Where Does Missing Water Go? Constraining Hydrologic Anomalies in Snow-Dominated Mountain Catchments

Daniel Hogan

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

Doctor of Philosophy

University of Washington

2026

Reading Committee:

Jessica Lundquist, Chair

Bart Nijssen

Erkan Istanbulluoglu

Program Authorized to Offer Degree:
Civil and Environmental Engineering

University of Washington

Abstract

Where Does Missing Water Go? Constraining Hydrologic Anomalies in Snow-Dominated Mountain Catchments

Daniel Hogan

Chair of the Supervisory Committee:
Jessica Lundquist
Civil and Environmental Engineering

This dissertation explores uncertainties that can drive hydrologic anomalies, which occur when observed streamflow falls far from projected values. Each chapter targets a different point where uncertainty enters the mountain water balance: how much snow is lost before it can melt, how accurately we measure precipitation, and how much storage shapes streamflow over multiple years.

In Chapter 2, we combine meteorological observations spanning site-to-synoptic scales from three field campaigns in a Colorado mountain valley to identify the events responsible for the majority of winter sublimation. We find that most sublimation occurs during a small number of discrete events coinciding with either short sunny, dry periods or long storm events with new snowfall and blowing snow. The number of these events varies considerably year to year with no significant long-term trend. Accurately representing seasonal sublimation requires capturing discrete events driven by identifiable synoptic and local meteorological conditions.

In Chapter 3, we compare precipitation measurements from 11 co-located gauges, supplemented by radar-derived and gridded products, to identify the dominant drivers of measurement variability at a mountain site. We find large discrepancies in total accumulation across gauges, with biases concentrated in a small number of winter events driven by power

failures, gauge burial, blowing snow, and high snowfall rates. For future deployments in snow-dominated mountain terrain, we recommend co-located storage-type gauges with low power requirements to ensure reliable, continuous measurements.

In Chapter 4, we use the Structure for Unifying Multiple Modeling Alternatives (SUMMA) model to evaluate whether representing long-term water storage improves simulated streamflow in two contrasting snow-dominated basins. Observed late-season low flows carried a memory signature that models without an explicit storage compartment could not reproduce, while adding a basin-appropriate compartment recovered it. These stores contributed disproportionately to low flows in dry years but did little to improve water-year volume accuracy. As mountain snowpack and groundwater decline, a model's ability to account for missing water will depend on how well it represents the dominant storage mechanism in the basin in question.

Together, these chapters show that much of this uncertainty traces back to conditions that average behavior obscures: discrete high-impact events in the case of sublimation and precipitation measurement, and slow storage dynamics that only become visible across years. As mountain systems experience increased variability, these results suggest that improving representations of the mountain water balance requires deliberately targeting what average conditions leave out.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Number Page

ACKNOWLEDGMENTS

As I sit writing this section on a late May evening, I struggle to even comprehend how I could communicate the gratitude I feel for all those who gave me mentorship, guidance, support, and love. But I will do my best to try by starting from the beginning. There is a quote from Henry David Thoreau that has guided me for many years: "If one advances confidently in the direction of his own dreams and endeavors, to lead the life he has imagined, he will meet with a success unexpected in common hours". The pursuit of my own "dreams and endeavors" would only be possible with the unwavering support from my parents, Sean and Lynn, who always pushed me to chase my passions. This is something I will never take for granted, and I will be eternally grateful for.

To my three sisters, Kelsey, Lauren, and Brighid, you have all brought me so much joy and taught me how I might navigate the real world by both advocating for myself and pursuing what I love. I can only hope I can return the favor. And to my Granne, your interest in the natural world that you captured through your art has always inspired me to slow down and try to observe the beauty in the world every day. I am forever grateful for this lesson.

To my closest friends both new and old — Jordi, Sam, Allen, Christian, Clinton, Eli, Will(s), Lucas — thank you for keeping me sane and always being there for a laugh, an adventure, or just an ear to listen.

To my elementary and high school teachers and university professors, especially Professor Jim Smith from Princeton, thank you for teaching me the educational groundwork so I could pursue my interests in weather, water, and wilderness. And to my coaches and teammates over the years, most notably Marty Crotty, Andrew Blair, and Brian de Regt, you taught me how to apply myself for the pursuit of excellence. Practice makes permanent.

To my advisor, Dr. Jessica Lundquist, thank you for bringing me into this research space. I always dreamed of seeking out answers to the mysteries of the natural world, but never felt I had a way to access it. You opened that door for me, providing me with the opportunity to dive deeper into my passion for understanding the interconnectedness of the natural world and do research in one of the world's most beautiful places. Your expertise and guidance have been instrumental in my growth as a scientist, researcher, and writer. I feel so fortunate to have had you as an advisor.

I also gratefully acknowledge my funding support for this work from the National Science Foundation (Award 2139836), Sublimation of Snow, and the Department of Energy Earth Systems Science (Grant DE–SC0024075), Seasonal Cycles Unravel Mysteries of Missing Mountain Water. Without this, I would never have had the opportunity to conduct such a wide breadth of research.

To colleagues I met along the way, most notably Dan Feldman and billy barr. Dan, you have always been a welcoming voice in a world of stern scientists. I look forward to continuing to work with you. billy, your devotion to observing the natural world and living a simple life is inspiring. To my medical care team during my time at UW, you have kept me in the game, especially in these final 2 years, which included a bulging disk, an epilepsy diagnosis, and a genetic heart condition diagnosis. These were (and still are) difficult challenges to face, but you made addressing them straightforward. Thank you.

To Denali, I couldn't have imagined that a little Covid foster pup would turn into such a fundamental part of my adult life. You make me smile on a daily basis and are the ultimate buddy in everything I do. Plenty of frisbee sessions and mountain frolics await you, my boy!

And last, to my wife, Grace. This is likely the hardest acknowledgment to put into words. To adequately do so, I would have to write another 200 page dissertation. So, I will say simply: thank you and I love you. Your endless support these last five years is unmatched and your insatiable sense of adventure has allowed us to engage our shared love

for the outdoors. I look forward to supporting you in the same way for the many future decades we will have together.

From all these people and experiences, the fundamental lesson I have learned is that grit and hard work, driven by passion for something you love can get you pretty damn far in this world. So get after it.

Chapter 1

INTRODUCTION

1.1 Background

Mountain snowpacks serve as the primary seasonal water reservoir for many water-stressed regions across the world, storing winter precipitation and releasing it gradually through spring and summer when downstream demand is highest. Snow water equivalent (SWE) has long served as the primary predictor for water resource forecasts. However, this historically reliable relationship between snowpack and streamflow periodically breaks down. For example, recent years in the Upper Colorado River Basin, which supplies water to over 40 million people [*Wheeler et al.*, 2022], offer a stark illustration: in 2020 and 2021, streamflow reached only 65% and 38% of normal despite near-normal snowpacks [*Goble and Schumacher*, 2023]. These hydrologic anomalies, sometimes called "missing water" [*Lapides et al.*, 2022; *Hogan and Lundquist*, 2024], carry direct consequences for reservoir operations, water allocations, and ecosystem flows.

Attributing such anomalies to physical processes requires balancing streamflow with its sources (rain and snow) and sinks (evapotranspiration, sublimation, storage change). But each term in this balance carries its own measurement and modeling uncertainty, making it difficult to isolate any single process. This dissertation explores three of those uncertainties in three separate chapters and develops strategies for reducing them.

The first is uncertainty in sublimation, the direct phase change of snow to water vapor, which represents a loss of both mass and energy from the snowpack. The estimated impact from snow sublimation on the alpine water balance ranges widely, with prior work finding anywhere from 1–50% of snow is lost through this process [*Strasser et al.*, 2008; *Mott et al.*, 2018]. This has led it to be defined as a key variable of uncertainty in hydrologic models [*Slater et al.*, 2001]. Sublimation intensity is controlled by the vapor pressure gradient, wind exposure, and available energy, meaning it peaks under very different conditions: high net radiation during warm, dry periods, and blowing snow during storms [*Reba et al.*, 2012; *Sigmund et al.*, 2022]. Because these conditions are intermittent and extreme, direct observations are rare and difficult to sustain in mountain terrain [*Svoma*, 2016].

The second uncertainty lies in precipitation. Although the most vital component of the water balance [*Larson and Peck*, 1974; *Moulin et al.*, 2009], precipitation is notoriously difficult to measure accurately in mountains: sparse gauge networks, wind-driven undercatch, and instrument failures all introduce errors that propagate into every downstream application [*Kochendorfer et al.*, 2022; *Lundquist et al.*, 2019]. When measurements go wrong and go uncorrected, the apparent water balance deficit may reflect bad inputs as much as real hydrologic processes.

The third uncertainty lies in the importance of water held in a basin across several years, a term we refer to as long-term storage. Recent work has shown that storage conditions in one year can affect streamflow for several subsequent years, a phenomenon referred to as hydrologic memory [*Brooks et al.*, 2021; *Opie et al.*, 2020]. Unlike snow or precipitation, long-term water stores, like deep groundwater or perennial snow fields, are rarely observed directly at the basin scale, so its influence must usually be inferred from streamflow itself. Simpler operational and research models typically do not represent this behavior [*Zhang et al.*, 2026; *Fowler et al.*, 2020], and whether they benefit from doing so is unclear, especially in mountain basins where storage takes very different forms.

In Chapter 2, we characterize the meteorological conditions driving sublimation variability in a Colorado River headwater catchment, providing observational constraints on a loss term whose magnitude and variability have been difficult to pin down at the basin scale. In Chapter 3, we leverage a rare co-located deployment of precipitation gauges to diagnose when and why precipitation measurement accuracy degrades in mountains, identifying which gauge types are most resilient during high-impact events and offering practical strategies to reduce input uncertainty. In Chapter 4, we examine how adding long-term storage compartments to a hydrologic model changes its representation of streamflow, identifying which components of the hydrograph depend on storage and when representing it matters most. Together, these chapters do not resolve the missing water problem, but they work toward a more honest accounting of what we can and cannot yet attribute, and offer a clearer path forward for reducing the uncertainties that stand in the way.

1.2 Outline

We used a combination of measurements from the three concurrent field campaigns and hydrologic modeling to complete three research objectives, which are organized into three chapters, Chapters 2–4, outlined below. In Chapter 5, we conclude this dissertation by summarizing the objectives completed in Chapters 2–4 and discussing the broader impacts of the work, open questions, and future areas for research.

Chapter 2 *Synoptic-Scale Systems Control Total Winter Sublimation at an Alpine Site*

- **Objective**: Investigate the most extreme sublimation events throughout two winter seasons and isolate the dominant synoptic-scale processes that enhance or limit these events to determine what drives sublimation variability.

- This work was published in *Journal of Hydrometeorology* and is reproduced from the following citation.

  Hogan, D., Schwat E., Gutmann, E., Vano, J. A., and Lundquist, J. D. (2026). Synoptic-Scale Systems Control Total Winter Sublimation at an Alpine Site. Journal of Hydrometeorology. https://doi.org/10.1175/JHM-D-25-0143.1

Chapter 3 *Precipitation Measurement Uncertainty in Mountains: A Gauge Intercomparison from the SAIL, SPLASH, and SOS Campaigns*

- **Objective**: Intercompare precipitation measurements from several instruments over a two-year period in a mountain environment to examine when and why measurements diverge from one another. Use these results to provide deployment recommendations for future precipitation measurement campaigns in mountains.

- This chapter has been submitted to the *Journal of Oceanic and Atmospheric Technology.*

  Hogan, D., Schwat, E., Gutmann, E., Feldman, D.R., and Lundquist, J. D. Precipitation Measurement Uncertainty in Mountains: A Gauge Intercomparison from the SAIL, SPLASH, and SOS Campaigns.

Chapter 4 *Are models forgetting something? Representing multi-year snow and subsurface storage improves simulated low-flow persistence*

- **Objective**: Use a single modeling platform to test whether observed streamflow records carry a signature of long-term storage memory, and whether a model can reproduce that signature without representing storage explicitly. Add a basin-appropriate long-term storage compartment to each configuration and track how and where it changes simulated streamflow. Test how results differ across two physiographically distinct snow-dominated catchments, and identify the conditions under which representing storage matters most.

- This chapter is in preparation for submission to *Water Resources Research.*

  Hogan, D., Gutmann, E., Carroll, R., and Lundquist, J. D. Are models forgetting something? Representing multi-year snow and subsurface storage improves simulated low-flow persistence

Chapter 2

SYNOPTIC-SCALE SYSTEMS CONTROL TOTAL WINTER SUBLIMATION AT AN ALPINE SITE

2.1 Introduction

Sublimation reflects a loss of both mass, in the form of water vapor, and energy, in the form of latent heat, from the snowpack to the atmosphere as snow particles change phase [*Schmidt*, 1972]. Its intensity is chiefly controlled by the vapor pressure gradient between the surface and atmosphere, the wind speed and exposure, and the magnitude of the available energy supply from either net radiation or turbulent sources [*Mott et al.*, 2018]. Given these different driving forces, vastly different weather conditions can cause sublimation to increase. For example, during relatively warm, dry conditions, sublimation tends to peak when net radiation is maximized [*Reba et al.*, 2012; *Sexstone et al.*, 2016]. Alternatively, blowing snow during high winds leads to the highest rates of sublimation [*MacDonald et al.*, 2018; *Sigmund et al.*, 2022; *Schmidt*, 1982; *Bliss et al.*, 2011; *Reba et al.*, 2012]. Strong winds entrain snow into the atmosphere, where suspended snow sublimates efficiently even when other drivers are less favorable [*Sigmund et al.*, 2022; *Strasser et al.*, 2008]. These conditions frequently occur during passing storms, when fresh snowfall lowers the wind speed threshold for blowing snow [*Guyomarc'h and Mérindol*, 1998; *Vionnet et al.*, 2013].

Since sublimation is maximized under such distinct and sometimes extreme meteorological conditions, it is perhaps unsurprising that direct observations of the process remain difficult and relatively rare. Just how much snow is lost to sublimation remains highly uncertain: estimates based on measurements and models have ranged from 1–90% of annual snowfall [*Svoma*, 2016; *Strasser et al.*, 2008] and largely depend on how and where these estimates are made. Accurately measuring it requires fine-scale observations of the local meteorological and snow conditions. Yet, these observations are often unavailable and difficult to maintain in mountainous terrain outside of dedicated field campaigns [*Svoma*, 2016; *Lundquist et al.*, 2024].

Among available measurement methods, properly installed and well-monitored open-path eddy-covariance (EC) measurements are considered the most reliable for quantifying turbulent exchanges in alpine terrain [*Lundquist et al.*, 2024; *Reba et al.*, 2012; *Sexstone et al.*, 2016]. However, these measurements have limitations. They are sensitive to both data quality criteria and post-processing methods in complex terrain and can be impacted by blowing snow and snowfall as particles pass through the EC system [*Stiperski and Rotach*, 2016; *Bintanja*, 2001a; *Sigmund et al.*, 2022]. Last, EC measurements cannot differentiate blowing snow sublimation from surface sublimation [*Reba et al.*, 2012; *Groot Zwaaftink et al.*, 2013]. Likewise, modeling efforts face their own challenges. There is substantial uncertainty in forcing data and surface parameterizations [*Sigmund et al.*, 2022; *Xia et al.*, 2017; *Musselman et al.*, 2015], and relatively few models represent important processes like blowing snow [*Lundquist et al.*, 2024; *Slater et al.*, 2001; *Sharma et al.*, 2023]. These limitations also have real consequences when estimating sublimation's influence on water resources. For instance, hydrologic modeling work by *Xiao et al.* [2018] attributed 18% of the Colorado River's streamflow decline over the past century to increases in winter sublimation; however, they mainly attributed this increase to unrealistic wind speed estimates in the historic forcing dataset they used. Consequently, sublimation remains a poorly constrained component of the snow mass balance, with few direct observations to characterize its variability over long periods [*Sexstone et al.*, 2018; *Liston and Sturm*, 2004; *Reba et al.*, 2012].

Despite these uncertainties, many studies have tried to estimate how sublimation responds to changing climate conditions, although estimates of its response vary, too. Several modeling efforts suggest that warmer temperatures increase sublimation by increasing the vapor pressure gradient between the snow and the air above it [*Cullen et al.*, 2014; *Scaff et al.*, 2024; *Guo et al.*, 2023; *Harpold et al.*, 2012; *Harpold and Brooks*, 2018]. However, other research argues that warming may instead reduce sublimation by shortening the snow season and limiting the frequency of blowing snow [*Rasouli et al.*, 2015; *Pomeroy and Gray*, 1994; *Sexstone et al.*, 2018]. Results from observation-based studies, although limited in their spatial and temporal extent, support a third possibility: that sublimation totals are sensitive to the frequency of high-magnitude sublimation events that are discrete and episodic [*Bintanja*, 2001b; *Bliss et al.*, 2011; *Cline*, 1997; *Hood et al.*, 1999; *Jackson and Prowse*, 2009; *Zhang et al.*, 2004]. This suggests that processes operating from the surface to synoptic scales control these events. However, many of these researchers explicitly discuss how they lacked the necessary data to make these larger-scale connections [*Hood et al.*, 1999; *Jackson and Prowse*, 2009; *Reba et al.*, 2012; *Sexstone et al.*, 2016]. Therefore, the frequency and character of these events have not been explored in detail, and we still do not know how much they contribute to winter sublimation.

This research gap points to a need to better characterize the most impactful sublimation events. Similar to past work that linked different snow mass balance components to broader atmospheric patterns [*Grundstein and Leathers*, 1999; *Cline*, 1997; *McGinnis*, 2000; *McCabe*, 1994; *Dettinger et al.*, 2011], our objective is to identify what drives these events to better understand and constrain winter sublimation in mountainous terrain. To accomplish this objective, we used data from three co-located field campaigns in a Colorado mountain valley to address three research questions (RQs):

1. **RQ1** How much sublimation occurred during large sublimation events during the 2022 and 2023 winter seasons?

2. **RQ2** What were the site-, valley-, and synoptic-scale characteristics during them?

3. **RQ3** How can the synoptic-scale characteristics we identify be used to model long-term trends in the frequency and duration of these events?

To answer these questions, we integrated observations across site-, valley-, and synoptic-scales to identify and characterize the largest sublimation events. Using the dominant features of these events, we trained a statistical model to estimate the occurrence of these events over the past four decades. This analysis provides the necessary context to understand the importance of discrete, episodic sublimation events to winter sublimation and highlights the potential for using synoptic-scale conditions to estimate long-term variability in sublimation.

2.2 Methods

We used data collected during the Surface Atmosphere Integrated Field Laboratory (SAIL), Study of Precipitation, the Lower Atmosphere and Surface for Hydrometeorology (SPLASH), and Sublimation of Snow (SOS) field campaigns. The SAIL and SPLASH campaigns spanned October 2021 through June 2023, and the SOS campaign spanned October 2022 through June 2023. These campaigns used a diverse set of observations to gain insight into surface-atmosphere interactions in a mountainous and seasonally snow-covered environment [*de Boer et al.*, 2023; *Feldman et al.*, 2023; *Lundquist et al.*, 2024].

2.2.1 Study Area

These campaigns primarily took place in the upper East River Basin (UERB), a headwater catchment of the Colorado River. The UERB has garnered significant research interest and investment due to its hydrologic importance, long observational record, and low levels of human impact [*Hubbard et al.*, 2018]. It lies high in the Rocky Mountains of central Colorado on the western slope of the Continental Divide and is oriented northwest-to-southeast (Figure 2.1A and B). Its terrain consists of steep topographic relief and seasonal snow cover that lasts for six or seven months of the year. Water year (1 October - 30 September) precipitation measured at the valley bottom averages 916 mm, 75% of which falls as snow [*Prather et al.*, 2023]. The average annual temperature is 2° C and median peak snow water equivalent (SWE) is 373 mm [*Natural Resources Conservation Service,*, 2025]. The basin's vegetation is characterized by mixed stands of aspen, subalpine fir, and Engelmann spruce along the valley slopes up to treeline, while the valley bottom is dominated by fescue grasses and sagebrush with sparse tree cover [*Cox et al.*, 2025]. This valley bottom vegetation is generally buried during continuous snow cover. Measurements from each campaign were collected at three

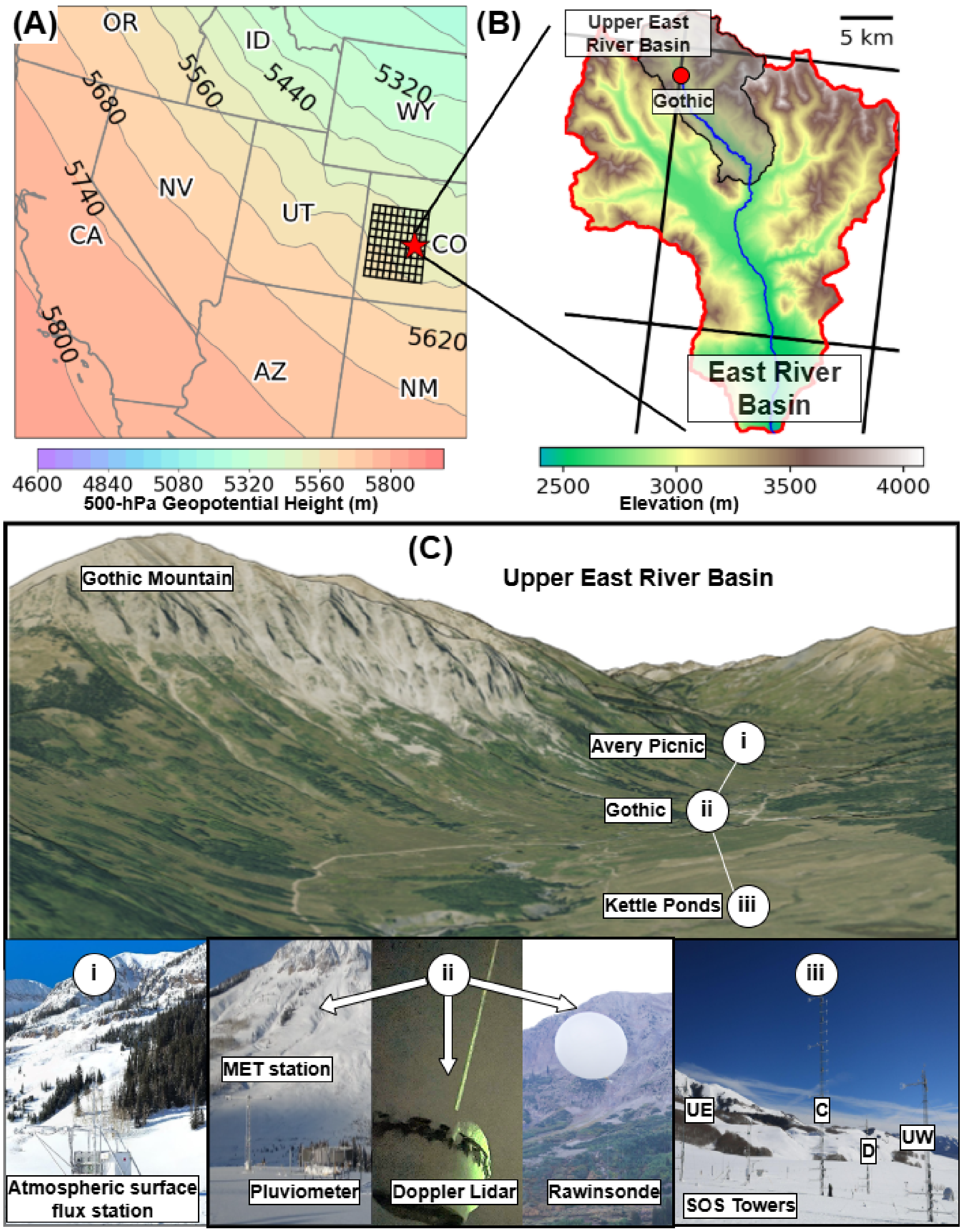

Figure 2.1: (A) Study location within the western United States. Grid-boxes indicate the spatial extent of reanalysis data used to model sublimation event occurrence. Colors and contours indicate an example of ERA5 geopotential height data at 500-hPa. (B) Inset of the East River Basin and UERB. (C) Study site locations with photos of the instruments used in this study at (i) Avery Picnic, (ii) Gothic, and (iii) Kettle Ponds.

sites along the valley bottom: Avery Picnic, Gothic, and Kettle Ponds.

Site Descriptions

Avery Picnic The Avery Picnic (AP) site (38.972° N, 106.997° W at 2934 m MSL) was the upper-most site (Figure 2.1Ci). The valley is approximately 250-m wide at this point. The up-valley fetch is unobstructed for approximately 1-km and has a shallow 3° gradient. The SPLASH campaign installed instrumentation at AP to observe near-surface meteorological conditions and surface energy exchanges [*Cox et al.*, 2023, 2025]. *Cox et al.* [2025] describes this instrumentation and data processing steps in further detail.

Gothic The Gothic site (38.956°N, 106.988°W at 2886 m MSL) sits at the southeast base of Gothic Mountain and is located approximately 2-km down-valley from AP (Figure 2.1Cii). Gothic hosted the Department of Energy's Atmospheric Radiation Management Mobile Facility under the SAIL campaign where several instruments were installed to study near-surface and valley-scale conditions [*Feldman et al.*, 2023].

Kettle Ponds The Kettle Ponds (KP) site (38.975°N, 106.973° W at 2861 m MSL) lies about 2-km southeast of Gothic near the center of the valley, which is approximately 1500-m wide at this point (Figure 2.1Ciii). The site is characterized by flat to rolling terrain with a shallow, 4°-slope and an unobstructed fetch for approximately 2-km along the valley [*Lundquist et al.*, 2024].

The SOS campaign partnered with the NSF National Center for Atmospheric Research to deploy the Integrated Surface Flux System at KP, which included a suite of instruments mounted on four towers (three 10-m tall and one 20-m tall). We used these instruments to observe the seasonal evolution of near-surface meteorology, the surface energy balance, and the snowpack [*Lundquist et al.*, 2024]. SPLASH also measured the surface energy balance at KP [*de Boer et al.*, 2023], and we used these measurements to gap-fill missing data.

2.2.2 Site-scale analysis

Instrumentation and Data

We focus on the 1 December to 31 March 2021-2022 and 2022-2023 winter seasons to analyze conditions strictly during snow accumulation and prior to spring melt. AP data only covered winter 2022 due to limited turbulent flux observations in 2023 [*Cox et al.*, 2025], whereas SOS data were limited to winter 2023, the only winter that campaign operated.

Instrument specifications are compiled in Table 2.1 and discussed in *Lundquist et al.* [2024], *Feldman et al.* [2023], and *de Boer et al.* [2023]. Unless otherwise noted, all data were resampled to 30-minute averages/sums, and measurement heights are relative to snow-free ground.

Near-Surface Meteorology At KP, a ventilated and shielded hygrothermometer on the central (C) tower measured temperature and relative humidity. Since temperatures were rarely above freezing each winter, relative humidity was computed with respect to ice [*Huang*, 2018]. Atmospheric pressure was measured by nanobarometers atop each tower. We adjusted these measurements following a hydrostatic model of the atmosphere to estimate surface pressure. At AP, air temperature, pressure, and humidity were measured using a probe on the flux station boom (Table 2.1). Vapor pressure deficit (VPD) was calculated from these observations at both sites. Horizontal wind speed and direction were calculated from the u and v components of the sonic anemometers.

Since neither KP nor AP data spanned both winters continuously, we used observations from SAIL's Meteorology System (MET) at Gothic to compare surface meteorology between winters. The MET included standard sensors measuring surface wind speed, wind direction, air temperature, relative humidity, and atmospheric pressure [*Kyrouac and Tuftedal*, 2024].

Precipitation Precipitation was measured at Gothic by a weighing bucket pluviometer with a double Alter wind shield and aggregated to 30-minute totals (Figure 2.1Ciii). Compared to other precipitation measurements available, this pluviometer most closely matched daily totals from the long-term Gothic gauge, a record consistently maintained by the same observer, billy barr [sic], since the mid-1970s [*Prather et al.*, 2023]. During winter 2023, four fluidless snow pillows were also deployed at KP to directly measure SWE. We used these measurements to determine if the snow pillows could detect sublimation-driven decreases in SWE.

Surface Energy Balance At KP, heated pyranometers and pyrgeometers on the SOS downwind (D) tower measured incoming and outgoing shortwave and longwave radiation. SPLASH also deployed a surface energy balance monitoring system at KP [*de Boer et al.*, 2023]. Because of its proximity to the SOS radiometer and strong agreement in net radiation (R_{net}) during winter 2023 (Pearson's r=0.90, RMSE=17 W m^{-2}), we used this system to gap-fill missing SOS radiometer data (3% of the data record). The AP site was similarly outfitted with pyranometers and pyrgeometers [*Cox et al.*, 2025]. Solar radiation data were dropped whenever outgoing solar radiation exceeded incoming radiation, likely due to snow covering the sensor.

All surface energy balance variables follow a consistent sign convention: positive (incoming) fluxes represent energy directed into the snow surface, while negative (outgoing) fluxes represent energy loss from the snow surface. We report R_{net} as the sum of net shortwave and net longwave radiation.

Surface Temperature At KP and AP, snow surface skin temperature (T_s) was measured by infrared sensors and corrected for surface emissivity following:

$$T_s = \left(\frac{\sigma T_c^4 - (1-\varepsilon) LW_{out}}{\varepsilon \sigma} \right)^{0.25}, \tag{2.1}$$

where ε is the surface emissivity, σ is the Stefan-Boltzmann constant (5.670×10^{-8} W m^{-2} K^{-4}), T_c is the sensor temperature, and LW_{out} is the outgoing longwave radiation. The snow emissivity for the sensor's spectral range was set to 0.99 [*Warren*, 1982].

Blowing Snow Blowing snow mass flux (g m^2 s^{-1}) at KP was measured by two, 1-m long acoustic blowing snow sensors [*Trouvilliez et al.*, 2015] that were vertically stacked along the upwind east (UE) tower base between 0 and 2-m. A blowing snow threshold of 1 g m^2 s^{-1} [*Trouvilliez et al.*, 2015] was used to distinguish periods of strong blowing snow from calmer ones.

Turbulent Fluxes and Sublimation Turbulent fluxes of sensible (H_s) and latent heat (H_l) were calculated using the EC method [*Foken et al.*, 2012; *Reba et al.*, 2009]. At each site, EC systems included three-dimensional sonic anemometers and open-path gas analyzers. Raw EC data were recorded at 20 Hz. Although a 10-m measurement height is preferable for estimating total sublimation [*Schwat et al.*, 2025], this was not feasible because the only available measurement at AP was at 4.6 m. To maintain consistency across sites, we therefore averaged the four 3-m EC measurements from the SOS towers at KP (Table 2.1).

To prepare the 20 Hz data for flux calculations, the planar fit method was used to correct for sonic anemometer tilt [*Wilczak et al.*, 2001]. We then calculated turbulent fluxes using a 30-minute Reynolds averaging length and applied standard corrections [*Schotanus et al.*, 1983; *Webb et al.*, 1980]. *Schwat et al.* [2025] and *Cox et al.* [2025] discuss the specific post-processing and data quality of the SOS and SPLASH EC datasets, respectively, in further detail.

After averaging, we gap-filled missing data (5% for SOS and 9% for SPLASH) using two strategies that depended on the duration of each gap. For gaps shorter than 1 hour, we used linear interpolation. For longer gaps, we used the mean diurnal variation method with a 3-day moving window average [*Finnigan et al.*, 2003]. Lastly, we excluded flux data whenever more than 0.5-mm of precipitation fell within 30-minutes to reduce the influence of precipitation on flux measurements.

While blowing snow can introduce additional uncertainty into EC measurements, data quality remained high during most blowing snow events, except one (21 December-22 December 2022). For that event, sublimation totals were likely undermeasured [*Schwat et al.*, 2025].

We calculated sublimation from the corrected H_l measurements in units of W m^{-2}, which we converted to sublimation rate (S) in millimeters of SWE per hour (mm hr^{-1}) following:

$$S = \frac{H_l \Delta t}{\rho_w L_s}, \tag{2.2}$$

where Δt is the time interval in seconds (3600 seconds), ρ_w is the density of water (1000 kg m^{-3}), and L_s is the latent heat of sublimation at 0°C (2835 kJ kg^{-1}).

Table 2.1: Observed variables, instruments, and measurement details. Sampling intervals denoted with an asterisk (*) were sampled more frequently but averaged to these intervals on the data logger.

Variable	Instrument	Height (m)	Site	Period	Units	Sampling Interval
Temperature	Vaisala PTU307	2.89	AP	Winter 2022	°C	10 min*
	Vaisala HMP155	2	Gothic	Both winters		1 min*
	Sensirion SHT85	3	KP	Winter 2023		5 min*
Humidity	Vaisala PTU307	2.79	AP	Winter 2022	%RH	10 min*
	Vaisala PTB330	2	Gothic	Both winters		1 min*
	Sensirion SHT85	3	KP	Winter 2023		5 min*
Pressure	Vaisala PTU307	2.79	AP	Winter 2022	hPa	10 min*
	Vaisala PTB330	1	Gothic	Both winters		1 min*
	Paroscientific 6000	10	KP	Winter 2023		5 min*
Precipitation	OTT Pluvio2-L	–	Gothic	Both winters	mm	1 min
Snow water equivalent	Alpine Hydromet FSP5	–	KP	Winter 2023	mm	5 min
Wind speed & direction	Metek uSonic-3	4.62	AP	Winter 2022	m s^{-1}, °	20 Hz
	RM Young 05103/05106	10	Gothic	Both winters		1 min
	CSAT3	3	KP	Winter 2023		20 Hz
Shortwave radiation	Hukseflux SR30-D1	3	AP	Winter 2022	W m^{-2}	10 min*
	Kipp & Zonen CM21	10	KP	Winter 2023		5 min*
Longwave radiation	Hukseflux IR20	3	AP	Winter 2022	W m^{-2}	10 min*
	Kipp & Zonen CG4	10	KP	Winter 2023		5 min*
Surface temperature	Apogee IRT	3	AP	Winter 2022	°C	10 min*
	Apogee SIF-111	9	KP	Winter 2023		5 min*
Sensible heat flux	Metek uSonic-3	4.62	AP	Winter 2022	W m^{-2}	20 Hz
	EC150	3	KP	Winter 2023		20 Hz
Latent heat flux	LI-COR 7500DS/Metek	4.27	AP	Winter 2022	W m^{-2}	20 Hz
	EC150	3	KP	Winter 2023		20 Hz
Blowing snow	FlowCapt4	0–2	KP	Winter 2023	g m^{-2} s^{-1}	5 min*
Rawinsonde	Vaisala RS-41SGP	–	Gothic	Both winters	–	12 hrs
Vertical velocity	Doppler lidar	–	Gothic	Both winters	m s^{-1}	10 min*

Large sublimation events

We performed a peaks-over-threshold analysis to identify sustained periods of elevated sublimation. This involved testing combinations of intensity and duration thresholds to distinguish meaningful events. To determine suitable thresholds, we first resampled the EC data to a coarser temporal resolution, which served as a simple low-pass filter to minimize sensitivity to short-term variability within the time series. We found a 3-hour averaging period was best because it both smoothed high-frequency fluctuations and retained signals from changing weather patterns. Once resampled, two distinct categories of large sublimation events emerged that best balanced the intensity and duration criteria: (1) short-duration, high-intensity ("short") events, defined by 3-hour mean sublimation rates exceeding the 95th seasonal quantile ($q_{0.95}$) for up to 12 hours, and (2) long-duration, moderate-intensity ("long") events, where rates exceeded the 75th quantile ($q_{0.75}$) for more than 12 hours (Figure 2.9).

Meteorological comparison during large sublimation events

To assess the characteristics of these events at the site-scale, we compared the mean diurnal cycles of select surface observations for each sublimation event type to the average conditions over the entire study period. Although short events are defined as lasting less than 12 hours, we analyzed data from the full day on which they occurred to capture the conditions surrounding the events.

2.2.3 Valley-scale analysis

Instrumentation and Data

We used data from the twice-daily rawinsondes and vertical velocity variance statistics from the Doppler lidar to evaluate conditions within the valley atmosphere [*Feldman et al.*, 2023]. The rawinsondes were launched at around 00 UTC and 12 UTC (0500 and 1700 MST) each day from Gothic (right photo in Figure 2.1Cii), and provided a vertical profile of temperature, humidity, wind speed, and direction from the surface to the troposphere.

The Doppler lidar in Gothic (center photo in Figure 2.1Cii) was used to remotely monitor the wind field within the valley atmosphere. In its vertically staring mode, it measured profiles of vertical velocity (w) at a 1-second temporal resolution. SAIL processed these data into a 10-minute summary statistics dataset from which we used the w variance (σ_w^2). We resampled the σ_w^2 data to 30-minute resolution with 30-m vertical range gate spacing. Data product specifications and quality control details are discussed in *Newsom et al.* [2019].

Turbulence Signals from Vertical Velocity Variance

σ_w^2 values indicate the turbulent intensity at each range gate. Prior work has shown Doppler lidar σ_w^2 estimates closely match EC observations [*Bonin et al.*, 2017]. We resampled the σ_w^2 data using three range-gates to a vertical resolution of 90-m, which reduces error in the variance estimates but maintains a useful vertical resolution [*Hogan et al.*, 2009]. We also limited the data in the vertical to the lowest 2 km above the valley floor because low signal-to-noise ratios reduced data quality higher up [*Shukla et al.*, 2019].

To assess how σ_w^2 differed by sublimation event type, we combined data from both years and compared mean σ_w^2 vertical profiles during large sublimation events to the entire study period. We did the same for vertical profiles of wind speed and direction using averaged data from the 00 UTC (afternoon) soundings.

Inversion Detection and Characterization

We used the rawinsonde temperature profiles to characterize temperature inversion structure. Valley temperature inversions strongly impact energy distribution throughout mountain valleys. When inversions are deep and persistent, which is common during periods of extensive snow cover, surface conditions (e.g. temperature, wind speed, humidity) decouple from the free atmosphere conditions. In contrast, when inversions are weak or shallow, these same surface conditions often change alongside the conditions aloft [*Whiteman and Doran*, 1993; *Lundquist et al.*, 2008; *Whiteman*, 1982].

For each sounding, we identified inversions wherever the vertical potential temperature gradient ($\frac{d\theta}{dz}$) was positive for > 100 m. We then calculated inversion depth (D) and bulk intensity (I), following *Gilson et al.* [2018]. D was calculated using the heights of the bottom (z_{bottom}) and top (z_{top}) of the inversion, defined as the heights where $\frac{d\theta}{dz}$ changed sign from negative to positive (z_{bottom}) and then back again (z_{top}). If the inversion reached the surface, as was common, z_{bottom} was set to the rawinsonde launch elevation (2861 m).

We quantified I as:

$$I = \theta_{z_{top}} - \theta_{z_{bottom}}, \tag{2.3}$$

where $\theta_{z_{top}}$ and $\theta_{z_{bottom}}$ are the potential temperatures at the top and bottom of the inversion, respectively. An inversion was defined as strong if D was greater than 100-m and I greater than 2 K [*Whiteman*, 1982]. To gauge how well connected the valley atmosphere was to conditions aloft, we compared the summary statistics of D and I during each type of large sublimation event and the entire study period.

2.2.4 Synoptic-scale analysis

To link synoptic-scale conditions to large sublimation event occurrences in the UERB, we used the information gained from the site- and valley-scale analyses, along with atmospheric reanalysis data, to develop a binary classification model. Here, we focused exclusively on long-duration events because they consistently coincided with synoptically active periods, making them easier to distinguish from background conditions in the reanalysis record. We tested several models of varying complexity and chose the one that performed best based on evaluation metrics suited for imbalanced classification problems. The chosen model, random forest, is a machine learning algorithm that builds an ensemble of decision trees using bootstrapped samples of the data and random subsets of features to improve predictive accuracy through averaging [*Breiman*, 2001]. Our results only focus on this selected model; however, the training and evaluation strategy was the same for all developed models. A full description and performance comparison is provided in Appendix B.

Reanalysis Inputs and Target Labels

The model's input data came from the European Centre for Medium-Range Weather Forecasting (ECMWF) Reanalysis v5 (ERA5) [*Hersbach et al.*, 2020]. ERA5 assimilates meteorological and remotely sensed observations into the ECMWF modeling system to provide a global and temporally complete representation of the atmosphere. Its dataset includes numerous meteorological variables at multiple pressure levels and is available at a 1-hour temporal and 31-km (0.25°) spatial resolution from 1940 onward.

To reduce data volume, we analyzed a spatially and temporally constrained subset of ERA5 data at the 500-hPa pressure level, which limits surface influences from complex terrain and represents the large-scale circulation well [*Hersbach et al.*, 2020]. We chose the vector wind (u, v), geopotential height, temperature (T), specific humidity (q), and relative vorticity (ζ) variables and downloaded them using the ECMWF Climate Data Store API [*Copernicus Climate Change Service*, 2023]. These variables capture key atmospheric processes relevant to sublimation: winds and geopotential heights describe flow patterns aloft; temperature and specific humidity indicate the thermal energy and moisture availability; and relative vorticity signals cyclonic activity associated with approaching or departing weather systems. From these variables, we also calculated temperature advection (A_T), moisture advection (A_q), and vorticity advection (A_ζ), defined as the dot product of the vector wind and the gradient in each property. These advection terms help characterize the presence of upstream thermal, moisture, and vorticity advection not evident in the raw fields alone. We excluded precipitation from this analysis due to known limitations in ERA5 sub-daily precipitation over complex terrain [*Lavers et al.*, 2022] and to independently evaluate the relationship between observed snowfall and large sublimation events.

Next, we limited our analysis period to the satellite era (1979–present) to avoid discontinuities due to differing data assimilation methods and restricted the data to a 2.5°-by-2.5°area centered 1°west of the UERB (Figure 2.1A). To reduce the dimensionality of the dataset, we computed the spatial mean of each variable across this subset region. Finally, we restricted the dataset to December–March and resampled it to 3-hour intervals to match the resolution of large sublimation events.

We trained and evaluated our model using a target dataset built using long sublimation event observations from the 2022 and 2023 winter seasons. To prepare the target dataset, we constructed a binary time series that indicated when these events occurred at our sites.

Predicting large sublimation event occurrence

Model training, evaluation, and selection To develop the model, we limited our input data to match the target dataset and partitioned each into a 90% training set and 10% hold-out validation set. Since sublimation events were rare relative to non-events, we used stratified 10-fold cross-validation to preserve class imbalance and reduce sensitivity to temporal patterns, using the same 90-10 train/test split for each fold. We used GridSearchCV from the python package scikit-learn [*Pedregosa et al.*, 2011] to tune select hyperparameters within a defined hyperparameter space. Further training details are provided in the code repository provided in the data availability statement.

We evaluated model performance using the F1-score and average precision calculated from the cross-validation results and compared the mean evaluation metrics and balanced accuracy for each model (Appendix B). Among the models we tested, the random forest model performed best on the independent validation set (Figure 2.12), with an F1-score of 0.64, average precision of 0.73, and balanced accuracy of 79%. Because RQ3 seeks to identify changes in the frequency and duration of large events over entire seasons rather than individual events, we found this performance adequate to address RQ3.

Applying the model We applied the trained model to the reanalysis dataset (December 1979-–March 2025) to analyze trends and variability in long sublimation event occurrences. For each winter, we summed the total number of hours classified as long events (referred to as "event hours"). To assess trends, we used the Mann-Kendall test to detect monotonic changes in event hours at a significance level of $\alpha = 0.05$. To evaluate variability, we compared the monthly average distribution of event hours and examined how this distribution shifted decade-to-decade within the winter season.

2.3 Results

2.3.1 Site-scale analysis

The average wind speed and direction matched closely across winters (Table 2.2), with north-westerly down-valley winds occurring 45–65% of the time (Figure 2.2). However, other meteorological variables had large differences. When compared to winter 2022, winter 2023 had a markedly lower mean temperature, lower mean VPD, and 36% more precipitation (Table 2.2). Another noteworthy difference was how precipitation accumulated across each winter (Figure 2.3). In winter 2022, 42% (153 mm) of the season's precipitation fell during a stormy 10-day period in late December 2021. This massive event was followed by two dry months before ending with a few late winter storms and another extended dry period. In contrast, precipitation accumulated more consistently during winter 2023 due to a relatively continuous sequence of low pressure systems throughout the winter with fewer extended dry periods.

Table 2.2: Winter meteorological conditions for 2022 and 2023. Temperature, vapor pressure deficit, and wind speed are reported as the mean ± standard deviation. Cumulative precipitation and cumulative sublimation are provided as sum totals.

Winter Season	Temperature (°C)	VPD (Pa)	Wind Speed (m s^{-1})	Cumulative Precipitation (mm)	Cumulative Sublimation (mm)
2022	-6.8 ± 4.8	164 ± 101	2.3 ± 1.1	363	21.5
2023	-8.2 ± 4.2	127 ± 66	2.2 ± 1.2	494	22.0

Despite these substantial differences in precipitation, cumulative sublimation was nearly identical, with 21.5 mm in 2022, and 22.0 mm in 2023 (Table 2.2). As a result, sublimation losses represented 5.9% and 4.4% of seasonal precipitation for 2022 and 2023, respectively. Even with these comparable totals, sublimation totals accumulated differently between these two years as each winter was punctuated by distinct sublimation increases over short periods

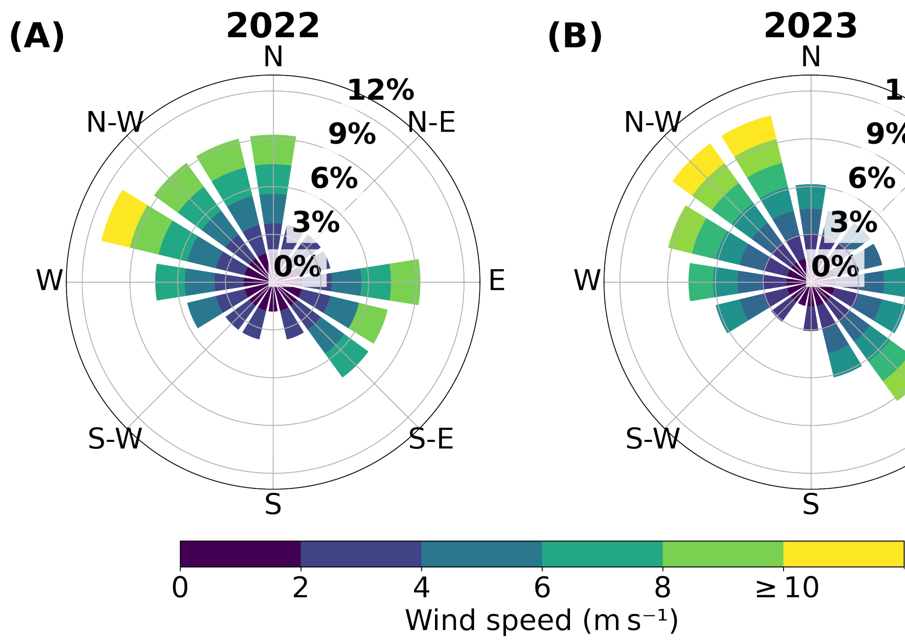

Figure 2.2: Comparison of wind roses for the 2022 (a) and 2023 (b) winters from Gothic MET station.

(Figure 2.3).

Blowing Snow and Sublimation

Most blowing snow occurred during or soon after snowfall, accounting for 17% and 41% of all blowing snow observations, respectively (Figure 2.4A). However, 12% still occurred in the 12 hours leading up to precipitation events. These observations not only emphasize the importance of new snow to enhance blowing snow, but also the influence an approaching storm can have on blowing snow even prior to new snowfall.

Sublimation rates generally increased with higher wind speeds and blowing snow fluxes, but the importance of blowing snow differed between short and long events. High rates of blowing snow were more common during long events (squares in Figure 2.4B) than short events (triangles), with 71% of all strong blowing snow flux measurements (≥ 1 g m^{-2} s^{-1}) occurring during long events.

