperature (like wind and humidity), we suggest that it is appropriate for its purpose here as an input for the reservoir operations model, and its performance compared to historical observations adequate (Table C2). For more details and results on the evaporation model, see Text C3.

4.3.3.3 Model parametrization, calibration, and validation

There are three static parameters in the model that are calibrated for each reservoir based on historical data for the WYs 2001-2023 time period: the factors *a* and *b* which determine hedging thresholds h_1 and h_2 (Text C2), and a flood control factor *fc*. We use grid search parameter optimization to test values between 0 to 1 for each parameter and find the combination for each reservoir that optimized the average Kling-Gupta efficiency (KGE) of simulated weekly reservoir outflow and storage for each reservoir model. This way, we find the parameter values which most closely mimic the operating decisions that determine historical storage and outflows. We validate the model with historical reservoir storage and outflow observations for WYs 2001-2023. For both the calibration and validation exercises, we use historical observed inflow and temperature from the post-processed daily historical forcings (bias-corrected ERA5-WRF; Rahimi et al., 2022) as model input.

4.3.4 Climate and water supply metrics

We compute a suite of annual (water year) metrics describing the climate, hydrology, and water supply for each reservoir and its upstream watershed over both the historical and future time horizons for each GCM (Table 4.2). For each metric over each future time period and reservoir, we compute 25 years * 7 GCMs = 175 projected values. The climate and hydrology metrics are computed from VIC model inputs and outputs. Volumes such as peak SWE (snow water equivalent) or fall and winter precipitation are computed as the product of the watershed-average modeled rate and the drainage area (Table 4.1). The water supply metrics are calculated from the simulated reservoir storage and outflow for each reservoir and each GCM. Region-wide metric values are summed from the individual reservoirs or watersheds in each region.

Table 4.2. List and description of annual metrics. A separate metric is computed for each year, reservoir, watershed, and GCM.

Metric		**Unit**	**Description**
Climate and hydrology	peak SWE	TAF	Basin-integrated (sum) pixel-wise peak SWE
	peak SWE day	day	Basin-integrated (average) day when pixel-wise peak SWE occurs
	OM precip	TAF	Basin-integrated (sum) total precipitation for October-March (fall and winter)
	AS precip	TAF	Basin-integrated (sum) total precipitation for April-September (spring and summer)
	annual inflow	TAF	Total annual (water year) inflow
	inflow centroid	day	Day when 50% of annual (water year) inflow has occurred
Water supply	water deliveries	TAF	June-September total outflow from the reservoir
	end of year storage	TAF	Reservoir storage on September 30th of each water year
	OM outflow	TAF	October-March total outflow from the reservoir
	AS outflow	TAF	April-September total outflow from the reservoir

We note that water deliveries are defined here as June through September total reservoir outflow because this encompasses the most crucial time period for irrigation and municipal water demands. For consistency and simplicity, we hereafter label and refer to metrics by the name of the reservoir at its outlet (i.e., the Upper Sacramento watershed will be referred to as “Shasta” or “sha”).

4.4 Results

4.4.1 Validation of the reservoir operations model

We validate the reservoir operations model with historical observations of weekly reservoir storage and outflow over WYs 2001-2023 (Fig. 4.4, Fig. 4.5). This evaluation period includes a mix of dry, normal, and wet years. We measure the performance of the simulations with the Kling-Gupta efficiency (KGE). The KGE for weekly simulated reservoir storage ranges from 0.57 to 0.92 and averages 0.78 (Fig. 4.4). The daily storage R^2 values are comparable to those achieved with the CALFEWS model (Zeff et al., 2021) and another reservoir simulation model (Steinschneider et al., 2023) over the same reservoirs (Table C3). For weekly outflow, the minimum, maximum, and average KGE is 0.74, 0.83, and 0.93 (Fig. 4.5). Generally, model performance is strongest in the Sacramento region reservoirs (Shasta, Oroville, New Bullards Bar, Folsom; average storage KGE of 0.86, outflow KGE of 0.85), and poorest in the Tulare region (Pine Flat, Kaweah, Success, and Isabella reservoirs; average storage KGE of 0.68, outflow KGE of 0.82) (Fig. 4.4, Fig. 4.5). The reservoirs in the Tulare region are smaller and managed by the USACE primarily for flood control and irrigation deliveries (Table 4.1). Operators may decide to enact flood control releases before the reservoir reaches its top of conservation level based on forecasts or other information; this makes it difficult for the reservoir operations model which does not incorporate forecasts to capture observed drops in storage. The Tulare region also includes recharge operations, where on-demand releases are made for irrigation and groundwater recharge in the wet season (winter and spring). Incorporating recharge operations more explicitly into the reservoir model could improve its performance in these reservoirs.

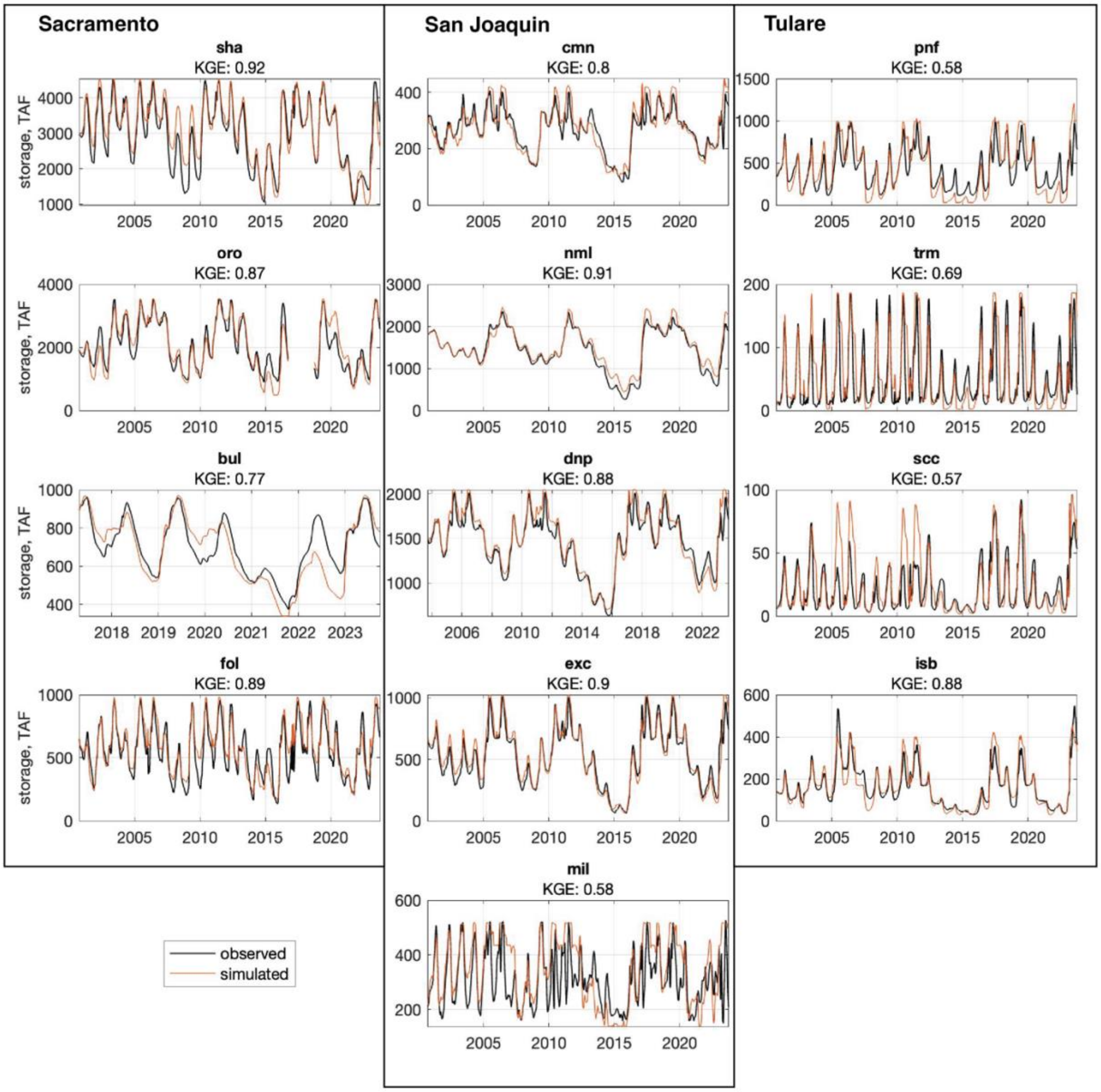


Figure 4.4. Simulated and observed weekly reservoir storage over the validation period.

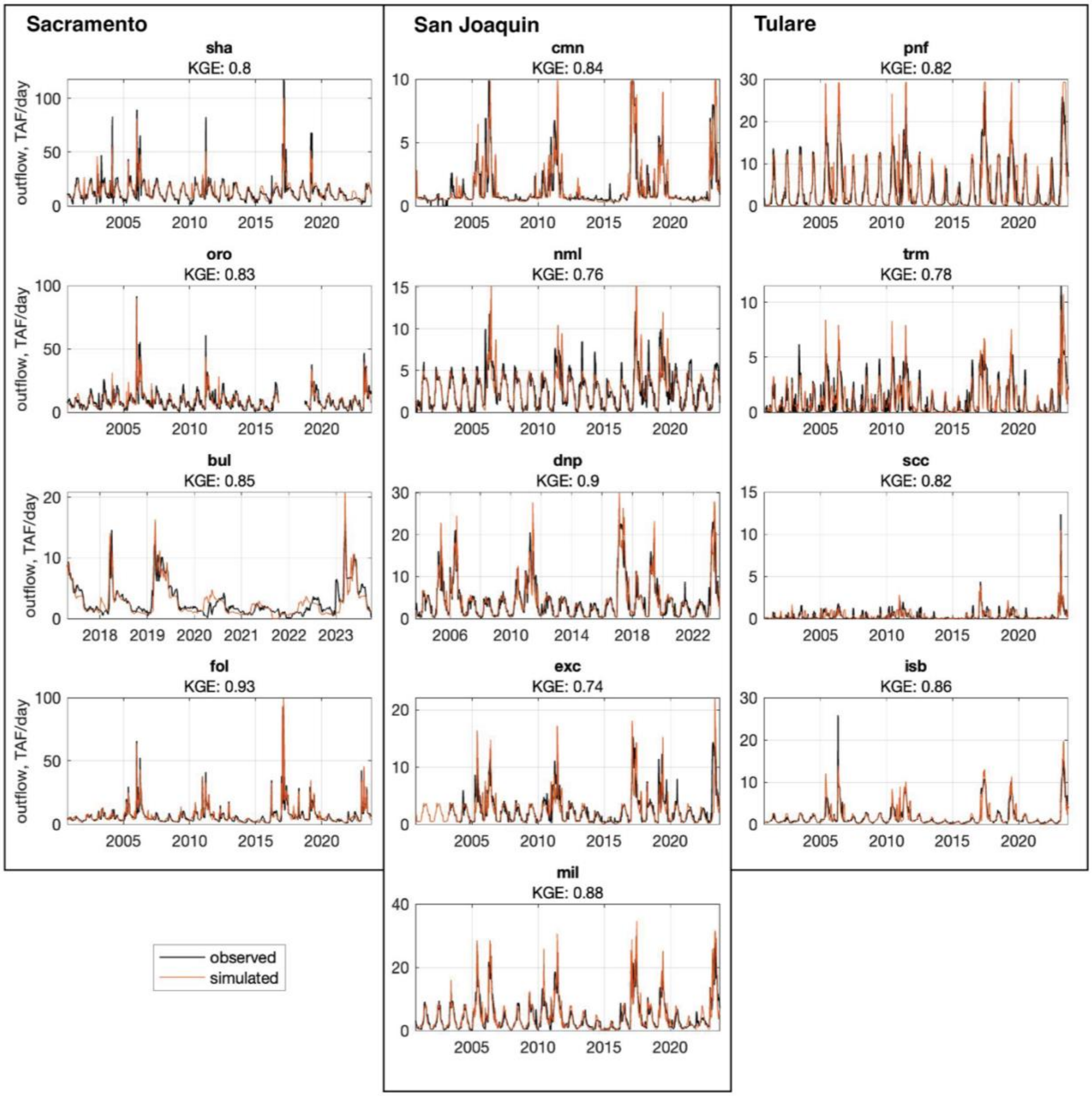


Figure 4.5. Simulated and observed weekly average reservoir outflow over the validation period

Outside of the Tulare region, Millerton reservoir has the poorest performance. Previous studies (Zeff et al., 2021; Steinschneider et al., 2023) similarly recorded poor performance of reservoir operations models here. Millerton is smaller than most other reservoirs and subject to flashy flows especially during the winter. It should be noted that the model simulates outflows from Millerton, which includes releases to the San Joaquin River and both the Friant-Kern and Madera canals, much better (KGE of 0.88) (Fig. 4.5).

We note that due to data availability, the validation of New Bullards Bar reservoir storage and outflow is conducted over the shorter (relative to the other reservoirs) subset of WYs 2017-2023, and of Don Pedro reservoir over WYs 2003-2023. Also, WYs 2017 and 2018 are removed from the validation of Oroville storage and outflow, due to the historic spill and dam crisis that occurred in February 2017.

4.4.2 Projected changes in snow and water supply metrics

We evaluate the projected changes in snow and water supply across Sierra Nevada regions and watersheds by comparing metric values from GCM model results across the future time periods and the historical time period. For each time period, we aggregate metric values across all GCMs and years.

4.4.2.1 Projected changes in snow volume and timing

Total peak SWE over all regions is expected to decrease from the historical average by 6,792 TAF (-44 %, standard deviation ± 31%) by the end of the century (Fig. 4.6a). The greatest loss in snowpack is expected in the Sacramento region (-64%) and the Yuba watershed upstream of New Bullards Bar reservoir (-65%), but all watersheds are projected to lose at least 16% of their peak SWE volume by the end of the century (Fig. 4.6a). By volume (4,654 TAF), the total loss in the Sacramento region is 3 and 7 times greater than the projected losses in the San Joaquin and Tulare region, respectively. The magnitude of projected snowpack declines found here is consistent with previous studies (for example, projected average of 30-64% less by the end of the century under mitigation and business-as-usual scenarios in Sun et al. (2018)).

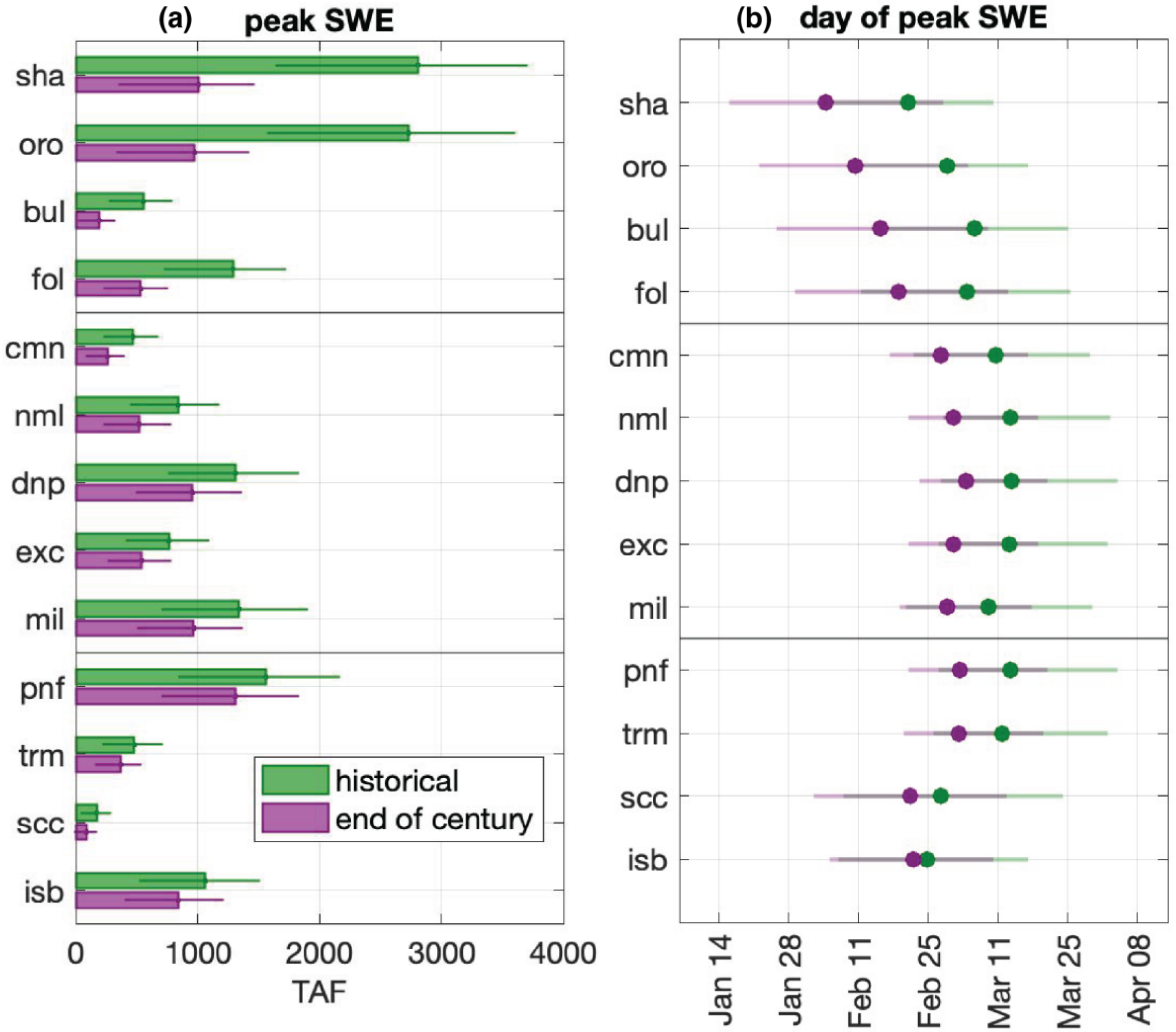


Figure 4.6. Comparison of the (a) peak SWE and (b) day of peak SWE values across the study domain for the historical time period and the end of century period. In (a), the height of the bar corresponds to the average value, and the line indicates the interquartile range (IQR). In (b), lines indicate the IQR for each watershed at the end of the century and historical time periods. Black horizontal lines separate watersheds into three regions.

On average over all watersheds, the timing of peak SWE (indicating snowmelt onset) is expected to occur 11 ± 22 days earlier than the historical average, shifting on average from March 5th to February 22nd. The greatest shifts in timing occur in the Sacramento region, ranging from an average of 14 days (Folsom) to 19 days (New Bullards Bar) (Fig. 4.6b). In contrast, the shifts in timing of peak SWE range from an average of 3 days (Isabella) to 10 days (Pine Flat) in the Tulare region (Fig. 4.6b). Note that peak SWE tends to occur earlier in the Sacramento region than the rest of the Sierra Nevada in both the historical and end of century time period. The significant shifts in timing of peak SWE in the Sacramento region are consistent with the significant decreases in peak SWE volume (Fig. 4.6a).

4.4.2.2 Projected changes in inflow

Summed over all regions, average annual inflow by the end of the century is projected to increase by 771 TAF, or 3% from the historical average (Fig. 4.7a). This small and positive shift implies that rainfall and other seasonal mechanisms (such as evapotranspiration or baseflow/soil moisture) may generate enough runoff to compensate for snowpack declines on an annual scale. The high standard deviation (± 46%) on the domain-wide average indicates that this result is variable and uncertain. The greatest average regional shift is projected for the Tulare region (11 ± 56%), and the least for the Sacramento region (0.5 ± 48%) (Fig. 4.7a). Oroville reservoir is the only one expected to experience a decrease in average annual inflow (-1.5%, Fig. 4.7a).

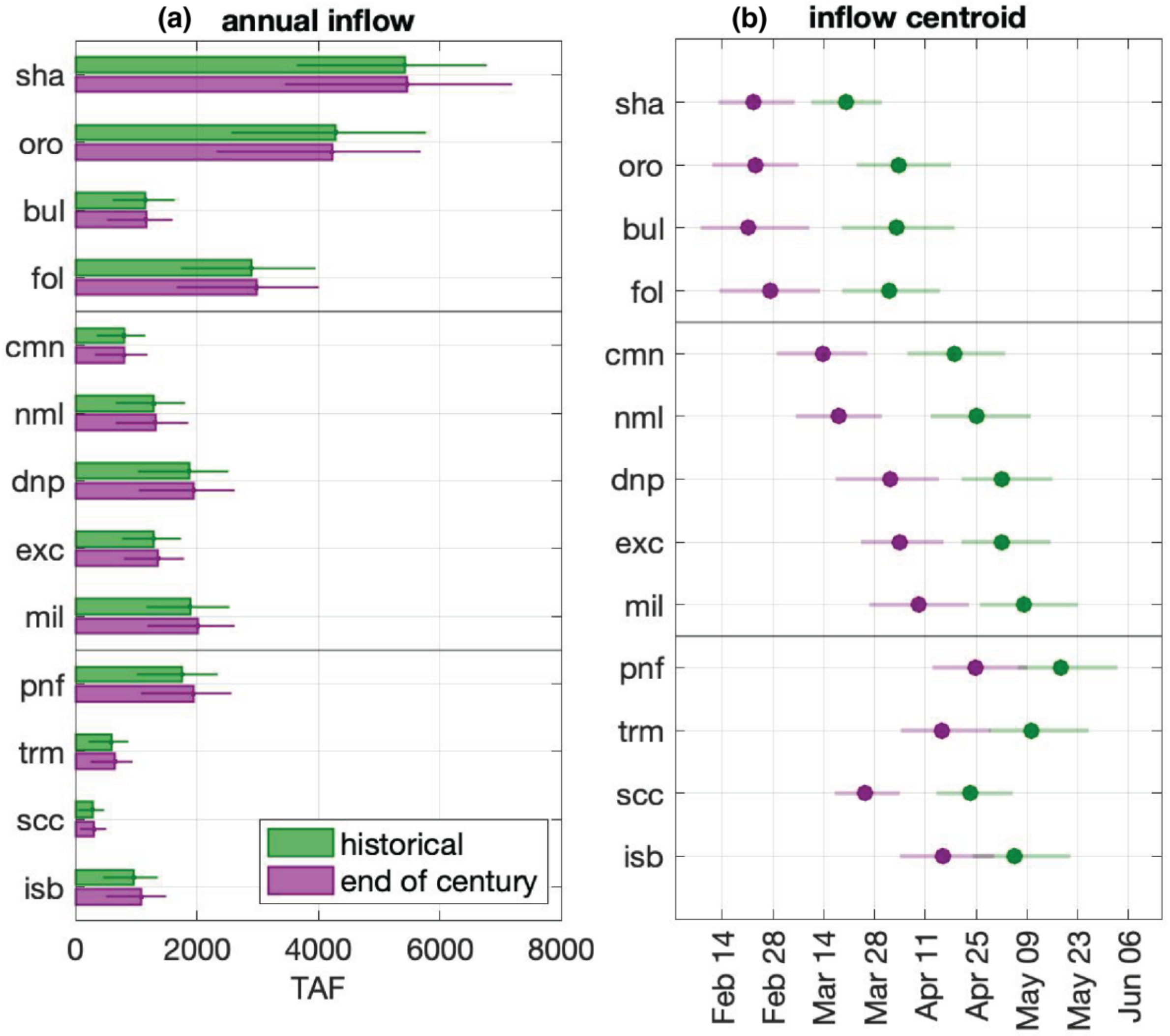


Figure 4.7. Comparison of the (a) annual inflow and (b) inflow centroid values across the study domain for the historical time period and the end of century period. In (a), the height of the bar corresponds to the average value, and the line indicates the interquartile range (IQR). In (b), lines indicate the IQR for each watershed at the end of the century and historical time periods. Black horizontal lines separate watersheds into three regions.

