Integrating Observations and Modeling to Understand

Snowmelt-Driven Streamflow Dynamics During the Colorado River Basin Drought

by

Swastik Ghimire

A Thesis

Presented in Partial Fulfillment

of the Requirements for the Degree

Master of Science

Approved April 2025 by the

Graduate Supervisory Committee:

Enrique R. Vivoni, Chair

Giuseppe Mascaro

Ruijie Zeng

ARIZONA STATE UNIVERSITY

May 2025

ABSTRACT

The Colorado River Basin (CRB), a critical freshwater source for the southwestern United States, has experienced a prolonged drought since 2000, reducing water supply and lowering streamflow efficiency. Despite near-average snowpack conditions measured by snow water equivalent (*SWE*) in recent years, observed streamflow has been markedly below expected levels, raising a critical question: Where does all the snowmelt go? This research employs the Variable Infiltration Capacity (VIC) model and observations from instrument networks and remote sensing to investigate three hypothesized factors causing this mismatch: inadequate spatial representativeness of ground networks, springtime climate conditions, and antecedent soil moisture conditions. First, it was found that the April 1st *SWE* averaged across 104 SNOTEL stations used by the Colorado Basin River Forecasting Center (CBRFC) may under- or over-estimate the overall snowpack in the Upper CRB (UCRB) in wet and dry years, respectively, as these sites only represent areas with a high mean *SWE* (>40 mm) covering merely 30% of the snow-producing regions. Results also indicate that hot and dry April-May-June (AMJ) weather conditions can substantially reduce the positive effects of average *SWE* on April 1st, with variations in AMJ climate alone causing a variation of 11.33 to 19.64 km^3 in annual streamflow volumes. Additionally, dry soils before snow onset can diminish the positive effects of an average snowpack, with initial soil moisture conditions producing a variation of 9.44 to 17.89 km^3 in annual streamflow when other factors remain constant. Prolonged and intense droughts deplete deep soil moisture, especially in the UCRB, which recovers slowly over multiple years and has reduced baseflow over time. Together, under a constant snowpack, it was found that antecedent soil moisture accounts for approximately 74% of variability for UCRB streamflow, followed by AMJ precipitation (~24%) and temperature (~2%). Conversely, precipitation (~72%) is the primary driver of streamflow variability in the Lower CRB. These findings highlight the complex interplay of factors affecting the snow-to-flow relationship in the CRB and suggest the need for more comprehensive monitoring and forecasting approaches that incorporate soil moisture dynamics and seasonal climate predictions.

ACKNOWLEDGMENTS

This work was completed under the project "Managing the Colorado River as an Infrastructure Asset: Fusing Remote Sensing and Numerical Modeling in the Operations of the Central Arizona Project" funded by the NASA Applied Sciences Water Resources Program. I am sincerely grateful to my supervisor, Dr. Enrique Vivoni, for his invaluable guidance, unwavering support, and expertise throughout this research journey. His mentorship has been instrumental in shaping not only this thesis but also my growth as a researcher. I am deeply grateful to my mentor, Dr. Zhaocheng Wang, whose insightful feedback, patience, and encouragement have been vital to the completion of this work. I extend my appreciation to my committee members, Dr. Giuseppe Mascaro and Dr. Ruijie Zeng, for their constructive criticism, thought-provoking questions, and valuable suggestions that significantly enhanced the quality of this research.

This would not have been possible without my family's unconditional love and support back in Nepal. Despite the physical distance, their faith in my abilities and emotional support have been my constant strength and motivation. Special thanks to Suraj Tiwari, Kshitij Dahal, and Shraddha Sharma, who have been more than friends—they have been family through thick and thin. Their unwavering support, late-night discussions, and encouragement during challenging times have been invaluable. I am also thankful to the entire Nepali community here, whose presence made this foreign land feel like home. I want to acknowledge my colleagues and friends who made the workplace a joy to be in each day: Jose, Xinyu, Akram, Nour, Laura, Aashish, Binaya, and Johnny.

Their camaraderie, intellectual discussions, and shared laughter created an environment conducive to creativity and productivity. I would also like to mention my friends, who continue to be my pillars over the last 10 years: Tek, Shlok, Yubin, and especially Supriya, who has been with me through all the highs and lows and continues to tolerate the child in me patiently. Despite the distance, her presence remains a constant source of strength in my life.

I would like to express my gratitude to Research Computing at Arizona State University for providing high-performance computing resources that have supported the research findings presented in this paper. Finally, I am grateful to the staff at the Central Arizona Project (CAP) for supporting this research. Specifically, I am grateful to Vineetha Kartha, Nolie Templeton, Orestes Morfin, and Joshua Randall, who always supported my work and provided valuable feedback. This journey has been transformative for me as a student, researcher, and person, and I am thankful to everyone who has participated in it.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

1. INTRODUCTION

The hydrology of the western United States is characterized by seasonal mountain snowpacks, with an annual water storage of approximately 200 km^3 or 162 million-acre-feet (MAF; see Appendix A for all acronyms) (Siirila-Woodburn et al., 2021; Sturm et al., 2017). Within this hydrological system, the peak snow water equivalent (*SWE*) occurs in late winter to early spring, followed by a temperature-driven ablation season that sustains streamflow and groundwater recharge, ultimately contributing to 70% of the total runoff in the mountainous areas of the western US (Li et al., 2017). The Colorado River Basin (CRB), which encompasses the high mountain ranges of the Rocky Mountains to the arid desert plains of the south, exemplifies this snow-dependent regime as approximately 85% of the total streamflow originates from snowmelt in the headwater areas (Christensen et al., 2004; Lukas & Payton, 2020). Furthermore, peak snowpack in the CRB stores approximately 18 MAF of water, equivalent to 70% of Lake Powell's capacity, highlighting the critical role of peak snowpack in sustaining streams, reservoirs, and overall water availability in the basin (Lukas & Payton, 2020; Miller & Piechota, 2011).

The CRB stands as a critical lifeline for the American Southwest, providing a major source of water supply for over 40 million people, vast agricultural lands, and intricate industrial and ecological systems across seven U.S. states and northern parts of Mexico (Christensen et al., 2004; Wheeler et al., 2022). However, since 2000, the basin has experienced an unprecedented, prolonged drought with cascading effects on soil conditions and streamflow volumes (Udall & Overpeck, 2017). Persistent multi-year low snowpack magnitudes have also reduced inflows to reservoirs and led to groundwater depletion (Castle et al., 2014). Further, the long-term snow-to-flow relation, once a reliable predictor of streamflow, has deviated from historical observations, suggesting fundamental shifts in hydrological processes in the region (Hogan & Lundquist, 2024; Udall & Overpeck, 2017). As a result, streamflow has become increasingly unpredictable under current drought conditions, and studies suggest that a future with low to no snow will further exacerbate these challenges (Livneh et al., 2017; Livneh & Badger, 2020).

Snow measurements have traditionally guided water management in the basin, with 88% of Upper Colorado River Basin (UCRB) stakeholders ranking ground-based snow measurements (SNOTEL) as their most valuable metric while making decisions about drought (McNie, 2014). Institutions responsible for water supply forecasts, such as the Natural Resources Conservation Service (NRCS) and the Colorado Basin River Forecast Center (CBRFC), draw on a network of monitoring stations, including SNOTEL sites, to estimate how much runoff to expect (Lukas & Payton, 2020; USDA, 2024). However, this relationship between snowpack and streamflow, especially when relying on snow information from SNOTEL alone, has become increasingly less predictable as the relationship between these variables is changing under drought conditions. This shifting relationship is clearly illustrated in Figure 1, which compares relative April 1 SWE measurements from the 104 CBRFC SNOTEL stations with relative Lake Powell inflows (both as percentages of 1991-2020 means, a standard approach used by the water managers). While pre-2000 data points (blue) frequently plot above the 1:1 line, post-

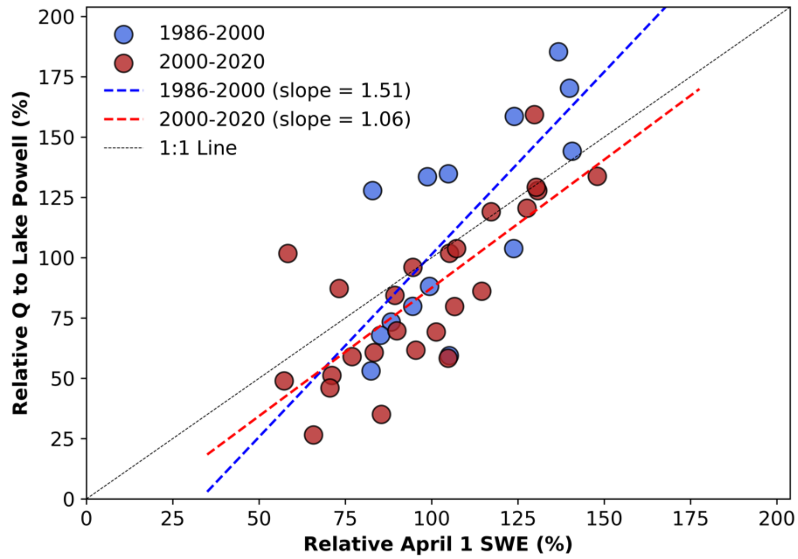


Figure 1. Relationship between relative April 1 SWE from 104 SNOTEL stations and relative inflow to Lake Powell (% of 1991-2020 mean). Blue and red circles represent pre and post-2000 data points, respectively. The black line shows the regression relationship across the entire record (1986-2020).

2000 measurements (red) predominantly fall below it, demonstrating that equivalent snowpack percentages generated reduced streamflow volumes during the post-2000 drought period. This quantifiable decrease in runoff efficiency indicates a significant alteration in the basin's hydrological processes that undermines the predictive capability of forecasting methods based primarily on SNOTEL network data.

Recent years have further exposed limitations in snow-based forecast approaches. For instance, in 2020 and 2021, near-average snowpack (104% and 84% of the 1991-2020 conditions, respectively) did not translate to expected streamflow volumes (71% and 51% of the 1991-2020 normal, respectively) in the UCRB, resulting in lower than expected inflows into Lake Powell (58% and 35% of normal, respectively) (CBRFC, 2024). Following this, the U.S. Bureau of Reclamation (USBR) declared its first Tier 1 shortage on the Colorado River in August 2021, implementing cuts in January 2022 (ADWR, 2022b). Arizona faced an 18% reduction (512,000 acre-feet), with smaller cuts to Nevada and Mexico. By August 2022, worsening conditions triggered a Tier 2 status, deepening Arizona's cuts to 21% and expanding restrictions to additional water users (ADWR, 2022a). Water cuts during periods of adequate snowpack have driven water managers to reevaluate their hydrologic understanding of the CRB and to seek improved datasets, modeling, and tools for improved decision-making.

Due to shifting temperature regimes and altered precipitation patterns, snowpack in the CRB has been declining, and the timing of peak *SWE* has shifted earlier in the last 40 years, affecting the length of the snowmelt season (Harpold & Molotch, 2015). A decrease in snowpack magnitude (or snow drought) results in a non-linear increase in the uncertainties in the water supply forecast (Boardman et al., 2024). As the CRB hydrology evolves under the pressure of prolonged drought and climatic shifts, traditional water management approaches reliant on ground-based snowpack measurements and limited observational networks are no longer sufficient (Goble & Schumacher, 2023; Heldmyer et al., 2023; Hogan & Lundquist, 2024; Livneh & Badger, 2020). As a result, factors beyond peak snowpack magnitude are increasingly likely to influence how much snowmelt reaches streams and reservoirs. Spring conditions, particularly those occurring from April through June, play an expanding role as soil moisture (*SM*), evaporation rates, and vegetation water demand can significantly alter runoff efficiency (Miller et al., 2014; Woodhouse et al., 2016). For example, in 2015, despite very low *SWE*, inflows were higher than expected due to a late-season surge in precipitation. Conversely, persistent precipitation deficits into the spring season can propagate drought conditions over multiple seasons and years, with low *SM* further reducing the portion of snowmelt that translates into streamflow (Orth & Seneviratne, 2013). These interlinked processes and their variability introduce and shape the basin hydrologic response.

Moreover, the spatial and temporal constraints of ground-based observations imply that decision-makers are often working with an incomplete picture of the overall water budget in the CRB. This limitation has encouraged the development and application of remote sensing technologies in water management in order to capture the full hydrological variability across landscapes. In that regard, advancements in remote sensing, such as NASA's Gravity Recovery and Climate Experiment (GRACE; Tapley et al., 2004) and Soil Moisture Active Passive (SMAP; Entekhabi et al., 2010) satellite missions, offer opportunities to supplement and, in some cases, surpass the insights gained from traditional ground data. GRACE provides information on terrestrial water storage changes that were previously difficult to detect (Famiglietti et al., 2011), while SMAP captures surface *SM* data critical for understanding runoff efficiency and drought propagation. Integrating these data into decision-making frameworks could enable a more comprehensive assessment of current drought conditions.

Prior studies have sought to identify the most critical factors influencing streamflow generation during droughts in and around the CRB, yielding varying conclusions. CBRFC (2020) conducted a sensitivity analysis using a hydrological model and found that precipitation had the strongest influence on seasonal and annual flows, followed by *SM* (dominant during Oct-Dec) and evapotranspiration (*ET*), while temperature effects were minimal (<1%). Recently, a study by Goble & Schumacher (2023) used sub-surface modeling with random forest statistical forecasting to find that, on average, antecedent *SM* and groundwater storage data show limited utility in seasonal water supply forecasts for western Colorado, though precipitation and temperature measurements after peak snowpack significantly enhanced forecast accuracy. In contrast, Alam et al. (2024) employed land surface modelling to demonstrate that incorporating antecedent winter *SM* significantly improved spring runoff forecasts, particularly in high-elevation interior (cold and dry) catchments of the western U.S. Lapides et al. (2022) used observations to specifically analyze the 'missing snowmelt' after the drought in 2021 and concluded that incorporating antecedent root-zone *SM* deficits into prediction models could significantly decrease the overprediction of snowmelt runoff. Supporting this, Koster et al. (2023) used satellite-based *SM* observations from SMAP to show that late-fall *SM* conditions positively correlate with subsequent spring streamflow.

This study uses the Variable Infiltration Capacity (VIC) hydrological model to determine the major streamflow drivers in the CRB. The VIC model is well-established for CRB research, having been applied in prior studies to evaluate snow dynamics, *SM*, and climate impacts (Christensen et al., 2004; Vano et al., 2012; Wang et al., 2024; Whitney et al., 2023a; Xiao et al., 2022). Unlike earlier work, which often combines statistical methods, observational data, or different modeling tools, this study employs a controlled numerical experiment approach using the physical process-based framework in VIC. For example, one of the experiments in this study ensures that snowpack conditions remain consistent across simulations, isolating the specific effects of spring climate anomalies and multi-year *SM* deficits on streamflow. Additionally, the spatially distributed snowpack modeling in VIC allows us to test whether ground-based methods (e.g., SNOTEL station data) used operationally by agencies like the CBRFC accurately represent simulated basin-wide conditions.

To enhance the model accuracy, VIC was calibrated using the latest Parameter-elevation Relationships on Independent Slopes Model (PRISM; Daly et al., 2008) climate forcings, an important improvement in CRB studies, as studies in this region typically rely on older datasets, further detailed in section 2.2.1. Updated snow and vegetation parameterizations (detailed in section 2.3.3) were incorporated to reflect current basin conditions better. Model outputs were further validated against independent satellite-derived *SM* (SMAP) and terrestrial water storage (GRACE) data. This comprehensive calibration and validation framework was developed to increase confidence in the model's ability to represent real-world processes and make substantive improvements to existing CRB modeling approaches.

The observed changes in snowpack-streamflow relationships demonstrated in Figure 1 underscore the necessity for advanced hydrological understanding in the CRB. The systematic reduction in runoff efficiency during the post-2000 drought period suggests additional factors are playing an expanding role in driving streamflow within the basin's hydrological system. As a result, the overarching goal of this study is to utilize the calibrated VIC model to answer three research questions (see research framework in Figure 2):

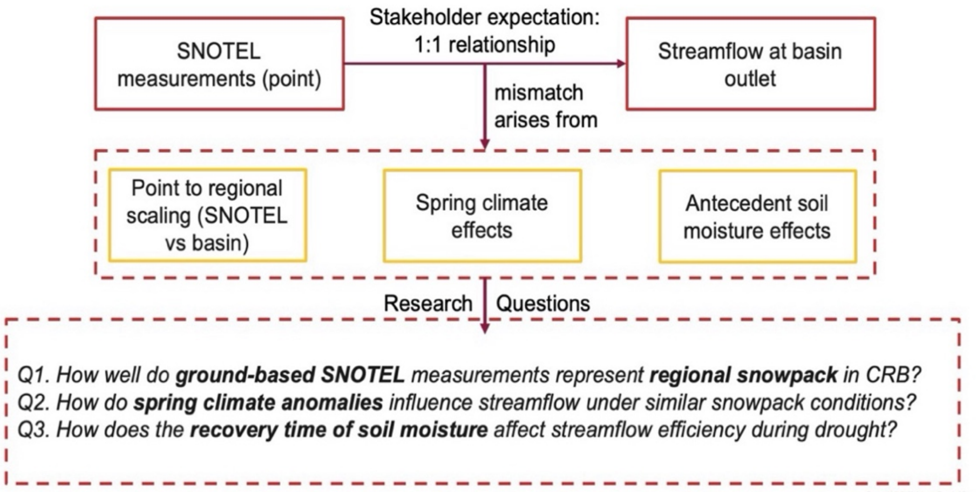


Figure 2. Research framework of this study.

1. How well do ground-based SNOTEL measurements represent regional snowpack conditions in CRB?
2. How do spring climate anomalies influence streamflow under similar snowpack conditions?
3. How does the recovery time of soil moisture affect streamflow efficiency during drought?

Ultimately, bridging the gap between traditional monitoring practices, cutting-edge satellite observations, modeling tools, and understanding the physical mechanisms that drive hydrological responses during drought is essential for effective water management. This research addresses this need by examining the three critical questions mentioned above. By addressing these knowledge gaps, this work aims to provide water managers with a more comprehensive understanding of basin hydrology and support the development of adaptive strategies in the face of climate uncertainty.

2. STUDY AREA, DATASETS, AND METHODS

2.1 Study Area

The CRB is a large watershed (~640,000 km²) extending from the Rocky Mountains to the Gulf of California. The basin spans seven U.S. states (Wyoming, Colorado, Utah, New Mexico, Arizona, Nevada, and California) and two Mexican states (Sonora and Baja California). The basin can be subdivided into the Upper Colorado River Basin (UCRB) and the Lower Colorado River Basin (LCRB). The division was established by the 1922 Colorado River Compact, with Lees Ferry, Arizona, serving as the point of division (The Colorado River Compact, 1922). The Upper Basin states include Wyoming, Colorado, Utah, and New Mexico. The Lower Basin consists of Arizona, California, and Nevada. The river continues south beyond the US border into Mexico, eventually reaching its delta region at the Gulf of California. As shown in Figure 3, the CRB can be further delineated into eight hydrologic sub-basins: four in the UCRB (Green, Upper Colorado, Glen Canyon, and San Juan) and four in the LCRB (Grand Canyon, Little Colorado, Lower Colorado, and Gila). Whitney et al. (2023a) established these 'analysis sub-basins' corresponding to subdivisions by the USGS (Table 1).

The CRB exhibits a significant topographic variation, with elevations ranging from sea level at the Gulf of California to over 4,000 m in the Rocky Mountains. This elevation gradient creates distinct ecological and hydroclimatic zones throughout the basin (Lukas & Payton, 2020). The UCRB is characterized by mountainous terrain with

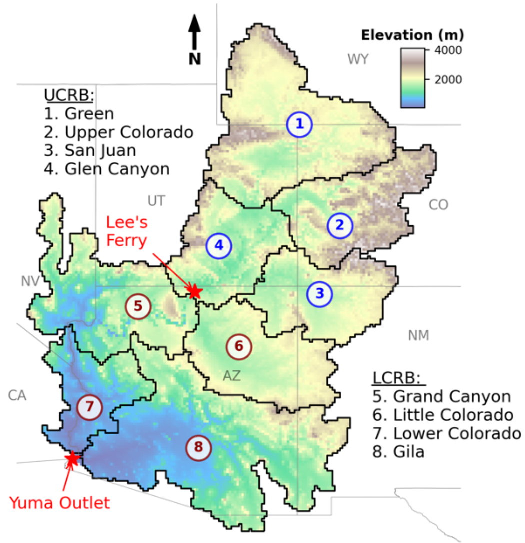


Figure 3. Study area map of the CRB showing elevation, major sub-basins within the watersheds, along with Lees Ferry and outlet at Yuma.

high-elevation snowpack zones above 2500 m, mid-elevation transition zones (1500-2500 m), and valley bottoms. The LCRB features canyon-dominated landscapes in the northern portion, lower-elevation desert plains, and isolated mountain ranges in the south

The basin is characterized by considerable spatial and temporal precipitation variability. Some mountain headwaters receive over 1000 mm of precipitation annually, with mean annual temperatures well below 0 °C, whereas the driest desert valleys receive approximately 100 mm of precipitation per year, with maximum daily temperatures exceeding 49 °C (Lukas & Payton, 2020; using Livneh et al. (2013) and PRISM datasets). While the UCRB receives the majority of its precipitation as winter snowfall from November through April, both the UCRB and LCRB experience significant summer rainfall from the North American Monsoon from July through September.

The Colorado River and its tributaries form a drainage network with a natural mean annual flow of approximately 18.2 km^3 measured at Lees Ferry (1906-2017 mean).

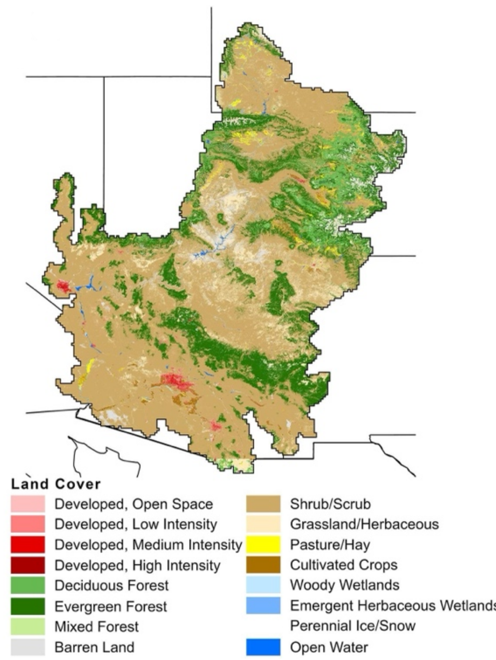


Figure 4. Land cover map of the CRB, from Bohn & Vivoni, (2019).

The UCRB contributes approximately 85% of the total runoff (Christensen et al., 2004) , while the LCRB tributaries provide minimal flow due to arid conditions. The natural flow regime has been substantially altered by water infrastructure development, including major storage facilities like Glen Canyon Dam (forming Lake Powell) and Hoover Dam (forming Lake Mead), which, combined together, can hold 21.97 km^3 of water - approximately 80% of the entire current reservoir storage capacity of the basin.

Land cover analysis from Figure 4 reveals that the basin is predominantly characterized by shrub/scrub ecosystems, followed by evergreen and deciduous forests. Herbaceous vegetation appears in scattered patches, while wetland ecosystems (both woody and emergent herbaceous) and open water bodies constitute a very small portion of the total area, reflecting the arid to semi-arid climate conditions that dominate the region. Notable concentrations of expanding metropolitan areas like Phoenix, Tucson,

Table 1. The major subbasins in the CRB and their corresponding USBR gauge, along with their characteristics. Streamflow refers to naturalized outputs from USBR (2024) (1980- 2020 mean except for CRB and Gila, which have 1980-2010 mean).

Basin	Location and Operation	Area (km^2)	Elevation (m)	Naturalized Streamflow (km^3/year)
CRB	Gage (09429490 + 09520500)	629,547	1,660	20.58
UCRB	Gage 09380000	283,401	2,090	17.60
LCRB	CRB – UCRB	346,146	1,289	2.98
Green	Gage 09315000	105,879	2,114	6.44
Upper Colorado	Gage 09180500	62,484	2,430	8.17
Glen Canyon	UCRB- Green- Upper Colorado- San Juan	55,850	1,787	0.68
San Juan	Gage 09379500	59,188	1,962	2.31
Grand Canyon	Gage 09421500 – Little Colorado- UCRB	80,023	1,400	0.86
Little Colorado	Gage 09402000	68,477	1,904	0.16
Lower Colorado	Gage (09429490 – 09421500)	42,001	672	-0.17
Gila	Gage 09520500	1,129	1,129	1.52

and Las Vegas are also observed, creating significant water demand within this predominantly shrub-dominated landscape.

2.2. Ground and Remote Sensing Products

In this study, ground observation networks, gridded products, and remote sensing datasets (Table 2) were used to capture hydrological variables. Ground observations provide accurate point measurements, while gridded datasets offer continuous spatial coverage through interpolation, and remote sensing captures large-scale patterns in areas with limited ground data. These datasets, including climate forcings, snowpack, soil moisture, streamflow, and terrestrial water storage change products, played different roles in the framework. Some were used to drive model forcings or served as calibration or validation targets, while others were used to perform independent model comparisons. All these datasets are stored in a Dropbox data repository, detailed in Appendix B.

2.2.1 Climate Forcings

Livneh et al. (2015) has been the preferred forcing dataset for hydrological modeling, especially in the southwestern U.S. and the CRB (Alam et al., 2024; Bennett et al., 2018; Heldmyer et al., 2023; Wang et al., 2024; Whitney et al., 2023a). This gridded meteorological dataset provides precipitation (P), temperature (T), and other variables across the US, Mexico, and Southern Canada at a spatial resolution of 1/16° (~ 6 km) from 1950-2013. The dataset was extended to 2018 through the efforts of Su et al. (2021). However, there were a couple of issues with using this dataset in our study. The objectives of this work require data from recent years, with the need to obtain timely (around monthly) updates as well. Moreover, a study from Shulgina et al. (2023) suggested that Livneh et al. (2015) dataset uses a fixed lapse rate of 6.5 °C/km, which can introduce elevation-dependent biases to the temperature interpolations. A strong bias in the minimum daily temperature (T_{min}) results in substantial cold biases at high elevations, which could result in an overestimation of snow accumulation in certain months.

In light of these issues, the meteorological forcings were transitioned using the products from PRISM (Daly et al., 2008). Studies have shown the PRISM dataset to be less biased, in agreement with other independent products, and spatially consistent in its

Table 2. Description of datasets used in this study. *P*: Precipitation, *T*: Temperature, *SWE*: Snow Water Equivalent, *SM*: Soil Moisture, *Q*: Streamflow, *TWSA*: Terrestrial Water Storage Anomalies.

Source	Type	Variables	Resolution
Livneh et al. (2015)	Gridded	*P*, *T*	6 km, daily (1950-2018)
PRISM	Gridded	*P*, *T*	4 km, daily (1981-2024)
SNOTEL	Ground	*SWE, SM*	Daily (1981-2024) for *SWE*, varies for *SM*
SNODAS	Modeled	*SWE*	1km, daily (2003-2024)
SMAP	Remote Sensing	*SM*	9km, daily (2015- 2024)
USBR	Ground	*Q*	Monthly (1981-2024)
GRACE	Remote Sensing	*TWSA*	Monthly (2002-2024)

long-term trends (Henn et al., 2018; Shulgina et al., 2023). This product leverages a knowledge-based interpolation approach that accounts for factors such as elevation and coastal proximity to produce high-resolution (typically 4 km) *P* and *T* grids. PRISM is available at various temporal resolutions (daily, monthly, annual) and covers a long historical record (monthly data from 1895–present, daily data from 1981–present). PRISM is updated frequently to support running the VIC model with the latest climate conditions. 'Provisional' data becomes available the very next day and can be used in the model for preliminary analysis, while 'stable' data is released after approximately 6-7 months of quality control and processing.

The PRISM dataset was obtained from the PRISM Climate Group (University of Oregon) FTP server (see details in Appendix C), and post-processing was performed to convert it into the intended format for analysis. This included reformatting, unit conversions, and spatial resampling to the model resolution. Detailed procedures and corresponding codes are provided in Appendix C.1. Further comparisons were also made to evaluate PRISM forcings with ground datasets (section 3.1).

2.2.2 Snow Products

Snow observations from multiple sources were compared, including SNOTEL ground observations, SNODAS (Snow Data Assimilation System) modeled data, and the University of Arizona (UA) gridded dataset. SNOTEL (Snow Telemetry) is a network of automated data collection sites operated by the Natural Resources Conservation Service (NRCS) (Serreze et al., 1999). Primarily located in remote, high-mountain watersheds across the western U.S., these stations measure and transmit information on snow depth, *SWE*, *T*, and *P*. There are over 900 SNOTEL sites in total, out of which 200 fall inside the CRB (160 in UCRB, 40 in LCRB) (Figure 5). Bulk download of long-term data from all these stations is facilitated with an R-programming-based package called 'snotelr' (Hufkens et al., 2024) (see Appendix C.2).

SNODAS is a modeling and data integration framework developed by the National Operational Hydrologic Remote Sensing Center (NOHRSC) (Barrett, 2003). It merges ground-based (including SNOTEL), airborne, and satellite observations with numerical weather and snowpack models to produce daily, high-resolution (1 km) estimates of snow cover, snow depth, and SWE across the conterminous U.S. The data availability is from September 2003 to the present. Similarly, the UA snow dataset provides daily 4 km gridded estimates of SWE and snow depth from 1981 to present (Broxton et al., 2019). It is generated by combining in-situ snow observations from the

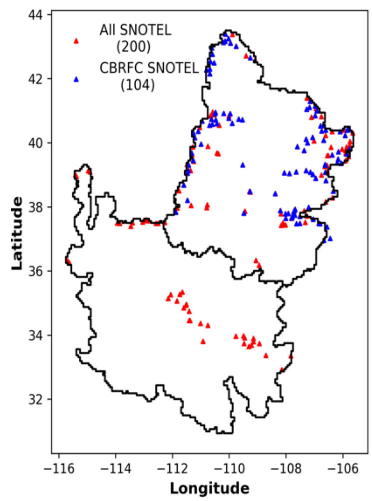


Figure 5. Spatial map showing the location of in-situ SNOTEL as well the 104 CBRFC ground stations in the CRB. Boundaries shown are UCRB (top) and LCRB (bottom).

SNOTEL network and the Cooperative Observer Network (COOP) network from the National Weather Service with gridded temperature and precipitation fields from PRISM.

2.2.3 Soil Moisture Products

A number of ground *SM* sensor networks were explored to check the data availability and quality: SNOTEL (118 sites), Soil Climate Analysis Network (SCAN) (16 sites), AmeriFlux (5 sites), and United States Climate Reference Network (USCRN) (7 sites). However, only SNOTEL sites (Figure 6) were evaluated due to their substantial number of stations compared to the others. Despite their value for point measurements, these networks presented several limitations for a basin-wide analysis. First, their spatial distribution proved inadequate for capturing the hydrological variability in the basin, particularly in the LCRB, where station density is sparse. Second, data gaps and temporal

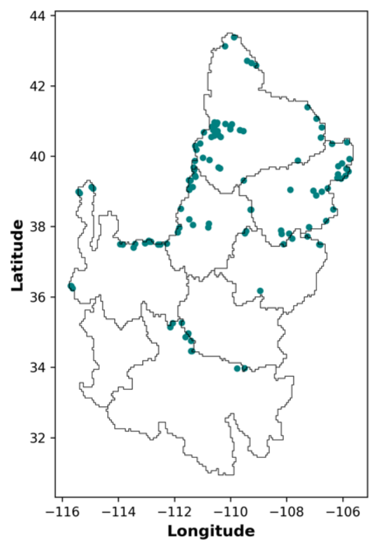


Figure 6. Spatial location of in-situ SNOTEL *SM* networks across the CRB.

inconsistencies were common among the available stations, complicating long-term analyses. Third, most stations lacked measurements at greater soil depths (e.g., 100 cm), limiting our ability to evaluate deep *SM* dynamics. These constraints necessitated the use of remote sensing products for a consistent basin-wide evaluation.

To address these limitations, NASA's Level 4 (L4) surface and rootzone *SM* (m^3/m^3) from the SMAP (Soil Moisture Active Passive) project (Entekhabi et al., 2010; Reichle et al., 2019) was used. The L4 data product is based on the assimilation of SMAP observation into a customized version of NASA Goddard Earth Observing System- Land Data Assimilation System (GEOS-5 LDAS). The SMAP spacecraft carries two L-band instruments, a radar (active) and a radiometer (passive), to leverage their complementary strength. The radiometer records microwave emissions from the top 5 cm in the soil with a spatial resolution of 36 km, while the radar provides backscatter measurements at a 3 km resolution. The combined instrumentation enables SMAP to generate highly accurate global *SM* at a 9 km resolution. The temporal coverage of the mission is from March 31, 2015. The raw measurements are taken every 2 to 3 days. There were gaps between 19 June to 23 July 2019 and August 2020 to Sept 2022 when the satellite went into safe mode. Latency is ~7 days, but the mean latency for the L4 product is around 2-3 days.