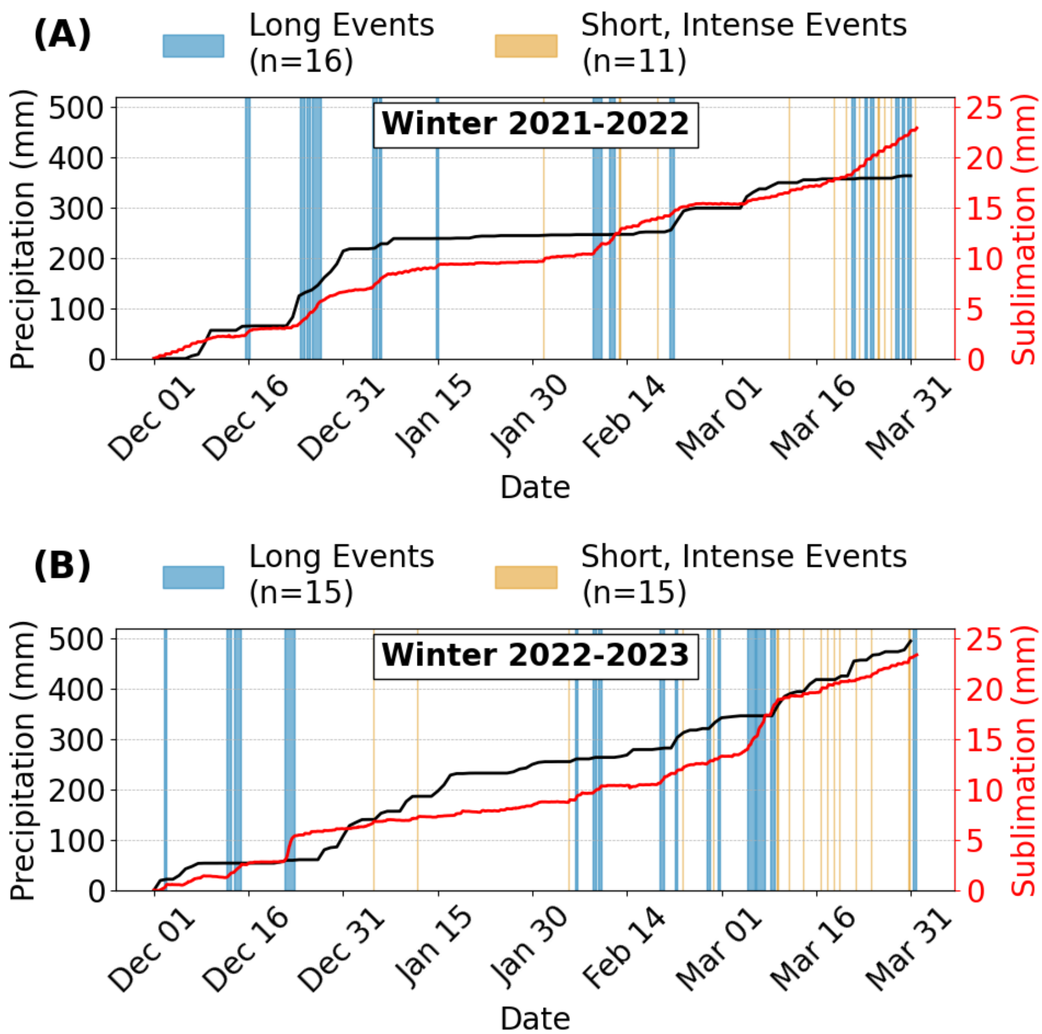

Figure 2.3: Comparison of cumulative precipitation (black line) and cumulative sublimation (red line) accumulation for the (A) 2022 and (B) 2023 winters. Long (blue bars) and short (orange bars) indicate when sublimation events occurred. Total event numbers are shown in the legend, indicated by n.

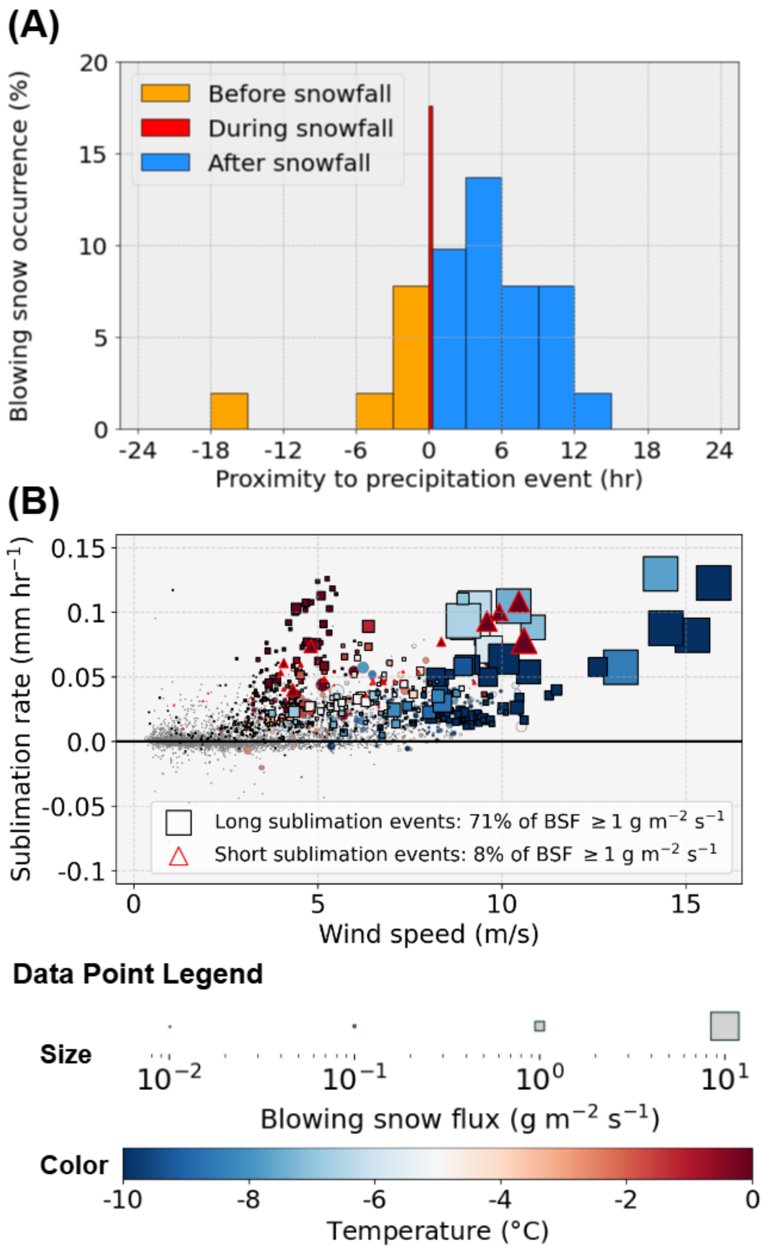

Figure 2.4: Relationships between wind speed, blowing snow and sublimation during winter 2023. (A) Blowing snow in the 24 hours surrounding snowfall at KP. Each vertical bar represents the percentage of time that blowing snow was observed before (orange bars) and after (blue bars) snowfall. The vertical red bar represents how often blowing snow was measured during snowfall. (B) 30-min wind speed plotted against sublimation rate at KP. Points are colored by air temperature and sized by average blowing snow flux (see data point legend). Points classified as strong blowing snow flux (≥ 1 g m^{-2} s^{-1}) during large and short sublimation events are outlined in black and red, respectively; all others are outlined in gray.

Large Sublimation Events

Winter 2022 had a total of 16 long events that averaged 22 hours in duration. These events accounted for 42% of the total winter sublimation (Table 2.3). In comparison, winter 2023 had 14 long events that were slightly longer on average (23 hours) and made up 49% of that winter's sublimation. In both winters, long sublimation events lasted for less than 12% of the winter. In contrast, short events were more numerous in 2023 than in 2022. They made up 15% of total sublimation in each winter, but occurred over only 2.5% of each winter's duration (Table 2.3). Together, long and short sublimation events accounted for the majority of winter sublimation (57% in 2022 and 64% in 2023), despite occurring during only 14% of each winter season, on average.

Table 2.3: Comparison of long and short sublimation events during the winter seasons of 2022 and 2023. The table shows the average duration, the percentage of total sublimation contributed, and the percentage of the winter season's length associated with each type of event.

	Long events		**Short events**	
	2022	2023	2022	2023
Average duration (hours)	22	23	4.5	3.6
% of total sublimation	42%	49%	15%	15%
% of winter length	11%	12%	2.5%	2.5%

The short events clustered mostly in the later portion of each winter season, whereas the long events were more evenly distributed throughout (Figure 2.3). Short events became more common as solar angles increased and temperatures rose during the transition to spring. However, long events appeared to arise from different processes: most (63% in 2022 and 79% in 2023) started within a day after measurable precipitation and coincided with high winds (Figure 2.4B). This relationship aligns with the observation that wind speed showed its strongest correlation with precipitation during or shortly after precipitation fell (Figure 2.10). To better understand these processes, we next compare event meteorology to wintertime averages across the diurnal cycle.

Meteorological comparison during large sublimation events

The diurnal cycles of surface and air temperature, VPD, wind speed, and blowing snow flux differed across sublimation event types and the study period average (Figure 2.5). During long events, surface and air temperatures were near the top quartile of the wintertime average and were especially warm at night (Figure 2.5B). The same was true for VPD, wind speed, and blowing snow flux. Each lacked any clear diurnal pattern, remained entirely above the study period inter-quartile range (IQR), and had especially large differences from the average at night (Figures 2.5F, H, and J).

During short events, the diurnal pattern in these observations was much clearer. Each followed the average winter pattern but with greater amplitude (Figures 2.5A, E, G, and I). Surface and air temperatures rose above the winter average during the day and were colder at night. The same was true for the patterns in VPD, wind speed, and blowing snow flux: increases during the day and decreases at night. However, the VPD pattern was more extreme than either the wind speed or blowing snow flux patterns during the day, as VPD values rose above the top quartile.

During both event types, the diurnal R_{net} pattern qualitatively followed that of the study period average; however, its range was amplified during short events, with R_{net} far above the top quartile during the day and below the bottom quartile at night (Figures 2.5C). This exaggerated pattern suggests short events mainly occurred under clear skies. On the other hand, the R_{net} during long events matched the study period average quite closely (Figure 2.5D).

During short events, the diurnal pattern of turbulent fluxes also qualitatively matched the average winter pattern (Figure 2.5C). However, relative differences were apparent throughout the day. For instance, as temperature peaked in the mid-afternoon, H_s increased relative to the average as R_{net} rapidly decreased. Additionally, H_l values fell outside the IQR for much of the day, while the diurnal H_l pattern was more exaggerated and peaked slightly

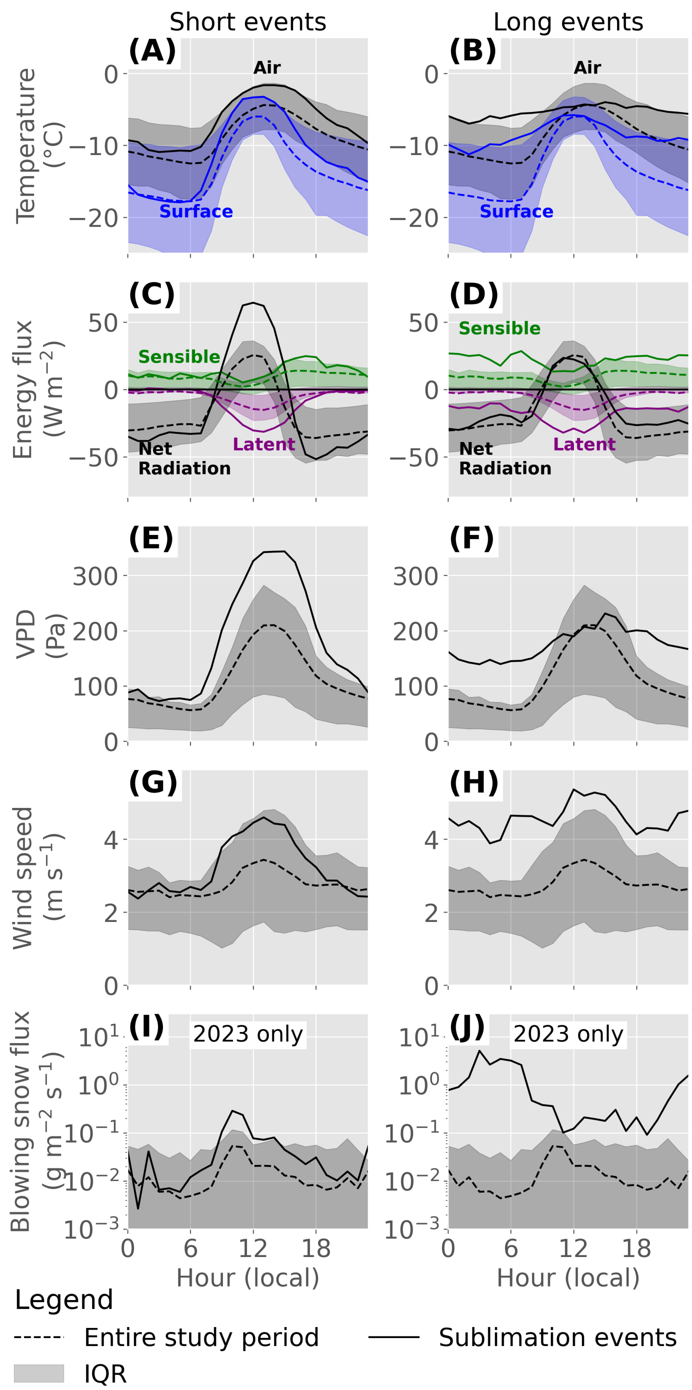

Figure 2.5: Mean diurnal surface and air temperature, energy fluxes, vapor pressure deficit, wind speed, and blowing snow flux for short (A, C, E, G, I) and long (B, D, F, H, J) sublimation events (solid lines) compared to averages from the entire study period (dashed lines). Shaded regions show the inter-quartile range (IQR) during the entire study period.

later compared to the average.

Contrary to short events, the diurnal patterns in the turbulent fluxes were very different during long events (Figure 2.5D). Both H_s and H_l were elevated in the absolute and generally balanced one another at night, as positive H_s directed energy towards the snow surface and negative H_l took energy away from it. However, this pattern shifted during the day. H_s and H_l were less balanced as H_s decreased, while H_l remained elevated in the absolute. This shift happened when R_{net} flipped from negative to positive. At that point, R_{net} began to supplant the energy provided by H_s to continue to support and enhance the large negative H_l values. Even so, both H_s and H_l sat far outside the IQR of average conditions throughout the entire diurnal cycle.

Although these events had clear patterns in their near-surface meteorology, we could not identify a corresponding relationship between sublimation and SWE change at the KP snow pillows. Even during the largest events, sublimation-driven SWE changes were too small to detect relative to competing signals from recent snowfall and wind redistribution.

2.3.2 Valley-scale analysis

Turbulence signals from vertical velocity variance

We observed distinct differences in the mean σ_w^2 profiles between sublimation event types and the study period average (Figure 2.6A). During average conditions, σ_w^2 increased slightly from the surface through the valley atmosphere. Once above the mean ridgeline elevation (605 m), σ_w^2 decreased slightly. We observed a more distinct pattern to the σ_w^2 profile during sublimation events.

During short events, mean σ_w^2 values increased from 0.6 m^2 s^{-2} near the surface to a max of 2.9 m^2 s^{-2} at around 250 m above the surface. Values then decayed to 1.0 m^2 s^{-2} above this level. This same pattern occurred during long events but was more extreme: mean σ_w^2 values rapidly increased from 1.0 m^2 s^{-2} near the surface to a max of 4.6 m^2 s^{-2} near the mean ridgeline elevation before decreasing, as well.

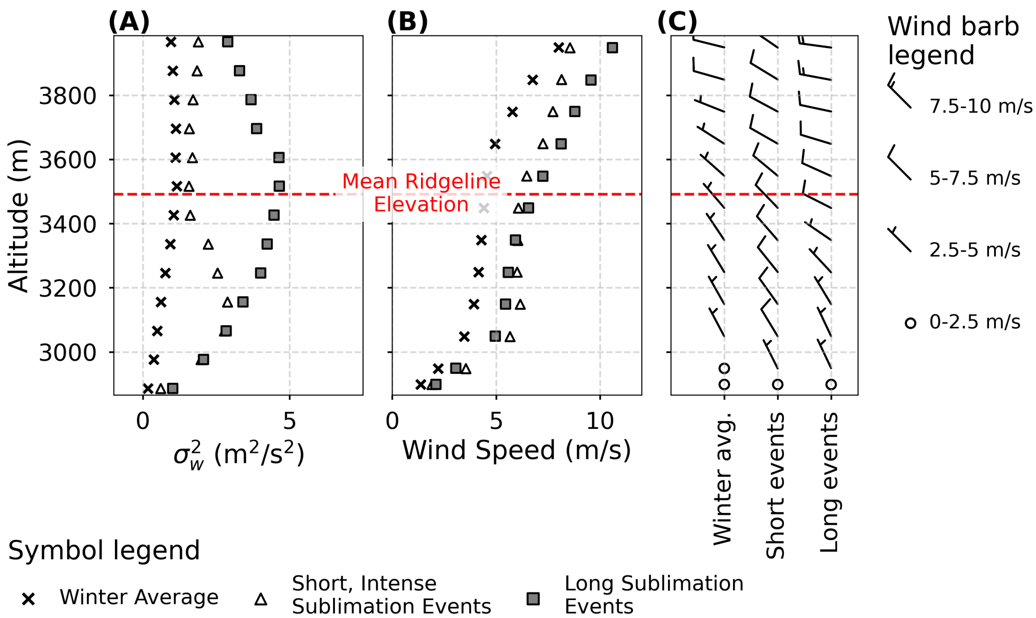

Figure 2.6: Mean profiles of (A) vertical velocity variance, (B) wind speed, and (C) wind direction. Mean ridgeline elevation, which was 605 m above the valley floor (2900 m), is indicated by the dashed red line. In (A) and (B), mean winter conditions are marked with an X, long sublimation events are marked by squares, and short sublimation events are marked by triangles. (B) and (C) use data from the 00 UTC (1700 MST) soundings that were then averaged for each event type.

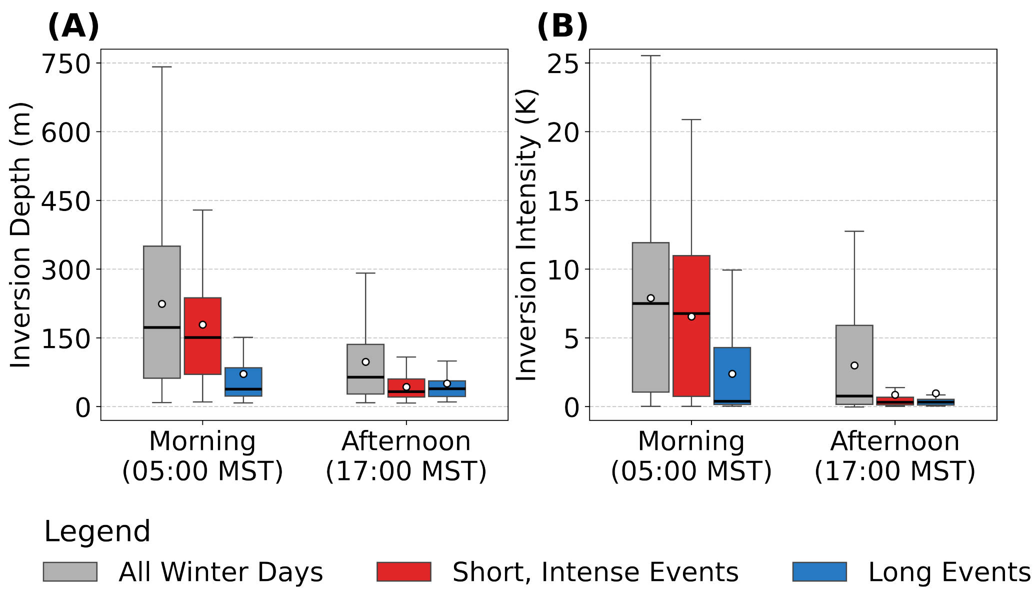

Figure 2.7: Box plots of the average inversion depth (A) and inversion intensity (B) within the valley between the morning and afternoon rawinsonde releases for the entire study period, days with long sublimation events, and days with short sublimation events. The thick black line and white dot represent the median and mean values, respectively. Each colored box encompasses the IQR, and the caps at the end of each thin line indicate the 5th (bottom) and 95th (top) percentile values.

The profiles of wind speed and direction show similar characteristics to the σ_w^2 profiles (Figure 2.6B and C), with wind speeds greater during sublimation events compared to average conditions. This wind speed difference grew with altitude and was greatest during long events. As a result, wind shear was greater during these events, even though the wind directional change was similar between event types (Figure 2.6C).

Inversion Detection and Characterization

Strong temperature inversions were common during each winter, occurring in 64% and 37% morning and afternoon soundings, respectively. They were more common in 2022 ($n = 53$) than in 2023 ($n = 42$). Inversions generally decreased in depth and weakened in the afternoon (Figure 2.7). The median morning inversion depth and intensity were 224 m and 7.9°C, compared to the afternoon inversion depth and intensity of 98 m and 3.0°C.

Morning inversions during long sublimation events were comparatively smaller and weaker, with average depth and intensity of 71 m and 1.0°C, respectively. In the afternoon, inversions (if present) were weaker and shallower still. However, on mornings when short events occurred, inversions resembled those during the study period average, but they typically broke down by the time of the afternoon sounding, indicating greater daytime destabilization.

2.3.3 Synoptic-scale analysis

Predicting large sublimation event occurrence

The 500-hPa wind speed emerged as the most influential reanalysis variable from the random forest model's variable importance scores, contributing to 23% of the model's predictive power (Figure 2.13). Given its high importance, wind speed correlated closely with event hours (Spearman's r = 0.62). Since storm events are an important contributor to long sublimation events, we also compared the total December-March precipitation from billy barr's [sic] precipitation station to event hours. Interestingly, although precipitation was not included in our model, it showed a moderate, positive correlation with event hours (Spearman's r = 0.5), further supporting the idea that precipitation events and large sublimation events are related.

The modeled total sublimation event hours were highly variable from one winter to the next, ranging between 24 and 363 hours, with a mean of 120 hours and standard deviation of 73 hours (Figure 2.8A). As a result, any trends had to be large in order to be significant. Although Mann-Kendall results indicated slight increasing trends in both modeled event hours per winter and mean winter wind speed, these trends were insignificant: $p = 0.24$ for event hours and $p = 0.08$ for wind speed, respectively. Additionally, the average duration of individual events has not been changing ($p = 0.44$) (Figure 2.8D).

Modeled long sublimation event hours consistently peaked in December and January before declining in February and March. While this pattern has remained consistent, the

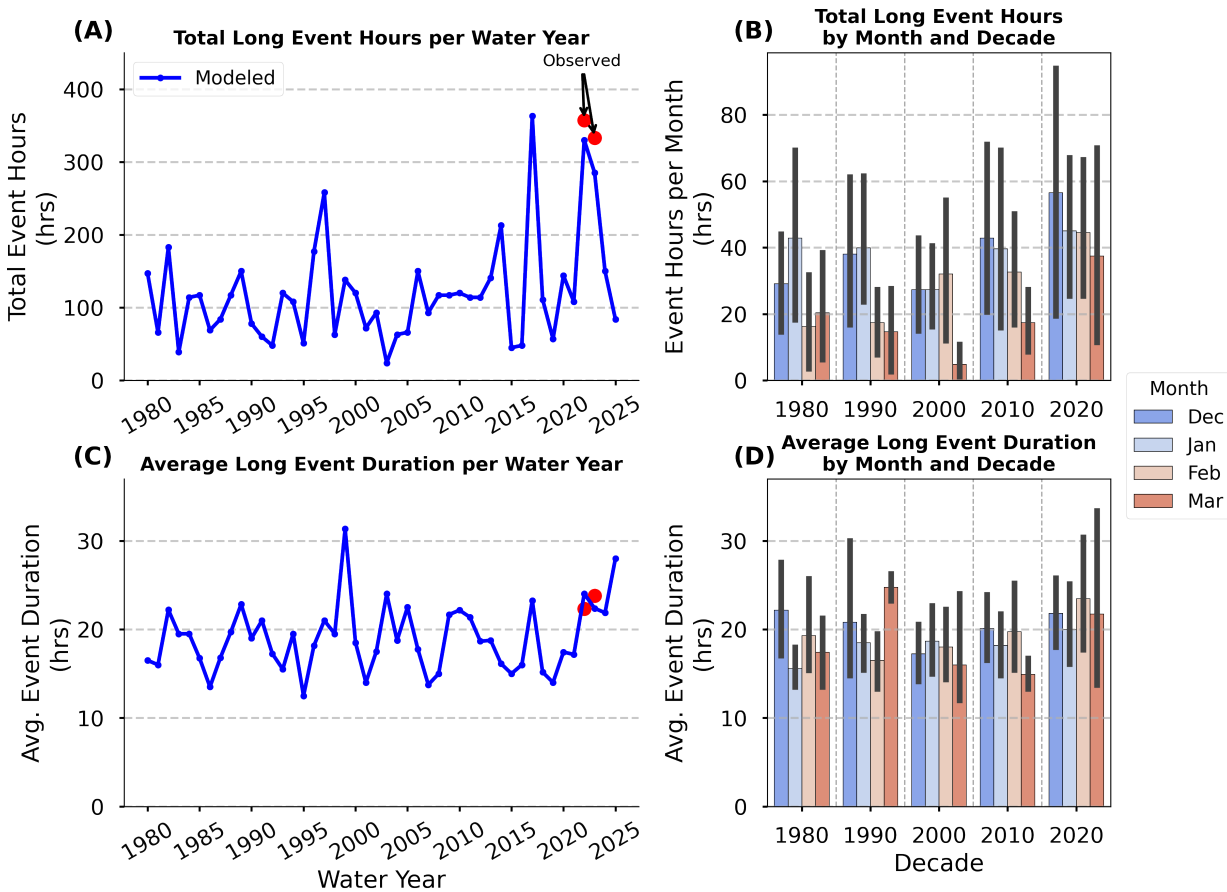

Figure 2.8: Model results from the trained random forest model. (A) Total long event hours per water year. (B) Event hours by month across each decade. (C) Average duration of long sublimation events per water year. (D) The average duration of events by month and decade. Grey bars indicate the 95% confidence interval for each month within each decade.

distribution across months has become more uniform in recent decades as event hours have increased later in the winter since 2000, although uncertainty is large (Figure 2.8B). Nevertheless, the average duration of events, both within each winter and across decades, remained stable (Figure 2.8D).

2.4 Discussion

2.4.1 The importance of large sublimation events

Following the idea that high sublimation totals correspond with warm, dry and windy conditions [*Guo et al.*, 2023; *Grundstein and Leathers*, 1999; *Harpold and Brooks*, 2018], the warmer, drier, and slightly windier conditions in winter 2022 would suggest greater sublimation than in 2023. Thus, it was somewhat surprising that accumulation season sublimation resulted in similar totals across the two winters (Table 2.2). While these driving processes are still important, we found that average winter conditions did not control sublimation totals. Instead, roughly 60% of total sublimation occurred over just 14% of each winter examined, underscoring that sublimation is concentrated in a relatively small number of impactful events.

2.4.2 Sublimation event characteristics

Building on this, we used observations from site to synoptic-scales to address RQ2 by classifying events into two categories based on their duration and sublimation intensity: short with high intensity or long with moderate intensity. We found they each showed distinct surface and valley-scale characteristics. The short events were concentrated later in winter and occurred mainly during the afternoon, generally coinciding with warm, springlike conditions. They had two key features: a breakdown of the morning valley inversion and an ample supply of energy from high R_{net} under relatively warm and dry conditions. The inversion breakdown shifted the valley atmosphere from stable, where very strong temperature gradients (Figure 2.7B) limited near surface mixing, towards neutral, where higher winds and associated turbulence enabled mixing of drier, warm air to the surface. Due to the cold snow surface, inversion breakdown was likely caused by the top-down heating of the inversion layer which leads to subsidence, as observed by *Whiteman and McKee* [1982]. The energy supply for sublimation was clearly tied to R_{net} during these events, as the diurnal cycle of H_l closely followed that of R_{net} (Figure 2.5D). Because skies were generally clear during these events, incoming solar radiation dominated the R_{net} signal. As a result, these events ended shortly after sunset when R_{net} dropped.

Long sublimation events followed a different pattern. They occurred throughout the winter and were characterized by recent snowfall, windy conditions with blowing snow, weak inversions, and relatively balanced turbulent fluxes. Much of the blowing snow observed at KP occurred within 24 hours of new snowfall (Figure 2.4A), consistent with findings from other snowy regions [*Vionnet et al.*, 2013; *Gossart et al.*, 2017], and these events accounted for most strong blowing snow flux observations (Figure 2.4B). However, blowing snow during these events still varied (Figure 2.5J), and several events continued after blowing snow subsided, suggesting that factors beyond blowing snow were important for sustaining them.

One such factor may be top-down turbulence generated by strong westerly flow interacting with terrain along the Gothic Mountain ridgeline during long events (Figure 2.6C). Combined with the comparatively greater σ_w^2 aloft, this interaction likely enhanced the turbulent signal below the ridgeline (Figure 2.6B) and supported well-mixed conditions within the valley atmosphere *Schwat et al.* [2025].

These well-mixed conditions had two main effects. First, a neutrally stratified boundary layer developed that suppressed buildup of valley inversions and encouraged these events to persist (Figure 2.7). With a weaker near-surface temperature gradient, the strong winds aloft could mix relatively warm, turbulent air toward the surface. As a result, sublimation did not rely solely on energy from the R_{net} budget during these events: H_s contributed substantially. This was particularly evident at night, when high H_s values coincided with high blowing snow fluxes and generally balanced H_l values (Figure 2.5D and J), consistent with observations from other snow-covered regions [*Bintanja*, 2001a; *Aksamit and Pomeroy*, 2020]. Therefore, these events could persist as long as this alternative energy source remained available [*Bintanja*, 2001b].

2.4.3 Modeling long-term changes in sublimation events

To address RQ3, we trained a random forest classifier using observed long sublimation events and synoptic-scale reanalysis to estimate the frequency of these events since 1980. Model results showed substantial interannual variability but no significant trends in the total number, total duration, or average duration of long events (Figure 2.8A and C). The absence of significant trends in these long events suggests that increases in accumulation season sublimation alone are unlikely to have driven the observed streamflow decline in this region. Instead, recent research indicates that decreases in springtime precipitation, alongside associated increases in spring evapotranspiration, have played a more important role [*Hogan and Lundquist*, 2024]. Although our study excluded the spring period, during which evaporation and sublimation both occur, this interpretation is consistent with a greater frequency of conditions characteristic of the short event type we identified.

2.4.4 Limitations of analysis

While these findings provide valuable insights into the role discrete events play in driving sublimation, some limitations should be acknowledged when interpreting the results. First, our analysis relied on only two winters of data, which limits our ability to capture the full range of event types, especially if certain patterns did not occur during these years. For example, prior work has shown föehn wind events induce high sublimation rates in some snow-dominated regions [*Bliss et al.*, 2011; *Shestakova et al.*, 2022; *Hayashi et al.*, 2005; *Bintanja*, 2001b]. However, we found föehn wind events were uncommon in our dataset.

Second, all observations were collected at valley-bottom sites, which experience less sublimation than either wind-exposed ridgelines or forest canopies. Ridgelines often have high rates of blowing snow sublimation, while intercepted snow in forest canopies sublimates more efficiently [*Knowles et al.*, 2012; *Strasser et al.*, 2008; *Sexstone et al.*, 2018]. As a result, our estimates do not capture the spatial variability of sublimation across the basin. Even within the valley bottom, however, site-specific differences introduced additional uncertainty. For example, mean wind speeds at KP were 0.16 to 0.21 m s^{-1} higher than at AP over the two years. These higher winds at KP may have exposed that site to more blowing snow, potentially increasing sublimation relative to AP. Nevertheless, meteorological conditions that define large events, like recent snowfall, strong winds, and turbulent heat flux balance, remain important drivers regardless of measurement location.

Third, we limited our analysis to the accumulation season only. However, we acknowledge that melt season sublimation is an important contributor to total water year sublimation, and springtime weather likely shifts the dominant event type from long to short events as solar energy input increases, temperatures warm, and blowing snow decreases [*Schwat et al.*, 2025]. Yet, isolating the signal of direct sublimation becomes infeasible once snowmelt begins because EC systems will measure both sublimation and evaporation. Moreover, interannual variability in snow disappearance date alters how long sublimation can occur, further complicating comparisons across years [*Sexstone et al.*, 2018]. To avoid these issues, we focused on the accumulation season, during which sublimation is the primary mechanism of snow loss [*Liston and Sturm*, 2004]. Nevertheless, future work should seek to address how the distribution of event types changes during spring.

Last, the random forest model's ability to capture long-term trends may be limited by the short observational record, the small set of reanalysis variables, and our sole focus on long duration events. To minimize these first two limitations, we evaluated several models and selected the best-performing one; notably, all multi-variable models yielded similar variable importance metrics, suggesting our chosen predictors were appropriate. Even so, model performance could be improved by incorporating additional reanalysis variables or testing other pressure levels. We focused specifically on long-duration events because they contributed the most to total winter sublimation and were more closely linked to synoptic conditions. However, since we chose a binary classification model, we could only study event occurrence and not magnitude. A multi-class classification algorithm that separates sublimation events based on their magnitude could be an opportunity to test for trends in event magnitude. Additionally, short events may become increasingly important. Recent modeling work by *Scaff et al.* [2024] projects that conditions similar to those we identified during these events may become more common during winter in a warming climate.

2.4.5 Broader implications

Our findings offer new insight into what drives sublimation variability in a mountain basin. Large sublimation events spanned a small fraction of each winter yet contributed the majority of sublimation during the accumulation season. Thus, our ability to represent sublimation correctly is tied to our ability to identify and understand the processes that govern these events. Due to differences in terrain and weather patterns, we expect the distribution of event types to vary in other snow regimes; however, the event types themselves can still be used to identify large sublimation events elsewhere. For long events, storm-driven snowfall and wind created favorable conditions for blowing snow and turbulent mixing, which are key to sustaining elevated sublimation rates. For short events, high temperatures alone were insufficient to cause them. They also needed high net radiation, dry air, and a breakdown of valley inversions to produce higher latent heat fluxes. These mechanisms are consistent with findings from previous studies, which emphasize the importance of both meteorological forcing and surface conditions in driving sublimation [*Hood et al.*, 1999; *Xie et al.*, 2022].

Given their connection to winds and precipitation, long sublimation events are influenced by storm frequency and timing. While future changes to storm tracks remain uncertain across the Colorado region [*Bolinger et al.*, 2023], a shift toward fewer but stronger storms could amplify sublimation during individual events. Likewise, an earlier onset of spring-like conditions may increase the influence of short event frequency.

2.5 Conclusion

This study demonstrated that a small number of isolated events accounted for the majority of accumulation season sublimation in a mountain valley. Although each event was unique, we identified several key mechanisms that were consistently involved in some combination, with at least two typically present: strong wind-driven mixing from above, blowing snow following fresh snowfall, the breakdown of valley inversions, and enhanced turbulent exchange under clear skies. Note that air temperature alone is not a good predictor of sublimation rates, nor is it even in the top four. Therefore, a comprehensive dive into changing synoptic and mesoscale circulation patterns is necessary before drawing any conclusions or projections about long-term sublimation change. The event-specific conditions outlined here can guide the detection of similar events elsewhere, enabling a broader understanding of sublimation's spatial and temporal variability across mountainous terrain.

These results point to several areas for future work that span a range of scales. First, the role of top-down turbulence in complex terrain deserves further study. Since the mean ridgeline elevation consistently aligned with key features in vertical profiles of vertical velocity statistics during sublimation events, turbulence generated at ridges may play an important role on the valley-floor. Future research could try to determine the sources of turbulence during these events and relate them to physical mechanisms, something that has yet to be done in complex terrain [*Manninen et al.*, 2018]. Secondly, the breakdown of valley inversions in snow covered mountain valleys merits further study because this breakdown was a consistent feature of large sublimation events. Improving our ability to predict inversion formation and breakdown could enhance our ability to accurately detect these events. Lastly, our findings underscore the importance of resolving event-scale processes in models that simulate snowpack evolution and water availability. Therefore, if certain events are poorly captured in forcing data, or blowing snow conditions are not represented, sublimation may be grossly misrepresented.

Acknowledgments

We gratefully acknowledge funding support for this analysis from the National Science Foundation (Award 2139836), Sublimation of Snow, and the Department of Energy Earth Systems Science (Grant DE–SC0024075), Seasonal Cycles Unravel Mysteries of Missing Mountain Water. We wish to thank billy barr [sic] in particular for his decades work dedicating himself to meticulous observations. Our analysis would not be possibly without his work. Additionally, we thank the SOS, SAIL, and SPLASH project members and the winter 2021-2023 RMBL winter caretakers. Their dedicated and detailed work was instrumental in producing the data used in this analysis. We also would like to thank the Mountain Hydrology Research Group members at the University of Washington for their valuable feedback and insight.

Data Availability Statement

All data sources used in this analysis are publicly available. SPLASH data collected at the AP site can be accessed from this Zenodo repository via `https://zenodo.org/records/15283524`. SAIL data collected from the Gothic site (wind speed, wind direction, temperature, radiation, soundings, vertical velocity variance statistics) can be accessed using the DOE Atmospheric Radiation Measurement User Facility data portal via `https://www.arm.gov/research/campaigns/amf2021sail`. Precipitation observations from billy barr are available via `https://wrcc.dri.edu/cgi-bin/rawMAIN.pl?corbil`. Butte SNOTEL data can be accessed from `https://wcc.sc.egov.usda.gov/nwcc/site?sitenum=380`. SPLASH radiation data collected at KP can be accessed from this Zenodo repository via `https://zenodo.org/records/10011476`. SOS data collected at KP can be accessed from the project data repository at `https://data.eol.ucar.edu/master_lists/generated/sos/`. Data cleaning, processing, and analysis are detailed in the project's GitHub repository at `https://github.com/dlhogan97/Synoptic-Sublimation`.

2.6 Appendix A

2.6.1 Sublimation event occurrence in 2022 and 2023

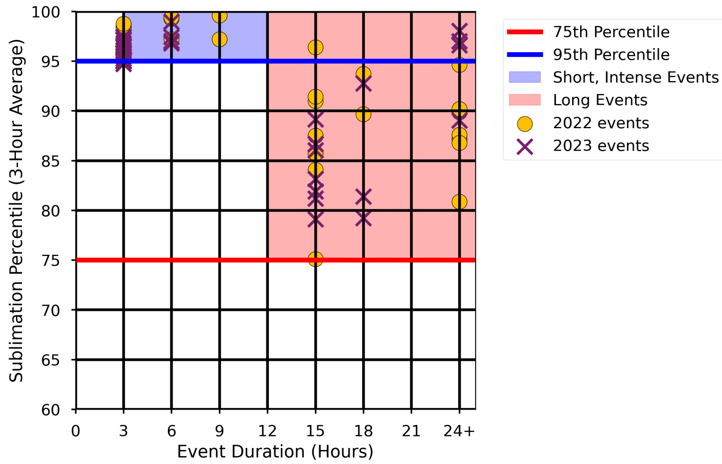

Figure 2.9: Duration and intensity (sublimation percentile) of events during 2022 (yellow circles) and 2023 (purple Xs). The shaded regions show the thresholds used to define short (blue) and long (red) sublimation events.

2.7 Appendix B

2.7.1 Modeling large sublimation events

These models included a single-variable and multi-variable logistic regression, random forest, and XGBoost. We provide a brief description of each model below.

Logistic Regression

Logistic regression is a classification algorithm used to predict the probability of a categorical dependent value. The method can be used to predict the probability of occurrence of a binary event utilizing a logit function. Logistic regression uses the maximum likelihood estimation approach to determine the parameters most likely to reproduce the observed output from the available variables.

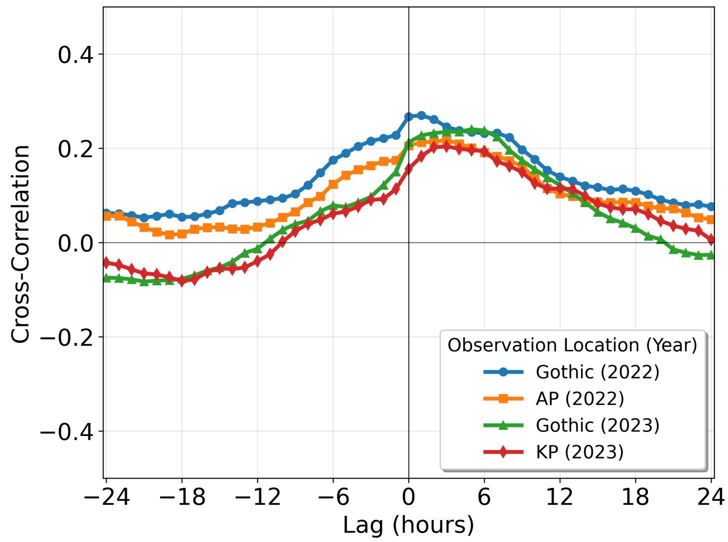

Figure 2.10: Time-lagged cross-correlation between hourly precipitation and wind speed. Positive values at positive time lags indicate higher wind speeds follow precipitation. Wind speed data for 2022 are from Gothic (red) and AP (blue), and for 2023 from Gothic (green) and KP (purple). Precipitation data are from the Gothic pluviometer.

Single-variable The single-variable model used only the 500-hPa wind speed as its predictive variable. This model was used as a simple baseline for comparison.

Multi-variable The multi-variable model used all variables mentioned in the Methods section.

Random Forest

See main text for a brief description.

XGBoost

XGBoost, is a machine learning algorithm that leverages gradient boosted decision trees, which can improve predictive accuracy by training decision trees sequentially, rather than individually [*Chen and Guestrin*, 2016]. The XGBoost model used all variables mentioned in the Methods section.

2.7.2 Evaluation metrics

We evaluated model performance using the F1-score and average precision, both suited for imbalanced classification problems. The F1-score, the harmonic mean of precision and recall, balances false positives and false negatives. Precision is the proportion of predicted events that were correct, while recall is the proportion of actual events that were correctly identified. Average precision summarizes the area under the precision-recall curve, highlighting performance on the minority class across all classification thresholds.

2.7.3 Model comparisons

The Receiver Operating Characteristic (ROC) curve plots recall against the false positive rate across classification thresholds. A high area under the ROC curve (AUC) indicates strong class separation; 1.0 is perfect, and 0.5 reflects random guessing. The Precision-Recall (P-R) curve is better suited for imbalanced datasets, showing the tradeoff between precision and recall. High area under the P-R curve reflects both accurate and comprehensive identification of true events.

Figures 2.11 and 2.12 show that the random forest model outperformed others in both these aspects of performance, so we selected it for the analysis.

2.7.4 Feature importance

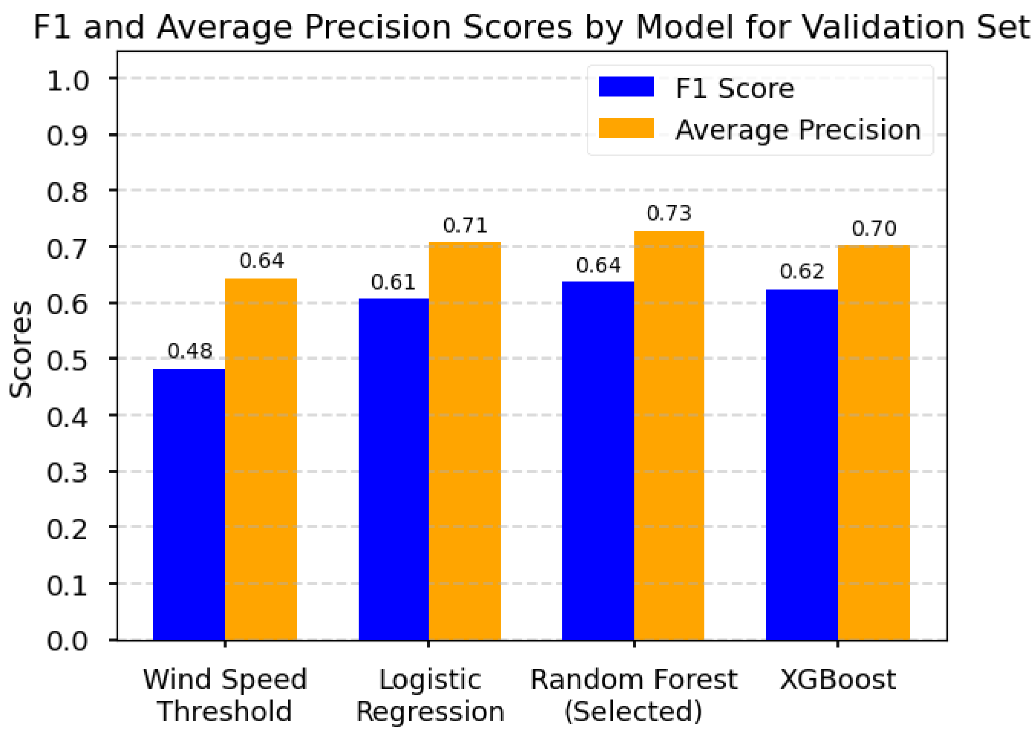

Figure 2.11: Comparison of (A) ROC and (B) Precision-Recall curves for each model. Area under each curve is shown in the legend.

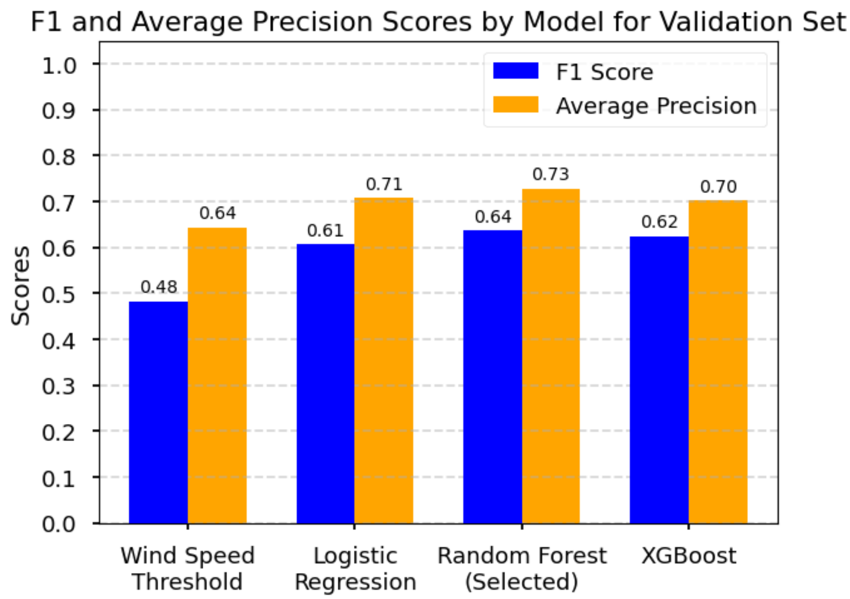

Figure 2.12: Comparison of F1 and average precision scores for each model using the validation dataset.

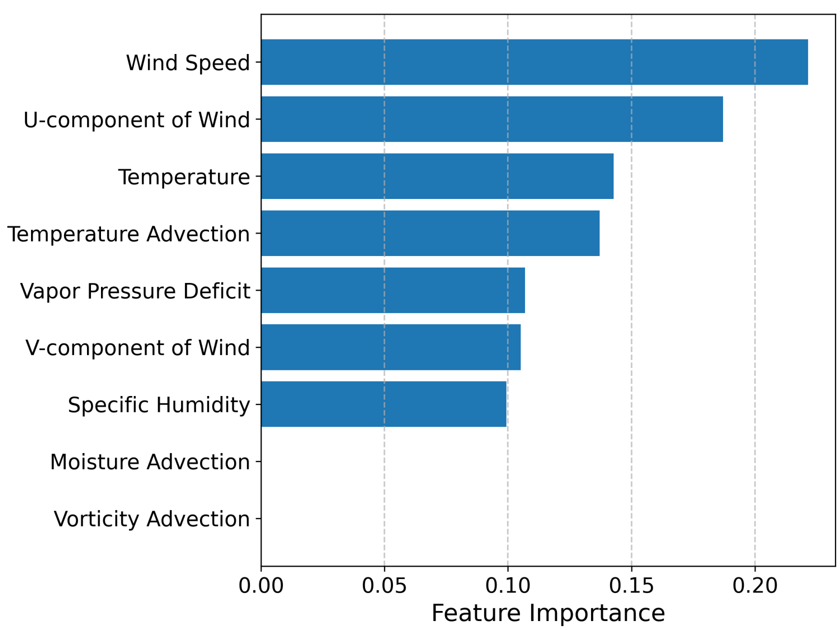

Figure 2.13: Comparison of fractional feature importance for each input variable to the random forest model.