A climate change impact in mountain environments that has been widely reported on is the projected and observed shift in runoff timing (e.g., Rauscher et al., 2008; Schwartz et al., 2017). Here, we observe that the inflow centroid (i.e., the day when half of total annual runoff has occurred) occurs on average 31 ± 16 days earlier by the end of the century (Fig. 4.7b). This number is consistent with the average value reported for the Sierra Nevada under a mitigation scenario in Schwartz et al. (2017). The greatest regional shift is projected in the Sacramento region (average -33 days), and the least in the Tulare region (average -24 days) (Fig. 4.7b). This is consistent with the finding in Fig. 4.6b that snowmelt in the northern watersheds is expected to occur earlier and with a more significant shift relative to the historical than in southern watersheds. The advancement of the inflow centroid is more than that of snowmelt onset (Fig. 4.6b); we suggest this is due to the concurrence of higher winter rainfall and earlier (and less) snowmelt.

4.4.2.3 Projected changes in water deliveries and reservoir storage

On average, total water deliveries (defined as June-September total outflow from the reservoir) are expected to decrease 1,849 TAF at the end of the century from a historical average of 9,993 TAF; this represents a 19% total loss over all reservoirs (Fig. 4.8a), with a standard deviation ± 18%. The greatest average percent loss is expected in the San Joaquin region (-24 ± 23%), then the Sacramento region (-17 ± 14%), and then the Tulare region (-15 ± 32%). All reservoirs are projected to experience reductions in water deliveries, ranging from -7% in Shasta to -51% in Camanche (Fig. 4.8a).

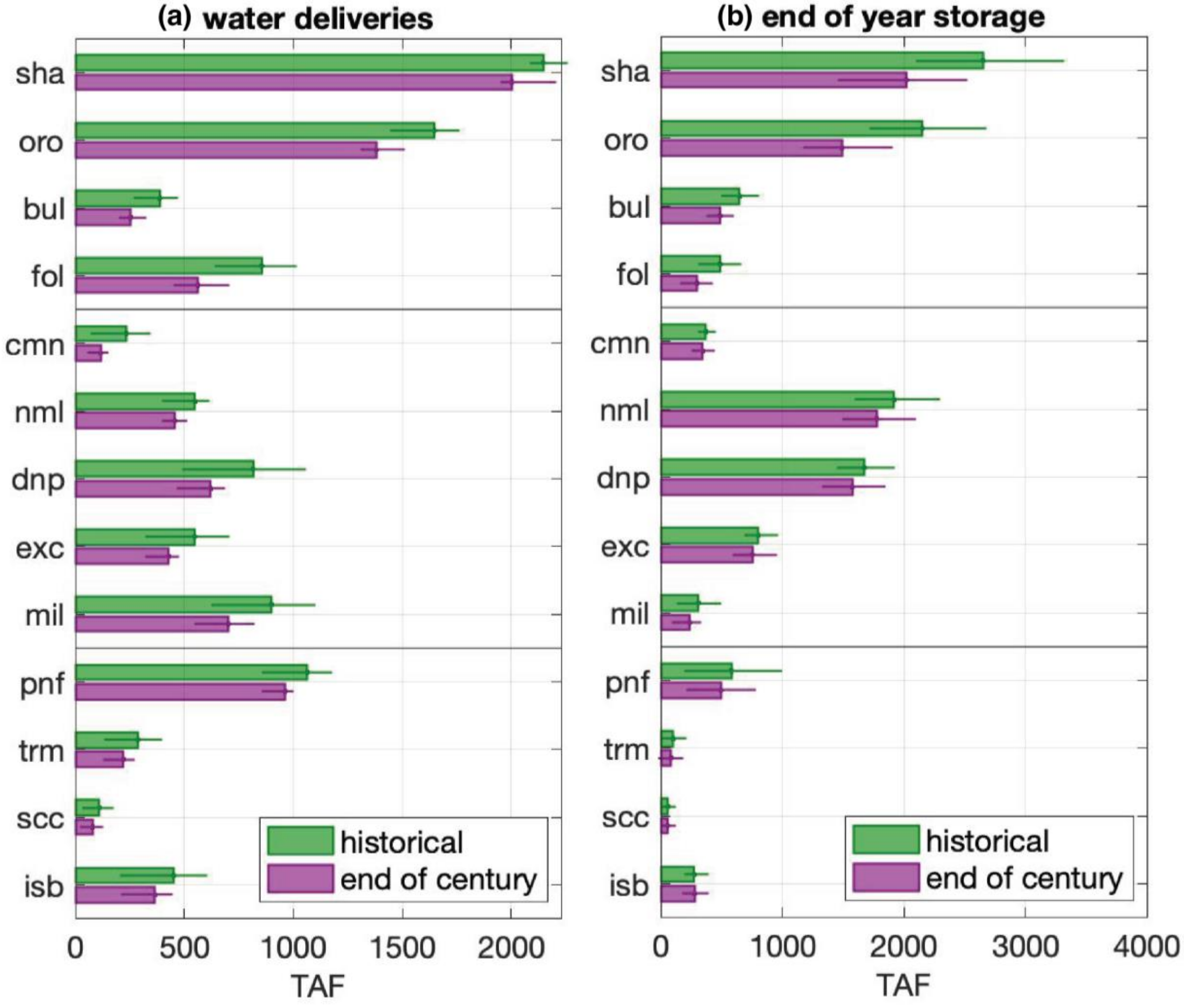


Figure 4.8. Comparison of the (a) water deliveries and (b) end of year storage values across the study domain for the historical time period and the end of century period. The height of the bar corresponds to the average value, and the line indicates the interquartile range (IQR). Black horizontal lines separate watersheds into three regions.

Over the same time period, the total end of year storage across all reservoirs is expected to decrease 2,140 AF from a historical average of 12,041 AF. So, on average, end of year storage would be at 54% of total capacity by the end of the century, down from 66% in the historical time period. This represents a 18 ± 20% total loss relative to the historical. The greatest regional decreases in storage are expected in the Sacramento (-28 ± 22%), followed by the Tulare (-10 ±

56%), and then the San Joaquin (-8 ± 17%) (Fig. 4.8b). All reservoirs except Isabella are projected to experience declines in end of year reservoir storage, ranging from an average of -2.5% in Success to -39% in Folsom by the end of the century (Fig. 4.8b).

The projected median (50th percentile) water deliveries and reservoir storage in the three regions become progressively less through the near future, mid century, and end of century time periods (Fig. 4.9). In most cases, the projected 10th and 90th percentiles also decrease by the end of the century. These percentiles represent the extreme dry and wet years that pose the most significant challenges to water management. These percentiles are computed for each time period from metric values ranging all GCMs and all years. Here, we will focus on the end of the century time period as it represents the most dramatic shifts. Notably, both the 10th and 90th percentile for water deliveries in the Sacramento region decrease more significantly than the median (Table 4.3); this suggests that the extreme dry years will become much more extreme, and the extreme wet years will become much less wet. This is consistent for the two largest reservoirs in the Sacramento region (Shasta and Oroville) (Table 4.3). On the other hand, over the Tulare region, total water deliveries and end of year storage actually increases for the driest years (<= 10th percentile) (Fig. 4.9c; Table 4.3); this suggests that the dry extremes could become less dry in these reservoirs. This is the case for three out of the four reservoirs in the Tulare region (Pine Flat, Kaweah, and Isabella). We suggest that this is due to projected increases in fall and winter precipitation, which are largest in the Tulare region watersheds.

Figure 4.9. Cumulative distribution functions (CDFs) showing water deliveries (left column) and end of year storage (right column) for the historical time period, near future, mid century, and end of century time periods in the (a) Sacramento region, (b) San Joaquin region, and (c) Tulare region. Horizontal lines indicate the 90th, 50th, and 10th percentiles. Region-wide metrics are summed from individual reservoirs in each region.

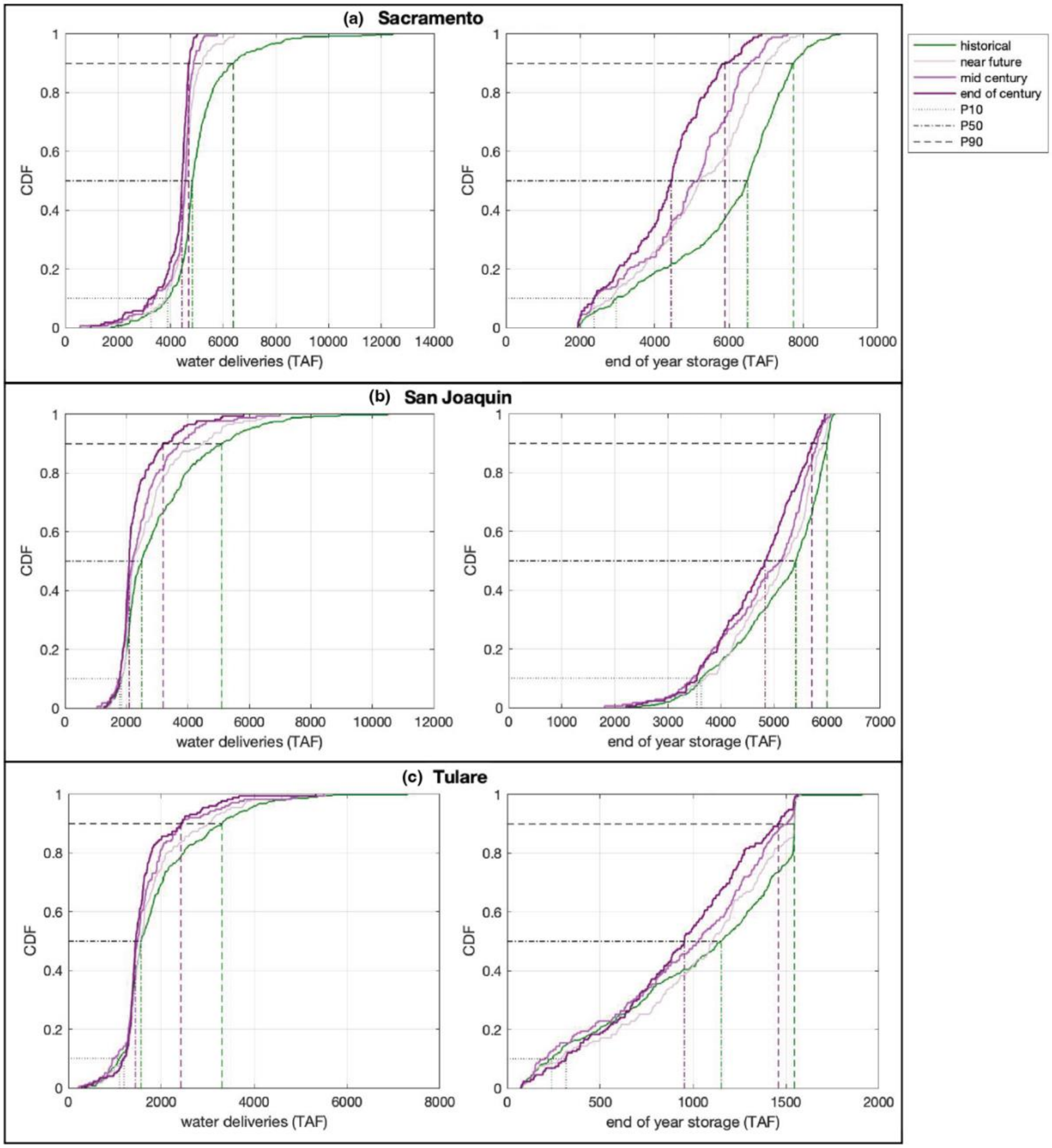

(a) Sacramento
(b) San Joaquin
(c) Tulare
CDF
water deliveries (TAF)
end of year storage (TAF)
0
0.2
0.4
0.6
0.8
1
0
2000
4000
6000
8000
10000
12000
14000
7000
1000
3000
5000
500
1000
1500
2000
historical
near future
mid century
end of century
P10
P50
P90

Table 4.3. Summary of the percent changes in the 10th, 50th, and 90th percentile values for water deliveries and end of year storage in the end of century time period relative to the historical time period, for the three regions and 13 reservoirs.

	Water deliveries			**End of year storage**		
	P10	P50	P90	P10	P50	P90
Sacramento	-16%	-8%	-27%	-20%	-32%	-24%
San Joaquin	-3%	-16%	-37%	-2%	-11%	-5%
Tulare	9%	-8%	-27%	32%	-17%	-6%
sha	-8%	0%	-9%	-14%	-28%	-21%
oro	-13%	-3%	-28%	-34%	-37%	-22%
bul	-18%	-26%	-45%	-18%	-30%	-22%
fol	-26%	-28%	-44%	-21%	-46%	-36%
cmn	-7%	-48%	-62%	-10%	-6%	-1%
nml	0%	0%	-40%	-3%	-10%	-5%
dnp	-4%	-16%	-33%	-2%	-9%	-2%
exc	0%	-17%	-31%	-11%	-8%	-1%
mil	-8%	-16%	-31%	0%	-28%	-20%
pnf	7%	-1%	-23%	62%	-22%	-8%
trm	5%	-13%	-33%	-7%	-30%	-4%
scc	-1%	-22%	-40%	-16%	-1%	-1%
isb	3%	-17%	-28%	44%	-3%	0%

4.4.3 Impact of projected changes in snowpack on water resources

The projected loss of region-wide peak SWE by volume at the end of the century is 3.2 and 3.7 times greater than the projected loss in total end of year reservoir storage and water deliveries, respectively. Across 12 out of 13 individual reservoirs and their watersheds, the projected percent decreases in SWE volume by the end of the century are higher than projected decreases in water supply and end of year reservoir storage. This suggests that the effects of declining snowpack volume on water supply may be attenuated by other factors such as rainfall and reservoir operations.

The future losses in reservoir storage and water deliveries occur despite a slight projected increase in total annual inflow in 12 out of 13 reservoirs; this implies that the current operations are not well suited to manage the projected shifts in timing and source of inflow.

We note that future precipitation totals in California are less certain across GCMs than the projected increases in temperatures. Dettinger et al. (2018) find that there is a 10% tendency towards a wetter Sierra Nevada across 10 observed climate models, but that precipitation trends remain uncertain. We also observe a wide interquartile range for future monthly winter precipitation across all three regions (Fig. 4.10). As such, the noted increases in average winter precipitation should still be considered uncertain.

4.4.3.1 Region-wide patterns

In Figure 4.10, we track how changes in monthly precipitation, SWE, and the resulting inflow can explain the shifts in timing and volume of monthly water availability (reservoir storage and outflow). Across the three regions, we observe a projected increase in winter precipitation, a decrease in SWE, and a shift to earlier snowmelt and inflow by the end of the century (Fig. 4.10). As a result, the monthly pattern of reservoir storage is projected to shift earlier in the water year, peak monthly outflow is projected to increase and occur earlier, and summer outflows are projected to decrease (Fig. 4.10). In effect, the shifts in precipitation and snowmelt amount and timing mean that the reservoirs under current operating rules are less able to capture and store winter and spring inflow for summer water supply. We note that the projected declines in reservoir outflow over the summer (on average, 19%, Fig. 4.8a) are less than the declines in inflow during those months (on average, -49%) in all regions; so, reservoir storage is still helping to provide water supply in those months and attenuate the impact of snowpack declines on water availability outcomes. We find distinctive geographic differences related to shifts in volume and the severity of the changes. Crucially, the lag time between peak monthly precipitation and peak SWE also differs regionally; these dynamics influence the timing of the inflow centroid and how inflow is distributed through the winter and spring. Here, we explore region-wide patterns; we explore reservoir-specific patterns in section 4.4.3.2.

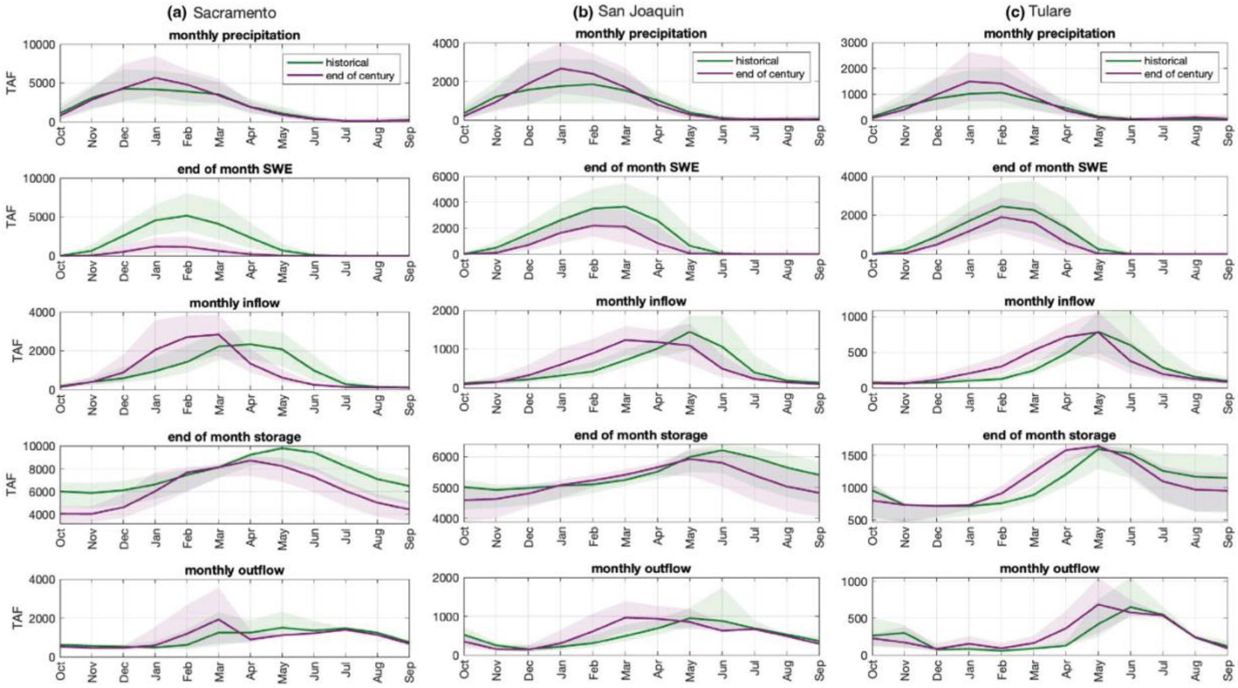


Figure 4.10. Monthly patterns in total precipitation, end of month SWE, total inflow, end of month reservoir storage, and total reservoir outflow for the (a) Sacramento, (b) San Joaquin, and (c) Tulare regions over the historical and end of century time periods**.** Solid lines indicate the median, and the shaded regions indicate the interquartile range (IQR).

In the Sacramento region, we see the most dramatic projected losses in SWE and earliest onset of snowmelt by the end of the century (Fig. 4.6, Fig. 4.10). This explains the significantly lower future monthly inflow in April-June, when historically snowmelt has pulsed through the watershed. That translates to lower-than-historical reservoir storage and outflow in the spring and summer months when key water deliveries are made (Fig. 4.10a). In effect, the historical monthly distribution for reservoir storage in the Sacramento region by the end of the century is lessened by ~1,000 TAF at its peak (April) and double that at its minimum (October), and shifted earlier by one month. Earlier in the water year, there is a significant projected increase in precipitation in January and February (890 – 1,485 TAF) (Fig. 4.10a). Because of the low snow accumulation during these months, we suggest most of this precipitation falls as rainfall and runs off into streamflow. This rainfall pulse coincides with earlier snowmelt as peak precipitation and monthly SWE both occur in January to increase monthly inflow relative to the historical in February and March (Fig. 4.10a). Much of this increased inflow is released from the reservoir for flood control, leading to the increase in February-March monthly outflow (Fig. 4.10a). Peak monthly outflow thus shifts two months earlier than the historical by the end of the century (May to March), and increases by about 400 TAF (Fig. 4.10a).