The L4 *SM* algorithm uses a Kalman filter data assimilation system to merge SMAP observations with *SM* estimates from NASA Catchment Land Surface Model (CLSM) to obtain the deep layer of *SM* (Reichle et al., 2019). CLSM is driven by observation-based surface meteorological data, and it encapsulates the vertical transfer of *SM* between surface and root zone reservoirs. The model then interpolates and extrapolates SMAP observations in time and space, producing 3-hourly estimates of *SM* at 9 km resolution. The surface and root-zone *SM* estimates have been demonstrated to meet the accuracy requirement of unbiased Root Mean Squared Error (RMSE) ≤ 0.04 m^3/m^3 at the 9-km and 36-km scales, making it suitable for basin-scale hydrological studies (Reichle et al., 2019).

2.2.4 Streamflow Products

The U.S. Bureau of Reclamation provides estimates of naturalized streamflow which indicate how much water would naturally flow through a river system if human influences, such as reservoir operations, irrigation diversions, and other consumptive uses, were absent (USBR, 2024). Incorporating such a product in a hydrological modeling framework such as that of VIC allows for a like-to-like comparison of the modelled streamflow outputs. Table 1 summarizes the key sub-basins in the CRB and the corresponding USBR gauge location at their outlet used for model evaluation.

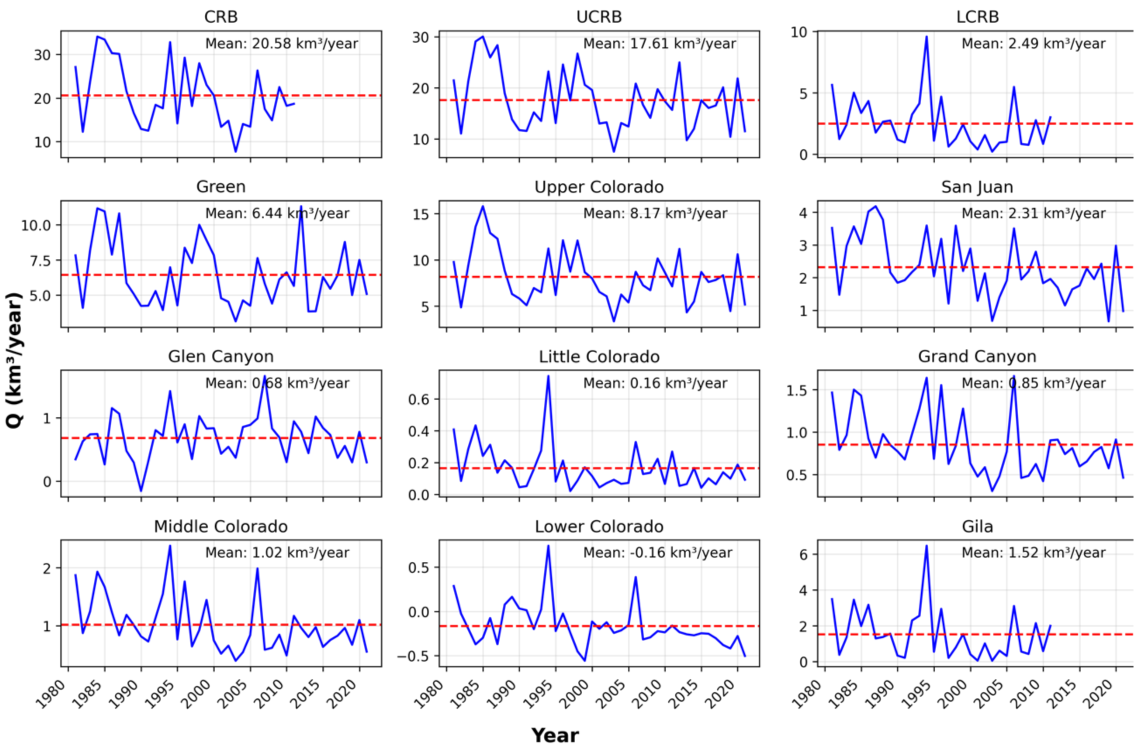


Figure 7. Time series of naturalized streamflow (USBR) for the major sub-basins in the CRB.

The USBR naturalized streamflow product has undergone updates since its use in Whitney et al. (2023a). This study implements the most recent version (2022), which includes refined accounting methods for reservoir evaporation and return flows. Figure 7 shows the time series of the naturalized streamflow for the major sub-basins reported in Table 1. The negative naturalized streamflow values observed in the Lower Colorado sub-basin represent statistical artifacts from the USBR dataset, which can occur due to estimation procedures when calculating natural flow conditions in highly managed river systems. In addition to the major sub-basins defined by USBR, Little Colorado and Grand Canyon sub-basins were combined into the Middle Colorado since this area is commonly used in management decisions in the LCRB (Whitney et al., 2023a).

2.2.5 Terrestrial Water Storage Products

GRACE (Gravity Recovery and Climate Experiment) estimates the amount of water stored in a region by observing the changes in Earth's gravitational field (Rodell & Famiglietti, 2001). Since April 2002, GRACE has provided monthly Terrestrial Water Storage Anomaly (*TWSA*) estimates: a sum of snow, surface water, *SM*, and groundwater, and continues under the GRACE-Follow On (FO) mission after 2018. Although its native resolution is around 300 km, techniques such as applying gain factors or advanced machine-learning approaches enable finer-scale resolutions. Integrating GRACE-based indicators into U.S. and North American Drought Monitors has provided valuable insights into large-scale drought conditions, and assimilating GRACE data into land surface models has been shown to improve groundwater simulations (Li et al., 2019). Beyond these operational applications, GRACE observations offer an independent means to validate hydrological models and gain further insights into groundwater changes, which are difficult to track across large regions (Jin & Feng, 2013). GRACE data was obtained from Physical Oceanography Distributed Active Archive Center (PODAAC) platform (NASA, 2023).

In this study, the gain factors were used to adjust the 3-degree or 300 km mascon areas to a 50 km resolution. Gain factors are multiplicative adjustments that correct for the smoothing introduced by large-scale averaging in GRACE data, helping recover finer local signals (Wiese et al., 2016). To derive these factors, the 0.5-degree resolution Community Land Model (CLM) was first "mascon averaged" to match the coarser GRACE scale, and then a least-squares fit was performed between the "mascon-averaged" hydrology model and the original GRACE data (Wiese et al., 2016). This downscaling approach allowed for a more detailed analysis of water storage variations across sub-basins within the CRB, which is particularly important for distinguishing regional drought impacts.

2.3 Hydrological Modelling and Calibration

2.3.1 The VIC Model

The VIC model (Liang et al., 1994) is a semi-distributed, large-scale hydrological model developed to simulate water and energy balances at the land surface. VIC uses a grid-based approach, dividing the land surface into grid cells. Within each cell, it can represent sub-grid variability (e.g., vegetation, soil) and thus can better capture spatial heterogeneity. The model operates without lateral water movement between grid cells, treating precipitation as the primary input to each individual cell. VIC is typically run at sub-daily time steps to capture diurnal fluctuations at various spatial resolutions, most commonly at a resolution of 1/16° or approximately 6 km grid cells. The model employs elevation bands to account for topographic effects on meteorological variables within each grid cell. This is particularly important in the UCRB, where significant elevation gradients exist within individual grid cells. By subdividing each grid cell into multiple elevation bands, VIC can more accurately represent the vertical variation of precipitation and temperature, which strongly influences snowpack dynamics. As demonstrated by Wang et al. (2024), this elevation band approach allows for more realistic simulation of precipitation phase partitioning (rain vs. snow) and temperature-dependent processes across different elevations within the same grid cell.

A core feature of VIC is the ‘variable infiltration curve’, which allows infiltration capacity to vary with *SM*, improving runoff and baseflow simulations. In addition to these features, VIC also has a multi-layer soil structure that tracks moisture storage and fluxes across different soil depths. VIC’s energy balance component also helps capture *ET* dynamics, providing a detailed picture of how *SM* evolves under different scenarios. In this study, the latest version of the model known as VIC-5 (Hamman et al., 2018) was used, which features improved code structure, enhanced computational capabilities, and a streamlined workflow, making it easier for users to set up, run, and analyze simulations. These features and capabilities make VIC an ideal tool to help achieve the objectives of this study.

There have been extensive studies supporting the application of the VIC model in the CRB that validate its use in this study. Christensen et al. (2004) used VIC to assess climate change impacts on the CRB, projecting significant reductions in streamflow and reservoir storage. The Bureau of Reclamation's 2012 Colorado River Basin Water Supply and Demand Study employed VIC as part of its modeling framework to evaluate future water availability scenarios, providing critical information for long-term water management planning in the basin (Bureau of Reclamation, 2012). VIC has proven particularly useful for drought analysis in the basin, with Mote et al. (2018) employing the model to understand snowpack decline and Xiao et al. (2018) employing it to assess the impacts of climate change on water resources in the UCRB. The Western Water Assessment's 2020 report utilized VIC simulations to examine the impacts of warming temperatures on Colorado River flows (Lukas & Payton, 2020). More recently, Whitney et al. (2023b) applied VIC to investigate the compounding effects of climate change and land cover modifications on basin hydrology, while Wang et al. (2024) focused on understanding and improving the model physics of precipitation partitioning in VIC, enhancing the representation of hydrological processes in the basin.

2.3.2 Model Dynamics

Since this study specifically investigates the role of *SM* in snowmelt-streamflow relationships, a detailed understanding of how VIC represents snow, *SM,* and streamflow dynamics is essential. The model treats the snowpack as consisting of two layers: a thin surface layer and a thicker underlying layer. It calculates the energy and mass balance at the ground surface in a manner consistent with other cold region land process models (Andreadis et al., 2009). Only the surface layer participates in energy exchange with the atmosphere and the forest canopy. The surface layer's energy balance is evaluated as:

$$\rho_w c_s \frac{dWT_s}{dt} = Q_r + Q_s + Q_e + Q_p + Q_m \quad , \tag{2.1}$$

where ρ_w is the density of the water, c_s is the specific heat of ice, W is the water equivalent and T_s is the temperature of the surface layer, Q_r is the net radiation flux, Q_s is the sensible heat flux, Q_e is the latent heat flux, Q_p is the energy flux advected to the snowpack by rain or snow, Q_m and is the energy flux given to the pack because of liquid water refreezing or removed from the pack during melt.

The model employs several key formulations to simulate infiltration, moisture redistribution, and baseflow generation. Water enters the top soil layer using a "variable infiltration capacity" formulation (Liang et al., 1994). Conceptually, infiltration capacity decreases with increasing *SM* and part of the incoming precipitation is partitioned into surface runoff when the infiltration limit is exceeded. Mathematically, the variation of the infiltration capacity within an area can be represented as:

$$i = i_m \left[1 - (1 - A)^{\frac{1}{b}} \right] \qquad , \tag{2.2}$$

where i and i_m are the infiltration capacity and its maximum value, respectively, A is the fraction of area for which the infiltration capacity is less than i, and b is the shape parameter for infiltration. Then, from (Liang et al., 1996), the direct runoff is estimated as:

$$Q_d \cdot \Delta t = P \cdot \Delta t - z_2 \cdot (\theta_s - \theta_2), i_0 + P \cdot \Delta t \geq i_m \qquad \text{, and} \tag{2.3}$$

$$Q_d \cdot \Delta t = P \cdot \Delta t - z_2 \cdot (\theta_s - \theta_2) + z_2 \cdot \theta_s \cdot \left[1 - \frac{i_0 + P\Delta t}{i_m}\right]^{1+b_i}, i_0 + P \cdot \Delta t \geq i_m \ , \tag{2.4}$$

where Q_d is the direct runoff, Δt is the time step which is taken as one hour in the model calculation, z_2 is the depth of the second layer, P is the precipitation rate, i_m is the maximum infiltration capacity, i_0 is a specific point infiltration capacity of i that corresponds to the soil moisture at that time step, $b_i$is the infiltration shape parameter which is a measure of the spatial variability of the infiltration capacity, θ_2 is the volumetric water content of layer 2, and θ_s is the soil porosity.

Once water infiltrates, VIC redistributes moisture through gravity-driven flow following unsaturated hydraulic conductivity relationships (Brooks and Corey, 1964):

$$Q_{12[N+1]} = K_s \left(\frac{W_{1[N+1]} - \theta_r}{W_1^c - \theta_r} \right)^{\frac{2}{Bp}+3} \qquad , \tag{2.5}$$

where K_s is the saturated hydraulic conductivity, θ_r is the residual moisture content, Bp is the pore size distribution index, $Q_{12[N+1]}$ is the drainage from layer 1 to layer 2, $W_{1[N+1]}$ is the SM content in layer 1 at the end of each time step, and W_1^c is the maximum SM content of layer 1.

Layers within the root zone lose moisture to transpiration, while the topmost layer can also lose moisture to soil evaporation. VIC's soil evaporation follows an ARNO-type formulation (Franchini & Pacciani, 1991), which is calculated in proportion to the fraction of average moisture content to the average storage capacity. Finally, drainage from the lowest soil layer is controlled by an ARNO baseflow equation (Franchini & Pacciani, 1991), often written as:

$$Q_b = \frac{\alpha\, D_m\, W(t)}{W_c} \text{ for } W <= Y \quad \text{, and} \tag{2.6}$$

$$Q_b = \frac{\alpha\, D_m\, W(t)}{W_c} + (1-\alpha)D_m \left[\frac{W(t)-Y}{W_c-Y}\right]^2 \text{ for } W >= Y \quad , \tag{2.7}$$

where Q_b is the baseflow, α is the fraction between 0 to 1 (ratio of D_s and W_s, $D_s <= W_s$), D_m is the maximum baseflow, Y is a moisture content threshold, W is the moisture in the third layer at the time t, and W_c is the maximum SM content. Equation 2.6 represents the linear portion of the baseflow, which occurs when SM content is lower or the soil is relatively dry. As the moisture in the deeper layer increases above a threshold, then the baseflow production is rapid as represented by the non-linear part of equation 2.7. The water balance in different soil layers is represented as:

$$W_l^+[N+1] = W_l^-[N+1] + (P - Q_d[N+1] - Q_{12}[N+1] - E_1) \cdot \Delta t, \tag{2.8}$$

where P represents precipitation, $N+1$ represents the bare soil class, $Q_d[N+1]$ is the direct runoff from the area, $Q_{12}[N+1]$ is the drainage from layer 1 to layer 2, $W_l^-[N+1]$is the soil moisture content in layer 1 at the beginning of the time step, $W_l^+[N+1]$ is the soil moisture content in layer 1 at the end of each time step, E_1 is the evaporation rate of layer 1 and Δt is the time step.

$$W_2^+[N+1] = W_2^-[N+1] + (Q_{12}[N+1] - Q_b[N+1] - E_2) \cdot \Delta t \quad , \qquad (2.9)$$

when $W_2^-[N+1] + (Q_{12}[N+1] - Q_b[N+1] - E_2) \cdot \Delta t < W_2^c$

where $Q_b[N+1]$ is the subsurface runoff, $W_2^-[N+1]$ is the soil moisture content at the beginning of the time step in layer 2, $W_2^+[N+1]$ is the layer 2 soil moisture content at the end of the current time, W_2^c is the maximum soil moisture content of layer 2 and E_2 is the evaporation rate of layer 2.

In the case when $W_2^-[N+1] + (Q_{12}[N+1] - Q_b'[N+1] - E_2) \cdot \Delta t > W_2^c$,

$$W_2^+[N+1] = W_2^c \quad , \qquad (2.10)$$

And,

$$Q_b''[N+1] = W_2^-[N+1] + (Q_{12}[N+1] - Q_b'[N+1] - E_2) \cdot \Delta t - W_2^c \ , \qquad (2.11)$$

The VIC model's complexity offers both advantages and limitations for this study. A simpler bucket model might capture the first-order monthly snow-streamflow dynamics with fewer parameters and computational requirements. However, such models would likely miss important energy balance components that affect snowmelt timing and rate, particularly during spring when radiation inputs become increasingly important (Marks &

Dozier, 1992). While VIC includes sophisticated representations of surface energy and radiation processes essential for accurate snowmelt simulation, its infiltration and baseflow generation schemes are relatively simplified compared to more physically based hydrological models. The variable infiltration curve approach, along with the ARNO baseflow formulation used in VIC, represents subsurface flow through empirical relationships rather than physically based equations of groundwater flow. This means that slow subsurface processes and deep layer storages, which are important in long-term drought/recovery analysis, may not be represented accurately. These simplifications make VIC computationally efficient for large basin applications but potentially less accurate in representing subsurface processes compared to fully distributed physically based models such as ParFlow (Maxwell et al., 2015). Despite these limitations, VIC's semi-distributed approach strikes a balance between physical process representation and computational efficiency.

2.3.3 Model Updates

The Colorado River Basin Supply and Demand Study conducted by the USBR in 2012 established a baseline for VIC modeling applications in the CRB, demonstrating the model capability to inform basin-wide water management decisions (Bureau of Reclamation, 2012). Building on this foundation, the VIC-5 framework has undergone significant enhancements through contributions from researchers at Arizona State University, improving its representation of key hydrological processes relevant to the CRB. Whitney et al. (2023b) expanded the model's applications by incorporating land cover change scenarios. Moreover, Xiao et al. (2022) used Moderate Resolution Imaging

Spectroradiometer (MODIS) land surface temperature and snow data to calibrate the vegetation parameters and snow processes. Further refinements were introduced by Wang et al. (2024), who demonstrated that implementing a wet-bulb temperature method provided more accurate estimations of snow conditions and subsequently improved the model's snow parameterization and precipitation phase partitioning algorithms. The current study employs a model version incorporating these improvements, representing the most advanced implementation of VIC for hydrological simulations within the CRB.

To systematically evaluate these improvements, four distinct model configurations were compared, each run over the entire basin from 1982 to 2018. Model performance was assessed by comparing simulated streamflow against naturalized flow records for the UCRB, along with its three major streamflow-producing sub-basins: Green, Upper Colorado, and San Juan. The four model configurations are as follows:

1. <u>Baseline:</u> This configuration represents the VIC implementation used in Whitney et al. (2023a) with Livneh et al. (2015) forcings. This configuration serves as the reference point against which subsequent improvements are evaluated.
2. <u>Updated Forcing Data (upd_forcing)</u>: This configuration retains the ‘Baseline’ VIC implementation but replaces the forcing with the PRISM dataset.
3. <u>Updated Snow Parameters (upd_snow)</u>: This configuration maintains the Livneh et al. (2015) forcings but incorporates the enhanced VIC model with improvements on:
    a. Implementation of a wet-bulb temperature-based method for snow/rain partitioning (Wang et al., 2024)

b. Modified downward longwave radiation to snowpack (Xiao et al., 2022)
c. Minimum wind speed threshold of 0.1 m/s to prevent computational artifacts
d. Compiler optimizations for improved performance on AMD CPU architecture

4. Updated Forcing and Snow (upd_forcing_snow): This configuration represents the most advanced implementation, combining both the enhanced VIC model structure (as in upd_snow) with the improved PRISM meteorological forcing data (as in upd_forcing).

2.3.4 Computational Implementation

The VIC model's computational demands required the simulations to be run on a high-performance computing environment. All simulations were performed on Arizona State University's Research Computing platform (Sol) (Jennewein et al., 2023) to accommodate the model's intensive processing requirements. Model implementation followed a structured procedure that included compilation with necessary environmental modules, preparation of input datasets (meteorological forcings, parameters, and initial conditions), configuration of simulation parameters, and execution through parallel processing.

The model configuration was specified through a global parameter file that defined simulation periods, input/output file locations, and selected output variables. Simulations were conducted using parallel processing techniques distributed across multiple computing nodes to maximize computational efficiency. This approach significantly reduced processing time, particularly for the extensive scenario-based experiments conducted in this study. Detailed technical specifications, including sample configuration files and execution scripts, are provided in Appendix D for reproducibility purposes.

2.3.5 Model Calibration

With the use of a new set of forcings (PRISM) and updates to the model, as discussed above, the model needed a comprehensive recalibration process. The calibration was done in two steps. First, the snow parameters were calibrated using SNOTEL products, then the soil parameters were calibrated with respect to the naturalized streamflow from USBR. The model outputs were then validated through independent comparisons against GRACE and SMAP observations.

The VIC model documentation outlines several approaches for snow module calibration, with three parameters typically requiring adjustment: (i) the maximum air temperature at which snowfall occurs, (ii) the minimum air temperature at which rainfall occurs, and (iii) the snow surface roughness. For this study, a modified approach was implemented that focused on snow albedo parameters to better align with ground observations. The snow albedo in VIC is represented as:

$$\alpha_{snow} = \alpha_{max} A^{t^B} \quad , \qquad (2.5)$$

where α_{snow} is the snow albedo, α_{max} is the fresh snow albedo, *t* is the time since last snowfall, and *A* and *B* are constants.

SNOTEL stations were used to calibrate the snow module. However, not all of the 200 stations in the basin were used. Only stations where the absolute elevation difference between the station and its corresponding VIC grid cell was less than 250 m were included, resulting in 163 suitable sites from the original 200. Of these, 88 stations contained complete records from 1982-2022. To evaluate whether the station selection criteria might introduce bias into the calibration process, a bootstrapping analysis was performed. Different subsets of stations (ranging from 5 to the full 200 stations in the basin) were randomly selected, and the normalized RMSE (ratio of actual RMSE to the mean value) for the mean daily SWE was calculated for each subset compared to the full set of observations. However, it should also be noted that SNOTEL stations have inherent location biases that affect their representativeness of basin-wide snow conditions. These stations are typically situated in forest openings on relatively flat terrain due to logistical constraints in installing and maintaining instrumentation on slopes or in dense vegetation (Meromy et al., 2013; Molotch & Bales, 2006). As a result, SNOTEL sites tend to overestimate SWE compared to surrounding terrain, particularly in areas with complex topography. Despite these limitations, the combined approach of elevation-based filtering and statistical validation through bootstrapping helped minimize potential biases while ensuring sufficient spatial coverage for robust model calibration.

The snow calibration focused on two specific objectives: (1) minimizing mean bias in elevation-based *SWE* comparisons, and (2) reducing mean bias in peak *SWE* for each water year. The parameters listed in Table 3 were tuned during the calibration process. While albedo parameters were adjusted uniformly across the basin through the global parameter file, snow roughness was calibrated at a sub-basin scale to account for regional differences in snow cover.

Table 3. Important snow parameters for calibration of snow module in VIC.

Parameter	Implication (with increase)
Snow albedo accumulation *A*	Increases snow albedo during accumulation phase
Snow albedo thawing *A*	Increases snow albedo during ablation phase
Snow albedo thawing *B*	Increases snow albedo during ablation phase
Snow roughness	Increases the rate of melting and increases runoff

Table 4. Important soil parameters for the calibration of streamflow in VIC.

Parameter	Range	Meaning/Implication
Ds	0-1	Fraction of Ds_{max} where nonlinear flow starts. High Ds means higher baseflow at lower water content in the lowest layer
Ds_{max}	0-~30	Maximum baseflow than can occur in the lowest layer
Ws	0 – 1	Fraction of maximum *SM* where nonlinear baseflow starts
b_{inf}	0 – 0.4	Controls rainfall/snowmelt partition into infiltration and runoff. Higher b_{inf} means lower infiltration and higher runoff
Soil Depth	Vary by layer	Thicker soil depth means slower runoff response due to longer retention of *SM*

The soil parameters influence the infiltration and, subsequently, runoff and baseflow produced from each cell. The parameters listed in Table 4 are considered to be the most important calibration parameters (Schaperow et al., 2021). The calibration process began with sensitivity analysis to determine how each sub-basin responded to parameter adjustments. Parameters were calibrated on a basin-by-basin basis through modifications to the vegetation parameter file. Calibration performance was evaluated by comparing simulated streamflow against naturalized streamflow at corresponding basin outlet gauges, as specified in Table 1.

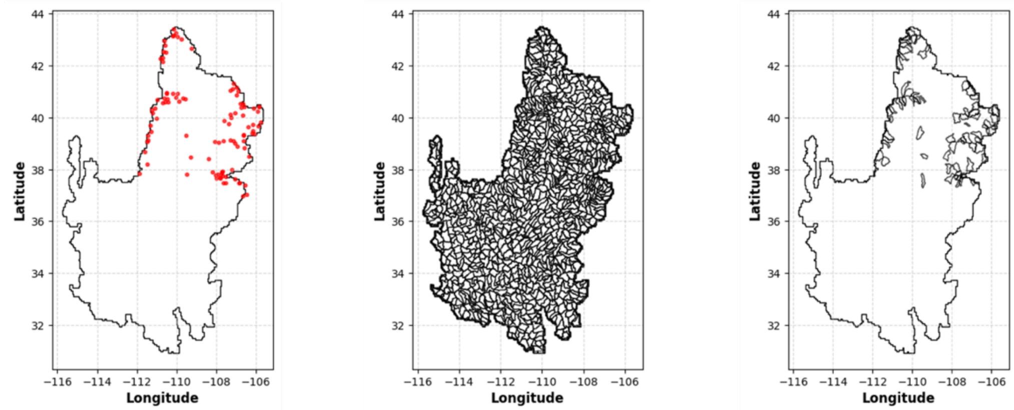
2.4 Design of Numerical Experiments

Following model calibration and validation, a series of numerical experiments were designed to systematically address the research questions outlined in Chapter 1. The experimental design employed controlled simulations where specific variables could be manipulated while holding others constant, allowing for the attribution of hydrological responses to individual factors. This approach enabled quantitative assessment of how ground-based snow measurements represent basin-wide conditions, how post-April climate affects streamflow generation, and how antecedent *SM* influences streamflow efficiency during drought periods.

2.4.1 Spatial Representation of SNOTEL Stations

This experiment addresses the first research question concerning potential uncertainties in snowmelt-driven streamflow prediction due to limitations in SNOTEL spatial representativeness. Current operational practice by water managers relies on an ensemble median from 104 SNOTEL stations, analyzed against a 1991-2020 climatological baseline to classify snowpack conditions as near, above, or below average. This experiment evaluates if this point-based approach accurately captures basin-wide snow conditions, particularly in the UCRB, and whether alternative approaches using VIC model outputs or gridded products might provide more representative estimates.

HUC10 watershed boundaries were downloaded and clipped to the CRB to align with the area of interest. For the comparison, the HUC10 polygons containing at least one



CBRFC SNOTEL | All HUC10s (mean size 600km^2) | Selected HUC10s

Figure 8. Spatial map showing a) 104 CBRFC SNOTEL stations (left), b) All HUC10 boundaries (middle), c) HUC10 boundaries with SNOTEL stations in the UCRB (right).

of the 104 SNOTEL stations were identified, resulting in 81 HUC10 watershed polygons (Figure 8). For each of these 81 polygons, time series data from the SNOTEL station located within the specific HUC10 watershed were extracted. In cases where multiple SNOTEL stations were present within a single watershed, the mean *SWE* values from those stations were computed. Similarly, the geometry of each HUC10 watershed was used to extract the mean *SWE* time series for VIC and SNODAS over the same region.

For each HUC10 watershed, three daily *SWE* time series were obtained: SNOTEL, VIC, and SNODAS. UA product was not compared here due to its limitations (section 3.1.2). Daily time series were resampled to mean yearly values for further analysis. The yearly mean values were then normalized by expressing them as percentages relative to their respective baselines. For the comparison between SNOTEL and VIC, 1991–2020 was selected as the baseline period. For the comparison between SNOTEL and SNODAS, the baseline period of 2004–2020 was selected due to SNODAS data availability. The following methods were employed to conduct the analysis:

i. It was determined whether the selected HUC10 boundaries adequately represent the corresponding SNOTEL stations by calculating correlation coefficients between point measurements and watershed averages.

ii. The ability of these boundaries to capture conditions across the Upper Basin was evaluated by calculating the long-term mean *SWE* for all HUC10 boundaries and filtering them based on specific long term daily *SWE* thresholds. The median of all watersheds for each water year was calculated, following a similar approach to the median of 104 stations, to enable direct comparison of the approaches.

iii. Statistical comparisons were conducted between the SNOTEL *SWE* observations and the modeled *SWE* outputs to identify discrepancies and evaluate whether the modeled *SWE* better represents snow conditions in the basin.

2.4.2 Effects of Spring Weather

This experiment addresses the second research question regarding uncertainty in snowmelt-streamflow relationships attributable to weather conditions following peak snowpack accumulation. Specifically, it isolates the influence of April-May-June (AMJ) weather on annual streamflow generation in the UCRB. Water year (WY) 2020 was selected as the experimental baseline due to its near-average snowpack conditions (104% of the 1991-2020 mean conditions) (CBRFC, 2024).

To test this hypothesis, calibrated VIC model simulations spanning 40 years (1984-2023) were utilized, and state files for each year were generated based on this set of calibrated outputs. The initial state file was created using a spin-up run from 1982-2019. All state variables, including initial *SM* at the start of winter and *SWE* before April

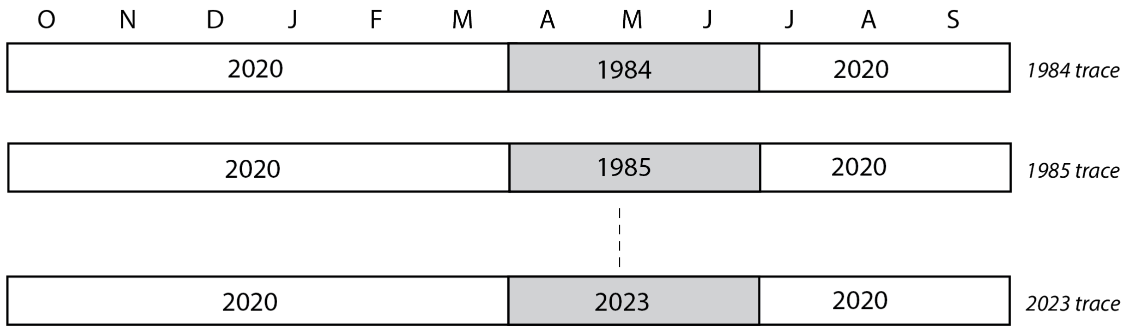


Figure 9. Setup for creating unique climate traces for WY 2020.

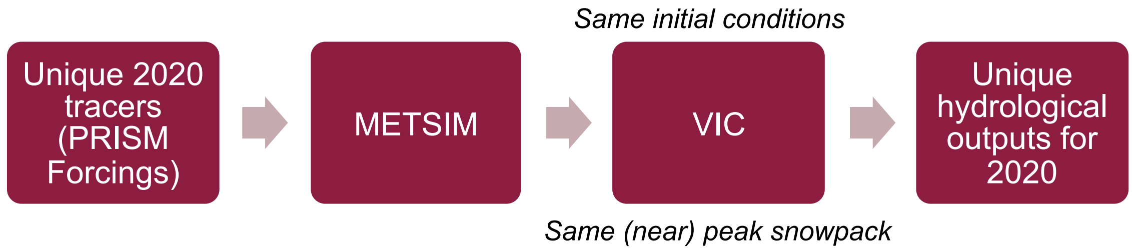


Figure 10. Design of the experiment to isolate AMJ climate effects on WY streamflow in the UCRB.

1, were held constant, ensuring that any variations in the hydrological outcomes could directly be attributed to the climatic conditions during the transition period (see experimental design, Figure 10). This controlled framework allowed for systematic quantification of streamflow sensitivity to AMJ climate under consistent antecedent *SM* and near-peak snowpack conditions.

In each simulation (1984-2023), the baseline AMJ P, T_{max}, and T_{min} were replaced by the corresponding climate data from each respective year, while all other variables remained as in 2020 (Figure 9). This process generated 40 unique traces, each representing the climatic conditions of a different year during the transition period, while keeping pre-April and post-June conditions identical to 2020. These modified climate datasets were processed through METSIM and the VIC model, producing 40 new VIC

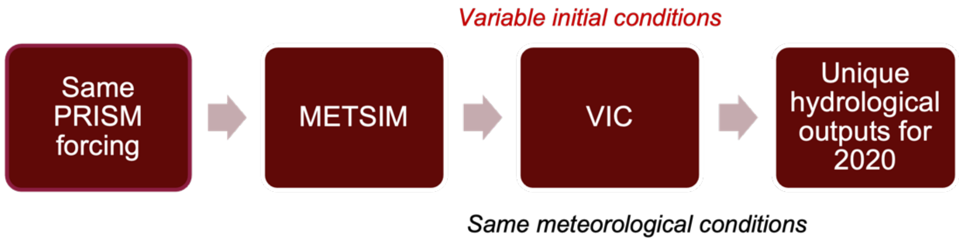


Figure 11. Design of experiment to isolate the impacts of antecedent SM on the WY streamflow in the UCRB.

output files. These outputs were analyzed to understand how variations in the transition period climate impact streamflow and other hydrological variables under consistent snowpack conditions.