Chapter 3

PRECIPITATION MEASUREMENT UNCERTAINTY IN MOUNTAINS: A GAUGE INTERCOMPARISON FROM THE SAIL, SPLASH, AND SOS CAMPAIGNS

3.1 Introduction

In mountain regions, precipitation measurements underpin weather, climate, flood, water supply and hydropower projections. They represent our most direct and verifiable determination of how much precipitation falls locally and are also vital to evaluate remotely sensed precipitation products and create reliable geostatistical precipitation datasets [*Beck et al.*, 2019; *Bytheway et al.*, 2024; *Daly et al.*, 2008; *Derin and Yilmaz*, 2014; *Gan et al.*, 2024; *Henn et al.*, 2018; *Lundquist et al.*, 2019]. However, accurate and representative precipitation measurements remain notoriously difficult to make in mountains [*Gultepe et al.*, 2014; *Kochendorfer et al.*, 2022; *Lundquist et al.*, 2019]. Sparse gauge distribution, harsh weather conditions, and instrument malfunctions all contribute to significant measurement uncertainty [*Gultepe et al.*, 2014; *Kidd et al.*, 2017; *Mankin et al.*, 2025; *Martinaitis et al.*, 2023; *Rasmussen et al.*, 2012; *Wayand et al.*, 2016]. Consequently, this uncertainty muddles our understanding of the true state of the hydrologic system [*Bárdossy et al.*, 2022; *Boardman et al.*, 2024; *Joseph et al.*, 2018; *Lundquist et al.*, 2015a]. This gap between measured and actual precipitation not only contributes to hydrologic model errors [*Larson and Peck*, 1974; *Moulin et al.*, 2009], but also has real operational consequences: near-normal measured precipitation in the Colorado River Basin was followed by far lower-than-expected streamflow in 2020 and 2021, threatening water and power supplies in the southwestern United States [*Goble and Schumacher*, 2023]. Examples like this raise a fundamental question: how much can we trust our mountain precipitation measurements?

Precipitation gauge intercomparison studies have been the primary tool for quantifying this uncertainty and understanding how gauge choice affects measurement accuracy [e.g., *Goodison et al.*, 1998; *Nystuen et al.*, 1996; *Rasmussen et al.*, 2012; *Boudala et al.*, 2014; *Smith et al.*, 2020]. However, relatively few studies have been conducted in mountains, and those that did were limited to single seasons or isolated storm events [*Smith et al.*, 2020; *Boudala et al.*, 2014; *Thériault et al.*, 2014]. Fewer still have isolated event-specific instrument failures that are difficult or impossible to correct. This leaves a critical gap in understanding how inter-gauge variability evolves across seasons and storm types in mountains.

In mountain regions, where precipitation primarily falls as snow, most measurement problems arise during snowfall [*Frei and Schär*, 1998; *Henn et al.*, 2018; *Sevruk*, 1997; *Kochendorfer et al.*, 2022]. Wind-driven undercatch is among the most well-documented of these problems, and correction methods have been developed and extensively evaluated [*Kochendorfer et al.*, 2017; *Smith et al.*, 2020; *Pierre et al.*, 2019; *Schnepper et al.*, 2023]. However, other environmental conditions during snowfall introduce errors that are harder to systematically correct. Blowing snow can contaminate measurements, variable snowflake characteristics can affect catch efficiency, and near-freezing temperatures increase snow wetness and adhesion, allowing snow to adhere to or accumulate on gauges, clog inlets, or melt and refreeze [e.g., *Fassnacht*, 2004; *Capozzi et al.*, 2021; *Thériault et al.*, 2012; *Rasmussen et al.*, 2012]. Wind shields and heating elements can mitigate some of these issues, but may be impractical to deploy at certain gauge sites [*Rasmussen et al.*, 2012]. Beyond these systematic issues, gauges are vulnerable to random failures with no viable correction. Intense winter storms can damage a gauge or interrupt power supply [*Do et al.*, 2025; *Martinaitis et al.*, 2015]. Biological fouling or mechanical failure can interfere with or miscalibrate a measurement [*Nystuen et al.*, 1996; *Savina et al.*, 2012; *Steiner et al.*, 1999; *Lundquist et al.*, 2015b]. Because mountain gauges are often remote and infrequently maintained, these issues can persist for long periods [*Fleming et al.*, 2023].

To better understand how these measurement challenges vary across instruments, seasons, and storm types, we present a multi-season, event-focused comparison of 11 precipitation gauges and a radar-derived precipitation product collected during three co-located field campaigns in a Colorado mountain valley: the Department of Energy's Surface and Atmospheric Integrated Field Laboratory (SAIL) campaign [*Feldman et al.*, 2023], the National Oceanic and Atmospheric Administrations' Study of Precipitation, the Lower Atmosphere and Surface for Hydrometeorology (SPLASH) [*de Boer et al.*, 2023], and the Sublimation of Snow (SOS) campaign [*Lundquist et al.*, 2024]. Unlike prior studies limited to single seasons or isolated storm events, this dataset spans both warm and cold seasons across multiple gauge types, allowing us to evaluate how inter-gauge variability evolves. From this dataset, we formed two hypotheses:

1. Precipitation measurement uncertainty is greatest during large winter storms that expose gauges to blowing snow, burial, or power failure.

2. Gauges with higher power demands and greater wind exposure exhibit larger and more variable biases than lower-power, sheltered instruments.

We concentrate our analysis on the most biased events to identify the meteorological and instrument-specific drivers of measurement uncertainty and conclude with recommendations for future gauge deployments in mountain environments.

3.2 Observations and Methods

3.2.1 Site Description

We conducted our study in the valley of the upper East River, a headwater basin of the Colorado River, USA (Figure 3.1A and B). Our study used data from SAIL, SPLASH, and SOS alongside reference measurements from a long-term observatory. The long-term site observer, billy barr [sic], has maintained several meteorological instruments and recorded daily weather observations within the upper East River valley since 1975 [*Faybishenko et al.*, 2021].

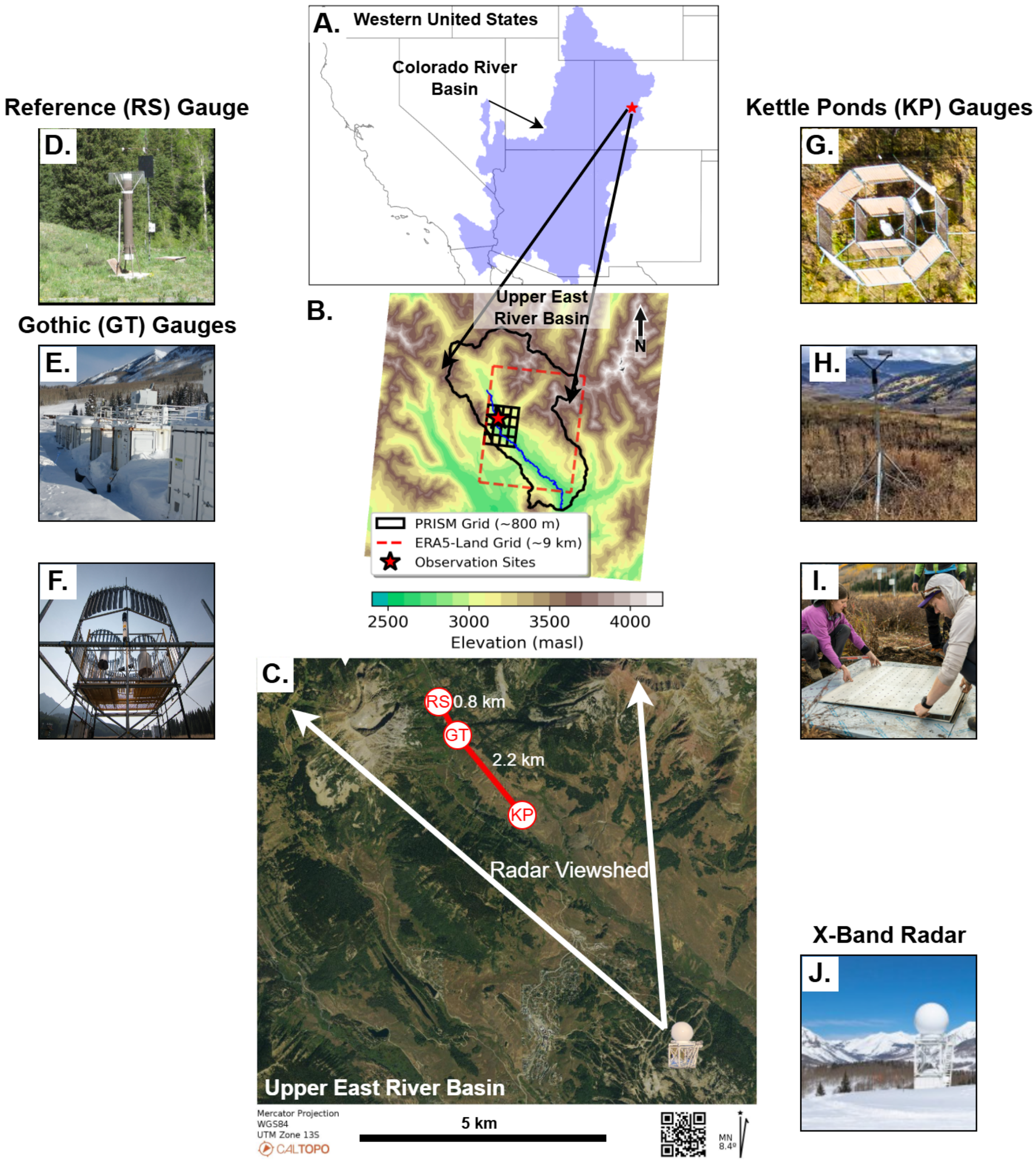

Figure 3.1: (A) The location of the upper East River Basin (red star) within the Colorado River Basin. (B) The upper East River Basin boundary shaded by elevation. The red star represents the approximate field site locations. The dashed red grid cell represents the size of the 9-km ERA5-Land grid cells. The black grid represents the 800-m PRISM grid cells. (C) Locations of the reference site (RS), Gothic Town (GT), and Kettle Ponds (KP) sites. The SAIL X-band radar's location is depicted in the lower right hand corner. Map retrieved from www.caltopo.com on December 12, 2025. Photos of (D) the reference gauge; (E) the location of the laser disdrometer, present weather detector, and optical rain gauge at GT; (F) the tipping bucket and pluviometer in an SDFIR wind shield at GT; (G) the KP pluviometer in a LPDF wind shield; (H) the KP laser disdrometer; and (I) the four fluidless snow pillows.

We used data from three field sites (Figure 3.1C): the reference site (RS) (38.96283° N, -106.99256° W at 2908 m.a.s.l.), the Gothic site (GT) (38.95676° N, -106.98807° W at 2886 m.a.s.l.), and the Kettle Ponds site (KP) (38.94167° N, -106.97222°W at 2858 m.a.s.l.). KP is located 2.2 km down-valley from GT and 3 km from RS. At these sites, the valley is approximately 1 km wide, with mountains rising to either side up to 3800 m.a.s.l. The RS site is located in a clearing surrounded by conifers and aspens on all sides. The area surrounding GT and KP is covered by short grasses and shrubs in warmer months. In winter, low lying vegetation is covered by snow.

Table 3.1: Instrument specifications including gauge type, model, location and field campaign, wind shield type (if applicable), deployment period, sampling interval, reported accuracy, spatial resolution (if applicable), and known measurement issues with relevant references. Asterisks (*) denote instruments requiring external power for heating or data logging.

Gauge / Product	**Model**	**Location(s) (Campaign)**	**Wind Shield**	**Deployment Period**	**Sampling Interval**	**Reported Accuracy**	**Spatial Resolution**	**Known Issues**
In-Situ Precipitation Gauges								
Storage Gauge *[Reference]*	—	RS	Single-Alter	Ongoing	10 min	—	—	Snow capping over orifice; wind-driven undercatch [*Serreze et al.*, 1999; *McGurk*, 1986].

Continued on next page

Table 3.1 continued from previous page

Gauge / Product	**Model**	**Location(s) & Campaign**	**Wind Shield**	**Deployment Period**	**Sampling Interval**	**Reported Accuracy**	**Spatial Resolution**	**Known Issues**
Weighing Bucket*	OTT Pluvio2	GT (SAIL); KP (SPLASH)	SDFIR (GT); LPDF (KP)	Sep 2021–Jun 2023	1 min; 30 min†	1%	—	Wind-driven undercatch [*Rasmussen et al.*, 2012]; wetting losses via evaporation or sublimation [*Groisman and Legates*, 1994; *Savina et al.*, 2012]; snow capping; burial in a deep snowpack; power-dependent heating.

Continued on next page

Table 3.1 continued from previous page

Gauge / Product	**Model**	**Location(s) & Campaign**	**Wind Shield**	**Deployment Period**	**Sampling Interval**	**Reported Accuracy**	**Spatial Resolution**	**Known Issues**
Tipping Bucket*	Novalynx 260-2500E-12	GT (SAIL)	SDFIR	Sep 2021–Jun 2023	1 min	3% liquid	—	Filter clogging and riming during heavy snowfall cause severe undercatch [*Nitu*, 2010; *Kochendorfer et al.*, 2022]; wind-driven undercatch; snowmelt timing errors during mixed-phase events.

Continued on next page

Table 3.1 continued from previous page

Gauge / Product	**Model**	**Location(s) & Campaign**	**Wind Shield**	**Deployment Period**	**Sampling Interval**	**Reported Accuracy**	**Spatial Resolution**	**Known Issues**
Optical Rain Gauge*	Optical Scientific ORG-815-DS	GT (SAIL)	—	Sep 2021–Jun 2023	1 min	5% liquid; 10% solid	—	Underperforms during solid precipitation and low precipitation rates [*Nystuen et al.*, 1996; *Kyrouac and Tuftedal*, 2024]; falsely registers blowing snow as precipitation [*Fassnacht*, 2004; *Capozzi et al.*, 2021]; sensitive to particle size and shape [*Kochendorfer et al.*, 2022; *Smith et al.*, 2020].

Continued on next page

Table 3.1 continued from previous page

Gauge / Product	**Model**	**Location(s) & Campaign**	**Wind Shield**	**Deployment Period**	**Sampling Interval**	**Reported Accuracy**	**Spatial Resolution**	**Known Issues**
Present Weather Detector*	Vaisala PWD22	GT (SAIL)	—	Sep 2021–Jun 2023	1 min	Not applicable	—	Reduced accuracy during frozen precipitation; classification errors during mixed-phase events; susceptible to blowing snow contamination [*Kyrouac and Tuftedal*, 2024].

Continued on next page

Table 3.1 continued from previous page

Gauge / Product	**Model**	**Location(s) & Campaign**	**Wind Shield**	**Deployment Period**	**Sampling Interval**	**Reported Accuracy**	**Spatial Resolution**	**Known Issues**
Laser Disdrometer*	OTT Parsivel2	GT (SAIL); KP (SPLASH)	—	Sep 2021–Jun 2023 (GT); Sep 2023– (KP)	1 min	5% liquid; 20% solid	—	Assumes liquid water density; overestimates snowfall without ρ_s–D correction [*Boudala et al.*, 2014]; blowing snow particles falsely registered as precipitation; wind speed reduces apparent fall velocity, affecting precipitation rate estimates.

Continued on next page

Table 3.1 continued from previous page

Gauge / Product	**Model**	**Location(s) & Campaign**	**Wind Shield**	**Deployment Period**	**Sampling Interval**	**Reported Accuracy**	**Spatial Resolution**	**Known Issues**
Snow Pillow	Alpine Hydromet FSP5 Fluidless	KP (SOS)	—	Oct 2022–Jun 2023	5 min	0.3% (SWE only)	—	Measures SWE only; not designed for event-scale precipitation timing [*Lundquist et al.*, 2015b]; wind redistribution complicates accumulation signal [*Lundquist et al.*, 2024]; valid only when fully snow-covered.
Gridded Precipitation Products								

Continued on next page

Table 3.1 continued from previous page

Gauge / Product	**Model**	**Location(s) & Campaign**	**Wind Shield**	**Deployment Period**	**Sampling Interval**	**Reported Accuracy**	**Spatial Resolution**	**Known Issues**
SQUIRE	CSU X-Band Radar (SAIL)	Gridcell at GT & KP (SAIL)	—	Dec 2021–Mar 2022; Dec 2022–Mar 2023	5 min	—	250 m	Z–S relationship uncertainty varies with snow crystal habit and storm type [*Matrosov et al.*, 2009; *Jackson et al.*, 2026]; winter months only; underperforms at high precipitation rates; ensemble spread reflects Z–S uncertainty.

Continued on next page

Table 3.1 continued from previous page

Gauge / Product	**Model**	**Location(s) & Campaign**	**Wind Shield**	**Deployment Period**	**Sampling Interval**	**Reported Accuracy**	**Spatial Resolution**	**Known Issues**
PRISM	—	Gridcell at GT & KP	—	Oct 2021–Sep 2023	1 day	—	800 m	Reference gauge incorporated in PRISM algorithm, inflating apparent agreement [*Daly et al.*, 2008, 2021]; sparse station network in complex terrain limits performance for small events [*Henn et al.*, 2018].
ERA5-Land	—	Gridcell at GT & KP	—	Oct 2021–Sep 2023	1 hr	—	∼9 km	Coarse resolution cannot resolve valley-scale precipitation variability; parameterized convection struggles with localized warm-season precipitation in complex terrain [*Beck et al.*, 2019].

Continued on next page

Table 3.1 continued from previous page

Gauge / Product	Model	Location(s) & Campaign	Wind Shield	Deployment Period	Sampling Interval	Reported Accuracy	Spatial Resolution	Known Issues
ASO SWE	Airborne Snow Observatories	3×3 km area (BB, GT, KP)	—	1 Apr 2023	Instant	—	50 m	Single flight limits temporal coverage; snow density derived from modeled values rather than direct measurement [*Painter et al.*, 2016]; used here as estimate of accumulated winter precipitation only.

* Instrument requires external power for heating elements or data logging.

† At GT, the "not-real-time" 30-min variable was used for the weighing bucket per manufacturer recommendation, as it undergoes additional proprietary processing to improve accuracy (Bartholomew, 2020).

BB = billy barr site; GT = Gothic Town site; KP = Kettle Ponds site; SDFIR = Small Double Fence Intercomparison Reference; LPDF = Low-Porosity Double Fence; SWE = snow water equivalent; Z–S = radar reflectivity–snowfall rate relationship; ρ_s–D = snow density–diameter relationship.

3.2.2 Reference site

Precipitation at RS was measured by a storage gauge surrounded by a single-Alter wind shield [*Alter*, 1937], which is the same gauge used by the snow telemetry (SNOTEL) network (Figure 3.1D). This gauge uses a pressure transducer to report changes in pressure from accumulated precipitation that is then converted to water depth. At the beginning of each water year (1 October - 30 September), it is refilled with an oil-antifreeze solution to prevent evaporation and freezing [*Serreze et al.*, 1999; *Wen et al.*, 2016; *McGurk*, 1986]. The gauge accumulates precipitation over the entire year and is emptied annually. The site overseer, billy barr [sic], has maintained the gauge since it was installed in 2015 (billy barr [sic], personal communication).

We chose to use this gauge as our reference for four reasons: its location in a protected forest clearing, low power requirements, long period of record, and meticulous maintenance. However, this gauge type is susceptible to snow "capping" over the gauge orifice and wind-driven undercatch [*Serreze et al.*, 1999; *McGurk*, 1986]. To assess the magnitude of these errors, we compared the gauge's measured precipitation to co-located manual snow water equivalent (SWE) and precipitation measurements (described in Appendix A) taken twice daily at 0800 and 1600 MST (local time) over our study period (billy barr, personal communication). The gauge and manual measurements were well correlated (Peason r = 0.94), with the gauge having a low mean bias of -0.5 mm per day compared to these manual measurements (Figure 3.6). Additionally, 73% of the reference gauge's daily totals were within 1 mm of the co-located manual measurements; on days when more than 10 mm of precipitation fell, the median absolute deviation was 2 mm. Overall, the gauge total was 5% less than the manual measurements over the analysis period. We treat this difference as the characteristic uncertainty bounds ($\pm$5%) of the reference gauge for the following comparisons.

3.2.3 Spatial variability between study sites

Because precipitation can vary substantially across short distances in mountains, we evaluated whether site-to-site differences were small enough to justify comparing gauges across sites. To do so, we calculated the coefficient of variation (CV) and inter-site measurement differences for two high-resolution gridded products: the 50-m Airborne Snow Observatories (ASO) SWE product from the 1 April 2023 LiDAR flight over the East River and the 250-m radar-derived Surface Quantitative Precipitation Estimation (SQUIRE) product from SAIL [*O'Brien et al.*, 2024; *Jackson et al.*, 2026]. ASO SWE estimates are derived from airborne LiDAR snow depth measurements and modeled snow density [*Painter et al.*, 2016]. Because precipitation only fell as snow between 1 December 2022 to 31 March 2023, and ASO flew on 1 April 2023, we used SWE as an estimate of total winter precipitation between these dates.

SQUIRE produces a gridded quantitative precipitation estimation (QPE) field using data from the SAIL X-band radar located down-valley on Mt. Crested Butte (Figure 3.1C) [*Jackson et al.*, 2026; *Grover et al.*, 2023]. It calculates liquid equivalent QPEs using an ensemble of radar-snowfall rate relationships ($Z - S$) described in further detail in the Appendix A. For SQUIRE, we calculated total precipitation between the same dates as for ASO using the WSR88D-IW ensemble member. Both ASO and SQUIRE are also evaluated in the comparative analysis below. These products each showed low spatial variability in the valley bottom (CV = 0.07 and 0.09 for SQUIRE and ASO, respectively; Figure 3.2), consistent with prior observations in similar terrain [*Silverman et al.*, 2013]. Although magnitudes differed significantly between products, each showed small relative differences between sites: each site's ASO SWE estimates overlapped within their uncertainty bounds (10%), and the SQUIRE differences were within 8% across sites, comparable to the reference gauge uncertainty of 5% (Figure 3.2). We therefore attribute gauge measurements differing from the reference gauge by more than 10% to instrument performance rather than spatial variability.

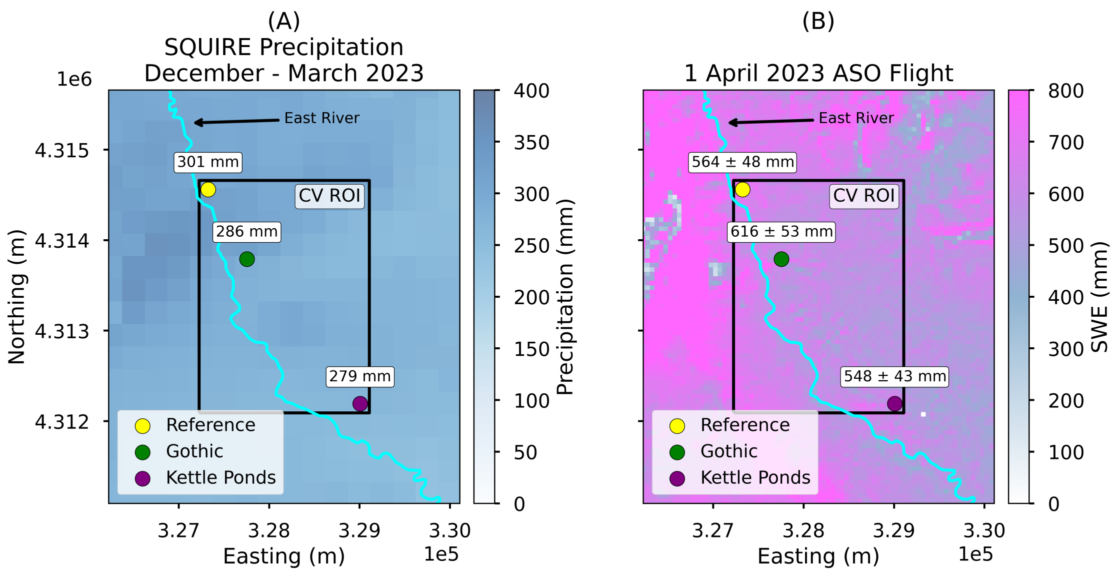

Figure 3.2: Comparison of (A) accumulated precipitation between sites from SQUIRE for December - March 2023 from the WS-D88IW ensemble member and (B) ASO SWE from the 1 April 2023 flight over the East River basin. The black box signifies the region of interest (ROI) used for spatial coefficient of variation (CV) calculation. For each subfigure, the measured precipitation and SWE at each field site's grid cell is shown. The standard deviation of the measurements in the 100-m surrounding each site's ASO grid cell is also included in (B). The East River is drawn (cyan line) for spatial context.

3.2.4 Gothic Town site (GT)

At GT, SAIL deployed five precipitation gauges that were maintained by on-site field technicians: an optical rain gauge; a present weather detector; a laser disdrometer; a heated weighing bucket pluviometer; and a heated tipping bucket [*Feldman et al.*, 2023]. The optical rain gauge measures the scintillation intensity from hydrometeors falling past an infrared beam and converts the signal to precipitation rate and type [*Kyrouac and Tuftedal*, 2024]. The present weather detector measures the scatter produced by hydrometeors falling past an infrared beam, then uses the scatter signal characteristics, magnitude, and ambient temperature to classify precipitation type and intensity [*Kyrouac and Tuftedal*, 2024]. The laser disdrometer measures the attenuation of a horizontal laser beam by falling hydrometeors to determine their size and fall velocity, from which precipitation rate and drop size distribution are derived assuming liquid water density [*Zhu and Wang*, 2025]. The location of these three optical sensors is shown in Figure 3.1E. The heated weighing bucket weighs precipitation using an airtight load-cell. It is filled with mineral oil to prevent evaporation or freezing. The heated tipping bucket funnels precipitation to a small cup that tips when filled. Each tip carries 0.254 mm of water [*Kyrouac and Tuftedal*, 2024]. A small Double Fence Intercomparison Reference (SDFIR) wind shield surrounded these two gauges, which were co-located (Figure 3.1F).

3.2.5 Kettle Ponds site (KP)

At KP, SPLASH deployed a weighing bucket and laser disdrometer (Figure 3.1G and H). The weighing bucket was surrounded by a Low-Porosity Double Fence wind shield (LPDF) [*Kochendorfer et al.*, 2023]. SOS deployed four fluidless snow pillows (Figure 3.1I), which measured accumulated snow water equivalent (SWE) [Figure 1I; *Lundquist et al.*, 2024]. Each snow pillow uses four load-cells to measure the weight of the overlying snowpack from which SWE is derived. To estimate cumulative precipitation from the snow pillows, we calculated the maximum difference between two consecutive SWE measurements, and only included consecutive measurements that increased. We only analyzed these data between December and March. Only snow (no rain) fell during this period.

To test whether blowing snow contaminated precipitation measurements, we used data from an acoustic blowing snow sensor deployed at KP during winter 2023 [*Lundquist et al.*, 2024].

3.2.6 Gridded Products

In addition to gauge measurements, we evaluated three gridded precipitation products: the 250-m SQUIRE 30-minute data from SAIL (winter only) that was described above, the Parameter-elevation Regressions on Independent Slopes Model (PRISM) 800-m daily data [*Daly et al.*, 2008], and the European Centre for Medium-Range Weather Forecasting (ECMWF) Reanalysis v5 land surface product (ERA5-Land) 0.1-arc minute (9 km) hourly data [*Muñoz-Sabater et al.*, 2021]. We evaluated SQUIRE because radar-derived QPE products like this are rarely available in complex terrain, and, to date, SQUIRE has only been evaluated against a single gauge type (Jackson et al., 2026). We included PRISM and ERA5-Land because they are widely used and publicly available datasets that provide independent precipitation estimates.

For PRISM and SQUIRE, we used the grid cell containing each study site. For ERA5-Land, we used a single grid cell because all sites fall within the same grid cell (Figure 3.1B). SQUIRE was only available for the 2022 and 2023 winter seasons (1 December and 31 March).

3.2.7 Data filtering and standardization

Before comparing measurements, we applied three filters to each gauge. First, we treated values with quality control or instrument status flags as missing. Second, we masked values above a maximum probable precipitation for each instrument's sampling interval (Table 3.1). We derived these thresholds from the 100-year recurrence interval at the nearby Butte SNOTEL gauge using the NOAA Atlas 14 Precipitation Frequency Data Server [*NOAA*, 2017]. Third, we set values below each instrument-specific minimum threshold to 0, as these likely reflected sensor noise (Table 3.1). We did not interpolate across or gap-fill periods of missing data, as doing so would obscure the instrument failures central to our analysis. These missing values were assigned to 0, as well. Figure 3.7 summarized the fraction of valid observations by month over the study period.

We resampled the gauge measurements and SQUIRE to 30-minute sums and converted the timeseries to local time. To match the daily PRISM data, we resampled the reference gauge and ERA5-Land data to daily sums and defined our "day" as 1200-1200 UTC [*Daly et al.*, 2021].

Laser disdrometer winter precipitation rate adjustments

Because the laser disdrometer assumes all particles have the density of liquid water, it often overestimates the precipitation rate during snowfall [*Battaglia et al.*, 2010; *Boudala et al.*, 2014; *Brandes et al.*, 2007]. To account for this overestimation, snowflake density (ρ_s) must be estimated. Prior work has demonstrated that ρ_s relates to snow particle size (D), assuming a uniform snow particle shape [*Ishizaka*, 1993], and several $\rho_s - D$ relationships have been developed to adjust laser disdrometer snowfall rates. These density estimates are limited by their simple assumptions of snow crystal habit, the random nature of snowflake aggregation and riming, and varying definitions of D [*Boudala et al.*, 2014]. To address some of these uncertainties, we used three different $\rho_s - D$ relationships following *Heymsfield et al.* [2004]; *Brandes et al.* [2007]; and *Holroyd and Jiusto* [1971]. We describe these relationships in further detail in Appendix A.

3.2.8 Comparison of gauge total precipitation

Statistical comparison

For clarity, we refer to any gauge other than the reference gauge as a test gauge. To evaluate test gauge differences relative to the reference, we used the normalized bias, which quantifies the general direction of the bias with respect to the reference and is insensitive to the magnitude of each individual observation. We calculated it for the entire study period and each year's December-March period and the 2022 June - September period. These seasonal comparisons allowed us to assess how gauge performance varied between snow-dominated and rain-dominated periods.

We also calculated the root mean squared error, standard deviation, and correlation coefficient of daily precipitation between each gauge and the reference. For brevity, we summarized these statistics in Figure 3.8.

Event-specific analysis

We defined discrete precipitation events using a minimum 30-minute accumulation threshold of 0.05 mm. An event began when either the test gauge or reference gauge measured precipitation above this threshold. An event ended when precipitation continuously measured below this threshold for 6-hours or more at both gauges. Each gauge's event total was calculated as the sum over the event time window.

For the gridded products, a precipitation event began when daily precipitation exceeded 0.25 mm for either the gridded product or the reference and ended on the first day precipitation was below this threshold. Event differences were calculated in the same way as for the test gauges.

3.3 Results

3.3.1 Study period precipitation differences across gauges

Gauge measurement totals varied dramatically over the study period, ranging from 882 mm to 2389 mm, with the reference sitting between these extremes at 1562 mm (Figure 3.3). With a coefficient of variation across gauges of 0.51 at GT, instrument-to-instrument variability far exceeded site-to-site variability, though missing data ranged from 0% to 5.3% of observations for individual gauges and likely contributed to this spread (Figure 3.7).

The weighing buckets at both sites measured closest to the reference (Figure 3.3). At GT, most other gauges exceeded the reference, with the unadjusted laser disdrometer and optical rain gauge showing the largest positive differences and the tipping bucket showing the most negative difference (Figure 3.3A). At KP, the unadjusted laser disdrometer again showed a large positive difference while the disdrometer with the *Heymsfield et al.* [2004] adjustment showed a large negative difference (Figure 3.3B). The other adjusted laser disdrometer totals were closer to the reference than at GT. At both sites, the SQUIRE ensemble members showed large negative differences, with the WSR88D-IW ensemble member deviating least, consistent with *Jackson et al.* [2026].

Between December and March 2023, the total KP snow pillow measurements ranged from 64% to 77% of the reference gauge total (not shown). The range of measurements between pillows was driven by local snow redistribution [*Lundquist et al.*, 2024]. Both gridded products performed well. At GT, ERA5-Land and PRISM exceeded the reference by 10% and 6%, respectively, though the latter was expected given that PRISM incorporates the reference in its algorithm. PRISM showed a slight negative bias at KP compared to the reference.

Pearson correlations between concurrent test and reference gauge measurements also varied widely (0.25-0.94; Figure 3.8), reflecting differences in each gauge's ability to resolve precipitation timing.

3.3.2 Seasonal bias comparison

Precipitation differences and inter-gauge variability were greatest during the snow-dominated season (Figure 3.3). At most gauges, the mean normalized bias during the December-March periods (winter) was more than that of the June-September period (summer), with the exceptions of ERA5-Land and the adjusted laser disdrometers (Figure 3.4). However, their direction (positive or negative) rarely changed between seasons, but there were exceptions. The normalized bias for the optical rain gauge and both unadjusted laser disdrometers changed from negative in summer 2022 to positive in both winters (Figure 3.4), indicating that these instruments tend to overestimate precipitation during snowfall.

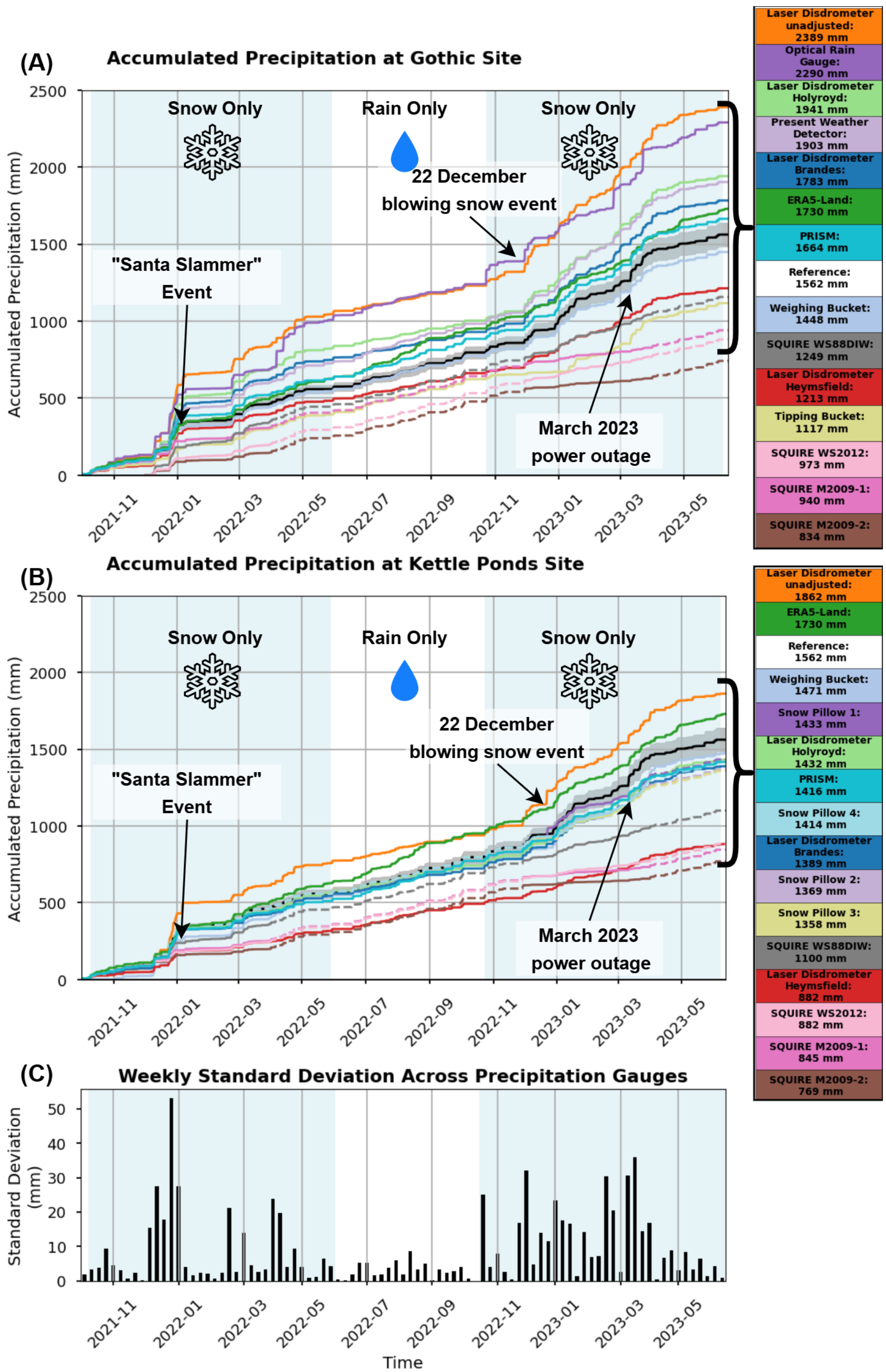
(A)
Accumulated Precipitation at Gothic Site
Snow Only
Rain Only
Snow Only
22 December
blowing snow event
"Santa Slammer"
Event
March 2023
power outage
Accumulated Precipitation (mm)
Laser Disdrometer unadjusted: 2389 mm
Optical Rain Gauge: 2290 mm
Laser Disdrometer Holyroyd: 1941 mm
Present Weather Detector: 1903 mm
Laser Disdrometer Brandes: 1783 mm
ERA5-Land: 1730 mm
PRISM: 1664 mm
Reference: 1562 mm
Weighing Bucket: 1448 mm
SQUIRE WS88DIW: 1249 mm
Laser Disdrometer Heymsfield: 1213 mm
Tipping Bucket: 1117 mm
SQUIRE WS2012: 973 mm
SQUIRE M2009-1: 940 mm
SQUIRE M2009-2: 834 mm
(B)
Accumulated Precipitation at Kettle Ponds Site
Snow Only
Rain Only
Snow Only
22 December
blowing snow event
"Santa Slammer"
Event
March 2023
power outage
Accumulated Precipitation (mm)
Laser Disdrometer unadjusted: 1862 mm
ERA5-Land: 1730 mm
Reference: 1562 mm
Weighing Bucket: 1471 mm
Snow Pillow 1: 1433 mm
Laser Disdrometer Holyroyd: 1432 mm
PRISM: 1416 mm
Snow Pillow 4: 1414 mm
Laser Disdrometer Brandes: 1389 mm
Snow Pillow 2: 1369 mm
Snow Pillow 3: 1358 mm
SQUIRE WS88DIW: 1100 mm
Laser Disdrometer Heymsfield: 882 mm
SQUIRE WS2012: 882 mm
SQUIRE M2009-1: 845 mm
SQUIRE M2009-2: 769 mm
(C)
Weekly Standard Deviation Across Precipitation Gauges
Standard Deviation (mm)
Time
2021-11
2022-01
2022-03
2022-05
2022-07
2022-09
2022-11
2023-01
2023-03
2023-05

Figure 3.3

Figure 3.3: Accumulated precipitation at (A) Gothic and (B) Kettle Ponds. The grey shaded region indicates the accumulated reference gauge uncertainty (±5%). Arrows point out the December 2021 "Santa Slammer" precipitation event, December 2022 blowing snow event, and March 2023 power outage. (C) 7-day standard deviation across precipitation gauges. In all plots, the light blue shaded areas signify periods when precipitation fell exclusively as snow. The tables on the right of (A) and (B) represent the total accumulated precipitation from each gauge. The color of each cell is the color of the line in each plot, except for the reference, which has a white cell, but a black line. Note that the snow pillow and SQUIRE data were only available during winter. Outside these seasons, we substituted their accumulations with the reference gauge measurements (dashed lines) for visualization purposes.

The laser disdrometer biases were greater at GT than KP, where winds were consistently higher. At higher wind speeds, particles are blown across the laser disdrometer beam at shallower angles, reducing their apparent fall velocity and thus the calculated precipitation rate. We saw direct evidence of this effect in the laser disdrometer data, where particle fall velocity decreased with increasing wind speed, regardless of particle size (Figure 3.9), suggesting that higher winds at KP partially offset the laser disdrometer's overestimation.

ERA5-Land had small winter biases but a large positive bias in summer 2022. This summertime difference likely reflects ERA5-Land's coarse (9 km) resolution and parameterization of convection, neither of which can resolve point-scale variability in a valley bottom. However, a single valley-bottom gauge can not capture the spatial variability of localized summer convection, which can vary significantly [*Suri and Azad*, 2024].

3.3.3 Causes of precipitation measurement uncertainty

We observed individual storms affected gauges differently, such that the same event could produce large biases in some gauges while minimally impacting others (Figure 3.3). In this section, we describe three of these events during which (1) blowing snow, (2) power failure, and (3) gauge burial significantly impacted the accuracy of certain gauges. We also highlight the stark differences that occurred during the largest precipitation event of the study period.

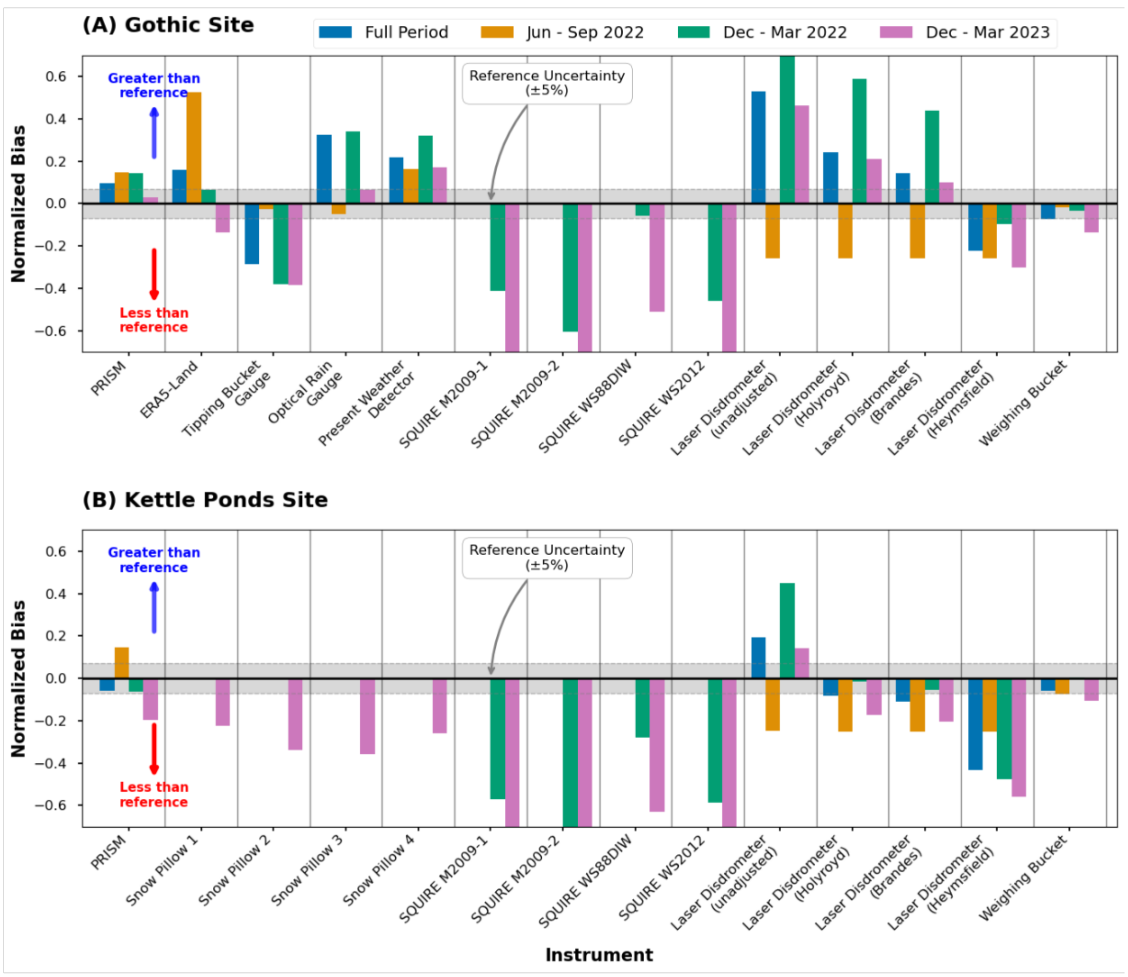

Figure 3.4: Normalized bias for each instrument and product at (A) Gothic and (B) Kettle Ponds.The colors indicate the period over which the normalized bias was calculated. The grey shaded region indicates the reference gauge uncertainty (±5%).

Large blowing snow event: 20 December 2022 - 22 December 2022

This event coincided with the highest sustained winds of the study period (see event pointed out in Figure 3.3). It began with low precipitation rates measured by all gauges at KP and GT and low wind speeds (Figure 3.11). Wind speed progressively increased and remained above 5 m s−1 after 18:00 (local time) on 21 December. At midnight on 22 December, laser disdrometer precipitation rates increased dramatically for over 12 hours, while the weighing bucket and reference gauge accumulation remained low. During this time, mean wind speed was 10.1 m s−1, and blowing snow rates increased dramatically at KP (Figure SI1). Blowing snow particles passing through the beam caused the laser disdrometers to falsely register precipitation, while the reference and weighing buckets showed no such response. By the end of the event, the unadjusted laser disdrometers measured 85 mm and 41 mm more than the reference at KP and GT, respectively. Although wind-driven undercatch may have caused the reference and weighing buckets to undermeasure precipitation, their close agreement with one another (4.6-6.6 mm) suggests their totals were reliable. Although blowing snow only occurred 10% of the time in winter 2023, 29-34% of the laser disdrometers' bias; 29% of the optical rain gauge's bias; and 23% of the present weather detector's bias were concentrated during those periods. Because blowing snow measurements were unavailable during winter 2022, these figures likely underestimate the true contribution from blowing snow to optical sensor bias over the full study period.

Power Failure: 10-11 March 2023

This event began early on 10 March 2023 as an atmospheric river advected warm air into the region under high winds (see event pointed out in Figure 3.3). Precipitation rates and values between the GT weighing bucket and reference gauge were nearly identical until power was cut off to the GT site between 23:00 and 15:30 on 11 March (Figure 3.12). This outage lasted for approximately 16 hours, causing the most significant precipitation difference between these two gauges (26.9 mm) for a single event over the entire campaign. Once power returned, the two gauges again matched closely until the end of the event. Overall, power outages were rare during precipitation, occurring just 1.5% of the time at GT; yet, they accounted for 31% of the weighing gauge's total bias, underscoring that even infrequent outages can dominate accumulated measurement error when they occur during large events. Rather than treating precipitation during outages as zero, these missing measurements could be interpolated using a neighboring gauge with an independent power supply, reinforcing the value of redundant instrumentation.

Gauge burial: 30 March 2023 - 01 April 2023

During this event, the reference gauge measured 18 mm of precipitation, while the KP weighing bucket measured 60 mm. This difference was caused by two spikes in precipitation rate (¿ 30 mm hr^{-1}) measured at the weighing bucket (Figure 3.5A). To understand why this occurred, we note that a large snow drift had built up around the gauge by early March 2023 (Figure 3.5C). This snow drift caused the LPDF wind shield to provide progressively less wind sheltering, making the gauge more susceptible to snow blowing into it. We observed two other events in March and April 2023 that showed a similar signal, albeit to a lesser degree. In total, the KP weighing bucket measured 76 mm more precipitation than the reference during these three events. However, the gauge was not buried in 2022 because peak snow depth was less in 2022 (172 cm) than in 2023 (243 cm), and we did not observe any of these types of events that year.