In the San Joaquin region, projected SWE declines are less severe than in the Sacramento region (Fig. 4.6, Fig. 4.10). As a result, while projected San Joaquin May-September inflow still decreases relative to the historical median, that difference is less than for Sacramento inflow (Fig. 4.10b). The reduced inflow still leads to drops in monthly storage (about -600 TAF) and outflow (maximum loss of 253 TAF in June) in the summer months. This reduces crucial water delivery for irrigation and other downstream needs. The 2-month lag time between projected peak precipitation (January) and peak monthly snowpack (March) in the San Joaquin region leads to future monthly inflow being more distributed over the January to May time period (Fig. 4.10b), as opposed to concentrated in January-March as it does in the Sacramento region. This explains why San Joaquin reservoirs are most full in May (Fig. 4.10b), whereas Sacramento reservoir storage peaks in April (Fig. 4.10a). Mirroring the advancement of snowmelt onset and inflow centroid, the future peak reservoir storage occurs earlier by one month and 64 TAF less in volume than the historical. In February through March, future reservoir storage is projected to be slightly above the historical median (by ~200 TAF) because of higher inflows. Similar to the

Sacramento region, monthly outflow in the San Joaquin region is projected to increase in February-April when flood control releases are made to maintain the historical flood pool under higher inflows (Fig. 4.10b). As a result of these flood releases, peak outflow is projected for March, two months before and 13 TAF higher than the historical May (Fig. 4.10b).

In the Tulare region, the projected increase in winter and spring inflows and decrease in summer inflows, and corresponding changes in reservoir storage and outflow, can similarly be explained by higher future winter precipitation and a lower SWE (Fig. 4.10c). Note here that the projected increases in winter precipitation are (proportionately) largest in the Tulare region, while the projected snowpack declines the smallest (Fig. 4.6, Fig. 4.10c). As opposed to the Sacramento region, there is a 1-month lag time between projected peak monthly precipitation (January) and peak SWE (February) in the Tulare region (Fig. 4.10c). The projected decline in snowpack (and thus snowmelt) is not enough to shift the month when peak monthly inflow occurs (May) in the Tulare region (Fig. 4.10c). The increases in winter precipitation also cause increases in December-March inflow and correspondingly, reservoir storage in February-April and outflow in January-May (Fig. 4.10c). As a result and contrary to other regions, the end of century peak reservoir storage in the Tulare occurs in the same month as the historical (May) and increases slightly in volume (45 TAF) (Fig. 4.10c). Reservoir outflow reaches its peak a month earlier (May) and with 36 TAF more than the historical median (Fig. 4.10c). Similar to other regions, the lower summer inflows (caused by less snowmelt) by the end of the century lead to a projected decrease in reservoir storage in June through October (-81-200 TAF).

4.4.3.2 Reservoir-specific patterns

We observe several common patterns across all reservoirs by the end of the century: projected decreases in peak SWE, earlier timing of snowmelt and inflow centroid, increases in October-March precipitation and annual inflow (except for Oroville), decreases in end of year reservoir storage (except for Isabella), increases in fall and winter outflow, and decreases in water supply (Fig. 4.6, Fig. 4.7, Fig. 4.8, Fig. 4.11c). These changes are already noticeable in the near future time period, and then get more severe by the end of the century (Fig. 4.11). However, the scale of those changes and the influence of projected shifts in climate and snowpack on water resources availability differ by geography (Fig. 4.11c). There is a clear north to south gradient, where the greatest reductions of snowpack and acceleration of snowmelt onset (day of peak SWE) and inflow centroid occur in the northern watersheds (Fig. 4.11). The opposite gradient, where the greatest shifts occur in the southern watersheds, exists for increases in annual inflow and in fall and winter precipitation (Fig. 4.11). Gradients are less clear for water management metrics, due to the individual characteristics and rules of each reservoir.

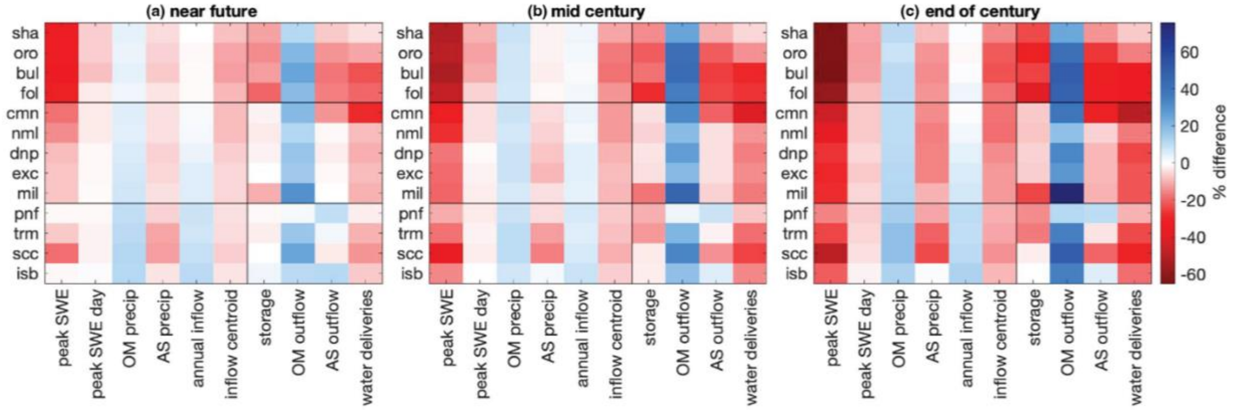


Figure 4.11. Percent changes from the historical average metric values in the (a) near future, (b) mid century, and (c) end of century time periods**.** Negative (red) values indicate a projected decrease in the future. Horizontal lines separate reservoirs in the three regions (top to bottom: Sacramento, San Joaquin, and Tulare). Metrics to the left of

the horizontal line relate to hydrology and climate, and metrics to the right describe reservoir operations and water resources.

In all reservoirs but one (Oroville), average annual inflow is projected to increase by the end of the century while both reservoir storage and water deliveries are expected to decrease. That is, a higher annual water input but reduced ability to meet water supply needs. This implies that the current reservoir operating rules are not well suited to deal with the projected shifts in source (i.e., snowmelt or rainfall) and thus timing of inflows.

Of all the reservoirs, nine (New Bullards Bar, Camanche, New Melones, Don Pedro, McClure, Kaweah, Success, and Isabella) have a greater projected decrease in water deliveries than in end of year storage (Fig. 4.11c). Most of these reservoirs are located in the San Joaquin and Tulare regions. The other five (Shasta, Oroville, Folsom, Millerton, Pine Flat) are instead more susceptible to decreases in end of year storage than water deliveries (Fig. 4.11c). We note that these last five reservoirs have an observed historical summer water supply storage that is greater than the region-wide average (Fig. 4.1b), suggesting a historical operational priority towards summer water deliveries that is carried into the future case. The differences between reservoirs are thus driven by factors such as size and elevation of the watershed and reservoir, reservoir operating rules, and climate (especially, the contribution of snowmelt to total inflow).

Notably, although the watershed upstream of the Shasta reservoir is projected to lose over 60% of its snowpack by the end of the century and average end of year storage is predicted to be 24% less, water supply is only reduced by 7% (Fig. 4.11c). Shasta is the largest reservoir by capacity (Table 4.1), and its watershed is more rainfall-dominated (on a historical average, peak SWE makes up only 22% of the watershed's total annual precipitation, the lowest ratio across all watersheds). This, combined with a coincident projected increase in winter precipitation, makes the reduction in snowpack not as impactful on water supply here. On the other hand, Camanche is projected to experience a 45% decrease in peak SWE by end of century and a 7% reduction in end of year reservoir storage, but a 51% reduction in water supply – the largest decrease across all the watersheds (Fig. 4.11c). Compared to Shasta, Camanche is more snow-dominated and has a much smaller capacity (Table 4.1). We also find that based on the historical calibration of the reservoir operations model, the threshold at which hedging (that is, restricting releases in order to keep water in the reservoir) is triggered in Camanche is much higher than any other reservoirs. So, we suggest that the historical reservoir operating rules in Camanche, particularly around hedging summer releases, are not well-suited to maintain historically-consistent water supply under future climate conditions. We note that Camanche is the reservoir with the least summer water supply storage (Fig. 1b), implying a lower historical priority for summer water deliveries.

The Sacramento region reservoirs are universally projected to experience a greater decrease in end of year storage than most other reservoirs (Fig. 4.11c). We suggest that this is driven by the projected earlier inflow centroid (the projected shifts of which are greatest in the Sacramento reservoirs), which is caused by an overlap of significantly earlier snowmelt and higher fall and winter precipitation for these reservoirs (Fig. 4.11c, Fig. 4.10a). The result is a concentration of inflow in the winter when flood control releases are large and less inflow in spring and summer when water supply releases are prioritized. Historical operating rules may also prioritize the release of summer water supply over maintaining carryover storage (Fig. 4.1b).

The reservoirs and watersheds in the Tulare region show less region-wide consistency in projected changes of both climate and water availability than the reservoirs in the Sacramento or San Joaquin regions (Fig. 4.11c). For example, the watershed upstream of Isabella reservoir is projected to experience no significant change in April-September precipitation, while the watersheds for Success and Kaweah are projected to experience the most severe declines (Fig. 4.11c). Notably for Isabella, the result is no projected change in average end of year reservoir storage by the end of the century (Fig. 4.11c). The Success and Kaweah watersheds are smaller, lower elevation, and more historically rainfall-dominated than the Isabella watershed, and so the projected temperature and precipitation trends will impact them differently. Pine Flat and Isabella are the only reservoirs across the domain to project an increase in April-September reservoir outflow (Fig. 4.11c). This is likely due to higher inflows and corresponding outflows in April-May, when snowmelt and increased spring precipitation coincide. It is worth noting that the performance of the reservoir operations model is more uncertain for reservoirs in the Tulare region due to poorer validation over historical observations (Fig. 4.4, Fig. 4.5).

We suggest that changes in flood control operations to increase storage capacity in winter months could mitigate the projected losses in reservoir storage and summer outflow, especially for reservoirs in the Sacramento and San Joaquin regions. Current flood control rules are designed based on historical patterns in precipitation and inflow to maximize storage capacity in the wet season in order to capture high flows and mitigate downstream flood risk. The structure of those rules is mismatched with a future climate where inflow occurs earlier in the year and is more rainfall- than snow-dominated. For example, in the Sacramento reservoirs, significant flood releases are projected in February and March by the end of the century. Under historical climate conditions, this would have been effective in maintaining reservoir capacity to accommodate snowmelt runoff in April-June, but in the future time period, this prematurely drains down the reservoir while anticipating a larger inflow that never comes. If instead the projected higher inflows in January-March were allowed to accumulate in the reservoir under a reduced flood pool constraint, more water would be available for releases later in the water year. We note that the tradeoff with a reduced flood pool constraint could be increased flood risk; finding operating rules that balance the water supply and flood control needs of a reservoir is a challenge that will continue and grow into the future.

The California Water Plan Update 2023 (CADWR, 2023) acknowledges the urgency and challenges of climate change for California's water systems, and outlines adaptation strategies that include efforts to improve forecasting, data, and management of water resources. A specific strategy is already in place in California for USACE to update reservoir operations manuals and their flood control rules to reflect a changing climate (CNRA, 2022).

4.4.4 Predictive power of snow metrics on water supply

Because of differences in climate, we expect the predictive power of snow metrics on water availability to vary geographically. In Figure 4.12 and Table 4.4, we quantify the Pearson correlation coefficient R between two snow metrics: peak SWE and peak SWE day (snowmelt onset), and two water availability metrics: water deliveries and end of year reservoir storage across the three regions and 13 reservoirs. Note that in the historical time period, and even more so in the future, the correlations between peak SWE and water availability in the Sacramento region are lower than in the San Joaquin and Tulare region (Fig. 4.12). This is explained by the more rainfall-influenced climate in the Sacramento region. Also, the correlations with peak SWE as the independent variable are more significant than with peak SWE day (Fig. 4.12). We acknowledge that the snow metrics are themselves inter-correlated, but quantifying and comparing the differences between the correlations across reservoirs and time periods provides insight into the relative predictive power of these two metrics.

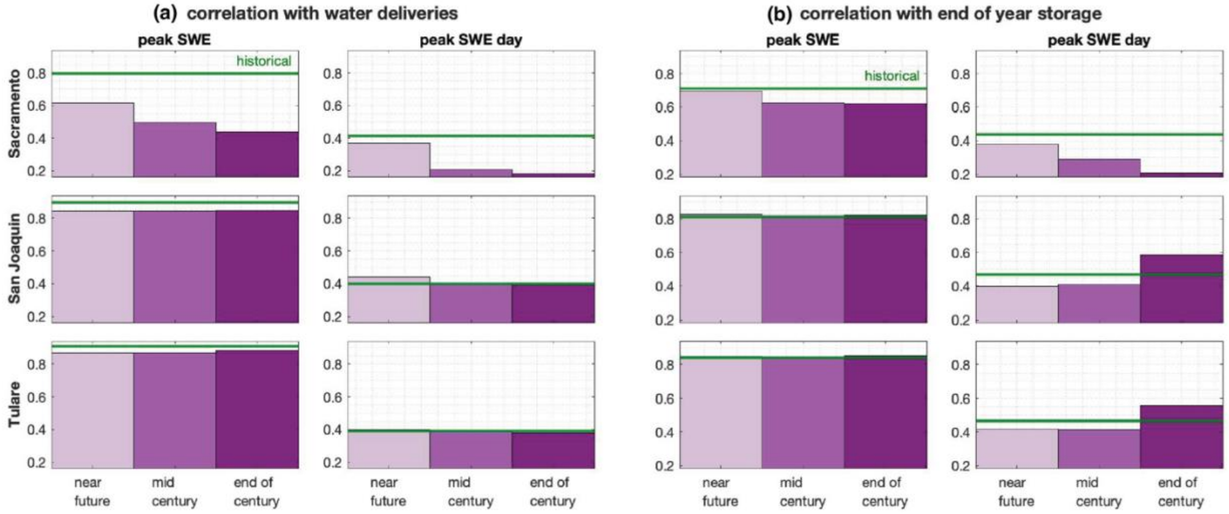


Figure 4.12. Correlation coefficient R quantifies the relationship between snow metrics and (a) water supply and (b) end of year storage across the three regions (rows) and time periods (bars). The horizontal green line indicates the correlations over the historical time period.

Table 4.4. Summary of the correlations between independent variables: peak SWE and day of peak SWE and dependent variables: water deliveries and end of year storage, over the historical and end of century time periods for the three regions and 13 reservoirs. Bolded values indicate the higher value between the historical and end of century time periods.

	Water deliveries					End of year storage			
	Peak SWE		Day of peak SWE			Peak SWE		Day of peak SWE	
	Historical	End of century	Historical	End of century		Historical	End of century	Historical	End of century
Sacramento	**0.8**	0.44	**0.41**	0.18		**0.71**	0.62	**0.43**	0.2
San Joaquin	**0.9**	0.85	**0.4**	0.39		0.81	**0.82**	0.47	**0.58**
Tulare	**0.91**	0.88	**0.39**	0.38		0.84	**0.85**	0.47	**0.56**
sha	**0.62**	0.34	**0.33**	0.2		**0.64**	0.53	**0.37**	0.25
oro	**0.78**	0.33	**0.35**	0.19		**0.7**	0.61	**0.4**	0.25
bul	**0.88**	0.58	**0.45**	0.36		**0.8**	0.67	**0.5**	0.39
fol	**0.85**	0.69	**0.44**	0.39		**0.82**	0.72	**0.47**	0.33
cmn	**0.87**	0.72	**0.43**	0.31		0.59	**0.66**	0.42	**0.47**
nml	**0.79**	0.69	**0.35**	0.3		0.75	**0.77**	0.44	**0.56**
dnp	**0.89**	0.82	0.37	0.37		0.79	**0.81**	0.45	**0.63**
exc	**0.89**	0.8	0.36	0.36		0.72	**0.74**	0.45	**0.62**
mil	**0.91**	0.88	**0.49**	0.47		**0.88**	0.84	**0.49**	0.4
pnf	**0.88**	0.82	0.37	**0.4**		0.87	**0.9**	0.47	**0.57**
trm	**0.92**	0.89	0.4	**0.44**		0.82	**0.83**	0.46	**0.56**
scc	**0.89**	0.81	**0.46**	0.44		**0.46**	0.37	0.36	0.36
isb	**0.9**	0.89	**0.42**	0.36		**0.67**	0.64	0.43	**0.46**

We see how the correlations of the snow metrics with both water deliveries and end of year storage decrease across all reservoirs in the Sacramento region into the future (Fig. 4.12, Table 4.4); this is consistent with the projected declines in snowpack and increases in rainfall triggering different runoff and reservoir operating responses. The decreases in these correlations are generally greater in the Sacramento reservoirs than the other reservoirs; this highlights that the mechanisms through which snowpack affects water supply (end of year storage and water deliveries) are greatly impacted by future climate change here.

The trends for the San Joaquin and Tulare regions and reservoirs are more mixed. In both regions, the future correlations between both peak SWE and day of peak SWE and water deliveries decreases slightly from the historical (Fig. 4.12a, Table 4.4); this is consistent with a projected decline in snowpack that is not as severe as that in Sacramento. In two reservoirs of the San Joaquin region, there is no change in this correlation, and in two reservoirs of the Tulare region, there is instead a slight increase in this correlation by the end of the century (Table 4.4). Conversely to the trend in Sacramento, the correlation between peak SWE and end of year storage increases slightly from the historical in the San Joaquin and Tulare regions (Fig. 4.12b). This is also the case in four out of five reservoirs in San Joaquin and two out of four reservoirs in the Tulare (Table 4.4). The greatest difference (relative to the Sacramento case and the historical) happens in the relationship between peak SWE day and end of year storage. This correlation notably increases by the end of the century in both the San Joaquin and Tulare regions, and in seven out of the nine reservoirs in these regions (Fig. 4.12b, Table 4.4). This signifies that the timing of snowmelt onset is a better indicator for end of year reservoir storage in those reservoirs by the end of the century than in the historical and for reservoirs in the Sacramento region. So, a later peak SWE day more closely correlates with higher end of year storage by the end of the century. We suggest that this is because over the future time period in those regions, the projected decreases in spring and summer precipitation means there is less variance in reservoir storage that is unrelated to changes in snow timing.

4.5 Conclusions

Using hydrology projections driven by seven downscaled GCMs and a simple reservoir operations model, we evaluate how changes in future hydrometeorology (notably, shifts in snowpack volume and timing) impact reservoir storage and water deliveries under current reservoir operating rules across 13 major reservoirs in the California Sierra Nevada.

We find that average total water deliveries (July-September reservoir outflows) are projected to decrease 19% (1,850 TAF) from the historical average by the end of the century (2075-2100) over the entire domain. Total end of year reservoir storage is also expected to decrease, by 18% (2,140 TAF). These losses occur despite a slight increase in domain-wide average annual inflow (+3% or 771 TAF). We find that this happens because, in general, current reservoir operating rules do not sufficiently allow reservoirs to capture the projected earlier inflows for later water supply. The inflow centroid (i.e., the day in the water year when half of annual runoff has occurred) is projected to happen an average of 31 days earlier across all reservoirs; this is driven by earlier snowmelt onset, lower spring snowmelt, and increased (but uncertain) winter precipitation. We find that, on average, region-wide snowpack is projected to decline by 44% (6,792 TAF) by the end of the century, and snowmelt onset accelerates by 11 days.

We find common patterns in projected conditions across all reservoirs: earlier snowmelt onset and lower peak SWE, more winter precipitation (especially January-February), earlier inflow centroid, increase in annual inflow (except for Oroville), decrease in end of year reservoir storage (except for Isabella), increase in fall and winter outflow, decrease in spring and summer outflow (except for Kaweah and Success), and decrease in water deliveries. However, the severity of these shifts differs geographically.

There is a strong north-south gradient in projected hydrometeorology changes by the end of the century, where Sacramento reservoirs have more drastic shifts in snowpack volume and timing and inflow centroid, and Tulare reservoirs have the greatest relative increases in winter precipitation and annual inflow. This follows the gradient from lower elevations and warmer, more rainfall-influenced climate in the north to higher elevations with more snowfall-dominated climate in the south.

However, water supply metrics vary more based on the characteristics and rules (i.e., flood pool timing and size, hedging priorities) of the reservoirs, and the interaction of those rules with upstream hydrometeorology shifts. Nine reservoirs primarily located in the San Joaquina and Tulare regions (New Bullards Bar, Camanche, New Melones, Don Pedro, McClure, Kaweah, Success, and Isabella) have a greater projected decrease in water deliveries than in end of year storage. The other five (Shasta, Oroville, Folsom, Millerton, Pine Flat) are instead more susceptible to decreases in end of year storage than water deliveries. Sacramento reservoirs show greater decreases in end of year storage than other reservoirs. The correlation between snow metrics (peak volume and snowmelt onset) and water supply metrics (water deliveries and end of year storage) decreases into the future in the Sacramento reservoirs, reflecting that the mechanisms through which snowpack affects water availability is greatly impacted by future climate change here. In most reservoirs of the Tulare and San Joaquin regions, the correlation between snowmelt onset and end of year storage instead increases into the future; this implies that there is less variance in reservoir storage in the future here that is unrelated to changes in snowmelt timing.

We note that the implications for reduced summer outflow and reservoir storage range broader than just water supply which was the focus here. Reduced summer outflows can impact ecosystem health such as fish habitat both with lower flows and warmer water temperatures; and changes in reservoir storage can reduce the availability of water for hydropower generation. Lower water deliveries can negatively impact food security and economic stability in regions like the Central Valley of California which relies on snowmelt for irrigation.

We conclude that the impact of projected snowpack declines on water availability from California Sierra Nevada reservoirs is partially attenuated by increased rainfall and reservoir storage, but that the current reservoir operating rules, specifically around seasonal flood pool constraints and hedging releases in the summer, are not optimal for maintaining water supply under a future climate. We suggest that changes in seasonal flood pool constraints to increase storage capacity in winter months could help mitigate the projected losses in reservoir storage and water deliveries, especially for reservoirs in the Sacramento and San Joaquin regions. The effectiveness of adjusting reservoir operating rules, for example reducing flood pool constraints in the winter, for maintaining water supply under a future climate needs to be evaluated. We note that the tradeoffs between protecting against floods and providing water supply becomes more pronounced under a future climate with earlier inflows. Future changes in flood risks, such as single-day peak inflows and outflows, was not assessed in this study but would be an important consideration for reservoir reoperations.