2.4.3 Effects of Initial Soil Moisture

This experiment addresses the third research question concerning the influence of antecedent *SM* conditions on snowmelt-streamflow relationships. October 1, which marks the beginning of the WY, was selected as the initialization point to evaluate how pre-winter *SM* status affects subsequent streamflow generation. This experimental design (Figure 11) complements the previous AMJ climate experiment while focusing on initial conditions rather than post-snowpack climate.

For this experiment, a set of 40 simulations was performed using WY 2020 as the baseline. In each simulation, the initial *SM* conditions from October 1, 2020 were replaced with conditions from another year in the historical record (1984-2023). All three soil layers in the VIC model (surface layer, root zone, and deep soil) were modified simultaneously by substituting the complete *SM* state from each historical year. This approach allowed us to systematically analyze how different antecedent *SM* conditions influence streamflow generation while keeping meteorological forcings identical across all simulations.

The resulting ensemble outputs were analyzed to quantify streamflow sensitivity to initial *SM* states and to evaluate how this sensitivity manifests differently in the surface runoff and baseflow components. This partition was particularly important for understanding how *SM* deficits might influence different streamflow generation mechanisms during drought periods, potentially explaining observed reductions in streamflow efficiency under adequate snowpack conditions.

2.4.4 Combined Effects of Post-April Climate and Initial Soil Moisture

This experiment builds upon the previous two experiments to investigate the interactive effects of antecedent *SM* conditions and post-April climate on streamflow generation. The experiment was designed to quantify how these two factors jointly influence snowmelt-streamflow relationships, addressing both the second and third research questions in an integrated framework.

To maintain computational feasibility, a strategic approach was implemented rather than conducting all possible combinations across 40 years (1,600 simulations). Five representative *SM* conditions were selected based on the 1991-2020 mean value of 123.56 mm: very wet (1998, +23.80% relative to mean), wet (2015, +10.29%), average (2017, +0.11%), dry (2013, -9.58%), and very dry (2021, -22.39%) (Figure 12). These five initial condition scenarios were each combined with the 40 different AMJ climate conditions from the historical record (1984-2023), using the same experimental framework as in Section 2.4.2. This design resulted in 200 unique simulations (5 initial

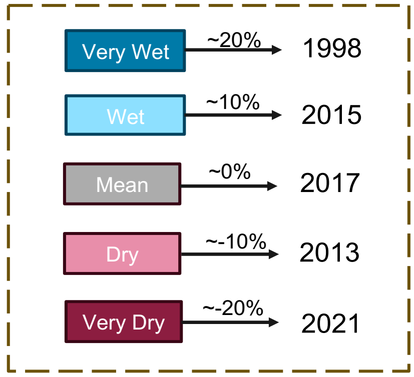


Figure 12. Selection of representative years for the combined *SM* and AMJ climate experiment. Changes are relative to the mean for 1991-2020. AMJ climate setup is the same as in Figure 9.

SM states × 40 AMJ climate conditions), with all other variables and forcings held constant as in WY 2020. Each simulation maintained consistent snowpack conditions to isolate the effects of *SM* and post-snowpack climate. This experimental approach enabled systematic quantification of streamflow sensitivity to the two variables simultaneously, allowing for evaluation of their relative importance and potential interaction effects.

To synthesize findings from the 200 simulations, a multiple regression approach using standardized anomalies was developed. This standardization ensured equal weighting of variables with different units and magnitudes. The analytical procedure followed several sequential steps:

i. Spatial Integration: Simulation outputs were clipped within the basin boundary to calculate water year streamflow totals (combined baseflow and surface runoff) in million acre-feet (MAF).

ii. Normalization of Response Variable: Each simulation streamflow value was converted to a percent deviation from the baseline mean condition. For example, a simulation yielding 8 MAF compared to a baseline mean of 10 MAF represented a -20% change, while 14 MAF corresponded to a +40% change. This normalized metric served as the predictand in the subsequent regression analysis.

iii. Standardization of Predictor Variables: Predictor variables (October 1 *SM*, AMJ *P* and *T*) were standardized using the ensemble (40-year) climatological means and standard deviations. This transformation converted each variable to dimensionless z-scores, allowing direct comparison of their relative influences. Spatial averaging was similarly constrained to the UCRB.

2.4.5 Drought Experiments

While the previous experiments focused on single-year responses, this final set of experiments extends the analysis to multi-year timescales to investigate *SM* memory and persistence during extended drought conditions. This experiment builds on the *SM* sensitivity analysis but addresses longer hydrological responses that are relevant for understanding drought propagation and recovery in the CRB. The experimental design employed a series of hypothetical scenarios to isolate specific drought-related mechanisms, particularly focusing on deep *SM* dynamics and baseflow generation. One baseline scenario (representing historical conditions) and four alternative simulations were developed to quantify different aspects of *SM* memory:

1. Initial Condition Experiment: This scenario extended the *SM* analysis to *SM* memory by replacing the initial conditions of WY 2020 (80% of normal) with those from WY 1997 (140% of normal). This was done to simulate an alternative trajectory to see how wet conditions just before the start of WY2020 would persist under the meteorological forcings of 2020-2024.

2. Snowmelt Impact Experiments: These experiments investigated how replacing dry periods with wet periods of different durations affects deep *SM* and baseflow persistence. The experiments focused on observing the resulting changes in deep *SM* storage and baseflow generation. Three temporal scenarios were implemented:

   a) Short-term response: WY 2020 forcing was replaced with water year 2023 (high snowpack) while maintaining constant forcing for other years in the WY 2020-2024 simulation period.

   b) Mid-term response: Forcings from WY 2012-2013 were replaced with those from 1996-1997, with simulations spanning WY 2012-2024.

   c) Long-term response: Forcings from WY1999-2003 were replaced with those from 1993-1997, with simulations extending from WY1999-2024.

The primary objective was to quantify how long the effects of increased moisture input (from replacing dry years with wet years) would persist in the deep soil layer and subsequently influence baseflow generation under drought conditions.

2.5 Statistical Metrics

The following statistical metrics were used to evaluate model performance, quantify relationships between variables, and assess errors in the various analyses conducted throughout this study. These metrics provide a framework for objective comparison between observed and simulated data, enabling robust assessment of model reliability and the strength of relationships between hydrological variables. Python's *scipy.stats* (Virtanen et al., 2020) module was used for the calculation of these metrics.

(i) Correlation Coefficient (CC): CC measures the strength and direction of linear association between two variables. Values range from -1 to 1, where 1 indicates perfect positive correlation, -1 indicates perfect negative correlation, and 0 indicates no correlation.

(ii) Coefficient of Determination (R^2): R^2 represents the proportion of variance in the dependent variable that is predictable from the independent variable. Values range from 0 to 1, with higher values indicating better explanatory power of the model.

(iii) Root Mean Square Error (RMSE): RMSE measures the average magnitude of errors between predicted and observed values. Lower values indicate better model performance.

(iv) Bias: Bias quantifies the average tendency of predicted values to deviate from observed values. A positive bias indicates systematic overprediction, while a negative bias indicates underprediction.

(v) Nash-Sutcliffe Efficiency (NSE): NSE is a normalized metric that determines the relative magnitude of residual variance compared to observed data variance. Values range from $-\infty$ to 1, where 1 indicates perfect agreement, and zero indicates that the model predictions are as accurate as the mean of observed data. Negative values indicate that the observed mean is a better predictor than the model.

3. RESULTS AND DISCUSSION

3.1 Evaluation of Ground and Remote Sensing Products

In this section, the different datasets used in this study were evaluated to understand their accuracy, biases, and limitations. This evaluation was essential for interpreting subsequent model outputs and ensuring that any conclusions drawn were based on reliable data. First, meteorological forcing datasets interpolated from ground stations were compared. Subsequently, the spatial and temporal patterns in soil moisture and terrestrial water storage observations from remote sensing were analyzed to gain additional information about basin conditions.

3.1.1 Comparison of Forcing Products

The comparison between the Livneh et al. (2015) and PRISM datasets from 1981-2018 reveals important differences in their representation of key meteorological variables across elevation gradients. Biases were calculated as the difference between monthly mean values from Livneh et al. (2015) minus PRISM, averaged over the entire period. The analysis of the bias patterns between these datasets reveals distinct elevation-dependent relationships across daily P, T_{max}, and T_{min} variables (Figure 13). The P biases are relatively small across all elevation bands, with slightly negative values (red) appearing at the higher elevations and slightly positive values (blue) at some lower elevations. This indicates generally good agreement between datasets for P.

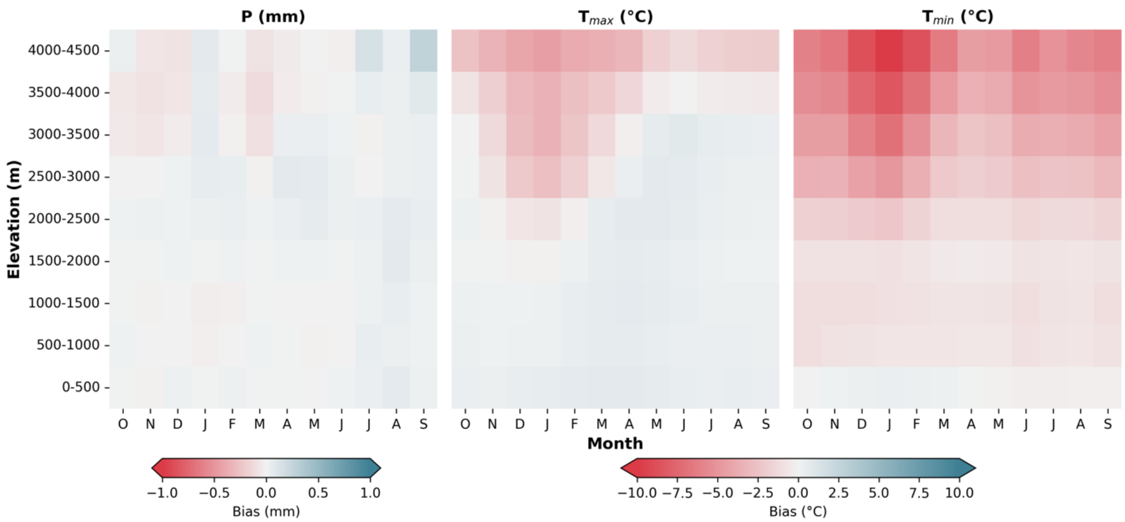


Figure 13. Comparison of biases in *P,* T_{max}, and T_{min} variables for Livneh et al. (2015) and PRISM products across elevation ranges.

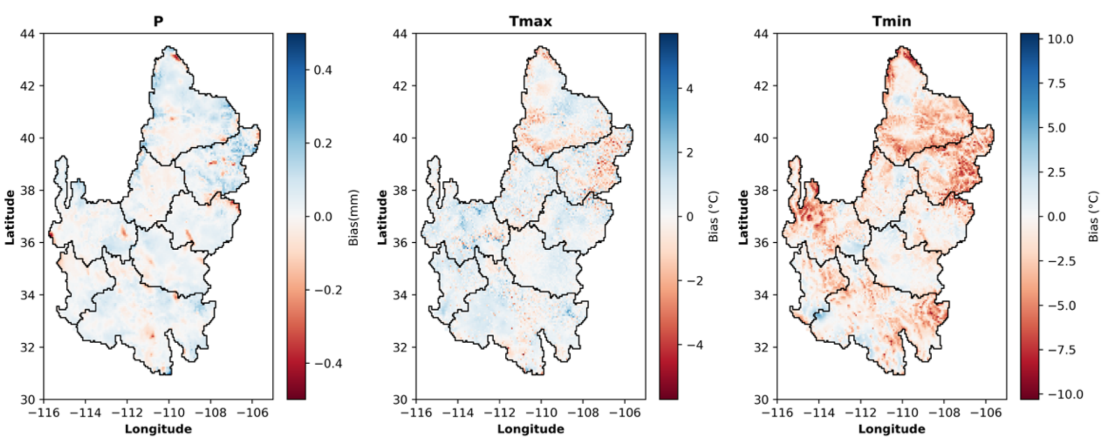


Figure 14. Spatial distribution of the biases (obtained as Livneh et al. (2015) minus PRISM) across the CRB for *P,* T_{max}, and T_{min}.

The T_{max} bias patterns reveal interesting patterns, particularly at elevations above 2500 m, where substantial negative biases (~2-3 °C) occur in the winter months from November through April. Interestingly, these pronounced biases were not evident in the spatial analysis of UCRB sub-basins (Green, Upper Colorado) (Figure 14). This discrepancy may be explained by examining the mean elevations of those sub-basins

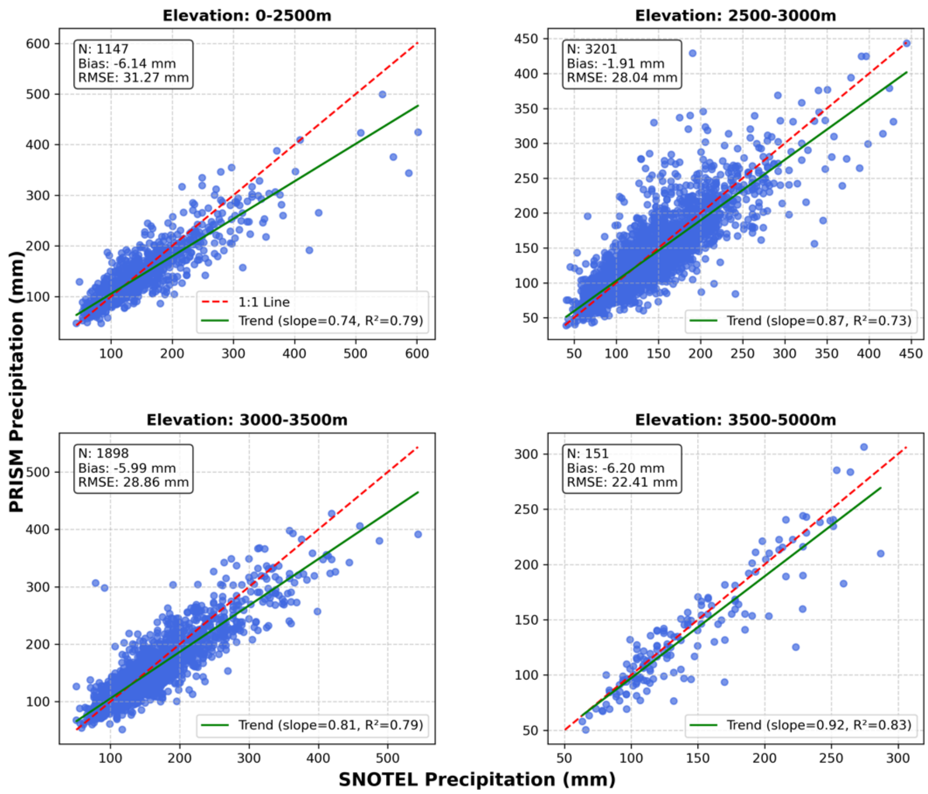


Figure 15. Comparison of PRISM and SNOTEL precipitation measurements across elevation bands in the CRB. Each point represents a station-to-pixel comparison for one year.

(Upper Colorado: 2430 m, Green: 2114 m), which suggests that basin-wide averaging likely dampened the magnitude of these elevation-specific biases.

For T_{min}, a strong negative bias pattern emerges across all elevation bands and months. This bias reaches up to -8 to -10 °C in the highest elevation bands in the winter months. This indicates that Livneh et al. (2015) consistently underestimates minimum temperatures relative to PRISM. These minimum temperature biases are particularly important for cold-season processes like snow accumulation and melt, as they control phase partitioning of precipitation and the persistence of snowpack throughout the winter season.

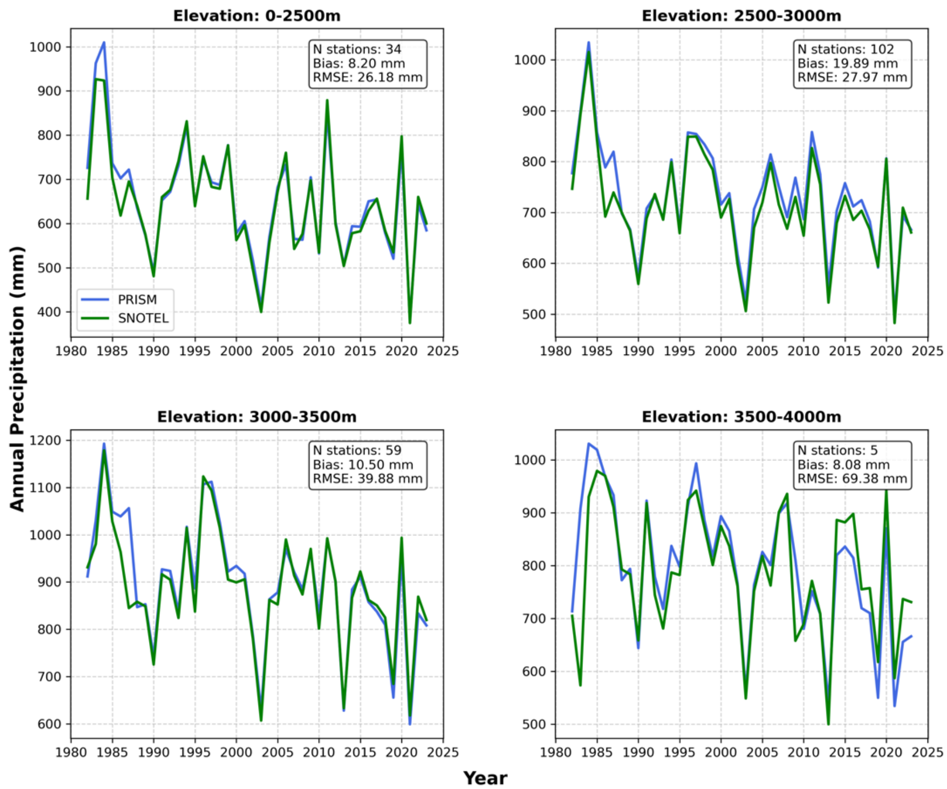


Figure 16. Comparison of PRISM and SNOTEL precipitation measurements across elevation bands in the CRB. Yearly values are obtained by averaging total annual precipitation over all stations.

The observed temperature biases can be attributed to methodological differences in the construction of the datasets. The Livneh et al. (2015) dataset employs a fixed lapse rate of 6.5 °C/km for temperature interpolation, introducing systematic elevation-dependent biases (Shulgina et al., 2023). This approach mainly affects T_{min} estimates, where the fixed positive lapse rate generates substantial cold biases at high elevations. Such biases have implications for hydrological modeling, leading to overestimating snow accumulation during certain months. These findings strongly support the decision to use PRISM rather than Livneh et al. (2015) as the primary forcing dataset in this study, as its variable lapse rate approach better captures the complex temperature patterns across the diverse topography in the CRB. The improved representation of temperature gradients is crucial for accurately simulating snowpack dynamics and subsequent snowmelt processes, which are central to understanding streamflow generation in the basin.

In addition to comparing PRISM with the Livneh et al. (2015) dataset, PRISM was evaluated against SNOTEL ground observations across different elevation bands (Figure 15 and Figure 16). Scatter plot analysis (Figure 15) shows good agreement between PRISM and SNOTEL precipitation, with R^2 values of 0.73-0.83 and consistent negative biases (-5.91 to -6.20 mm) across all elevation ranges. When examining time series data (Figure 16), an interesting contradiction appears. While scatter plots show PRISM underestimating precipitation at individual stations, the time series shows PRISM often overestimating precipitation when averaged across stations. This discrepancy results from spatial averaging effects and the influence of outliers. The scatter plot considers each station-pixel pair individually, while the time series averages across all stations within an elevation band, allowing extreme or relatively higher values to influence the results. Other discrepancies also arise due to substantial elevation mismatches between stations and their corresponding grid cells. For example, the Black Mesa station (3525m) corresponds to a PRISM grid cell with a mean elevation of 2845m, creating significant comparison challenges due to orographic effects.

For temperature, PRISM T_{min} values are systematically lower than SNOTEL observations. This discrepancy relates to changes in SNOTEL instrumentation as documented by Oyler et al. (2015). SNOTEL T_{min} observations increased across the western U.S. during sensor relocation and equipment replacement campaigns, with new sensors exhibiting warm bias at colder temperatures. PRISM documentation indicates these biases were addressed in their processing workflow, further supporting the use of PRISM for basin-wide modeling (PRISM Climate Group, 2025).

3.1.2 Comparison of Snow Products

SWE measurements from SNOTEL ground stations were compared with two gridded datasets (SNODAS and UA) to evaluate the accuracy of gridded snow products in the CRB. This comparison was conducted across four elevation bands (0-2500 m, 2500-3000 m, 3000-3500 m, and 3500-5000 m) for the period 2004-2021 (overlap period of SNODAS and UA) to assess how well these products represent snowpack conditions at different elevations.

Figure 17 shows the time series comparison of annual mean daily *SWE* values for each elevation band. General agreement in temporal patterns was observed among all three datasets, with both gridded products capturing the year-to-year variability recorded by SNOTEL measurements. However, notable differences in magnitude are apparent. The SNODAS product (red line) is found to moderately underestimate SWE across all elevation bands compared to SNOTEL observations (dotted black line), with biases ranging from -4.25 mm at mid-elevations (3000-3500 m) to -7.62 mm at lower elevations (0-2500 m). More substantial underestimation is shown by the UA product (blue line), particularly at higher elevations where snowpack is most significant. Bias values for UA ranged from -10.73 mm at lower elevations to -29.69 mm at the 3000-3500m band. This systematic underestimation suggests limitations in the UA product's ability to represent snowpack in mountainous terrain, which may impact hydrological modeling efforts that

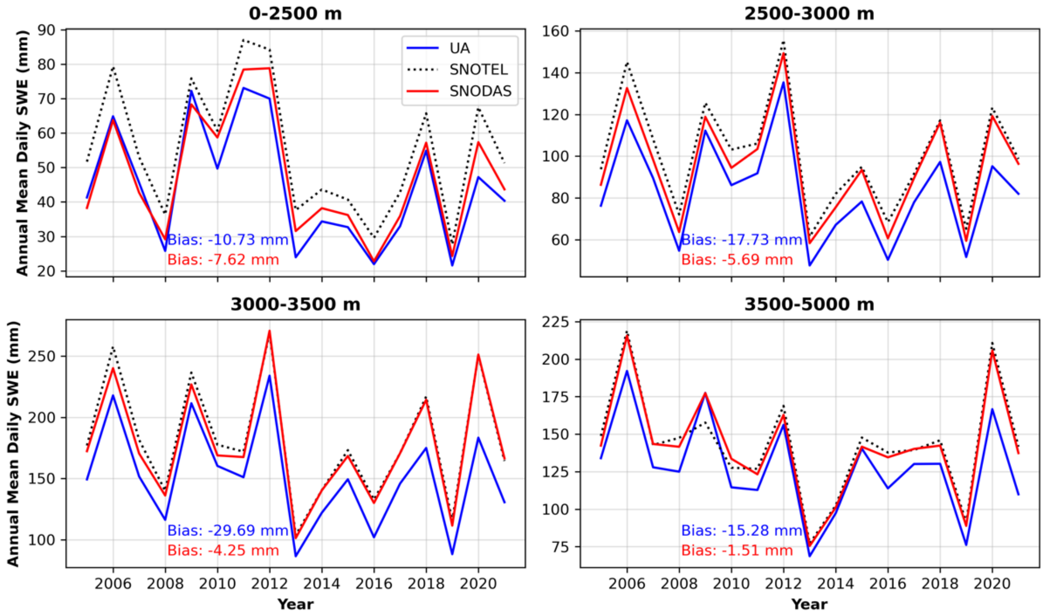


Figure 17. Comparison of SNOTEL, SNODAS and UA annual mean daily *SWE* across elevations (2004-2021).

rely on accurate snow input data; therefore, the UA product will not be used for further analysis in this study.

Through this comparative analysis, valuable insights were gained regarding the relative performance of gridded snow products. While both products could potentially serve as references, this study ultimately focuses on SNOTEL observations for model calibration due to their direct measurement approach and established reliability in operational forecasting contexts. However, understanding these inter-dataset differences remains valuable for contextualizing the snow representations across the CRB and may inform future work where gridded products could complement point-based observations.

3.1.3 Spatio-temporal analysis of SMAP

Before analyzing its spatio-temporal trends, the reliability of the SMAP dataset was evaluated through comparison with available ground observations. Comparing SMAP with field observations presented challenges due to the spatiotemporal constraints of ground *SM* networks. For this validation, surface *SM* observations from 116 SNOTEL stations (Figure 6) were compared with corresponding SMAP surface *SM* data at their collocated pixel. Surface measurements were selected for this validation as they represent SMAP's direct measurements and also due to a limited number of observations that include a rootzone depth of 100 cm. The validation metrics included correlation, root mean square error (RMSE), mean absolute error (MAE), and bias.

As shown in Figure 18, SMAP captures *SM* patterns with moderate success. The correlation histogram reveals that most stations exhibit correlation values between 0.3 and 0.7, indicating a moderate relationship between satellite and ground observations. RMSE and MAE distributions demonstrate relatively low errors, with most values clustering near 0.1 m^3/m^3 or below, suggesting SMAP represents soil moisture conditions reasonably well across the basin. The bias distribution centers approximately around zero but skews slightly negative, indicating a minor systematic underestimation of *SM* by SMAP at some locations.

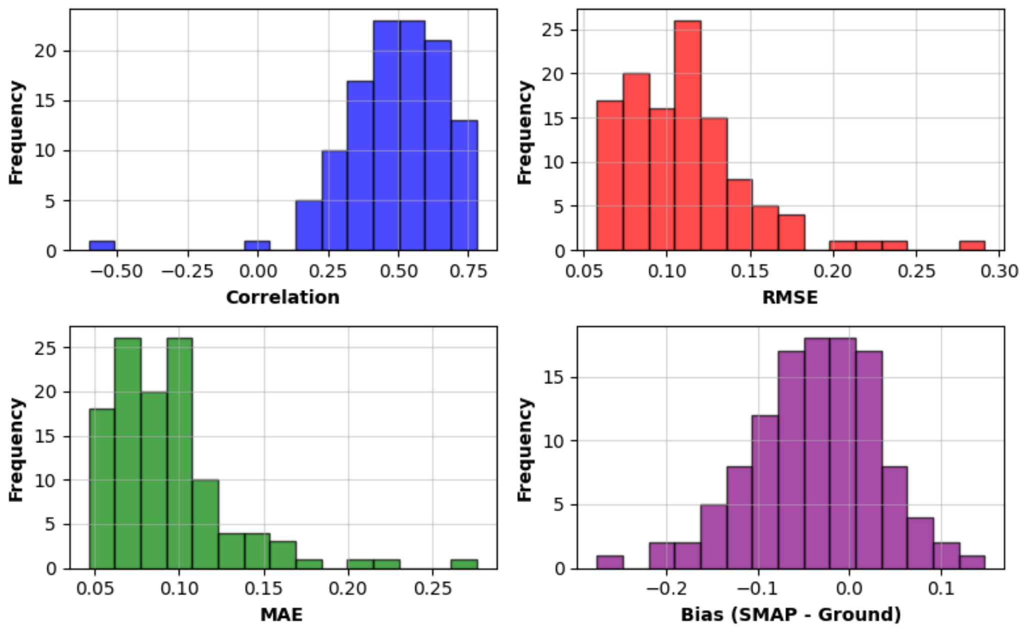


Figure 18. Histograms for error metrics for surface *SM* for 116 ground observations vs SMAP.

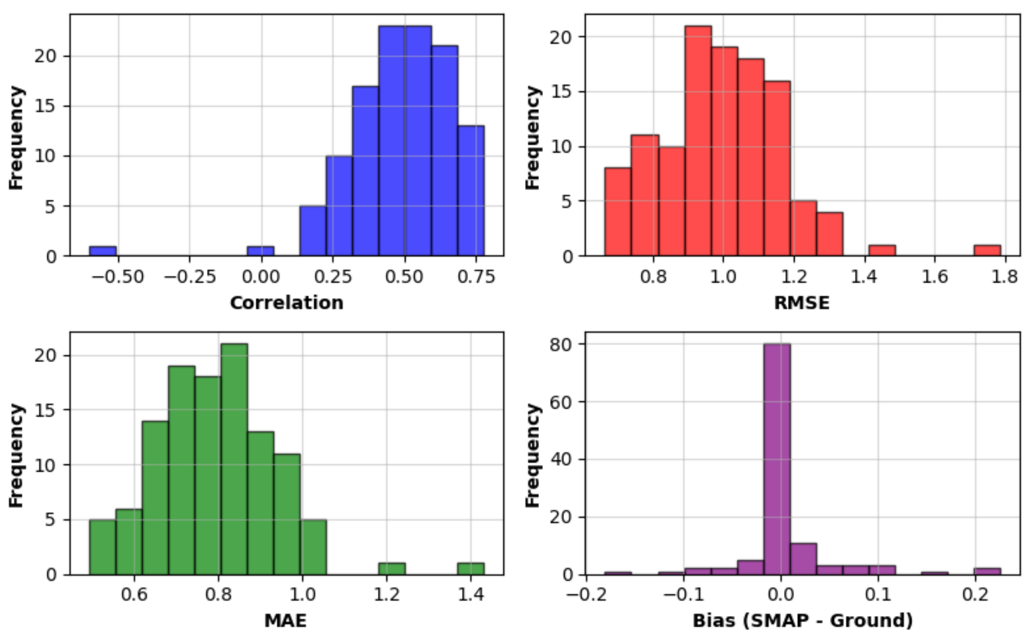


Figure 19. Histograms for error metrics for surface *SM* for 116 ground observations vs SMAP in terms of standardized anomalies.

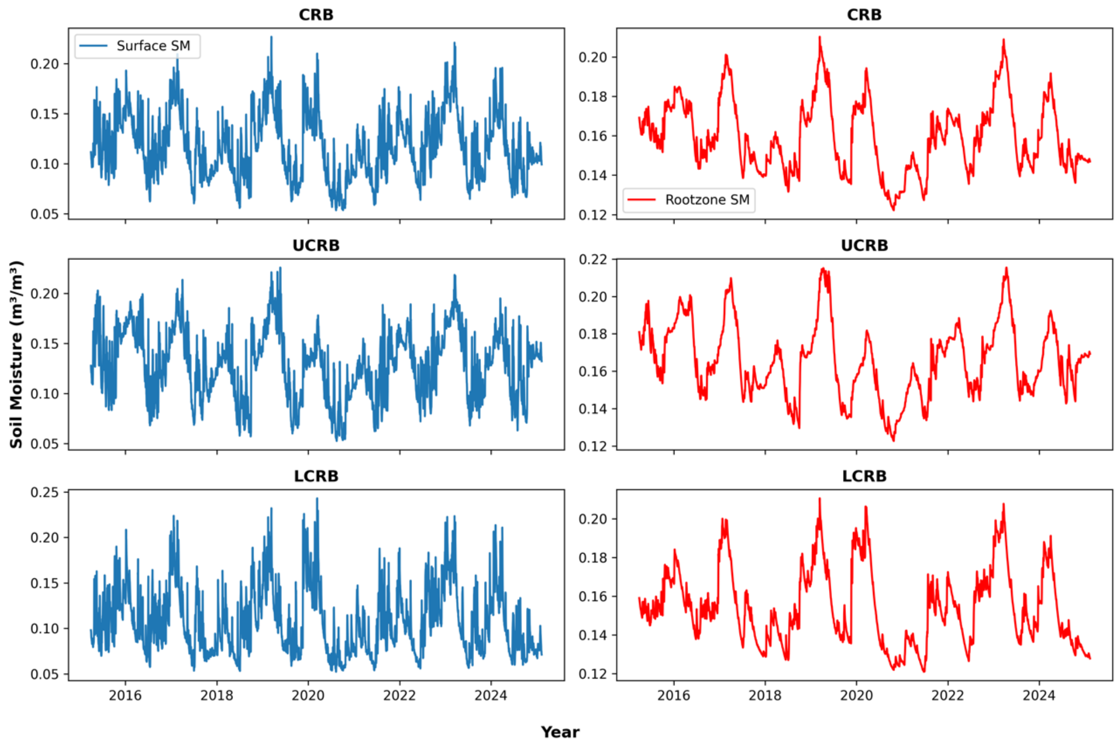


Figure 20. Time series of SMAP *SM* (m^3/m^3) over sub-basins (CRB, UCRB and LCRB in rows) and depth (surface and rootzone in columns).