'Santa Slammer' Precipitation Event: 23 December 2021 - December 31 2021

The 'Santa Slammer', a series of large storms in December 2021, was both the most hydrologically impactful and most uncertain event of the study period. It contributed 21% (164 mm) of the 2022 water year total precipitation (772 mm) but also produced the largest spread in measured precipitation across all gauges (43–311 mm) (Figure 3.3). Much of this spread occurred during the passage of four low-pressure systems, which coincided with periods of intense precipitation that exposed certain gauges to their specific measurement vulnerabili-

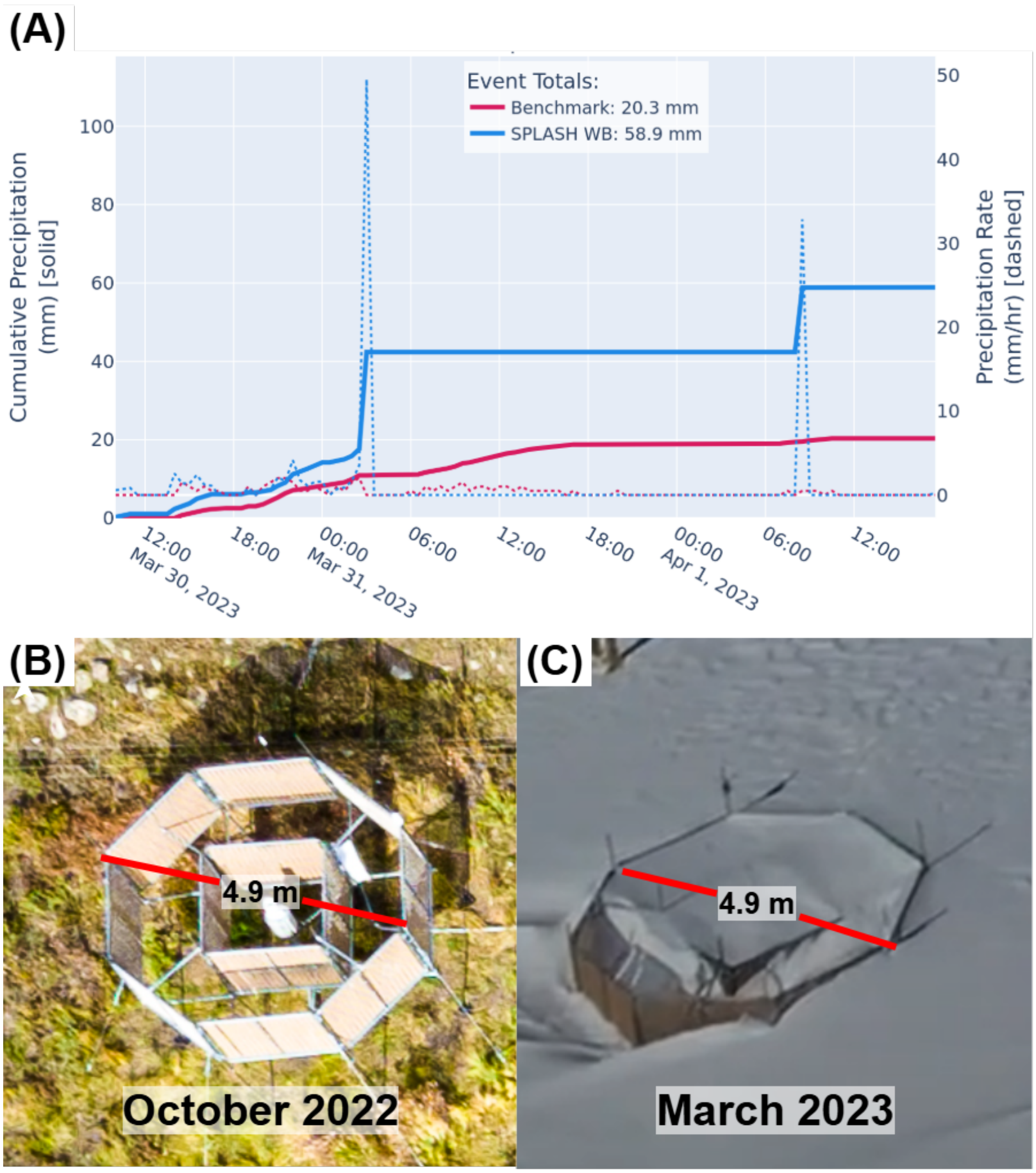

Figure 3.5: Time series for the most biased event for the KP weighing bucket. (A) Cumulative event precipitation (solid line, left axis) and precipitation rate (dashed line, right axis) for the weighing bucket (blue) and reference gauge (red); (B) wind speed (solid line, left axis) and direction (grey dots, right axis); (D) LPDF weighing bucket prior to snowfall in October 2022 (photo courtesy of Mark Stone). (D) LPDF weighing bucket buried by snow, March 2023 (photos courtesy of Jeremy Snyder).

ties. For example, the tipping bucket measured significantly less than the weighing bucket and reference gauges, which SAIL technicians attributed to snow accumulation on the filter screen that had blocked meltwater from reaching the tipping mechanism. Meanwhile, the reference gauge periodically recorded unrealistically high precipitation rates consistent with snow capping, while the optical sensor measurements were likely inflated by blowing snow (Figure 3.13).

Event-scale contributions to total precipitation biases

To assess how bias was distributed across events, we ranked all precipitation events by total absolute bias for each gauge. Although 139 unique precipitation events occurred, the 10 most biased events accounted for 38–61% of total bias and 20–60% of total precipitation across all gauges, indicating that a small number of storms dominated both bias and water year accumulation (Figure 3.10). These most biased events were concentrated in winter, primarily in December and March (Figure 3.10C), when precipitation fell exclusively as snow and monthly totals were highest. Consistent with this result, measurement differences between gauges were greatest during the largest storms (Figure 3.14). Event bias scaled linearly with event size, although the strength of this relationship varied by gauge type.

3.4 Discussion

3.4.1 Causes of precipitation measurement uncertainty during SAIL, SPLASH, and SOS

The significant inter-gauge variability we observed during SAIL, SPLASH, and SOS (Figure 3) was consistent with prior gauge intercomparison studies [e.g. *Rasmussen et al.*, 2012; *Smith et al.*, 2020; *Goodison et al.*, 1998] and far exceeded spatial variability between sites (Figure 3.2). Critically, the majority of these differences were concentrated during a few large winter storms (Figure 3.10), which exposed gauges to their specific measurement vulnerabilities. Because gauges did not respond uniformly to these events, many event-specific biases were driven by instrument-specific or environmental factors that cannot be systematically corrected for. We profiled three such factors (blowing snow, power failure, and gauge burial) unique to three different gauges that had severe impacts on their measurement accuracy.

Blowing snow

Optical sensors overestimated snowfall during blowing snow events because they cannot distinguish falling snow from wind-transported particles. In snow-dominated mountain environments, blowing snow is common, especially during and surrounding storms [*Hogan et al.*, 2026b; *Schwat et al.*, 2025; *Wagner et al.*, 2022; *Sugiura et al.*, 2003], making this a persistent source of uncertainty. SNOTEL stations site selection, in forest clearings [*Nolin et al.*, 2021], seeks to manage this issue for snowpack monitoring. Still, the relationship between wind speed and bias was not straightforward: the laser disdrometer bias was actually lower at the KP site than the GT site. We attribute this to the negative relationship between particle fall velocity and wind speed, which reduced calculated precipitation rates as winds increased (Figure B2). Blowing snow sensors alongside wind measurements can help flag affected periods for post-hoc correction, and installing optical sensors above the blowing snow layer, which rarely exceeds 10 m, could reduce this bias.

Power failure and gauge burial

While the weighing buckets performed comparably to the reference gauge, they remained vulnerable to unpredictable, site-specific failures. At GT, an 11-hour power outage during a major storm caused a large measurement gap (Figure S2 in the Supporting Information). Although outages were rare, they accounted for a disproportionate share of the GT weighing bucket's total bias, underscoring that infrequent failures can dominate accumulated error. In mountains, these outage events are most likely to coincide with large winter storms [*Do et al.*, 2025], precisely when accurate measurements matter most. Furthermore, the GT site, as well as the SAIL instruments, were specifically hardened for outages, including Uninterruptible Power Supplies (UPSs), technician support, and limited backup power. Yet outages still occurred and must be planned for.

At KP, the LPDF wind shield became buried in March 2023, allowing snow to collapse or blow directly into the gauge. While these wind shields are effective at reducing wind-driven undercatch, they can inadvertently act as snow fences, causing snow to accumulate around them. This effect has also been observed in other snowy regions (Matthew Sturm, personal communication). Due to the large spatial footprint of WMO-standard wind shields, finding a suitable, unobstructed location for snow measurement can be difficult, and burial risk should be considered when selecting gauge placement in deep snowpack environments.

3.4.2 Gridded products

PRISM and ERA5-Land agreed reasonably well with the reference over the full study period and during the largest events (Figure 3.3), but performed poorly for smaller events (Figure 3.8) and exhibited seasonal variability (Figure 3.4). Yet, PRISM's agreement is partly expected given the reference gauge is included in its algorithm. ERA5-Land's large summer bias (Figure 3.4) likely reflects its coarse (9 km) resolution and parameterized convection; given the high spatial variability in convective precipitation, it should not be expected to resolve point-scale variability in a valley bottom. The SQUIRE ensemble captured smaller precipitation events well but struggled during high precipitation rates, with variable performance across ensemble members. Consistent with *Jackson et al.* [2026], the WSR88D-IW member performed best. $Z - S$ relationship uncertainty, which varies with snow crystal habit and storm type [*Matrosov et al.*, 2009], likely drove these discrepancies, and can likely be mitigated with empirical choices of $Z - S$ relationships for large storms [*Jackson et al.*, 2026].

3.4.3 Why large events matter

Instrument failures are consequential because measurement uncertainty is highest during large events when accuracy matters most. Measurements varied widely between gauges during the largest precipitation events (Figure 11), yet discrete, large events account for a substantial fraction of annual precipitation in snow-dominated mountain regions [*Lundquist* *et al.*, 2015a; *Guan et al.*, 2010]. The East River Basin is no exception: approximately 30% of water year precipitation falls over an average of just 10 days (Figure 3.15). Uncertainty during these events therefore has an outsized impact on the accuracy of accumulated precipitation totals. We also note that winter precipitation during our study period fell exclusively as snow; mixed-phase events, which are more common in maritime mountain environments, would introduce additional challenges not captured here [*Boudala et al.*, 2014; *Currier et al.*, 2017; *Wayand et al.*, 2016].

3.4.4 Gauge recommendations

All gauges suffer from systematic biases that make measuring ”true” precipitation exceptionally challenging, and mountain environments introduce additional hurdles [*Sevruk*, 1997; *Kidd et al.*, 2012]. In an era of data-driven approaches that depend on high-quality observations, “incorrect observations are worse than no observations” [*Lundquist et al.*, 2019]. We note that bias corrections exist for many gauges evaluated here [e.g. *Kochendorfer et al.*, 2017; *Rasmussen et al.*, 2012], but our analysis intentionally excluded them to provide a baseline comparison of raw performance. Each gauge type has its own strengths and limitations. Disdrometers, for instance, are valuable for measuring precipitation rate and particle size distribution for radar or satellite calibration rather than as standalone precipitation records. $\rho_s - D$ adjustments can reduce their uncertainty during snowfall, but overestimation during blowing snow remains unavoidable without independent flagging. Of the $\rho_s - D$ we evaluated, the *Holroyd and Jiusto* [1971] and *Brandes et al.* [2007] adjustments performed best, consistent with *Boudala et al.* [2014]. Snow pillows are well suited for point SWE measurements but are not designed for event-scale precipitation, especially in exposed locations where wind redistribution complicates the signal [*Lundquist et al.*, 2024]. Tipping buckets are not recommended for snowfall measurement given well-documented issues with filter clogging, riming, and wind scouring [*Nitu*, 2010; *Kochendorfer et al.*, 2022].

We attribute the reliability of the reference storage gauge to its sheltered siting, sufficient height, low power demand, and self-contained power supply. This gauge has the added advantages of low cost, a small spatial footprint, and adaptability across snow climates [*Fleming et al.*, 2023; *Serreze et al.*, 1999]. Its close agreement with WMO-standard gauges demonstrated that careful site selection can compensate for a suboptimal wind shield. For future deployments, we recommend co-locating a sheltered storage-type gauge with an independent power source alongside any other sensor of interest. If a sheltered site is infeasible, we recommend deploying two weighing gauges surrounded by a double fence wind shield (e.g. DFIR, SDFIR, LPDF) at a site less susceptible to building snow drifts, which tend to form in the same locations year over year [*Sturm and Wagner*, 2010]. Such sites could be identified prior to deployment using high-resolution satellite imagery acquired during the melt season when these drift patterns are clearly visible.

Gauge intercomparison studies like ours represent a best-case scenario, with redundant and rigorously maintained instruments. Yet, we still observed several instrument failures, underscoring that no mountain precipitation record should be treated as ground truth without scrutiny. Field campaigns in complex terrain inevitably face greater challenges, and failures during high-impact events may go unreported. We commend SAIL, SPLASH, and SOS for their transparency in documenting failures and providing open data access, which allowed us to identify problems that would otherwise have been difficult to diagnose.

3.5 Conclusion

We compiled two years of precipitation measurements from 11 gauges, a radar-derived product, and two independent gridded products in a Colorado mountain valley and evaluated their performance against a reference storage gauge. We found gauge type exerted a stronger control on bias than differences in site location. The largest inter-gauge differences occurred during winter snow storms, particularly in December and March, when event precipitation totals were highest. A small number of these storms accounted for a disproportionate fraction of both total bias and water year accumulation, underscoring that uncertainty was episodic rather than continuous.

Precipitation measurement uncertainty was dominated by snow-related processes that created challenges for different gauge types. Optical gauges were highly sensitive to blowing snow and snow density causing overestimation of precipitation; weighing bucket gauges, which performed best overall, faced issues with power disruptions and gauge burial, causing undermeasurement; snow pillows struggled at the event scale due to snow redistribution; and tipping bucket gauges severely underrepresented snowfall due to filter blockage during large precipitation events.

Some of these measurement challenges also influenced how well the gridded products represented precipitation. The SQUIRE product performed well during small events, but struggled to capture high precipitation rates. Further development of this product using event specific $Z-S$ relationships would offer a great advancement in distributed precipitation measurements in mountains. PRISM and ERA5-Land closely matched the reference total but struggled to resolve smaller events and exhibited some seasonal variability in performance. These products are therefore best suited for seasonal to annual analyses in complex terrain.

To most accurately measure precipitation in snow-dominated mountain areas, gauges must be optimized to capture snowfall. We found the reference gauge, similar to those used by the SNOTEL network, remains reliable when installed at sufficient height and in sheltered locations to minimize drifting and undercatch. Careful site selection is critical, and meteorological observations (e.g., wind speed, temperature, blowing snow) alongside an awareness of larger scale meteorological conditions, such as frontal passages, can help identify periods when uncertainty increases.

By identifying the meteorological conditions and event types that produce the largest biases, we can diagnose when and why measurement accuracy degrades. This event-based perspective helps identify which gauges are most resilient during high-impact events and when precipitation records are most vulnerable to instrument-specific failure. Such information provides a framework for prioritizing use of certain gauge types, guiding site selection, and flagging periods when measurements warrant greater scrutiny.

Distributed ground-based precipitation networks in mountains remain foundational for detecting hydroclimatic change, evaluating remotely sensed products, and forcing hydrologic models. But the key take-away here is that reliable measurements ultimately require more than just a single gauge. Several decisions can mitigate issues: deliberate siting to reduce instrument-specific biases; redundant gauge deployments to guard against failures during critical precipitation events; and co-located meteorological observations to flag periods of elevated uncertainty. Yet perhaps most importantly, a willingness to document and share when measurements go wrong, as we have sought to do here, can allow the research community to acknowledge the strengths and limitations of certain technologies across locations and storm types. Doing so will ultimately help the community rein in precipitation measurement uncertainty and improve the scientific return from future deployments.

Data Availability Statement

SAIL data is publicly available from the SAIL data repository via `https://armgov.svcs.arm.gov/research/campaigns/amf2021sail` and was accessed using the ARM Atmospheric Community Toolkit (`https://arm-doe.github.io/ACT/`). SPLASH weighing bucket and meteorological data were downloaded from *Meyers* [2024]. SPLASH laser disdrometer data was accessed from *Ballard et al.* [2023]. Reference gauge data was accessed from the Desert Research Institute repository for the billy barr [sic] station via `https://wrcc.dri.edu/cgi-bin/rawMAIN.pl?corbil`. Manual snow measurements were accessed through personal communication with billy barr [sic]. SOS data was accessed from the SOS data repository via `https://www.eol.ucar.edu/field_projects/sos`. PRISM data was downloaded from the PRISM Group, Oregon State University, `https://prism.oregonstate.edu`, accessed 12 Dec 2025. ERA5-Land data was accessed from the Copernicus Climate Change Service (C3S) (2022) using the Climate Data Store API [*ECMWF*, 2025](accessed on 23 Oct 2025). SWE measurements and related data collected and made available by Airborne Snow Observatories, Inc. and funded by the US Department of Energy, Office of Science, Biological and Environmental Research under Contract No. DE-AC02-05CH1123. All data used in this analysis is aggregated and available via `10.5281/zenodo.20116301`. All analysis code is available via `https://github.com/dlhogan97/S3-precipitation-rodeo.git`.

3.6 Appendix A

3.6.1 Manual snow water equivalent (SWE) measurements from billy barr

Manual SWE measurements are recorded at approximately 08:00 and 16:00 local time each day. First, new snow depth is measured using a ruler on a snow board that is cleared after the previous measurement. Next, an empty standard 8-inch diameter National Weather Service rain gauge is weighed and placed over the newly fallen snow. A thin metal sheet is slid beneath the gauge to collect the fresh snow, and the gauge and snow are then weighed together. Finally, the total mass is subtracted by the mass of the empty gauge to get the mass

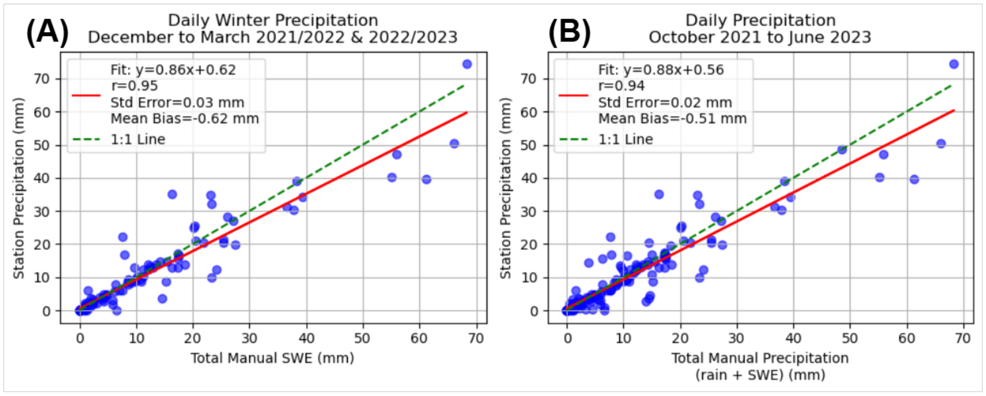

Figure 3.6: (A) Daily manual SWE observations from billy barr [sic] against daily storage gauge precipitation measurements for the December-March 2021/2022 and 2022/2023 periods. The dashed line tracks the 1:1 line, the red line indicates the best-fit regression. (B) As in the left panel, but for all manual precipitation measurements (snow + rain) from October 2021 - June 2023.

of the collected snow (in grams), which is converted to its water equivalent (in centimeters). Daily SWE totals represent the sum of the morning and afternoon measurements (billy barr, personal communication).

3.6.2 Study period data availability

3.6.3 SQUIRE

SQUIRE calculates a quantitative precipitation estimation (QPE) field from the X-band precipitation radar. To do so, it first transforms the radar data from radar antenna coordinates to a Cartesian grid using the Corrected Moments in Antenna Coordinates (CMAC) value added product from ARM. CMAC has four main steps. First, it corrects for both beam blockage and signal attenuation, which can significantly limit radar data quality in mountainous terrain. Next, it calculates liquid equivalent QPEs using an ensemble of radar-snowfall rate relationships ($Z - S$). Then, it maps these QPEs onto a 250-m grid using nearest-neighbor interpolation [*Grover et al.*, 2023]. Last, it selects the lowest vertical level

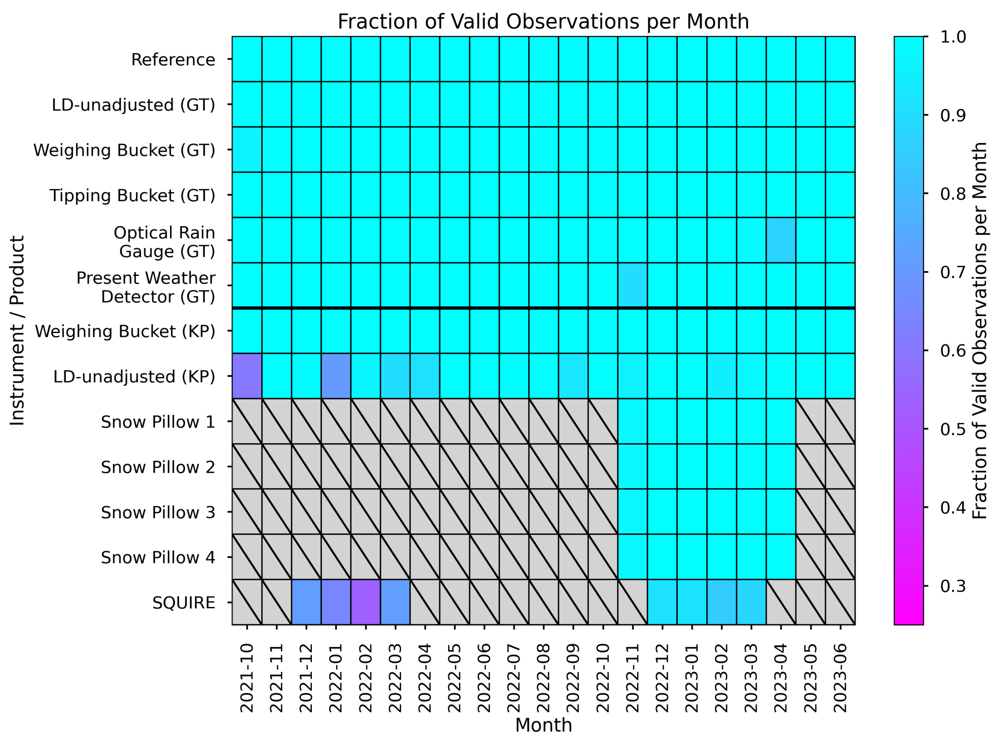

Figure 3.7: Monthly fraction of valid (not flagged or missing) observations per month. Grey, crossed-out cells indicate times when the gauge/product was inactive.

Table 3.2: Z–S relationships from various sources. Taken from [*O'Brien et al.*, 2024]

Source	Z(S)	A Coefficient	B Coefficient	Radar Band
Wolfe and Snider (2012)	$Z = 110S^2$	110	2	S
WSR-88D High Plains	$Z = 130S^2$	130	2	S
Braham (1990) 1	$Z = 67S^{1.28}$	67	1.28	X
Braham (1990) 2	$Z = 114S^{1.39}$	114	1.39	X

not blocked by terrain to locate the surface, allowing locations higher or further from the radar to be included. The $Z - S$ relationships that constituted the SQUIRE ensemble, listed in Table 3.2 were applied by *Jackson et al.* [2026] and *O'Brien et al.* [2024] and included as value added products (VAPs) in the dataset.

3.6.4 Laser disdrometer $\rho_s - D$ relationships

The laser disdrometer calculates the liquid equivalent snowfall rate (R_s) based on the mass (m_s), fall speed (V), and number concentration (N) of snowflakes within 32 discrete particle size bins (D_i):

$$R_s = 3.6 \sum_{i=3}^{32} m_s(D_i)\, V(D_i)\, N(D_i) \tag{3.1}$$

where m_s depends on both D_i and ρ_s:

$$m_s(D) = \frac{\pi \rho_s(D)}{6} D^3 \tag{3.2}$$

Three main limitations complicate these density estimates. First, snow crystal habit, which depends on atmospheric conditions aloft, can dramatically change both between and within storms, affecting snowflake density. Second, these $\rho_s - D$ relationships are also subject to the random nature of snowflake aggregation and riming. Third, different $\rho_s - D$ relationships define D in distinct ways, which increases uncertainty in their application [*Boudala et al.*, 2014]. The $\rho_s - D$ relationships we applied here are described in Table 3.3.

Table 3.3: ρ_s–D relationships used in this analysis.

Study	$\rho_s - D$ **relationship**
Holroyd and Jiusto [1971]	$\rho_s(D) = 0.17D^{-1}$
Brandes et al. [2007]	$\rho_s(D) = 0.178D^{-0.922}$
Heymsfield et al. [2004]	$\rho_s(D) = 0.104D^{-0.95}$

3.7 Appendix B

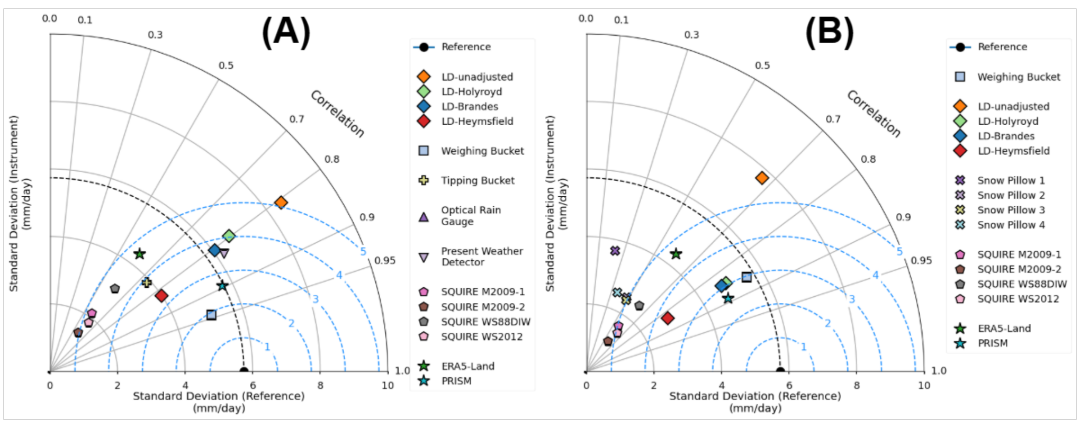

Figure 3.8: Taylor diagrams for each gauge relative to the reference (black dot on x-axis) for (A) Gothic and (B) Kettle Ponds. The angular axes show the correlation coefficient and the radial axes show the standard deviation. The black dashed line represents the standard deviation of the reference. The blue dashed radial lines are the RMSE values around the reference.

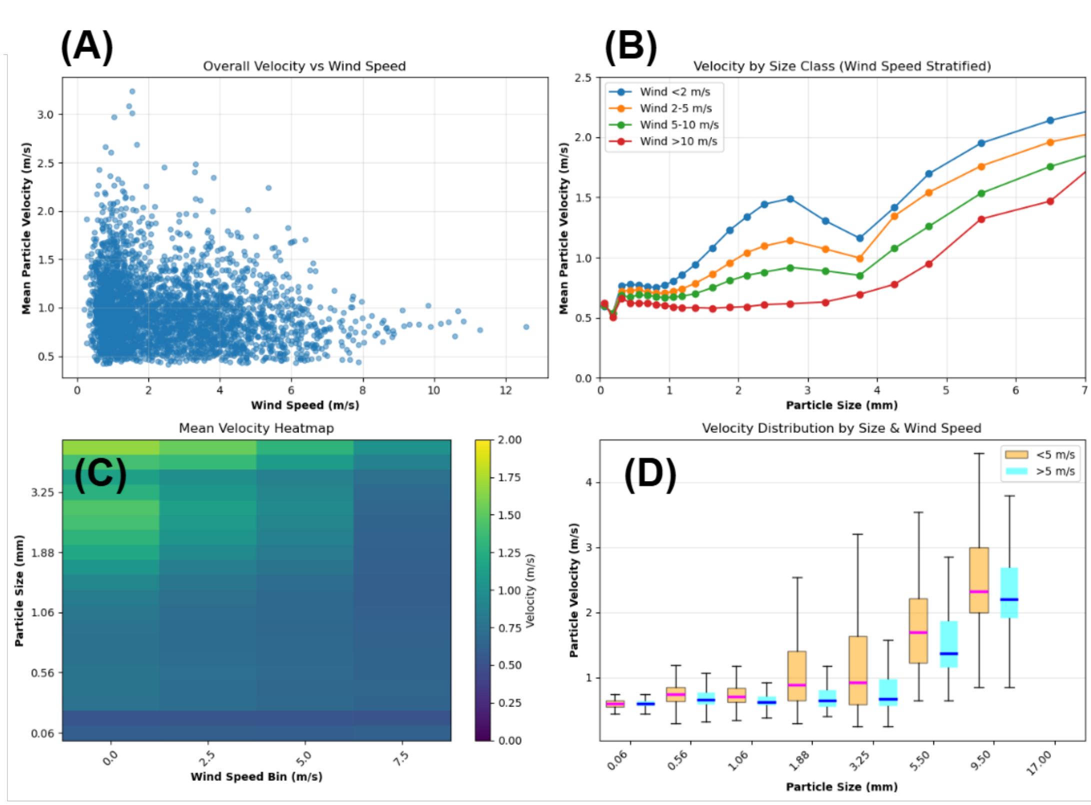

Figure 3.9: (A) Relationship between mean 10-minute wind speed and 10-minute mean particle fall velocity at the KP laser disdrometer. (B) Particle size (mm) against mean particle velocity for different wind speed thresholds. Note that the lowest fall velocities occur during the highest wind speeds. (C) Heatmap of wind speed against particle size. (D) Boxplots of particle size against particle fall velocities for wind speeds below (orange) and above (cyan) 5 m/s.

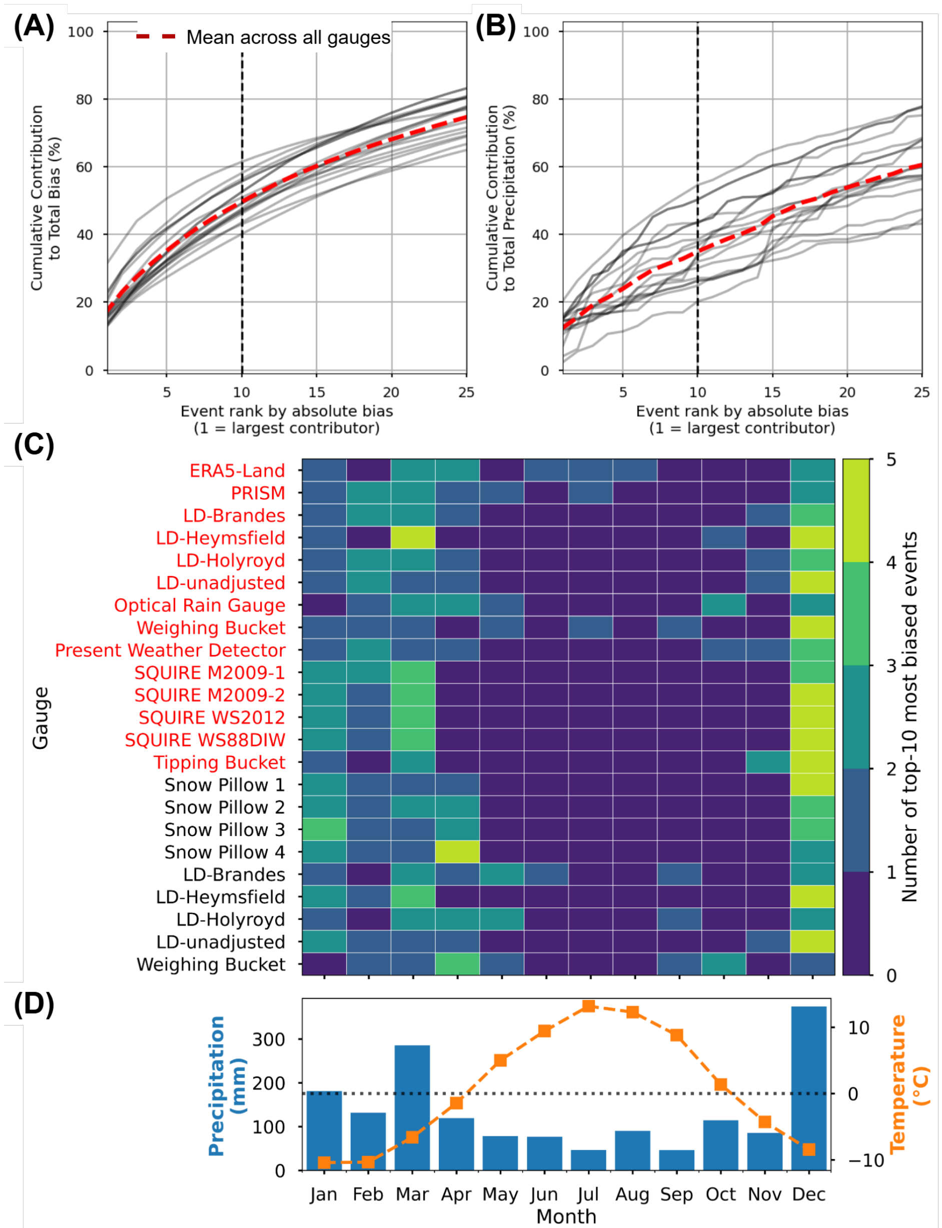

Figure 3.10: (A) Cumulative contribution of event bias to the total bias. (B) Cumulative contribution of event precipitation to total precipitation. In both plots, event rank signifies the order of the most biased events to the least biased events in ascending order (e.g. event 1 is the most biased). The dashed vertical line indicated the cumulative values for the 10th most biased event. (C) Heatmap showing the number of top-10 most biased events by month for each gauge. On the y-axis, instruments/products are labeled in red text for those at GT and black text for those at KP. (D) Monthly total precipitation from the reference gauge and average temperature for WYs 2022 and 2023.

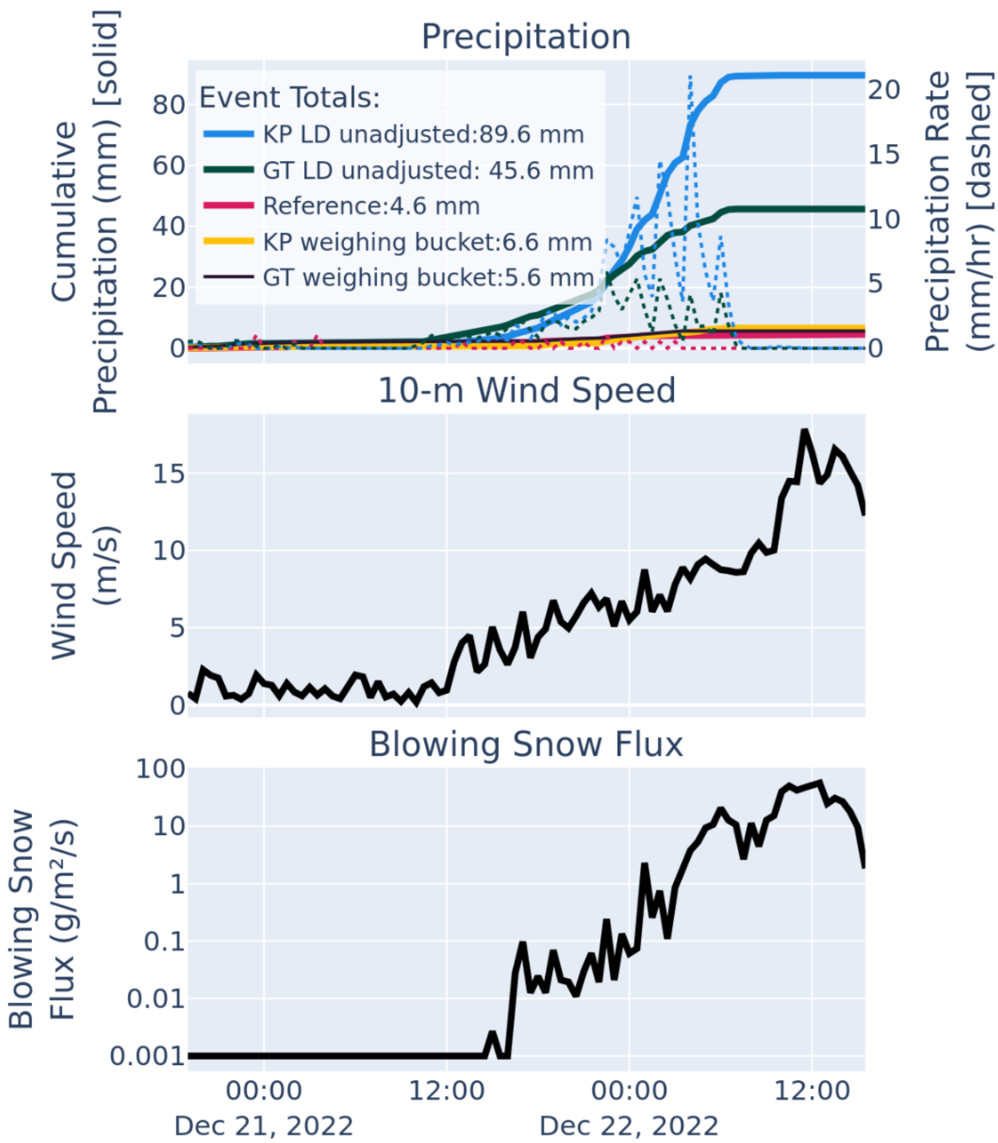

Figure 3.11: Meteorological time series of the December 21-22, 2022 blowing snow event. (Top) Cumulative precipitation (solid, left axis) and hourly precipitation rate (dashed, right axis) for the laser disdrometers at Kettle Ponds (blue) and Gothic (green), the reference gauge (red), and the weighing bucket gauges at Kettle Ponds (yellow), and Gothic (grey). (Middle) 10-m wind speed recorded at Kettle Ponds. (Bottom) Blowing snow flux measured by an acoustic blowing snow sensor at Kettle Ponds.

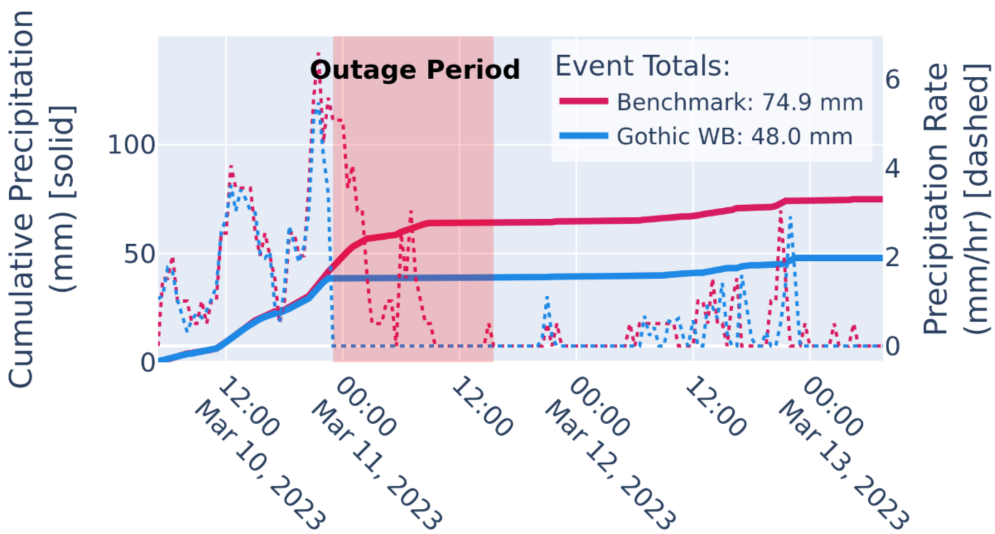

Figure 3.12: Cumulative precipitation (solid, left axis) and hourly precipitation rate (dashed, right axis) for the Gothic weighing bucket (blue) and the reference gauge (red) during the March power outage.

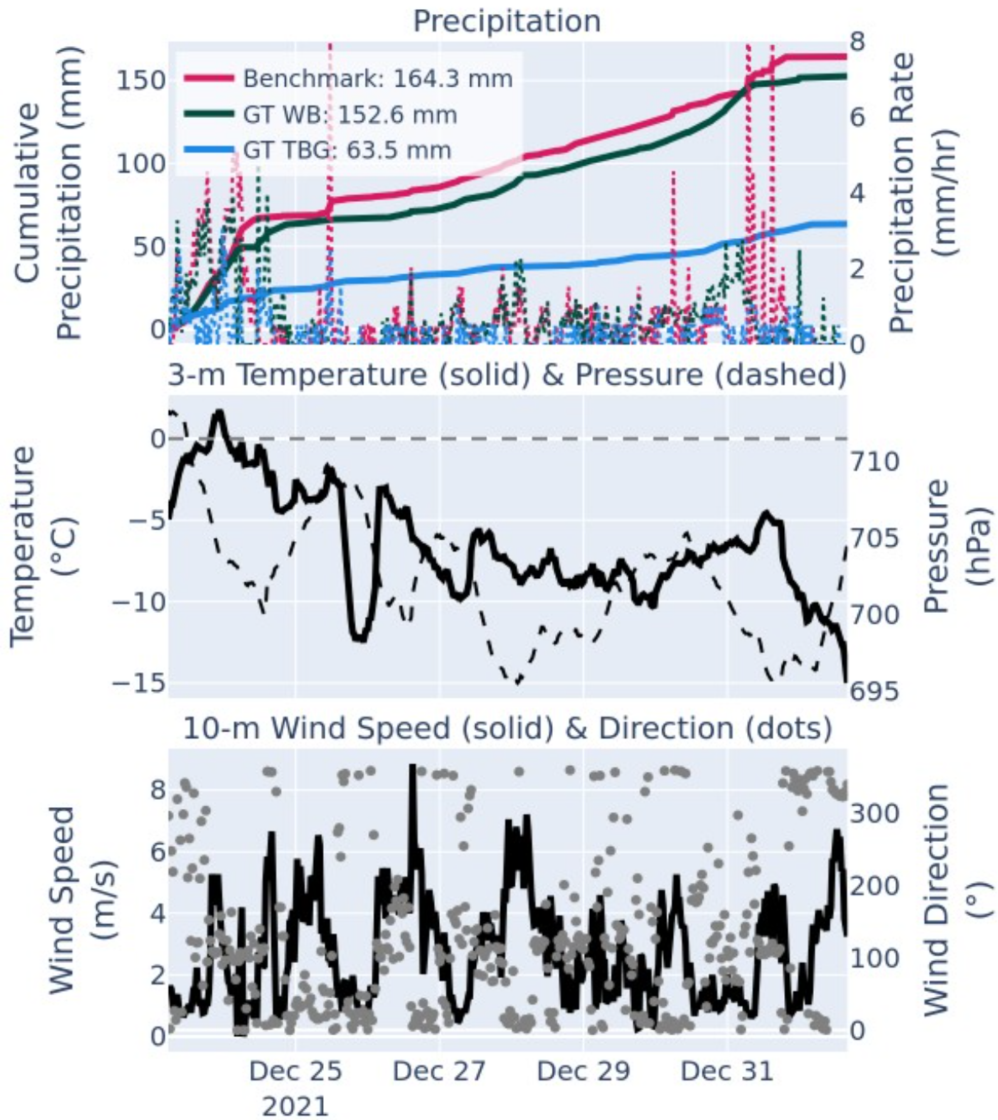

Figure 3.13: (Top) Cumulative event precipitation (solid line, left axis) and precipitation rate (dashed line, right axis) for the Gothic weighing bucket (blue), the Gothic tipping bucket gauge (light blue), and reference gauge (red). (Middle) Air temperature (left axis, solid line) and atmospheric pressure (right axis, dashed line) during the event recorded at Gothic. (Bottom) 10-m wind speed (solid line, left axis) and direction (grey dots, right axis) recorded at Gothic.