4.6 References

Bass, B., Rahimi, S., Goldenson, N., Hall, A., Norris, J., & Lebow, Z. J., 2023: Achieving Realistic Runoff in the Western United States with a Land Surface Model Forced by Dynamically Downscaled Meteorology. *Journal of Hydrometeorology*, 24(2), 269–283. https://doi.org/10.1175/JHM-D-22-0047.1.

California Department of Water Resources, 2023: California Water Plan update 2023. Retrieved from https://water.ca.gov/Programs/California-Water-Plan/Update-2023.

California Natural Resources Agency, 2022: California water supply strategy: Adapting to a hotter, drier future. Retrieved from https://resources.ca.gov/-/media/CNRA-Website/files/initiatives/water-resilience/CA-water-supply-strategy.pdf.

Cayan, D. R., Dettinger, M. D., Kammerdiener, S. A., Caprio, J. M., & Peterson, D. H., 2001: Changes in the Onset of Spring in the Western United States. *Bulletin of the American Meteorological Society*, 82(3), 399–415. https://doi.org/10.1175/1520-0477(2001)082<0399:CITOOS>2.3.CO;2

Christensen, N. S., Wood, A. W., Voisin, N., Lettenmaier, D. P., & Palmer, R. N., 2004: The Effects of Climate Change on the Hydrology and Water Resources of the Colorado River Basin. *Climatic Change*, 62(1–3), 337–363. https://doi.org/10.1023/B:CLIM.0000013684.13621.1f.

Cohen, J. S., Zeff, H. B., & Herman, J. D., 2020: Adaptation of Multiobjective Reservoir Operations to Snowpack Decline in the Western United States. *Journal of Water Resources Planning and Management*, 146(12), 04020091. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001300.

Dettinger, M., Alpert, H., Battles, J., Kusel, J., Saford, H., Fougeres, D., Knight, C., Miller, L., & Sawyer, S., 2018: Sierra Nevada Summary Report. *California's Fourth Climate Change Assessment* (Publication No. SUM-CCCA4-2018-004). California Energy Commission. Retrieved from https://www.energy.ca.gov/sites/default/files/2019-11/Reg_Report-SUM-CCCA4-2018-004_SierraNevada_ADA.pdf.

Draper, A. J., Munévar, A., Arora, S. K., Reyes, E., Parker, N. L., Chung, F. I., & Peterson, L. E., 2004: CalSim: Generalized model for reservoir system analysis. *Journal of Water Resources Planning and Management*, 130(6), 480–489. https://doi.org/10.1061/(ASCE)0733-9496(2004)130:6(480).

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E., 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016.

Georgakakos, A. P., Yao, H., Kistenmacher, M., Georgakakos, K. P., Graham, N. E., Cheng, F.-Y., Spencer, C., & Shamir, E., 2012: Value of adaptive water resources management in Northern California under climatic variability and change: Reservoir management. *Journal of Hydrology*, 412–413, 34–46. https://doi.org/10.1016/j.jhydrol.2011.04.038.

Gupta, R. S., Steinschneider, S., & Reed, P. M., 2024: Exploring Vulnerabilities in the California Water System under the Late Renaissance Megadrought and Climate Change. https://doi.org/10.22541/essoar.172901487.72592403/v1.

Huang, X., Hall, A. D., & Berg, N., 2018: Anthropogenic Warming Impacts on Today's Sierra Nevada Snowpack and Flood Risk: WARMING IMPACTS ON SNOWPACK AND FLOOD RISK. *Geophysical Research Letters*, 45(12), 6215–6222. https://doi.org/10.1029/2018GL077432.

Klos, P. Z., Link, T. E., & Abatzoglou, J. T., 2014: Extent of the rain-snow transition zone in the western US under historic and projected climate. *Geophysical Research Letters*, 41(13), 4560–4568. https://doi.org/10.1002/2014GL060500.

Knowles, N., Cronkite-Ratcliff, C., Pierce, D. W., & Cayan, D. R., 2018: Responses of Unimpaired Flows, Storage, and Managed Flows to Scenarios of Climate Change in the San Francisco Bay-Delta Watershed. *Water Resources Research*, 54(10), 7631–7650. https://doi.org/10.1029/2018WR022852.

Krantz, W., Pierce, D., Goldenson, N., & Cayan, D., 2021: Memorandum on evaluating global climate models for studying regional climate change in California. *Scripps Institute of Oceanography*. Retrieved from https://www.energy.ca.gov/sites/default/files/2022-09/20220907_CDAWG_MemoEvaluating_GCMs_EPC-20-006_Nov2021-ADA.pdf.

Li, Y., Zhao, G., Allen, G. H., & Gao, H., 2023: Diminishing storage returns of reservoir construction. *Nature Communications*, 14(1), 3203. https://doi.org/10.1038/s41467-023-38843-5.

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.

Lund, J. R., 1996: Developing seasonal and long-term reservoir system operation plans using HEC-PRM. *U.S. Army Corps of Engineers*. Retrieved from https://www.hec.usace.army.mil/publications/ResearchDocuments/RD-40.pdf.

Madani, K., Guégan, M., & Uvo, C. B., 2014: Climate change impacts on high-elevation hydroelectricity in California. *Journal of Hydrology*, 510, 153–163. https://doi.org/10.1016/j.jhydrol.2013.12.001.

Medellín-Azuara, J., Harou, J. J., Olivares, M. A., Madani, K., Lund, J. R., Howitt, R. E., Tanaka, S. K., Jenkins, M. W., & Zhu, T., 2008: Adaptability and adaptations of California's water supply system to dry climate warming. *Climatic Change*, 87(S1), 75–90. https://doi.org/10.1007/s10584-007-9355-z.

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). https://doi.org/10.1038/s41612-018-0012-1.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., & Xia, Y., 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, 116(D12), D12109. https://doi.org/10.1029/2010JD015139.

Null, S. E., Viers, J. H., & Mount, J. F., 2010: Hydrologic Response and Watershed Sensitivity to Climate Warming in California's Sierra Nevada. *PLoS ONE*, 5(4), e9932. https://doi.org/10.1371/journal.pone.0009932.

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., & Sanderson, B. M., 2016: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, 9(9), 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016.

Pierce, D. W., Cayan, D. R., & Thrasher, B. L., 2014: Statistical Downscaling Using Localized Constructed Analogs (LOCA). *Journal of Hydrometeorology*, 15(6), 2558–2585. https://doi.org/10.1175/JHM-D-14-0082.1.

Pierce, D. W., Cayan, D. R., Feldman, D. R., & Risser, M. D., 2023: Future Increases in North American Extreme Precipitation in CMIP6 Downscaled with LOCA. *Journal of Hydrometeorology*, 24(5), 951–975. https://doi.org/10.1175/JHM-D-22-0194.1.

Purkey, D. R., Joyce, B., Vicuna, S., Hanemann, M. W., Dale, L. L., Yates, D., & Dracup, J. A., 2008: Robust analysis of future climate change impacts on water for agriculture and other sectors: A case study in the Sacramento Valley. *Climatic Change*, 14.

Rheinheimer, D. R., Ligare, S. T., & Viers, J. H., 2012: Water-energy sector vulnerability to climate warming in the Sierra Nevada: Simulating the regulated rivers of California's west slope Sierra Nevada. *California Energy Commission* (Publication No. CEC-500-2012-016). Retrieved from https://escholarship.org/uc/item/8h62s4xg.

Rahimi, S., Krantz, W., Lin, Y., Bass, B., Goldenson, N., Hall, A., Lebo, Z. J., & Norris, J., 2022: Evaluation of a Reanalysis-Driven Configuration of WRF4 Over the Western United States From 1980 to 2020. *Journal of Geophysical Research: Atmospheres*, 127(4), e2021JD035699. https://doi.org/10.1029/2021JD035699.

Rauscher, S. A., Pal, J. S., Diffenbaugh, N. S., & Benedetti, M. M., 2008: Future changes in snowmelt-driven runoff timing over the western US. *Geophysical Research Letters*, 35(16). https://doi.org/10.1029/2008GL034424.

Ray, P., Wi, S., Schwarz, A., Correa, M., He, M., & Brown, C., 2020: Vulnerability and risk: Climate change and water supply from California's Central Valley water system. *Climatic Change*, 161(1), 177–199. https://doi.org/10.1007/s10584-020-02655-z.

Rhoades, A. M., Jones, A. D., & Ullrich, P. A., 2018: The Changing Character of the California Sierra Nevada as a Natural Reservoir. *Geophysical Research Letters*, 45(23). https://doi.org/10.1029/2018GL080308.

Schwartz, M., Hall, A., Sun, F., Walton, D., & Berg, N., 2017: Significant and Inevitable End-of-Twenty-First-Century Advances in Surface Runoff Timing in California's Sierra Nevada. *Journal of Hydrometeorology*, 18(12), 3181–3197. https://doi.org/10.1175/JHM-D-16-0257.1.

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

Steinschneider, S., Herman, J. D., Kucharski, J., Abellera, M., & Ruggiero, P., 2023: Uncertainty Decomposition to Understand the Influence of Water Systems Model Error in Climate Vulnerability Assessments. *Water Resources Research*, 59(1), e2022WR032349. https://doi.org/10.1029/2022WR032349.

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.

Su, L., Lettenmaier, D. P., Pan, M., & Bass, B., 2024: Improving runoff simulation in the Western United States with Noah-MP and variable infiltration capacity. *Hydrology and Earth System Sciences*, 28(13), 3079–3097. https://doi.org/10.5194/hess-28-3079-2024.

Sun, F., Berg, N., Hall, A., Schwartz, M., & Walton, D., 2019: Understanding End-of-Century Snowpack Changes Over California's Sierra Nevada. *Geophysical Research Letters*, 46(2), 933–943. https://doi.org/10.1029/2018GL080362.

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.

Tanaka, S. K., Zhu, T., Lund, J. R., Howitt, R. E., Jenkins, M. W., Pulido, M. A., Tauber, M., Ritzema, R. S., & Ferreira, I. C., 2006: Climate Warming and Water Management Adaptation for California. *Climatic Change*, 76(3–4), 361–387. https://doi.org/10.1007/s10584-006-9079-5.

Taylor, W., Brodeur, Z. P., Steinschneider, S., Kucharski, J., & Herman, J. D., 2024: Variability, Attributes, and Drivers of Optimal Forecast-Informed Reservoir Operating Policies for

Water Supply and Flood Control in California. *Journal of Water Resources Planning and Management*, 150(10), 05024010. https://doi.org/10.1061/JWRMD5.WRENG-6471.

U.S. Army Corps of Engineers, 1970: *Oroville Dam and Reservoir: Report on reservoir regulation for flood control*. Sacramento District, U.S. Army Corps of Engineers.

U.S. Army Corps of Engineers, 1977: *Shasta Dam and Lake: Report on reservoir regulation for flood control*. Sacramento District, U.S. Army Corps of Engineers.

U.S. Army Corps of Engineers, 1980: *New Melones Dam and Lake Stanislaus River, California: Report on reservoir regulation for flood control*. Sacramento District, U.S. Army Corps of Engineers.

U.S. Army Corps of Engineers, 1981: *New Exchequer Dam and Reservoir (Lake McClure) Merced River, California: Water control manual*. Sacramento District, U.S. Army Corps of Engineers.

U.S. Army Corps of Engineers, 1987: *Folsom Dam and Lake American River, California: Water control manual*. Sacramento District, U.S. Army Corps of Engineers.

U.S. Army Corps of Engineers, 2004: *New Bullards Bar Dam and Reservoir, North Yuba River, California: Water control manual*. Sacramento District, U.S. Army Corps of Engineers.

U.S. Army Corps of Engineers, 2017b: *Isabella Dam safety modification project temporary water control manual deviation draft supplemental environmental assessment*. Sacramento District, U.S. Army Corps of Engineers. Retrieved from https://www.spk.usace.army.mil/Portals/12/documents/civil_works/Isabella/Project%20Documents/IDSMP_WCM_Dev_SEA7_Sep2017.pdf?ver=2017-09-19-174749-487.

U.S. Army Corps of Engineers, 2017a: *Management of water control systems* (Rep. No. EM 1110-2-3600). U.S. Army Corps of Engineers. Retrieved from https://www.publications.usace.army.mil/portals/76/publications/engineermanuals/em_1110-2-3600.pdf.

VanRheenen, N. T., Wood, A. W., Palmer, R. N., & Lettenmaier, D. P., 2004: Potential Implications of PCM Climate Change Scenarios for Sacramento–San Joaquin River Basin Hydrology and Water Resources. *Climatic Change*, 62(1–3), 257–281. https://doi.org/10.1023/B:CLIM.0000013686.97342.55.

Vicuna, S., Maurer, E. P., Joyce, B., Dracup, J. A., & Purkey, D., 2007: The Sensitivity of California Water Resources to Climate Change Scenarios. *Journal of the American Water Resources Association*, 43(2), 482–498. https://doi.org/10.1111/j.1752-1688.2007.00038.x.

Willis, A. D., Lund, J. R., Townsley, E. S., & Faber, B. A., 2011: Climate Change and Flood Operations in the Sacramento Basin, California. *San Francisco Estuary and Watershed Science*, 9(2). https://doi.org/10.15447/sfews.2011v9iss2art3.

Zeff, H. B., Hamilton, A. L., Malek, K., Herman, J. D., Cohen, J. S., Medellin-Azuara, J., Reed, P. M., & Characklis, G. W., 2021: California's food-energy-water system: An open source simulation model of adaptive surface and groundwater management in the Central Valley. *Environmental Modelling & Software*, 141, 105052. https://doi.org/10.1016/j.envsoft.2021.105052.

CHAPTER 5

Conclusions and Future Work

5.1 Conclusions and key findings

This dissertation addresses two of the most important challenges in snow modeling: uncertainty in model parametrization and mountain precipitation bias. It showcases adjustments to both sources of uncertainty that yield significant improvements in SWE estimation.

The key contribution from Chapter 2 is a thorough and wide-ranging prognosis of the performance of snow-related model physics. Because the model forcings are both constrained by observations and consistent with the validation dataset, SWE estimation errors here are uniquely linked to model physics rather than a mix of model errors and forcing biases. The study includes a large number (199) of sites across varying climates and geographies and tests all snow-related physics options of the Noah-MP model. Key conclusions related to model performance, sensitivity, and differences across sites are:

1) The model's base case configuration, which matches the National Water Model (NWM), generally overestimates accumulated SWE (by 10%), predicts later peak SWE timing (5 days), and underestimates snow melt rate (-15%), likely due to inaccuracies in precipitation partitioning and albedo modeling. Changing the precipitation partitioning threshold, albedo model, and surface resistance to evaporation algorithm reduced errors at most sites.

2) Of the tested physics options, the precipitation partitioning scheme and surface drag parametrization had the most influence on SWE estimates during the accumulation and melt seasons, respectively. But switching out the source of forcing data had the greatest (negative) impact on model accuracy, highlighting the importance of accurate meteorological data for reliable SWE estimates.
3) Model performance and sensitivity vary by region, climate, and vegetation, with colder sites generally showing more accuracy in the accumulation season and warmer stations performing better in the melt season. No single model configuration worked best across all sites.

In Chapter 3, a novel mountain precipitation bias correction method is developed and proved effective for SWE estimation over a test domain. The originality comes from the use of a database of historical precipitation bias correction factors which are a byproduct of a SWE reanalysis framework and had not yet been applied towards SWE modeling in a real-time context. Key findings are:

1) Applying a spatially distributed historically informed precipitation bias correction reduces error (-57-58%), increases spatial correlation (43%), and decreases bias (-85-88%) in April 1st SWE. The strength of this approach lies in its ability to capture the climatological spatial distribution of bias.
2) Assimilation of snow depth observations further reduces error (-45%) by April 1st. However, the assimilation of fSCA measurements prior to April 1st more often degrades than improves SWE estimates due to the weak relationship between fSCA and SWE outside of the ablation season.

3) The improvements in SWE estimation provide value for snowmelt-driven streamflow predictions: using bias-corrected precipitation for modeling April 1st SWE reduces bias in predicted April to July runoff (-46-52%) and improves correlation between predicted snowmelt and observed streamflow (31-39%).

In Chapter 4, we evaluate how changes in future hydrometeorology shifts impact reservoir storage and water deliveries under current reservoir operating rules across 13 major reservoirs in the California Sierra Nevada, using hydrology projections driven by seven downscaled GCMs and an original reservoir operations model. Results reinforce findings from the literature, and additionally contribute a) a comprehensive comparison across the entire north-south gradient of the Sierra Nevada, and b) an update based on the latest climate and hydrology modeling. Key findings are:

1) On average, region-wide snowpack is projected to decline by 44% or 6,792 TAF by the end of the century, snowmelt onset accelerates by 11 days, and the inflow centroid accelerates by 31 days. There is a north-south gradient in projected hydrometeorology changes, where Sacramento reservoirs have the greatest shifts in snowpack volume and inflow centroid, and Tulare reservoirs have the greatest increases in winter precipitation and annual inflow.
2) Average water deliveries are projected to decrease 19% or 1,850 TAF by the end of the century over the entire domain, and total end of year reservoir storage by 18% or 2,140 TAF. These losses occur despite a slight increase in average domain-wide annual inflow (+3% or 771 TAF). We find that this happens because current reservoir operating rules

do not sufficiently allow reservoirs to capture the projected earlier inflows for later water supply.

3) The extent to which reservoirs moderate projected shifts in upstream hydrometeorology varies based the reservoir's characteristics, climate, and operating rules. Nine reservoirs mostly located in the San Joaquin and Tulare regions have a greater projected decrease in water deliveries than in end of year storage. The other five are instead more susceptible to decreases in end of year storage.

5.3 Potential for future work

The findings in Chapter 2 and Chapter 3 provide valuable insights on how to improve SWE estimation by adjusting model physics and bias-correcting input precipitation, respectively. Further research can expand on these methods and explore their scalability and applicability to an operational context. The potential pathways for future work are:

Regarding snow model physics:

1) Explore how the snow model responds to superposing changes in model configurations; for example, applying a new precipitation partitioning threshold and a different albedo model at once.
2) Consider the implications of using different model configurations for different site conditions (such as climate) or regions when running large-scale hydrologic applications like the National Water Model. That way, model physics representation in addition to typical user-defined parameters can be optimized to site conditions.

Regarding mountain precipitation bias correction:

1) Explore how the tested bias correction methods perform in other global mountain regions where historical reanalysis datasets have successfully estimated SWE. The opportunity is there to use historical SWE reanalysis datasets to simultaneously bias-correct and downscale globally available precipitation products in a reliable and real-time fashion.
2) Test more sophisticated methods such as machine learning for bias correction estimation. For example, a bias correction that updates over time based on conditions of the water year as they become known in real time would be useful in an operational context.
3) Evaluate if and how the assimilation of other sources of snow observations – such as ground-based SWE or snow depth measurements, passive microwave or p-band SWE measurements, or satellite-based albedo measurements – could further reduce errors and correct anomalies in SWE estimation when applied in real time.
4) Consider and test the impact of this bias correction on streamflow forecasts through spatially distributed hydrologic modeling and ensemble streamflow prediction (ESP).

To follow up on the assessment of the impact of future snow declines on water availability in California Sierra Nevada reservoirs in Chapter 5, further research directions can:

1) Evaluate the effectiveness of adjusting reservoir operating rules, for example decreasing flood pool constraints in the winter, for maintaining water supply under a future climate. Identify if, when, and how additional storage capacity or reservoir reoperations can increase resilience.
2) Assess future changes in flood risks, such as single-day peak inflows and outflows, especially with reservoir reoperations. We note that under a future climate, the tradeoffs

between maintaining water supply and protecting against flood risks become more pronounced.

3) Evaluate wider-ranging climate change scenarios and compare amongst more and less severe emissions pathways (e.g., SSP585 or SSP250). Future climate remains uncertain and a broader look at the range of possible outcomes can provide additional insight.
4) Identify the specific impacts of temperature or precipitation changes with a sensitivity analysis, for example by increasing historical temperature by incremental amounts and quantifying the resulting changes in water supply.
5) Consider the implications of projected reduced water availability on factors such as ecosystem health, food security, and hydropower generation.

Appendix A

Supplemental information for Chapter 2: Evaluation of Noah-MP snow simulation across site conditions in the Western US

Contents of this Appendix:

Texts A1, A2, A3, A4, A5, A6, A7

Figures A1, A2, A3, A4, A5, A6

Tables A1, A2

References

A1. Eco-regions

We assigned each station to an eco-region based on the Commission for Environmental Cooperation (CEC) Terrestrial Ecoregions Level III classification (Wilken et al., 2011). These eco-regions are defined by both data and expert opinion using a holistic range of diagnostic criteria including soils, physiography, water bodies, major vegetation type, land use and other human influences, and climates. For illustration purposes in this study, certain nearby eco-regions were combined because model behavior was similar. North Cascades, Klamath Mountains, and Cascades were joined to become “Cascades”; Columbia Mountains/Northern Rockies, Idaho Batholith, and Middle Rockies were joined to become “Northern Rockies”; and Northern Basin and Range and Central Basin and Range were joined to become “Basin and Range”. Four regions contained only a single station located close to the region boundary; these stations were added to the most nearby region. A single station in the Colorado Plateau was added to Wasatch and Unita Mountains region; a single station in Wyoming Basin was joined with the Northern Rockies; a single station in the Coast Range was added to the Cascades; and a single station in the Snake River Basin was joined with the Idaho Batholith (Northern Rockies).