Following Koster et al. (2009), standardized anomalies were also used for comparison since remote sensing algorithms often have their own physical mechanisms that influence absolute *SM* values. This approach normalizes both datasets, allowing for more meaningful comparison of temporal dynamics rather than absolute magnitudes. Figure 19 presents these standardized anomaly comparisons, which demonstrate improved agreement between SMAP and ground observations, with higher correlation values and reduced bias. These validation results provide sufficient confidence in the SMAP dataset for subsequent analysis.

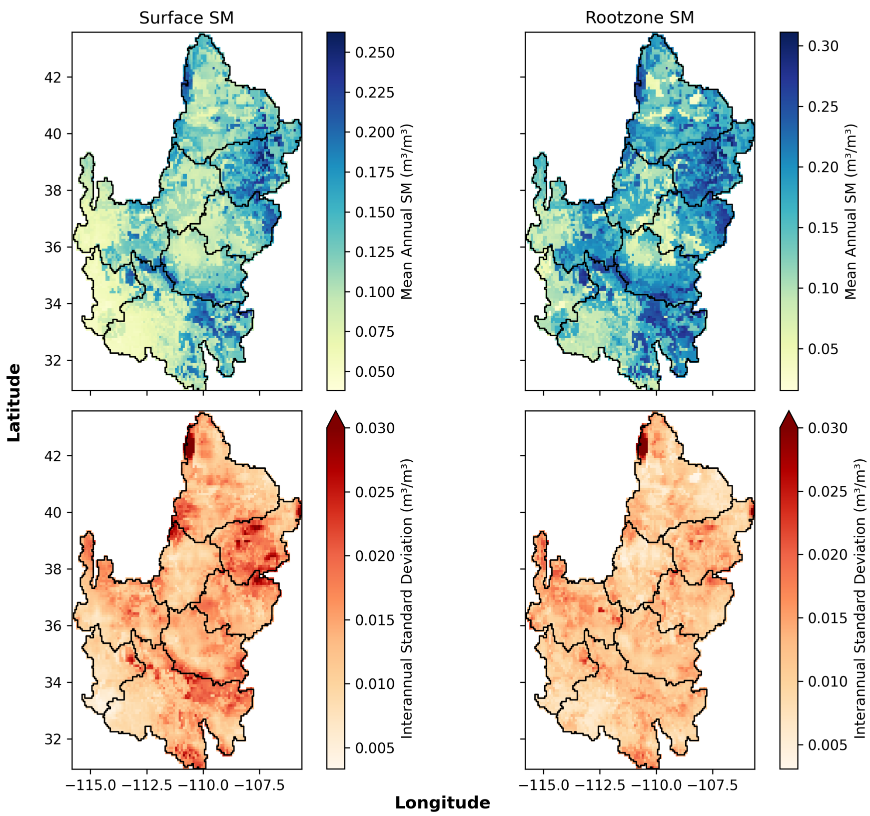


Figure 21. Mean annual values and standard deviation for surface and rootzone SM across 2015-2024.

Then, SMAP observations were analyzed to characterize *SM* dynamics across the CRB during 2015-2024. Figure 20 compares the SMAP surface (0-5 cm) and root zone (0-100 cm) *SM* time series across the basin over the time period. Surface *SM* shows high-frequency fluctuations, while root zone moisture displays more dampened oscillations with clearer seasonal patterns. Notable differences are observed between the UCRB, which maintains relatively higher moisture levels (0.16-0.18 m^3/m^3), and the LCRB, which experiences more pronounced dry periods, particularly during 2020-2022.

Further, the spatial maps (Figure 21; top panel) reveal the distribution of mean annual *SM* across the basin, with higher values concentrated in the northeastern mountainous regions and lower values in the southwestern deserts. The interannual variability maps (bottom panels) highlight areas experiencing greater year-to-year fluctuations. Figure 22 further illustrates the monthly climatology of *SM* across the basin divisions. All regions display a distinct seasonal cycle with peak moisture in the winter-spring months (January-May) followed by summer dry down, with the UCRB maintaining consistently higher moisture levels than the LCRB. The rootzone *SM* (bottom row) shows less pronounced seasonality than surface measurements (top row), demonstrating the buffering effect of deeper soil layers against short-term meteorological variations. Together, these analyses reveal important spatial and temporal patterns in *SM* that help explain hydrological responses across the basin. Root zone *SM* was selected for further analysis because it better represents the total water available (including *SM* from the surface layer) to vegetation and provides a more stable signal that captures persistent hydrological patterns relevant to drought propagation and streamflow generation.

The spatial analysis of SMAP (Figure 23a) reveals considerable heterogeneity in rootzone *SM* trends across the basin. Notable drying trends (red areas) are evident in the northwestern portion of the Green sub-basin, with *SM* decreasing at rates approaching -3 mm/yr with similar drying patterns appearing in the central regions (Glen Canyon and Grand Canyon). Conversely, several areas in the northeastern and central portions of the basin exhibit wetting trends (blue areas) of 1 to 3 mm/yr. When aggregated by sub-basin (Figure 23b), a clearer spatial pattern emerges, with northern sub-basins experiencing the most severe drying trends (exceeding -1.0 mm/yr for Green, Grand Canyon, and Glen Canyon), while southern regions show more moderate declines.

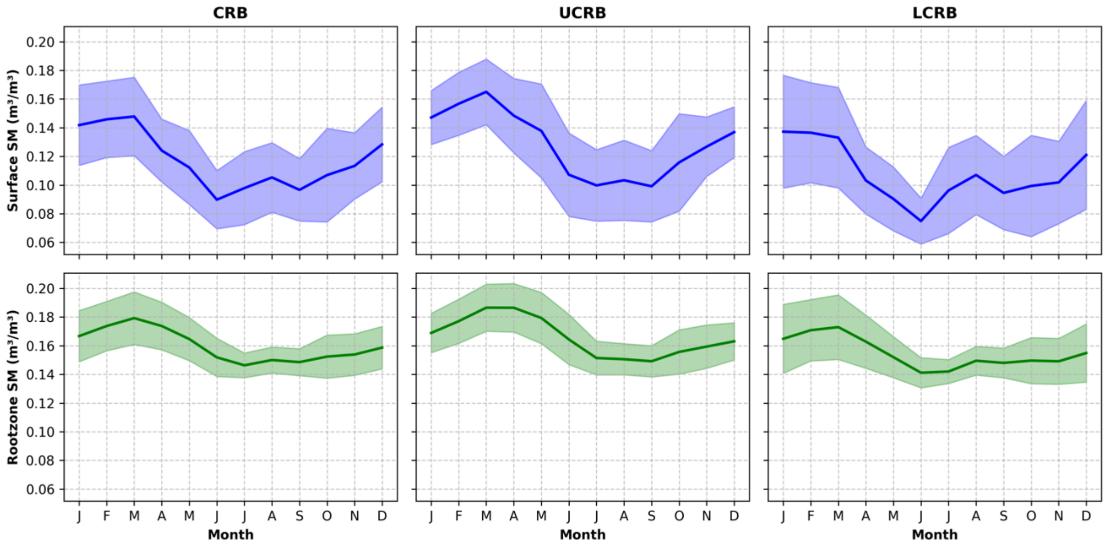


Figure 22. Mean monthly values averaged across 2015-2024 for the CRB, UCRB, and LCRB. Shaded lines represent one standard deviation value.

When overlaying these trends with the vegetation distribution (Figure 3), it is observed that Shrub/Scrub, which dominates the basin landscape (50.7% of the total area), exhibits a drying trend of nearly -0.9 mm/yr (Figure 24). This widespread vegetation class largely corresponds to the mixed pattern of moderate drying seen across the central and southern portions of the basin. Similarly, Evergreen Forest (16.2% of basin area) shows a comparable drying trend and is primarily distributed in the higher elevation regions where some of the stronger drying signals are visible. The most intense drying trends are observed in Perennial Ice/Snow (~-1.8 mm/yr) and Developed/High-Intensity areas (~-1.3 mm/yr), though these represent minimal portions of the basin (less than 0.1% each). Only Developed/Open Space areas, which comprise a negligible portion of the basin (less than 0.1%), exhibit a slight wetting trend (+0.1 mm/yr).

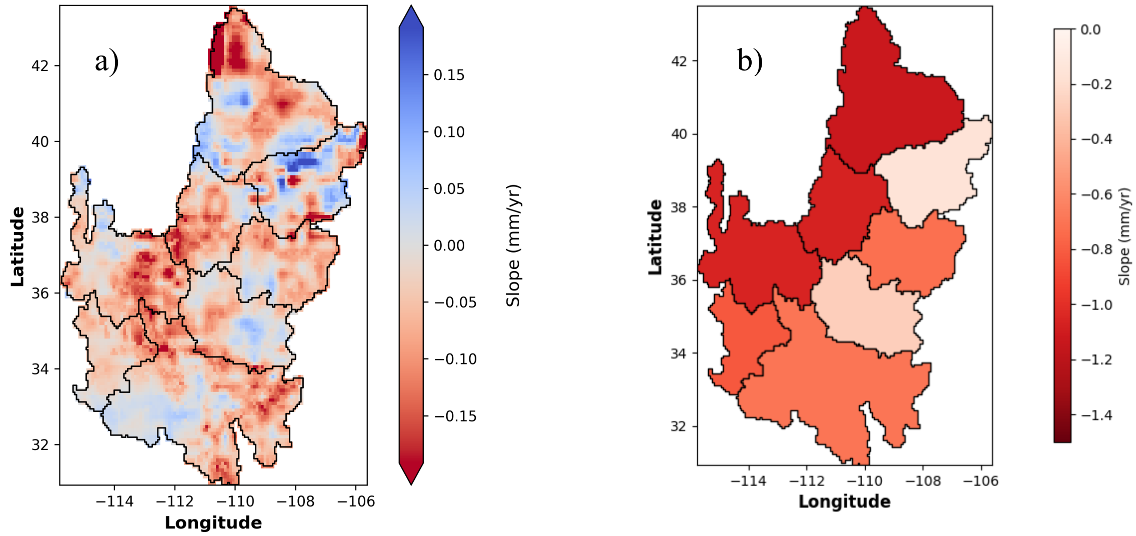


Figure 23. Trends in root zone *SM* (100cm) over the period of 2015-2024 shown by SMAP (a) for each pixel (b) averaged over sub-basins.

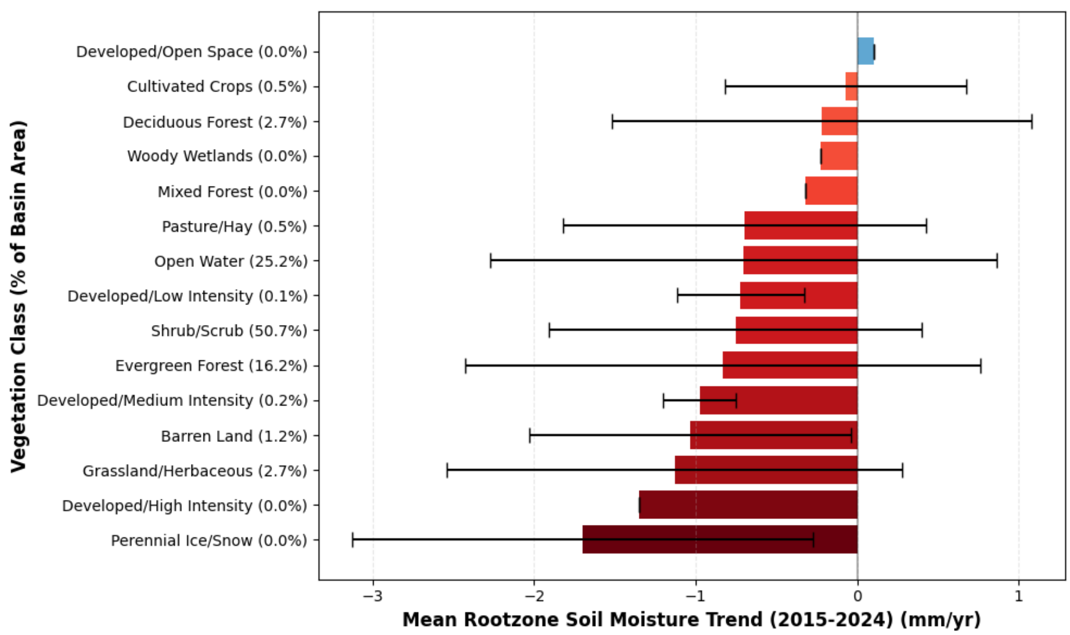


Figure 24. Trends in SMAP root zone *SM* (100cm) across different land cover types. Horizontal lines represent errors (one standard deviation).

These spatial patterns of *SM* change have important implications for water management, allowing for targeted interventions in high-risk areas. Additionally, these spatial patterns help detect potential "hotspots" where water loss through *SM* might trigger cascading hydrological impacts, including reduced baseflow to streams and declining groundwater recharge. The analysis also establishes a valuable baseline for model validation. Further, the contrast between time series patterns and spatial trend magnitudes presents an interesting hydrological dynamic. While the time series analysis showed that the LCRB experiences lower average *SM* values and more pronounced dry periods than the UCRB, the spatial trend analysis reveals that the UCRB sub-basins are drying faster. This suggests that although the LCRB maintains consistently lower *SM* conditions, the UCRB is undergoing more rapid moisture depletion over time. This pattern indicates that the historically wetter UCRB may be experiencing more significant hydrological changes under recent drought conditions.

It would be interesting to compare these *SM* patterns with other studies, but existing research in the CRB using SMAP remains limited. This analysis provides a unique and comprehensive spatio-temporal assessment of *SM* dynamics using SMAP across multiple sub-basins, along with an unprecedented approach of validation of the SMAP values with ground networks in the CRB. The approach of examining *SM* trends across different vegetation types and soil depths offers a more nuanced understanding of hydrological processes in the CRB that were often obscured by ground data limitations.

3.1.4 Spatio-temporal Analysis of GRACE

GRACE satellite observations (anomalies to 2004 – 2009 mean, a standard approach in GRACE analysis) were analyzed to understand terrestrial water storage changes across the CRB from 2002 to 2024. This time period captures both the long-term drought conditions and the intensification after 2020, providing insights into how water storage components have responded to these changing conditions.As seen in Figure 25, both UCRB and LRCB show concerning long-term declining trends in terrestrial water storage for the period of 2002-2024. The UCRB shows a modest decline with a trend of -0.48 km^3/year, resulting in a total depletion of approximately -10.71 km^3 over the study period. It exhibits considerable interannual variability, with several periods of recovery interspersed throughout the record. In contrast, the LCRB demonstrates a much more severe decline with a trend of -1.75 km^3/year, resulting in a total depletion of -39.37 km^3, which is more than three times that of the UCRB. These numbers are consistent with recent findings from Scanlon et al. (2025). The LCRB shows rapid storage decreases after 2012, with *TWSA* values depleting to approximately -40 km^3 during 2021-2022, representing the most severe drought conditions in the entire record. Time series of individual components of GRACE is presented in Appendix H.

Trends for individual sub-basins for the entire record are shown in the spatial map in Figure 26. Note that the northern sub-basins show the lowest rates of decline with trends around -0.1 to -0.2 km^3/year, while the central sub-basins display moderate declines with trends around -0.4 to -0.6 km^3/year. Southern sub-basins exhibit the most severe declines, with trends exceeding -0.8 km^3/year. The Gila and Lower Colorado sub-

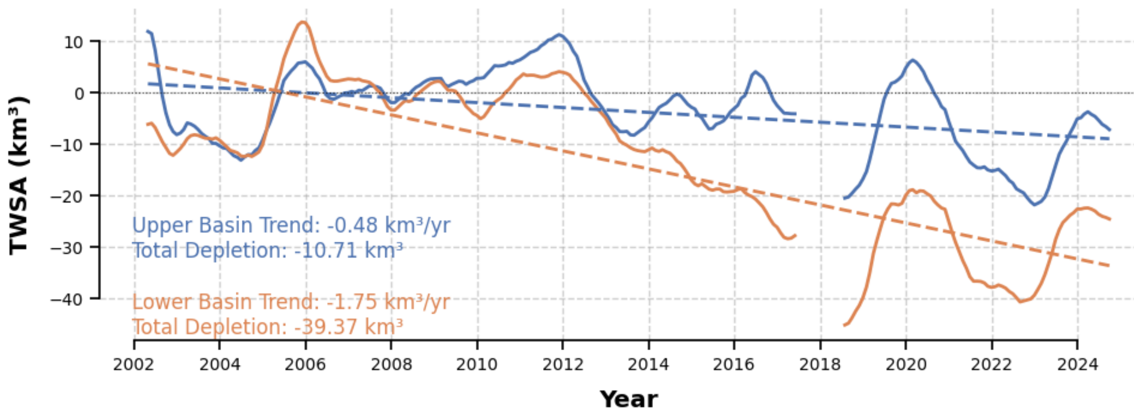


Figure 25. Water depletion (*TWSA*) from GRACE anomalies averaged over the UCRB (blue) and LCRB (orange). The lines represent water storage changes (km³) averaged over a 12-month rolling window.

basins are experiencing the most severe water storage declines. The spatial pattern of water loss corresponds with the division between the UCRB and LCRB shown in the time series, confirming that the LCRB is experiencing greater water losses than the UCRB.

Figure 27 shows the seasonal *TWSA* value (cm), comparing the pre-2020 drought period (2002-2019 mean, top row) with the post-2020 drought period (2020-2024 mean, bottom row). During the winter months (Jan-Mar), the pre-2020 period shows predominantly neutral to slight positive anomalies (light blue) in the northern regions. In contrast, the post-2020 period exhibits significant negative anomalies (red) in the southwestern portions of the basin. Similarly, for spring (Apr-Jun), the post-2020 drought period displays intensified negative anomalies, particularly in the southwestern regions. Summer and fall (Jul-Sep and Oct-Dec) show the most pronounced differences between the two time periods. The post-2020 period shows severe water deficits (dark red)

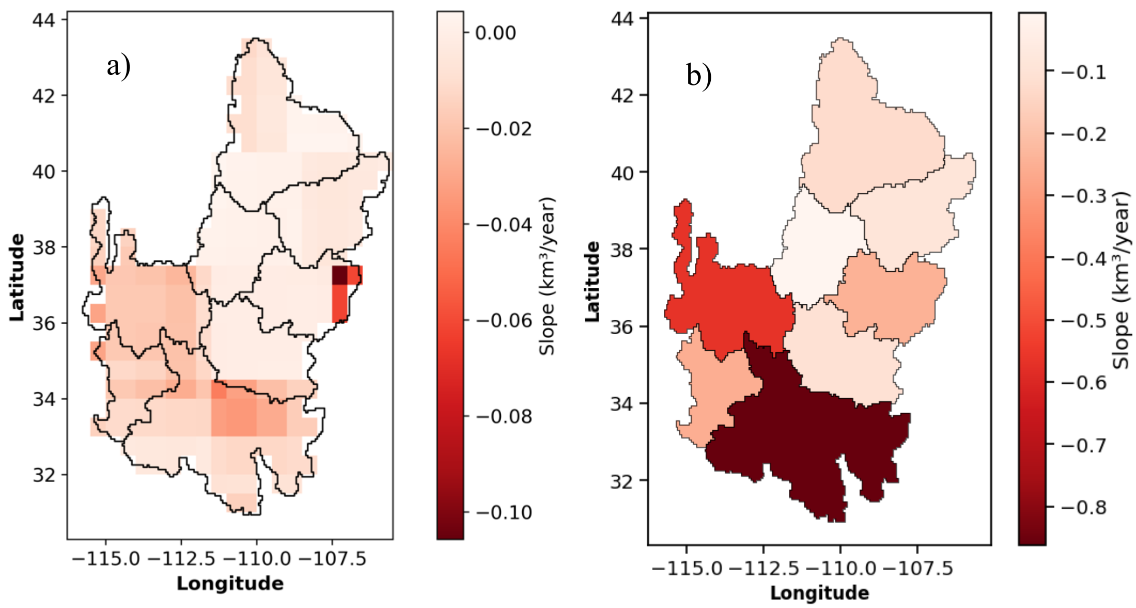


Figure 26. *TWSA* trends (km^3/year) in the CRB (2002- 2024) for (a) each GRACE pixel (50 km) (b) averaged over major sub-basins.

throughout most of the basin, with the southwestern regions experiencing the most extreme losses (exceeding -20 cm).

The spatial analysis reveals that drought impacts are not uniform across the basin. The LCRB consistently shows the most severe water losses across all seasons in the post-2020 period, with *TWSA* values frequently exceeding -15 to -20 cm. The UCRB generally maintains less severe anomalies, likely due to its higher elevation and different precipitation regimes. This regional difference in drought response aligns with findings from Castle et al. (2014), which documented greater groundwater depletion in the LCRB than in UCRB during the 2004-2013 drought period. The progression of negative anomalies throughout the seasonal cycle in the post-2020 period suggests cumulative water storage deficits that are not being replenished by the winter precipitation.

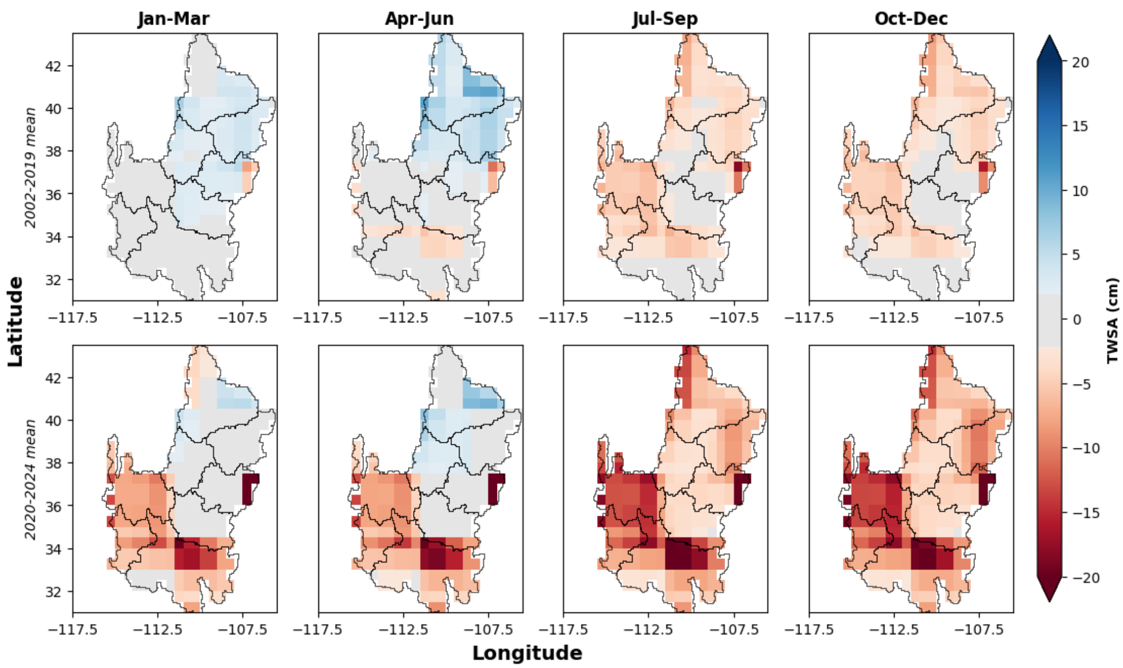


Figure 27. Spatial variation of GRACE *TWSA* across seasons (columns) and two periods (rows, 2002-2019 and 2020-2024).

3.2 Model Calibration

A calibration approach was implemented to ensure accurate representation of hydrological processes in the CRB. First, different VIC configurations were evaluated to determine the optimal setup. Snow parameters were then calibrated using SNOTEL observations, followed by soil parameter adjustments to improve streamflow simulations. The calibrated model was then validated against independent SMAP and GRACE observations. This framework established confidence in the model's ability to represent the complex hydrology in the CRB for further experimental analysis.

3.2.1 Comparison of VIC Configurations

The performance evaluation of four VIC model configurations across 1982-2018 (baseline forcings from Livneh et al. (2015) end in 2018) reveals significant improvements when both forcing data and snow parameterization are updated (Figure 28; Table 5). This comparison mainly focuses on the UCRB and its sub-basins since these regions generate more than 80% of the CRB's annual streamflow (Christensen et al., 2004). The combined updates (upd_forcing_snow) consistently demonstrate superior performance with high NSE values (>0.9), along with high correlation coefficients (CC > 0.95) and minimal bias (close to zero in the UCRB). Individual updates show mixed results: updated forcing data alone (upd_forcing) reduces model skill in all basins compared to the baseline, with particularly poor performance in the Upper Colorado sub-basin (NSE: 0.161). Meanwhile, updated snow parameterization alone (upd_snow) performs well only in the San Juan basin (NSE: 0.793) but exhibits degraded performance elsewhere, even producing negative NSE (-0.22) in the Green sub-basin.

The differences in model performance likely stem from how each modification affects streamflow generation processes. The upd_forcing configuration uses PRISM P and T inputs while using parameters calibrated for Livneh et al. (2015) forcings, causing discrepancies in the streamflow produced. In the upd_snow configuration, the wet-bulb temperature-based precipitation partitioning method modifies the snowfall fraction in ways that increase snow accumulation compared to the standard air temperature method (Wang et al., 2024). However, this increased snowfall required recalibration of soil parameters, hence the poor performance.

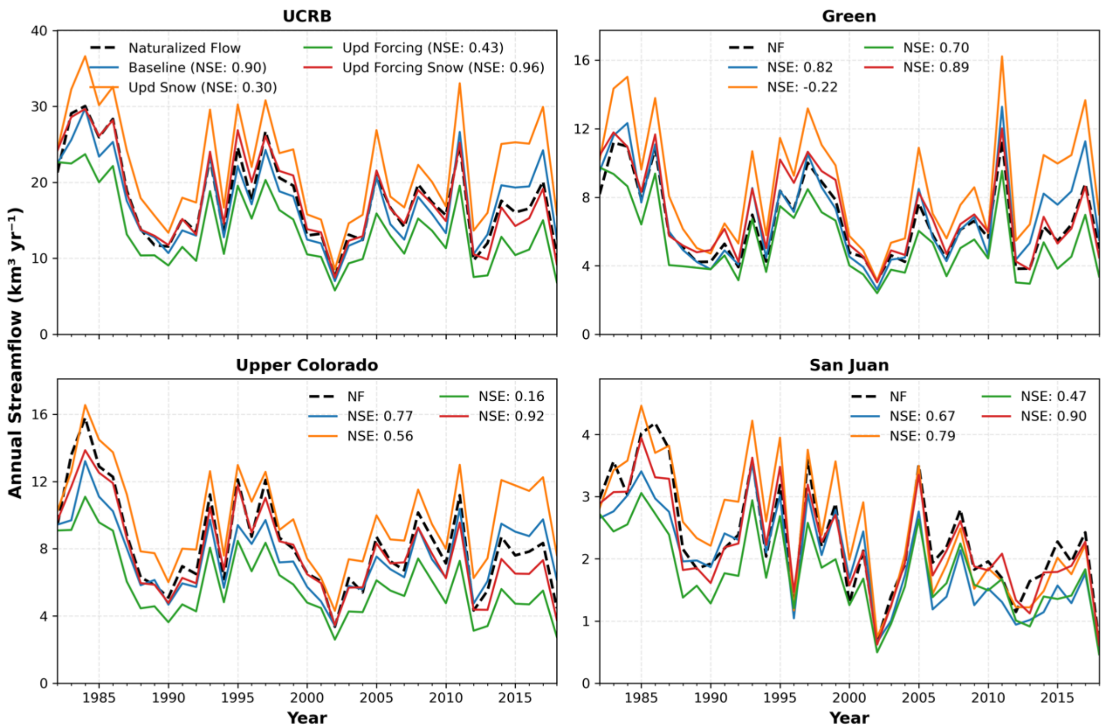


Figure 28. Comparison of VIC streamflow performance across different configurations of versions and forcings across 1982-2018.

The performance improvement in the combined approach likely results from complementary effects between the two updates that cancel out some of the biases in each individual component. The Livneh et al. (2015) dataset tends to produce colder temperatures at higher elevations, which likely resulted in inaccurate snow accumulation when using the standard air temperature-based precipitation partitioning method. By simultaneously replacing it with PRISM, which has less cold bias at high elevations, the combined approach creates a more balanced representation of precipitation phase partitioning and subsequent streamflow generation, hence a good performance is observed.

Table 5. Performance metrics for major sub-basins across different VIC configurations across 1982-2018.

Basin	Configuration	Bias (km^3/yr)	RMSE (km^3/yr)	CC	NSE
UCRB	baseline	-0.02	1.84	0.95	0.90
	upd_forcing	-0.23	4.32	0.97	0.43
	upd_forcing_snow	0.01	1.10	0.98	0.96
	upd_snow	0.24	4.82	0.95	0.30
Green	baseline	0.06	0.97	0.95	0.82
	upd_forcing	-0.16	1.25	0.96	0.70
	upd_forcing_snow	0.08	0.78	0.98	0.89
	upd_snow	0.34	2.54	0.95	-0.22
Upper Colorado	baseline	-0.08	1.35	0.92	0.77
	upd_forcing	-0.29	2.57	0.97	0.16
	upd_forcing_snow	-0.06	0.81	0.98	0.92
	upd_snow	0.19	1.87	0.93	0.56
San Juan	baseline	-0.15	0.51	0.91	0.67
	upd_forcing	-0.24	0.65	0.94	0.47
	upd_forcing_snow	-0.04	0.28	0.96	0.90
	upd_snow	0.05	0.41	0.92	0.79

The results demonstrate a clear utility in combining both updates, as neither improvement alone consistently enhances model performance. However, additional refinements are needed to improve snow accumulation and melt processes when compared to SNOTEL ground measurements and also to better capture the timing and magnitude of monthly streamflow patterns. These refinements, which will be addressed in later sections, will further enhance model performance and build confidence in the subsequent outputs produced.

3.2.2 Snow Parameters Calibration

Before calibration, bootstrapping analysis was done to evaluate potential biases introduced in the calibration process through the subset of selected stations with respect to the mean conditions represented by the entire set of 200 SNOTEL stations. Figure 29 shows that with the selection of 163 stations, potential error was maintained below 5% based on normalized RMSE. This confirms that the selected stations provide a reasonable balance between data quality and spatial representativeness. Further sensitivity analysis showed that stricter thresholds (50m, 100m) would limit the available stations to 52 and 96, respectively, which would increase the normalized RMSE to more than 10%. More lenient thresholds (500m) included nearly all stations (192) but potentially introduced elevation-related biases. Therefore, the chosen 250m threshold provided 163 stations, corresponding to acceptable error limits (below 5% normalized RMSE) and strong basin-wide representation while maintaining reasonable elevation correspondence.

The snow module calibration focused on modifying albedo parameters to improve the model representation of *SWE* in terms of timing and magnitude. The calibration process required adjustments to three key snow albedo parameters: (i) Snow albedo accumulation parameter *A* was increased from 0.94 to 0.98, (ii) Snow albedo thawing parameter *A* was increased from 0.80 to 0.92, and (iii) Snow albedo thawing parameter *B* was increased from 0.46 to 0.85. The parameter adjustments resulted in higher albedo values during both snow accumulation and ablation phases, which reduced the rate of simulated snowmelt and better matched observed conditions from SNOTEL. Figure 30 illustrates the outcomes of this calibration effort across four elevation bands (0-2500 m,

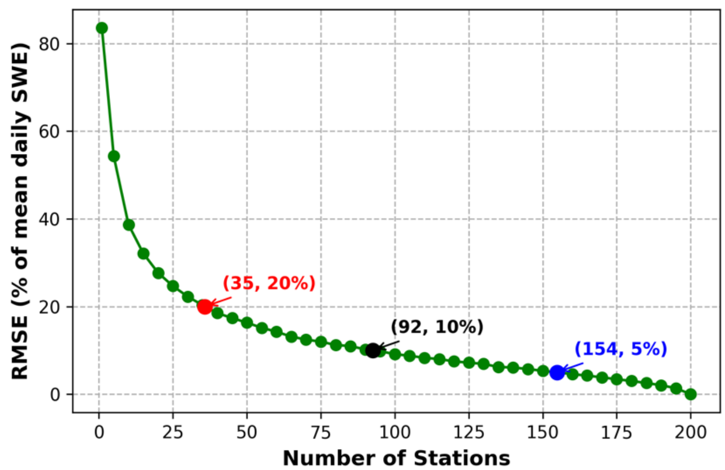


Figure 29. Bootstrapping analysis showing the relationship between the number of SNOTEL stations used for calibration and the normalized Root Mean Square Error (RMSE) expressed as a percentage of mean daily SWE.

2500-3000 m, 3000-3500 m, and 3500-5000 m) over the entire CRB. The mean *SWE* values were obtained by averaging daily *SWE* measurements for each water year across all SNOTEL stations and their corresponding collocated VIC grid cells within each elevation band.