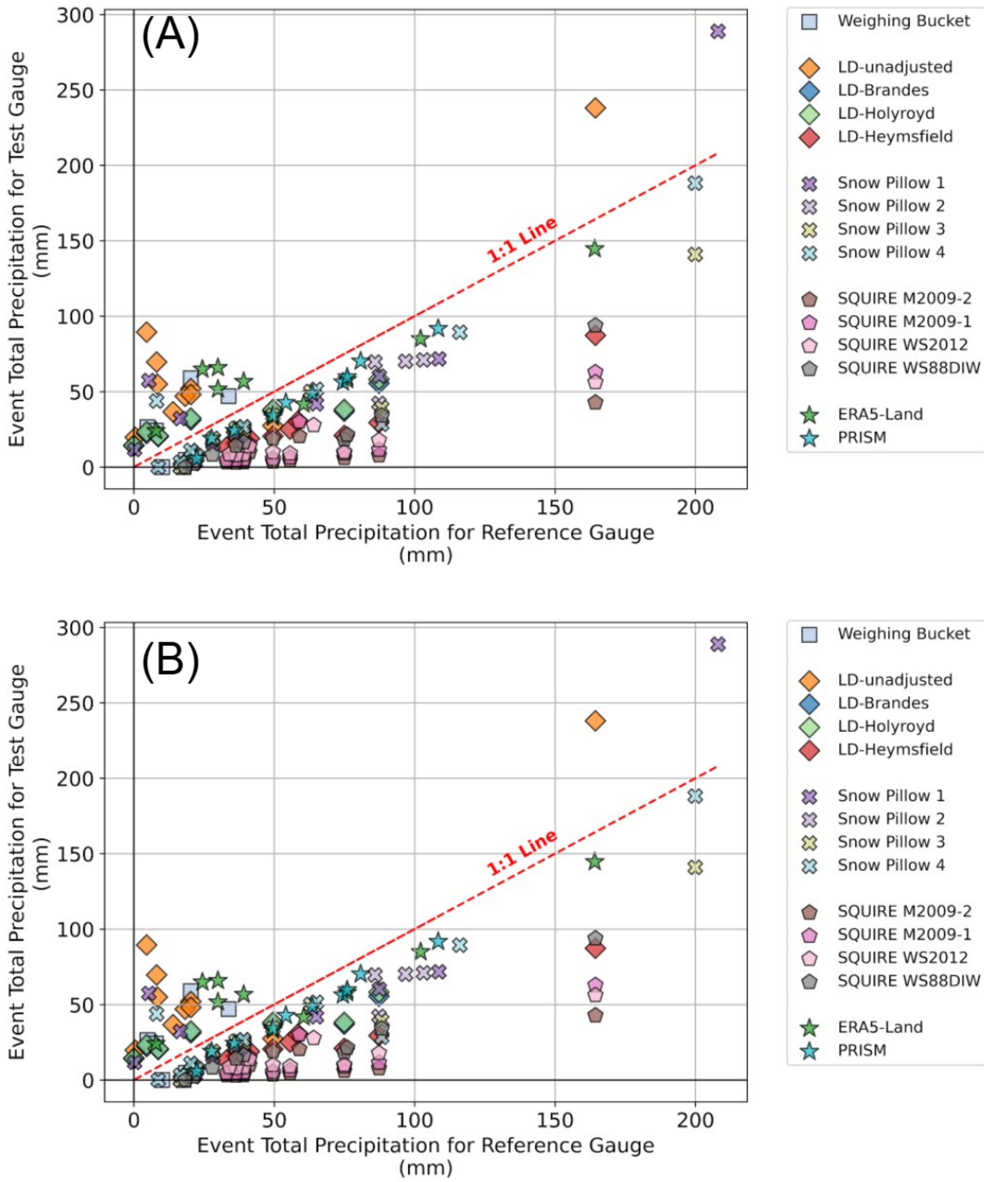

Figure 3.14: Event total precipitation at the reference gauge versus each test gauge for the 10 most biased events at (A) Gothic and (B) Kettle Ponds. Each symbol shape represents the general gauge/product type, while the color represents the specific value. The dashed red line represents the 1:1 line

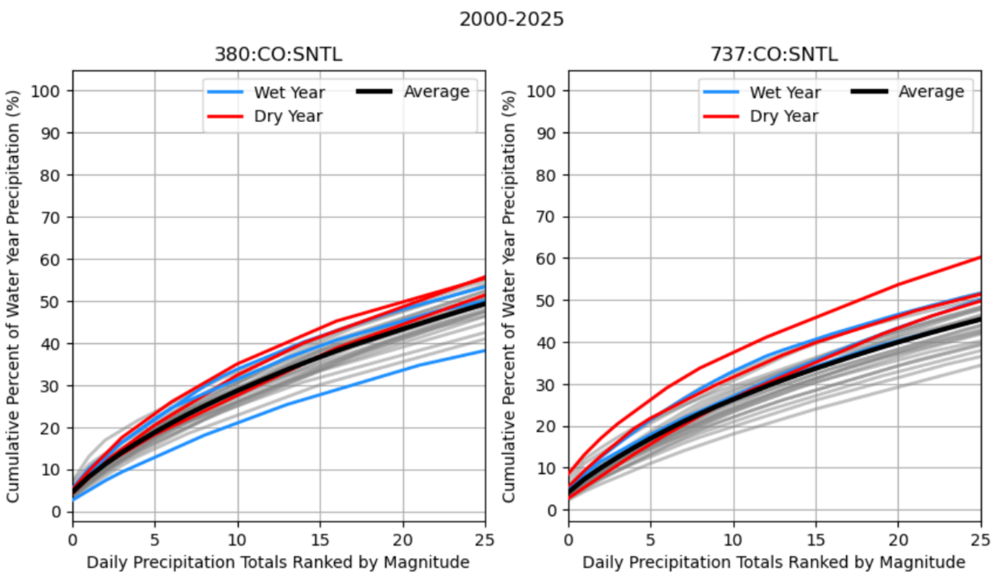

Figure 3.15: Daily precipitation ranked by magnitude against cumulative percent of water year total precipitation for the Butte (380) SNOTEL site (left) and Schofield Pass (737) SNOTEL site (right). Red (blue) lines are the 3 driest (wettest) years between 2000 and 2025 . The bold black line is the average over all years. Data downloaded from *USDA* [2024]

Chapter 4

ARE MODELS FORGETTING SOMETHING? REPRESENTING MULTI-YEAR SNOW AND SUBSURFACE STORAGE IMPROVES SIMULATED LOW-FLOW PERSISTENCE

4.1 Introduction

Mountain snowpacks serve as an invaluable seasonal water reservoir across the western United States, storing winter precipitation and releasing it gradually through spring and summer when downstream demand is highest [*Viviroli et al.*, 2020; *Li et al.*, 2017]. Snow water equivalent (SWE) has long served as a reliable early-season predictor of spring and summer streamflow in mountain regions, forming the backbone of operational water supply forecasts and water management planning [*Baker et al.*, 2021; *Mahanama et al.*, 2012; *Pagano et al.*, 2004]. Yet this historically reliable relationship periodically breaks down, with observed streamflow falling far outside of what accumulated SWE would predict [*Pagano and Garen*, 2005]. These hydrologic anomalies, recently referred to as "missing water" [*Hogan and Lundquist*, 2024; *Lapides et al.*, 2022; *Luce et al.*, 2013], carry direct consequences for reservoir operations, water allocations, and ecosystem flows, and are increasingly difficult to anticipate as snowpack becomes more variable [*Livneh and Badger*, 2020; *Qin et al.*, 2020; *Siirila-Woodburn et al.*, 2021]. Recent examples from the Upper Colorado River Basin (UCRB) include the anomalously low-runoff years of 2020 and 2021, when streamflow was only 65% and 38% of normal, respectively, even with near-normal snowpacks [*Goble and Schumacher*, 2023; *USDA NRCS*, 2026]. Anticipating years like these requires knowing where the water went.

In snow-dominated mountain basins, total water year (WY, 1 October to 30 September) streamflow (Q_{wy}) reflects the annual water balance partitioned among precipitation inputs, losses to sublimation and ET, and change in storage across soil moisture, groundwater, and perennial snow (ΔS). In a typical year, SWE captures enough of this variability to serve as the most skillful single Q_{wy} predictor. But in anomalous years, ET and ΔS take on outsized importance, and unlike SWE, both are difficult to measure at the basin scale [e.g., *Bales et al.*, 2006; *Rapp et al.*, 2020; *Tran et al.*, 2019]. Hydrologic models have therefore become the primary tool for attributing these anomalies to ET and ΔS, but only to the extent that a given model resolves the underlying process.

Recent investigations have centered on ET and soil-moisture related processes as explanatory variables. Increasing winter sublimation has been posed as a significant, though poorly constrained, source of water loss [*Gordon et al.*, 2022; *Xiao et al.*, 2018], but recent work found its magnitude and interannual variability too small to explain streamflow anomalies [*Hogan et al.*, 2026b; *Schwat et al.*, 2025]. Coincident decreases in spring precipitation and increases in springtime ET drive streamflow declines, as earlier snow disappearance extends the snow-free season and partitions a greater share of meltwater to ET rather than runoff [*Ghimire et al.*, 2026; *Hogan and Lundquist*, 2024; *Milly and Dunne*, 2020; *Palumbo et al.*, 2025]. Dry antecedent fall soils act similarly, as spring meltwater must first saturate them before contributing to runoff [e.g., *Harpold et al.*, 2017; *Lapides et al.*, 2022; *McNamara et al.*, 2005]. However, these processes operate within a single water year and cannot account for anomalies that persist over several years.

A separate class of processes connects streamflow anomalies to water storage conditions in prior years. We refer to this dependence on prior-year storage as long-term ($\geq$ year) storage memory [*Brooks et al.*, 2021; *Opie et al.*, 2020; *Orth and Seneviratne*, 2013; *Sutanto and Van Lanen*, 2022]. Memory has been detected in mountain basins using tracer-based estimates of water age [*Brooks et al.*, 2025; *Siirila-Woodburn et al.*, 2026] and lagged correlations between prior-year conditions (e.g., minimum flows, precipitation, temperature) and minimum flows in following years [*Alvarez-Garreton et al.*, 2021; *Berghuijs et al.*, 2025; *Iliopoulou et al.*, 2019; *Nippgen et al.*, 2016].

Whether these long-term storage compartments must be explicitly represented in models remains unresolved. *Rosenberg et al.* [2013] found that including long-term water storage in a semi-distributed model offered no improvement to streamflow forecast performance, while others found long-term storage was an important explanatory variable for streamflow decline in several different model frameworks [*Carroll et al.*, 2025; *Fowler et al.*, 2020; *Siirila-Woodburn et al.*, 2026]. The disagreement is at least partly architectural. Simple and complex structures alike perform well when calibrated locally [*Smith et al.*, 2004; *Troin et al.*, 2022], yet different modeling architectures produce substantially different sensitivities in ET and water storage when forced identically [*Bennett et al.*, 2019; *Chegwidden et al.*, 2019; *Elkouk et al.*, 2024; *Vano et al.*, 2012]. A model may therefore reproduce observed streamflow without reproducing the storage dynamics that generate it.

The influence of prior year hydrology depends on which storage compartment is important for streamflow generation in a given basin. In groundwater-influenced mountain basins, tracer studies indicate that groundwater aged 1 to 10 years contributes meaningfully to streamflow, and snow-dominated basins carry longer memory than rain-dominated ones [*Alvarez-Garreton et al.*, 2021; *Brooks et al.*, 2025; *Carroll et al.*, 2024, 2025; *Siirila-Woodburn et al.*, 2026; *Wolf et al.*, 2023, 2025]. Where groundwater storage is limited, perennial snow-patches and glaciers can serve a similar role, buffering interannual streamflow variability between wet and dry years [*Boardman et al.*, 2024; *Fountain and Tangborn*, 1985; *Hawkins et al.*, 2026; *Stock et al.*, 2017]. Basins with neither storage compartment show little evidence of memory [*Bilish et al.*, 2020; *de Lavenne et al.*, 2022]. In cases where prior-year storage conditions contribute to current-year streamflow, a relationship built on current-year snowpack alone may struggle when antecedent storage departs from normal.

Where these storage compartments exist, they can be progressively drawn down by multi-year drought or temperature rise, such that their capacity to sustain low flows diminishes over time [*Carroll et al.*, 2024, 2025; *Siirila-Woodburn et al.*, 2026; *Hawkins et al.*, 2026]. Drawdown may also alter the storage–low flow relationship itself: as storage declines and flow paths lengthen, the water sustaining streamflow becomes older, potentially extending rather than weakening memory.

Capturing either behavior requires a model that contains such a storage compartment. Until recently, these stores, such as deep groundwater, have been treated as minor contributors to streamflow, so simpler operational and research models typically represent only an active, shallow soil layer [*Brooks et al.*, 2025; *Rapp et al.*, 2020]. Where long-term storage behavior does emerge in these models, it typically arises through calibration, by repurposing components designed for other functions. Such models may match observed streamflow during calibration while producing storage that does not respond to sustained drought or recharge as a discrete storage compartment would [*Fowler et al.*, 2020; *Zhang et al.*, 2026].

Evidence that models struggle to represent multi-year storage memory comes largely from rainfall-dominated catchments, while work demonstrating its importance in mountain systems relies on models far more complex than those used in broader water resource management contexts. Whether a simpler model benefits from representing long-term storage explicitly is therefore unclear, especially in mountain basins where that storage may take very different forms. Here we use streamflow observations and a physically based, semi-distributed hydrologic model to address three questions in two contrasting snow-dominated basins: (1) Do observed streamflow records carry a signature of long-term storage memory, and can a model reproduce this signature without representing that storage explicitly? (2) How and where does adding a basin-appropriate long-term storage compartment change simulated streamflow? and (3) Under what conditions does representing storage matter most in these basins?

4.2 Methods

We conduct our study between 1981 and 2025 in two contrasting western US snow-dominated basins that have different dominant sources of long-term storage: the East River, CO, where deep groundwater dominates, and the Tuolumne River, CA, where limited groundwater storage leaves perennial snow patches as the principal carryover water store. Extensive prior work in both basins makes them useful experimental testbeds [e.g., *Carroll et al.*, 2019, 2024, 2025;

Siirila-Woodburn et al., 2026; *Lundquist et al.*, 2016, 2005, 2024]. We use the Structure for Unifying Multiple Modeling Alternatives (SUMMA) hydrologic model, which supports alternative storage representations within a single framework [*Clark et al.*, 2015a, b]. SUMMA has been used to investigate the impacts of model structure on components of the water balance [*Alden et al.*, 2026; *Cristea et al.*, 2022; *Currier et al.*, 2017; *Knoben et al.*, 2022; *Wayand et al.*, 2016]. We characterize memory from the streamflow record itself, using the lag autocorrelation of annual late season flows as a proxy for memory and the observed characteristic recession as a constraint on drainage timescale. We first compare these signatures against simulations from calibrated baseline models and from configurations in which a deep aquifer and a perennial snowpatch are added to the East and Tuolumne Rivers, respectively. We then evaluate how the state of each storage compartment directly relates to low flows. Finally, we compare water-year streamflow from each configuration against an April 1 SWE–streamflow regression to compare their relative performance.

4.2.1 Study sites

Tuolumne River Basin, California

The Tuolumne River basin above the Hetch Hetchy Reservoir is a snow-dominated catchment located in Yosemite National Park within California's central Sierra Nevada mountains. It covers approximately 774 km^2, with elevations ranging from 1170 to 3930 m (Figure 4.1). Basin-mean WY total precipitation averages 1160 mm per year [*Daly et al.*, 2008], though there is significant interannual variability, with a few large winter storms accounting for most of the annual precipitation [*Lundquist et al.*, 2015a]. At high elevations (>2000 m), snowfall dominates precipitation, while lower elevations lie in the rain-snow transition zone. Snowpack normally peaks between March and April and melts between May and July, but during extreme years, snow can persist into the following year at high elevations [*Boardman et al.*, 2024; *Hawkins et al.*, 2026]. Additionally, high elevation cirques harbor glaciers and persistent snow patches that support late summer flows [*Boardman et al.*, 2024; *Hawkins*

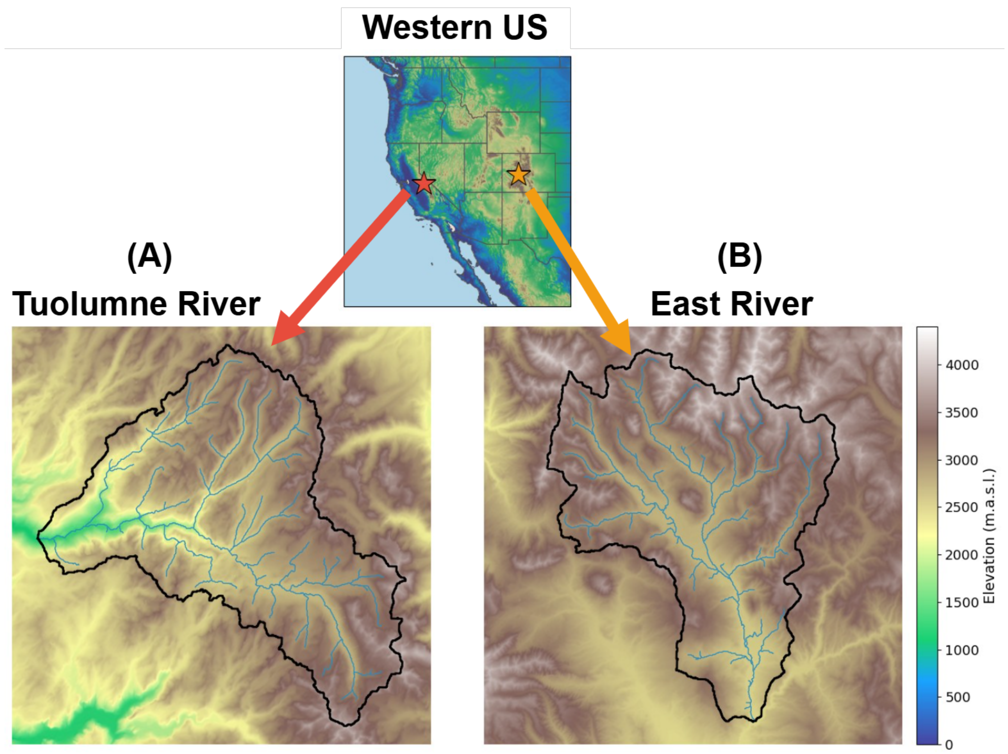

Figure 4.1: Locations and basin boundaries of the (A) Tuolumne (red star) and (B) East (orange star).

et al., 2026; *Lundquist et al.*, 2005].

The basin is underlain predominantly by granitic bedrock with shallow, coarse-textured soils in the upper parts of the basin and limited deep aquifer storage [*Lundquist et al.*, 2016; *Lundquist and Loheide II*, 2011]. Vegetation is characterized by mixed conifer forest at lower elevations transitioning to alpine meadow and bare granite at higher elevations.

Streamflow observations came from reconstructed flows into Hetch Hetchy reservoir that span 1970 to 2024 [Bruce McGurk, personal communication, 10 April 2026; *Hourdequin*, 2018; *Lundquist et al.*, 2016]. These reconstructed flows have uncertainties comparable to standard streamflow observations [*Lundquist et al.*, 2016, $\approx \pm 10\%$].

East River Basin, Colorado

The East River basin is a high-elevation, snow-dominated headwater catchment of the UCRB located in central Colorado that is representative of other high-elevation watersheds in the region [*Hubbard et al.*, 2018; *Slosson et al.*, 2025]. The basin drains 748 km^2 and spans an elevation range of 2,450 to 4,310 m (Figure 4.1B). Basin-mean WY precipitation averages 840 mm per year [*Daly et al.*, 2008], a majority of which falls as snow. The annual snowpack typically peaks between March and April and persists from May until June [*Prather et al.*, 2023].

Soils are predominantly sandy loam derived from mixed sedimentary and metamorphic parent material, with significant spatial variability in depth and hydraulic properties across the basin [*Carroll et al.*, 2018; *Hubbard et al.*, 2018]. Land cover transitions from low-lying grasses and shrubs at low elevations to subalpine forest dominated by conifers and aspens at mid-elevations to alpine above treeline ($\approx$ 3300 m) [*Carroll et al.*, 2020]. Groundwater plays a well-documented role in the basin's hydrologic cycle, sustaining late-season baseflow and introducing multi-year memory to the water balance [*Carroll et al.*, 2018, 2019, 2024; *Siirila-Woodburn et al.*, 2026]. Observed daily streamflow was acquired from the United States Geological Survey (USGS) gauge 09112500 [*Survey*, 2016].

4.2.2 Baseflow recession analysis

We characterize the recession curve using the observed daily streamflow record following the expression:

$$Q_t = Q_o k^t, \tag{4.1}$$

where Q_t is the discharge at time t, Q_o the initial discharge, and k is the recession constant [*Nathan and McMahon*, 1990]. Recessions that follow Eq. 4.1 plot as a straight line on a semi-log plot, with the slope equal to $\ln k$. From there, the characteristic drainage timescale, τ, is equal to $\frac{-1}{\ln k}$. While groundwater dynamics often behave in a nonlinear fashion and more complex functions exist to try to capture that nonlinearity [*Nathan and McMahon*, 1990; *Wittenberg*, 1999; *Kirchner*, 2009], we chose to use Eq. for its foundational use in hydrology [*Hall*, 1968].

Cleanly separating baseflow from other major streamflow components, like interflow, can be somewhat subjective, and baseflow itself may have several contributing sources (e.g., groundwater, residual snowmelt) that drain at different rates [*Hall*, 1968; *Nathan and McMahon*, 1990]. Nevertheless, faster draining components dominate only briefly within the total recession period before dropping out, leaving the slowest draining source to control the tail (Figure 4.2). We isolate this tail behavior using the matching strip method, in which recession segments from across the entire streamflow record are superimposed until they overlap to form a single composite curve, termed the master recession curve (MRC) [*Nathan and McMahon*, 1990; *Toebes and Strang*, 1964]. The slope of that MRC therefore approximates the basin's characteristic drainage timescale (Figure 4.2B).

4.2.3 Hydrologic Memory

We define streamflow memory as the autocorrelation between low flows in one year to low flows in following years [*Berghuijs et al.*, 2025]. We characterized it using the lag-j spearman autocorrelation (ρ_j) of late season low flows from the daily streamflow record. These low flows were defined as the minimum 7-day rolling mean streamflow ($Q_{7,min}$) between 1 August and

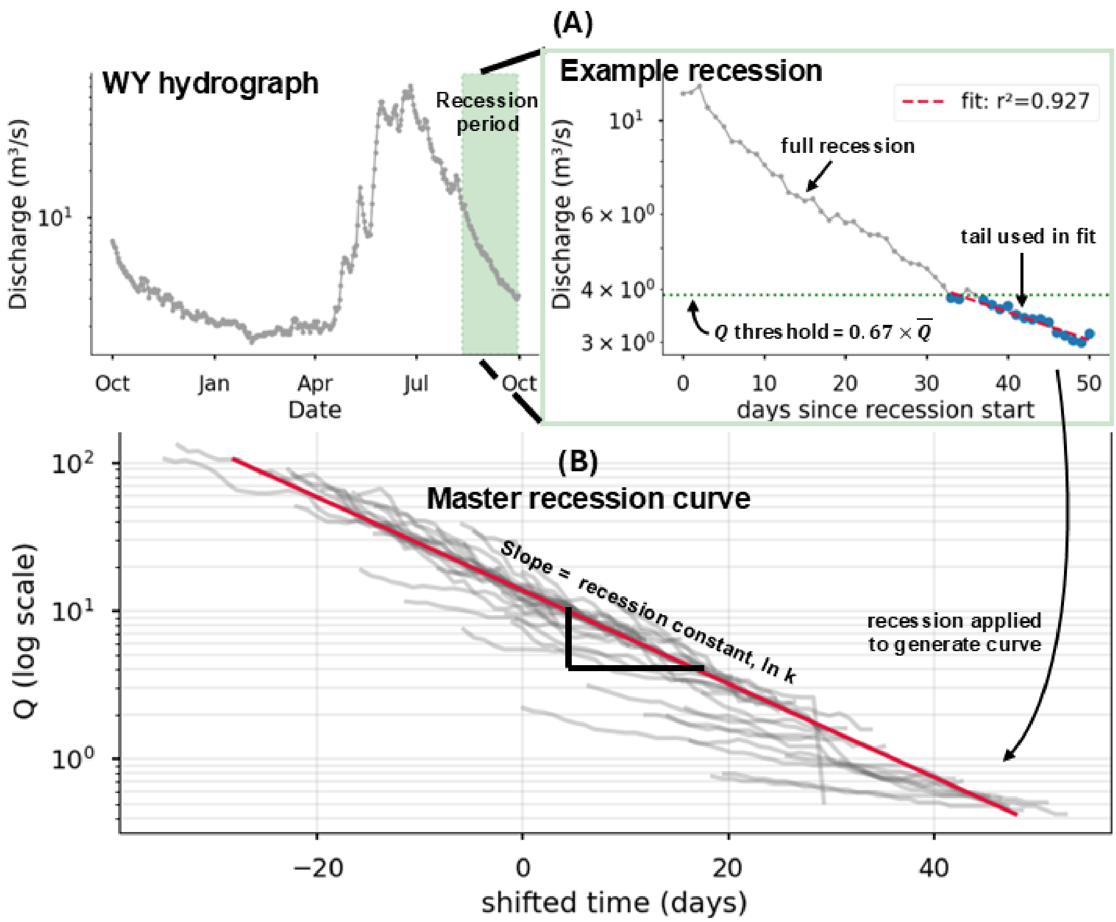

Figure 4.2: (A) Example baseflow recession curve from the East River basin. We only used the tail of each recession, defined as beginning when streamflow dropped below two-thirds of its mean during the full recession, to generate the master recession curve. (B) The master recession curve generated for the East River.

31 October. We chose this window to avoid potential artifacts from ice affecting streamflow measurements during winter. Each $Q_{7,min}$ timeseries was tested for a monotonic trend using Spearman's rank correlation against water year. Where the trend was significant ($p < 0.10$), we modeled it using a Theil-Sen estimator to detrend the timeseries, ensuring the detrended values remained strictly non-negative. If no significant trend was detected, the original data was preserved. ρ_j was then computed on the resulting series for lags $j = 1$ to 5 years given by:

$$\rho_j = \frac{\sum_{t=1}^{n-j}(Q_t - \bar{Q})(Q_{t+j} - \bar{Q})}{\sum_{t=1}^{n}(Q_t - \bar{Q})^2}, \tag{4.2}$$

where Q_t is the $Q_{7,min}$ for year t and $\bar{Q}$ is the mean $Q_{7,min}$.

To assess the significance of these autocorrelations, we compared each observed ρ_j value against a distribution of autocorrelations generated from 2,000 random permutations of the annual low-flow series that remove effects from temporal ordering. We considered an autocorrelation to be statistically significant at $p < 0.10$ [*Berghuijs et al.*, 2025], where p is the probability that a given permutation exceeds that of the observed value.

4.2.4 Model Configurations

SUMMA solves coupled energy and mass balance equations at the hydrologic response unit (HRU) level. HRUs can vary in shape or size and can combine many hydrologically similar, but spatially disconnected, areas [*Clark et al.*, 2015a]. In our case, we discretized HRUs using the intersection of elevation bands and aspect classes to capture the distinct orographic and topographic gradients that drive local variation in temperature, precipitation, and solar radiation (Figures 4.3A-D). This topographic information was extracted from the USGS 30-m 3D elevation program digital elevation model (DEM) using the TauDEM data [*Tarboton*, 2002; *Survey*, 2022]. Table 4.6) in Section 4.6.4 summarizes the attributes of each HRU.

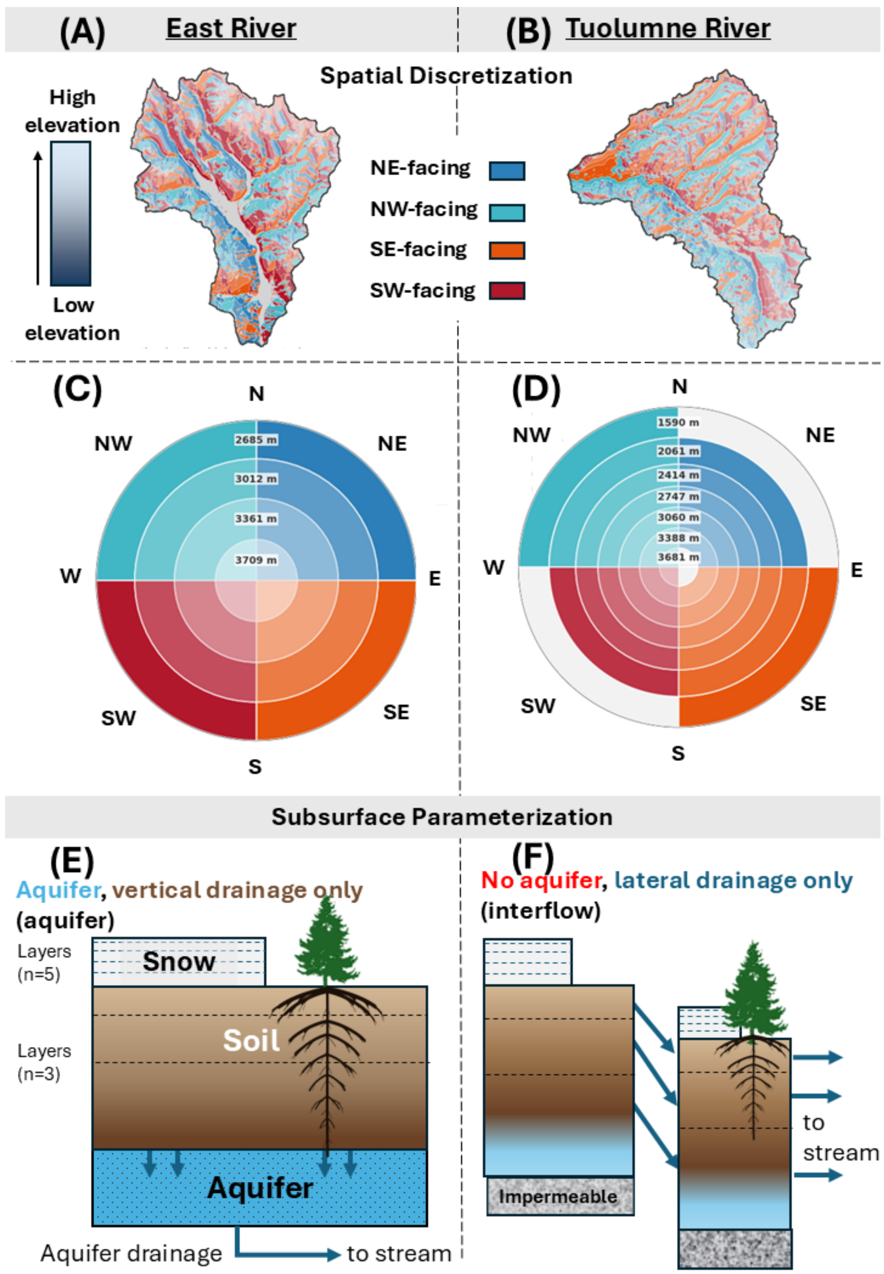

Figure 4.3: Spatial and subsurface discretizations for the East and Tuolumne Rivers. (A) and (B) show the HRU delineation over the outline of each basin. Shading decreases with elevation so that higher elevations appear lighter colored. Aspects are colored by dominant down-slope direction, with gray shaded HRUs representing flat HRUs. (C) and (D) depict the elevation bands as rings that increase towards the center and the four aspect directions. (E) and (F) are schematics depicting the subsurface parameterizations used for the East and Tuolumne Rivers, respectively.

Baseline Configuration

We applied subsurface schemes that align with the current understanding of subsurface flow in these two basins [*Carroll et al.*, 2024; *Lundquist and Loheide II*, 2011]. Key parameters for each scheme are listed in Table 4.7 in Section 4.6.6. In the East, we applied the "aquifer" scheme, which routes soil drainage into a storage reservoir that generates baseflow as a power law function of aquifer storage, with subsurface fluxes directed to the basin outlet [Figure 4.3C; *Clark et al.*, 2008]. SUMMA expresses this baseflow following:

$$b_f = b_{aq}(\frac{S}{S_f})^b, \tag{4.3}$$

where b_{aq} is a baseflow rate parameter, S_{aq} is the simulated aquifer storage state, S_f is an aquifer scaling factor, and b is the aquifer baseflow exponential parameter. For $b = 1$, this reduces to $Q_{bf} = S_{aq} \times k$ with $k = \frac{b_{aq}}{S_f}$, corresponding to the linear case of Eq 4.1.

In the Tuolumne, we applied the interflow scheme, which transfers water laterally downslope following a TOPMODEL transmissivity decay function, with HRUs connected from high to low elevation [4.3F; *Beven and Kirkby*, 1979]. Neither baseline configuration included an explicit water storage compartment designed solely to carry water across years.

Long-term storage configurations

We added a basin-appropriate long-term storage compartment to each baseline model without recalibrating it, so that any change in streamflow behavior was attributable to the imposed storage.

East River deep store To replicate a slower-draining storage compartment consistent with deep groundwater behavior in the East River basin [*Carroll et al.*, 2024], we increased the S_f parameter used in Eq. 4.3. Doing so damps the aquifer's drainage response: more water accumulates before baseflow increases, and storage remains higher for longer, sustaining baseflow through extended dry periods (Figure 4.4A). We applied this change to the HRUs

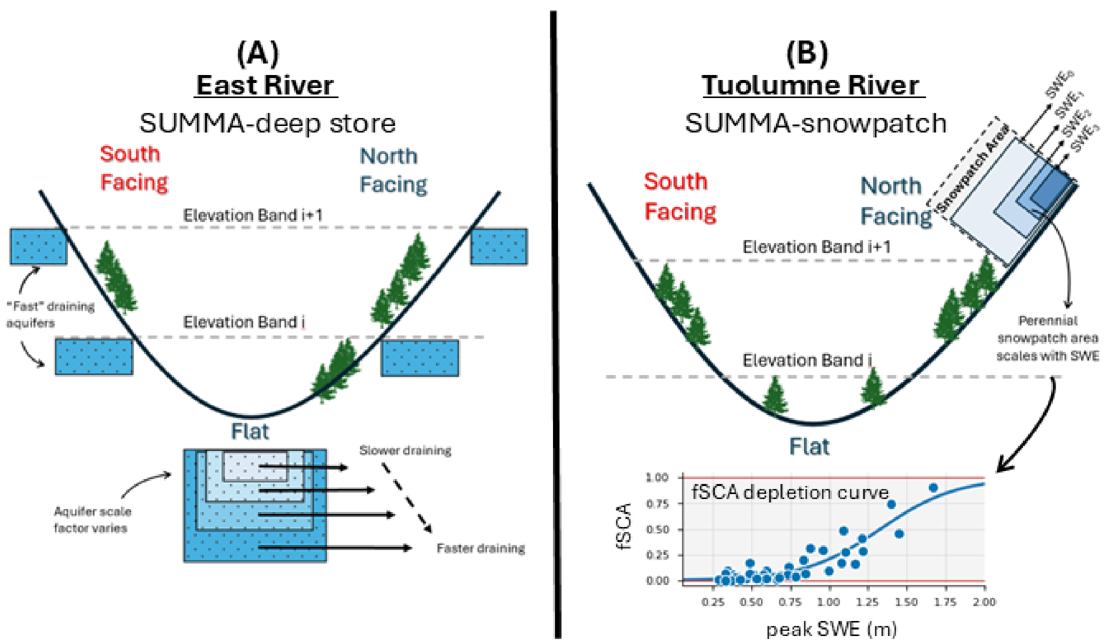

Figure 4.4: (A) Conceptual diagram of the East River's deep store configuration. The low-elevation HRUs have an aquifer with a larger scale factor, which controls the size and drainage rate of the aquifer. (B) Conceptual diagram of the Tuolumne River's snowpatch HRUs. The area of the snowpatches is designed to follow the Landsat-derived fSCA depletion curve, which relates changes in fSCA to changes in modeled peak SWE.

in the lowest elevation band, which covered 31% of the basin. We refer to this model output as "deep store."

Tuolumne River snowpatch configuration To represent the perennial snowpatches observed in the Tuolumne [*Hawkins et al.*, 2026], we subdivided the model's high-elevation, north-facing HRU (8.6 km^2, ≈1% of basin area) into ten sub-HRUs and prescribed their initial snow mass (Figure 4.4B). This was necessary because SUMMA carries no fractional snow-covered area (fSCA) information: each HRU is either fully snow-covered or snow-free, so a single snowpatch HRU could not represent a snowpatch that grows or shrinks between wet and dry years. We therefore treated the sub-HRUs as depth bins ordered by snowpatch persistence, with each covering a progressively smaller area (Table 4.1), so that shallower bins melted out first. In combination, the sub-HRUs could replicate fSCA evolution over time.

Table 4.1: Snowpatch sub-HRU discretization within the Tuolumne high-elevation, north-facing HRU. Bins are ordered by snowpatch persistence, with deeper bins covering progressively smaller areas.

Bin	Area (km^2)	Area (%)	Initial SWE (m)
1	0.04	0.5	20.0
2	0.04	0.5	6.0
3	0.09	1	4.0
4	0.17	2	2.5
5	0.34	4	1.2
6	0.69	8	0
7	1.38	16	0
8	2.07	24	0
9	2.07	24	0
10	1.72	20	0
Total	8.61	100	0.3[a]

[a]Area-weighted mean initial SWE across all bins.

Bin areas and initialized depths were interpolated from an empirical fSCA-SWE depletion curve we developed for the basin (Figure 4.4B). This curve related simulated peak SWE from the high, north-facing HRU (baseline configuration) to late-season (August-October) fSCA. Late-season fSCA was estimated from the available Landsat imagery between 1 August and 1 October for each year: we computed normalized differential snow index (NDSI) for each available scene covering the HRU, excluded images contaminated by clouds, classified snow cover from NDSI, and took the annual minimum fSCA across the remaining images. Years with fewer than three usable images in this late-season window (1986-1992) were excluded. This method is described further in the Appendix Section 4.6.1. The resulting fSCA depletion curve was made by fitting these data to a logistic regression ($r^2 = 0.83$), following *Bales et al.* [2015]. From there, we initialized each bin with enough SWE to reproduce the observed area-depth relationship (Table 4.1). We refer to this configuration as "snowpatch."

4.2.5 Model Forcing and Basin Attributes

SUMMA requires seven meteorological forcing variables: precipitation (mm), air temperature (K), incoming shortwave and longwave radiation (W m^{-2}), air pressure (hPa), specific humidity (kg/kg), and wind speed (m s^{-1}). Precipitation, air temperature, and incoming shortwave radiation were taken from the daily 4-km Parameter-elevation Regressions on Independent Slopes Model (PRISM) data [*Daly et al.*, 2008]. Hourly wind speed was taken from the European Center for Mid-Range Weather Forecasting (ECMWF) ERA5-Land reanalysis (9 km^2) [*Muñoz-Sabater et al.*, 2021]. Pressure, specific humidity, and longwave radiation were estimated using empirical relationships detailed in Section 4.6.2. Daily data were temporally disaggregated to hourly timesteps using the METSIM data package [*Bennett et al.*, 2020]. Forcing was distributed as the area-weighted mean over each HRU using the EASYMORE remapping tool [*Knoben et al.*, 2022].

Because precipitation is a dominant source of hydrologic model uncertainty [*Bárdossy et al.*, 2022], we compared PRISM winter precipitation gradients against Airborne Snow Observatories (ASO) lidar SWE estimates [*Painter et al.*, 2016]. PRISM's elevation gradient was accurate in the East but too weak in the Tuolumne, so we weighted Tuolumne winter precipitation across elevation bands to match the ASO SWE gradient while preserving total volume (Appendix Figure 4.12).

Land cover and soil type were extracted from the 2021 30-m National Land Cover Dataset (NLCD) [*Dewitz and Survey*, 2021] and the 250-m SoilGrids dataset [*Poggio et al.*, 2021]. For each HRU, we used the most common land cover and soil class contained within that HRU. Soil depth increased from 0.2 m at the highest elevation band to 1.0 m at the lowest [*Lundquist et al.*, 2016; *Wainwright et al.*, 2024]. Deeper soils in meadows and riparian areas were not represented in these configurations.

4.2.6 Model Calibration

Model parameters were calibrated per basin against observed streamflow using the differential evolution algorithm [*Storn and Price*, 1997], which is well-suited to high-dimensional, nonlinear parameter spaces where gradient-based methods risk finding local minima [*Mai et al.*, 2020]. We used a 10-year calibration period that spanned WYs 1993-2002, with a 3-year spin-up (1989-1992).

SUMMA specifies model parameters at the HRU level, but treating each HRU independently is both computationally inefficient and physically unreasonable [*Mai et al.*, 2020; *Newman et al.*, 2014]. We therefore implemented a spatially distributed parameter multiplier approach in which a base parameter set is defined for the basin, and calibrated weights scale parameters across HRUs. The effective parameter θ_{eff} for HRU i is:

$$\theta_{eff,i} = \theta_{base} \times M \times w_i, \tag{4.4}$$

where θ_{base} is the base parameter value, M is a basin-wide scalar calibrated by the optimizer, and w_i is the physically motivated spatial weight for i. Table 4.2 lists the optimized parameters and whether they were spatially variable.

For soil hydraulic conductivity (K), we assumed K was constant with elevation [*Rapp et al.*, 2020]. In the East River, the θ_{base} value for K and field capacity is estimated from in-situ observations [*Falco et al.*, 2024; *Wainwright et al.*, 2024]; in the Tuolumne, it was estimated for each HRU's assigned SoilGrids texture and bulk density using the ROSETTA pedotransfer function [*Schaap et al.*, 2001].

Rather than optimize for daily streamflow, we calibrated against three streamflow metrics, each expressed as a mean absolute error over the calibration period: the timing of the annual flow centroid, the sum of absolute monthly volume errors, and the low flow (August–October) volume error. These streamflow objectives constrain the seasonal distribution and late season behavior of flow, rather than just day-to-day variability. Full metric definitions are given in the Appendix Section 4.6.3.

Table 4.2: Model parameters optimized during calibration.

Parameter	Description	Spatiallyvariable
`k_soil`	Saturated soil hydraulic conductivity	No
`rootingDepth`	Depth of roots in the soil layers	Yes
`qSurfScale`	Surface runoff generation scaling parameter	Yes
`kAnisotropic`	Lateral to vertical hydraulic conductivity ratio	Yes
`zScale_TOPMODEL`	Depth scale for vertical drainage and transmissivity decay	Yes
`frozenPrecipMultip`	Frozen precipitation multiplier	Yes
`tempCritRain`	Rain–snow temperature threshold	No
`routingGammaScale`	Scale parameter in Gamma distribution for routing	No
East River only		
`aquiferScaleFactor`	Aquifer storage capacity scaling parameter	No
`aquiferBaseflowRate`	Maximum aquifer baseflow rate	No

4.2.7 Model Evaluation

Streamflow

In each basin, we ran a long-term simulation spanning WYs 1986–2025 with a 5-year spin-up period (1981-1985). Model streamflow performance was evaluated over WYs 2015-2024 using the mean bias in water year total flow volume, and in the 7-day maximum and minimum flows. In addition, we report streamflow Kling-Gupta efficiency (KGE), Nash-Sutcliffe efficiency, and log-KGE – which emphasizes low flows – in Table 4.8 in the Appendix Section 4.6.6, though these were not calibration targets.

ET and Snow

To confirm that simulated snow and ET were reasonably represented, we evaluated the spatial patterns of ET and the elevation distribution and timing of SWE across HRUs. For ET, we compared simulated water year total ET against OpenET's ensemble mean averaged over the basin [*Melton et al.*, 2022]. Despite known limitations in mountain terrain, OpenET provides the longest available continuous time series of distributed observation-based ET estimates in both basins [*Volk et al.*, 2024]. In the East River, we also compared simulated ET in the valley-bottom HRU against eddy covariance measurements collected there between November 2022 and June 2023 during the Sublimation of Snow campaign [*Lundquist et al.*, 2024; *Schwat et al.*, 2025].

For snow, we compared the elevation distribution of simulated SWE against ASO lidar-based SWE estimates [*Painter et al.*, 2016]. Both were aggregated to elevation bands, with simulated SWE taken from the model state on each available flight date. We also compared simulated SWE from the HRU containing a mid-elevation SNOTEL station against snow pillow observations at that station (Table 4.9, Appendix Section 4.6.6), evaluating peak SWE volume and melt duration (days from peak SWE to melt-out) over the modeled period. Last, we compared the time series of simulated late season snowpatch area to the Landsat-derived estimates between 1993 and 2025.

Streamflow Memory

To compare memory signatures, we calculated ρ_j for the observed and modelled $Q_{7,min}$ timeseries using lags of 1 to 5-years over the modeled period. We only detrended the observed Tuolumne $Q_{7,min}$ timeseries, which showed a decreasing $Q_{7,min}$ trend of -1.3% yr^{-1} (p=0.05). The East River did not show a significant trend.

To compare recession behavior between models and observations, we constructed the MRC for each timeseries from all filtered recessions, and took the reciprocal of the MRC slopes, which equals the characteristic recession timescale (τ). The specific details describing how recessions were filtered are described in the Appendix Section 4.6.4.

Long-term storage impacts on low flows

To test whether each basin's long-term storage compartment affects low flows independently of the magnitude of the snow year, we calculated the partial rank correlation between variables following:

$$\rho_{x_1,y|x_2} = \frac{\rho_{x_1,y} - \rho_{x_1,x_2}\,\rho_{y,x_2}}{\sqrt{\left(1-\rho^2_{x_1,x_2}\right)\left(1-\rho^2_{y,x_2}\right)}} \tag{4.5}$$

In this case, x_1 is the storage state (S) at the end of each water year (snowpatch area in the Tuolumne, aquifer storage in the East), x_2 is peak SWE of that same year, and y is $Q_{7,min}$ of the following year.

To further evaluate the relationship between the storage compartment state and low flows, we regressed log-transformed low flow on log-transformed storage size (S) using ordinary least squares (OLS).

$$\ln Q_{7,min} = \alpha + \varepsilon \ln S, \tag{4.6}$$

where the coefficient ε is an elasticity, representing the percent change in the following year's low flow per 1% change in S.

SWE-Q regression comparison for total WY streamflow

We compared total WY streamflow (Q_{wy}) from each SUMMA model configuration against an April 1 SWE-Q regression to test their performance relative to a common operational baseline. To represent the baseline prediction, we used OLS regression to relate observed Q_{wy} to April 1 SWE observations from the representative SNOTEL sites in each basin (Table 4.9 in Appendix Section 4.6.6) and calibrated the relationship over the same period as the SUMMA models. We evaluated all three models (baseline, deep store/snowpatch, and regression) against observed Q_{wy} using percent bias and root mean squared error (RMSE).

4.3 Results

4.3.1 Streamflow evaluation

In both basins, the simulated mean WY hydrographs closely mirrored the observations over the evaluation period (Figure 4.5A,B), while mean Q_{wy} biases were all below ±10% (Table 4.3). Although 7-day maximum and minimum flow biases were larger than the Q_{wy} biases (Table 4.3), each model could reasonably capture the observed hydrograph's flow patterns during more extreme wet and dry years (Figure 4.5C-F), which are more likely to have larger forcing errors that affect the total volume and timing of snowmelt-driven runoff [*Mizukami and Smith*, 2012]. Overall, there were minor differences between the calibration and evaluation periods (Table 4.3), indicating these models were not overfit during calibration. NSE, KGE, and log-KGE also indicated satisfactory performance (Table 4.8).

The low flow and high flow metrics showed the largest differences between the baseline and deep store models in the East River (Table 4.3). $Q_{7,min}$ bias increased from 3% to 9%, while $Q_{7,max}$ bias changed from -2% to -5% in the deep store model, indicating that more water was partitioned to low flows at the expense of peak flows (Table 4.3). Total low-flow volume bias decreased from 14% to 12% and log-KGE increased from 0.80 to 0.90, indicating that overall low-flow behavior was better captured in the deep store model (Table 4.3, Table 4.8).

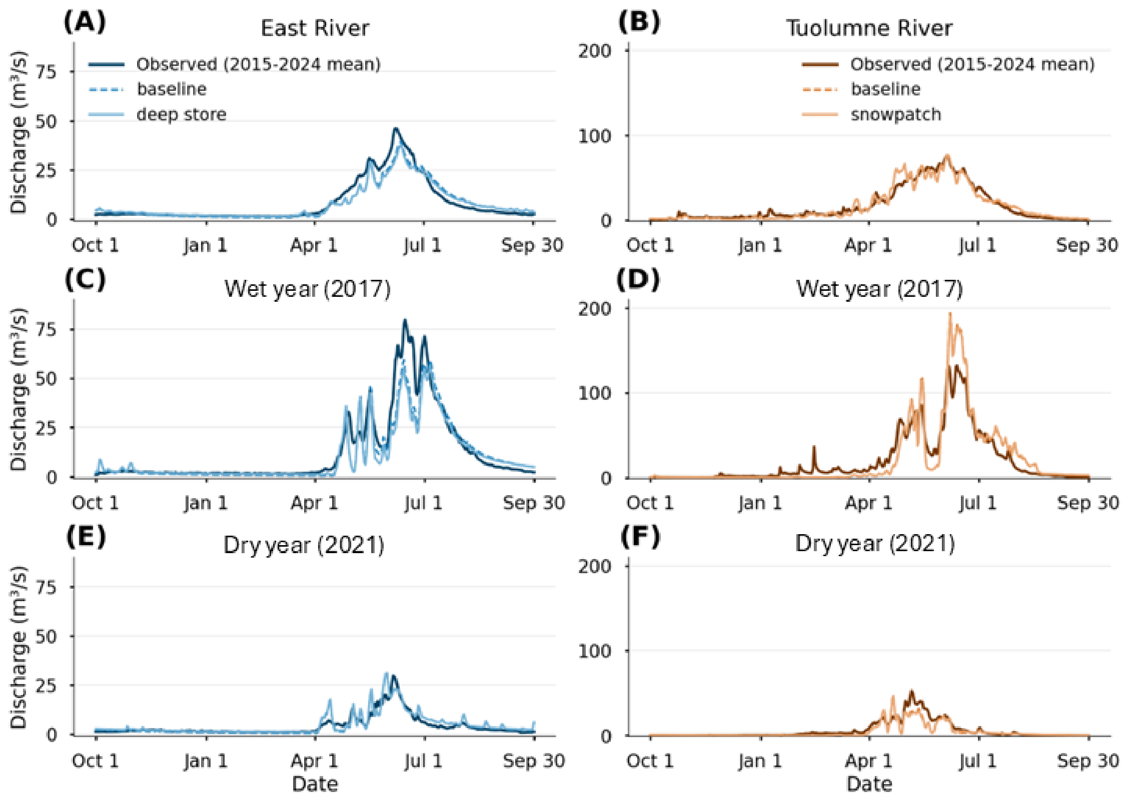

Figure 4.5: (A) East River mean WY hydrograph for the evaluation period (2015-2024) showing observed (solid, dark blue) and modeled (baseline = dashed, light blue; deep store = solid, light blue) streamflow. (B) Same as in (A) but for the Tuolumne.

Table 4.3: Calibration and evaluation metrics for baseline and store-added configurations. Calibration metrics correspond to the objectives optimized during calibration; evaluation metrics were not optimized.