A2. SNOTEL Meteorological Records QA/QC

Daily precipitation and SWE values were taken from the bias-corrected quality-controlled data product published by Yan et al. (2018). These data have undergone a three-stage quality control (QC) filter to eliminate outliers and erroneous or inconsistent observations. The quality-controlled precipitation data is then corrected for potential under-catch of snowfall, which has been widely observed at SNOTEL stations due to wind processes and wetting loss on collector walls (e.g., Livneh et al., 2014; Serreze et al., 1999; Sun et al., 2019).

While the Yan et al. (2018) data product also includes quality-controlled and bias-corrected daily temperature records, we chose to instead use hourly temperature data for this study because the Noah-MP model is run at an hourly time step. So, hourly temperature data was downloaded in raw form from the NRCS web portal for over 800 SNOTEL stations. A two-stage QC filter was applied to these records. First, outliers were removed based on global minimum/maximum thresholds of +39 °C and -50 °C (Livneh et al., 2014). Second, following the statistics-based approach used by Serreze et al. (1999) and Yan et al. (2018), values lying outside of +/- three standard deviations from the daily average were removed as outliers. We chose to compute these statistics at the daily level rather than hourly in order to include more data points for each day of the water year.

Only stations with less than 5% missing quality-controlled hourly temperature and daily precipitation and no missing daily SWE over the study period of record (water years 2007-2019) were selected for further study. Data gaps were then filled in order to generate complete records for model input. This process was applied to hourly temperature data over three steps. First, short-term gaps (identified as 5 continuous hours or less) were completed with linear interpolation, following the method in Sun et al. (2019). Second, long-term gaps (identified as 6 continuous hours or longer) were filled in by regressing each station data on nearby stations that have data available during those missing time steps. For this, linear regressions were fitted between each station data and the 10 closest stations that have greater than half of usable hourly temperature records. The neighboring station with the highest R^2 value is used first to predict the missing temperature data. Remaining missing data is filled in by using the station with the next highest R^2 value, and so on. Third, if data is still missing (in this case, an average of 9 hours in about a quarter of the stations), the remaining gaps are filled in with the station's climatological mean for that hour of the water year. Only one of the selected stations had gaps in daily precipitation data – this was filled in by regressing the station's precipitation records with the nearest 5 stations and selecting the one with the highest R^2 value to predict the missing values.

A warm bias at cold temperatures has been noted at SNOTEL stations in numerous studies; this has been attributed to erroneous conversion from voltages to °C (Harms et al., 2016; Currier et al., 2017; Oyler et al., 2015). So, we applied a linear equation that was developed by Harms et al., 2017 and subsequently applied in several studies (e.g., Currier et al., 2017; Sun et al., 2019) to correct this error:

$$\mathrm{T_{corr}} = 1.03 * \mathrm{T_{sntl}} - 0.9 \tag{1}$$

where T_{sntl} is the SNOTEL raw temperature observation (°C).

A3. AORC Forcings

The Analysis of Record for Calibration (AORC) forcings include: precipitation, temperature, specific humidity, terrain-level pressure, downward longwave and shortwave radiation, and west-east and south-north wind components (AORC version 1.1). The dataset is constructed from over a dozen individual datasets, including: North American Regional Reanalysis (NARR), North American Land Data Assimilation System **(**NLDAS-2), and National Centers for Environmental Prediction (NCEP) Global Data Assimilation System (GDAS); and was bias-corrected by gauge-based climatological datasets including PRISM, Livneh et al. (2015), Vose et al. (2014), and Hill et al. (2015). Compared to quality-controlled bias-corrected SNOTEL observations, AORC winter precipitation is on average 10.6% less, with most (82%) stations showing less precipitation in AORC than in the SNOTEL record (Fig. A1a, c). AORC winter temperatures are also on average lower than SNOTEL (average of -0.2°C), but the differences are more heterogenous across stations (Fig. A1b, d).

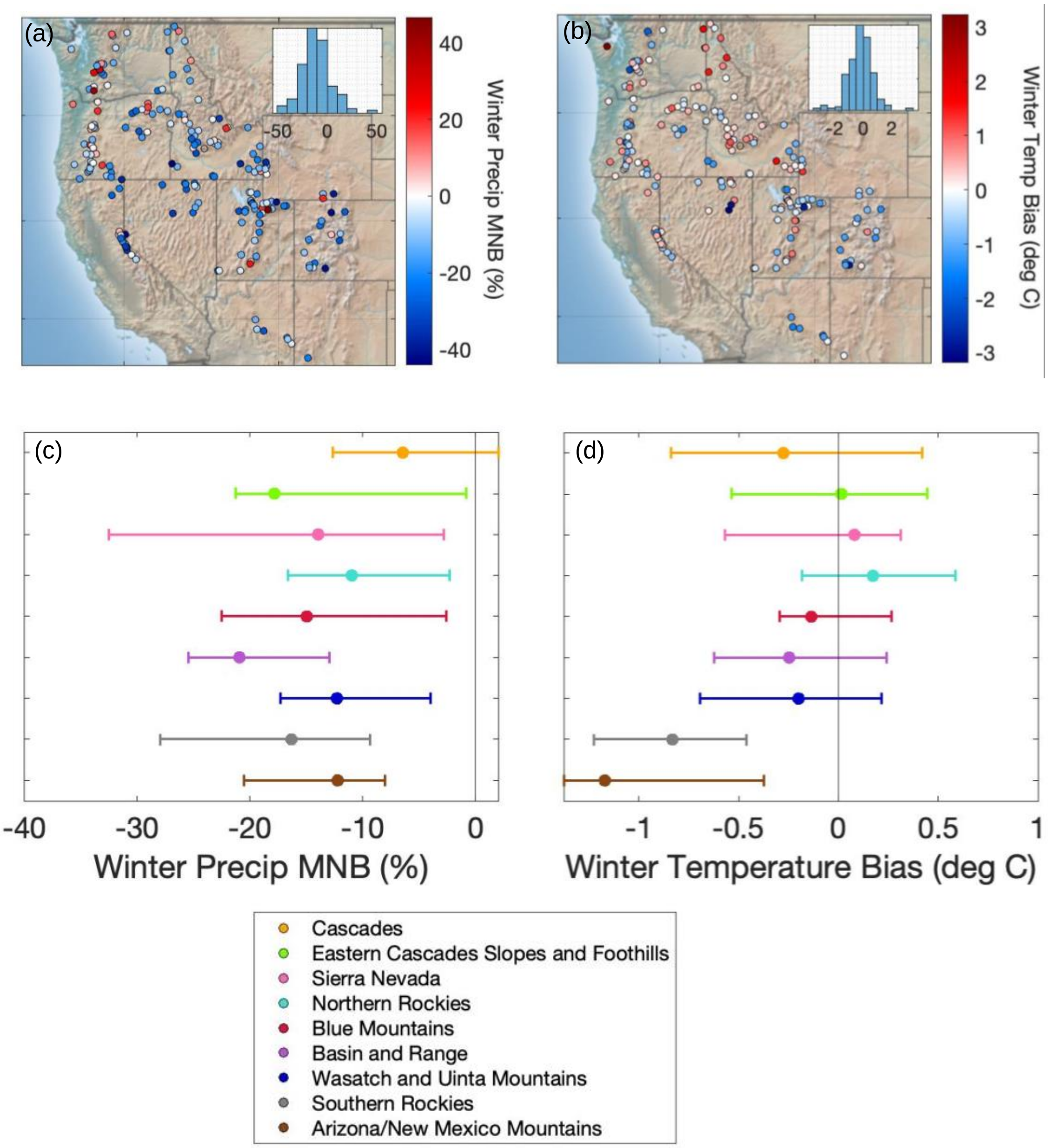


Figure A1. Bias between AORC and SNOTEL winter (Nov-March) precipitation (a) and winter temperature (b) across 199 sites in WYs 2007-2019. Bias in precipitation is computed as the mean normalized bias, and bias in air temperature is computed as the difference. These sites are classified into eco-regions, for which the bias for winter precipitation is shown in (c), and for winter temperature in (d). Circles mark the median of the subgroup, and the width of the line marks the interquartile range.

A4. Physics processes in alternative model experiments

Precip2.2 and Precip0 test alternative options for precipitation partitioning into snow and rain, and should primarily impact snow accumulation by changing input snowfall (Fig. A2). The base case option for snow/rain partitioning defines a prescribed linear snowfall fraction when air temperature is between 0.5 and 2.5 °C (Jordan 1991). Precip2.2 instead sets a fixed threshold for snow at 2.2 °C, while Precip0 uses 0 °C. Note that recent studies have explored precipitation partitioning with wet-bulb temperature rather than air temperature (Wang et al., 2019, Letcher et al., 2022) and have found that this improves model performance; the Wang et al. 2019 wet-bulb temperature-based precipitation partitioning scheme is now included in the latest version of Noah-MP (v5, He et al., 2023).

Alb tests the alternative option for snow surface albedo, impacting snowmelt by changing net radiation (Fig. A2). The base case uses BATS (Biosphere-Atmosphere Transfer Scheme, Dickinson et al., 1986, Yang et al., 1997), which calculates snow albedo for direct and diffuse radiation in visible and near-infrared broadband (Niu et al., 2011). The alternative uses CLASS (Canadian Land Surface Scheme), which computes snow albedo from fresh snow albedo and snow age. BATS with default parametrization has been shown to overestimate snow albedo (Niu et al., 2011; Abolafia-Rosenzweig et al., 2022).

ResisDrag and ResisEvap test alternative options for the surface layer drag coefficient and surface resistance to evaporation/sublimation processes, impacting snowmelt through the computation of surface energy fluxes (Fig. A2). The base case sets the surface resistance to evaporation/sublimation as a constant parameter (rsurf,snow = 50s/m) if the surface is snowy. The alternative used in ResisEvap instead employs the Sakaguchi and Zeng (2009) algorithm for surface resistance for all grid cells regardless of snowiness. This algorithm, defined for NCAR's CLM3.5, describes a surface resistance that explicitly represents the effects of plant litter cover, under-canopy stability, and turbulent resistance; as such, the surface resistance is often above 50s/m but varies by season, water content, and vegetation (Sakaguchi and Zeng 2009). This surface resistance algorithm has been found to decrease modeled latent heat flux in the Western US, but with no significant changes to snow depth (Sakaguchi and Zeng 2009).

The surface layer drag coefficient is determined either by the Monin-Obukhov similarity theory in the base case, or by the original Noah approach (Chen 1997) in ResisDrag. The Chen (1997) approach has been observed to produce a lower surface drag coefficient (e.g., Zhang et al., 2014), which would lead to a higher aerodynamic resistance and lower values for sensible and latent heat fluxes in ResisDrag.

TempSolv and TempLB use alternative options for the lower boundary condition of soil temperature and snow/soil temperature in the model's soil heat flux calculation, respectively, and are expected to impact melt processes via the dissipation of energy in the soil (Fig. A2). The lower soil temperature boundary condition is set by a read-from-file parameter in the base case. TempLB instead prescribes zero heat flux from the bottom of the soil column. The snow/soil temperature time scheme is a solver option rather than a physics option: in the base case, fractional snow cover is considered in the semi-implicit solution to the thermal diffusion equation, whereas it is not considered in the alternative (TempSolv). The thermal diffusion equation affects upper soil and lower snow layer temperatures.

DynVeg, the experiment for the dynamic vegetation option, affects both accumulation and melt processes (Fig. A2). In the base case, the dynamic vegetation module is turned off;

instead, parameters like leaf area index (LAI) and maximum vegetation fraction are based on ground- and satellite- observations. The dynamic vegetation module models prognostic vegetation growth (Dickinson et al., 1998), by combining Ball-Berry photosynthesis-based stomatal resistance with dynamic vegetation and allocating carbon to different parts of vegetation. Vegetation can influence snow processes by: intercepting snow, changing total albedo, changing heat flux with soil temperature, or re-emitting radiation downwards (Park and Park 2016). Based on how the dynamic vegetation module changes the parameters that affect these processes (for example, a larger LAI would intercept more snow), the difference in snow simulation between DynVeg and the base case varies by vegetation type.

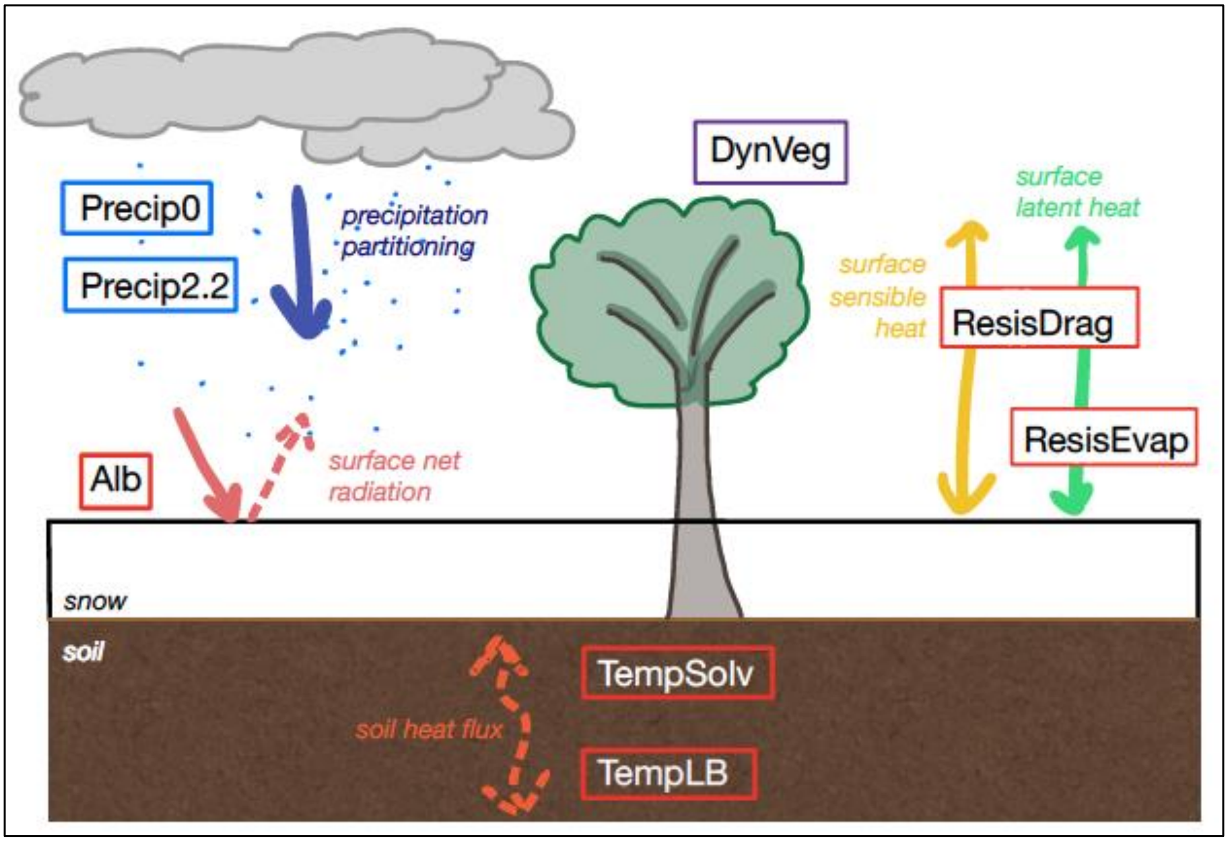


Figure A2. Schematic of a snow model with relevant model physics processes. Experiments tested in this study are boxed and placed near the relevant physics processes. Those labeled with a blue box are ones that primarily impact snow accumulation processes; those with a red box should impact snowmelt processes; and those with a purple box should impact both.

Table A1. List of physics options in Noah-MP, with indicators of usage with WRF-Hydro/NWM. Summarized from Gochis et al. (2018).

	NOAA NWM 2.0 options		WRF-Hydro recommended options	
Option	**Value**	**Definition**	**Value**	**Definition**
DYNAMIC_VEG_OPTION: options for dynamic vegetation	4	off (use table LAI; use maximum vegetation fraction)	4	off (use table LAI; use maximum vegetation fraction)
CANOPY_STOMATAL_RESISTANCE_OPTION: options for canopy stomatal resistance	1	Ball-Berry	1	Ball-Berry
BTR_OPTION: option for soil moisture factor for stomatal resistance	1	Noah (soil moisture)	1	Noah (soil moisture)
RUNOFF_OPTION: options for runoff and groundwater	3	Original surface and subsurface runoff (free drainage)	3	Original surface and subsurface runoff (free drainage)
SURFACE_DRAG_OPTION: options for surface layer drag coeff (CH & CM)	1	M-O	1	M-O
FROZEN_SOIL_OPTION: options for frozen soil permeability	1	Linear effects; more permeable (Niu and Yang 2006)	1	Linear effects; more permeable (Niu and Yang 2006)
SUPERCOOLED_WATER_OPTION: options for supercooled liquid water (or ice fraction)	1	No iteration (Niu and Yang 2006)	1	No iteration (Niu and Yang 2006)
RADIATIVE_TRANSFER_OPTION: options for radiation transfer	3	Two-stream applied to vegetated fraction (gap = 1-FVEG)	3	Two-stream applied to vegetated fraction (gap = 1-FVEG)
SNOW_ALBEDO_OPTION: option for ground snow surface albedo	1	BATS	2	CLASS
PCP_PARTITION_OPTION: options for partitioning precipitation into rainfall & snowfall	1	Jordan (1991)	1	Jordan (1991)
TBOT_OPTION: options for lower boundary condition of soil temperature	2	TBOT at ZBOT (8m) read from a file (original Noah)	2	TBOT at ZBOT (8m) read from a file (original Noah)
TEMP_TIME_SCHEME_OPTION: options for snow/soil temperature time scheme (only layer 1)	3	Semi-implicit; flux top boundary condition, but FSNO for TS calculation (generally improves snow; v 3.7)	1	Semi-implicit; flux top boundary condition
GLACIER_OPTION: options for glacier treatment	2	Ice treatment more like original Noah (slab)	2	Ice treatment more like original Noah (slab)
SURFACE_RESISTANCE_OPTION: options for surface resistance to evaporation/sublimation	4	Sakaguchi and Zeng (2009) for non-snow; rsurf=rsurf_snow for snow (set in MPTABLE); AD v.3.8	1	Sakaguchi and Zeng (2009)

Table A2. Mean characteristics of SNOTEL stations used in study, as grouped by eco-region. Climate and SWE variables are observed historical means for WYs 2007 to 2019. Winter is defined as months Nov-March. Winter precipitation (PPT) is the sum over that period, whereas winter temperature is the average over that period. Accumulated SWE is the sum of positive daily changes in SWE for the entire water year.

							Percent by vegetation type				
Region	**Number of stations**	**Elevation (m)**	**Winter PPT (mm)**	**Winter Temp (ºC)**	**Accumulated SWE (mm)**	**DOWY Peak SWE (day)**	*Barren or Sparsely Vegetated*	*Evergreen Needleleaf*	*Grassland*	*Shrubland*	*Deciduous Broadleaf Forest*
Cascades	39	1282	1233	-0.12	796	169 (Mar 19)	3%	34%	53%	8%	3%
Eastern Cascades Slopes and Foothills	11	1761	549	-1.04	480	157 (Mar 7)	9%	9%	55%	27%	0%
Sierra Nevada	12	2225	886	-0.58	734	168 (Mar 18)	33%	17%	17%	33%	0%
Northern Rockies	42	1975	634	-4.15	624	183 (Apr 2)	2%	21%	50%	24%	2%
Blue Mountains	17	1669	597	-1.81	533	170 (Mar 20)	6%	24%	65%	6%	0%
Basin and Range	24	2247	461	-1.82	480	174 (Mar 24)	29%	17%	21%	29%	4%
Wasatch and Uinta Mountains	30	2563	449	-3.76	512	180 (Mar 30)	0%	10%	60%	10%	20%
Southern Rockies	17	2997	465	-5.25	562	189 (Apr 8)	6%	18%	76%	0%	0%
Arizona/New Mexico Mountains	7	2500	375	1.75	264	135 (Feb 13)	14%	29%	43%	14%	0%

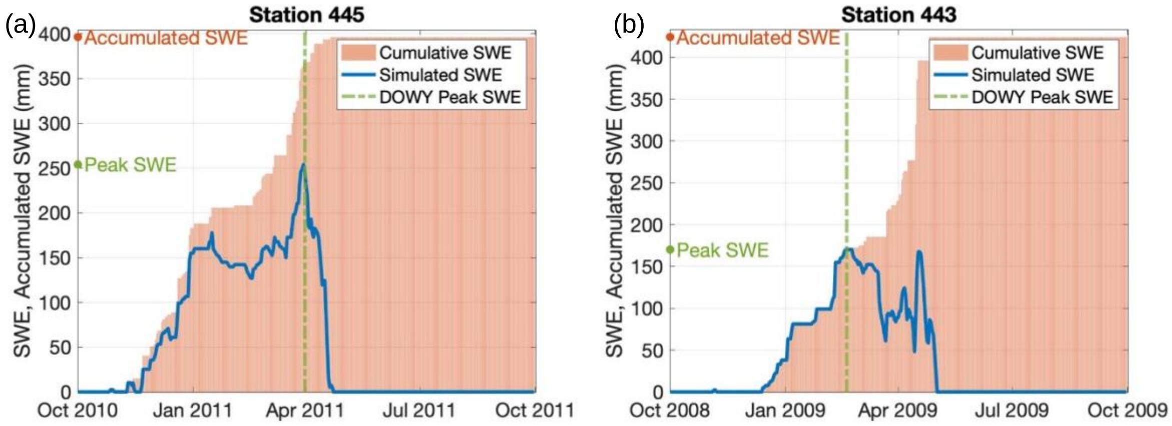


Figure A3. Illustrations of observed SWE from SNOTEL station-years when a single-day snow metric like peak SWE significantly underestimates the totality of SWE being produced over the course of the water year. DOWY for peak SWE is indicated with a vertical green line, and peak SWE depth is indicated with a blue dot on the y-axis. Cumulative accumulated snowmelt is illustrated with orange bars and yearly accumulated SWE is marked with an orange dot on the y-axis. (a) WY 2011 at SNOTEL station 445 shows a case of pre-peak SWE melt. (b). WY 2009 at SNOTEL station 443 shows a case of post-peak SWE accumulation.