The calibration substantially reduced the negative bias observed in the baseline simulation when compared to SNOTEL measurements, though with varied effects across elevation bands. In the lowest elevation band (0-2500 m), the bias shifted from -1.92 mm to +6.69 mm, indicating a slight overestimation. It is worth noting that since these albedo parameters were applied uniformly across the entire basin (as they are defined in the global parameter file), the calibration led to some overestimation in lower elevations while correcting the more significant underestimation in higher elevations.

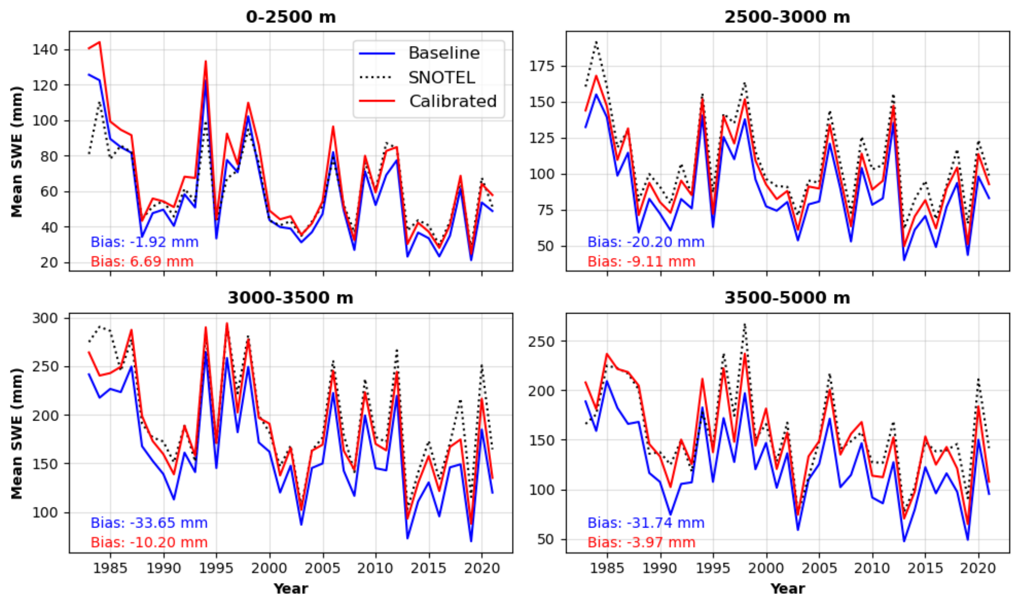


Figure 30. Comparison of mean *SWE* between observations (SNOTEL), baseline, and calibrated conditions across elevation bands (1984-2020).

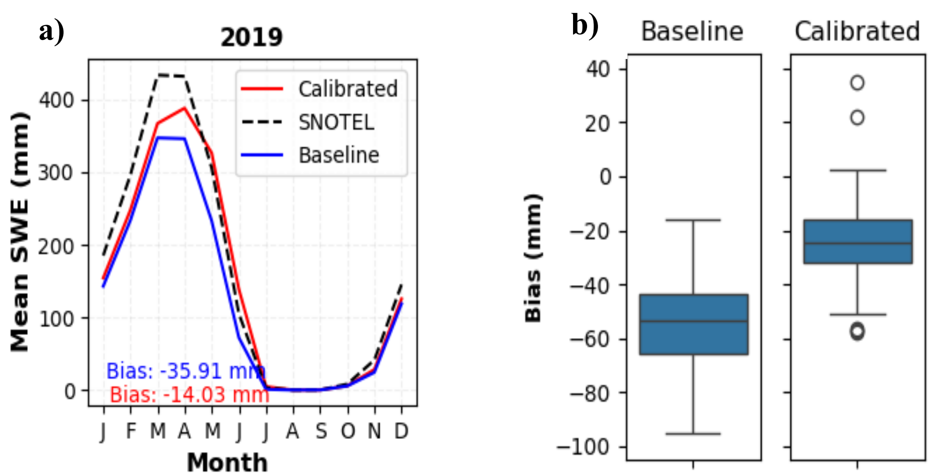


Figure 31. (a) Comparison of mean *SWE* across SNOTEL observations, and VIC baseline and calibrated versions for WY 2019 (see Appendix G for all WY comparisons) (b) Box plot summarizing peak *SWE* bias reduction for all years (1984-2020).

However, this overestimation is less critical since the majority of significant snow accumulation in the CRB occurs at higher elevations. More notable improvements were achieved in the 2500-3000 m band, where bias was reduced from -20.20 mm to -9.11 mm

(55% reduction). In the 3000-3500 m band, the bias decreased from -33.65 mm to -10.20 mm (69% reduction), while in the highest elevation band (3500-5000 m), the bias improved from -31.74 mm to -5.57 mm (82% reduction). These results demonstrate that the first calibration target (reducing elevation-based mean negative bias) was successfully achieved for the most hydrologically significant elevation bands.

The WY analysis for 2019 (Figure 31a) reveals that the calibrated model more accurately captures the timing and magnitude of *SWE* throughout the year. For example, in 2019, the mean bias during peak snow season decreased from -35.91 mm in the baseline simulation to -14.03 mm in the calibrated version, representing a 61% improvement. This addresses the second calibration target of reducing peak *SWE* bias by more than 50%. The box plots (Figure 31b) provide statistical evidence of the improvement. The baseline simulation shows a median bias of approximately -45 mm with values ranging primarily between -20 mm and -90 mm. After calibration, the median bias improved to approximately -25 mm with a narrower distribution, indicating both improved accuracy and precision in the model simulations. Details on the performance of each WY is shown in Appendix F.

Overall, the snow parameter calibration significantly improved the model's ability to represent snow processes throughout the CRB, particularly in the higher elevation bands where most of the basin's snow accumulation occurs. Both calibration targets were successfully achieved: elevation-based mean negative bias was substantially reduced in hydrologically important regions and peak *SWE* bias for each year was reduced by more than 50%. However, some biases still remain throughout the simulation period. Further calibration using solely albedo parameters was explored but presented two challenges: (i) the timing of peak *SWE* shifted in ways that were difficult to control, and (ii) excessive snow accumulation and subsequent high streamflow were produced, which complicated the calibration of soil parameters.

Instead, remaining snow biases were addressed through adjustments to snow roughness parameters (section 3.2.3) during the streamflow calibration phase with soil parameters. This approach offers some advantages since snow roughness can be adjusted specifically for individual areas or sub-basins, unlike the uniform application of albedo parameters across the entire basin. This basin-specific calibration strategy provides a better spatial control over snow processes.

3.2.3 Streamflow Calibration

This process built upon the previous snow parameter calibration by further refining the model's representation of streamflow. Table 6 shows that the initial parameters varied slightly across sub-basins, with most areas having similar starting values (D_2 = 0.7 m, Ds = 0.01, B_{inf} < 0.1, and snow roughness = 0.0005 m). The Ds_{max} parameter showed greater initial variation between sub-basins. As shown in Table 7, D_2 was increased across all basins, indicating that greater water storage capacity was needed to represent the basin's streamflow properly. Ds_{max} values were adjusted to represent baseflow contributions better, while increases in Ds were applied to allow for higher baseflow generation at lower *SM* levels. B_{inf} values were increased modestly across most basins, improving the partitioning between surface runoff and infiltration. Snow roughness showed the most substantial relative changes, and these adjustments directly

Table 6. Soil parameters in the major sub-basins before calibration.

Basin/Parameters	**D_2(m)**	**Ds_{max}**	**Ds**	**B_{inf}**	**Snow rough (m)**
Green	0.7	0.1- ~20	0.01	<0.1	0.0005
Upper Colorado	0.7	0.1- ~20	0.01-0.1	<0.1	0.0005
San Juan	0.7	~10-15	0.01	<0.1	0.0005
Glen Canyon	0.7	~5-10	0.01	<0.1	0.0005
LCRB	0.7	~5-10	0.01	0.01-0.4	0.0005

address the need for basin-specific snow parameter adjustments following the albedo calibration.

Figure 32 shows substantial improvements in long-term mean monthly simulated streamflow. The best improvement occurred in the Green River basin, where NSE increased from 0.40 to 0.94 during the calibration period and maintained a high value of 0.94 during validation. This correction significantly reduced the peak flow overestimation evident in the baseline simulation. The UCRB, which is the sum of streamflow from Green, Upper Colorado, San Juan, and Glen Canyon sub-basins, maintained a good performance, with NSE values improving from 0.82 to 0.97 during calibration and from 0.89 to 0.99 during validation. The key improvement was the model ability to capture better the timing and magnitude of peak flows, typically during the late spring and early summer months (May-July) due to snowmelt. The calibrated model also improved the representation of recession curves during the late summer and fall months.

Table 7. Soil parameters in the major sub-basins after calibration. '-' means no change was made.

Basin/Parameters	***D_2*(m)**	***Ds_{max}***	***Ds***	***B_{inf}***	**Snow rough (m)**
Green	2	1.5	0.4	0.14	0.003
Upper Colorado	1.3	-	0.05	0.1	0.001
San Juan	1	0.5	0.4	0.1	0.014
Glen Canyon	1	-	-	-	-
LCRB	1.5	1.5	0.005	-	-

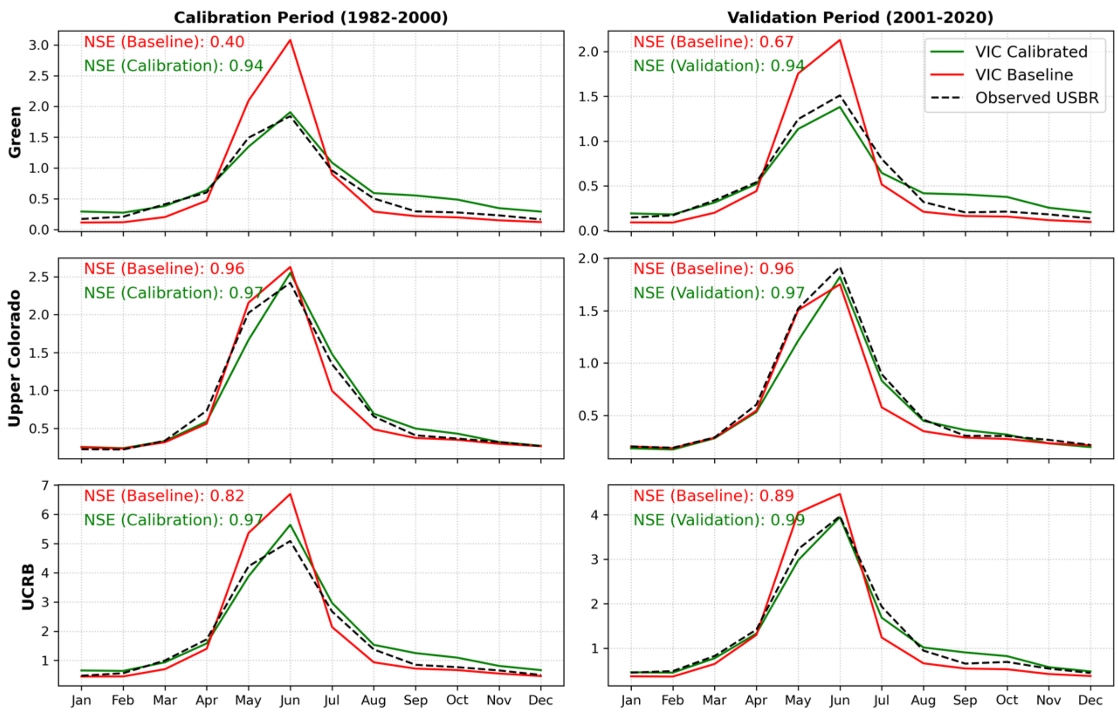


Figure 32. Assessment of baseline and calibrated VIC monthly mean streamflow in the major sub-basins in the CRB for calibration (1982-2000) and validation (2001-2020) periods.

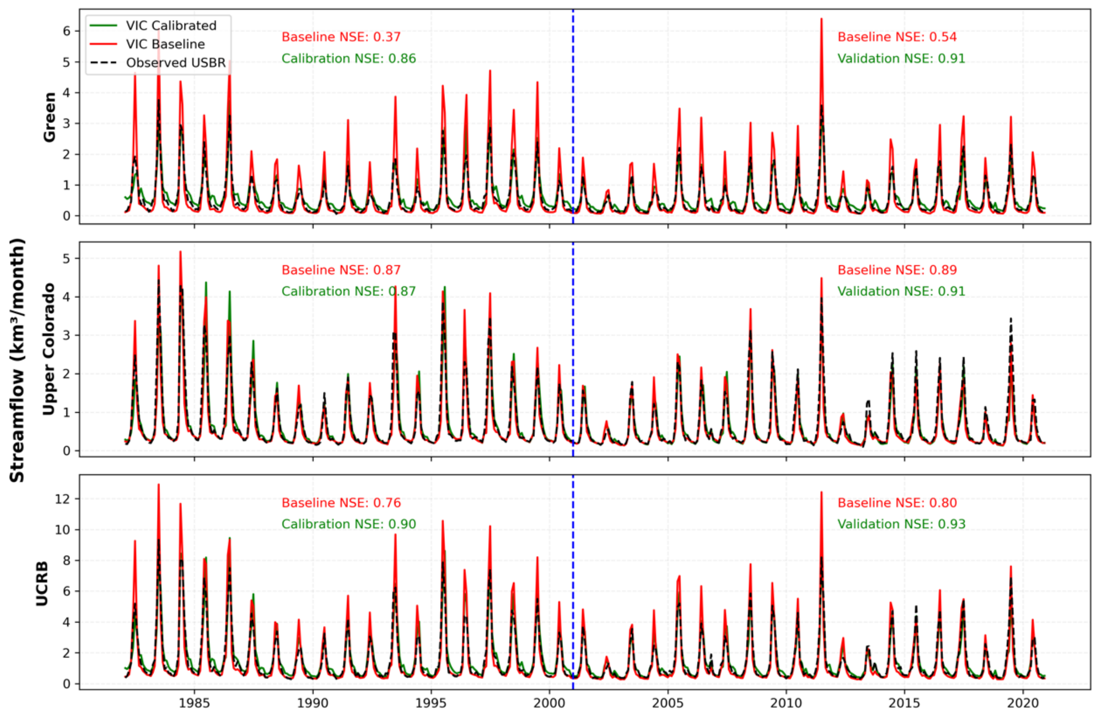


Figure 33. Assessment of baseline and calibrated VIC monthly streamflow time series in the major sub-basins in the CRB for calibration (1982-2000) and validation (2001-2020) periods.

Figure 33 illustrates the model's performance in capturing interannual variability over the entire simulation period. The calibration significantly improved the model's ability to represent peak flows across all basins, with the Green River again showing the most substantial improvement (baseline NSE = 0.37 to calibrated NSE = 0.86 during calibration; baseline NSE = 0.54 to calibrated NSE = 0.91 during validation). The time series analysis reveals that the calibrated model effectively captures the year-to-year streamflow patterns variability, including high-flow and low-flow years. The consistent performance during both calibration and validation periods suggests that the parameter adjustments are robust and not overfitted to specific conditions. Interestingly, the validation period exhibits slightly better performance metrics than the calibration period.

This improvement can be attributed to the calibration approach that balanced representation of both peak and low flows, rather than optimizing solely for long-term monthly averages, as commonly seen in many studies.

Figure 34 confirms the model's ability to accurately represent annual streamflow volumes across the entire CRB (NSE = 0.87) and its major sub-basins. The UCRB shows the strongest performance (NSE = 0.91), while the LCRB exhibits the weakest but still acceptable performance (NSE = 0.65) given the fact that LCRB contributes minimally to streamflow production in the CRB. Despite some limitations in capturing extreme peak flows in the Green basin and lower performance in the LCRB, the overall model performance is robust across the CRB. The calibrated model effectively represents streamflow patterns with high NSE values, demonstrating its reliability for hydrological applications.

The hierarchical calibration approach proved highly effective, with snow parameters first adjusted basin-wide, followed by sub-basin-specific refinements to soil parameters and snow roughness. This calibrated version of the model represents the dynamics of the CRB well and stands out for its applicability for future climate and water resource studies in the region. Appendix G shows the standard model outputs, which detail the basin's hydrological outputs and spatio-temporal trends across months, seasons, and years. These outputs show similar spatial patterns and magnitudes of key hydrological variables (*P*, *SWE*, *ET*, etc.) as reported in Whitney et al. (2023a), providing additional validation of the model's performance.

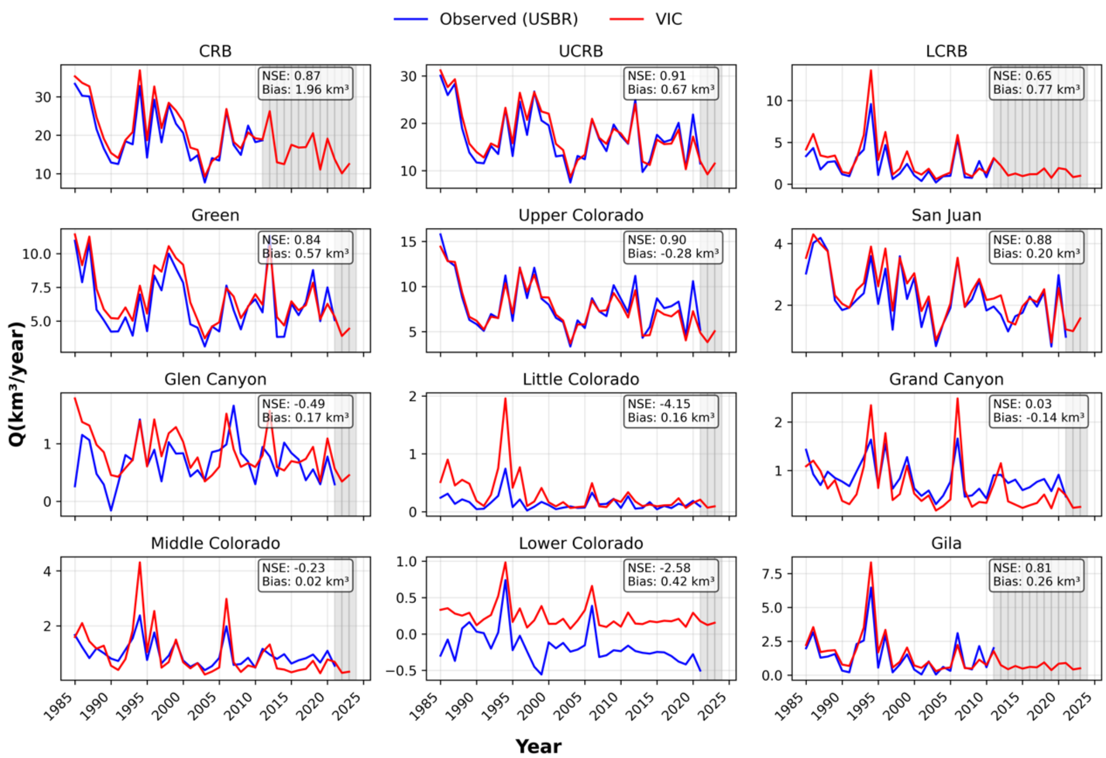


Figure 34. Assessment of VIC yearly streamflow in the major sub-basins in the CRB for the 40-year period (1984-2023) (CRB and Lower Basin have comparisons till 2010 because naturalized streamflow from the Gila Basin outlet was only available till 2010). Shaded area represents period where naturalized streamflow data was not available.

3.2.4 Comparisons of VIC Outputs with Earth Observations

The time series comparisons presented in Figure 35 and Figure 36 illustrate the comparison between VIC model soil *SM* outputs and SMAP across the CRB from 2015 to 2024. For this analysis, the VIC model outputs (originally at 6 km resolution) were upscaled to match SMAP's coarser 9 km resolution (using first-order conservation mapping with the ‘cdo’ operator in Linux), ensuring a consistent spatial comparison between datasets. Considering the fact that this comparison is model-to-model, using standardized anomalies (equation 3.1) is a better way since models have their own physical mechanisms that influence or define the mean *SM*, which might be slightly different from observations (Koster et al., 2009). This approach eliminates systematic biases and focuses on relative temporal patterns rather than absolute values. For reference, the mean *SM* values over the entire CRB between the two datasets: surface *SM* (SMAP: 0.12 ± 0.03 m^3/m^3, VIC: 0.132 ± 0.07 m^3/m^3) and rootzone *SM* (SMAP: 0.16 ± 0.018 m^3/m^3, VIC: 0.095 ± 0.019 m^3/m^3).

$$Z_i = \frac{SM_i - \overline{SM}}{\sigma_{SM}} \tag{3.1}$$

where Z_i represents standardized anomaly at time *i*, SM_i represents actual *SM* value at time *i*, $\overline{SM}$ represents mean *SM* over 2015-2024, and σ_{SM} represents standard deviation of SM_i.

The temporal patterns between VIC (10 cm depth) and SMAP (5 cm depth) surface *SM* demonstrate substantial agreement, with R^2 values of 0.71, 0.69, and 0.74 for the entire CRB basin, UCRB, and LCRB, respectively. Both datasets capture similar seasonal variations, suggesting that the VIC model reasonably represents the temporal dynamics of surface *SM* observed by SMAP. This level of agreement is particularly noteworthy considering the inherent challenges in comparing model-derived soil moisture with satellite observations, which operate at different spatial and temporal resolutions. Comparison for rootzone *SM* (SMAP:100 cm depth; VIC: varies by sub-basin up to 150 cm depth) reveals an even stronger agreement, with a high R^2 (0.84) for the entire CRB and LCRB (0.92). The stronger agreement observed in rootzone *SM* may be attributed to several factors. First, deeper soil layers exhibit less temporal variability compared to surface layers, as they are less directly influenced by rapid atmospheric

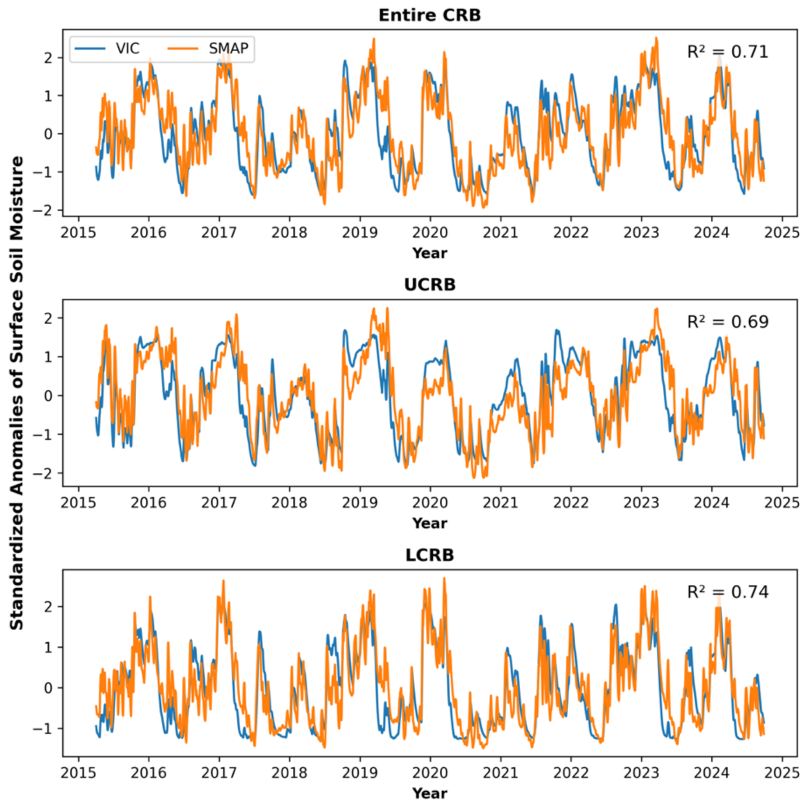


Figure 35. Comparison of standardized anomalies of VIC Layer 1 (surface) *SM* outputs with SMAP surface *SM*. Values are shown as weekly rolling means.

forcing and have slower response times. This reduced variability makes it easier for models to capture the dominant temporal patterns. Second, SMAP's rootzone *SM* is derived through data assimilation techniques using a land surface model, which may introduce some model consistency with VIC's physical representation. Additionally, rootzone *SM* integrates water storage over a larger soil volume, potentially averaging out small-scale discrepancies that are more pronounced in surface layers.

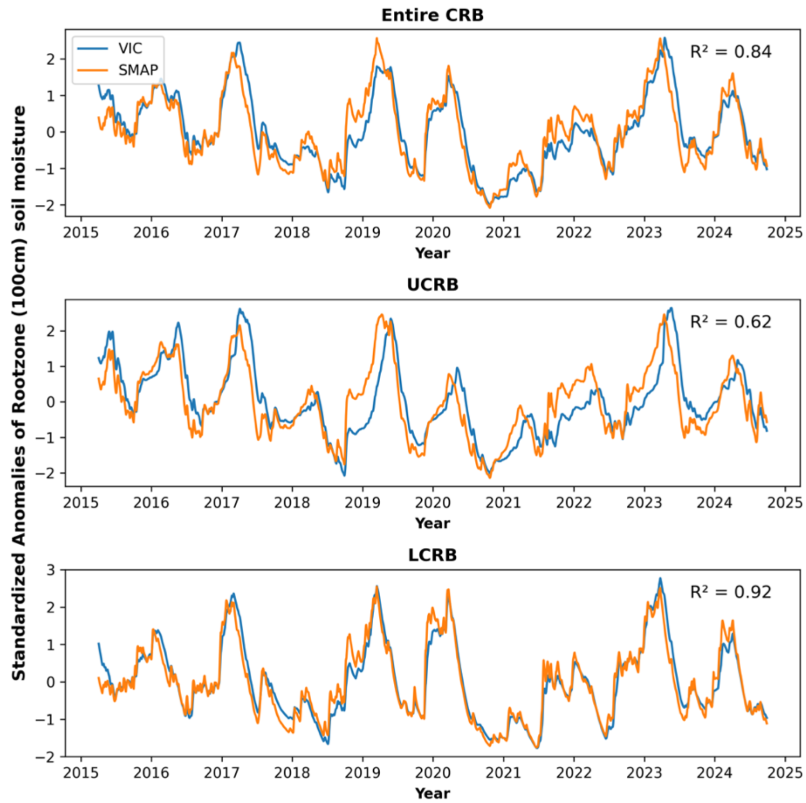


Figure 36. Comparison of standardized anomalies of VIC Layer 1+2 (rootzone) *SM* outputs with SMAP rootzone SM. Values are shown as weekly rolling means.

In the UCRB, the rootzone *SM* comparison reveals an interesting phenomenon. While the magnitudes of standardized anomalies between VIC and SMAP are similar, there appears to be a small temporal lag effect in several periods, particularly visible in Figure 36. This temporal offset may be related to differences in how the two modeling approaches represent water movement through soil layers in complex terrain. Both VIC and the SMAP Level 4 algorithm (which uses the CLSM) employ different soil parameterizations and hydraulic conductivity formulations that can result in varying rates of vertical moisture transport. Unlike the LCRB, where both models show strong agreement ($R^2 = 0.92$), the UCRB's complex topography and substantial snowmelt

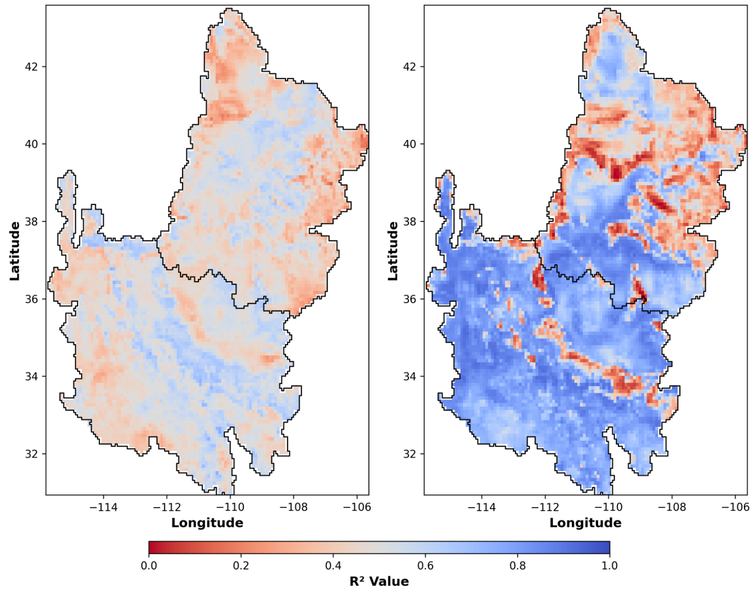


Figure 37. Spatial R^2 between corresponding grids from VIC and SMAP for (a) Surface *SM* (left) (b) Rootzone *SM* (right).

contribution likely amplify these differences. The stronger agreement in the LCRB can be attributed to its more homogeneous terrain, less snowmelt-dominated hydrology, and arid conditions where soil moisture is more directly coupled to precipitation events. This highlights how regional landscape characteristics influence agreement between different modeling approaches.

The analysis was further extended to compare spatial correlations between VIC and SMAP standardized anomalies at the 9 km SMAP pixel resolution for the period of 2015-2024. For surface *SM* (Figure 37), the correlation pattern displays moderate agreement across most of the basin, with some regions showing higher R^2 values (blue areas). The spatial pattern is heterogeneous, with correlation coefficients generally ranging from 0.4 to 0.8. The rootzone *SM* comparison demonstrates stronger spatial agreement across much of the basin, particularly in the LCRB, where large areas show R^2

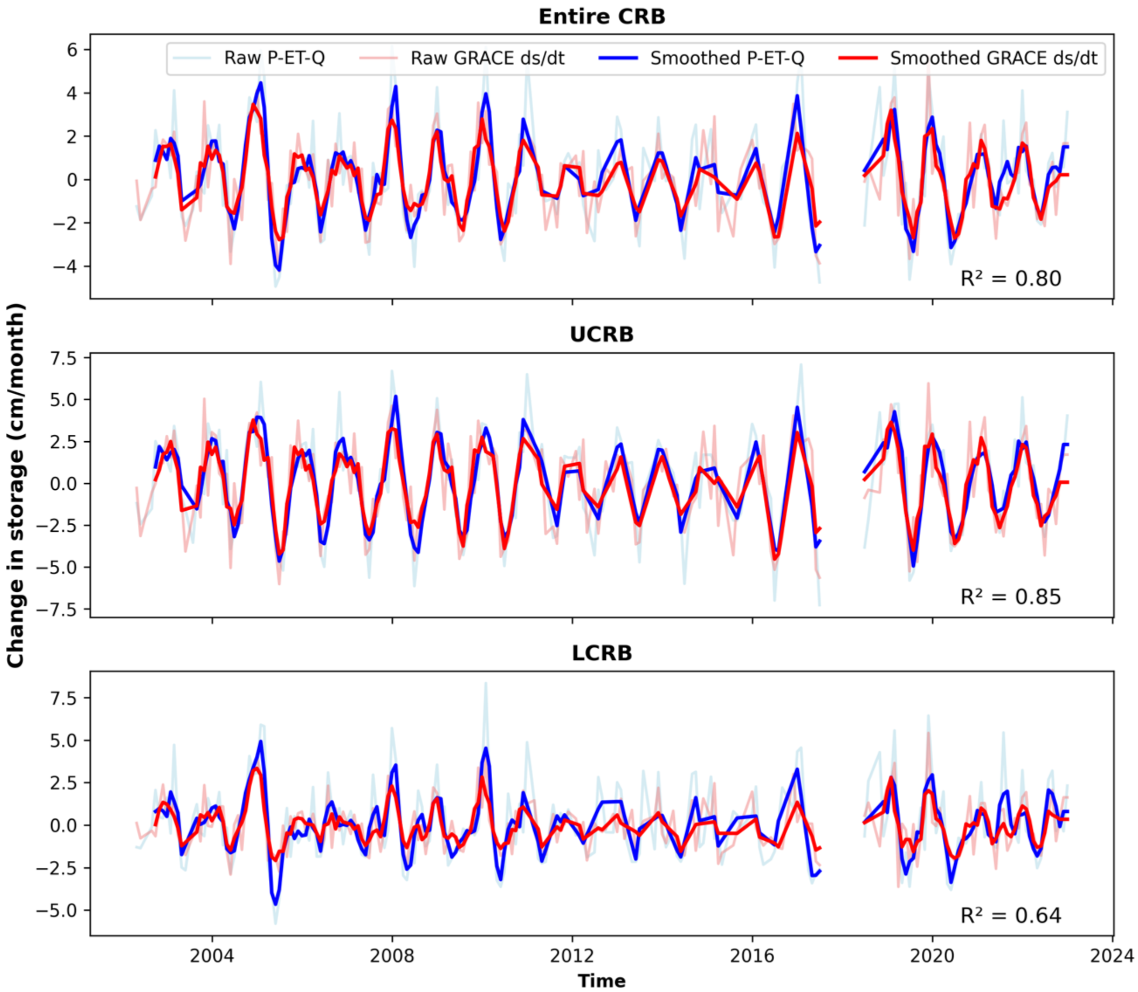


Figure 38. Comparison of GRACE change in storage with water balance from VIC. Data is smoothed to quarterly means to reduce noise.

values exceeding 0.8. Notably, the regions showing lower correlation values correspond strongly with high-elevation, snow-dominated areas of the basin. The results demonstrate that while the VIC model captures temporal patterns of *SM* variability across the CRB, spatial differences in correlation values exist between the two datasets, particularly for surface *SM* and in snow-dominated environments.