		Calibration metrics			Evaluation metrics		
Period	Model	Mean abs. monthly bias (%)	COM bias (days)	Low-flow volume bias (%)	Q_{wy} bias (%)	$Q_{7,max}$ bias (%)	$Q_{7,min}$ bias (%)
East River							
Calibration	Baseline	10	1	6	3	11	0
	Deep store	10	0	0	−7	16	−6
Evaluation	Baseline	9	7	14	−5	−2	3
	Deep store	9	5	12	−7	−5	9
Tuolumne River							
Calibration	Baseline	13	1	17	4	17	9
	Snowpatch	12	1	19	4	15	9
Evaluation	Baseline	12	−7	9	−1	−15	6
	Snowpatch	13	−6	10	−1	−18	6

Table 4.4: Low-flow contributions from the HRUs where a long-term store was added, compared against the same HRUs in the baseline configuration. Target HRUs are the valley-bottom HRUs in the East River and the high-elevation, north-facing HRU in the Tuolumne. Dry and wet years are defined as the top and bottom quartile of water year streamflow totals.

	Target HRUs (% basin	Change in low-flow contribution vs. baseline (%)		Share of basin-wide low flows (%)	
Model	area)	Dry years	Wet years	Dry years	Wet years
East River					
Deep store	31	+40	−13	30	28
Baseline	31	—	—	19	22
Tuolumne River					
Snowpatch	1	+62	+15	14	7
Baseline	1	—	—	1	0

The effect of the deep store on low flows in the East River was most pronounced during dry years: within the HRUs where it was applied, the added storage compartment increased late season flows by 40% during dry years but decreased them by 13% during wet years, relative to the baseline (Table 4.4). This is consistent with groundwater buffering behavior.

In the Tuolumne, differences between the baseline and snowpatch models were muted across all skill metrics (Table 4.3, Table 4.8). However, the snowpatch contributed substantially more to late-season low flows in dry years: it accounted for 14% of basin-wide August through October runoff, compared to only 1% from the same HRUs in the baseline model (Table 4.4).

4.3.2 ET and SWE evaluation

Both modeled and observed ET generally decreased with elevation in both basins, demonstrating the transition from lower elevation vegetated terrain to barren high-elevation terrain. The baseline and long-term storage models showed no discernible difference between one another at all elevations (Figure 4.6A, B). In the East River, simulated ET was consistently lower than OpenET, though OpenET has been found to be biased high in forested and riparian terrain [*Volk et al.*, 2024].

In the Tuolumne, the lowest elevation band showed the largest discrepancy of all bands, which may reflect error in either the model or the OpenET reference. In SUMMA, simulated ET is highly sensitive to land cover type, so an inappropriate land class assignment could meaningfully bias ET. Conversely, OpenET's estimate may be affected by the deep canyon located in this elevation band, a terrain feature that could increase satellite retrieval error. To further evaluate these error magnitudes, we compared the baseline, deep store, and OpenET monthly ET totals to eddy covariance flux measurements taken in the East River's valley bottom between November 2022 and May 2023 (Figure 4.6C). Though both SUMMA models and OpenET were higher than the point observations, the SUMMA models better matched the point observation in both the magnitude and pattern of monthly ET.

In both basins, the models closely captured the elevation distribution and volume of ASO

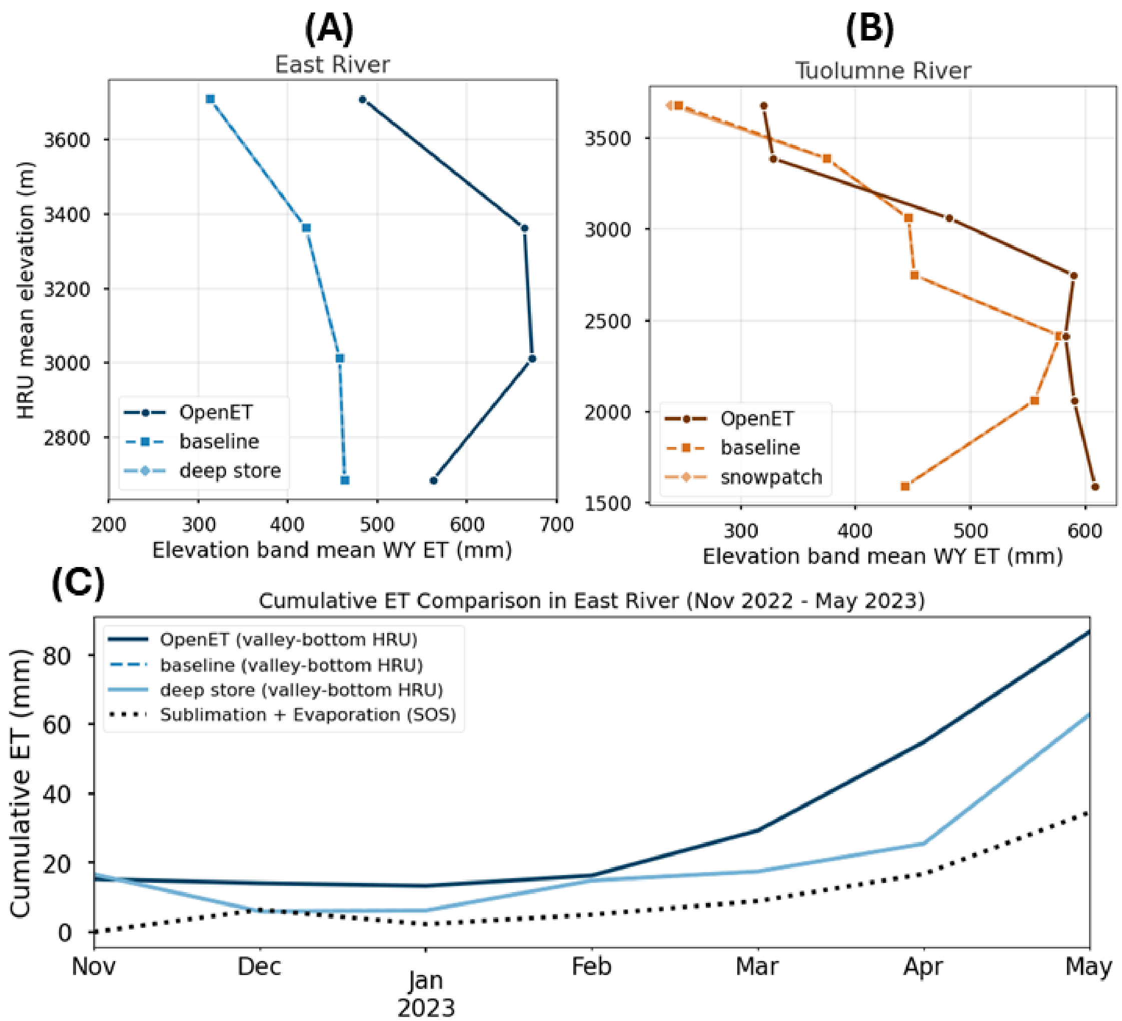

Figure 4.6: Comparisons of total WY ET from SUMMA simulations (red) and OpenET (black) averaged across elevation bands in the (A) East and (B) Tuolumne basins averaged between WY 2015-2024. (C) Monthly cumulative ET for the November 2022 – May 2023 period from eddy covariance observations in the East River valley bottom (black, dashed) compared against ET from the HRU containing the flux tower for SUMMA (red, dotted), and OpenET (black, solid).

SWE (Figure 4.7A, B). Basin-wide SWE absolute biases on these flight dates were 5.9% in the East and 6.2% in the Tuolumne. In the East, SWE from the baseline and deep store models were identical because initialized SWE was unchanged. This was also the case in the Tuolumne, except for the highest elevation band, where we placed the snowpatches (Figure 4.7B). Although the stark difference at this band was expected, ASO may underestimate SWE over these perennial snowpatches because lidar-derived SWE retrievals require a bare ground reference to calculate SWE depth. Nevertheless, due to the small area of the snowpatch HRUs, this added snow had little effect on the basin-wide SWE bias, which increased by only 1.9% in the snowpatch model.

4.3.3 Observed and modeled streamflow lagged autocorrelation

Significant lag-1 autocorrelation (ρ_1) in $Q_{7,min}$ was present in both the observed record ($\rho_1 =$ +0.26 in the East, +0.27 in the Tuolumne) and the long-term store models ($\rho_1 = +0.31$ in the East-deep store, +0.29 in the Tuolumne-snowpatch), but not in the baseline models, which stayed inside the 90% confidence envelope at every lag (Figure 4.8). The East's observations were also significant at lag 2 ($\rho_2 = +0.26$).

After the first lag, ρ declined in both the models and observations, flipping from positive to negative by the third or fourth lag (Figure 4.8). Notably, the observed lag-5 autocorrelations were the largest in magnitude among all lags and significant in both basins (ρ_5 = -0.49 in the East, -0.31 in the Tuolumne). The baseline models did not reproduce these strong negative autocorrelations at any lag, but the store-added models captured some of this behavior: the Tuolumne snowpatch model showed a significant autocorrelation at lag 4 (ρ_4 = -0.29), and the East deep store model showed moderate, though insignificant, negative autocorrelation at lag 5 (ρ_5 = -0.27). We address this result further in the Discussion.

The MRC analysis revealed that observed recession timescales (τ) were notably shorter in the Tuolumne (16 days) than in the East (28 days), where groundwater contributions are significant. In both baseline models, τ was roughly two-times longer than the observations. These longer timescales occur because each model contains only a single subsurface store,

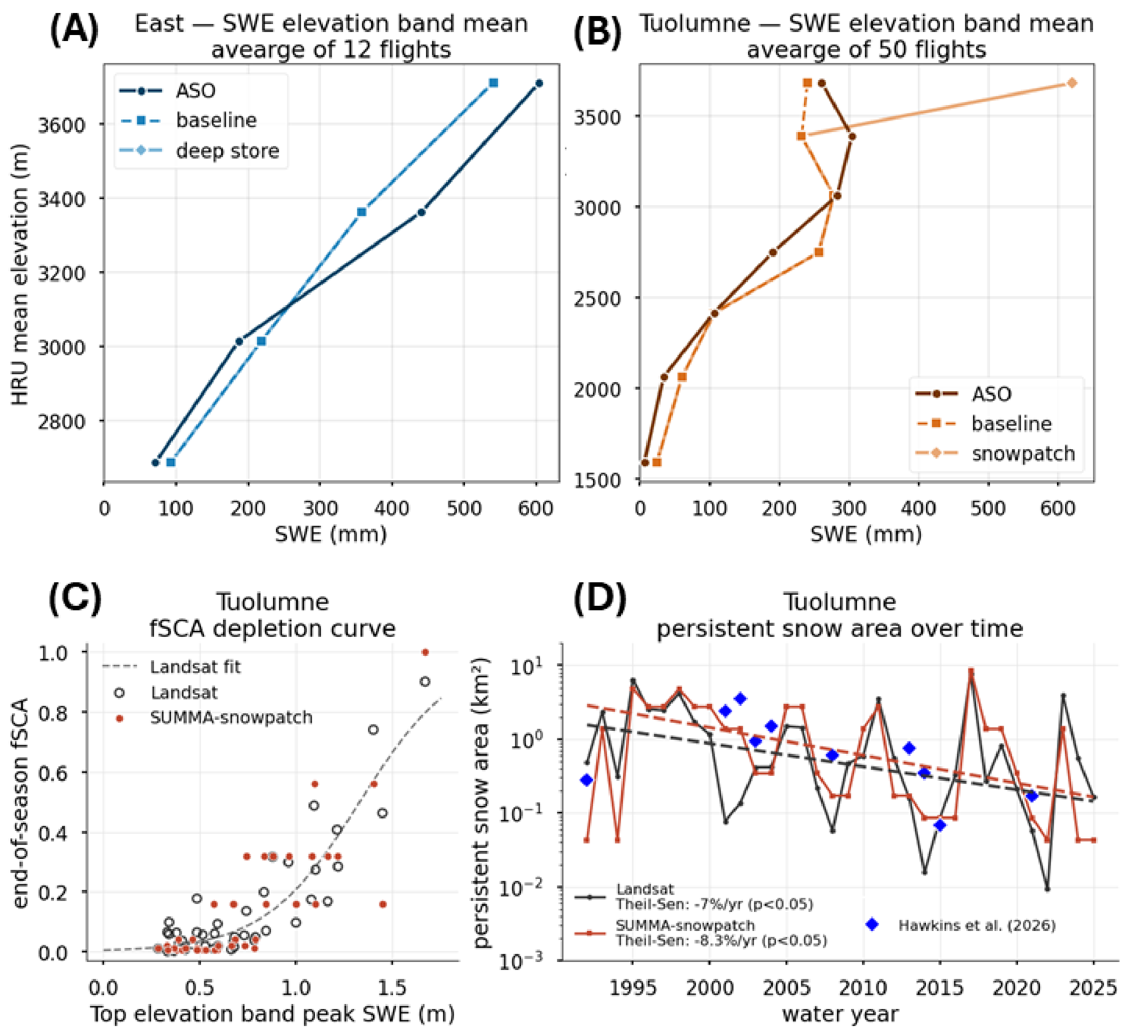

Figure 4.7: (A) Baseline (squares) and deep store (diamonds) elevation distribution of SWE compared to ASO (circles); (B) as in (A) but for the Tuolumne basin. The values reflect the average SWE value across all available ASO flight dates in each basin. (C) The Tuolumne fSCA depletion curve, which compares the minimum Aug-Oct Landsat (open circles) and SUMMA-snowpatch (orange dots) fSCA to the top elevation band's peak SWE each year. (D) Evolution of snowpatch area over time for Landsat (black line), SUMMA-snowpatch (orange line), and the *Hawkins et al.* [2026] measurements from Yosemite National Park (blue diamonds). The significant decreasing trends in snowpatch area (dashed lines) were calculated using the Theil-Sen estimator. Note the log-scale on the y-axis was used to emphasize years with low snowpatch area.

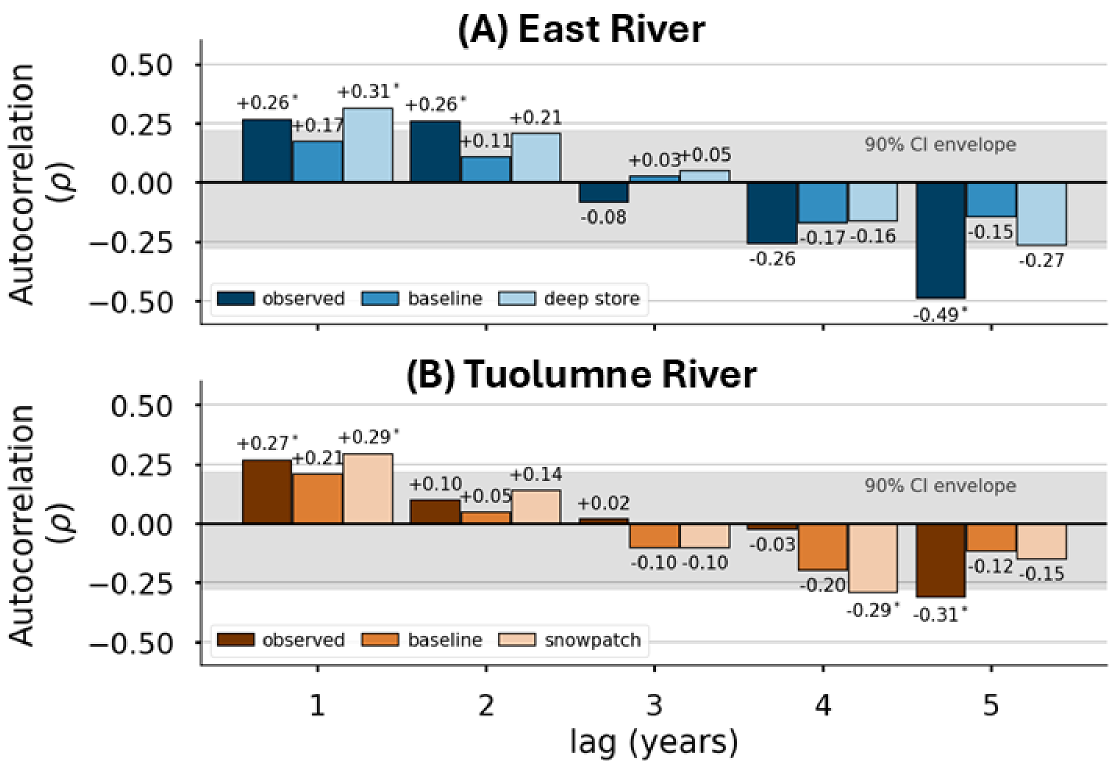

Figure 4.8: (A) East River lag-j year Q7,min autocorrelations from streamflow observations (dark blue), baseline model (medium blue), and the deep store model (light blue). (B) As in (A), but for the Tuolumne, with streamflow observations (maroon), baseline model (orange), and the deep store model (peach). Grey shaded area denoted the 90% confidence interval

which must blend faster soil drainage and interflow processes with relatively slower subsurface releases. However, adding an explicit long-term storage compartment did little to bring τ closer to observations, and instead pushed τ toward the store's underlying release rate (Figure 4.9). In the East, the addition of a slow, deep aquifer lowered the overall basin-mean drainage rate, moving τ further from the observations (Figure 4.9B). In the Tuolumne, additional meltwater from the added snowpatch sped up the basin-wide recession, though only marginally, given the store's small contribution to overall streamflow (Figure 4.9D).

4.3.4 Impacts of long-term storage compartments on low flows

The prior water year's storage state was positively correlated with the following year's $Q_{7,min}$ in both basins, but these relationships were only significant in the models with explicit long-term stores (Table 4.5, row 1). Because peak SWE partly controls how much storage increases in a given year, we also computed partial correlations to remove its influence on the storage state (Table 4.5, row 2). Doing so increased the correlations in all cases, indicating that antecedent peak SWE obscured the relationship between storage state and the following year's low flows.

Table 4.5: Relationships between end-of-year storage state (S) in year t and low flows in year $t+1$. The partial correlation removes the influence of peak SWE in year t from both variables. Elasticity (ε) is the slope of a log–log OLS regression, giving the percent change in low flow per 1% change in storage. Asterisks denote significance at $p < 0.10$.

	East River		Tuolumne River	
	Baseline	Deep store	Observed	Snowpatch
$\rho\ (S,\ Q_{7,min}^{t+1})$	+0.10	+0.47*	+0.22	+0.42*
Partial $\rho \mid \mathrm{SWE}_{peak}$	+0.21	+0.67*	+0.23	+0.53*
Elasticity, ε (% per %)	+0.03	+0.33*	+0.12	+0.44*

In the East River, antecedent end-of-WY aquifer storage state in the deep store model was significantly correlated with the following year's low flows (partial ρ =+0.67), while the

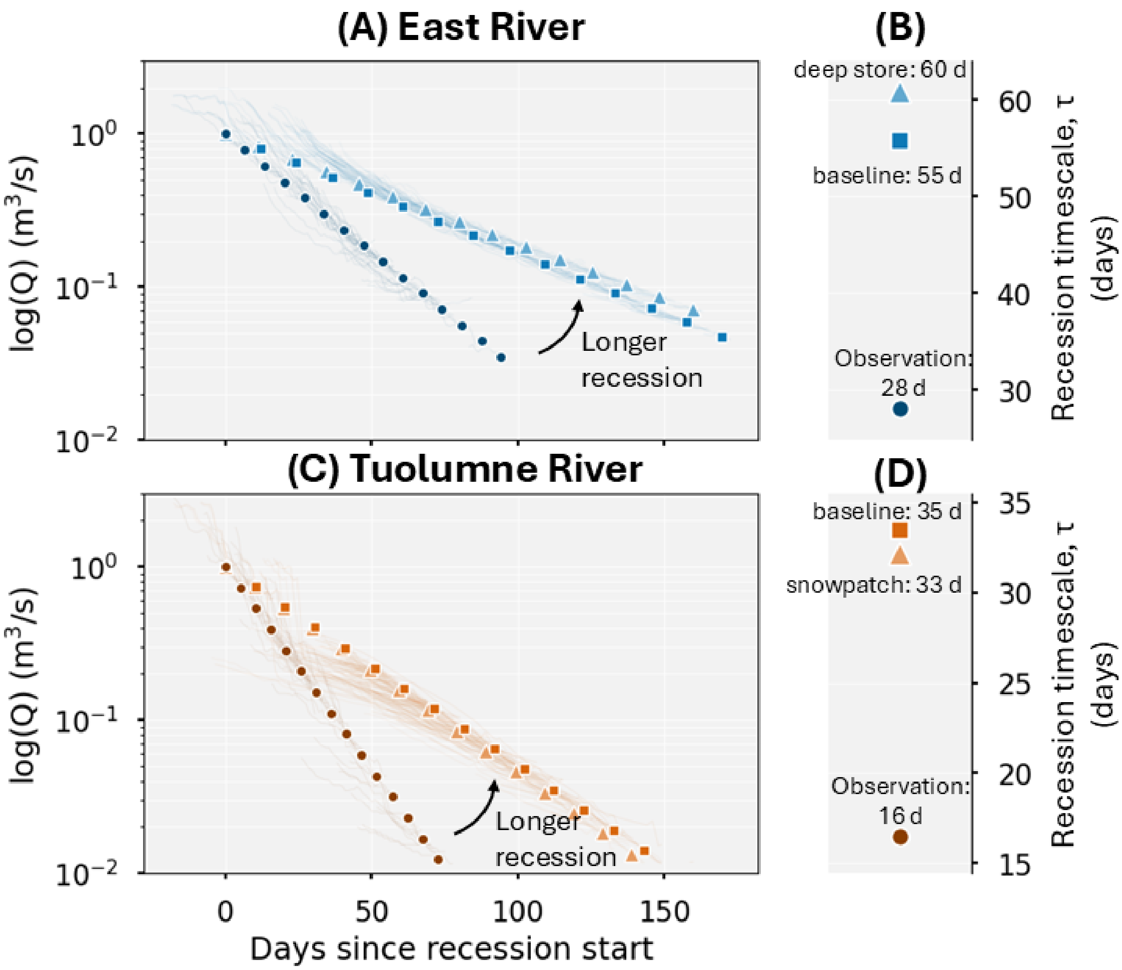

Figure 4.9: (A) Master recession curves (MRC) for the observations (circle), baseline model (square), and deep store model (triangle) in the East River. All recessions used in the construction of the MRC are lightly shaded behind the curves. (B) Recession timescales calculated from the MRC slopes. (C) As in (A) but for the Tuolumne. (D) As in (B) but for the Tuolumne.

baseline model showed a positive but insignificant relationship (partial ρ =+0.21). Low flows also showed a significant elasticity with antecedent storage in the deep store model: a 1% decrease in end-of-year aquifer storage was associated with a 0.33% decrease in the following year's low flow (ε = +0.33; Table 4.5, row 3).

In the Tuolumne, we were able to compare model snowpatch area against observations. Modeled antecedent snowpatch area was significantly correlated (partial ρ = +0.53) and showed significant elasticity (ε =+0.44) with modeled low flow. The observed record also showed a positive, but non-significant, relationship for both metrics (partial ρ = +0.23, ε = +0.12) over the available observational record (1993-2025). Both observed snowpatch area and low flows also declined over this period (-7.9% yr^{-1} and -1.3% yr^{-1}, respectively).

4.3.5 Modeled WY streamflow comparison to SWE-Q regression

In both basins, the SUMMA models modestly outperformed the SWE-Q regressions in reproducing observed Q_{wy} over the 1986-2025 period, with lower RMSE and comparable mean bias (Figure 4.10). However, when we isolated the driest years in the record, differences in performance widened considerably (subset boxes in Figure 4.10): bias in the SWE–Q regression reached +25% in the East and +37% in the Tuolumne. The SUMMA models reduced this to 9 to 11% in the East, and to -22 to -23% in the Tuolumne, where the models over-corrected but still reduced absolute bias. These low flow years included WY 2020 and WY 2021, the anomalously low-runoff years that occurred despite near-normal snowpack, as highlighted in the Introduction (Section 4.1).

Within the SUMMA configurations, the models with an added long-term storage compartment performed nearly identically to the baseline in both basins (Figure 4.10), indicating the store's contribution to interannual memory neither enhanced nor reduced Q_{wy} accuracy. As shown in Section 4.3.4, and discussed in Section 4.4.2, these storage models did impact modeled low flows, just not water year totals.

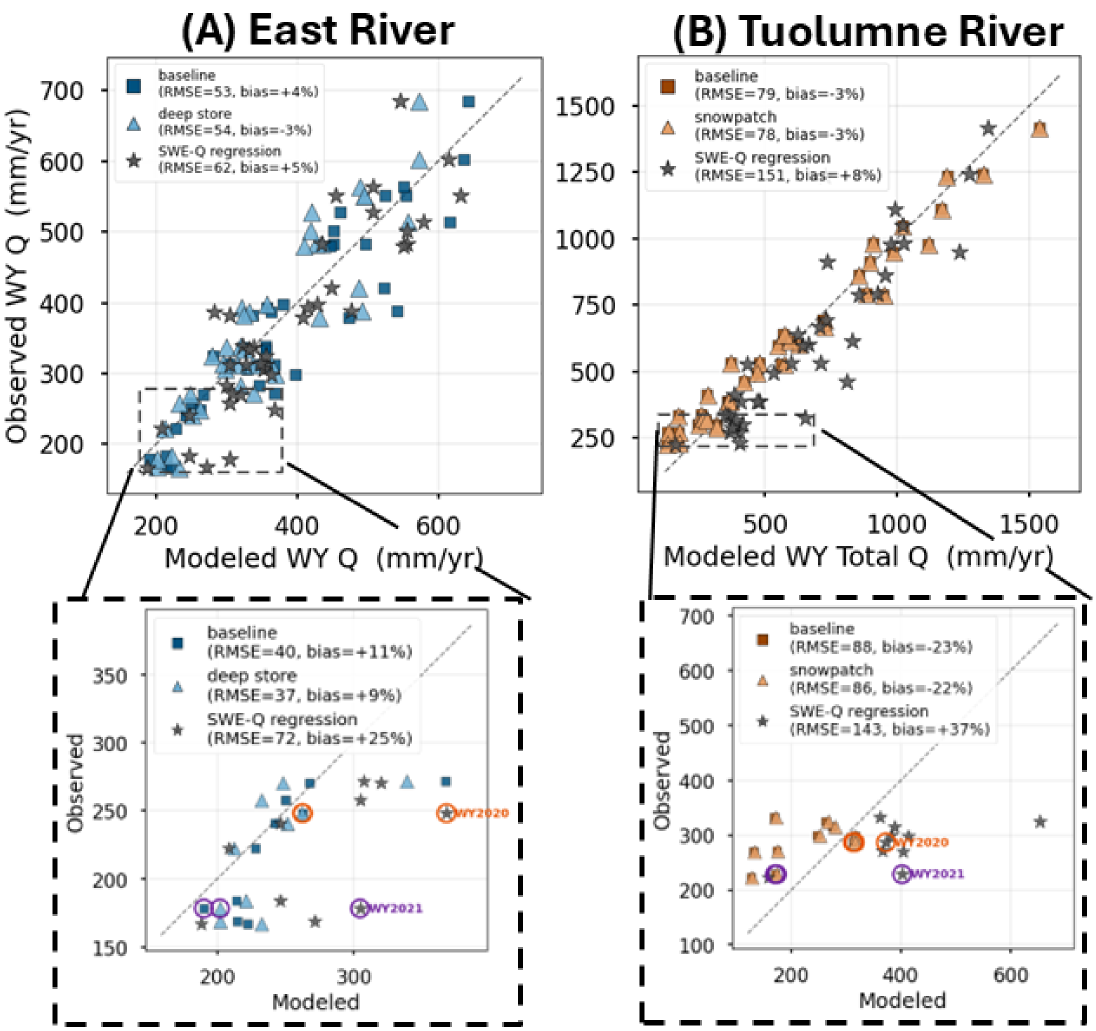

Figure 4.10: Modeled WY streamflow against observed for the analysis period (1986-2025) for the baseline (square), deep store/snowpatch (triangle), and SWE-Q regression (star) models for the (A) East River and (B) Tuolumne River. The dashed boxes and zoomed figures capture the bottom quartile of water year flows.

4.4 Discussion

4.4.1 How does adding a long-term storage compartment affect model performance?

In both the East and Tuolumne basins, overall model performance was largely unaffected by the addition of a long-term storage compartment, with only modest changes in water year streamflow and low-flow biases (Table 4.3; Figure 4.5). However, during dry years, this storage contributed significantly more water to low-flows than the baseline models (Table 4.4).

The two other dominant water balance components we evaluated, ET and snow, were also nearly identical between models (Figures 4.6 and 4.7). Additionally, the MRC analysis revealed that adding a storage compartment did not improve the representation of recession behavior, with both models substantially overestimating the observed characteristic recession timescales (Figure 4.9). The most notable difference between the models was in how they represented multi-year hydrologic memory.

Observed low flows showed significant autocorrelation with the following year's low flows, aligning with *Berghuijs et al.* [2025], who found 1-year memory in low flows to be a common feature of streams across the globe. This lag-1 memory signature was only reproduced in the models with the added long-term storage compartment (Figure 4.8). Combined with the similarity in model performance metrics, this result agrees with *Zhang et al.* [2026], who showed that adding a memory component to a model does not necessarily alter its performance.

In contrast, storage compartments added here failed to recreate the observed significant memory signature at longer lags. Observations showed significant autocorrelations at the 2- and 5-year lags in the East and the 5-year lag in the Tuolumne, which the store-added models did not reproduce, though their values were closer to observations than the baseline (Figure 4.8). In the East, more sophisticated subsurface models that explicitly represent bedrock circulation depth and geologic heterogeneity have captured longer memory signals in these settings [*Carroll et al.*, 2024; *Siirila-Woodburn et al.*, 2026]. In the Tuolumne, we represented only visible perennial snowfields, though the upper elevations also contain rock glaciers with unknown subsurface ice reserves that likely contribute to storage [*Hawkins et al.*, 2026; *Millar and Westfall*, 2008]. Streamflow response is sensitive to subsurface conceptualization [*Rapp et al.*, 2020], so the discrepancy at longer lags likely reflects the simplicity of our subsurface and long-term storage representations [*Fowler et al.*, 2020].

Even so, the 5-year signal may not reflect catchment storage at all. The 5-year rolling mean of standardized low-flow modeled and observed anomalies closely tracked the same rolling mean applied to observed peak SWE in both basins (Figure 4.13 in the Appendix Section 4.6.5), suggesting a periodic climate signal imprints itself on the record rather than a strictly storage-controlled process. *Wolf et al.* [2025] reported similar periodicity in low flows in UCRB headwater basins and suggested certain climate teleconnections could be responsible, though disentangling this signal from catchment memory would require a longer record than our 40-year analysis period and is beyond the scope of this study.

4.4.2 How does long-term storage state relate to low flows?

While identifying that memory is present within the streamflow record indicates a relationship between storage state and streamflow, these lag autocorrelations do not reveal the source of this memory by themselves. Directly relating these low flows to the storage features that we tested provides a clearer picture of how multi-year hydrologic memory is generated in these basins.

Similar to the low-flow memory results, end-of-WY long-term storage state carries information about the following year's low flows, but only models with an explicit storage compartment express it (Table 4.5). In the East, only the deep store model tied aquifer storage to the next year's low flow. The contrast in elasticities was sharper still. For the deep store model, a 1% change in antecedent storage yielded a 0.33% change in low flows, while the baseline model had minimal influence (Table 4.5). Though we were limited to a model intercomparison in the East, this result corroborates prior work which highlights the strong sensitivity of streamflow to storage state in the UCRB [*Brooks et al.*, 2025, 2021;

Siirila-Woodburn et al., 2026].

In the Tuolumne, we were able to compare the snowpatch model results to observations. The modeled snowpatch area was significantly related to the following year's low flows, and low flows were highly sensitive to changes in snowpatch area, while the observed snowpatch showed a weaker, insignificant relationship for both metrics (Table 4.5). While this insignificant relationship precludes attributing low flow changes directly to snowpatch area change, we note that both late-season low flow and snowpatch area declined at rates of -1.3% yr^{-1} and -7.9% yr^{-1}, respectively, over the overlapping observational record (1993-2024). The observed snowpatch trend was consistent with the -6.5% yr^{-1} in snowpatches of Yosemite National Park reported by *Hawkins et al.* [2026]. Applying the observed elasticity to the snowpatch trend predicts a low-flow decline of -1.0% yr^{-1}, close to the observed trend. Because the elasticity and both trends derive from the same short record, this is a consistency check rather than independent confirmation. Still, it indicates that snowpatch loss is of sufficient magnitude to account for the observed decline in late-season flow.

4.4.3 When is including a long-term storage compartment most important?

As expected, the April 1 SWE–Q relationship is strong in the average year in both basins (Figure 4.10), supporting its continued use as a water supply predictor [*Pagano et al.*, 2004]. The two basins differ, however, in how tightly they follow it. Greater scatter in the East River indicates that terms other than snowpack contribute more to interannual variability there, consistent with a larger groundwater contribution to streamflow. The tighter relationship in the Tuolumne is consistent with a system where surface processes dominate the annual water balance while subsurface storage plays a smaller role. This contrast supports the long-term storage compartment representations we selected for each basin.

The SWE-Q relationship weakens during both the wettest and driest years, but for different reasons. In wet years, precipitation measurement uncertainty likely dominates [*Bárdossy et al.*, 2022; *Hogan et al.*, 2026a], and our models performed no better than the regression. While prediction errors in these years still impact reservoir and hydropower management, they carry lower consequences for water supply decisions. In dry years, the models outperformed the regression: absolute bias decreased from 25% to 9% in the East and from 37% to 23% in the Tuolumne. A SWE-Q regression lumps all non-snowpack terms into its residual, so in years when SWE does not dominate, ET and storage change surface as significant sources of error. A model that solves each term individually is better positioned in exactly those years. As snowpack declines and becomes more variable across the western US, the conditions under which SWE alone is a sufficient predictor will occur less often [*Livneh and Badger*, 2020; *Modi et al.*, 2022; *Musselman et al.*, 2021; *Siirila-Woodburn et al.*, 2021].

That said, adding a long-term storage compartment did not improve water-year volume accuracy relative to the baseline models (Figure 4.10). Thus, the memory signal we recovered is not a dominant source of "missing water" when hydrologic anomalies are large. Late-season low flows are a small fraction of annual volume, so relying solely on a storage mechanism cannot close the observed error gap in anomalous years. Even so, we found that these stores do affect the late summer and autumn flows when water is scarcest, with potential consequences for aquatic habitat and late-season water allocations that are not captured by annual volume metrics [*Leathers et al.*, 2024; *Lundquist and Roche*, 2009].

In any single year, how a storage compartment is represented may not matter, but over longer periods, they may respond to sustained conditions very differently. For example, a snowpatch responds rapidly to both extremely wet and dry years, recovering quickly in high-accumulation years and declining rapidly in low ones (Figure 4.7C), meaning it can be lost entirely under a sustained dry period. It is also directly observable, with extent measurable from high-resolution satellite imagery [*Hawkins et al.*, 2026]. A deep aquifer instead integrates conditions over longer periods, drawing down and refilling slowly [*Carroll et al.*, 2024; *Siirila-Woodburn et al.*, 2026]. Distributed observations of aquifer conditions are rarer, and most analyses rely on coarse satellite products such as GRACE, which are accurate at large scales but cannot resolve the spatial heterogeneity of mountain systems [*Liaqat et al.*, 2026; *Opie et al.*, 2020; *Wang et al.*, 2026]. Representing long-term storage across basins with different physiographic characteristics therefore requires correspondingly different conceptual models, and different observational strategies to constrain them.

Signals of declining long-term storage have received increased attention [*Carroll et al.*, 2024; *Siirila-Woodburn et al.*, 2026; *Hawkins et al.*, 2026; *Thomas and Famiglietti*, 2019; *Opie et al.*, 2020], and our results suggest these declines may become increasingly visible in the streamflow record as they accumulate. In the Tuolumne, observed low flows and snowpatch area declined together, while peak SWE showed no significant trend, a divergence that would be difficult to interpret without accounting for storage. Where direct storage observations are unavailable, prior-year conditions offer a practical alternative: last year's low flow carries information about this year's late-season flow that snowpack alone does not.

4.4.4 Study limitations

We highlight four main limitations of our study. First, our representations of long-term storage remain conceptual. As noted above, prior studies in the East River have used models that track multi-year storage and groundwater age more directly [*Carroll et al.*, 2025, 2024; *Siirila-Woodburn et al.*, 2026]. But our models were constrained by what SUMMA can express: because the aquifer and interflow subsurface schemes in SUMMA are mutually exclusive, the East configuration omits the lateral interflow between HRUs, an important water transfer pathway there [*Carroll et al.*, 2019]. However, our goal was to illustrate how these processes can be represented in a simpler modeling framework, not to replicate those results.

Second, the snowpatch implementation approximates fSCA by discretizing a single high-elevation HRU into depth bins. This reproduces the observed fSCA depletion relationship by construction (Figure 4.7C), so agreement between modeled and observed fSCA at a given SWE is not independent validation; nevertheless, the end-of-WY fSCA timeseries comparison is independent (Figure 4.7D). Additionally, the limited number of depth bins we used (10) imposed a static minimum snow-covered area that would persist until the depth bin fully melted out, limiting the representation of area change during extreme drought.

Third, we ignore certain processes that may impact the representation of storage. We do not account for summer precipitation, which can be significant in the East River [*Carroll et al.*, 2020]. Its influence on soil moisture and late-season low flows could dilute a long-term memory signature. We also do not evaluate influences from soil moisture more broadly, although prior work has shown soil moisture memory is largely contained within a single water year [*Ghimire et al.*, 2026; *Alvarez-Garreton et al.*, 2021], so it falls outside the multi-year signal we targeted here, but anomalous soil moisture remains a plausible secondary contributor we did not constrain. Finally, we do not represent subsurface ice storage [*Millar and Westfall*, 2008], which may impart a memory signature not captured here.

Fourth, our analysis rests on a single memory metric derived from observed streamflow over a 40-year record. Both constraints limit statistical power at longer lags. For example, with roughly 35 independent pairs at lag 5, we cannot distinguish a genuine multi-year storage signal from climate periodicity or sampling bias. A longer record, or an independent storage observation in the East, could better address these inferences.

In spite of these limitations, this work represents a first attempt to evaluate different sources of multi-year storage in a deliberately simple modeling framework in mountains. Where the quantity of interest is low flow, particularly in dry years, investing in a basin-appropriate representation of long-term storage is worthwhile. For total water-year volume, a simpler model appears sufficient, and still outperforms a SWE–streamflow regression.

4.5 Conclusion

This study set out to determine how adding a long-term water storage compartment to a semi-distributed, physically based hydrologic model affects its representation of streamflow memory in two physiographically distinct snow-dominated basins in the western United States, the East and Tuolumne Rivers. We applied a storage compartment that matches our conceptual understanding of each basin – perennial snowpatches in the Tuolumne and a deep, slowly draining aquifer in the East – and characterized memory using the lag autocorrelation of annual 7-day August–October low flows.

Three findings emerged. First, the baseline models could not replicate the observed 1-year lag memory in either basin, while the store-added configurations could. However, neither model reproduced the observed 5-year lag memory signal, indicating that our conceptual representations still miss features contributing to multi-year memory, though periodic variability in precipitation may partly explain this longer cycle.

Second, end-of-year storage state in both store-added models was positively correlated with the following year's low flows, evidence of buffering between prior-year storage and subsequent streamflow. In the Tuolumne, Landsat-derived snowpatch area showed the same 1-year lag correlation with observed low flows. Observed snowpatch area and observed low flows declined together at -7.9% yr^{-1} and -1.3% yr^{-1} respectively, consistent with the modeled sensitivity, but the limited observational record warrants caution in this direct interpretation. Constraining the observed snowpatch–low flow relationship more tightly would benefit from using the nested streamflow observations in the Yosemite Hydroclimate Network [*Lundquist et al.*, 2016], though measurement error in small headwater catchments is a concern for low flows.

Third, adding a storage component had minimal effect on water-year volume accuracy relative to the baseline, yet both model types substantially outperformed the SWE–Q regression in the driest years, reducing absolute bias by 64% in the East and 38% in the Tuolumne relative to the regression. This underscores the value of representing all water balance components when snowpack-streamflow errors are highest.

Several future research opportunities could expand on this work. Applying each storage compartment type to the opposite basin would test whether a configuration tuned to one storage mechanism degrades where physiography differs. Comparing across a range of subsurface complexity would identify the minimum representation needed to capture relationships demonstrated with more sophisticated models. Extended drought and recovery experiments have been performed for groundwater but not for perennial snow, and a direct comparison of their response would clarify how differently these storage compartments behave during extreme conditions.

Because storage memory was not a dominant source of the water-year volume errors we examined, it is unlikely to explain most hydrologic anomalies. But as temperatures continue to rise and droughts persist, both stores will draw down, and in the Tuolumne, snowpatches may eventually disappear entirely. Models without these stores cannot represent those changes and will instead attribute the resulting decline in late-season flow to other processes, obscuring what truly happened. Determining which long-term storage mechanism dominates in a basin, and representing it explicitly, allows for multi-year changes to be physically accounted for.

Data Availability

Model code and configuration files are publicly available on GitHub (`https://github.com/dlhogan97/CONFLUENCE-uwmtnhydro`), with repository structure modeled after the CWARHM workflow [*Knoben et al.*, 2022] (`https://github.com/CH-Earth/CWARHM`). SUMMA source code is available vis `https://github.com/CH-Earth/summa`. Simulated output will be archived on ESS-DIVE upon publication. Meteorological forcing data were obtained from PRISM (`https://prism.oregonstate.edu`) and ERA5-Land [*Muñoz-Sabater et al.*, 2021] (`https://doi.org/10.24381/cds.e2161bac`). Land cover was sourced from the 2021 NLCD (`https://www.mrlc.gov`) [*Dewitz and Survey*, 2021] and soil data from SoilGrids (`https://soilgrids.org`) [*Poggio et al.*, 2021]. Basin topography was derived from the USGS 3DEP 30-m DEM (`https://www.usgs.gov/3d-elevation-program`) [*Survey*, 2022]. Lidar SWE measurements and related data collected and made available by Airborne Snow Observatories, Inc [*Painter et al.*, 2016]. East River streamflow was obtained from USGS gauge 09112500 (`https://waterdata.usgs.gov`). Reconstructed Tuolumne streamflow was provided by Bruce McGurk (personal communication; see also *Lundquist et al.* [2016]; *Henn et al.* [2018]). SNOTEL observations are available through the USDA Natural Resources Conservation Service (`https://www.nrcs.usda.gov/wps/portal/wcc/home`). OpenET data were accessed via `https://openetdata.org` [*Melton et al.*, 2022; *Volk et al.*, 2024]. Observed soil properties from the East River were taken from *Falco et al.* [2024] and *Wainwright et al.* [2024].

4.6 Appendix

4.6.1 Landsat fSCA estimate methodology

Landsat imagery came from the Collection 2 Level-2 surface-reflectance archive, accessed through the Microsoft Planetary Computer STAC catalog, which provides a continuous 30 m record from Landsat 5, 7, 8, and 9 beginning in 1985 [*Earth Resources Observation and Science (EROS) Center*, 2020a, b, c]. We retained scenes acquired between 1 August and 1 October to focus on capturing perennial snow. Scenes with more than 35% cloud cover were excluded. For the remaining scenes, we required that at least 95% of the highest elevation band fall within the scene footprint and that at least 70% of that band remain cloud-free after masking cloud, cirrus, and cloud shadow using the Landsat QA band. For each remaining scene we computed the Normalized Difference Snow Index from the green and shortwave-infrared (1.6 μm) bands and classified a pixel as snow where NDSI exceeded 0.40 and near-infrared reflectance exceeded 0.11 [*Hall et al.*, 1995]. Fractional snow-covered area (fSCA) is then the share of valid pixels classified as snow. Multiplying it by the high elevation band's area gives the snow area in km^2. We note that this is a conservative estimate, as we did not account for mixed pixels. Each water year is represented by the minimum snow area across its usable scenes, giving an end-of-summer value, and although the record extends back to 1985, we restricted our analysis to water years 1993–2025, when scene availability was consistently high enough for the annual minimum to be reliable. We independently evaluated the resulting Landsat fSCA estimates with ASO fSCA data from the same elevation band. To do so, we gathered all available filtered Landsat images within 6-days of a Tuolumne ASO flight date (n=38) to compare fSCA values (Figure 4.11). We saw good agreement between these two products (R^2 = 0.81, RMSE = 0.13). The small number of late season flights meant lower fSCA values were more uncertain. ASO generally measured higher fSCA values during these periods, but the strong relationship overall gave us confidence in this method.

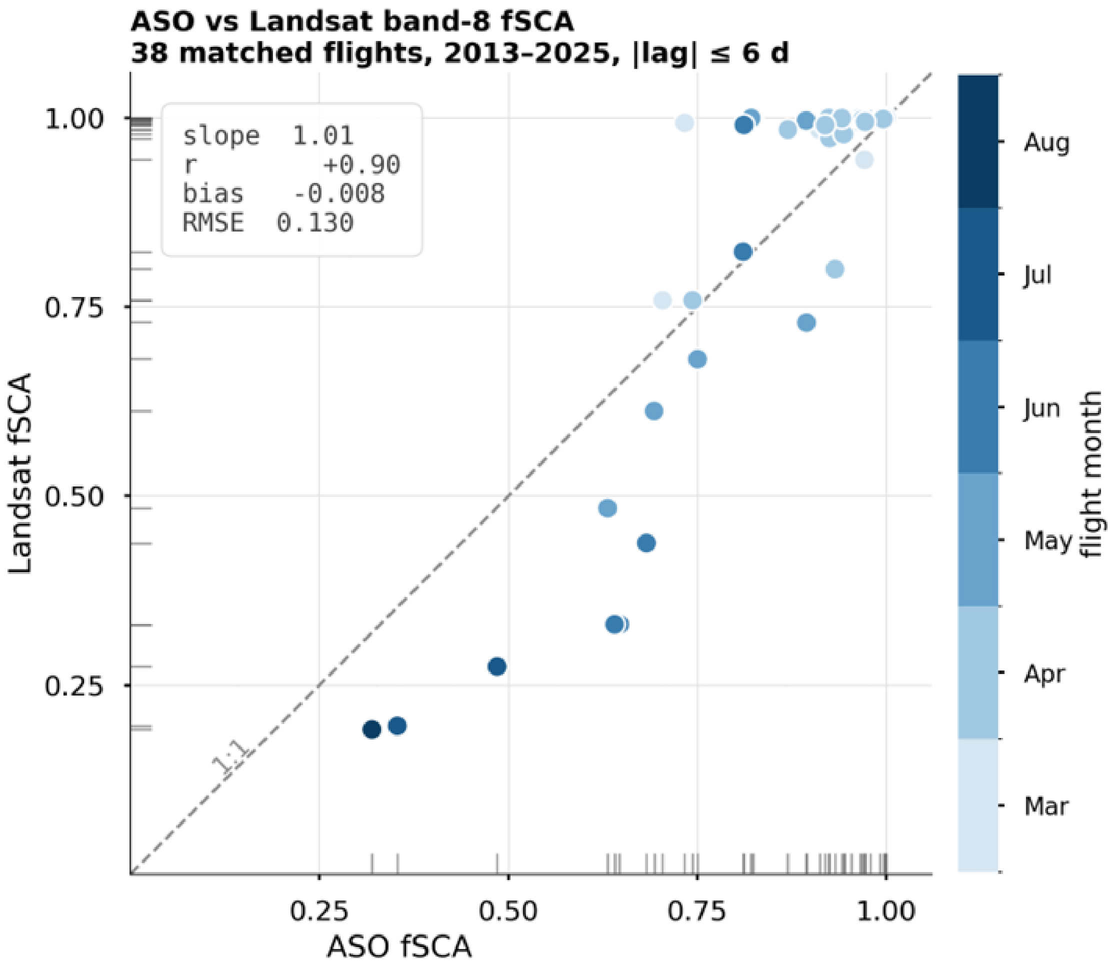

Figure 4.11: Scatterplot of ASO fSCA measurements and Landsat fSCA estimates between 2013 and 2025. Only filtered Landsat images taken within 6-days of an ASO flight were used. This resulted in 38 usable data points. Scatter points are colored by the month each flight occurred. The ticks on each axis indicate the location of each data point within each product.