A5. Removing propagated uncertainty from daily changes in SWE

The daily accumulation and melt rates are computed as the average positive or negative changes in daily SWE over the water year. Because we are comparing modeled results to observed measurements, we have to acknowledge differences in precision between the two. The SNOTEL snow pillow's precision is 0.254 mm, which we take as the detection limit. The propagated uncertainty from this single-day detection limit into values of daily changes (i.e., $SWE_{n+1} - SWE_n$) is computed as:

$$s_{\Delta SWE} = \sqrt{\left(s_{SWE_n+1}\right)^2 + \left(s_{SWE_n}\right)^2} \tag{2}$$

based on the rules of propagating uncertainty when adding or subtracting measurements (i.e., Kirchner, J. 2001). Here, we are replacing the typical standard deviation with the daily detection limit when describing the original measurement uncertainty. So, with a 2.54 mm uncertainty in the original measurement, the propagated uncertainty is 3.58 mm. Daily SWE changes below this value were set to zero in both simulated and observed records to maintain consistency, because the in situ sensor may not be capable of deriving differences below this level.

A6. Application of Kolmogorov-Smirnov (KS) test to evaluate model sensitivity

We applied the Kolmogorov-Smirnov (KS) test to evaluate model sensitivity. This test has been utilized to assess model sensitivity for numerous hydrology model studies (e.g., He et al., 2011; Sun et al., 2019). In general, this statistical test is used to decide if a sample comes from a population with a specific distribution. Applied as a two-sample test, KS can be used to test whether two underlying probability distributions differ. In this case, the KS statistic is computed as the maximum vertical distance between the two empirical distribution functions:

$$\mathrm{KS} = \sup_{x} \left| F_{1,n}(x) - F_{2,m}(x) \right| \tag{3}$$

where sup is the supremum function, and $F_{1,n}$ and $F_{2,m}$ are the empirical distribution functions of the first and second sample (Chakravarti et al., 1997).

In this study, for each station and for each snow metric, we apply a two-sample KS test with the yearly snow metric values from the base case and from the experiment as the two sets of inputs. For example, Fig. A4 illustrates how the KS statistic is computed at Station 302 for the accumulated SWE metric, between the base case and Precip0. The empirical distributions include all yearly accumulated SWE metrics for the base case at that station in blue, and for the Precip0 experiment in orange. The maximum distance between the curves is indicated with the black arrow, and equals the KS statistic. KS values range from 0 to 1, with higher values indicating greater sensitivities. We used a minimum KS threshold value of 0.5 to identify sensitivity because it yields statistically significant results at p-value < 0.1. So, stations with a

KS statistic equal to or greater than 0.5 were considered sensitive to that alternative model configuration for that snow metric.

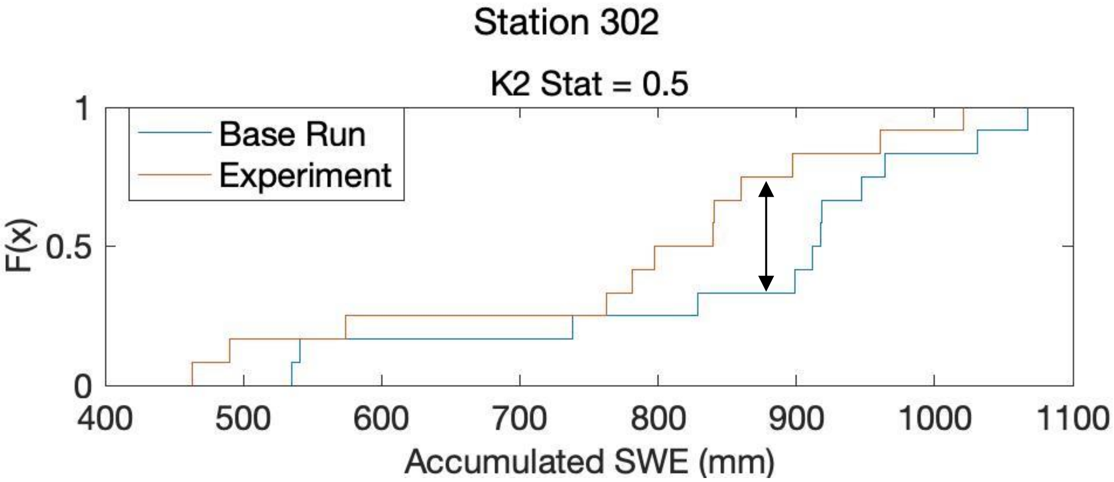
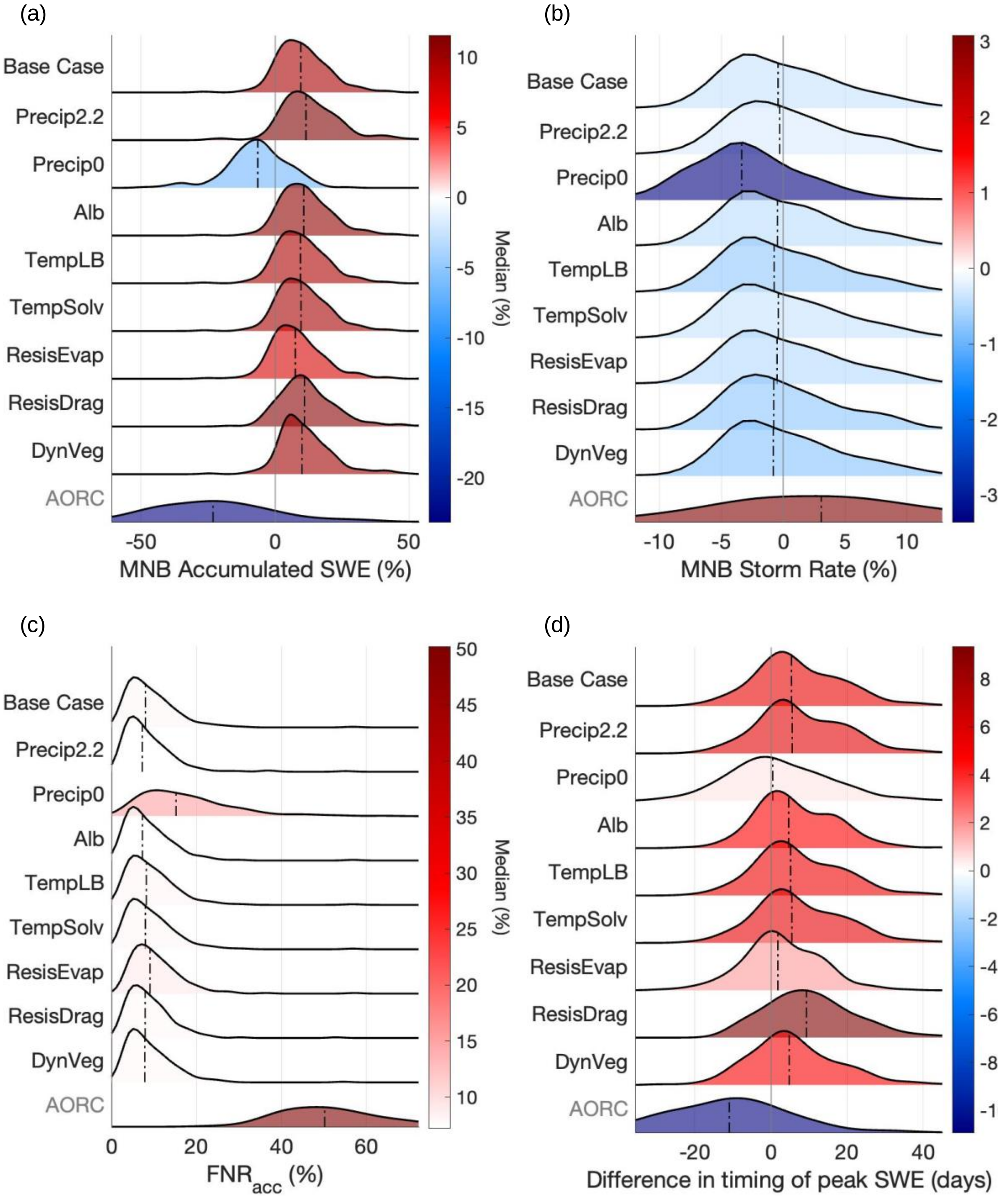


Figure A4. Example of computation of KS statistic, on the accumulated SWE metric and between the base case and Precip0 experiment at Station 302. The black arrow indicates the value of the KS statistic.

A7. Model performance across model configurations

Noah-MP predictions tend to overestimate observed accumulated SWE across most sites and for all model configurations except for the Precip0 and AORC experiments (Fig. A5a). All model configurations but the AORC experiment shows a model underestimation in storm rate, and a model overestimation in timing of peak SWE (Fig. A5b, d). The AORC experiment, and to a lesser degree, the Precip0 experiment, shows a high FNR for accumulation days (median of 50% and 15%, respectively), whereas the rest of the experiments have FNRs of less than 10% (Fig. A5e). The model consistently underestimates daily average melt rate across all model configurations (Fig. A5f). AORC and Precip0 show the highest FNR for melt days, but the FNR for melt days is higher across all experiments (between 35% and 54%) (Fig. A5f), suggesting that the model fails at simulating observed melt events more frequently than it does for accumulation events.


(a)
Base Case
Precip2.2
Precip0
Alb
TempLB
TempSolv
ResisEvap
ResisDrag
DynVeg
AORC
-50
0
50
MNB Accumulated SWE (%)
10
5
0
-5
-10
-15
-20
Median (%)
(b)
Base Case
Precip2.2
Precip0
Alb
TempLB
TempSolv
ResisEvap
ResisDrag
DynVeg
AORC
-10
-5
0
5
10
MNB Storm Rate (%)
3
2
1
0
-1
-2
-3
Median (%)
(c)
Base Case
Precip2.2
Precip0
Alb
TempLB
TempSolv
ResisEvap
ResisDrag
DynVeg
AORC
0
20
40
60
FNR_{acc} (%)
50
45
40
35
30
25
20
15
10
Median (%)
(d)
Base Case
Precip2.2
Precip0
Alb
TempLB
TempSolv
ResisEvap
ResisDrag
DynVeg
AORC
-20
0
20
40
Difference in timing of peak SWE (days)
8
6
4
2
0
-2
-4
-6
-8
-10
Median (days)

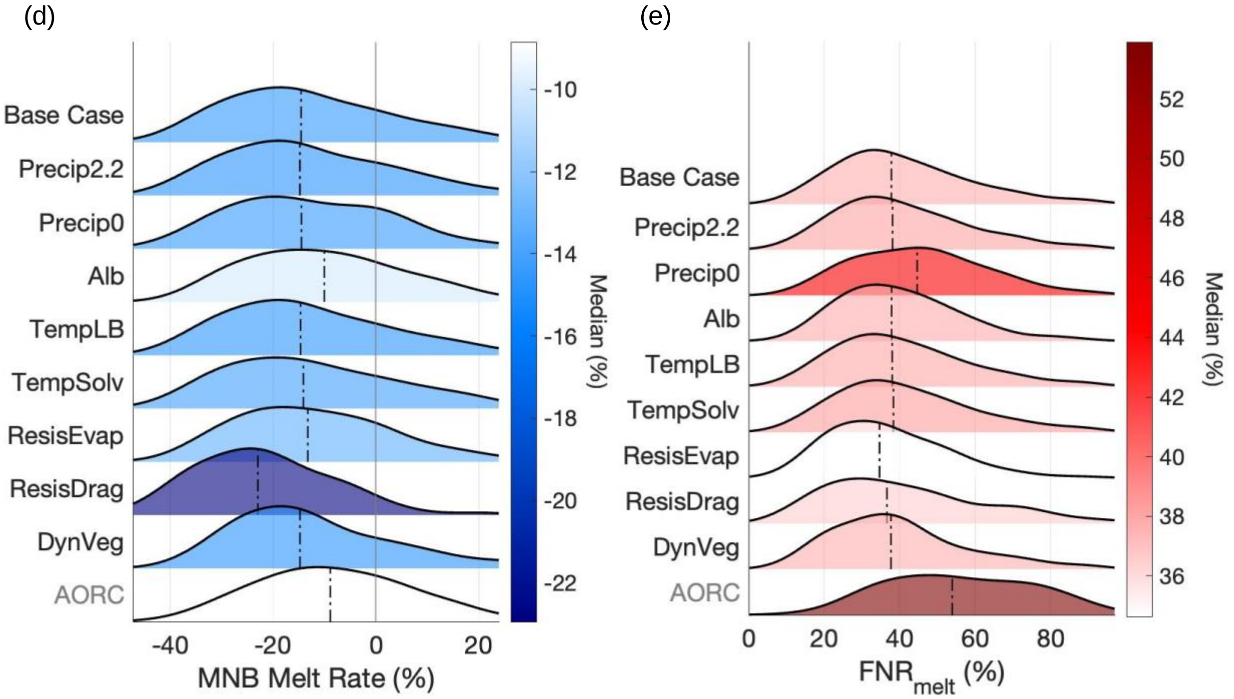


Figure A5. Distributions of model performance across SNOTEL stations and nine model configurations for four snow metrics. (a) mean normalized bias (MNB) of accumulated SWE, (b) MNB of storm rate, (c) average false negative rate (FNR) for accumulation days, (d) difference in timing of peak SWE, (e) MNB of daily melt rate, and (f) average FNR for melt days. Bias metrics are computed for each station-year with reference to the observed SNOTEL records, and then averaged for each station over the time period. The color of the distribution and dashed horizontal line corresponds to the median metric value for each model configuration. A blue (red) color indicates the model configuration produces a lower (higher) median metric value than the observation.

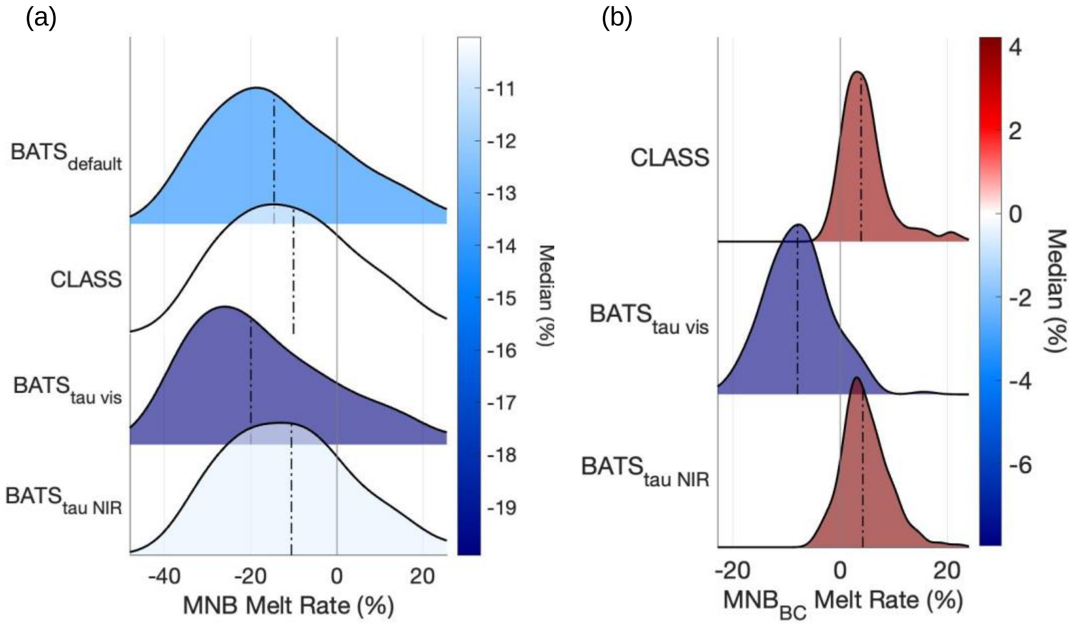


Figure A6. (a) Distributions of model performance across SNOTEL stations and four model configurations related to snow albedo, as described by mean normalized bias (MNB) of melt rate. BATS$_{\text{default}}$ refers to the Base Case (Table 2.2 in main text) with all default parameters. CLASS refers to the Alb experiment, which utilizes the CLASS albedo scheme. BATS$_{\text{tau\_vis}}$ and BATS$_{\text{tau\_NIR}}$ refer to experiments with the BATS albedo model but with the snow age parameter τ_0 adjusted to values optimized in Abolafia-Rosenzweig et al. (2022). A blue (red) color indicates the model configuration produces a lower (higher) median metric value than the observation. (b) Distributions of changes in the MNB of melt rate relative to the base case across SNOTEL stations and the three alternative snow albedo-related experiments. Bias metrics (MNB$_{\text{BC}}$, Table 2.2) are computed for each station-year with reference to the base case, and then averaged for each station over the time period. The distribution color and the dashed horizontal line correspond to the median bias value for each experiment. A red (blue) color indicates the model configuration produces a higher (lower) median value than the base case.

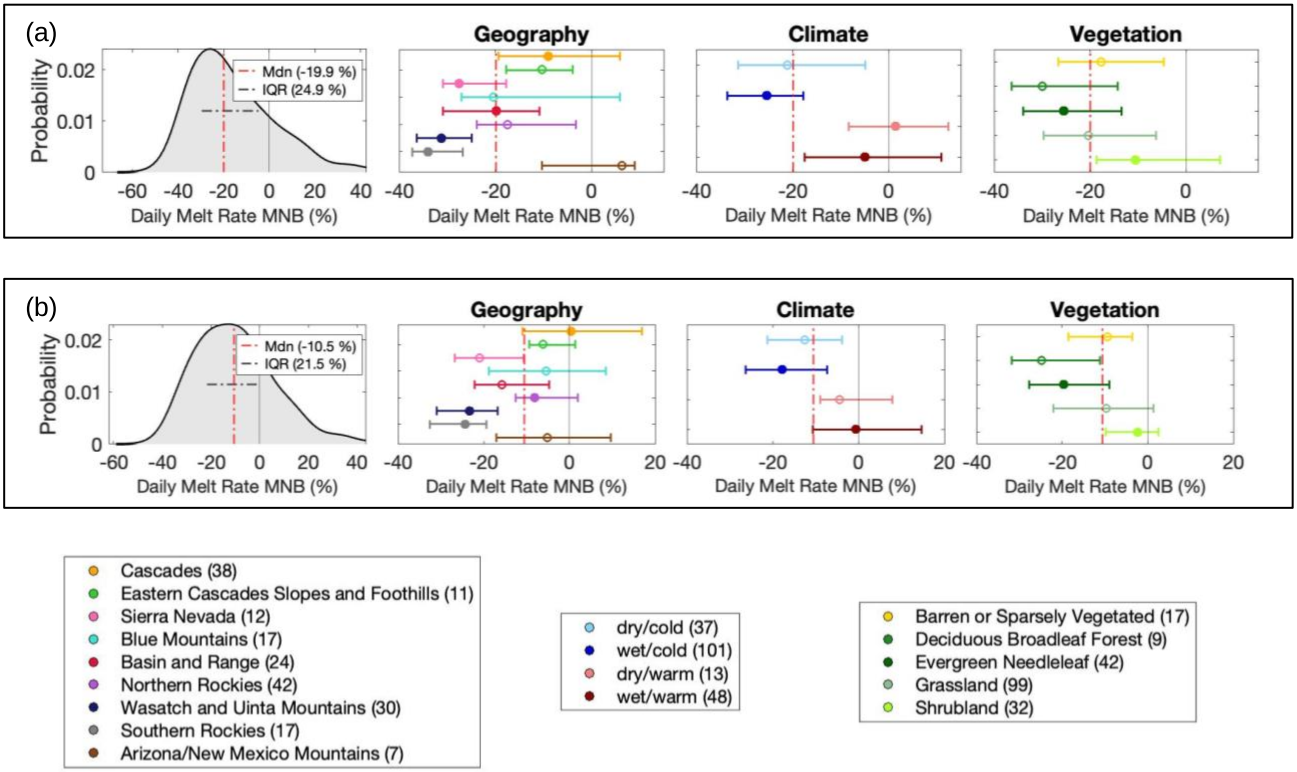


Figure A7. Model performance (MNB in daily melt rate) across all stations in (a) $\text{BATS}_{\text{tau\_vis}}$ and (b) $\text{BATS}_{\text{tau\_NIR}}$, as compared to SNOTEL SWE observations. The leftmost column of panels shows a smoothed histogram of the performance metrics across stations. A vertical dashed red line indicates the median metric value, and a horizontal dashed gray line indicates the interquartile range (IQR). The performance metrics are separated by geographic region in the second column; by climate subgroup in the third; and by vegetation type in the fourth. Circles mark the median of the subgroup, and the width of the line marks the interquartile range. If the subgroup has a filled-in circle, it is considered significantly different (p-value < 0.05) from the other subgroups. The number of stations in each subgroup is noted in the legend entries.

References

Abolafia-Rosenzweig, R., He, C., McKenzie Skiles, S., Chen, F., & Gochis, D., 2022: Evaluation and Optimization of Snow Albedo Scheme in Noah-MP Land Surface Model Using In Situ

Spectral Observations in the Colorado Rockies. *Journal of Advances in Modeling Earth Systems*, 14(10), https://doi.org/10.1029/2022MS003141.

Chakravarti, I.M., Laha, R.G., & Roy, J., 1967: *Handbook of Methods of Applied Statistics, Volume I*, John Wiley and Sons, pp. 392-394.

Chen, F., Janjić, Z., & Mitchell, K., 1997: Impact of Atmospheric Surface-layer Parameterizations in the new Land-surface Scheme of the NCEP Mesoscale Eta Model. *Boundary-Layer Meteorology*, 85(3), 391–421, https://doi.org/10.1023/A:1000531001463.