Figure 38 compares the GRACE satellite-derived change in storage (*ds/dt*) with the water balance residual (*P-ET-Q*) from the VIC model across the CRB from 2002-2023. The analysis reveals strong agreement between these independent datasets, with correlation coefficients of $R^2 = 0.80$ for the entire basin, $R^2 = 0.85$ for the UCRB, and a moderate $R^2 = 0.64$ for the LCRB. Both datasets effectively capture seasonal cycles, with storage changes typically ranging between ±2.5 cm/month, as well as significant inter-annual variations corresponding to wet and drought periods throughout the study period. The regional analysis reveals interesting differences in storage dynamics between the UCRB and LCRB. The UCRB generally shows larger amplitude in seasonal cycles, reflecting the dominant role of snowmelt processes. Meanwhile, the LCRB exhibits somewhat muted seasonality, particularly during certain periods (e.g., 2007-2012).

The agreement between these independent datasets provides additional validation of the VIC model's ability to capture water balance components across the CRB. While water storage changes weren't directly calibrated as an explicit target, the soil parameter adjustments made during streamflow calibration (particularly soil depth modifications) indirectly influenced the model's storage capacity. This comparison suggests that the calibrated parameterization adequately represents the seasonal and inter-annual water balance dynamics.

3.3 Numerical Experiments

Following model calibration and validation, a series of numerical experiments were conducted to address the research questions by isolating and quantifying the impacts of specific factors on snowmelt-driven streamflow dynamics in the CRB.

3.3.1 Spatial Representation of SNOTEL Stations

This experiment evaluated how well the current network of SNOTEL stations (104 used by CBRFC) represents basin-wide snow conditions in the UCRB through a comparison to spatially distributed model outputs across HUC10 watersheds in the UCRB (Figure 8c). HUC10 evaluations and comparisons with VIC pixels for suitability of use in the CRB are presented in Appendix E. Figure 39 shows a scatter plot comparison of SNOTEL median values with corresponding VIC and SNODAS median values. Each point represents a yearly relative *SWE* value with respect to the baseline (1991-2020 for SNOTEL vs. VIC, as used by water managers; 2004-2020 for SNOTEL vs. SNODAS, as SNODAS starts in 2004). The SNOTEL-VIC comparison shows that most values align closely along the 1:1 line with a strong R^2 value of 0.93, suggesting that the VIC reasonably represents the overall snow conditions at the SNOTEL stations contained in the HUC10 watersheds. This highlights the ability of the HUC10 watersheds to capture deviations from the historical median *SWE*, even if the absolute magnitudes may differ due to the inherent differences between point measurements and watershed-averaged values. However, the comparison between SNOTEL and SNODAS reveals a different trend. In recent years, SNODAS consistently overestimates *SWE* values across the SNOTEL stations when aggregated to the HUC10 watershed scale. This

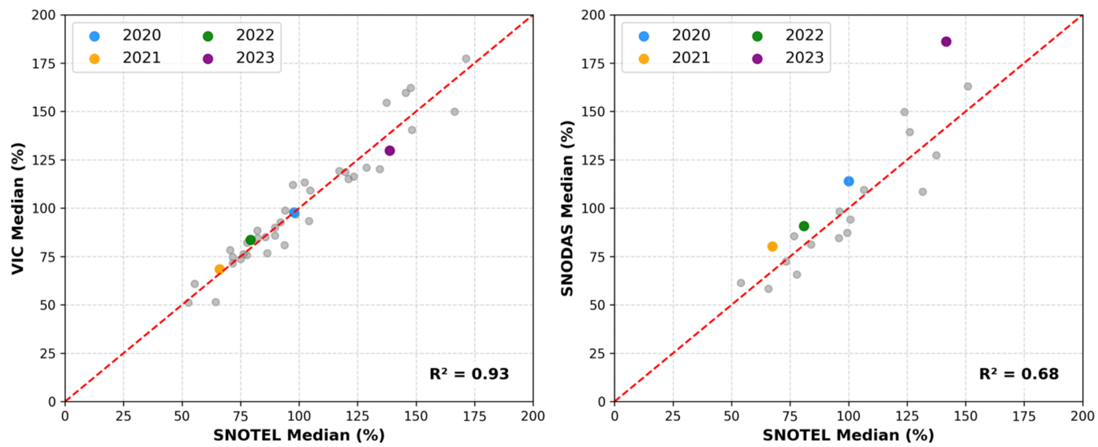


Figure 39. Comparison of relative median *SWE* values between (a) SNOTEL vs. VIC (Baseline: 1991-2020) (b) SNOTEL vs. SNODAS (Baseline: 2004-2020).

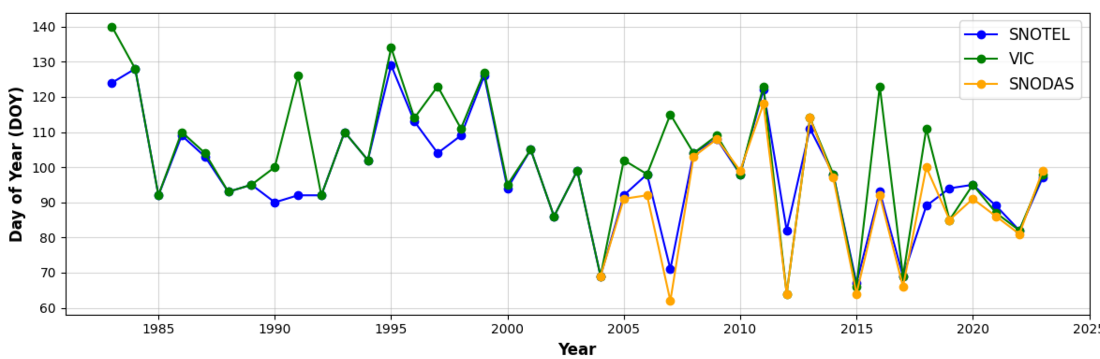


Figure 40. DOY of peak *SWE* across SNOTEL, VIC, and SNODAS datasets.

overestimation suggests that while SNODAS provides spatially comprehensive snow data, it may introduce biases, especially in representing *SWE* magnitudes over larger areas (the area of SNODAS pixel is 1 km^2 while the mean area of HUC10 boundaries in this analysis was 600 km^2). Considering these findings, VIC *SWE* outputs offer a better alternative for capturing *SWE* dynamics across watersheds than SNODAS.

The timing of peak *SWE* (Figure 40) (median of all SNOTEL stations and corresponding HUC10 watersheds) shows general correspondence across the datasets in

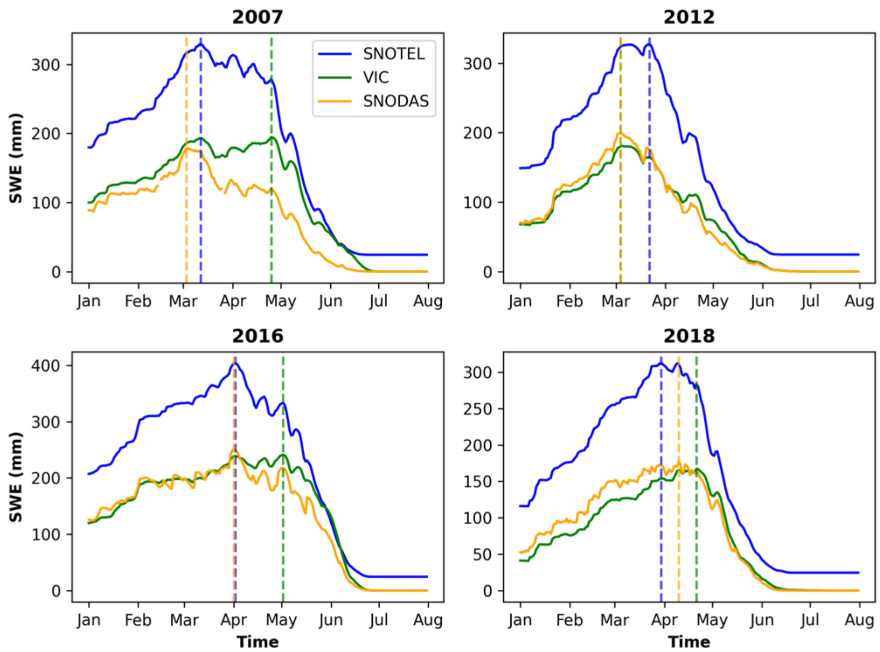


Figure 41. Example of years when timing of SNOTEL, SNODAS and VIC peak *SWE* did not align with each other.

most years, especially in recent periods. However, certain years (2007, 2012, 2016, and 2018) show notable discrepancies. This inconsistency is explored in Figure 41.

In 2007 and 2016, VIC exhibited unique double-peak patterns with limited snowmelt after the initial peak (which matched with SNOTEL), and a secondary peak of similar magnitude was formed in late spring. In 2012, all datasets aligned on the first peak, but a secondary peak in SNOTEL created temporal differences. In 2018, SNOTEL peaked earlier than both VIC and SNODAS. This timing difference may be partly attributed to spatial aggregation effects, as point measurements can respond faster to local conditions compared to watershed-averaged values. Overall, VIC *SWE* outputs, when aggregated using selected HUC10 boundaries, effectively capture the magnitude of relative *SWE* from point measurements and demonstrate strong agreement with the

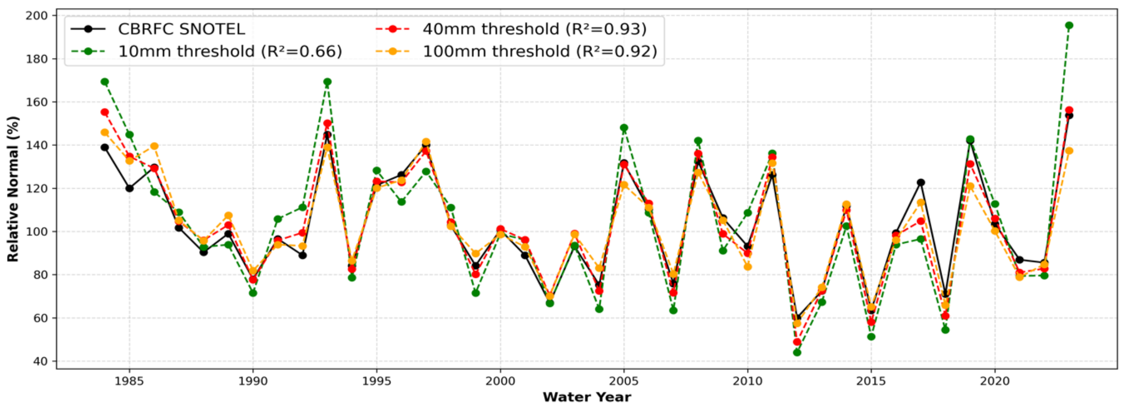


Figure 42. Time series of SNOTEL and selected HUCS from VIC based on thresholds.

timing of peak *SWE* in most years, making them reliable proxies for SNOTEL stations in representing *SWE* dynamics across the basin.

The next part of the analysis aimed to determine whether these selected HUC10 watersheds accurately represent the overall snow conditions of the entire basin. A threshold-based filtering approach was developed to evaluate how well different subsets of the basin capture basin-wide snow conditions. First, the long-term mean daily *SWE* (1991-2020 mean) for all HUC10 watersheds in the UCRB was calculated from VIC *SWE* outputs. Then, these watersheds were filtered based on specific mean *SWE* thresholds (10 mm, 25 mm, and 40 mm) to identify areas with significant snow accumulation. As shown in Figure 42, relative values from SNOTEL networks match well with thresholds greater than or equal to 40 mm ($R^2 > 0.9$). This correlation was calculated by comparing yearly median *SWE* values from the SNOTEL stations against the median *SWE* from HUC10 watersheds filtered at different thresholds. In contrast, a lower R^2 (0.66) is observed for the lower threshold value of 10 mm.

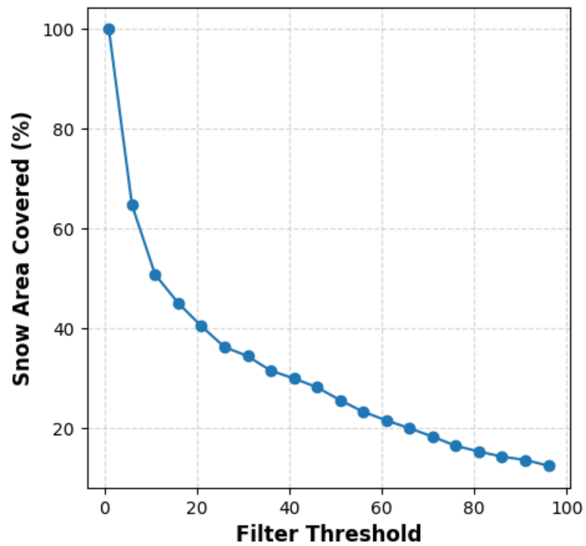


Figure 43. Snow area covered calculated using different thresholds of mean daily *SWE* (1991-2020) in the HUC10 watersheds located in the UCRB.

This approach revealed important insights about the representativeness of the snow areas. Figure 43 illustrates that the percentage of snow-covered areas represented decreases as the filter threshold increases. At the 40 mm threshold, which was found to provide comparable *SWE* values to the 81 SNOTEL-containing HUC10s based on correlation analysis, only about 30% of the total snow-producing area in the UCRB is represented. This indicates that high-snow regions are predominantly captured by the current operational approach using SNOTEL stations, while significant lower-snow areas that have significant snow cover may potentially be overlooked.

The spatial distribution of watersheds at different thresholds (Figure 44) further confirms this finding, showing that the 81 HUC10 watersheds (comparable to a 40 mm threshold) miss substantial snow-covered regions, particularly in the Upper Colorado sub-basin. Lower thresholds like 10 mm or 25 mm provide much broader coverage of snow conditions across the basin. This highlights biases in the current methodology that the SNOTEL-based estimates tend to overestimate *SWE* in dry years and underestimate it in wet years, which is also evident in the time series plot in Figure 42.

The April 1 *SWE* analysis, a critical date for water management decision-making in the CRB, revealed further biases in the current operational approach. To evaluate these biases, VIC *SWE* outputs filtered at different thresholds (10 mm and 40 mm) were compared with the median *SWE* values from SNOTEL. While SNOTEL-based estimates and VIC with a 40 mm threshold aligned closely (Figure 45), significant discrepancies were identified when comparing SNOTEL values against VIC outputs with a 10 mm threshold. During drier years (bottom left), SNOTEL-based estimates consistently overestimated April 1 *SWE* compared to the more spatially extensive 10 mm threshold. Conversely, during wetter years (top right), SNOTEL-based estimates consistently underestimated April 1 *SWE*. This pattern suggests that while the 40 mm threshold provides better statistical agreement with SNOTEL measurements, the 10 mm threshold provides a more representative characterization of overall basin-wide snow conditions because it includes a larger portion of the snow-producing areas that SNOTEL stations don't adequately represent.

The bias in the current methodology could lead to overconfidence in water availability during drought conditions and underestimation during wet periods. These findings suggest that the 10 mm threshold provides a more representative characterization of basin-wide snow conditions than the current operational SNOTEL network. For improved assessment of snow conditions, VIC *SWE* outputs filtered at lower thresholds could supplement existing SNOTEL networks, offering a more spatially comprehensive representation of snow conditions across the basin and potentially reducing systematic biases in wet and dry years.

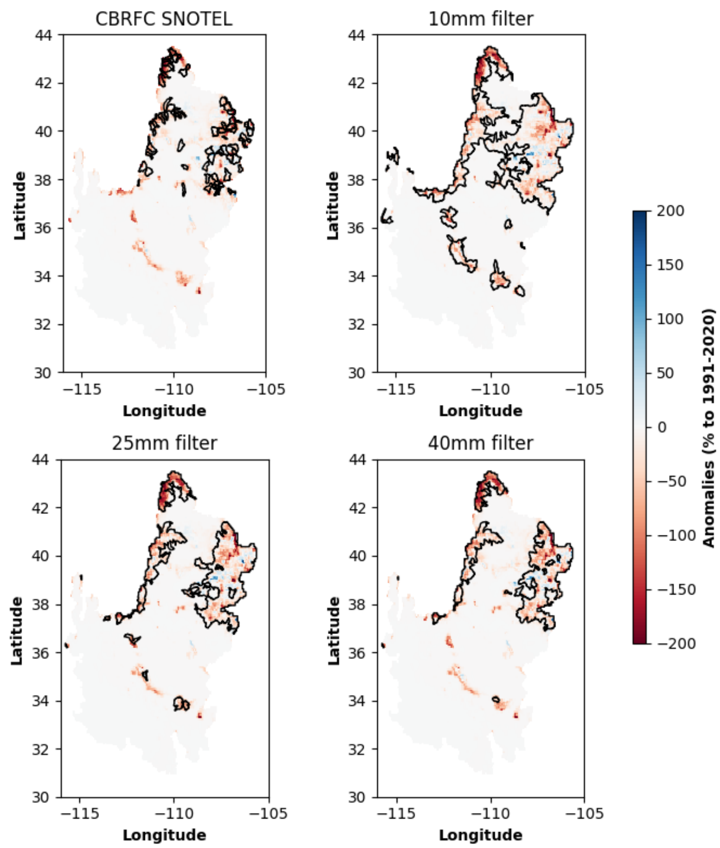


Figure 44. Comparison of the HUC10 watersheds (polygons) selected using different thresholds, overlaid on *SWE* conditions in the basin (Example of *SWE* data for April 1, 2022 anomalies relative to 1991-2020 April 1 mean).

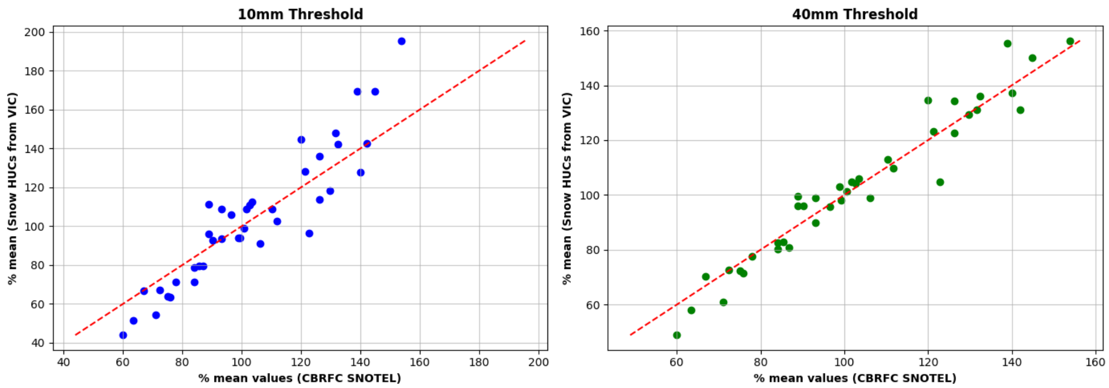


Figure 45. Comparison of using different thresholds for April 1 Relative *SWE* comparison.

3.3.2 Effects of Spring (AMJ) Weather

This experiment isolated the impacts of AMJ conditions on the UCRB streamflow generation while maintaining consistent initial *SM* along with winter and near-peak snowpack conditions of WY 2020. The simulation results reveal substantial variability in streamflow outcomes driven solely by AMJ weather conditions. Figure 46 illustrates that across the set of 40 simulations, the resulting annual WY streamflow volumes ranged from 11.33 to 19.64 km^3, which is a difference of 8.31 km^3. While subsetting only the summer (June-September or JJAS) months, values ranging from 3.74 to 13.37 km^3 were observed across the ensemble simulations, indicating that large streamflow sensitivity due to AMJ weather occurs during the summer months. This seasonal difference in sensitivity can be attributed to the temporal proximity between AMJ conditions and summer streamflow. Spring conditions directly influences the primary snowmelt season, with immediate effects on summer flows. This temporal relationship is particularly evident in years with extreme spring weather anomalies, where summer streamflow response showed amplified deviations from mean conditions compared to WY totals. It is also seen that despite having average snowpack conditions, the AMJ weather of WY 2020 itself represented an extreme case within the historical record, with total *P* (56.90 mm) being lower than most years and mean *T* (12.09 °C) value being higher than the mean within the historical AMJ record. Further, Figure 47 demonstrates the strong relationship between AMJ climate variables and streamflow in the UCRB, showing spring precipitation strongly drives (R^2 = 0.91) the WY streamflow. Temperature, although it has a relatively weak (R^2 = 0.58) relationship, still impacts the streamflow produced in the UCRB.

Figure 48 compares two analyses showing how AMJ weather affects streamflow. Actual conditions are the VIC modeled outputs where all factors vary naturally (snowpack, soil moisture, and climate), while constant snowpack refers to the controlled experiment where snowpack and initial *SM* were held constant at 2020, and AMJ weather was varied. *T* anomalies (difference from the 1991-2020 average in °C) are represented by the horizontal axis in both plots, while precipitation anomalies (difference from the 1991-2020 average) are shown on the vertical axis. The resulting streamflow (as a percent of the average) is indicated by the colors, with blue representing higher streamflow and red representing lower streamflow.

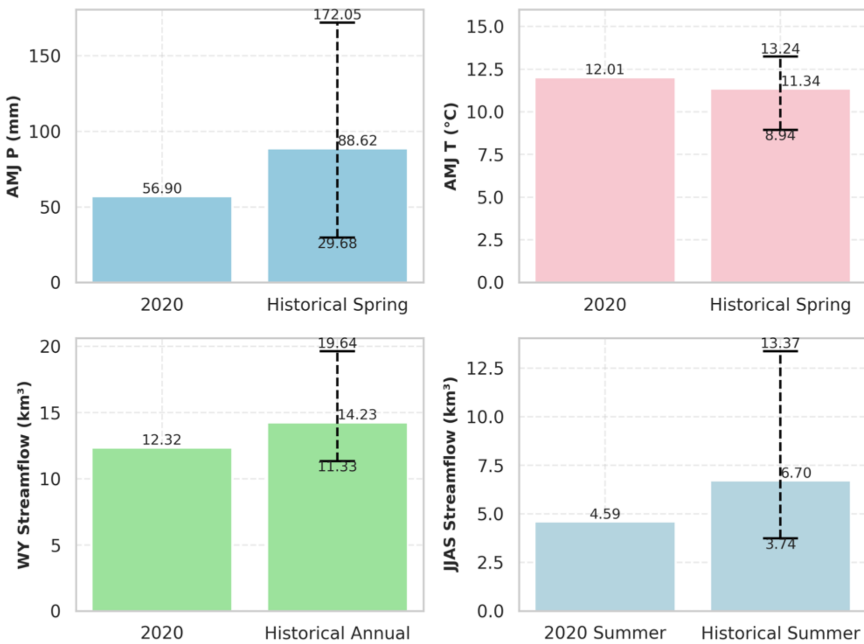


Figure 46. Range of variability of Spring (AMJ) *P*, *T* and UCRB WY and Summer (JJAS) streamflow in the historical record compared to actual conditions in WY 2020. Vertical lines in the right bars show maximum, mean and minimum values on record (1984-2023).

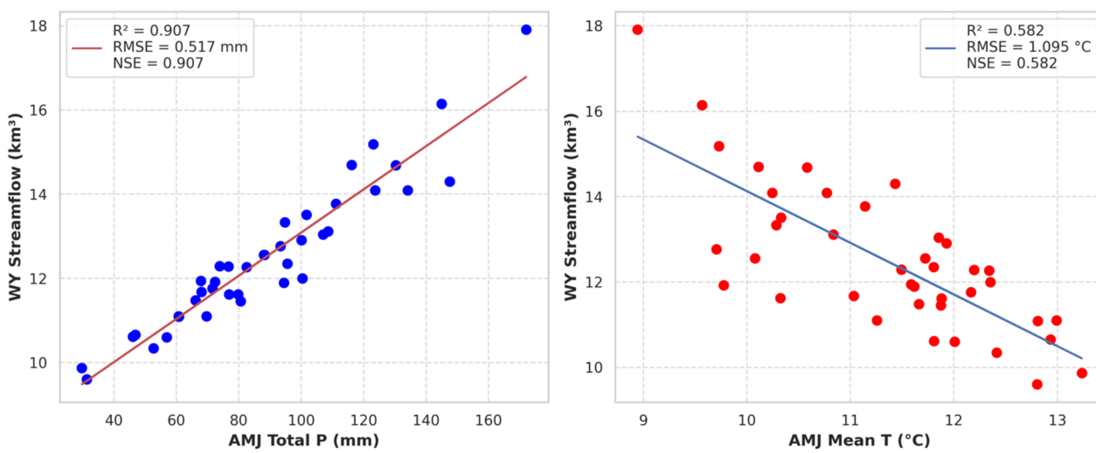


Figure 47. Scatter plots showing relationships between AMJ P with WY streamflow (left) along with AMJ T with WY streamflow (right) for the UCRB

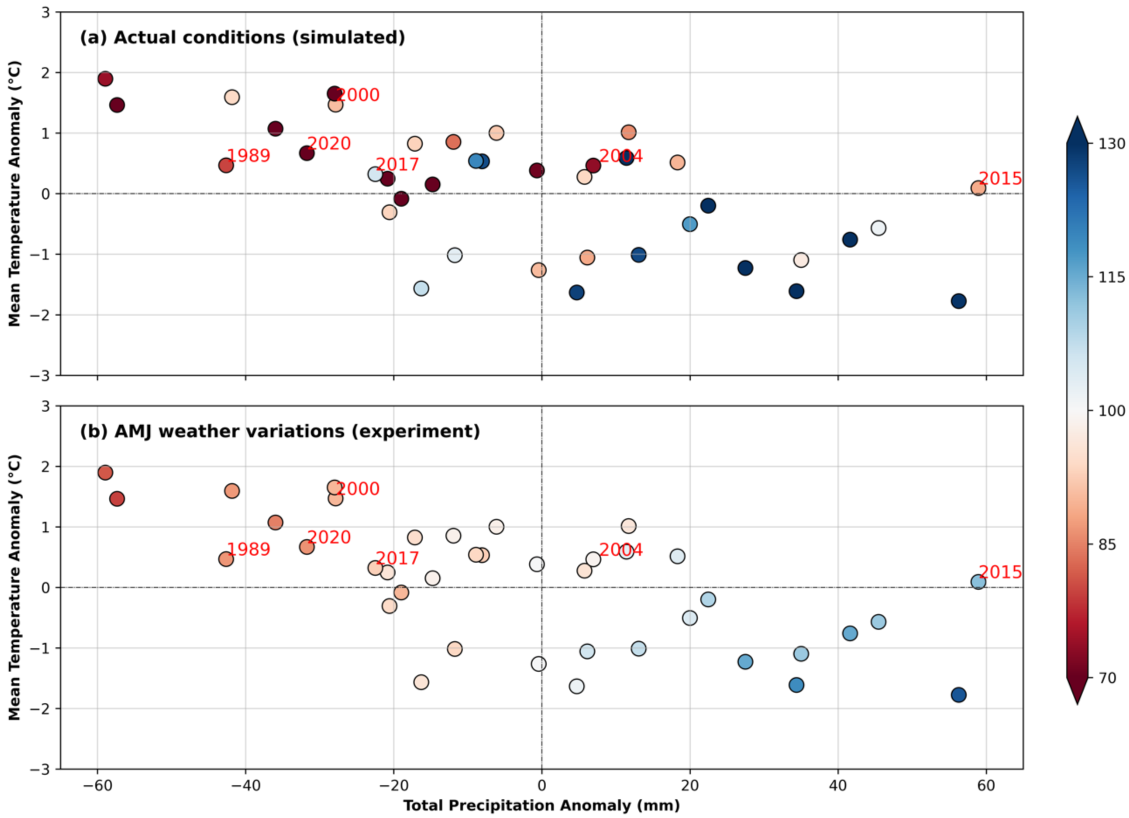


Figure 48. Relative WY streamflow compared to AMJ climate anomalies (Baseline: 1984-2023) for (a) Actual conditions (b) Constant snowpack experiments.

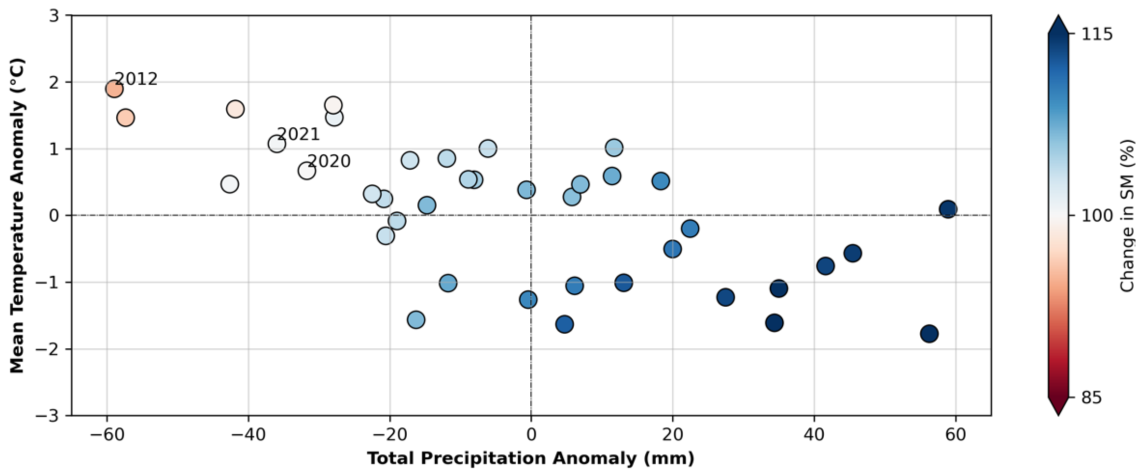


Figure 49. Change in SM (% of value after WY 2020 with respect to value before WY 2020) caused by variations in AMJ P and T.

From the scatter plot of actual conditions (Figure 48a), an underlying pattern can be observed, but it is obscured by considerable noise because multiple factors are changing simultaneously. However, in Figure 48b, a much clearer pattern is revealed: years with cool-wet spring conditions (lower right quadrant) consistently produced higher streamflow (blue colors, often exceeding 125% of average), while warm-dry springs (upper left quadrant) led to significantly reduced streamflow (red colors, often below 75% of average). This pattern highlights how dramatically spring weather can alter the amount of streamflow produced, even when starting with identical snowpack conditions. For example, a year with an average snowpack but a warm, dry spring (like 2020) can produce streamflow similar to a below-average snowpack year. Conversely, cool, wet spring conditions can boost streamflow well above what would be expected from snowpack alone.

The weather conditions during the AMJ period not only influence streamflow directly but also affect soil moisture storage for the following WY. Figure 49 illustrates this relationship by showing how AMJ weather anomalies impact October 1 *SM* conditions (of the following year) relative to that WY 2020. Most years (represented by blue circles) show *SM* improvements compared to 2020 conditions, demonstrating how favorable spring climate can positively affect basin hydrology. However, several years, including WY 2012, 2020, and 2021, show continued *SM* deficits, and are characterized by warm-dry spring conditions (upper-left quadrant), reinforcing the notion that unfavorable AMJ climate can propagate drought conditions through reduced *SM* recharge. This *SM* carryover effect creates an important feedback mechanism where spring weather anomalies influence not only current year streamflow but also set initial conditions for the following year's hydrological response.

The findings have important implications for water management in the CRB. Currently, many water supply forecasts in the region rely heavily on April 1st snowpack measurements to predict seasonal runoff (Goble & Schumacher, 2023). These results suggest that incorporating spring climate projections into these forecasts could substantially improve their accuracy, potentially helping water managers better anticipate actual water availability and make more informed decisions in their operations.