4.6.2 Model Forcing Derivations

Pressure was calculated using HRU elevations, air temperature, and an assumed constant lapse rate of 0.0065 K m^{-1} [*Chegwidden et al.*, 2019] (Eq. 4.7).

$$p = 101.325 \times \exp\left(\frac{-zg}{8.3143(T + 0.5z\Gamma)}\right), \tag{4.7}$$

where p is pressure in hPa, T is air temperature (K), g is the acceleration due to gravity, z is elevation above mean sea level, and Γ is the lapse rate (-0.0065 K m^{-1}).

We estimated specific humidity (q) by calculating saturation vapor pressure (e_s) using air temperature (T; °C) (Eq. 4.8) and subtracting PRISM's vapor pressure deficit (VPD) to estimate vapor pressure (e_o) (eq. 4.9), from which we calculate q (eq. 4.10).

$$e_s = 6.1078(17.27T + 237.3) \tag{4.8}$$

$$e_o = es - VPD \tag{4.9}$$

$$q = \frac{0.622e_o}{p - 0.378e_o} \tag{4.10}$$

Longwave radiation was estimated using the Dilley & O'Brien method [*Dilley and O'brien*] (eq. 4.11), which has been shown to perform well in mountainous regions [*Currier et al.*, 2017; *Alden et al.*, 2026].

$$LW = 59.38 + 113.7(\frac{T_a}{273.16})^6 + 96.96\sqrt{\frac{465 \cdot e_0}{2.5 \cdot T_a}} \tag{4.11}$$

4.6.3 Streamflow Evaluation Metrics

Centroid timing error:

Centroid timing is calculated as

$$t_Q = \frac{\sum t_i Q_i}{\sum Q_i}, \tag{4.12}$$

where t_i are the days since 1 October and Q_i is daily flow. We report the mean difference between the modeled and observed values across the evaluation period ($n = 10$).

$$E_{\text{com}} = \frac{\sum \left(t_{Q,i}^{\text{sim}} - t_{Q,i}^{\text{obs}}\right)}{n} \tag{4.13}$$

Sum of absolute monthly volume errors:

$$E_{\text{mon}} = \frac{\sum_m \left|Q_m^{\text{sim}} - Q_m^{\text{obs}}\right|}{\sum_m Q_m^{\text{obs}}} \tag{4.14}$$

where Q_m^{sim} and Q_m^{obs} are the simulated and observed volume in month m, summed over all years. Only months with more than 4 mm bias are included in the numerator to avoid large biases from low-flow months.

Low-flow (August–October) volume error:

$$E_{\text{aug-oct}} = \frac{\sum_{m=8}^{10} \left|Q_m^{\text{sim}} - Q_m^{\text{obs}}\right|}{\sum_{m=8}^{10} Q_m^{\text{obs}}} \tag{4.15}$$

4.6.4 Matching strip method

Following recommendations from *Nathan and McMahon* [1990], we filtered recessions before assembling the master recession curve. We identified contiguous falling limbs and excluded segments affected by rain or snowmelt inputs, data gaps, or quickflow contamination. The first five days after each streamflow increase were dropped to remove the quickflow-dominated early limb, and the seasonal window began 14 days after each basin's representative SNOTEL SWE melted out in each water year to exclude the rising freshet limb. Segments shorter than five days were discarded. These screens reflect the assumption underlying the matching strip method: each segment must represent pure drainage. Rain or melt during a segment adds storage mid-recession and biases the recession constant low, and in snow-dominated basins this contamination is worst during the melt season, motivating our filter of the melt season. Restricting the fit to the tail avoids the early limb, which reflects near-surface drainage rather than the slow store the master curve is meant to characterize. Each surviving segment was fit with a linear regression over its late-time tail, defined as the days below 0.67× the segment's mean flow (Figure 4.2), and retained only if the fit gave $r^2 \geq 0.30$ and a positive recession constant. Conforming segments were ranked longest-first and collapsed onto a single master curve by overlaying each curve onto a semi-log plot and a master recession curve is fit to the tails of the resulting curves [*Nathan and McMahon*, 1990; *Toebes and Strang*, 1964].

4.6.5 Appendix Figures

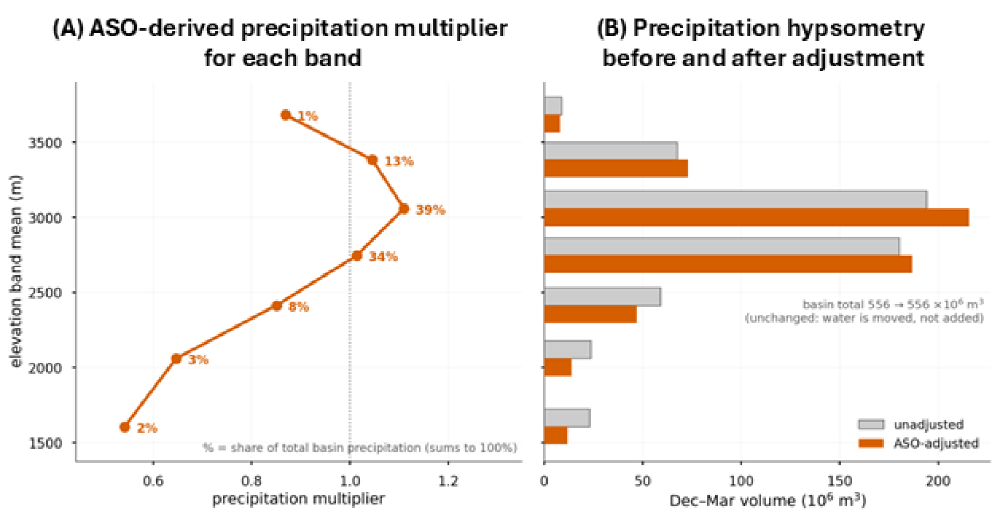

Figure 4.12: ASO-based precipitation redistribution for the Tuolumne. (A) Multiplier applied to each elevation band, derived from mean ASO SWE and blended with the original field; labeled percentages are each band's share of total basin precipitation after adjustment. (B) Mean December–March precipitation volume per band, before and after adjustment. The basin total is unchanged as the correction only moves water between elevations.

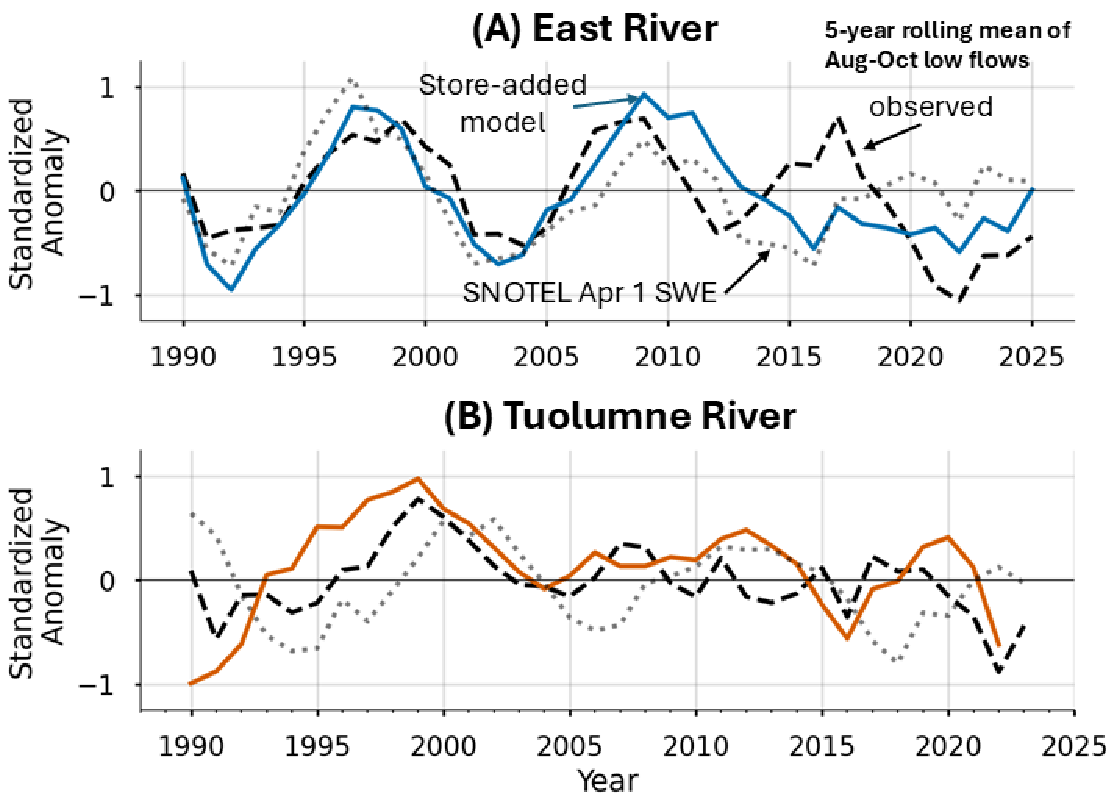

Figure 4.13: 5-year rolling mean of standard anomalies between 1986-2025 in observed low-flows (black, dashed), store-added modeled low-flows (blue/orange, dashed) streamflow and SNOTEL April 1 SWE (grey, dotted) for the (A) East and (B) Tuolumne River.

4.6.6 Appendix Tables

Table 4.6: Hydrological response unit (HRU) attributes for the elevation–aspect distributed configurations. HRUs are formed by intersecting equal-interval elevation bands with four aspect quadrants (NE, 0–90°; SE, 90–180°; SW, 180–270°; NW, 270–360°), plus a flat class assigned where the mean slope is below 5°. Elevation and aspect are area-weighted HRU means; aspect is measured clockwise from north. Soil classes follow the ROSETTA texture classification and land classes follow the NLCD.

	Elev.	Area	Area	Elevation	Aspect			
HRU	band	(km^2)	(%)	(m a.s.l.)	(°)	Class	Soil class	Land class
East River — 18 HRUs, 748.3 km^2 total, 372 m elevation bands								
1	4	16.44	2.2	3683	49	NE	Loam	Barren or sparsely vegetated
2	4	21.57	2.9	3695	138	SE	Loam	Open shrubland
3	4	33.93	4.5	3735	220	SW	Loam	Barren or sparsely vegetated
4	4	11.90	1.6	3698	307	NW	Loam	Barren or sparsely vegetated
5	3	53.30	7.1	3357	48	NE	Loam	Evergreen needleleaf forest
6	3	56.54	7.6	3372	137	SE	Loam	Open shrubland
7	3	71.21	9.5	3363	223	SW	Loam	Evergreen needleleaf forest
8	3	52.86	7.1	3353	312	NW	Loam	Evergreen needleleaf forest
9	2	13.02	1.7	2989	159	Flat	Loam	Open shrubland
10	2	72.31	9.7	3011	46	NE	Loam	Evergreen needleleaf forest
11	2	59.66	8.0	3012	139	SE	Loam	Open shrubland
12	2	78.84	10.5	3014	222	SW	Loam	Deciduous broadleaf forest
13	2	54.23	7.2	3016	318	NW	Loam	Evergreen needleleaf forest
14	1	52.04	7.0	2650	169	Flat	Loam	Permanent wetland
15	1	31.60	4.2	2710	46	NE	Loam	Open shrubland
16	1	19.75	2.6	2705	135	SE	Loam	Open shrubland
17	1	31.91	4.3	2700	226	SW	Loam	Open shrubland
18	1	17.14	2.3	2698	313	NW	Loam	Open shrubland
Tuolumne River — 28 HRUs, 774.5 km^2 total, 347 m elevation bands								
1	8	4.37	0.6	3686	226	SW	Sandy loam	Barren or sparsely vegetated
2	8	8.61	1.1	3679	348	NW	Sandy loam	Barren or sparsely vegetated

Continued on next page

Table 4.6 continued from previous page

HRU	Elev. band	Area (km^2)	Area (%)	Elevation (m a.s.l.)	Aspect (°)	Aspect Class	Soil class	Land class
3	7	4.47	0.6	3350	177	Flat	Sandy loam	Open shrubland
4	7	15.89	2.1	3396	42	NE	Sandy loam	Barren or sparsely vegetated
5	7	17.70	2.3	3387	136	SE	Sandy loam	Barren or sparsely vegetated
6	7	33.12	4.3	3391	227	SW	Sandy loam	Open shrubland
7	7	25.96	3.4	3389	315	NW	Sandy loam	Barren or sparsely vegetated
8	6	23.33	3.0	3040	218	Flat	Sandy loam	Open shrubland
9	6	39.04	5.0	3071	47	NE	Sandy loam	Open shrubland
10	6	64.76	8.4	3057	136	SE	Sandy loam	Open shrubland
11	6	68.23	8.8	3067	227	SW	Sandy loam	Open shrubland
12	6	75.22	9.7	3058	314	NW	Sandy loam	Evergreen needleleaf forest
13	5	29.36	3.8	2730	247	Flat	Sandy loam	Evergreen needleleaf forest
14	5	39.07	5.0	2749	42	NE	Sandy loam	Evergreen needleleaf forest
15	5	61.26	7.9	2745	139	SE	Sandy loam	Open shrubland
16	5	56.52	7.3	2758	221	SW	Sandy loam	Evergreen needleleaf forest
17	5	70.11	9.1	2747	314	NW	Sandy loam	Evergreen needleleaf forest
18	4	4.67	0.6	2409	253	Flat	Sandy loam	Evergreen needleleaf forest
19	4	12.47	1.6	2424	31	NE	Sandy loam	Open shrubland
20	4	19.85	2.6	2410	148	SE	Sandy loam	Open shrubland
21	4	17.34	2.2	2410	216	SW	Sandy loam	Open shrubland
22	4	22.58	2.9	2419	321	NW	Sandy loam	Evergreen needleleaf forest
23	3	4.60	0.6	2057	25	NE	Loamy sand	Evergreen needleleaf forest
24	3	9.22	1.2	2058	150	SE	Sandy loam	Open shrubland
25	3	6.65	0.9	2060	217	SW	Sandy loam	Open shrubland
26	3	9.69	1.3	2067	327	NW	Loamy sand	Evergreen needleleaf forest
27	2	17.94	2.3	1513	159	SE	Sandy loam	Open shrubland
28	2	12.47	1.6	1702	321	NW	Sandy loam	Evergreen needleleaf forest

Table 4.7: Constant model parameterizations

Decision name	Decision value	Decision description	Value description
`soilCatTbl`	`ROSETTA`	Soil category dataset	Merged ROSETTA table with STAS-RUC
`vegeParTbl`	`MODIFIED_IGBP_MO`	Vegetation category dataset	MODIS 20-category dataset
`soilStress`	`NoahType`	Soil moisture control on stomatal resistance	Threshold linear function of volumetric liquid water content
`stomResist`	`Jarvis`	Function for stomatal resistance	Ball–Berry
`num_method`	`itertive`	Numerical method	Iterative solver
`fDerivMeth`	`analytic`	Method for flux derivatives	Analytic derivatives
`LAI_method`	`specified`	Method to determine LAI and SAI	LAI/SAI computed from green vegetation fraction
`f_Richards`	`mixdform`	Form of Richards' equation	Mixed form
`hc_profile`	Varies by scheme	Hydraulic conductivity profile	Power-law profile or constant
`bcUpprTdyn`	`nrg_flux`	Upper boundary condition for thermodynamics	Energy flux
`bcLowrTdyn`	Varies by scheme	Lower boundary condition for thermodynamics	Zero flux

Continued on next page

Table 4.7 continued from previous page

Decision name	Decision value	Decision description	Value description
bcUpprSoiH	liq_flux	Upper boundary condition for soil hydrology	Liquid water flux
bcLowrSoiH	zeroFlux	Lower boundary condition for soil hydrology	Zero flux
veg_traits	CM_QJRMS1988	Vegetation roughness length and displacement heights	Choudhury & Monteith (1988)
canopyEmis	difTrans	Canopy emissivity	Function of diffuse transmissivity
snowIncept	lightSnow	Snow interception	Maximum interception capacity inverse function of new snow density
windPrfile	logBelowCanopy	Wind profile through canopy	Logarithmic profile below vegetation canopy
astability	louisinv	Stability function	Inverse power function (Louis, 1979)
canopySrad	BeersLaw	Canopy shortwave radiation	Beer's Law (as implemented in VIC)
alb_method	varDecay	Albedo representation	Variable decay rate (Dickinson et al., 1993)
compaction	anderson	Snow compaction algorithm	Semi-empirical method (Anderson, 1976)
snowLayers	CLM_2010	Snow layer combination/division	CLM type: rules depend on layer index

Continued on next page

Table 4.7 continued from previous page

Decision name	**Decision value**	**Decision description**	**Value description**
`thCondSnow`	`jrdn1991`	Thermal conductivity in snow	Jordan (1991)
`thCondSoil`	`funcSoilWet`	Thermal conductivity in soil	Function of soil wetness
`spatial_gw`	`localColumn`	Spatial representation of groundwater	Separate groundwater in each local soil column
`subRouting`	`timeDlay`	Sub-grid routing method	Time-delay histogram

Table 4.8: Model performance metrics for baseline and store-added configurations across calibration, evaluation, and full simulation periods.

Period	Model	KGE	NSE	log-KGE
East River				
Calibration	Baseline	0.90	0.76	0.83
	Deep store	0.85	0.78	0.80
Evaluation	Baseline	0.83	0.71	0.80
	Deep store	0.76	0.78	0.90
Full period	Baseline	0.86	0.76	0.70
	Deep store	0.87	0.76	0.71
Tuolumne River				
Calibration	Baseline	0.74	0.60	0.61
	Snowpatch	0.76	0.61	0.61
Evaluation	Baseline	0.76	0.62	0.70
	Snowpatch	0.77	0.64	0.70
Full period	Baseline	0.74	0.54	0.33
	Snowpatch	0.75	0.56	0.33

Table 4.9: SNOTEL site information

Attribute	**East River**	**Tuolumne River**
Site	Butte	Dana Meadows
Location	38.9°N, −106.95°W	37.898°N, −119.259°W
Elevation	3106 m	2975 m
SWE–Q regression (m, intercept)	m = 0.9753, intercept = 8.32	m = 1.6871, intercept = 118.47
Variance explained (r^2)	0.77	0.90

Chapter 5

CONCLUSIONS

5.1 Summary

In this dissertation, we examined several sources of uncertainty that can influence where "missing" mountain water originates. We did this through investigations of sublimation variability, precipitation measurement uncertainty, and long-term water storage dynamics.

We used a combination of measurements and models across a range of scales, from point measurements of sublimation, to valley-scale and synoptic-scale observations, to distributed precipitation measurements, to basin-wide domains in the East and Tuolumne Rivers. In combination, these strategies allowed us to explore several hydrometeorological processes in great detail over complex terrain.

In Chapter 2, we combined two years of meteorological observations spanning site-to-synoptic scales from three field campaigns in a Colorado mountain valley and identified the events that produced the most sublimation, classifying them by duration and intensity. We found that approximately 60% of winter sublimation occurred during a small number of discrete events (14% of the season). These events coincided with either short sunny, dry periods or long storm events that brought new snowfall followed by blowing snow. Using these classified events and synoptic-scale reanalysis, we trained a random forest classifier to estimate long sublimation event occurrences and test whether they increased over the past four decades. The model performed well (average precision = 0.73; balanced accuracy = 79%), and results were most strongly associated with winter mean 500-hPa wind speed and total observed precipitation. Event occurrence varied considerably year to year with no significant long-term trend. In all, our findings show that accurately representing seasonal sublimation requires capturing discrete events driven by identifiable synoptic and local meteorological conditions.

In Chapter 3, we performed a precipitation gauge measurement intercomparison to elucidate the dominant drivers of precipitation measurement variability in a mountain location. We compiled two years of precipitation measurements from 11 gauges deployed during the Surface Atmosphere Integrated Field Laboratory (SAIL); Study of Precipitation, Lower Atmosphere, and Surface for Hydrometeorology (SPLASH); and Sublimation of Snow (SOS) campaigns, supplemented by one radar-derived product and two independent gridded products. We evaluated measurement performance against a well-maintained reference gauge, identified the meteorological and instrument-specific drivers of measurement bias, and determined which gauges were most robust in this harsh environment. We found a large discrepancy in total accumulations across gauges, with individual gauges recording totals that ranged from 56% to 153% of the reference gauge. However, these biases relative to the reference were not distributed uniformly over time. Across all instruments, 40-60% of the total bias was concentrated in only 10 events that occurred primarily during winter. These event-specific biases disproportionately reflected instrument- and environment-related limitations, with dominant contributors including power issues, gauge burial, blowing snow, and high snowfall rates. For future precipitation measurement studies in mountainous, snow-dominated terrain, we recommend at least two co-located storage-type gauges with low power requirements be deployed to provide reliable, continuous measurements.

In Chapter 4, we found that late-season low flows in snow-dominated basins are sustained in part by water stored across multiple years, producing hydrologic "memory" – where storage conditions in one year influence streamflow in the next. We used the Structure for Unifying Multiple Modeling Alternatives (SUMMA) model to test whether adding a basin-appropriate long-term storage compartment could reproduce this behavior in two contrasting basins: the East River, Colorado, where deep groundwater is important, and the Tuolumne River, California, where perennial snowpatches matter. Observed late-season low flows showed significant one-year lag autocorrelation in both basins ($\rho_1 = +0.26$ and $+0.27$), which baseline models lacking a long-term storage compartment could not reproduce. Explicitly modeling multi-year storage recovered this signature ($\rho_1 = +0.31$ and $+0.29$), and in both basins end-of-year storage was positively correlated with the following year's low flows. These storage compartments also contributed disproportionately to low flows during dry years, supplying 30% of basin-wide low flows in the East and 14% in the Tuolumne, compared with 19% and 1% in the baseline models. Both baseline and store-added configurations substantially outperformed a snowpack–streamflow regression in the driest years, but adding a storage component did not further improve water-year volume accuracy. A simpler model appears sufficient for representing total water year volume and still outperforms a snowpack–streamflow regression. Where the quantity of interest is low flow, particularly in dry years, investing in a basin-appropriate representation of long-term storage is worthwhile, and this matters most as these long-term stores decline. A model that does not represent them cannot show that decline, and will instead attribute the resulting loss in late-season flow to other processes.

5.2 Broader Impacts, Future Directions, and Open Questions

Much work has been dedicated to addressing the uncertainties in water resource forecasts. But as these uncertainties become increasingly consequential in a changing climate [*Livneh and Badger*, 2020; *Siirila-Woodburn et al.*, 2021; *Ramírez Molina et al.*, 2026; *Sturm et al.*, 2017], we must be willing to question the assumptions that have been the backbone of water resource management over the past century [*Pagano et al.*, 2004; *Milly et al.*, 2008]. To do so, we must continue to explore causes for uncertainties in all aspects of the water balance.

Each of the scientific advancements outlined in Chapters 2–4 has implications for future scientific research or water resource management.

In Chapter 2, we identified the drivers of sublimation variability in a Colorado River headwater catchment, contributing to the conversation about sublimation's role in explaining hydrologic anomalies in the basin. The dominant event types are likely applicable to other mountain regions, though their relative frequency will vary with snow regime, terrain, and weather patterns. Long events are sustained by storm-driven snowfall and wind, which favor blowing snow and turbulent mixing; short events require high net radiation, dry air, and a breakdown of valley inversions. Both mechanisms are consistent with prior work emphasizing the joint role of meteorological forcing and surface conditions in driving sublimation [*Hood et al.*, 1999; *Xie et al.*, 2022]. Future research could try to determine the sources of turbulence during these events and relate them to physical mechanisms, something that has yet to be done in complex terrain [*Manninen et al.*, 2018].

Given their connection to winds and precipitation, long sublimation events are influenced by storm frequency and timing. While future changes to storm tracks remain uncertain across the Colorado region [*Bolinger et al.*, 2023], a shift toward fewer but stronger storms could amplify sublimation during individual events. Likewise, an earlier onset of spring-like conditions may increase the influence of short event frequency. Both these interactions should be front of mind for future studies investigating sublimation changes in coming decades.

Secondly, the breakdown of valley inversions in snow covered mountain valleys merits further study because this breakdown was a consistent feature of large sublimation events. Improving our ability to predict inversion formation and breakdown could enhance our ability to accurately detect these events. Lastly, our findings underscore the importance of resolving event-scale processes in models that simulate snowpack evolution and water availability. Therefore, if certain events are poorly captured in forcing data, or blowing snow conditions are not represented, sublimation may be grossly misrepresented.

In Chapter 3, we compared precipitation measurements at a mountain site using instruments rarely deployed together. By identifying the meteorological conditions that produce the largest biases, we diagnose when and why measurement accuracy degrades – revealing which gauges are most resilient during high-impact events and when records are most vulnerable to instrument-specific failure. This provides a framework for prioritizing gauge types, guiding site selection, and flagging periods warranting greater scrutiny.

Ground-based precipitation networks remain foundational for detecting hydroclimatic change, evaluating remotely sensed products, and forcing hydrologic models – but reliable measurements require more than a single gauge. Several practices can mitigate issues: deliberate siting to reduce instrument-specific biases, redundant deployments to guard against failure during critical events, and co-located meteorological observations to flag elevated uncertainty. Perhaps most importantly, a willingness to document and share measurement failures allows the community to honestly characterize the strengths and limitations of different technologies, ultimately helping to rein in precipitation measurement uncertainty and improve returns from future deployments.

However, questions remain as to the direct consequence of different measurement strategies on hydrologic model output. It is well known that precipitation is the primary form of uncertainty in streamflow forecasts, but future work should be dedicated to testing how hydrologic model performance differs when forced with precipitation from different gauge types.

In Chapter 4, we determined how adding a long-term water storage compartment to a semi-distributed, physically based hydrologic model affected its representation of streamflow memory in two physiographically distinct snow-dominated basins in the western United States, the East and Tuolumne Rivers. We applied a storage compartment that matches our conceptual understanding of each basin – perennial snowpatches in the Tuolumne and a deep, slowly draining aquifer in the East – and characterized memory using the lag autocorrelation of annual 7-day August–October low flows.

Three findings emerged. First, the baseline models could not replicate the observed 1-year lag memory in either basin, while the store-added configurations could. However, neither model reproduced the observed 5-year lag memory signal, indicating that our conceptual representations still miss features contributing to multi-year memory, though periodic variability in precipitation may partly explain this longer cycle.

Second, end-of-year storage state in both store-added models was positively correlated with the following year's low flows, evidence of buffering between prior-year storage and subsequent streamflow. In the Tuolumne, Landsat-derived snowpatch area showed the same 1-year lag correlation with observed low flows.

Third, adding a storage component had minimal effect on water-year volume accuracy relative to the baseline, yet both model types substantially outperformed a snowpack-streamflow regression in the driest years, reducing absolute bias by 64% in the East and 38% in the Tuolumne relative to the regression. This underscores the value of representing all water balance components when snowpack-streamflow errors are highest.

Several future research opportunities could expand on this work. Applying each storage compartment type to the opposite basin would test whether a configuration tuned to one storage mechanism degrades where physiography differs. Comparing across a range of subsurface complexity would identify the minimum representation needed to capture relationships demonstrated with more sophisticated models. Extended drought and recovery experiments have been performed for groundwater but not for perennial snow, and a direct comparison of their response would clarify how differently these storage compartments behave during extreme conditions.

Because storage memory was not a dominant source of the water-year volume errors we examined, it is unlikely to explain most hydrologic anomalies. But as temperatures continue to rise and droughts persist, long-term stores will draw down, and in the Tuolumne, snowpatches may eventually disappear entirely. Models without representations of these long-term stores cannot capture those changes and will instead attribute the resulting decline in late-season flow to other processes, obscuring the cause of decline. Determining which long-term storage mechanism dominates in a basin, and representing it explicitly, allows for multi-year changes to be physically accounted for.

5.3 Final Thoughts

Water is a substance of contrast. It can freeze solid, flow freely, or vanish into the atmosphere even in the same watershed on the same day. It is abundant, yet scarce; life-giving, yet destructive. These contrasts draw us as hydrologists into exploring how water interacts with the natural and built environment. Yet, it still holds many mysteries, as our current representation of the hydrologic system can still break down spectacularly. As the climate continues to change, we must motivate ourselves to continue advancing our understanding of the inter-connectedness of the hydrologic system from the atmosphere to bedrock. Nevertheless, we have come a long way. Back in 1861, an understanding of groundwater flow was considered to be "secret, occult and concealed" and "practically impossible" to comprehend (Ohio Supreme Court, Frazier v. Brown, 12 Ohio St. 294). It behooves us to keep advancing the field so that processes we currently believe to be "practically impossible to comprehend" become, in time, common knowledge.

BIBLIOGRAPHY

Aksamit, N. O., and J. W. Pomeroy, Warm-air entrainment and advection during alpine blowing snow events, *The Cryosphere*, *14*(9), 2795–2807, doi:10.5194/tc-14-2795-2020, 2020.

Alden, C., B. Sullender, J. Stimberis, and J. Lundquist, Higher Temperatures Lead to More Melt-Freeze Crusts in Snowpacks in Cooler Regions of the Pacific Northwest, *ARC Geophysical Research*, *2*(1), doi:10.5149/ARC-GR.2451, 2026.

Alter, J. C., Shielded Storage Precipitation Gages, *Monthly Weather Review*, *65*(7), 262–265, doi:10.1175/1520-0493(1937)65⟨262:SSPG⟩2.0.CO;2, 1937.

Alvarez-Garreton, C., J. P. Boisier, R. Garreaud, J. Seibert, and M. Vis, Progressive water deficits during multiyear droughts in basins with long hydrological memory in Chile, *Hydrology and Earth System Sciences*, *25*(1), 429–446, doi:10.5194/hess-25-429-2021, 2021.

Baker, S. A., B. Rajagopalan, and A. W. Wood, Enhancing Ensemble Seasonal Streamflow Forecasts in the Upper Colorado River Basin Using Multi-Model Climate Forecasts, *JAWRA Journal of the American Water Resources Association*, *57*(6), 906–922, doi: 10.1111/1752-1688.12960, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1752-1688.12960, 2021.

Bales, R. C., N. P. Molotch, T. H. Painter, M. D. Dettinger, R. Rice, and J. Dozier, Mountain hydrology of the western United States, *Water Resources Research*, *42*(8), doi:10.1029/2005WR004387, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2005WR004387, 2006.

Bales, R. C., R. Rice, and S. B. Roy, Estimated Loss of Snowpack Storage in the Eastern

Sierra Nevada with Climate Warming, *Journal of Water Resources Planning and Management*, *141*(2), 04014,055, doi:10.1061/(ASCE)WR.1943-5452.0000453, 2015.

Ballard, B., D. Gottas, and A. White, NOAA PSL OttDisdrometerRaw KettlePonds for SPLASH, doi:10.5281/zenodo.10368926, 2023.

Battaglia, A., E. Rustemeier, A. Tokay, U. Blahak, and C. Simmer, PARSIVEL Snow Observations: A Critical Assessment, *Journal of Atmospheric and Oceanic Technology*, doi:10.1175/2009JTECHA1332.1, 2010.

Beck, H. E., M. Pan, T. Roy, G. P. Weedon, F. Pappenberger, A. I. J. M. van Dijk, G. J. Huffman, R. F. Adler, and E. F. Wood, Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS, *Hydrology and Earth System Sciences*, *23*(1), 207–224, doi:10.5194/hess-23-207-2019, 2019.

Bennett, A., B. Nijssen, G. Ou, M. Clark, and G. Nearing, Quantifying Process Connectivity With Transfer Entropy in Hydrologic Models, *Water Resources Research*, *55*(6), 4613–4629, doi:10.1029/2018WR024555, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018WR024555, 2019.

Bennett, A. R., J. J. Hamman, and B. Nijssen, MetSim: A Python package for estimation and disaggregation of meteorological data, *Journal of Open Source Software*, *5*(47), 2042, doi:10.21105/joss.02042, 2020.

Berghuijs, W. R., R. A. Woods, B. J. Anderson, A. L. Hemshorn de Sánchez, and M. Hrachowitz, Annual memory in the terrestrial water cycle, *Hydrology and Earth System Sciences*, *29*(5), 1319–1333, doi:10.5194/hess-29-1319-2025, 2025.

Beven, K. J., and M. J. Kirkby, A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant, *Hydrological Sciences Bulletin*, *24*(1), 43–69, doi:10.1080/02626667909491834, _eprint: https://doi.org/10.1080/02626667909491834, 1979.

Bilish, S. P., J. N. Callow, and H. A. McGowan, Streamflow variability and the role of snowmelt in a marginal snow environment, *Arctic, Antarctic, and Alpine Research*, *52*(1), 161–176, doi:10.1080/15230430.2020.1746517, 2020.

Bintanja, R., Modelling snowdrift sublimation and its effect on the moisture budget of the atmospheric boundary layer, *Tellus A: Dynamic Meteorology and Oceanography*, *53*(2), 215–232, doi:10.3402/tellusa.v53i2.12189, _eprint: https://doi.org/10.3402/tellusa.v53i2.12189, 2001a.

Bintanja, R., Snowdrift Sublimation in a Katabatic Wind Region of the Antarctic Ice Sheet, *Journal of Applied Meteorology*, *40*, 1952–1966, 2001b.

Bliss, A. K., K. M. Cuffey, and J. L. Kavanaugh, Sublimation and surface energy budget of Taylor Glacier, Antarctica, *Journal of Glaciology*, *57*(204), 684–696, doi:10.3189/002214311797409767, 2011.

Boardman, E. N., et al., Sources of seasonal water supply forecast uncertainty during snow drought in the Sierra Nevada, *JAWRA Journal of the American Water Resources Association*, *60*(5), 972–990, doi:10.1111/1752-1688.13221, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1752-1688.13221, 2024.

Bolinger, R., J. Lukas, R. Schumacher, and P. Goble, Climate change in Colorado, *Tech. Rep. 3*, Colorado State University Libraries, 2023.

Bonin, T. A., A. Choukulkar, W. A. Brewer, S. P. Sandberg, A. M. Weickmann, Y. L. Pichugina, R. M. Banta, S. P. Oncley, and D. E. Wolfe, Evaluation of turbulence measurement techniques from a single Doppler lidar, *Atmospheric Measurement Techniques*, *10*(8), 3021–3039, doi:10.5194/amt-10-3021-2017, 2017.

Boudala, F. S., G. A. Isaac, R. Rasmussen, S. G. Cober, and B. Scott, Comparisons of Snowfall Measurements in Complex Terrain Made During the 2010 Winter Olympics in Vancouver, *Pure and Applied Geophysics*, *171*(1), 113–127, doi:10.1007/s00024-012-0610-5, 2014.

Brandes, E. A., K. Ikeda, G. Zhang, M. Schönhuber, and R. M. Rasmussen, A Statistical and Physical Description of Hydrometeor Distributions in Colorado Snowstorms Using a Video Disdrometer, *Journal of Applied Meteorology and Climatology*, doi:10.1175/JAM2489.1, 2007.

Breiman, L., Random Forests, *Machine Learning*, *45*(1), 5–32, doi:10.1023/A:1010933404324, 2001.

Brooks, P. D., et al., Groundwater-Mediated Memory of Past Climate Controls Water Yield in Snowmelt-Dominated Catchments, *Water Resources Research*, *57*(10), e2021WR030,605, doi:10.1029/2021WR030605, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021WR030605, 2021.

Brooks, P. D., et al., Groundwater dominates snowmelt runoff and controls streamflow efficiency in the western United States, *Communications Earth & Environment*, *6*(1), 341, doi:10.1038/s43247-025-02303-3, 2025.

Bytheway, J. L., W. R. Currier, M. R. Abel, K. Mahoney, and R. Cifelli, Evaluation of Wintertime Precipitation Estimates and Forecasts in the Mountains of Colorado, *Journal of Hydrometeorology*, doi:10.1175/JHM-D-23-0158.1, 2024.

Bárdossy, A., C. Kilsby, S. Birkinshaw, N. Wang, and F. Anwar, Is Precipitation Responsible for the Most Hydrological Model Uncertainty?, *Frontiers in Water*, *4*, doi:10.3389/frwa.2022.836554, 2022.

Capozzi, V., C. Annella, M. Montopoli, E. Adirosi, G. Fusco, and G. Budillon, Influence of Wind-Induced Effects on Laser Disdrometer Measurements: Analysis and Compensation Strategies, *Remote Sensing*, *13*(15), 3028, doi:10.3390/rs13153028, 2021.

Carroll, R. W. H., L. A. Bearup, W. Brown, W. Dong, M. Bill, and K. H. Willlams, Factors controlling seasonal groundwater and solute flux from snow-dominated basins, *Hydrological Processes*, *32*(14), 2187–2202, doi:10.1002/hyp.13151, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.13151, 2018.

Carroll, R. W. H., J. S. Deems, R. Niswonger, R. Schumer, and K. H. Williams, The Importance of Interflow to Groundwater Recharge in a Snowmelt-Dominated Headwater Basin, *Geophysical Research Letters*, *46*(11), 5899–5908, doi:10.1029/2019GL082447, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019GL082447, 2019.

Carroll, R. W. H., D. Gochis, and K. H. Williams, Efficiency of the Summer Monsoon in Generating Streamflow Within a Snow-Dominated Headwater Basin of the Colorado River, *Geophysical Research Letters*, *47*(23), e2020GL090,856, doi:10.1029/2020GL090856, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL090856, 2020.

Carroll, R. W. H., R. G. Niswonger, C. Ulrich, C. Varadharajan, E. R. Siirila-Woodburn, and K. H. Williams, Declining groundwater storage expected to amplify mountain streamflow reductions in a warmer world, *Nature Water*, *2*(5), 419–433, doi:10.1038/s44221-024-00239-0, 2024.

Carroll, R. W. H., A. H. Manning, and K. H. Williams, The Role of Bedrock Circulation Depth and Porosity in Mountain Streamflow Response to Prolonged Drought, *Geophysical Research Letters*, *52*(4), e2024GL112,927, doi:10.1029/2024GL112927, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2024GL112927, 2025.

Chegwidden, O. S., et al., How Do Modeling Decisions Affect the Spread Among Hydrologic Climate Change Projections? Exploring a Large Ensemble of Simulations Across a Diversity of Hydroclimates, *Earth's Future*, *7*(6), 623–637, doi:10.1029/2018EF001047, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018EF001047, 2019.

Chen, T., and C. Guestrin, XGBoost: A Scalable Tree Boosting System, in *Proceedings of the*

22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, doi:10.1145/2939672.2939785, arXiv:1603.02754 [cs], 2016.

Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay, Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, *Water Resources Research*, *44*(12), doi:10.1029/2007WR006735, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2007WR006735, 2008.

Clark, M. P., et al., A unified approach for process-based hydrologic modeling: 1. Modeling concept, *Water Resources Research*, *51*(4), 2498–2514, doi:10.1002/2015WR017198, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2015WR017198, 2015a.

Clark, M. P., et al., A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies, *Water Resources Research*, *51*(4), 2515–2542, doi:10.1002/2015WR017200, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2015WR017200, 2015b.

Cline, D. W., Snow surface energy exchanges and snowmelt at a continental, midlatitude Alpine site, *Water Resources Research*, *33*(4), 689–701, doi:10.1029/97WR00026, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/97WR00026, 1997.

Copernicus Climate Change Service, ERA5 hourly data on pressure levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), doi:10.24381/cds.bd0915c6, 2023.

Cox, C., M. Gallagher, J. Intrieri, B. Butterworth, T. Meyers, and O. Persson, Atmospheric Surface Flux Station #30 measurements (level 2 Processed), Study of Precipitation, the Lower Atmosphere and Surface for Hydrometeorology (SPLASH), September 2021-July 2023, doi:10.5281/zenodo.14641866, 2025.

Cox, C. J., et al., Continuous observations of the surface energy budget and meteorology over the Arctic sea ice during MOSAiC, *Scientific Data*, *10*(1), 519, doi:10.1038/s41597-023-02415-5, 2023.

Cristea, N. C., A. Bennett, B. Nijssen, and J. D. Lundquist, When and Where Are Multiple Snow Layers Important for Simulations of Snow Accumulation and Melt?, *Water Resources Research*, *58*(10), e2020WR028,993, doi:10.1029/2020WR028993, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020WR028993, 2022.

Cullen, N. J., T. Mölg, J. Conway, and K. Steffen, Assessing the role of sublimation in the dry snow zone of the Greenland ice sheet in a warming world, *Journal of Geophysical Research: Atmospheres*, *119*(11), 6563–6577, doi:10.1002/2014JD021557, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2014JD021557, 2014.

Currier, W. R., T. Thorson, and J. D. Lundquist, Independent Evaluation of Frozen Precipitation from WRF and PRISM in the Olympic Mountains, *Journal of Hydrometeorology*, doi:10.1175/JHM-D-17-0026.1, 2017.

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris, Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States, *International Journal of Climatology*, *28*(15), 2031–2064, doi:10.1002/joc.1688, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.1688, 2008.

Daly, C., et al., Challenges in Observation-Based Mapping of Daily Precipitation across the Conterminous United States, *Journal of Atmospheric and Oceanic Technology*, doi:10.1175/JTECH-D-21-0054.1, 2021.

de Boer, G., et al., Supporting Advancement in Weather and Water Prediction in the Upper Colorado River Basin: The SPLASH Campaign, *Bulletin of the American Meteorological Society*, pp. 1853–1874, doi:10.1175/BAMS-D-22-0147.1, 2023.

de Lavenne, A., V. Andréassian, L. Crochemore, G. Lindström, and B. Arheimer, Quantifying multi-year hydrological memory with Catchment Forgetting Curves, *Hydrology and Earth System Sciences*, *26*(10), 2715–2732, doi:10.5194/hess-26-2715-2022, 2022.

Derin, Y., and K. K. Yilmaz, Evaluation of Multiple Satellite-Based Precipitation Products over Complex Topography, *Journal of Hydrometeorology*, doi:10.1175/JHM-D-13-0191.1, 2014.

Dettinger, M., F. Ralph, T. Das, P. Neiman, and D. Cayan, Atmospheric Rivers, Floods and the Water Resources of California, *Water*, *3*, 445–478, doi:10.3390/w3020445, 2011.

Dewitz, and U. G. Survey, National Land Cover Database (NLCD) 2019 Products, doi: https://doi.org/10.5066/P9KZCM54, 2021.

Dilley, A. C., and D. M. O'brien, Estimating downward clear sky long-wave irradiance at the surface from screen temperature and precipitable water, doi:10.1002/qj.49712454903.

Do, V., L. B. Wilner, N. M. Flores, H. McBrien, A. J. Northrop, and J. A. Casey, Spatiotemporal patterns of individual and multiple simultaneous severe weather events co-occurring with power outages in the United States, 2018–2020, *PLOS Climate*, *4*(1), e0000,523, doi:10.1371/journal.pclm.0000523, 2025.

Earth Resources Observation and Science (EROS) Center, Landsat 4-5 thematic mapper level-2, collection 2, doi:10.5066/P9IAXOVV, 2020a.

Earth Resources Observation and Science (EROS) Center, Landsat 7 enhanced thematic mapper plus level-2, collection 2, doi:10.5066/P9C7I13B, 2020b.

Earth Resources Observation and Science (EROS) Center, Landsat 8-9 operational land imager / thermal infrared sensor level-2, collection 2, doi:10.5066/P9OGBGM6, 2020c.

ECMWF, Climate Data Store (CDS) Application Program Interface, 2025.

Elkouk, A., et al., Toward Understanding Parametric Controls on Runoff Sensitivity to Climate in the Community Land Model: A Case Study Over the Colorado River Headwaters, *Water Resources Research*, *60*(12), e2024WR037,718, doi:10.1029/2024WR037718, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2024WR037718, 2024.

Falco, N., et al., Vegetation classification map and covariates associated with NEON AOP survey, East River, CO 2018, 2024.

Fassnacht, S. R., Estimating Alter-shielded gauge snowfall undercatch, snowpack sublimation, and blowing snow transport at six sites in the coterminous USA, *Hydrological Processes*, *18*(18), 3481–3492, doi:10.1002/hyp.5806, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.5806, 2004.

Faybishenko, B., R. Versteeg, G. Pastorello, D. Dwivedi, C. Varadharajan, and D. Agarwal, Quality Assurance and Quality Control (QA/QC) of Meteorological Time Series Data for Billy Barr, East River, Colorado USA, doi:10.15485/1823516, 2021.

Feldman, D. R., et al., The Surface Atmosphere Integrated Field Laboratory (SAIL) Campaign, *Bulletin of the American Meteorological Society*, *104*(12), E2192–E2222, doi:10.1175/BAMS-D-22-0049.1, 2023.

Finnigan, J. J., R. Clement, Y. Malhi, R. Leuning, and H. Cleugh, A Re-Evaluation of Long-Term Flux Measurement Techniques Part I: Averaging and Coordinate Rotation, *Boundary-Layer Meteorology*, *107*(1), 1–48, doi:10.1023/A:1021554900225, 2003.

Fleming, S. W., L. Zukiewicz, M. L. Strobel, H. Hofman, and A. G. Goodbody, SNOTEL, the Soil Climate Analysis Network, and water supply forecasting at the Natural Resources Conservation Service: Past, present, and future, *JAWRA Journal of the American Water Resources Association*, *59*(4), 585–599, doi:10.1111/1752-1688.13104, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1752-1688.13104, 2023.

Foken, T., R. Leuning, S. R. Oncley, M. Mauder, and M. Aubinet, Corrections and Data Quality Control, in *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*, edited by M. Aubinet, T. Vesala, and D. Papale, pp. 85–131, Springer Netherlands, Dordrecht, doi:10.1007/978-94-007-2351-1_4, 2012.

Fountain, A. G., and W. V. Tangborn, The Effect of Glaciers on Streamflow Variations, *Water Resources Research*, *21*(4), 579–586, doi:10.1029/WR021i004p00579, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/WR021i004p00579, 1985.

Fowler, K., W. Knoben, M. Peel, T. Peterson, D. Ryu, M. Saft, K.-W. Seo, and A. Western, Many Commonly Used Rainfall-Runoff Models Lack Long, Slow Dynamics: Implications for Runoff Projections, *Water Resources Research*, *56*(5), e2019WR025,286, doi:10.1029/2019WR025286, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019WR025286, 2020.

Frei, C., and C. Schär, A precipitation climatology of the Alps from high-resolution rain-gauge observations, *International Journal of Climatology*, *18*(8), 873–900, doi:10.1002/(SICI)1097-0088(19980630)18:8⟨873::AID-JOC255⟩3.0.CO;2-9, _eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291097-0088%2819980630%2918%3A8%3C873%3A%3AAID-JOC255%3E3.0.CO%3B2-9, 1998.

Gan, Y., Y. Zhang, C. Kongoli, and M. Pan, The Role of Forcing and Parameterization in Improving Snow Simulation in the Upper Colorado River Basin Using the National Water Model, *Water Resources Research*, *60*(8), e2023WR035,303, doi:10.1029/2023WR035303, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2023WR035303, 2024.

Ghimire, S., E. R. Vivoni, and Z. Wang, Fall Soil Moisture Modulates Snow-Streamflow Dynamics in the Colorado River Basin, *Water Resources Research*, *62*(7), e2025WR042,871, doi:10.1029/2025WR042871, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2025WR042871, 2026.

Gilson, G. F., H. Jiskoot, J. J. Cassano, and T. R. Nielsen, Radiosonde-Derived Temperature Inversions and Their Association With Fog Over 37 Melt Seasons in East Greenland, *Journal of Geophysical Research: Atmospheres*, *123*(17), 9571–9588, doi:10.1029/2018JD028886, 2018.

Goble, P. E., and R. S. Schumacher, On the Sources of Water Supply Forecast Error in Western Colorado, *Journal of Hydrometeorology*, doi:10.1175/JHM-D-23-0004.1, 2023.

Goodison, B., P. Louie, and D. Yang, WMO solid precipitation measurement intercomparison, *Technical Document WMO/TD-No. 872*, World Meteorological Organization, Geneva, Switzerland, 1998.

Gordon, B. L., P. D. Brooks, S. A. Krogh, G. F. S. Boisrame, R. W. H. Carroll, J. P. McNamara, and A. A. Harpold, Why does snowmelt-driven streamflow response to warming vary? A data-driven review and predictive framework, *Environmental Research Letters*, *17*(5), 053,004, doi:10.1088/1748-9326/ac64b4, 2022.