Currier, W. R., Thorson, T., & Lundquist, J. D., 2017: Independent Evaluation of Frozen Precipitation from WRF and PRISM in the Olympic Mountains. *Journal of Hydrometeorology*, 18(10), 2681–2703, https://doi.org/10.1175/JHM-D-17-0026.1.

Dickinson, R. E., Henderson-Sellers, A., & Kennedy, P. J., 1986: Biosphere-atmosphere transfer scheme (BATS) for the NCAR community climate model (NCAR Tech. Note NCAR/TN-38+STR, Vol. 82). National Center for Atmospheric Research.

Dickinson, R. E., Shaikh, M., Bryant, R., & Graumlich, L., 1998: Interactive Canopies for a Climate Model. *Journal of Climate*, 11(11), 2823–2836, https://doi.org/10.1175/1520-0442(1998)011<2823:ICFACM>2.0.CO;2

Gochis, D.J., M. Barlage, A. Dugger, K. FitzGerald, L. Karsten, M. McAllister, J. McCreight, J. Mills, A. RafieeiNasab, L. Read, K. Sampson, D. Yates, W. Yu, 2018: The WRF-Hydro modeling system technical description, (Version 5.0). NCAR Technical Note. Retrieved from https://ral.ucar.edu/sites/default/files/public/WRF-HydroV5TechnicalDescription.pdf.

Harms, D., Weeks, J., Lea, J., 2016: Comparing SNOTEL extended air temperature sensor and equations to an NSIT certified sensor in an environmental chamber. *84th Annual Western Snow Conference*, Seattle, WA.

He, C., Valayamkunnath, P., Barlage, M., Chen, F., Gochis, D., Cabell, R., Schneider, T., Rasmussen, R., Niu, G.-Y., Yang, Z.-L., Niyogi, D., & Ek, M., 2023: The Community Noah-MP Land Surface Modeling System Technical Description Version 5.0, http://dx.doi.org/10.5065/ew8g-yr95.

He, M., Hogue, T. S., Franz, K. J., Margulis, S. A., & Vrugt, J. A., 2011: Characterizing parameter sensitivity and uncertainty for a snow model across hydroclimatic regimes. *Advances in Water Resources*, 34(1), 114–127, https://doi.org/10.1016/j.advwatres.2010.10.002.

Hill, D. F., N. Bruhis, S. E. Calos, A. Arendt, and J. Beamer, 2015: Spatial and temporal variability of freshwater discharge into the Gulf of Alaska. *J. Geophys. Res.*, 120, 634–646, https://doi.org/10.1002/2014JC010395.

Jordan, R. E., 1991: A one-dimensional temperature model for a snow cover: Technical documentation for SNTHERM, 89. Retrieved from https://erdc-library.erdc.dren.mil/jspui/bitstream/11681/11677/1/SR-91-16.pdf.

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., & Wood, E. F., 2015: Inroads of remote sensing into hydrologic science during the WRR era: REMOTE SENSING. *Water Resources Research*, 51(9), 7309–7342, https://doi.org/10.1002/2015WR017616.

Kirchner, J., 2001: Data Analysis Toolkit #5: Uncertainty Analysis and Error Propagation. Berkeley Seismology Laboratory. University of California. http://seismo.berkeley.edu/~kirchner/eps_120/Toolkits/Toolkit_05.pdf.

Letcher, T., Minder, J., and Naple, P., 2022: Understanding and improving snow processes in Noah-MP over the Northeast United States via the New York State Mesonet. Engineer Research and Development Center (U.S.), http://dx.doi.org/10.21079/11681/45060.

Livneh, B., Deems, J. S., Schneider, D., Barsugli, J. J., and Molotch, N. P., 2014: Filling in the gaps: Inferring spatially distributed precipitation from gauge observations over complex terrain. *Water Resources Research*, 50(11), 8589–8610, https://doi.org/10.1002/2014WR015442.

Livneh, B., T. Bohn, D. Pierce, F. Munoz-Arriola, B. Nijssen, R. Vose, D. Cayan, and L. Brekke, 2015: A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013. *Sci Data*, 2150042, https://doi.org/10.1038/sdata.2015.42.

Niu, G.-Y., and Yang, Z.-L., 2006: Effects of frozen soil on snowmelt runoff and soil water storage at a continental scale. *Journal of Hydrometeorology*, 7(5), 937–952. https://doi.org/10.1175/JHM538.1.

Niu, G.-Y., Z.-L. Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, E. Rosero, M. Tewari, and Y. Xia, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res.*, 116(D12). https://doi.org/10.1029/2010JD015139.

National Oceanic and Atmospheric Administration (NOAA), National Weather Service (NWS), Office of Water Prediction (OWP), Silver Spring, MD., Analysis of Record for Calibration Version 1.1 - Sources, Methods, and Verification. Retrieved from https://hydrology.nws.noaa.gov/aorc-historic/Documents/AORC-Version1.1-SourcesMethodsandVerifications.pdf.

Oyler, J. W., S. Z. Dobrowski, A. P. Ballantyne, A. E. Klene, and S. W. Running, 2015: Artificial amplification of warming trends across the mountains of the western United States. *Geophysical Research Letters*, 42(1), 153-161, https://doi.org/10.1002/2014GL062803.

Park, S., and S. K. Park, 2016: Parameterization of the snow-covered surface albedo in the Noah-MP Version 1.0 by implementing vegetation effects. *Geosci. Model Dev.*, 9(3), 1073–1085. https://doi.org/10.5194/gmd-9-1073-2016.

Sakaguchi, K., and X. Zeng, 2009: Effects of soil wetness, plant litter, and under-canopy atmospheric stability on ground evaporation in the Community Land Model (CLM3.5): New schemes for CLM3.5 soil evaporation. *J. Geophys. Res. Atmos.*, 114(D1). https://doi.org/10.1029/2008JD010834.

Serreze, M. C., Clark, M. P., Armstrong, R. L., McGinnis, D. A., and Pulwarty, R. S., 1999: Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resources Research*, 35(7), 2145–2160, https://doi.org/10.1029/1999WR900090.

Sun, N., Yan, H., Wigmosta, M. S., Leung, L. R., Skaggs, R., and Hou, Z., 2019: Regional Snow Parameters Estimation for Large-Domain Hydrological Applications in the Western United States. *Journal of Geophysical Research: Atmospheres*, 124(10), 5296–5313. https://doi.org/10.1029/2018JD030140.

Vose, R.S., and Coauthors, 2014: Improved historical temperature and precipitation time series for US climate divisions. *J. Appl. Meteor. Climatol.*, 53, 1232-1251, https://doi.org/10.1175/JAMC-D-13-0248.1.

Wang, Y., Broxton, P., Fang, Y., Behrangi, A., Barlage, M., Zeng, X., & Niu, G., 2019: A Wet-Bulb Temperature-Based Rain-Snow Partitioning Scheme Improves Snowpack Prediction Over the Drier Western United States. *Geophysical Research Letters*, 46(23), 13825–13835, https://doi.org/10.1029/2019GL085722.

Wiken, E., Jiménez Nava, F., and Griffith, G., 2011: North American Terrestrial Ecoregions—Level III. Commission for Environmental Cooperation, Montreal, Canada.

Yan, H., N. Sun, M. Wigmosta, R. Skaggs, Z. Hou, and R. Leung, 2018: Next-generation intensity-duration-frequency curves for hydrologic design in snow-dominated environments. *Water Resources Research*, 54(2), 1093–1108, https://doi.org/10.1002/2017WR021290.

Yang, Z.-L., and Dickinson, R. E., 1996: Description of the Biosphere-Atmosphere Transfer Scheme (BATS) for the soil moisture workshop and evaluation of its performance. *Global Planet. Change*, 13, 117–134, https://doi.org/10.1016/0921-8181(95)00041-0.

Zhang, G., Zhou, G., Chen, F., Barlage, M., & Xue, L., 2014: A Trial to Improve Surface Heat Exchange Simulation through Sensitivity Experiments over a Desert Steppe Site. *Journal of Hydrometeorology*, 15, 664-684, https://doi.org/10.1175/JHM-D-13-0113.1.

Appendix B

Supplemental information for Chapter 3: Improved modelling of mountain snowpacks with spatially distributed precipitation bias correction derived from historical reanalysis

Contents of this Appendix

Texts B1, B2

Figures B1, B2, B3, B4, B5, B6, B7, B8

B1: Assimilation window for fSCA assimilation

The time window for fSCA assimilation is from the snow onset date to 1 April. The snow onset date is defined for each pixel-year as the first day when 3 consecutive days have maximum snow cover; we use this date to exclude fSCA observations that occur during early season snow accumulation events. We find that, without later ablation-season fSCA observations to correct the prior SWE estimates, these early-season fSCA observations tend to add more noise than value to the reanalysis. So, the assimilation window generally includes the period of time when the pixel is fully snow-covered, which may include some dips in snow cover relating to winter melt events as exemplified on days 40-50 in Fig. 3.3b and some portion of the ablation season depending on whether ablation happens before 1 April at that pixel.

B2: Streamflow observations

A daily time series of streamflow observations at the TGC (Tuolumne River at Grand Canyon) gage was constructed from records downloaded from the CDEC (California Data Exchange Center) platform. 15-min data was available for water years (WYs) 2009-2021, and daily data was available for WYs 2015-2021. In an effort to optimize data availability, we filled gaps in the 15-min data less than 3 hours long with linear interpolation and aggregated it to a daily time step. From this constructed daily time series, we filled gaps less than 5 days long with a spline interpolation (for example, in mid-April 2021 in Fig. B1). Any remaining gaps in the daily constructed time series were filled with available daily data from CDEC. We found that the constructed daily time series from 15-min data matched the daily data from CDEC exactly on overlapping days (for example, WY 2021 in Fig. 1). These steps yielded a complete daily time series of streamflow observations for April-July in WYs 2009-2021, except for WY 2019 when several weeks of data in April are still missing (Fig. B2).

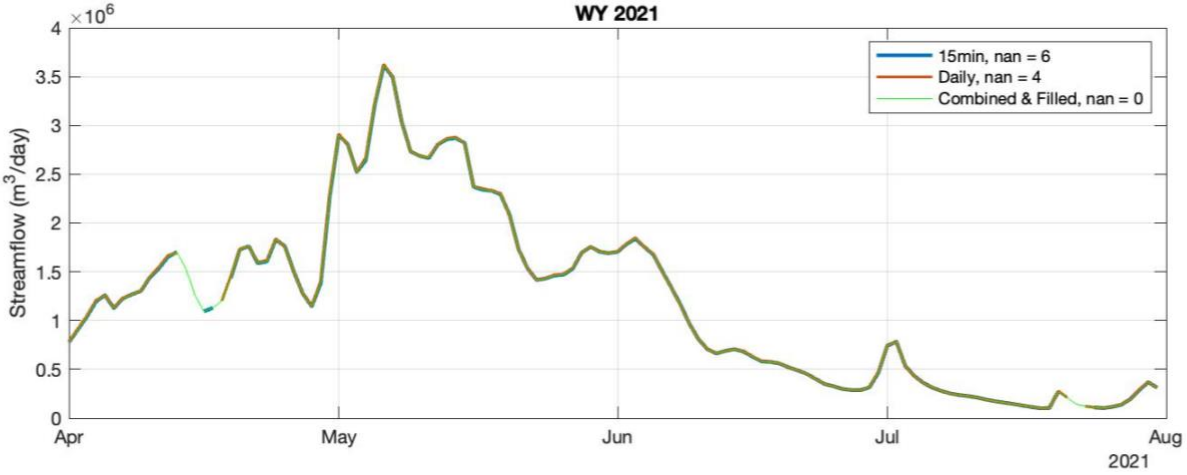


Figure B1. Time series of streamflow observations at the TGC gauge for April-July 2021. The blue line tracks the daily-aggregated 15-min data; the orange line tracks the daily data downloaded from CDEC; and the green line tracks the post-processed gap-filled time series.

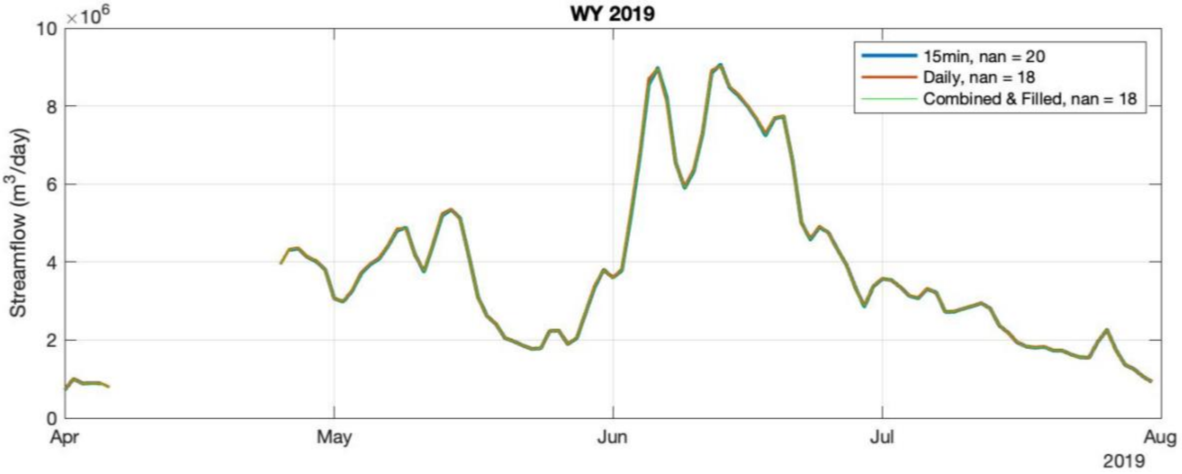


Figure B2. Time series of streamflow observations at the TGC gauge for April-July 2019. The blue line tracks the daily-aggregated 15-min data; the orange line tracks the daily data downloaded from CDEC; and the green line tracks the post-processed gap-filled time series.

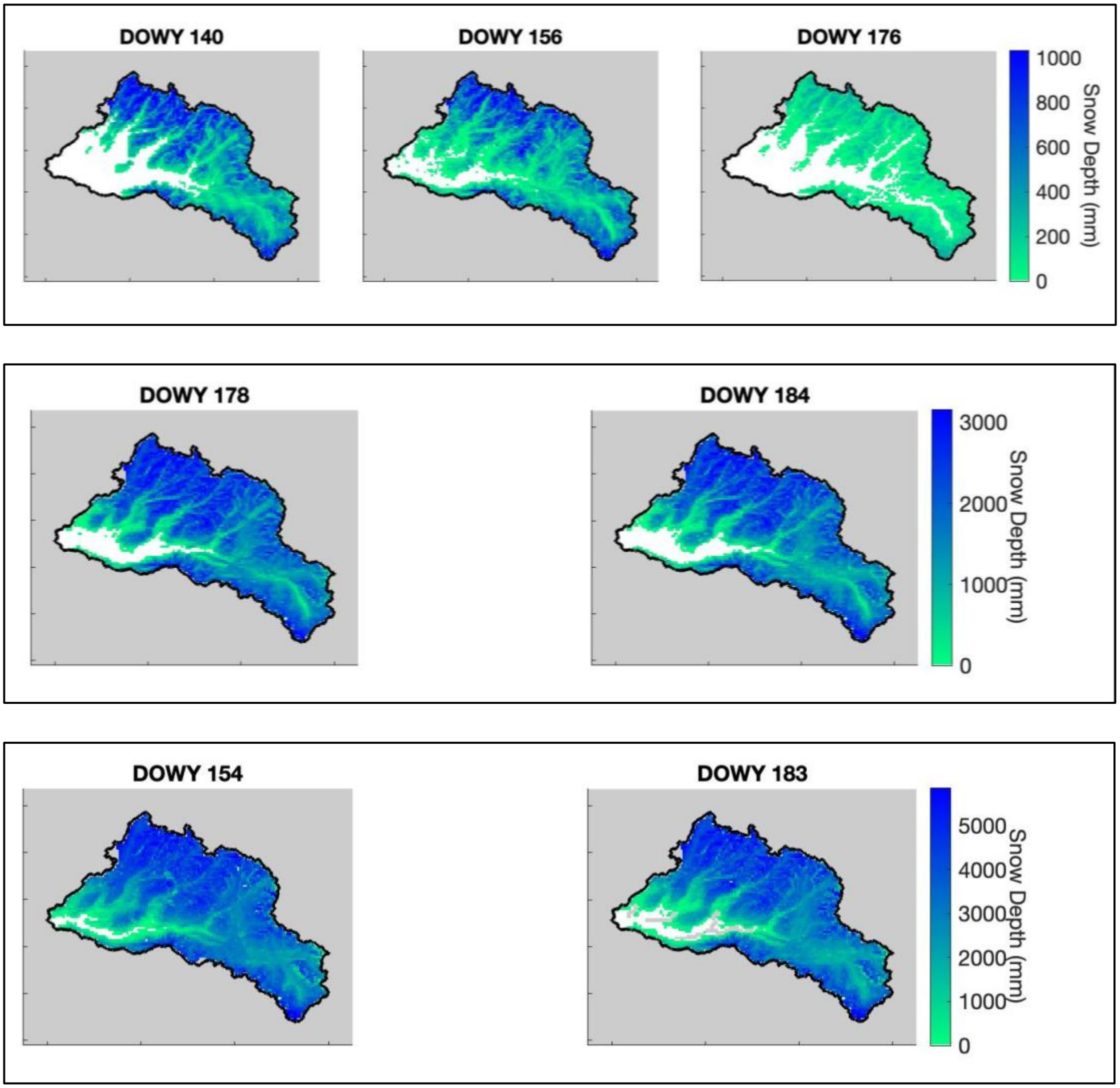


Figure B3. Maps showing the ASO-derived snow depth measurements over the Hetch Hetchy watershed used in data assimilation for (top) 2015, (middle) 2016, and (bottom) 2017.

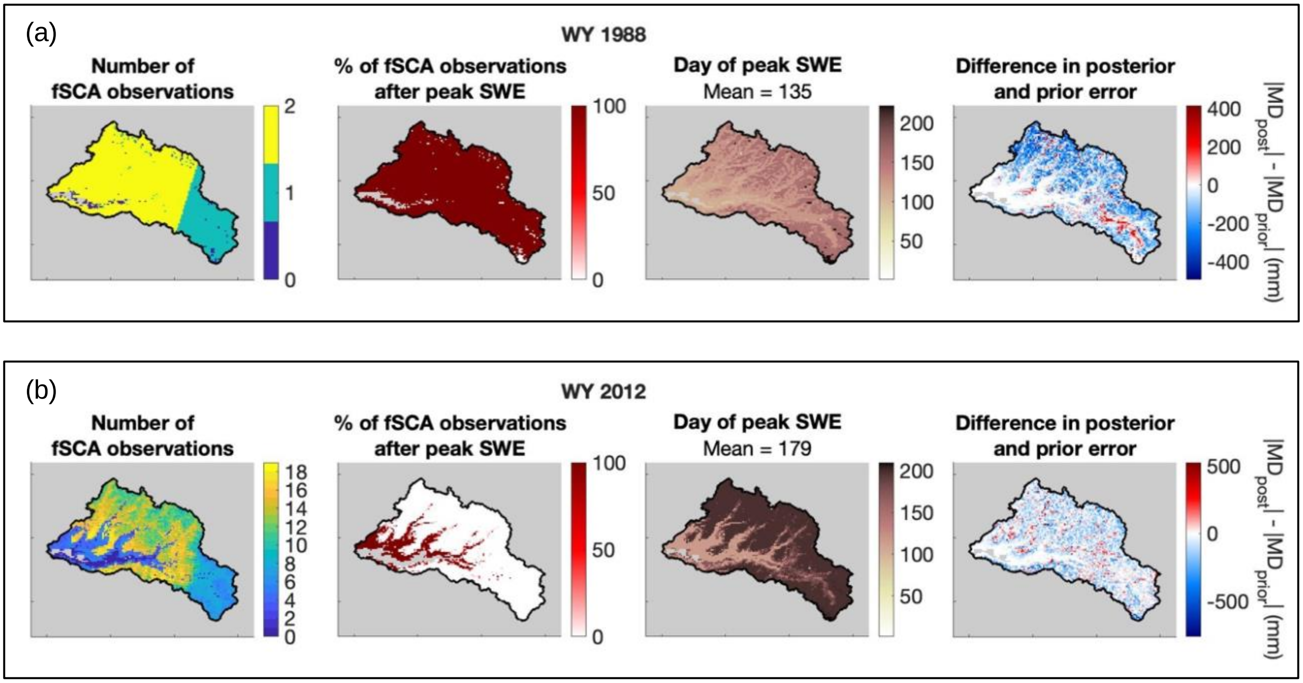


Figure B4. For the Case B + fSCA experiment, these maps illustrate (left to right) the number of assimilated fSCA observations; the percent of those fSCA observations that occur after pixel-wise peak SWE; the day of peak SWE; and the difference in posterior and prior error (where error is defined as the absolute difference between the SWE estimate and the historical reference) (a) WY 1998 and (b) WY 2012. In the right-most map, pixels colored in blue indicate a reduction in error with fSCA assimilation, and red indicates an increase in error. Note that the watershed-scale NRMSD was reduced with fSCA assimilation for both water years shown.