3.3.3 Effects of Initial Soil Moisture

This analysis isolating initial *SM* effects reveals that antecedent conditions alone can drive substantial variability in annual UCRB streamflow outcomes. WY 2020 was selected as the baseline for this experiment as well to maintain consistency with the AMJ weather analysis and because it provided average snowpack conditions while exhibiting below-average initial soil moisture (90% of 1991-2020 mean). As shown in Figure 50, while maintaining identical meteorological conditions across all simulations, annual streamflow volumes ranged from 9.44 to 17.89 km^3, representing a 1.9 times or 8.45 km^3 difference between minimum and maximum values due solely to initial *SM* variations. JJAS streamflow exhibited less sensitivity to initial SM conditions, with values ranging from 3.55 to 6.43 km^3 across the simulations, with an actual 2020 value of 4.59 km^3 compared to the historical mean of 4.69 km^3. WY 2020 exhibited below-average initial *SM* (213.25 mm) compared to the historical mean (235.19 mm), corresponding with below-average streamflow (12.32 km^3 vs. historical mean of 12.81

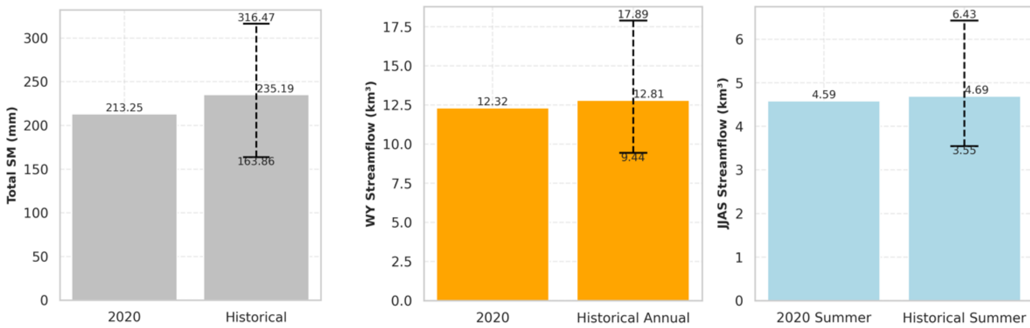


Figure 50. Range of variability of initial *SM,* WY and summer streamflow in the ensemble members compared to actual conditions in WY 2020.

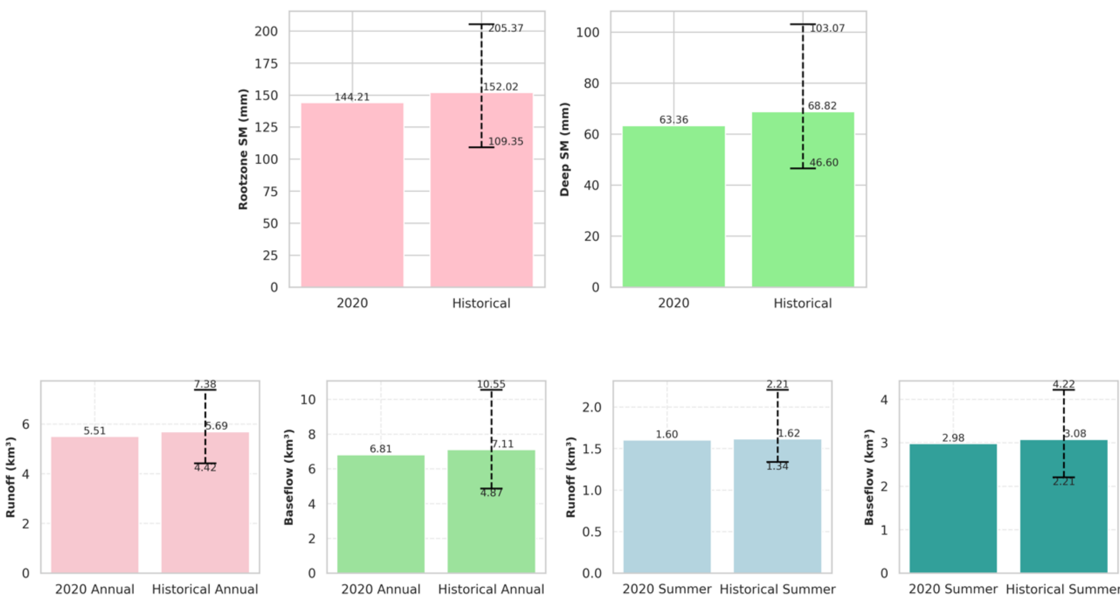


Figure 51. Range of variability of rootzone and deep *SM*, runoff, and baseflow in the ensemble members compared to actual conditions in WY and Summer 2020.

km^3). This pattern indicates that the dry antecedent conditions in 2020 contributed to reduced streamflow outcomes.

The component analysis (Figure 51) reveals distinct mechanisms through which *SM* affects streamflow generation. For clarity, *SM* in the VIC model is represented in multiple layers: the surface *SM* (first soil layer, approximately 10 cm depth), the rootzone

SM (second soil layer, approximately 100-150 cm depth), and the deep *SM* (in the lowest soil layer, 3-4.5 m depth). These different layers are parametrized to produce different components of streamflow, as surface and rootzone *SM* produce runoff (surface flow), and deep *SM* is responsible for baseflow (subsurface flow). It is observed that rootzone *SM*, despite showing noticeable variation between 2020 (144.21 mm) and the historical mean (152.02 mm), produced only small differences in total surface runoff (2020: 5.51 km^3 vs. historical mean: 5.69 km^3). In contrast, deep *SM* variations had a proportionally larger impact on streamflow outcomes. The relatively modest difference in deep *SM* between 2020 (63.36 mm) and the ensemble mean (68.82 mm) resulted in a larger difference in baseflow contribution (2020: 6.81 km^3 vs. mean: 7.11 km^3). Analysis of the variation across all simulations indicates that approximately 67% of the total streamflow variation was attributable to changes in baseflow rather than surface runoff. Summer runoff ranged from 1.34 to 2.21 km^3 (actual 2020: 1.60 km^3, historical mean: 1.62 km^3), while summer baseflow varied from 2.21 to 4.22 km^3 (actual 2020: 2.98 km^3, historical mean: 3.08 km^3). This partitioning reveals that baseflow variations accounted for a larger portion of summer streamflow sensitivity to initial soil moisture conditions. The greater sensitivity of streamflow to deep *SM* compared to rootzone *SM* can be explained by the fundamental hydrologic processes represented in the VIC model, as deep *SM* contributes to baseflow generation through the ARNO baseflow formulation (Equations 2.3 and 2.4), creating a sustained, slow release of water to streams throughout the year.

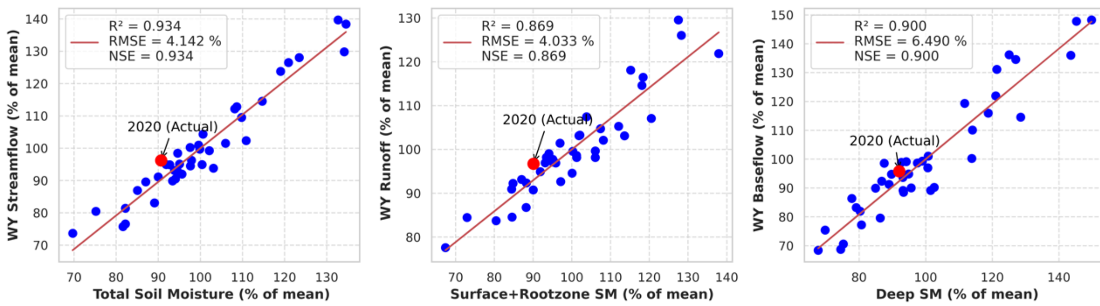


Figure 52. Scatter plot showing relationship between initial SM conditions (October 1) (total, surface+rootzone, deep) with WY streamflow, runoff and baseflow respectively.

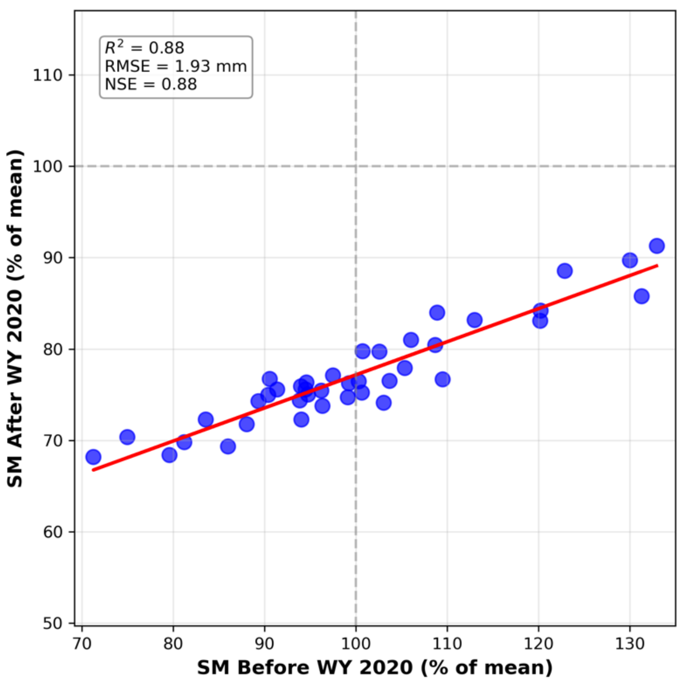


Figure 53. Scatter plot of the relationship between SM before the start of WY and its effect on the SM at the end of the WY.

These findings, when compared with the AMJ climate experiment results, reveal an important contrast in temporal sensitivity patterns. For annual streamflow, both AMJ weather and antecedent soil conditions show similar magnitudes of influence, as soil moisture variations produced an 8.45 km³ range (9.44 to 17.89 km³), while AMJ climate variations resulted in an 8.31 km^3 range (11.33 to 19.64 km^3). However, for JJAS streamflow, a distinct difference emerges. AMJ climate variations generated a summer streamflow range of 3.74 to 13.37 km^3, which is way greater than the range (3.55 to 6.43 km^3) produced by *SM* variations. This heightened summer sensitivity to spring climate can be attributed to temporal proximity - spring precipitation directly precedes and influences the summer runoff season, creating a more immediate hydrological response. In contrast, soil moisture influence remains relatively consistent across seasons, primarily sustaining baseflow throughout the year

These findings underscore the significance of *SM* conditions, particularly in deeper layers, for the CRB's hydrological response. The relationship observed in the simulations reveals how antecedent *SM* states can substantially influence annual streamflow outcomes even when meteorological conditions remain identical (Figure 52). Figure 52 illustrates strong positive correlations between initial *SM* conditions and hydrological responses, with especially high R^2 values for deep *SM* and baseflow (0.9), highlighting the often-overlooked role of subsurface water storage as a key factor in determining WY streamflow volumes. Further, the persistence of initial *SM* conditions throughout the water year under 2020's meteorological conditions is demonstrated in Figure 53. This scatter plot shows the relationship between initial *SM* and the resulting soil moisture at the end of the following WY (both expressed as a % of 1984-2023 mean). Even scenarios starting with above-average initial *SM* experienced significant drying, ending the WY at 80-90% of the long-term mean, while drier starting conditions showed even greater depletion, illustrating how WY 2020 created a systematic drying effect. This pattern suggests a potential feedback mechanism where available *SM* supports increased *ET*, further contributing to soil water depletion throughout the *WY,* without any substantial input. The results suggest that *SM* variations alone can account for significant differences in annual streamflow, emphasizing the importance of considering these conditions when assessing the basin's hydrological state.

3.3.4 Combined Effects of Post-April Climate and Initial Soil Moisture

The 200 simulations across varying combinations of initial *SM*, *P*, and *T* from 1984-2023 reveal distinct patterns in streamflow generation within the UCRB. However, the 200 simulations do not cover every possible combination of conditions. A multiple regression approach was developed to fill these gaps and provide a continuous representation across the full parameter space (equation 3.2). The values were obtained after standardizing all variables (equation 3.1, using 1984-2023 mean and standard deviations), and then, relative change in UCRB WY streamflow with respect to 1984-2023 mean was calculated (for example, if a simulation output was 70% of the mean value, relative change = -30%). To enable direct comparison of their effects, the following relationship was identified for the UCRB:

$$\Delta Q = -23.44 + 12.90\, SM + 7.31\, P - 1.80\, T \quad , \qquad (3.2)$$

where *SM, P,* and *T* represent the standardized anomalies of October 1 *SM*, AMJ *P*, and AMJ *T*, respectively and ΔQ is the relative change in UCRB streamflow. This equation has high explanatory power ($R^2 = 0.89$) and was used to interpolate streamflow responses for conditions not directly simulated. Figure 43 displays these results as a heatmap

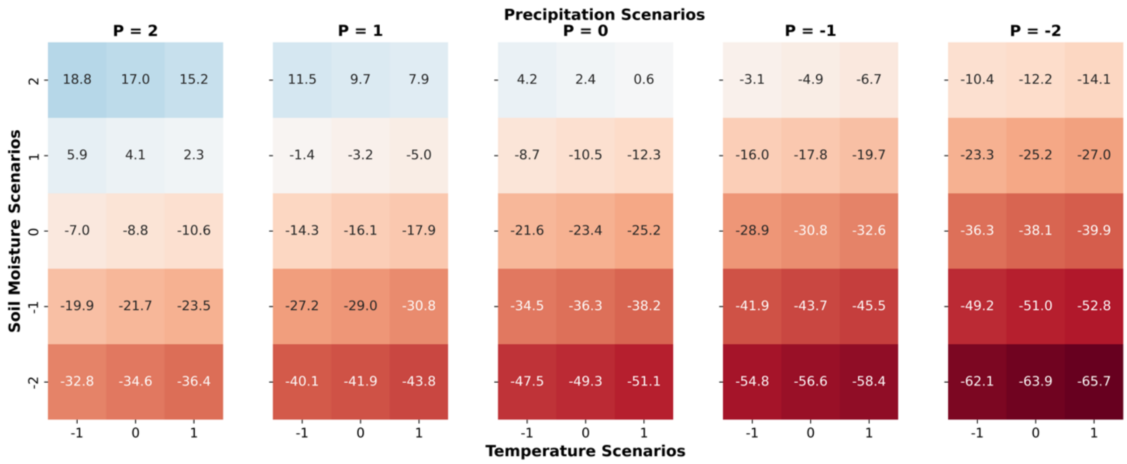


Figure 54. Change in streamflow (%) across a range of *SM*, *P*, and *T* conditions from the 200 simulations. Other values were filled using equation 3.2.

showing streamflow responses, with fall and winter meteorological conditions held constant at the 2020 value. The heatmap reveals clear gradients in streamflow response, with higher *SM* and *P* combinations yielding increased streamflow (blue regions) while drier conditions lead to reduced streamflow (red regions).

A key insight from Figure 54 is where the color starts changing from red to blue. For example, even under a near-normal snowpack in 2020, it would have taken either (i) wet Oct 1 *SM* and wet AMJ *P* or (ii) normal Oct 1 *SM* with very wet AMJ *P* to have near-average streamflow in the UCRB. This relationship (equation 3.2) reveals that antecedent *SM* exerts the strongest influence on streamflow response, with a standardized coefficient approximately twice as larger than that of *P* and around 7 times the magnitude of *T*. The negative intercept term (-23.4%) suggests a systematic baseline reduction in streamflow relative to the historical mean, meaning that even average soil and climatic conditions in 2020 would not have been enough to yield a near-to-average streamflow. It is important to emphasize that this relationship is specific to WY 2020 as the baseline and should not be applied generically to other water years without appropriate adjustments.

The Condition Number (CN) (Strang, 1976) was calculated as a diagnostic measure to address potential multicollinearity concerns in the multiple linear regression analysis. This metric evaluates how well the regression model can distinguish between the effects of different predictor variables. The CN is calculated by taking the ratio of the largest to smallest singular values obtained from the singular value decomposition of the predictor matrix (Strang, 1976). High CN indicates potential instability in the coefficient estimation process, where minor perturbations in input data could result in disproportionate fluctuations in regression coefficients. While standardized variables already reduce collinearity issues, formal verification was performed to ensure model stability. The calculated CN of 2.33 is substantially below the threshold of 30, confirming a stable regression model. The relative contribution of each variable was calculated by dividing the square of its standardized coefficient by the sum of squared standardized coefficients for all variables. Using this method, soil moisture accounted for approximately 74.6% of the explained variance in streamflow, followed by precipitation (23.9%) and temperature (1.46%).

This analysis was subsequently extended to each sub-basin to provide spatial context and identify potential regional differences in variable importance, shown in Figure 55. For each sub-basin, the 200 simulations were analyzed using sub-basin-specific streamflow as the dependent variable and sub-basin-averaged Oct 1 *SM*, AMJ *P*, and AMJ *T* as predictor variables. A notable spatial pattern emerged, with *SM* states

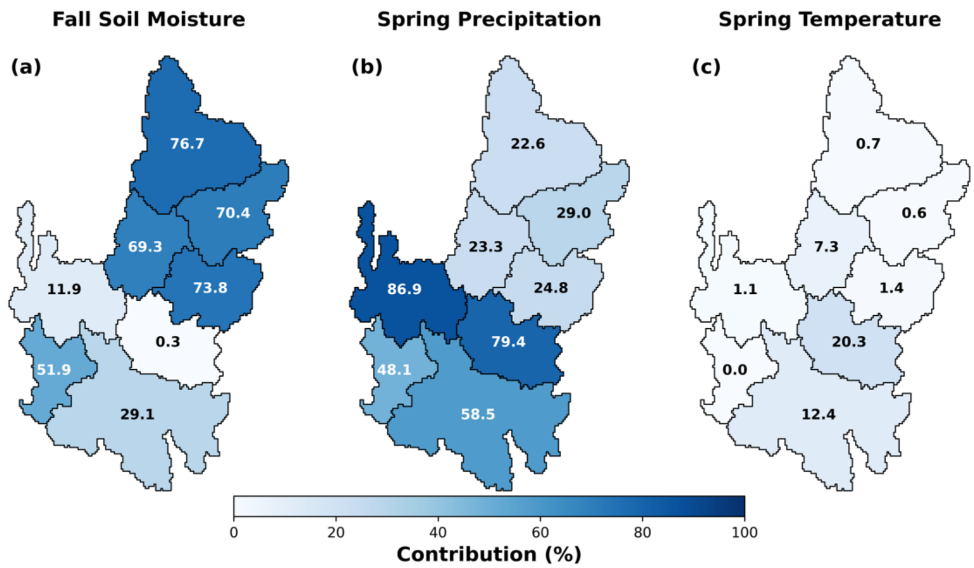


Figure 55. Relative contribution of each of the variables (*SM, P, T*) to change in streamflow across major sub-basins of the CRB.

exerting dominant influence in the UCRB sub-basins, while climate factors play a more significant role in the LCRB sub-basins. Particularly interesting is the enhanced importance of temperature in the Little Colorado and Gila Basins compared to other regions, possibly due to their lower elevations and warmer mid-latitude mountain climate which increases their sensitivity to temperature-driven *ET*. The results of the experiment are summarized in Table 8.

These findings suggest that water management in the CRB should account for the dominant role of *SM* in the UCRB. This oversight may explain why water managers have consistently overestimated expected runoff despite average snowpack conditions in recent years. Current operational systems should consider antecedent *SM* monitoring, particularly for the UCRB, where it explains approximately two-thirds of streamflow variability when compared with AMJ *P* and *T* under similar standardized anomalies.

Table 8. Summary of regression coefficients, $R^{2,}$ and CN from the regression for all sub-basins in the CRB.

Sub-Basin	Const	Coeff_P	Coeff_T	Coeff_SM	R^2	CN
CRB	-24.11	5.33	-3.10	9.59	0.84	1.95
UCRB	-23.44	7.31	-1.80	12.90	0.89	2.33
LCRB	-28.48	4.72	-1.44	2.52	0.35	1.85
Green	-17.07	7.74	-1.36	14.26	0.86	2.48
Upper Colorado	-26.04	8.28	-1.18	12.90	0.92	2.57
San Juan	-34	8.37	-2.0	14.45	0.88	1.58
Glen Canyon	-20.82	8.85	-4.96	15.27	0.85	1.86
Grand Canyon	-33.75	7.07	-0.8	2.62	0.28	1.89
Lower Colorado	5.33	6.32	-0.02	6.56	0.86	1.39
Little Colorado	-32.30	4.01	-2.03	0.25	0.30	1.49
Gila	-30.78	3.97	-1.83	2.80	0.33	1.63

3.3.5 Drought Experiments

This experiment investigated the persistence of *SM* anomalies and their impacts on streamflow generation during extended drought periods through a series of controlled simulations with modified initial conditions and meteorological forcings, defined as the 'initial condition experiment'. Figure 56 illustrates the difference between these two scenarios: a baseline using actual WY 2020 initial conditions (approximately 90% of normal) and an alternative scenario where these conditions were replaced with wetter initial conditions from WY 1997 (approximately 140% of normal). Both scenarios then used the same meteorological forcings from 2020-2024.

The temporal evolution of these differences reveals contrasting responses in different soil layers. Rootzone *SM* (upper layers) initially showed benefits from the wetter starting conditions but quickly converged toward the baseline as the dry meteorological conditions of 2020-2021 became dominant. This rapid adjustment demonstrates the limited memory of upper soil layers, which respond more directly to

Figure 56. Difference in *SM* levels (left column) and streamflow components (right column) across 2020-2024 from the initial *SM* change experiment. The difference is between the simulation and baseline.

atmospheric forcing than initial conditions. In contrast, deep *SM* maintained a prolonged positive anomaly despite experiencing identical dry meteorological conditions, demonstrating the extended memory effect of deeper subsurface storage. This persistent moisture difference directly translated to sustained baseflow generation throughout the five-year period, showing how initial wet conditions can provide a multi-year buffer against subsequent dry meteorological conditions. While this buffering effect gradually diminishes over time, the multi-year persistence indicates that wet periods preceding drought can provide extended hydrological resilience through deep *SM* memory.

While the initial condition experiment examined how pre-existing *SM* states influence subsequent drought periods, the snowmelt impact experiment investigated how introducing wet periods of varying durations affects recovery from drought. To investigate this question, a series of controlled simulations were conducted where forcing data from historically wet years were substituted into different drought periods. Three scenarios were implemented: (1) a short-term scenario where a single dry year (2020) was replaced with a wet year (2023), (2) a mid-term scenario where two dry years (2012-2013) were replaced with two wet years (1996-1997), and (3) a long-term scenario where five dry years (1999-2003) were replaced with five wet years (1993-1997). In each case, the simulation continued with actual historical forcing data after the replacement period to track the persistence of any hydrological changes.

Figure 57 and Figure 58 reveal distinct recovery patterns across soil layers and streamflow components. Recovery is defined as the time when the simulation outputs return back to baseline outputs. For rootzone *SM,* recovery times were relatively short (10.9, 8.0, and 3.5 months for long-term, mid-term, and short-term interventions, respectively), indicating a short response to atmospheric conditions. Deep *SM* exhibited significantly longer recovery periods (28.0, 22.1, and 21.2 months), demonstrating longer hydrological memory. The streamflow components also showed differences in recovery times. Surface runoff recovered relatively quickly (3, 2, and 2 years), reflecting its closer connection to recent precipitation and upper soil moisture. Baseflow, however, exhibited the longest recovery times (5, 4, and 4 years), indicating that streamflow benefits from wet periods can persist long after meteorological conditions have returned to normal. This indicates that while surface conditions might quickly reflect current meteorological conditions, baseflow can remain influenced by conditions from several years prior.

These findings provide mechanistic insights into the recent streamflow anomalies observed in the CRB. The low streamflow response to near-average snowpack in WYs 2020 and 2021 can be attributed to the combination of long-term deep *SM* depletion reducing the baseflow component, while dry spring weather conditions simultaneously reduced the surface runoff component. These reductions resulted in overall low streamflow despite adequate snowpack. In contrast, the above-average snowpack of 2023 produced a good amount of streamflow because the rootzone *SM* filled quickly, generating substantial surface runoff, as deeper soil layers also continued to gradually recharge.

These findings address a central question of this thesis: where does snowmelt go during drought periods? The results demonstrate that during prolonged drought, deep soil moisture reservoirs become increasingly depleted, meaning a significant portion of the snowmelt must first replenish these deficits before contributing to streamflow. This soil moisture deficit creates a situation where even years with average snowpack (like 2020-2021) produce below-average streamflow.

The key takeaway from this experiment is that drought recovery in the CRB operates on different timescales for different components of the hydrological system. While rootzone *SM* can recover relatively quickly (within months), deep *SM* and baseflow require years to fully recover from prolonged drought conditions. This persistence explains why the basin doesn't immediately return to normal streamflow patterns after a single wet year. Instead, multiple consecutive wet years are necessary to fully recharge the deep soil moisture deficit created by long-term drought, which in turn

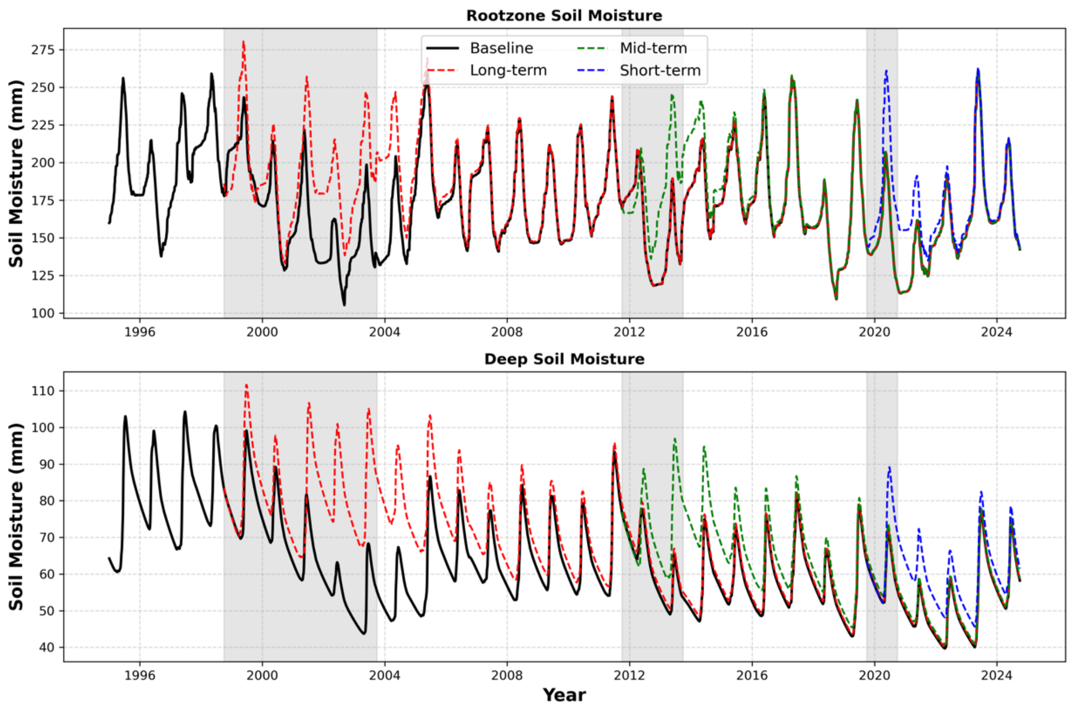


Figure 57. Temporal evolution of *SM* on the CRB under different experimental scenarios. Gray-shaded areas indicate time periods where meteorological forcing was replaced with wet years.

is needed to restore the baseflow component of streamflow. This understanding provides valuable context for water managers interpreting the basin's response to recent climate variations and helps explain why snowpack-streamflow relationships have become less predictable during the ongoing drought.

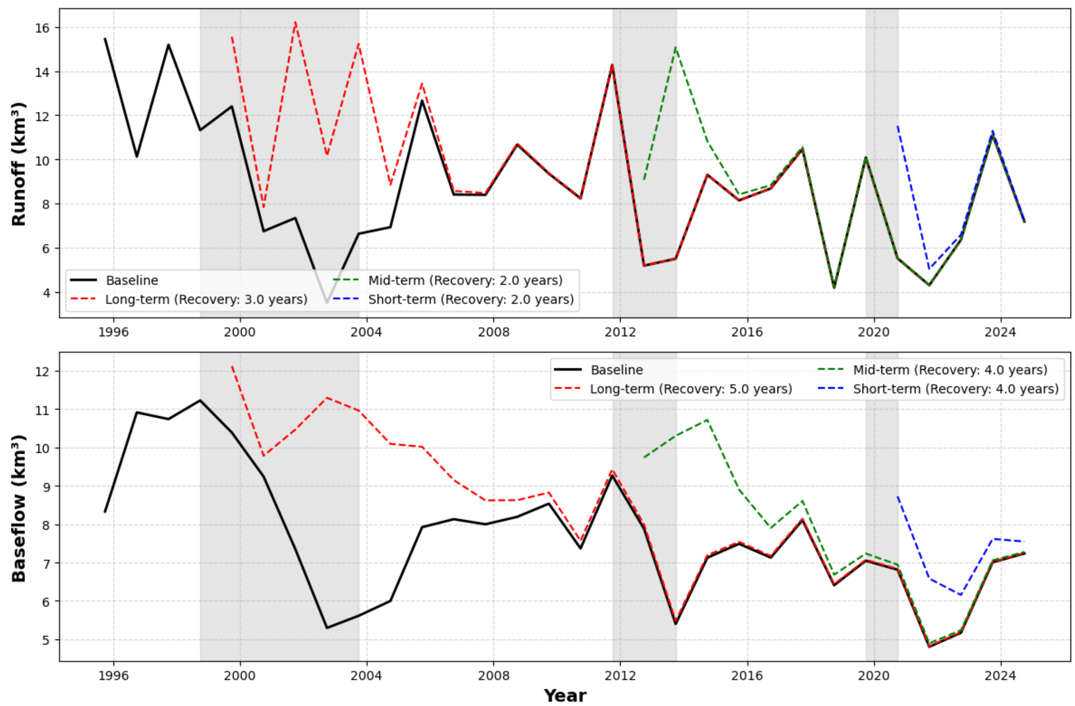


Figure 58. Temporal evolution of streamflow components on the CRB under different experimental scenarios. Gray-shaded areas indicate time periods where meteorological forcing was replaced with wet years.

3.3.6 Evapotranspiration Responses

Having established the influence of various hydrological components on streamflow generation, it is necessary to investigate the fate of water that does not contribute to streamflow during drought recovery periods. Figure 59 illustrates the relationships between April-September (AMJJAS) *ET* and three critical variables: AMJ total *P*, AMJ mean *T*, and initial *SM*. These results are derived from the experiments shown in sections 3.3.2 (AMJ climate experiment) and 3.3.3 (initial *SM* experiment).

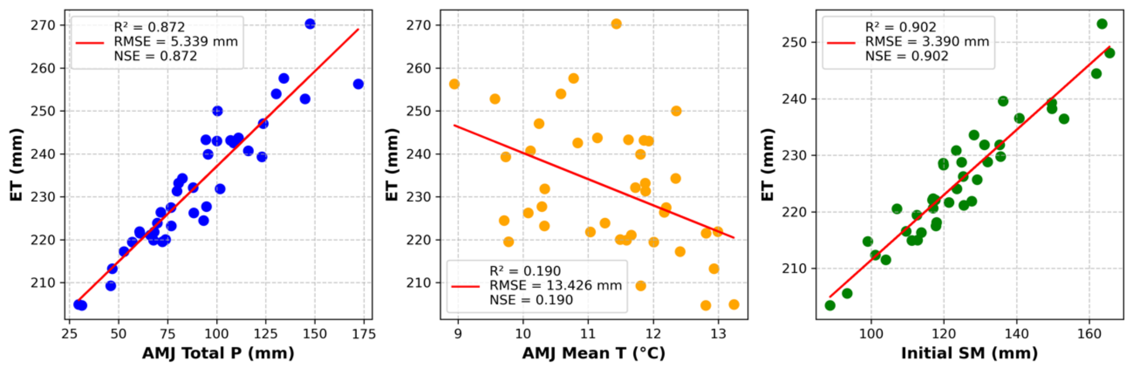


Figure 59. Relationship between AMJJAS *ET* and AMJ total *P* (left), AMJ mean *T* (middle), and initial *SM* (right). Each point represents a water year from 1984-2023.

The analysis reveals a strong positive correlation between AMJ *P* and *ET* (R^2 = 0.872), indicating that substantial portions of spring *P* are redirected to atmospheric losses. This relationship is mediated by *SM*, as *P* must first infiltrate the soil before becoming available for *ET* processes. As *SM* increases from *P* inputs, vegetation responds with increased transpiration, and direct soil evaporation also intensifies, resulting in greater overall *ET* fluxes. The relationship between AMJ *T* and *ET* exhibits a negative trend with a low correlation (R^2 = 0.190). This pattern can be attributed to the co-occurrence of higher *T* with lower *P* in the historical record, resulting in water-limited conditions where *ET* is constrained by moisture availability rather than energy input. Despite increased atmospheric demand during warmer periods, actual *ET* decreases when insufficient water is available in the soil profile, highlighting the water-limited system in the CRB.

Most significantly, the strong relationship between initial *SM* and *ET* (R^2 = 0.902) demonstrates how antecedent moisture conditions control the partitioning of available water. This relationship helps explain the "missing snowmelt" phenomenon observed in the CRB. When soils begin the WY with significant moisture deficits due to prolonged drought conditions, a substantial portion of snowmelt is allocated to replenishing these deficits rather than generating runoff. As *SM* is replenished, this water becomes accessible for plant transpiration and soil evaporation, effectively creating a diversion pathway that redirects potential streamflow from snowmelt toward atmospheric losses through *ET*. This mechanism explains why years with average snowpack can produce below-average streamflow when antecedent soil conditions are dry, as observed in water years 2020 and 2021.