Gossart, A., N. Souverijns, I. V. Gorodetskaya, S. Lhermitte, J. T. M. Lenaerts, J. H. Schween, A. Mangold, Q. Laffineur, and N. P. M. van Lipzig, Blowing snow detection from ground-based ceilometers: application to East Antarctica, *The Cryosphere*, *11*(6), 2755–2772, doi:10.5194/tc-11-2755-2017, 2017.

Groisman, P. Y., and D. R. Legates, The Accuracy of United States Precipitation Data, *Bulletin of the American Meteorological Society*, *75*(2), 215–228, doi:10.1175/1520-0477(1994)075⟨0215:TAOUSP⟩2.0.CO;2, 1994.

Groot Zwaaftink, C. D., R. Mott, and M. Lehning, Seasonal simulation of drifting snow sublimation in Alpine terrain, *Water Resources Research*, *49*(3), 1581–1590, doi:10.1002/wrcr.20137, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wrcr.20137, 2013.

Grover, M. A., J. R. O'Brien, R. C. Jackson, Z. S. Sherman, S. M. Collis, B. A. Raut, A. Theisen, M. Tuftedal, and D. Feldman, SAIL Field Campaign X-Band Precipitation

Radar Surface Quantitative Precipitation Estimation (SQUIRE) Value-Added Product Report, *Tech. Rep. DOE/SC-ARM-TR–287*, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States). Atmospheric Radiation Measurement (ARM), Data Center, doi:10.2172/1998075, 2023.

Grundstein, A. J., and D. J. Leathers, A spatial analysis of snow–surface energy exchanges over the northern Great Plains of the United States in relation to synoptic scale forcing mechanisms, *International Journal of Climatology*, *19*(5), 489–511, doi:10.1002/(SICI)1097-0088(199904)19:5⟨489::AID-JOC373⟩3.0.CO;2-J, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291097-0088%28199904%2919%3A5%3C489%3A%3AAID-JOC373%3E3.0.CO%3B2-J, 1999.

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

Gultepe, I., G. A. Isaac, P. Joe, P. A. Kucera, J. M. Theriault, and T. Fisico, Roundhouse (RND) Mountain Top Research Site: Measurements and Uncertainties for Winter Alpine Weather Conditions, *Pure and Applied Geophysics*, *171*(1), 59–85, doi:10.1007/s00024-012-0582-5, 2014.

Guo, S., Y. Li, and R. Chen, Climate Warming Favoring Sublimation on a Westerlies-Controlled Glacier of Tibetan Plateau, *Journal of Geophysical Research: Atmospheres*, *128*(24), e2023JD039,332, doi:10.1029/2023JD039332, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2023JD039332, 2023.

Guyomarc'h, G., and L. Mérindol, Validation of an application for forecasting blowing snow, *Annals of Glaciology*, *26*, 138–143, doi:10.3189/1998AoG26-1-138-143, 1998.

Hall, D. K., G. A. Riggs, and V. V. Salomonson, Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data, *Remote Sensing of Environment*, *54*(2), 127–140, doi:10.1016/0034-4257(95)00137-P, 1995.

Hall, F. R., Base-Flow Recessions—A Review, *Water Resources Research*, *4*(5), 973–983, doi:10.1029/WR004i005p00973, 1968.

Harpold, A., P. Brooks, S. Rajagopal, I. Heidbuchel, A. Jardine, and C. Stielstra, Changes in snowpack accumulation and ablation in the intermountain west, *Water Resources Research*, *48*(11), 501, doi:10.1029/2012WR011949, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2012WR011949, 2012.

Harpold, A. A., and P. D. Brooks, Humidity determines snowpack ablation under a warming climate, *Proceedings of the National Academy of Sciences*, *115*(6), 1215–1220, doi:10.1073/pnas.1716789115, 2018.

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

Hawkins, J., R. Hallnan, and G. M. Stock, Multidecadal area changes of perennial snowpatches in Yosemite National Park, CA, USA, *Arctic, Antarctic, and Alpine Research*, *58*(1), 2655,505, doi:10.1080/15230430.2026.2655505, _eprint: https://doi.org/10.1080/15230430.2026.2655505, 2026.

Hayashi, M., T. Hirota, Y. Iwata, and I. Takayabu, Snowmelt Energy Balance and Its Relation to Foehn Events in Tokachi, Japan, *Journal of the Meteorological Society of Japan. Ser. II*, *83*(5), 783–798, doi:10.2151/jmsj.83.783, 2005.

Henn, B., A. J. Newman, B. Livneh, C. Daly, and J. D. Lundquist, An assessment of differences in gridded precipitation datasets in complex terrain, *Journal of Hydrology*, *556*, 1205–1219, doi:10.1016/j.jhydrol.2017.03.008, 2018.

Hersbach, H., et al., The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049, doi:10.1002/qj.3803, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3803, 2020.

Heymsfield, A. J., A. Bansemer, C. Schmitt, C. Twohy, and M. R. Poellot, Effective Ice Particle Densities Derived from Aircraft Data, *Journal of the Atmospheric Sciences*, *61*(9), 982–1003, doi:10.1175/1520-0469(2004)061⟨0982:EIPDDF⟩2.0.CO;2, 2004.

Hogan, D., and J. Lundquist, Recent Upper Colorado River Streamflow Declines Driven by Loss of Spring Precipitation, doi:10.5281/zenodo.11043101, 2024.

Hogan, D., E. Schwat, E. Gutmann, D. Feldman, and J. D. Lundquist, Precipitation Measurement Uncertainty in Mountains: A Gauge Intercomparison from the SAIL, SPLASH, and SOS Campaigns, doi:10.22541/essoar.15004476/v1, 2026a.

Hogan, D., E. Schwat, E. Gutmann, J. Vano, and J. D. Lundquist, Synoptic-Scale Systems Control Total Winter Sublimation at an Alpine Site, *Journal of Hydrometeorology*, *27*(4), 529–547, doi:10.1175/JHM-D-25-0143.1, 2026b.

Hogan, R. J., A. L. M. Grant, A. J. Illingworth, G. N. Pearson, and E. J. O'Connor, Vertical velocity variance and skewness in clear and cloud-topped boundary layers as revealed by Doppler lidar, *Quarterly Journal of the Royal Meteorological Society*, *135*(640), 635–643, doi:10.1002/qj.413, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.413, 2009.

Holroyd, E. W., and J. E. Jiusto, Snowfall from a Heavily Seeded Cloud, *Journal of Applied Meteorology*, *10*(2), 266–269, 1971.

Hood, E., M. Williams, and D. Cline, Sublimation from a seasonal snowpack at a continental, mid-latitude alpine site, *Hydrological Processes*, *13*(12-13), 1781–1797, doi:10.1002/(SICI)1099-1085(199909)13:12/13⟨1781::AID-HYP860⟩3.0.CO;2-C, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291099-1085%28199909%2913%3A12/13%3C1781%3A%3AAID-HYP860%3E3.0.CO%3B2-C, 1999.

Hourdequin, M., Climate Change, Climate Engineering, and the 'Global Poor': What Does Justice Require?, *Ethics, Policy & Environment*, *21*(3), 270–288, doi:10.1080/21550085.2018.1562525, _eprint: https://doi.org/10.1080/21550085.2018.1562525, 2018.

Huang, J., A Simple Accurate Formula for Calculating Saturation Vapor Pressure of Water and Ice, *Journal of Applied Meteorology and Climatology*, *57*, 1265–1272, doi:10.1175/JAMC-D-17-0334.1, 2018.

Hubbard, S. S., et al., The East River, Colorado, Watershed: A Mountainous Community Testbed for Improving Predictive Understanding of Multiscale Hydrological–Biogeochemical Dynamics, *Vadose Zone Journal*, *17*(1), 1–25, doi:10.2136/vzj2018.03.0061, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2136/vzj2018.03.0061, 2018.

Iliopoulou, T., et al., A large sample analysis of European rivers on seasonal river flow correlation and its physical drivers, *Hydrology and Earth System Sciences*, *23*(1), 73–91, doi:10.5194/hess-23-73-2019, 2019.

Ishizaka, M., An accurate measurement of densities of snowflakes using 3-D microphotographs, *Annals of Glaciology*, *18*, 92–96, doi:10.3189/S0260305500011319, 1993.

Jackson, R., et al., Surface Quantitative Precipitation Estimates (SQUIRE) of Snow Water Equivalent from the Surface Atmospheric Integrated Field Laboratory, *Journal of Atmospheric and Oceanic Technology*, *43*(1), 77–93, doi:10.1175/JTECH-D-25-0023.1, 2026.

Jackson, S. I., and T. D. Prowse, Spatial variation of snowmelt and sublimation in a high-elevation semi-desert basin of western Canada, *Hydrological Processes*, *23*(18), 2611–2627, doi:10.1002/hyp.7320, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.7320, 2009.

Joseph, J., S. Ghosh, A. Pathak, and A. K. Sahai, Hydrologic impacts of climate change: Comparisons between hydrological parameter uncertainty and climate model uncertainty, *Journal of Hydrology*, *566*, 1–22, doi:10.1016/j.jhydrol.2018.08.080, 2018.

Kidd, C., P. Bauer, J. Turk, G. J. Huffman, R. Joyce, K.-L. Hsu, and D. Braithwaite, Intercomparison of High-Resolution Precipitation Products over Northwest Europe, *Journal of Hydrometeorology*, doi:10.1175/JHM-D-11-042.1, 2012.

Kidd, C., A. Becker, G. J. Huffman, C. L. Muller, P. Joe, G. Skofronick-Jackson, and D. B. Kirschbaum, So, How Much of the Earth's Surface Is Covered by Rain Gauges?, *Bulletin of the American Meteorological Society*, doi:10.1175/BAMS-D-14-00283.1, 2017.

Kirchner, J. W., Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, *Water Resources Research*, *45*(2), doi:10.1029/2008WR006912, 2009.

Knoben, W. J. M., et al., Community Workflows to Advance Reproducibility in Hydrologic Modeling: Separating Model-Agnostic and Model-Specific Configuration Steps in Applications of Large-Domain Hydrologic Models, *Water Resources Research*, *58*(11), e2021WR031,753, doi:10.1029/2021WR031753, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021WR031753, 2022.

Knowles, J. F., P. D. Blanken, M. W. Williams, and K. M. Chowanski, Energy and surface moisture seasonally limit evaporation and sublimation from snow-free alpine tundra, *Agricultural and Forest Meteorology*, *157*, 106–115, doi:10.1016/j.agrformet.2012.01.017, 2012.

Kochendorfer, J., T. P. Meyers, M. E. Hall, S. D. Landolt, J. Lentz, and H. J. Diamond, A new reference-quality precipitation gauge wind shield, *Atmospheric Measurement Techniques*, *16*(22), 5647–5657, doi:10.5194/amt-16-5647-2023, 2023.

Kochendorfer, J., et al., The quantification and correction of wind-induced precipitation measurement errors, *Hydrology and Earth System Sciences*, *21*(4), 1973–1989, doi:10.5194/hess-21-1973-2017, 2017.

Kochendorfer, J., et al., How Well Are We Measuring Snow Post-SPICE?, *Bulletin of the American Meteorological Society*, doi:10.1175/BAMS-D-20-0228.1, 2022.

Kyrouac, J., and M. Tuftedal, Surface Meteorological System (MET) Instrument Handbook, *ARM Tech. Rep. DOE/SC-ARM/TR-086*, U.S. Department of Energy, Atmospheric Radiation Measurement user facility, Richland, Washington, doi:10.2172/1007926, 2024.

Lapides, D. A., W. J. Hahm, D. M. Rempe, J. Whiting, and D. N. Dralle, Causes of Missing Snowmelt Following Drought, *Geophysical Research Letters*, *49*(19), e2022GL100,505, doi:10.1029/2022GL100505, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2022GL100505, 2022.

Larson, L. W., and E. L. Peck, Accuracy of precipitation measurements for hydrologic modeling, *Water Resources Research*, *10*(4), 857–863, doi:10.1029/WR010i004p00857, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/WR010i004p00857, 1974.

Lavers, D. A., A. Simmons, F. Vamborg, and M. J. Rodwell, An evaluation of ERA5 precipitation for climate monitoring, *Quarterly Journal of the Royal Meteorological Society*, *148*(748), 3152–3165, doi:10.1002/qj.4351, _eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.4351, 2022.

Leathers, K., D. Herbst, G. de Mendoza, G. Doerschlag, and A. Ruhi, Climate change is poised to alter mountain stream ecosystem processes via organismal phenological shifts, *Proceedings of the National Academy of Sciences*, *121*(14), e2310513,121, doi:10.1073/pnas.2310513121, 2024.

Li, D., M. L. Wrzesien, M. Durand, J. Adam, and D. P. Lettenmaier, 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, doi:10.1002/2017GL073551, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2017GL073551, 2017.

Liaqat, M. U., S. Camici, F. Leopardi, J. Gaona, and L. Brocca, Beyond GRACE: Evaluating the benefits of NGGM and MAGIC for precipitation estimation over Europe, *Hydrology and Earth System Sciences*, *30*(15), 4969–4983, doi:10.5194/hess-30-4969-2026, 2026.

Liston, G., and M. Sturm, The role of winter sublimation in the Arctic moisture budget, *Hydrology Research*, *35*(4-5), 325–334, doi:10.2166/nh.2004.0024, 2004.

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

Luce, C. H., J. T. Abatzoglou, and Z. A. Holden, The Missing Mountain Water: Slower Westerlies Decrease Orographic Enhancement in the Pacific Northwest USA, *Science*, *342*(6164), 1360–1364, doi:10.1126/science.1242335, 2013.

Lundquist, J., and J. Roche, Climate change and water supply in western national parks, *Park science*, *26*(1), 31–33, 2009.

Lundquist, J., M. Hughes, E. Gutmann, and S. Kapnick, Our Skill in Modeling Mountain Rain and Snow is Bypassing the Skill of Our Observational Networks, *Bulletin of the American Meteorological Society*, *100*(12), 2473–2490, doi:10.1175/BAMS-D-19-0001.1, 2019.

Lundquist, J. D., and S. P. Loheide II, How evaporative water losses vary between wet and dry water years as a function of elevation in the Sierra Nevada, California, and critical factors for modeling, *Water Resources Research*, *47*(3), doi:10.1029/2010WR010050, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2010WR010050, 2011.

Lundquist, J. D., M. D. Dettinger, and D. R. Cayan, Snow-fed streamflow timing at different basin scales: Case study of the Tuolumne River above Hetch Hetchy, Yosemite, California, *Water Resources Research*, *41*(7), doi:10.1029/2004WR003933, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2004WR003933, 2005.

Lundquist, J. D., N. Pepin, and C. Rochford, Automated algorithm for mapping regions of cold-air pooling in complex terrain, *Journal of Geophysical Research: Atmospheres*, *113*(D22), 107, doi:10.1029/2008JD009879, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2008JD009879, 2008.

Lundquist, J. D., M. Hughes, B. Henn, E. D. Gutmann, B. Livneh, J. Dozier, and P. Neiman, High-Elevation Precipitation Patterns: Using Snow Measurements to Assess Daily Gridded Datasets across the Sierra Nevada, California, *Journal of Hydrometeorology*, *16*(4), 1773–1792, doi:10.1175/JHM-D-15-0019.1, 2015a.

Lundquist, J. D., N. E. Wayand, A. Massmann, M. P. Clark, F. Lott, and N. C. Cristea, Diagnosis of insidious data disasters, *Water Resources Research*, *51*(5), 3815–3827, doi:10.1002/2014WR016585, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014WR016585, 2015b.

Lundquist, J. D., et al., Yosemite Hydroclimate Network: Distributed stream and atmospheric data for the Tuolumne River watershed and surroundings, *Water Resources Research*, *52*(9), 7478–7489, doi:10.1002/2016WR019261, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016WR019261, 2016.

Lundquist, J. D., et al., Sublimation of Snow, *Bulletin of the American Meteorological Society*, *105*(6), E975–E990, doi:10.1175/BAMS-D-23-0191.1, 2024.

MacDonald, M. K., J. W. Pomeroy, and R. L. H. Essery, Water and energy fluxes over northern prairies as affected by chinook winds and winter precipitation, *Agricultural and Forest Meteorology*, *248*, 372–385, doi:10.1016/j.agrformet.2017.10.025, 2018.

Mahanama, S., B. Livneh, R. Koster, D. Lettenmaier, and R. Reichle, Soil Moisture, Snow, and Seasonal Streamflow Forecasts in the United States, *Journal of Hydrometeorology*, *13*(1), 189–203, doi:10.1175/JHM-D-11-046.1, 2012.

Mai, J., J. R. Craig, and B. A. Tolson, Simultaneously determining global sensitivities of model parameters and model structure, *Hydrology and Earth System Sciences*, *24*(12), 5835–5858, doi:10.5194/hess-24-5835-2020, 2020.

Mankin, K. R., S. Mehan, T. R. Green, and D. M. Barnard, Review of gridded climate products and their use in hydrological analyses reveals overlaps, gaps, and the need for a more objective approach to selecting model forcing datasets, *Hydrology and Earth System Sciences*, *29*(1), 85–108, doi:10.5194/hess-29-85-2025, 2025.

Manninen, A. J., T. Marke, M. Tuononen, and E. J. O'Connor, Atmospheric Boundary Layer Classification With Doppler Lidar, *Journal of Geophysical Research: Atmospheres*, *123*(15), 8172–8189, doi:10.1029/2017JD028169, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2017JD028169, 2018.

Martinaitis, S. M., S. B. Cocks, Y. Qi, B. T. Kaney, J. Zhang, and K. Howard, Understanding Winter Precipitation Impacts on Automated Gauge Observations within a Real-Time System, *Journal of Hydrometeorology*, *16*(6), 2345–2363, doi:10.1175/JHM-D-15-0020.1, 2015.

Martinaitis, S. M., S. Lincoln, D. Schlotzhauer, S. B. Cocks, and J. Zhang, A Temporal Gauge Quality Control Algorithm as a Method for Identifying Potential Instrumentation Malfunctions, *Journal of Atmospheric and Oceanic Technology*, doi:10.1175/JTECH-D-22-0038.1, 2023.

Matrosov, S. Y., C. Campbell, D. Kingsmill, and E. Sukovich, Assessing Snowfall Rates from X-Band Radar Reflectivity Measurements, *Journal of Atmospheric and Oceanic Technology*, *26*(11), 2324–2339, doi:10.1175/2009JTECHA1238.1, 2009.

McCabe, G. J., Relationships between atmospheric circulation and snowpack in the Gunnison River basin, Colorado, *Journal of Hydrology*, *157*(1), 157–175, doi:10.1016/0022-1694(94)90103-1, 1994.

McGinnis, D. L., Synoptic controls on upper Colorado River basin snowfall, *International Journal of Climatology*, *20*(2), 131–149, doi:10.1002/(SICI)1097-0088(200002)20:2⟨131::AID-JOC465⟩3.0.CO;2-H, _eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291097-0088%28200002%2920%3A2%3C131%3A%3AAID-JOC465%3E3.0.CO%3B2-H, 2000.

McGurk, B. J., Precipitation and snow water equivalent sensors: an evaluation, in *54th Annual Western Snow Conference*, vol. 54, pp. 71–80, Western Snow Conference, Phoenix, Arizona, 1986.

McNamara, J. P., D. Chandler, M. Seyfried, and S. Achet, Soil moisture states, lateral flow, and streamflow generation in a semi-arid, snowmelt-driven catchment, *Hydrological Processes*, *19*(20), 4023–4038, doi:10.1002/hyp.5869, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.5869, 2005.

Melton, F. S., et al., OpenET: Filling a Critical Data Gap in Water Management for the Western United States, *JAWRA Journal of the American Water Resources Association*, *58*(6), 971–994, doi:10.1111/1752-1688.12956, 2022.

Meyers, T., SPLASH 10 Meter Eddy Covariance Data for 2021, 2022, and 2023, doi:10.5281/zenodo.10558503, 2024.

Millar, C. I., and R. D. Westfall, Rock glaciers and related periglacial landforms in the

Sierra Nevada, CA, USA; inventory, distribution and climatic relationships, *Quaternary International*, *188*(1), 90–104, doi:10.1016/j.quaint.2007.06.004, 2008.

Milly, P. C. D., and K. A. Dunne, Colorado River flow dwindles as warming-driven loss of reflective snow energizes evaporation, *Science*, *367*(6483), 1252–1255, doi:10.1126/science.aay9187, 2020.

Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer, Stationarity Is Dead: Whither Water Management?, *Science*, *319*(5863), 573–574, doi:10.1126/science.1151915, 2008.

Mizukami, N., and M. B. Smith, Analysis of inconsistencies in multi-year gridded quantitative precipitation estimate over complex terrain and its impact on hydrologic modeling, *Journal of Hydrology*, *428-429*, 129–141, doi:10.1016/j.jhydrol.2012.01.030, 2012.

Modi, P. A., E. E. Small, J. Kasprzyk, and B. Livneh, Investigating the Role of Snow Water Equivalent on Streamflow Predictability during Drought, *Journal of Hydrometeorology*, *23*(10), 1607–1625, doi:10.1175/JHM-D-21-0229.1, 2022.

Mott, R., V. Vionnet, and T. Grünewald, The Seasonal Snow Cover Dynamics: Review on Wind-Driven Coupling Processes, *Frontiers in Earth Science*, *6*, 197, 2018.

Moulin, L., E. Gaume, and C. Obled, Uncertainties on mean areal precipitation: assessment and impact on streamflow simulations, *Hydrology and Earth System Sciences*, *13*(2), 99–114, doi:10.5194/hess-13-99-2009, 2009.

Musselman, K. N., J. W. Pomeroy, R. L. H. Essery, and N. Leroux, Impact of windflow calculations on simulations of alpine snow accumulation, redistribution and ablation, *Hydrological Processes*, *29*(18), 3983–3999, doi:10.1002/hyp.10595, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.10595, 2015.

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

Muñoz-Sabater, J., et al., ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, *Earth System Science Data*, *13*(9), 4349–4383, doi:10.5194/essd-13-4349-2021, 2021.

Nathan, R. J., and T. A. McMahon, Evaluation of automated techniques for base flow and recession analyses, *Water Resources Research*, *26*(7), 1465–1473, doi:10.1029/WR026i007p01465, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/WR026i007p01465, 1990.

Natural Resources Conservation Service,, Snowpack Telemetry Network (SNOTEL) for Butte (380) station, 2025.

Newman, A. J., M. P. Clark, A. Winstral, D. Marks, and M. Seyfried, The Use of Similarity Concepts to Represent Subgrid Variability in Land Surface Models: Case Study in a Snowmelt-Dominated Watershed, doi:10.1175/JHM-D-13-038.1, 2014.

Newsom, R., C. Sivaraman, T. Shippert, and L. Riihimaki, Doppler Lidar Vertical Velocity Statistics Value-Added Product, *Tech. Rep. DOE/SC-ARM-TR-149*, U.S. Department of Energy, Atmospheric Radiation Measurement user facility, Richland, Washington., doi:10.2172/1238068, 2019.

Nippgen, F., B. L. McGlynn, R. E. Emanuel, and J. M. Vose, Watershed memory at the Coweeta Hydrologic Laboratory: The effect of past precipitation and storage on hydrologic response, *Water Resources Research*, *52*(3), 1673–1695, doi:10.1002/2015WR018196, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2015WR018196, 2016.

Nitu, R., WMO-CIMO survey on the methods and instruments for the measurement of solid precipitation at Automated Weather Stations, in *Proceedings of the WMO CIMO TECO Conference*, Helsinki, Finland, 2010.

NOAA, Precipitation frequency data server, 2017.

Nolin, A. W., E. A. Sproles, D. E. Rupp, R. L. Crumley, M. J. Webb, R. T. Palomaki, and E. Mar, New snow metrics for a warming world, *Hydrological Processes*, *35*(6), e14,262, doi:10.1002/hyp.14262, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.14262, 2021.

Nystuen, J. A., J. R. Proni, P. G. Black, and J. C. Wilkerson, A Comparison of Automatic Rain Gauges, *Journal of Atmospheric and Oceanic Technology*, *13*(1), 62–73, doi:10.1175/1520-0426(1996)013⟨0062:ACOARG⟩2.0.CO;2, 1996.

O'Brien, J. R., M. Grover, R. C. Jackson, Z. S. Sherman, S. M. Collis, A. Theisen, B. A. Raut, M. Tuftedal, and D. Feldman, Colorado State University (CSU) X-Band Precipitation Radar Plan Position Indicator Data Processed with Corrected Moments in Antenna Coordinates (CMAC) Value-Added Product Report, *ARM Tech. Rep. DOE/SC-ARM–TR-313*, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States). Atmospheric Radiation Measurement (ARM) Data Center, doi:10.2172/2483333, 2024.

Opie, S., R. G. Taylor, C. M. Brierley, M. Shamsudduha, and M. O. Cuthbert, Climate–groundwater dynamics inferred from GRACE and the role of hydraulic memory, *Earth System Dynamics*, *11*(3), 775–791, doi:10.5194/esd-11-775-2020, 2020.

Orth, R., and S. I. Seneviratne, Propagation of soil moisture memory to streamflow and evapotranspiration in Europe, *Hydrology and Earth System Sciences*, *17*(10), 3895–3911, doi:10.5194/hess-17-3895-2013, 2013.

Pagano, T., and D. Garen, A Recent Increase in Western U.S. Streamflow Variability and Persistence, *Journal of Hydrometeorology*, *6*(2), 173–179, doi:10.1175/JHM410.1, 2005.

Pagano, T., D. Garen, and S. Sorooshian, Evaluation of Official Western U.S. Seasonal Water Supply Outlooks, 1922–2002, *Journal of Hydrometeorology*, *5*(5), 896–909, doi:10.1175/1525-7541(2004)005⟨0896:EOOWUS⟩2.0.CO;2, 2004.

Painter, T. H., et al., 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, doi:10.1016/j.rse.2016.06.018, 2016.

Palumbo, D., S. Gangopadhyay, and U. Lall, Precipitation, moderated by spring temperature and vegetation, drives runoff efficiency in the Upper Colorado River Basin, USA, *Communications Earth & Environment*, *7*(1), 115, doi:10.1038/s43247-025-03136-w, 2025.

Pedregosa, F., et al., Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, *12*(85), 2825–2830, 2011.

Pierre, A., S. Jutras, C. Smith, J. Kochendorfer, V. Fortin, and F. Anctil, Evaluation of Catch Efficiency Transfer Functions for Unshielded and Single-Alter-Shielded Solid Precipitation Measurements, *Journal of Atmospheric and Oceanic Technology*, doi:10.1175/JTECH-D-18-0112.1, 2019.

Poggio, L., L. M. de Sousa, N. H. Batjes, G. B. M. Heuvelink, B. Kempen, E. Ribeiro, and D. Rossiter, SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty, *SOIL*, *7*(1), 217–240, doi:10.5194/soil-7-217-2021, 2021.

Pomeroy, J., and D. Gray, Sensitivity of snow relocation and sublimation to climate and surface vegetation, *IAHS Publications-Series of Proceedings and Reports-Intern Assoc Hydrological Sciences*, *223*, 213–226, publisher: Wallingford [Oxfordshire]: IAHS, 1981-, 1994.

Prather, R. M., N. Underwood, R. M. Dalton, B. Barr, and B. D. Inouye, Climate data from the Rocky Mountain Biological Laboratory (1975–2022), *Ecology*, *104*(11), e4153, doi:10.1002/ecy.4153, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ecy.4153, 2023.

Qin, Y., et al., Agricultural risks from changing snowmelt, *Nature Climate Change*, *10*(5), 459–465, doi:10.1038/s41558-020-0746-8, number: 5, 2020.

Ramírez Molina, A. A., G. Tootle, Z. Sun, and J. Fu, The Future of Snowpack Drought in the Upper Colorado River Basin (USA), *Hydrology*, *13*(4), 100, doi:10.3390/hydrology13040100, 2026.

Rapp, G. A., L. E. Condon, and K. H. Markovich, Sensitivity of Simulated Mountain Block Hydrology to Subsurface Conceptualization, *Water Resources Research*, *56*(10), e2020WR027,714, doi:10.1029/2020WR027714, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020WR027714, 2020.

Rasmussen, R., et al., How Well Are We Measuring Snow: The NOAA/FAA/NCAR Winter Precipitation Test Bed, *Bulletin of the American Meteorological Society*, doi:10.1175/BAMS-D-11-00052.1, 2012.

Rasouli, K., J. W. Pomeroy, and D. G. Marks, Snowpack sensitivity to perturbed climate in a cool mid-latitude mountain catchment, *Hydrological Processes*, *29*(18), 3925–3940, doi:10.1002/hyp.10587, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.10587, 2015.

Reba, M. L., T. E. Link, D. Marks, and J. Pomeroy, An assessment of corrections for eddy covariance measured turbulent fluxes over snow in mountain environments, *Water Resources Research*, *45*(4), W00D38, doi:10.1029/2008WR007045, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2008WR007045, 2009.

Reba, M. L., J. Pomeroy, D. Marks, and T. E. Link, Estimating surface sublimation losses from snowpacks in a mountain catchment using eddy covariance and turbulent transfer calculations, *Hydrological Processes*, *26*(24), 3699–3711, doi:10.1002/hyp.8372, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.8372, 2012.

Rosenberg, E. A., E. A. Clark, A. C. Steinemann, and D. P. Lettenmaier, On the contribution of groundwater storage to interannual streamflow anomalies in the Colorado River basin, *Hydrology and Earth System Sciences*, *17*(4), 1475–1491, doi:10.5194/hess-17-1475-2013, 2013.

Savina, M., B. Schäppi, P. Molnar, P. Burlando, and B. Sevruk, Comparison of a tipping-bucket and electronic weighing precipitation gage for snowfall, *Atmospheric Research*, *103*, 45–51, doi:10.1016/j.atmosres.2011.06.010, 2012.

Scaff, L., S. A. Krogh, K. Musselman, A. Harpold, Y. Li, M. Lillo-Saavedra, R. Oyarzún, and R. Rasmussen, The Impacts of Changing Winter Warm Spells on Snow Ablation Over Western North America, *Water Resources Research*, *60*(5), 2023WR034,492, doi:10.1029/2023WR034492, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2023WR034492, 2024.

Schaap, M. G., F. J. Leij, and M. T. van Genuchten, rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions, *Journal of Hydrology*, *251*(3), 163–176, doi:10.1016/S0022-1694(01)00466-8, 2001.

Schmidt, R. A., *Sublimation of Wind-transported Snow: A Model*, Rocky Mountain Forest and Range Experiment Station, Forest Service, U.S. Department of Agriculture, google-Books-ID: ndB586FaY2UC, 1972.

Schmidt, R. A., Vertical profiles of wind speed, snow concentration, and humidity in blowing snow, *Boundary-Layer Meteorology*, *23*(2), 223–246, doi:10.1007/BF00123299, 1982.

Schnepper, T., J. Groh, H. H. Gerke, B. Reichert, and T. Pütz, Evaluation of precipitation measurement methods using data from a precision lysimeter network, *Hydrology and Earth System Sciences*, *27*(17), 3265–3292, doi:10.5194/hess-27-3265-2023, 2023.

Schotanus, P., F. Nieuwstadt, and H. De Bruin, Temperature measurement with a sonic anemometer and its application to heat and moisture fluxes, *Boundary-Layer Meteorology*, *26*(1), 81–93, doi:10.1007/BF00164332, 1983.

Schwat, E., D. Hogan, K. T. Paw U, C. J. Cox, B. J. Butterworth, E. Gutmann, J. A. Vano, and J. D. Lundquist, Estimating Snow Sublimation in Complex Terrain: A Season of Intensive Field Measurements and the Role of Vertical Water Vapor Flux Divergence, *Journal of Hydrometeorology*, *26*(10), 1455–1473, doi:10.1175/JHM-D-25-0022.1, 2025.

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

Sevruk, B., Regional dependecy of precipitation-altitude relationships in the Swiss Alps, *Climatic Change*, *36*(3), 355–369, doi:10.1023/A:1005302626066, 1997.

Sexstone, G., D. Clow, D. I. Stannard, and S. R. Fassnacht, Comparison of methods for quantifying surface sublimation over seasonally snow-covered terrain, *Hydrological Processes*, *30*(19), 3373–3389, doi:10.1002/hyp.10864, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.10864, 2016.

Sexstone, G., D. Clow, S. Fassnacht, G. Liston, C. Hiemstra, J. Knowles, and C. Penn, Snow Sublimation in Mountain Environments and Its Sensitivity to Forest Disturbance and Climate Warming, *Water Resources Research*, *54*, 1191 – 1211, doi:10.1002/2017WR021172, 2018.

Sharma, V., F. Gerber, and M. Lehning, Introducing CRYOWRF v1.0: multiscale atmospheric flow simulations with advanced snow cover modelling, *Geoscientific Model Development*, *16*(2), 719–749, doi:10.5194/gmd-16-719-2023, 2023.

Shestakova, A. A., D. G. Chechin, C. Lüpkes, J. Hartmann, and M. Maturilli, The foehn effect during easterly flow over Svalbard, *Atmospheric Chemistry and Physics*, *22*(2), 1529–1548, doi:10.5194/acp-22-1529-2022, 2022.

Shukla, K. K., D. V. Phanikumar, R. K. Newsom, N. Ojha, K. NiranjanKumar, N. Singh, S. Sharma, V. R. Kotamarthi, and K. K. Kumar, Investigations of vertical wind variations at a mountain top in the Himalaya using Doppler Lidar observations and model simulations, *Journal of Atmospheric and Solar-Terrestrial Physics*, *183*, 76–85, doi:10.1016/j.jastp.2018.12.011, 2019.

Sigmund, A., J. Dujardin, F. Comola, V. Sharma, H. Huwald, D. B. Melo, N. Hirasawa, K. Nishimura, and M. Lehning, Evidence of Strong Flux Underestimation by Bulk Parametrizations During Drifting and Blowing Snow, *Boundary-Layer Meteorology*, *182*(1), 119–146, doi:10.1007/s10546-021-00653-x, 2022.

Siirila-Woodburn, E. R., et al., 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, doi:10.1038/s43017-021-00219-y, number: 11, 2021.

Siirila-Woodburn, E. R., et al., Warming and snow loss increase reliance on old groundwater in a Colorado River headwater, *Nature Geoscience*, pp. 1–7, doi:10.1038/s41561-026-01945-y, 2026.

Silverman, N. L., M. P. Maneta, S.-H. Chen, and J. T. Harper, Dynamically downscaled winter precipitation over complex terrain of the Central Rockies of Western Montana, USA, *Water Resources Research*, *49*(1), 458–470, doi:10.1029/2012WR012874, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2012WR012874, 2013.

Slater, A. G., et al., The Representation of Snow in Land Surface Schemes: Results from PILPS 2(d), *Journal of Hydrometeorology*, *2*(1), 7–25, doi:10.1175/1525-7541(2001)002⟨0007:TROSIL⟩2.0.CO;2, 2001.

Slosson, J. R., I. J. Larsen, M. J. Winnick, J. M. Marmolejo-Cossío, and K. H. Williams, Linking Surface Processes, Solute Generation, and CO2 Budgets Across Lithological and Land Cover Gradients in Rocky Mountain Watersheds, *Water Resources Research*, *61*(4), e2023WR036,850, doi:10.1029/2023WR036850, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2023WR036850, 2025.

Smith, C. D., A. Ross, J. Kochendorfer, M. E. Earle, M. Wolff, S. Buisán, Y.-A. Roulet, and T. Laine, Evaluation of the WMO Solid Precipitation Intercomparison Experiment (SPICE) transfer functions for adjusting the wind bias in solid precipitation measurements, *Hydrology and Earth System Sciences*, *24*(8), 4025–4043, doi:10.5194/hess-24-4025-2020, 2020.

Smith, M. B., D.-J. Seo, V. I. Koren, S. M. Reed, Z. Zhang, Q. Duan, F. Moreda, and S. Cong, The distributed model intercomparison project (DMIP): motivation and experiment design, *Journal of Hydrology*, *298*(1), 4–26, doi:10.1016/j.jhydrol.2004.03.040, 2004.

Steiner, M., J. A. Smith, S. J. Burges, C. V. Alonso, and R. W. Darden, Effect of bias adjustment and rain gauge data quality control on radar rainfall estimation, *Water Resources Research*, *35*(8), 2487–2503, doi:10.1029/1999WR900142, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/1999WR900142, 1999.

Stiperski, I., and M. W. Rotach, On the Measurement of Turbulence Over Complex Mountainous Terrain, *Boundary-Layer Meteorology*, *159*(1), 97–121, doi:10.1007/s10546-015-0103-z, 2016.

Stock, G., R. S. Anderson, T. H. Painter, B. Henn, and J. D. Lundquist, Impending Loss of Little Ice Age Glaciers in Yosemite National Park, Seattle, Washington, USA, 2017.

Storn, R., and K. Price, Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces, *Journal of Global Optimization*, *11*(4), 341–359, doi:10.1023/A:1008202821328, 1997.

Strasser, U., M. Bernhardt, M. Weber, G. E. Liston, and W. Mauser, Is snow sublimation important in the alpine water balance?, *The Cryosphere*, *2*(1), 53–66, doi: 10.5194/tc-2-53-2008, 2008.

Sturm, M., and A. M. Wagner, Using repeated patterns in snow distribution modeling: An Arctic example, *Water Resources Research*, *46*(12), doi:10.1029/2010WR009434, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2010WR009434, 2010.

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

Sugiura, K., D. Yang, and T. Ohata, Systematic error aspects of gauge-measured solid precipitation in the Arctic, Barrow, Alaska, *Geophysical Research Letters*, *30*(4), doi:10.1029/2002GL015547, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2002GL015547, 2003.

Suri, A., and S. Azad, Optimal placement of rain gauge networks in complex terrains for monitoring extreme rainfall events: a review, *Theoretical and Applied Climatology*, *155*(4), 2511–2521, doi:10.1007/s00704-024-04856-3, 2024.

Survey, U. G., National Water Information System data available on the World Wide Web (USGS Water Data for the Nation),, 2016.

Survey, U. G., 3D Elevation Program 1-Arc Second Resolution Digital Elevation Model, 2022.

Sutanto, S. J., and H. A. J. Van Lanen, Catchment memory explains hydrological drought forecast performance, *Scientific Reports*, *12*(1), 2689, doi:10.1038/s41598-022-06553-5, 2022.

Svoma, B. M., Difficulties in Determining Snowpack Sublimation in Complex Terrain at the Macroscale, *Advances in Meteorology*, *2016*, 9695,757, doi:10.1155/2016/9695757, 2016.

Tarboton, D. G., Terrain Analysis Using Digital Elevation Models (TauDEM), 2002.

Thomas, B. F., and J. S. Famiglietti, Identifying Climate-Induced Groundwater Depletion in GRACE Observations, *Scientific Reports*, *9*(1), 4124, doi:10.1038/s41598-019-40155-y, number: 1, 2019.

Thériault, J. M., R. Rasmussen, K. Ikeda, and S. Landolt, Dependence of Snow Gauge Collection Efficiency on Snowflake Characteristics, *Journal of Applied Meteorology and Climatology*, doi:10.1175/JAMC-D-11-0116.1, 2012.

Thériault, J. M., K. L. Rasmussen, T. Fisico, R. E. Stewart, P. Joe, I. Gultepe, M. Clément, and G. A. Isaac, Weather observations on Whistler Mountain during five storms, *Pure and Applied Geophysics*, *171*(1), 129–155, doi:10.1007/s00024-012-0590-5, 2014.

Toebes, C., and D. D. Strang, On recession curves, 1. Recession equations, *J. Hydrol. New Zealand*, *3*(2), 2–15, https://webstatic.niwa.co.nz/library/HHPP8.pdf, 1964.

Tran, A. P., J. Rungee, B. Faybishenko, B. Dafflon, and S. S. Hubbard, Assessment of Spatiotemporal Variability of Evapotranspiration and Its Governing Factors in a Mountainous Watershed, *Water*, *11*(2), 243, doi:10.3390/w11020243, number: 2, 2019.

Troin, M., J.-L. Martel, R. Arsenault, and F. Brissette, Large-sample study of uncertainty of hydrological model components over North America, *Journal of Hydrology*, *609*, 127,766, doi:10.1016/j.jhydrol.2022.127766, 2022.

Trouvilliez, A., F. Naaim-Bouvet, H. Bellot, C. Genthon, and H. Gallée, Evaluation of the FlowCapt Acoustic Sensor for the Aeolian Transport of Snow, *Jounral of Atmospheric and Oceanic Technology*, *32*, 1630–1641, doi:10.1175/JTECH-D-14-00104.1, 2015.

USDA, Air and Water Database, 2024.

USDA NRCS, Snowpack Telemetry (SNOTEL) report generator., 2026.

Vano, J. A., T. Das, and D. P. Lettenmaier, Hydrologic Sensitivities of Colorado River Runoff to Changes in Precipitation and Temperature, *Journal of Hydrometeorology*, *13*(3), 932–949, doi:10.1175/JHM-D-11-069.1, 2012.

Vionnet, V., G. Guyomarc'h, F. Naaim Bouvet, E. Martin, Y. Durand, H. Bellot, C. Bel, and P. Puglièse, Occurrence of blowing snow events at an alpine site over a 10-year period: Observations and modelling, *Advances in Water Resources*, *55*, 53–63, doi:10.1016/j.advwatres.2012.05.004, 2013.

Viviroli, D., M. Kummu, M. Meybeck, M. Kallio, and Y. Wada, Increasing dependence of lowland populations on mountain water resources, *Nature Sustainability*, *3*(11), 917–928, doi:10.1038/s41893-020-0559-9, 2020.

Volk, J. M., et al., Assessing the accuracy of OpenET satellite-based evapotranspiration data to support water resource and land management applications, *Nature Water*, *2*(2), 193–205, doi:10.1038/s44221-023-00181-7, 2024.

Wagner, D. N., et al., Snowfall and snow accumulation during the MOSAiC winter and spring seasons, *The Cryosphere*, *16*(6), 2373–2402, doi:10.5194/tc-16-2373-2022, 2022.

Wainwright, H. M., B. Dafflon, E. R. Siirila-Woodburn, N. Falco, Y. Wu, I. Breckheimer, and R. W. H. Carroll, Model and remote-sensing-guided experimental design and hypothesis generation for monitoring snow-soil–plant interactions, *Frontiers in Water*, *5*, doi:10.3389/frwa.2023.1220146, 2024.

Wang, Z., S. Ghimire, K. M. Whitney, G. Mascaro, M. Xiao, H. Yue, and E. R. Vivoni, Revisiting the application of variable infiltration capacity (VIC) model in the Colorado River Basin using SMAP and GRACE, *Scientific Reports*, *16*(1), 15,890, doi:10.1038/s41598-026-47430-9, 2026.

Warren, S. G., Optical properties of snow, *Reviews of Geo-*

physics, *20*(1), 67–89, doi:10.1029/RG020i001p00067, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/RG020i001p00067, 1982.

Wayand, N. E., J. Stimberis, J. P. Zagrodnik, C. F. Mass, and J. D. Lundquist, Improving simulations of precipitation phase and snowpack at a site subject to cold air intrusions: Snoqualmie Pass, WA, *Journal of Geophysical Research: Atmospheres*, *121*(17), 9929–9942, doi:10.1002/2016JD025387, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016JD025387, 2016.

Webb, E. K., G. I. Pearman, and R. Leuning, Correction of flux measurements for density effects due to heat and water vapour transfer, *Quarterly Journal of the Royal Meteorological Society*, *106*(447), 85–100, doi:10.1002/qj.49710644707, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49710644707, 1980.

Wen, Y., A. Behrangi, B. Lambrigtsen, and P.-E. Kirstetter, Evaluation and Uncertainty Estimation of the Latest Radar and Satellite Snowfall Products Using SNOTEL Measurements over Mountainous Regions in Western United States, *Remote Sensing*, *8*(11), doi:10.3390/rs8110904, company: Multidisciplinary Digital Publishing Institute Distributor: Multidisciplinary Digital Publishing Institute Institution: Multidisciplinary Digital Publishing Institute Label: Multidisciplinary Digital Publishing Institute, 2016.

Wheeler, K. G., B. Udall, J. Wang, E. Kuhn, H. Salehabadi, and J. C. Schmidt, What will it take to stabilize the Colorado River?, *Science*, *377*(6604), 373–375, doi:10.1126/science.abo4452, 2022.

Whiteman, C. D., Breakup of Temperature Inversions in Deep Mountain Valleys: Part I. Observations., *Journal of Applied Meteorology*, *21*, 270–289, doi:10.1175/1520-0450(1982)021⟨0270:BOTIID⟩2.0.CO;2, 1982.

Whiteman, C. D., and J. C. Doran, The Relationship between Overlying Synoptic-Scale

Flows and Winds within a Valley, *Journal of Applied Meteorology and Climatology*, *32*(11), 1669–1682, doi:10.1175/1520-0450(1993)032⟨1669:TRBOSS⟩2.0.CO;2, 1993.

Whiteman, C. D., and T. B. McKee, Breakup of Temperature Inversions in Deep Mountain Valleys: Part II. Thermodynamic Model, *Journal of Applied Meteorology and Climatology*, *21*(3), 290–302, doi:10.1175/1520-0450(1982)021⟨0290:BOTIID⟩2.0.CO;2, 1982.

Wilczak, J. M., S. P. Oncley, and S. A. Stage, Sonic Anemometer Tilt Correction Algorithms, *Boundary-Layer Meteorology*, *99*(1), 127–150, doi:10.1023/A:1018966204465, 2001.

Wittenberg, H., Baseflow recession and recharge as nonlinear storage processes, *Hydrological Processes*, *13*(5), 715–726, doi:10.1002/(SICI)1099-1085(19990415)13:5⟨715::AID-HYP775⟩3.0.CO;2-N, 1999.

Wolf, M. A., L. R. Jamison, D. K. Solomon, C. Strong, and P. D. Brooks, Multi-Year Controls on Groundwater Storage in Seasonally Snow-Covered Headwater Catchments, *Water Resources Research*, *59*(6), e2022WR033,394, doi:10.1029/2022WR033394, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2022WR033394, 2023.

Wolf, M. A., L. Jamison, C. Strong, and P. D. Brooks, Periodic Variability in Baseflow in Headwater Streams of the Upper Colorado River: Implications for Runoff Efficiency, *JAWRA Journal of the American Water Resources Association*, *61*(2), e70,017, doi:10.1111/1752-1688.70017, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1752-1688.70017, 2025.

Xia, Y., D. Mocko, M. Huang, B. Li, M. Rodell, K. E. Mitchell, X. Cai, and M. B. Ek, Comparison and Assessment of Three Advanced Land Surface Models in Simulating Terrestrial Water Storage Components over the United States, *Journal of Hydrometeorology*, *18*(3), 625–649, doi:10.1175/JHM-D-16-0112.1, 2017.

Xiao, M., B. Udall, and D. P. Lettenmaier, On the Causes of Declining Colorado River

Streamflows, *Water Resources Research*, *54*(9), 6739–6756, doi:10.1029/2018WR023153, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018WR023153, 2018.

Xie, Z., Y. Ma, W. Ma, Z. Hu, G. Sun, and Y. Wang, Analysis of Multiyear Blowing Snow Occurrences in the French Alps, *Journal of Hydrometeorology*, *24*, 3–19, doi:10.1175/JHM-D-22-0029.1, 2022.

Zhang, Y., K. Suzuki, T. Kadota, and T. Ohata, Sublimation from snow surface in southern mountain taiga of eastern Siberia, *Journal of Geophysical Research: Atmospheres*, *109*, D21,103, doi:10.1029/2003JD003779, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2003JD003779, 2004.

Zhang, Z., K. Fowler, and M. Peel, Can Conceptual Rainfall-Runoff Models Capture Multi-Annual Storage Dynamics?, *Water Resources Research*, *62*(3), e2025WR042,226, doi:10.1029/2025WR042226, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2025WR042226, 2026.

Zhu, Z., and D. Wang, Laser Disdrometer (LDIS) Instrument Handbook, *ARM Tech. Rep. DOE/SC-ARM-TR-137*, Department of Energy Atmospheric Radiation Measurement (ARM) User Facility, 2025.