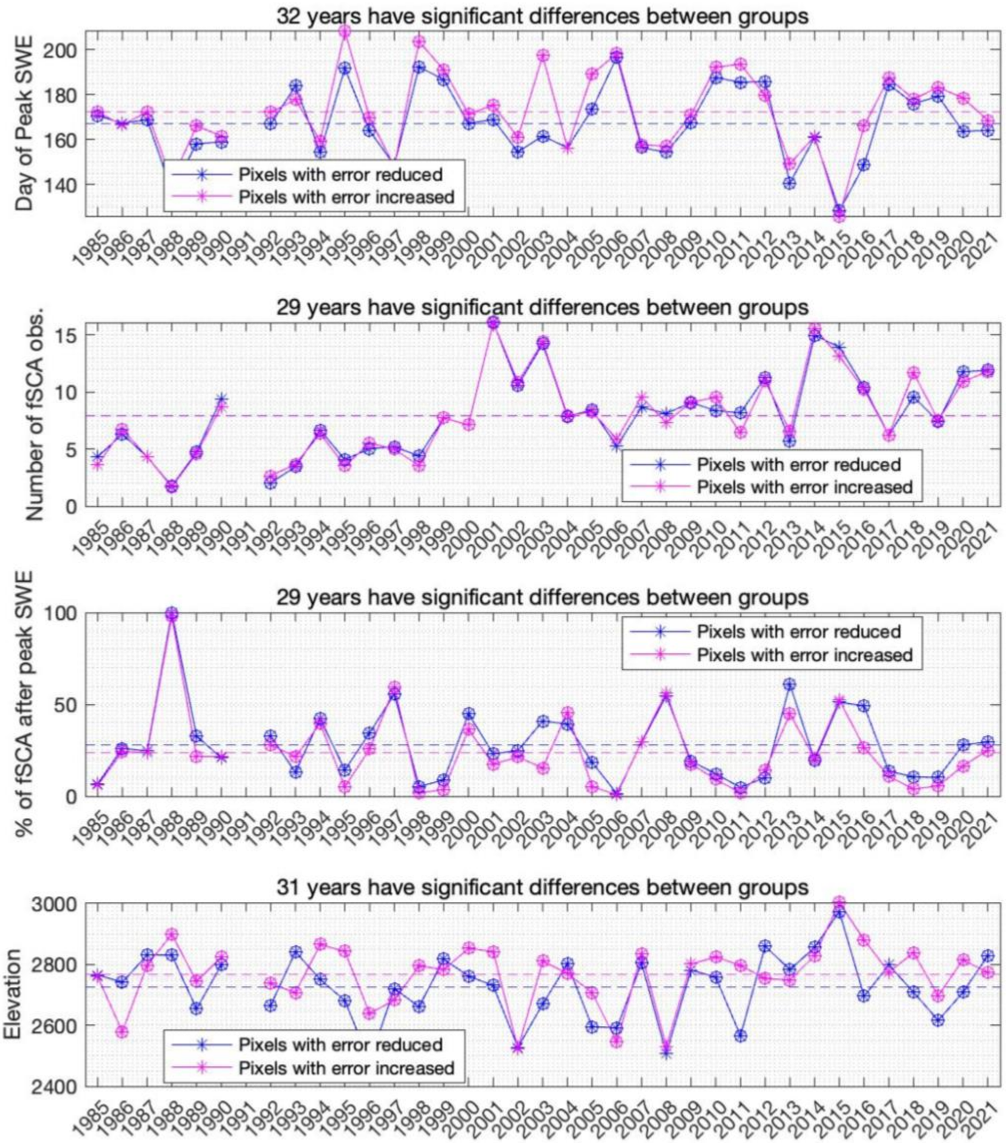


Figure B5. For the Case B + fSCA experiment, these time series track average values of 4 key metrics for model pixels with error reduced by fSCA assimilation versus pixels with error increased, where error is defined as the absolute difference with respect to reference 1 April SWE. A circle surrounds the yearly data point for years in which a statistically significance difference (p-value < 0.05 in a t-test) exists between the two groups of pixels. The dashed horizontal lines indicate the long-term average metric value for the two pixel groups. The metrics included are, from top to bottom: the day of peak SWE, the number of fSCA observations assimilated, the percent of those observations that occur after peak SWE, and elevation.

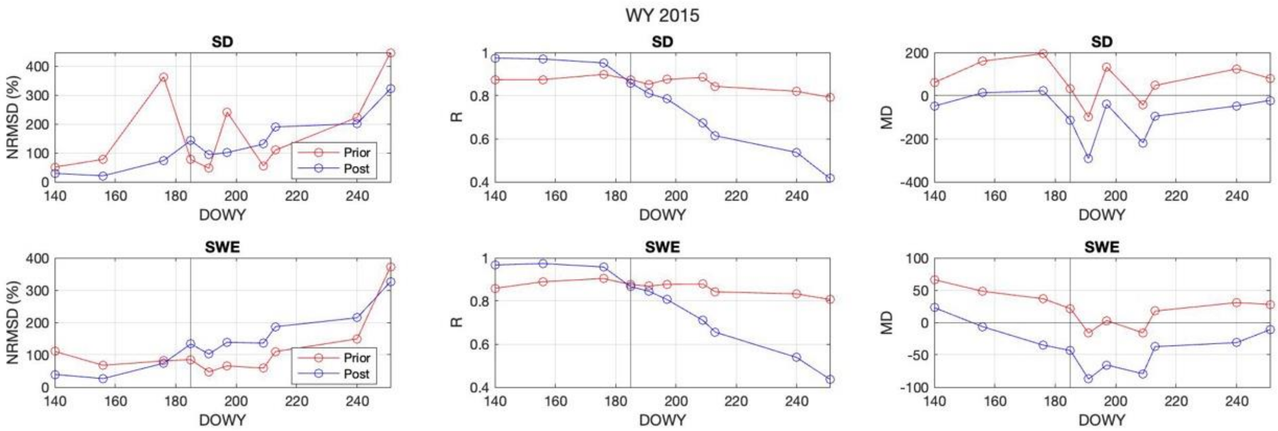


Figure B6. For WY 2015, a summary of the performance metrics (from left to right: NRMSD, R, and MD) for both prior (red) and posterior (blue) SWE and snow depth estimates in the Case B + SD experiment. Here, the reference is the ASO snow depth (top row) and ASO SWE (bottom row), and the metrics are evaluated on each day when ASO obtained measurements. The black vertical line indicates DOWY 185, the validation day for this year. The three prior ASO snow depth observations are assimilated into the posterior estimates in the Case B + SD experiment.

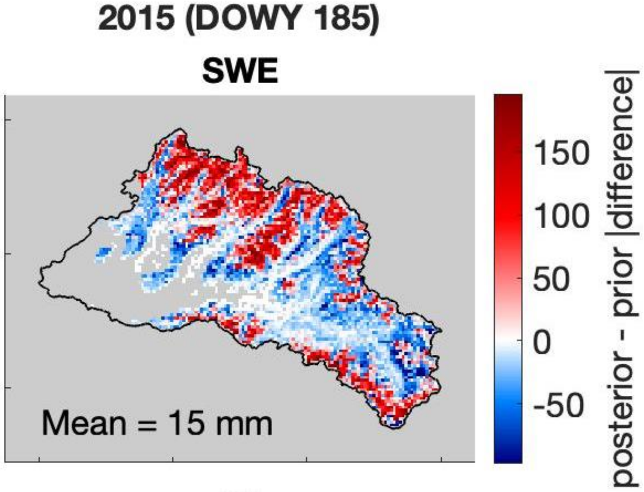


Figure B7. A map showing the change from the absolute difference relative to the ASO reference in prior SWE estimates to posterior SWE estimates for the Case B + SD experiment (|posterior – reference| – |prior – reference|), on WY 2015 DOWY 185. Areas in red indicate pixels where the posterior difference is greater than the prior difference. Note that pixels where both observed and simulated SWE is 0 are greyed out.

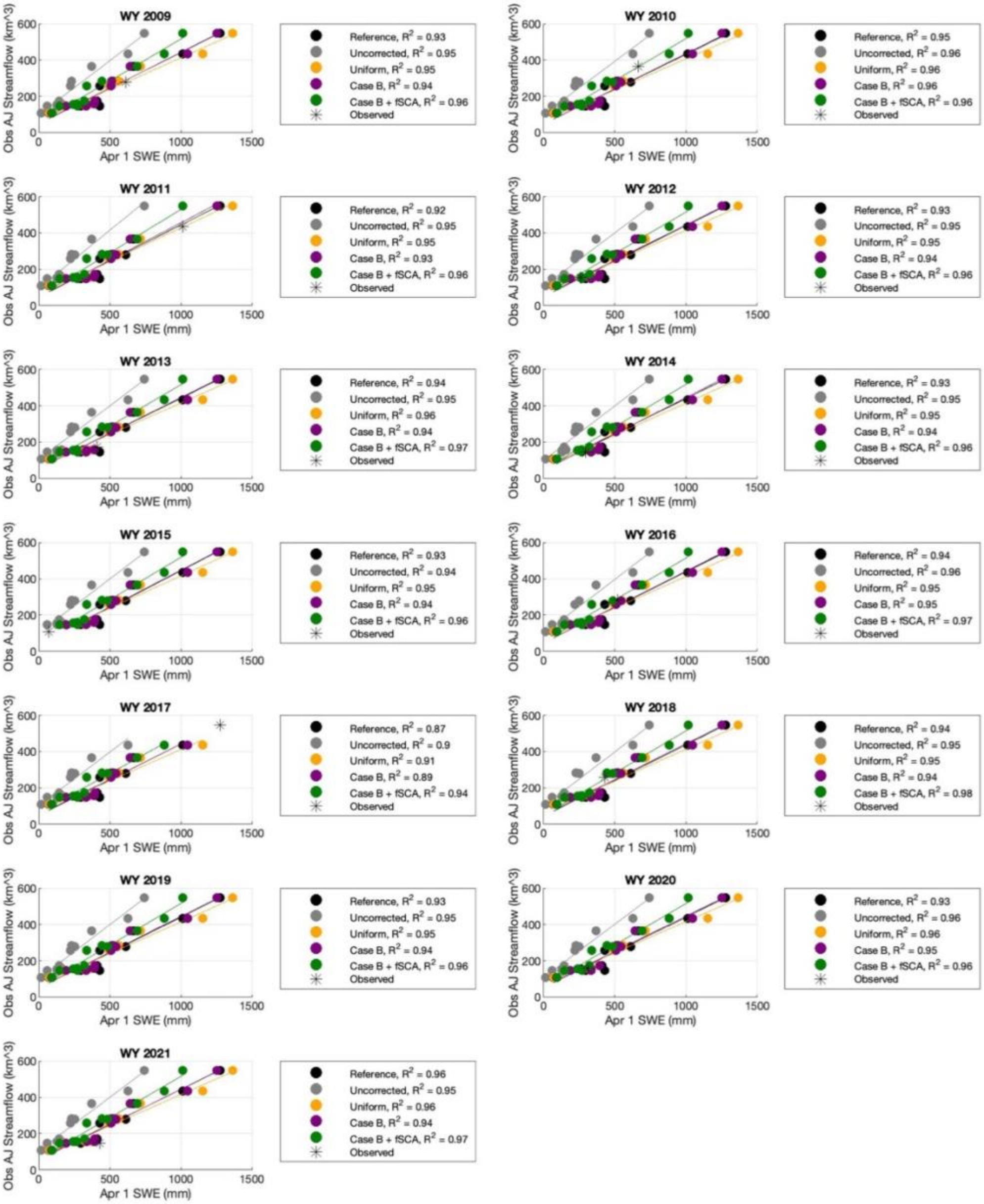


Figure B8. For WY 2009-2021, scatter plots showing basin-averaged 1 April SWE from the historical reference and four experiments, and observed AJ streamflow volume. Each dot represents a different water year. Solid lines indicate the regression lines for that experiment. The R^2 for each linear regression is noted in the legend. Note that for a given year, that year's data is excluded from the regression. Instead, the reference 1 April SWE and observed AJ streamflow is plotted with a black asterisk symbol.

Appendix C

Supplemental information for Chapter 5: Impact of climate warming and snowpack declines on California Sierra Nevada reservoir storage and water deliveries

Contents of this Appendix:

Texts C1, C2, C3

Figures C1, C2

Tables C1, C2, C3

References

C1: Selection of GCM subset

We select a subset of seven downscaled GCMs from the 14 that were used to force the VIC model to best match historical California climate. The selection criteria are high rankings in process-based and climate metrics from Krantz et al. (2021). These metrics compare GCM climate to ERA5-derived climate in California over the historical reference period of 1979-2014 (Krantz et al., 2021). Process-based metrics quantify the models' ability to capture large-scale patterns including: northern hemisphere circulation, blocking, wind shear, extreme precipitation, Santa Ana winds, and El Niño

Southern Oscillation. Local climate metrics measure seasonal and annual patterns in temperature and precipitation. A higher metric ranking corresponds to a better fit. So, we select GCMs from those used to drive the VIC model that have high rankings in both local climate and process-based metrics (Fig. C1). We prioritized a higher ranking in local climate metrics because these are more suited to statistical downscaling methods (such as LOCA, the method chosen for this study) (Krantz et al., 2021).

Figure C1. Annotated Figure 6 from Krantz et al. (2021). We add red boxes to highlight the selected GCMs for this study.

C2: Hedging

Hedging is applied as a reservoir operating rule to restrict releases when reservoir storage is low. Fig. C2 illustrates how this rule is applied. We use a 2-point hedging rule with three conditions:

1). If storage is greater or equal to h_2, release the target demand.

2). If storage is less than h_1, release all the water available.

3). If storage is between h_1 and h_2, make a restricted release:

$$R_t = m(A_t - h_1) + h_1 \tag{1}$$

where for day t, R is the prescribed release, A is the available storage (including the inflow gain and evaporation loss), and m is computed as:

$$m = (T_t - h_1)/(h_2 - h_1) \tag{2}$$

where T is the target demand. The thresholds h_1 and h_2 are chosen such that $0 \leq h_1 \leq D$ and $T \leq h_2 \leq K + T$, where K is the reservoir capacity.

In this application of hedging, we define h_1 and h_2 as time-varying thresholds, such that:

$$h_{2,t} = a\,(K + T_t) \tag{3}$$

$$h_{1,t} = b\,(T_t) \tag{4}$$

where factors a and b are static parameters ranging from 0 to 1. We calibrate these parameters for each reservoir.

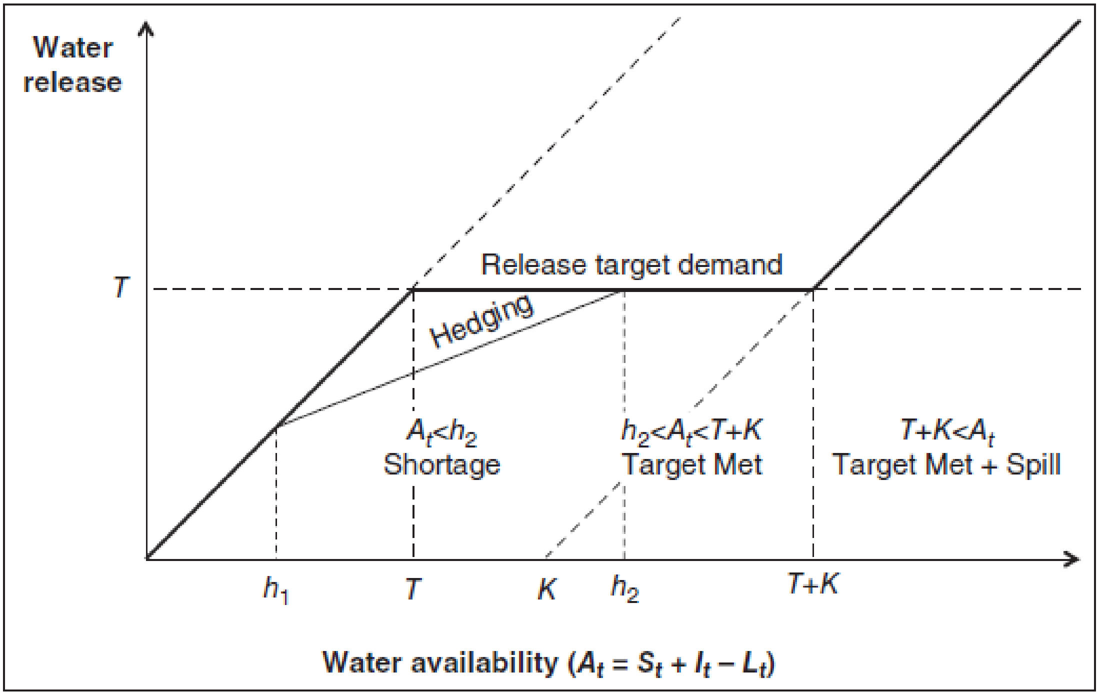


Figure C2. Schematic illustrating the application of a two-point hedging rule with standard operating policy, from Figure 130.2 in Vijay et al., 2017.

C3: Evaporation model

To build a statistical model that predicts evaporation losses from the reservoir, we follow three steps: 1). compilation of historical records for observed evaporation rate; 2). definition of a linear regression that predicts evaporation rate from temperature; 3). definition of a linear regression that predicts storage area from simulated storage.

We pull the best available data for historical evaporation rates from the California Data Exchange Center (CDEC). The source of observations is tabulated for each reservoir in Table C1. For four reservoirs, we use observed evaporation rates from a nearby reservoir sensor due to limited data availability. When the evaporation rate is recorded as a volume (cfs), we convert it to a depth (in/day) by dividing the volume with the predicted storage area. We remove obvious outliers and negative values from the record.

We build a linear regression to predict the evaporation rate (in/day) from temperature. This regression uses observed evaporation rate (Table C1) as the dependent variable, and temperature from the grid cell closest to the reservoir location as the independent variable. To build the regression equation, we pull temperature from the final post-processed daily historical forcings (bias-corrected ERA5-WRF; Rahimi et al., 2022) used in the calibration of the VIC land surface model for WYs 2001-2020. The coefficient of determination for these regressions are tabulated in Table C2. When applied to the historical and future time periods, the input to this regression is the temperature from the grid cell closest to the reservoir location in the bias-corrected GCM meteorology.

The second regression model predicts area from reservoir storage based on monthly measurements compiled by the Global Reservoir Storage dataset (Li et al., 2023). This dataset assembles reservoir water areas from a multi-source satellite dataset and monthly storage using bathymetric maps for years 1999 to 2018. The coefficient of determination for these regression equations are tabulated in Table C2.

To estimate evaporation volume, we multiply the predicted evaporation rate with the predicted storage area. For calibration and validation purposes, we use observed storage rates; and for the GCM results, we use simulated storage. We compare the predicted daily evaporation volumes with available observed values for Shasta, Oroville, Folsom, New Melones, and Millerton over the validation period of WYs 2001-2020. Weekly Kling-Gupta Efficiency (KGE) scores for this comparison are tabulated in Table C1. Note that for WYs 2021-2023 in the validation of simulated reservoir storage and outflow, we apply median historical evaporation because the available temperature input data only went to 2020.

Table C1. The source of historical evaporation observations for each reservoir is indicated as: CDEC station ID (sensor number).

Reservoir	**Source of observations**	**Description**
sha	SHA (74)	Computed lake evaporation from Shasta [cfs]
oro	ORO (74)	Computed lake evaporation from Oroville [cfs]
bul	ORO (74)	Computed lake evaporation from Oroville [cfs]
fol	FOL (74)	Computed lake evaporation from Folsom [cfs]
cmn	NHW (175)	Satellite-observed evaporation rate from New Hogan [in/day]
nml	NML (74)	Computed lake evaporation from New Melones [cfs]
dnp	NML (74)	Computed lake evaporation from New Melones [cfs]
exc	NML (74)	Computed lake evaporation from New Melones [cfs]
mil	MIL (74)	Computed lake evaporation from Millerton [cfs]
pnf	PFW (175)	Satellite-observed evaporation rate from Pine Flat [in/day]
trm	LKW (175)	Satellite-observed evaporation rate from Kaweah [in/day]
scc	SCW (175)	Satellite-observed evaporation rate from Success [in/day]
isb	IWS (175)	Satellite-observed evaporation rate from Isabella [in/day]

Table C2. Coefficient of determination for the two linear regression models, and weekly KGE scores for the validation period WYs 2001-2020. Note that validation is performed in reservoirs where available observed records of evaporation volume exist.

	Coefficient of determination (R^2)		**Simulated to observed validation (weekly)**
Reservoir	**Temperature to evaporation**	**Storage to area**	**KGE**
sha	0.64	0.999	0.85
oro	0.73	0.998	0.88
bul	0.7	0.999	
fol	0.59	0.999	0.81
cmn	0.74	0.997	
nml	0.71	0.994	0.89
dnp	0.72	0.998	
exc	0.71	0.989	
mil	0.68	0.999	0.86
pnf	0.7	0.992	
trm	0.78	0.971	
scc	0.74	0.967	
isb	0.73	0.995	

Table C3. Comparison of R^2 values for daily simulated reservoir storage from this study and two examples from the literature.

	Present study	**Steinschneider et al., 2023**	**Zeff et al., 2021**
Validation period:	WYs 2001-2023	WYs 1997-2021	WYs 1996-2016
Reservoir	**daily storage R^2**		
sha	0.88	0.86	0.92
oro	0.87	0.75	0.94
bul	0.78	0.8	0.87
fol	0.81	0.82	0.86
cmn	0.83		
nml	0.95	0.91	0.96
dnp	0.94	0.9	0.87
exc	0.94	0.79	0.93
mil	0.5	0.66, 0.56*	0.56
pnf	0.77	0.87*	
trm	0.57	0.45*	
scc	0.57	0.85*	
isb	0.84	0.91*	

* shared over email communication by Jonathan Herman (October 2024), for the validation period of WYs 2014-2023

References

Li, Y., Zhao, G., Allen, G. H., & Gao, H. (2023). Diminishing storage returns of reservoir construction. *Nature Communications*, *14*(1), 3203. https://doi.org/10.1038/s41467-023-38843-5

Rahimi, S., Krantz, W., Lin, Y., Bass, B., Goldenson, N., Hall, A., Lebo, Z. J., & Norris, J. (2022). Evaluation of a reanalysis-driven configuration of WRF4 over the western United States from 1980 to 2020. *Journal of Geophysical Research: Atmospheres*, *127*(4), e2021JD035699. https://doi.org/10.1029/2021JD035699

Singh, V. P. (2017). Reservoir operation. In *Handbook of applied hydrology* (2nd ed., 130.3). McGraw-Hill Education. Retrieved from https://www.accessengineeringlibrary.com/content/book/9780071835091/toc-chapter/chapter130/section/section14