3.3.7 Historical Evidence: Where Does All the Snowmelt Go?

The analysis of historical data provides a valuable opportunity to validate the mechanisms identified in the numerical experiments through real-world examples. The years from 1984 to 2023 encompass various combinations of initial *SM*, AMJ *P*, and AMJ *T* conditions that help explain the observed patterns in snowmelt-streamflow relationships and address the central question of where snowmelt goes when it doesn't produce expected streamflow. Figure 60 illustrates the relationship between snowpack and resulting streamflow for the same period. The vertical axis represents the difference between relative *SWE* percentage and relative Lake Powell inflow percentage, where positive values indicate that snowpack produced less streamflow than expected, and negative values show cases where streamflow exceeded what would be predicted by snowpack alone. The color of each data point represents the relative inflow as a percentage of normal while the size of each marker indicates snowpack conditions, with larger markers for wet conditions (>120% of normal), medium markers for normal

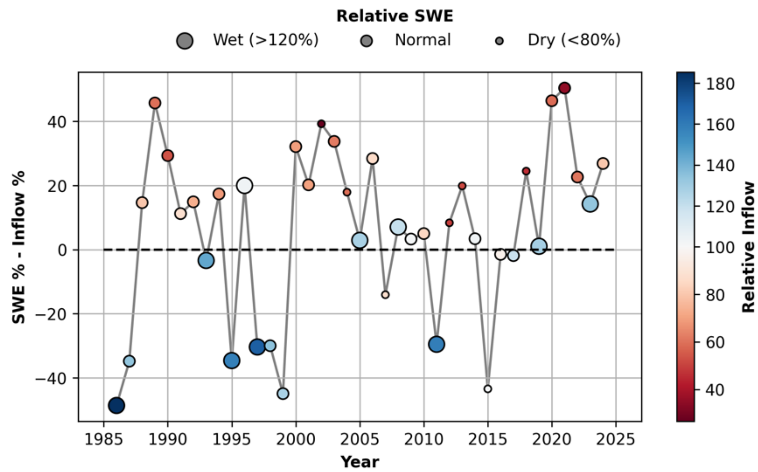


Figure 60. Difference between relative *SWE* and resulting inflow percentages (values shown in Figure 1) in the CRB (1984-2023). Marker colors represent the relative inflow percentage, and marker sizes indicate *SWE* conditions.

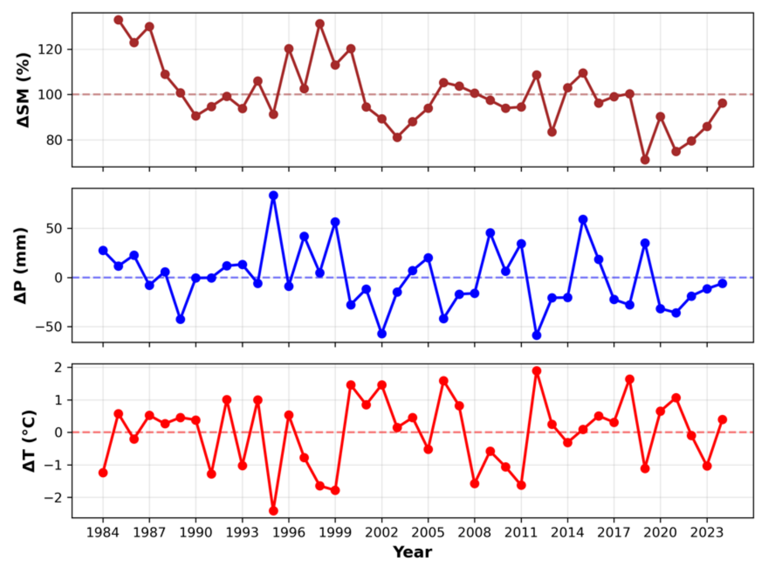


Figure 61. Historical anomalies in the CRB (1984-2023) showing anomalies of initial *SM* (%), AMJ *P* (mm), and AMJ *T* (°C) over 1984-2023.

conditions, and small markers for dry conditions (<80% of normal). Figure 61 presents the historical record of initial *SM* (Δ*SM* in %), spring *P* (Δ*P* in mm), and spring *T* (ΔT in °C) anomalies across the 40-year period.

The years with near-normal snowpack (80-120% of average) but divergent streamflow outcomes are particularly informative for testing the experimental findings. In 2021, the snowpack was 84% of normal but produced only 51% of the normal inflow. This is because the initial *SM* was extremely low (72% of normal), spring *P* was well below average (-30 mm), and mean spring *T* was significantly above normal (+1.7 °C). This combination created conditions where a substantial portion of snowmelt went toward replenishing *SM* deficits rather than contributing to streamflow. Similar patterns were observed in 1990, 2002 and 2020 (the test year of this study) as well. However, the difference in 1989 (104% of normal) is attributed to spring *P* deficits (-42.6 mm) rather than initial *SM* deficits (100% of normal). Another interesting case is for the year 2000, where snowpack was normal (100%) with wet initial conditions (120%). However, the spring was both very dry (-27.7) and hot (+1.45 °C), resulting in below average inflows (73% of normal).

Conversely, certain years demonstrate more efficient snowmelt-to-streamflow conversion. In 2005, despite near-normal snowpack (96%) and below-normal *SM* (88%), inflows reached 101% of normal levels. This can be attributed to above-average spring precipitation (+40 mm) and near-normal temperatures, which supplemented snowmelt contribution. Similarly, 1992 showed above-normal snowpack (112%) and proportionately higher streamflow (127%), facilitated by near-normal *SM* and cooler *T* that maintained efficient conversion. Examples of years having near normal snowpack, below normal streamflow but still producing near normal inflows due to cool/wet springs include 1995 and 2011. Conversely, examples wet initial conditions offsetting normal/warm-dry spring conditions include the years 1998 and 1987.

Building on the insights from numerical experiments, Figure 62 shows how drought signals propagate through different components of the hydrological system over the historical period 1984–2023. The time series of rootzone *SM*, surface runoff, deep *SM*, and baseflow (averaged over the UCRB) collectively reveal the memory and response dynamics of the CRB system. Four distinct case study periods are highlighted to demonstrate the sequential and cumulative nature of hydrological drought propagation. In the WY 2000, wet initial conditions in the deep *SM* layer buffered the system against a hot and dry spring, sustaining baseflow even as rootzone *SM* and runoff declined. This illustrates how deep *SM* can act as a drought buffer, sustaining subsurface flow in the face of meteorological anomalies. By contrast, the 2005 period highlights the asynchronous recovery of different system components. While favorable spring conditions facilitated rapid rebound in rootzone *SM* and runoff, deep *SM* and baseflow only began to recover gradually. This emphasizes the longer timescales required for deep storage layers to recharge, even when surface moisture conditions improve.

The 2010–2011 period shows the inverse behavior: despite initially dry rootzone conditions, cool and wet spring conditions triggered a strong and immediate surface hydrologic response, with runoff and rootzone *SM* increasing rapidly. Deep *SM* and baseflow followed, but slowly, reflecting their lagged but cumulative response to persistent wetting. Finally, the 2020–2022 multi-year drought period demonstrates the full extent of system memory and drought propagation. Across these years, sustained dry

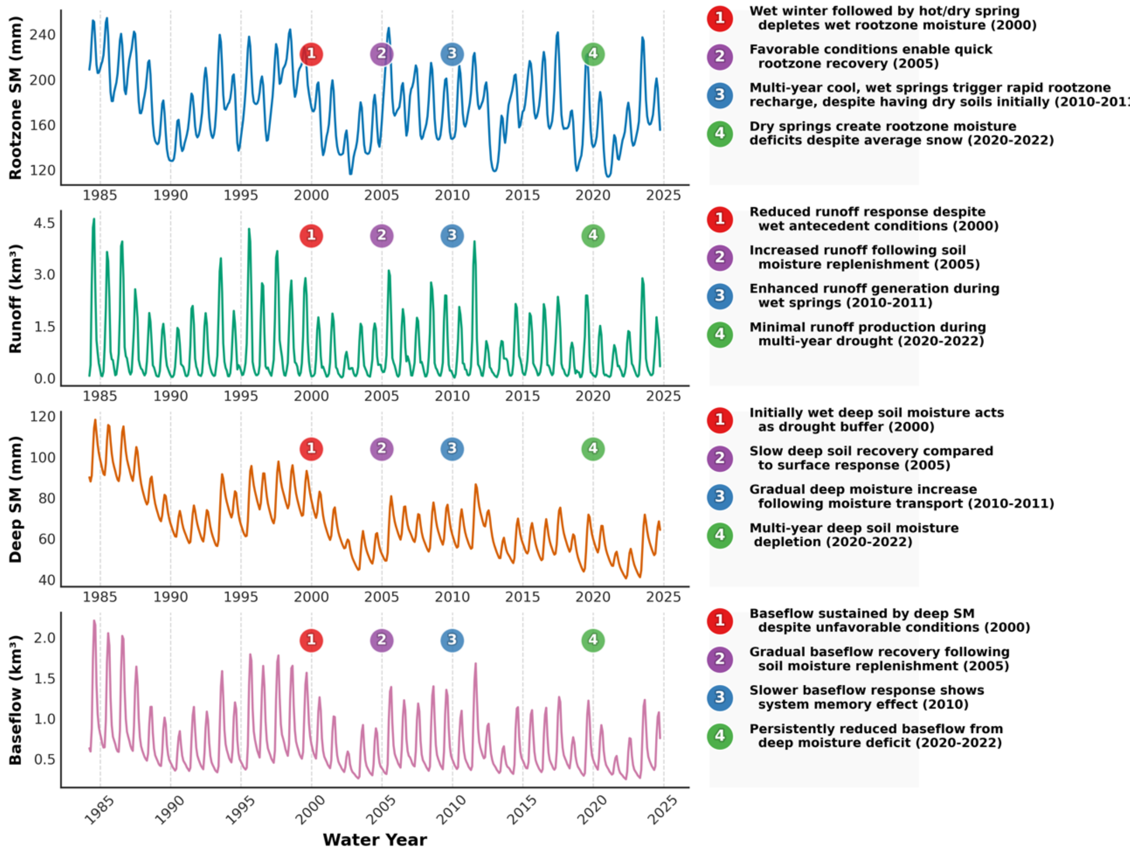


Figure 62. Soil moisture memory and drought propagation mechanisms in the CRB (1984-2024). Four major hydrological patterns are identified by numbered markers: (1) buffering effect of wet antecedent conditions (2000), (2) recovery dynamics (2005), (3) rapid response to wet springs despite dry antecedent conditions (2010-2011), and (4) prolonged deep moisture deficit during severe drought (2020-2022).

conditions led to the depletion of both rootzone and deep *SM*, which in turn suppressed runoff and baseflow generation despite near-average snowpack in the years to follow. This provides a mechanistic explanation for the "missing snowmelt" phenomena, as snowmelt during prolonged droughts is largely absorbed by the soil layers to replenish deficits, delaying or diminishing streamflow response.

Together, these case studies underscore the critical role of deep *SM* as both a buffer and a bottleneck in drought recovery. Surface components such as runoff and rootzone moisture may recover within a single season, but deep *SM* and baseflow often require multiple wet years to fully recover. This disconnect in recovery timescales explains why snowpack–streamflow relationships have become increasingly unreliable in recent decades and why streamflow remains suppressed long after meteorological drought conditions appear to have improved or when winter snowpack remains near-normal. These findings provide strong mechanistic support for the central hypothesis of this thesis: that deep soil moisture memory, shaped by preceding drought conditions, plays a decisive role in shaping streamflow anomalies and long-term water availability in the CRB.

4. CONCLUSIONS AND FUTURE WORK

The complex relationship between snowmelt and streamflow in the CRB during drought conditions was investigated in this research, focusing on three key questions: (1) the representativeness of SNOTEL measurements for basin-wide snowpack conditions, (2) the influence of spring weather on streamflow generation, and (3) the role of soil moisture in controlling streamflow efficiency. To ensure the robustness of the analysis, the model was calibrated in two phases: first, the snow module was calibrated against SNOTEL observations to improve snow accumulation and melt processes, and subsequently, soil parameters were fine-tuned using naturalized streamflow data from USBR to enhance runoff and baseflow representation. Validation using SMAP soil moisture and GRACE terrestrial water storage observations further reinforced confidence in the model outputs. By integrating multiple observational datasets with process-based modeling, this study provided significant insights into the fundamental question: Where does all the snowmelt go during drought periods?

First, it was demonstrated that the current operational approach using 104 SNOTEL stations in the UCRB provides a limited perspective of basin-wide snow conditions. These stations predominantly represent areas with above-average snow accumulation (mean daily $SWE > 40$ mm), covering only about 30% of the snow-producing area in the UCRB. This spatial bias leads to systematic errors in snowpack estimation, with overestimation during dry years and underestimation during wet years. This finding highlights the need for more comprehensive monitoring networks or model-based approaches that capture the spatial variability of snow conditions across the UCRB.

Second, through controlled numerical experiments, it was revealed that AMJ weather conditions substantially influence annual streamflow outcomes, when winter and near-peak snowpack conditions remains constant. The simulations showed that AMJ weather patterns alone can cause more around 8.31 km^3 differences (19.64 - 11.33 km^3) in annual streamflow volumes in the UCRB. Precipitation during this period strongly influenced annual streamflow generation ($R^2 = 0.91$) more than temperature ($R^2 = 0.58$), though both factors were identified as significant contributors to the overall water balance. Analysis of seasonal differences revealed that summer or JJAS streamflow showed greater sensitivity to AMJ climate variations than WY totals, with values ranging from 3.74 to 13.37 km^3. This summer sensitivity demonstrates the direct and immediate impact of spring climate on the primary snowmelt season due to temporal proximity, highlighting the importance of incorporating late-season climate forecasts into water management planning.

Third, antecedent *SM* conditions at the start of the water year were shown to produce nearly a twofold difference of 8.45 km^3 (17.89 - 9.44 km^3) in annual streamflow volumes in the UCRB when all other factors remained constant. Deep *SM* was found to be particularly influential, with approximately 67% of streamflow variability attributed to baseflow rather than surface runoff. Summer (JJAS) streamflow analysis revealed how *SM* impacts extend through different seasons, with summer baseflow (ranging from 2.21 to 4.22 km^3) showing greater sensitivity than summer runoff (ranging from 1.34 to 2.21 km^3). This demonstrates a key difference between climate and initial *SM* influences: while spring climate has a more immediate impact on summer flows due to temporal proximity, soil moisture provides a sustained influence throughout the entire water year through baseflow generation. The regression analysis confirmed *SM* as the dominant control on streamflow in the UCRB, contributing about 74% to overall variability, followed by precipitation (24%) and temperature (1.7%). This demonstrates that *SM* is substantially more sensitive and important than AMJ weather for determining streamflow outcomes in the UCRB. In contrast, the LCRB showed greater precipitation sensitivity, which accounted for around 70% of streamflow variability.

The *ET* analysis provided critical insights into the fate of "missing snowmelt" during drought periods. A strong positive correlation between AMJ *P* and *ET* (R^2 = 0.872) indicated that substantial portions of spring *P* are redirected to atmospheric losses. Moreover, the strong relationship between initial *SM* and *ET* (R^2 = 0.902) demonstrated how antecedent *SM* controls the partitioning of available water. When soils begin the water year with significant moisture deficits due to prolonged drought, a substantial portion of snowmelt is allocated to replenishing these deficits rather than generating runoff. As *SM* is replenished, this water becomes accessible for plant transpiration and soil evaporation, effectively creating a diversion pathway that redirects potential streamflow toward atmospheric losses. This mechanism explains why years with average snowpack can produce below-average streamflow when antecedent soil conditions are dry.

The final set of experiments showed that deep *SM* plays a critical role in sustaining streamflow during and after drought across a multi-year timescale. While upper soil layers respond rapidly to atmospheric conditions and have limited memory, deep *SM* retains anomalies for years, providing a multi-year buffer through enhanced baseflow. Recovery simulations further revealed that surface *SM* and runoff rebound relatively quickly, but deep *SM* and baseflow recover over much longer timescales. These findings explain why average snowpack years like 2020–2021 still yielded low streamflow: depleted deep *SM* absorbed much of the snowmelt before it could contribute to runoff. In contrast, improved streamflow in 2023 was due to above-average snowpack and subsequent snowmelt under favorable spring conditions, which led to rootzone soil moisture recharge and large surface runoff. Overall, the results underscore that full hydrological recovery from drought in the CRB requires multiple consecutive wet years, highlighting the need to consider *SM* memory.

Analysis of the historical record (1984-2023) provided real-world validation of these mechanisms through several case examples. The year 2021 exemplified how multiple factors can combine to reduce streamflow efficiency: despite 84% of normal snowpack, initial *SM* was extremely low (72% of normal), spring *P* was well below average (-30 mm), and mean spring *T* was significantly above normal (+1.7°C), resulting in just 51% of normal inflow. Similar patterns were observed in 1990, 2002, and 2020. Examination of multi-year patterns revealed four distinct hydrological responses across the historical record: (1) deep *SM* initially buffering snowmelt response during the 2000 wet-to-dry transition, (2) asynchronous recovery of different system components following favorable conditions in 2005, (3) rapid response to wet springs despite dry antecedent conditions during 2010-2011, and (4) prolonged deep moisture deficit during the severe 2020-2022 drought period. These historical cases demonstrate that while surface components may recover within a single favorable season, deep *SM* and baseflow often require multiple wet years to fully recover, explaining why snowpack-streamflow relationships have become increasingly unreliable in recent decades.

There are several limitations of this study that should be acknowledged. First, the model outputs are highly sensitive to calibration parameters, particularly those governing subsurface processes. For instance, the maximum baseflow rate (Ds_{max}) was calibrated to match observed streamflow, but this parameter can exert a strong influence on baseflow dynamics. Second, snow processes were calibrated using point-based station data, which may not adequately capture the spatial variability of snow accumulation and melt across the heterogeneous terrain of the CRB. Third, while the controlled experiments provided valuable insights into streamflow sensitivity, they primarily focused on the WY 2020. As such, the broader applicability of these findings to different years with varying climate conditions remains uncertain. These limitations suggest that future studies incorporating spatially distributed calibration strategies, multi-year experiment frameworks, and a more comprehensive representation of model uncertainty would further strengthen the robustness and generalizability of the conclusions presented here.

These findings explain the recent "missing snowmelt" phenomenon in the CRB, where near-average snowpack has failed to produce expected streamflow volumes. The combination of dry antecedent soil conditions, unfavorable spring weather, and long-term moisture deficits was shown to fundamentally alter the snow-to-flow relationship, requiring water managers to reconsider traditional forecasting approaches. From a practical perspective, several pathways for improving water management in the CRB are suggested from the outcomes: (1) expanding snow monitoring capabilities to capture the full spatial variability of basin-wide conditions, (2) developing more sophisticated seasonal forecasting systems that account for spring weather patterns, and (3) incorporating deep soil moisture monitoring into operational decision-making frameworks through the use of remote sensing products like SMAP and GRACE or regionally calibrated hydrological model like VIC.

There are several promising directions for future research that emerge from this study. First, developing operational products that combine ground-based SNOTEL observations with model outputs and remote sensing data could provide more comprehensive and less biased assessments of basin-wide snow conditions. This could be achieved by assimilating SNOTEL-derived *SWE* data into distributed hydrological or land surface models using data assimilation techniques such as the Ensemble Kalman Filter (Micheletty et al., 2022). Further, the method of 'averaging' 104 SNOTEL stations assumes equal influence of all SNOTEL stations across the domain. Introducing a spatial weighting scheme based on factors such as elevation, proximity to snow accumulation zones, and historical correlation with basin-wide *SWE* could refine the influence of each station during basin-wide snow condition assessment. Such weighting could be implemented within data assimilation frameworks or through statistical bias correction methods to emphasize those most critical for snowpack dynamics. This approach would help create a more accurate and spatially representative picture of snow distribution, ultimately improving both research outputs and operational capabilities.

Second, expanding the use of emerging remote sensing products could offer new insights into the water balance components during drought. While this study focused primarily on ground networks, SMAP, and GRACE, incorporating additional datasets related to *ET* and surface water dynamics could help provide a more comprehensive picture. For instance, ECOSTRESS and OpenET products offer spatially distributed estimates of *ET*, which could be valuable for validating modeled surface fluxes and understanding water losses during drought recovery. Similarly, the upcoming SWOT (Surface Water and Ocean Topography) mission will provide high-resolution observations of surface water extent and storage, addressing a key limitation in current GRACE-based decomposition methods, which lack explicit spatial representation of surface water. Once available, SWOT data could be integrated with other satellite products to create a better spatial understanding of groundwater conditions.

Third, future work could explore the generalizability of the experimental findings by expanding the controlled simulations across a wider range of hydrological years and climate conditions. This study focused primarily on WY 2020 and a few selected recovery scenarios, which provided clear mechanistic insights but may not capture the full range of interannual variability observed in the CRB. Conducting a more comprehensive set of experiments across multiple drought and non-drought periods would help assess the consistency of soil moisture memory effects and recovery trajectories under different combinations of snowpack, spring climate, and antecedent conditions. This could also enable a probabilistic characterization of drought recovery timescales and help identify threshold behaviors, such as the minimum wet period duration required to recharge deep soil layers fully. Such insights would strengthen the application of this framework in an operational context.

Finally, exploring the interactions between land cover change and the snowmelt-streamflow relationship could extend this study's findings. Whitney et al. (2023b) demonstrated that forest disturbances in the CRB could partially offset climate-driven streamflow efficiency declines by increasing snowpack in open areas and reducing *ET*. The study found that "higher ground snowpack and associated snowmelt increases in late winter to early spring months were the primary cause for streamflow increase" following forest disturbance. This mechanism would interact with the spring weather sensitivity identified in the current study, as exposed snowpack in disturbed areas responds strongly to AMJ *T* and *P*. Gleason et al. (2019) found that post-disturbance landscapes experience faster snowmelt rates, which could alter moisture transport to deep soil layers and its memory. Future research incorporating both vegetation disturbance and recovery processes into the VIC modeling framework would provide valuable insights into how these factors might further influence the snow-to-flow relationship during drought periods in the CRB.

As the basin faces increasing hydroclimatic variability and water stress, the findings from this study underscore the importance of considering not only seasonal meteorological inputs but also long-term *SM* conditions. In addressing the question of where snowmelt goes during drought years, this work highlights that snowpack alone is insufficient to explain streamflow outcomes, as subsurface conditions, particularly deep soil moisture deficits, play a critical role in modulating how much snowmelt ultimately contributes to runoff. These insights point to the need for forecasting and management strategies that move beyond a sole focus on winter snowpack and instead incorporate a more integrated understanding of antecedent conditions and seasonal climate variability to better anticipate drought impacts and recovery trajectories.

REFERENCES

ADWR. (2022a). *Colorado River Shortage FAQ*. Phoenix, AZ: Arizona Department of Water Resources. Available at: https://www.azwater.gov/sites/default/files/2022-12/2022-08-colorado-river-shortage-faq.pdf (Accessed 09/15/2025)

ADWR. (2022b). *Colorado River Shortage, 2022 Fact Sheet*. Phoenix, AZ: Arizona Department of Water Resources. Available at: https://www.azwater.gov/sites/default/files/media/ADWR-CAP-FactSheet-CoRiverShortage-081321.pdf (Accessed 09/15/2025)

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

Andreadis, K. M., Storck, P., & Lettenmaier, D. P. (2009). Modeling snow accumulation and ablation processes in forested environments. *Water Resources Research*, 45(5), W05429. https://doi.org/10.1029/2008WR007042

Barrett, A. (2003). *National Operational Hydrologic Remote Sensing Center SNOw Data Assimilation System (SNODAS) Products at NSIDC*. (NSIDC Special Report 11) (p.19). Boulder, CO: National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences. https://nsidc.org/sites/default/files/nsidc_special_report_11.pdf

Bennett, K. E., Bohn, T. J., Solander, K., McDowell, N. G., Xu, C., Vivoni, E., & Middleton, R. S. (2018). Climate-driven disturbances in the San Juan River sub-basin of the Colorado River. *Hydrology and Earth System Sciences*, 22(1), 709–725. https://doi.org/10.5194/hess-22-709-2018

Boardman, E. N., Renshaw, C. E., Shriver, R. K., Walters, R., McGurk, B., Painter, T. H., Deems, J. S., Bormann, K. J., Lewis, G. M., Dethier, E. N., & Harpold, A. A. (2024). Sources of seasonal water supply forecast uncertainty during snow drought in the Sierra Nevada. *Journal of the American Water Resources Association*, 60(5), 972–990. https://doi.org/10.1111/1752-1688.13221

Broxton, P., Zeng, X., & Dawson, N. (2019). Daily 4 km gridded SWE and snow depth from assimilated in-situ and modeled data over the conterminous US. (NSIDC-0719, Version 1) [Dataset]. National Snow and Ice Data Center. Available at: https://doi.org/10.5067/0GGPB220EX6A (Accessed 08/02/2023)

Bureau of Reclamation. (2012). *Colorado River Basin Water Supply and Demand Study: Executive Summary* (p.34). Washington, DC: U.S. Department of the Interior, Bureau of Reclamation. https://www.usbr.gov/watersmart/bsp/docs/finalreport/ColoradoRiver/CRBS_Executive_Summary_FINAL.pdf

Castle, S. L., Thomas, B. F., Reager, J. T., Rodell, M., Swenson, S. C., & Famiglietti, J. S. (2014). Groundwater depletion during drought threatens future water security of the Colorado River Basin. *Geophysical Research Letters*, 41(16), 5904–5911. https://doi.org/10.1002/2014GL061055

CBRFC. (2024). CBRFC Conditions Dashboard. [Dataset]. Colorado Basin River Forecast Center. Available at: https://www.cbrfc.noaa.gov/lmap/lmap.php (Accessed 08/10/2024)

Christensen, N. S., Wood, A. W., Voisin, N., Lettenmaier, D. P., & Palmer, R. N. (2004). The effects of climate change on the hydrology and water resources of the Colorado River Basin. *Climatic Change*, 62(1), 337–363. https://doi.org/10.1023/B:CLIM.0000013684.13621.1f

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

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N., McDonald, K. C., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., … Van Zyl, J. (2010). The Soil Moisture Active Passive (SMAP) Mission. *Proceedings of the IEEE*. 98(5), 704–716. https://doi.org/10.1109/JPROC.2010.2043918

Franchini, M., & Pacciani, M. (1991). Comparative analysis of several conceptual rainfall-runoff models. *Journal of Hydrology*, 122(1), 161–219. https://doi.org/10.1016/0022-1694(91)90178-K

Goble, P. E., & Schumacher, R. S. (2023). On the sources of water supply forecast error in western Colorado. *Journal of Hydrometeorology*, 24(12), 2321-2332 https://doi.org/10.1175/JHM-D-23-0004.1

Hamman, J. J., Nijssen, B., Bohn, T. J., Gergel, D. R., & Mao, Y. (2018). The Variable Infiltration Capacity model version 5 (VIC-5): Infrastructure improvements for new applications and reproducibility. *Geoscientific Model Development*, 11(8), 3481–3496. https://doi.org/10.5194/gmd-11-3481-2018

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

Heldmyer, A. J., Bjarke, N. R., & Livneh, B. (2023). A 21st-Century perspective on snow drought in the Upper Colorado river basin. *Journal of the American Water Resources Association*, 59(2), 396–415. https://doi.org/10.1111/1752-1688.13095

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

Hogan, D., & Lundquist, J. D. (2024). Recent Upper Colorado River streamflow declines driven by loss of spring precipitation. *Geophysical Research Letters*, 51(16), e2024GL109826. https://doi.org/10.1029/2024GL109826

Hufkens, K., Hill, D., Schneider, D., & Potter, N. (2024). bluegreen-labs/snotelr: Correcting metric conversions (v1.5.2). [Software]. Zenodo. Available at: https://doi.org/10.5281/zenodo.7012728. (Accessed 02/05/2024)

Jennewein, D. M., Lee, J., Kurtz, C., Dizon, W., Shaeffer, I., Chapman, A., ... & Yalim, J. (2023). The sol supercomputer at Arizona State University. In *Practice and Experience in Advanced Research Computing 2023: Computing for the Common Good* (pp. 296-301). https://doi.org/10.1145/3569951.3597573

Jin, S., & Feng, G. (2013). Large-scale variations of global groundwater from satellite gravimetry and hydrological models, 2002–2012. *Global and Planetary Change*, 106, 20-30. https://doi.org/10.1016/j.gloplacha.2013.02.008

Koster, R. D., Guo, Z., Yang, R., Dirmeyer, P. A., Mitchell, K., & Puma, M. J. (2009). On the nature of soil moisture in land surface models. *Journal of Climate*, 22(16), 4322–4335. https://doi.org/10.1175/2009JCLI2832.1

Koster, R. D., Liu, Q., Crow, W. T., & Reichle, R. H. (2023). Late-fall satellite-based soil moisture observations show clear connections to subsequent spring streamflow. *Nature Communications*, 14(1), 3545. https://doi.org/10.1038/s41467-023-39318-3

Lapides, D. A., Hahm, W. J., Rempe, D. M., Whiting, J., & Dralle, D. N. (2022). Causes of missing snowmelt following drought. *Geophysical Research Letters*, 49(19), e2022GL100505. https://doi.org/10.1029/2022GL100505

Li, B., Rodell, M., Kumar, S., Beaudoing, H. K., Getirana, A., Zaitchik, B. F., de Goncalves, L. G., Cossetin, C., Bhanja, S., Mukherjee, A., Tian, S., Tangdamrongsub, N., Long, D., Nanteza, J., Lee, J., Policelli, F., Goni, I. B., Daira, D., Bila, M., … Bettadpur, S. (2019). Global GRACE data assimilation for groundwater and drought monitoring: Advances and challenges. *Water Resources Research*, 55(9), 7564–7586. https://doi.org/10.1029/2018WR024618

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

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres*, 99(D7), 14415–14428. https://doi.org/10.1029/94JD00483

Liang, X., Wood, E. F., & Lettenmaier, D. P. (1996). Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification. *Global and Planetary Change*, *13*(1-4), 195-206. https://doi.org/10.1016/0921-8181(95)00046-1

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

Livneh, B., Badger, A. M., & Lukas, J. J. (2017). Assessing the robustness of snow-based drought indicators in the Upper Colorado river basin under future climate change. *World Environmental and Water Resources Congress 2017*, 511–525. https://doi.org/10.1061/9780784480618.051

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., Cayan, D. R., & Brekke, L. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013. *Scientific Data*, 2(1), 150042. https://doi.org/10.1038/sdata.2015.42

Lukas, J., & Payton, E. (2020). *Colorado River Basin Climate and Hydrology: State of the Science* (p.531). Boulder, CO: Western Water Assessment, University of Colorado Boulder. https://scholar.colorado.edu/concern/reports/8w32r663z

Marks, D., & Dozier, J. (1992). Climate and energy exchange at the snow surface in the alpine region of the Sierra Nevada: 2. Snow cover energy balance. *Water Resources Research*, 28(11), 3043-3054. https://doi.org/10.1029/92WR01483

Maxwell, R. M., Condon, L. E., & Kollet, S. J. (2015). A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3. *Geoscientific model development*, 8(3), 923-937. https://doi.org/10.5194/gmd-8-923-2015.

McNie, E. (2014). *Evaluation of the NIDIS Upper Colorado River Basin Drought Early Warning System* (p.44). Boulder, CO: Western Water Assessment, University of Colorado, Boulder. https://wwa.colorado.edu/sites/default/files/2021-07/Evaluation%20of%20the%20Nidis%20Upper%20Colorado%20River%20Basin%20Drought%20Early%20Warning%20System.pdf

Meromy, L., Molotch, N. P., Link, T. E., Fassnacht, S. R., & Rice, R. (2013). Subgrid variability of snow water equivalent at operational snow stations in the western USA. *Hydrological Processes*, 27(17), 2383-2400. https://doi.org/10.1002/hyp.9355

Micheletty, P., Perrot, D., Day, G., & Rittger, K. (2022). Assimilation of ground and satellite snow observations in a distributed hydrologic model for water supply forecasting. *Journal of the American Water Resources Association*, 58(6), 1030–1048. https://doi.org/10.1111/1752-1688.12975

Miller, W. P., & Piechota, T. C. (2011). Trends in western U.S. snowpack and related Upper Colorado river basin streamflow. *Journal of the American Water Resources Association*, 47(6), 1197–1210. https://doi.org/10.1111/j.1752-1688.2011.00565.x

Molotch, N. P., & Bales, R. C. (2006). SNOTEL representativeness in the Rio Grande headwaters on the basis of physiographics and remotely sensed snow cover persistence. *Hydrological Processes: An International Journal*, 20(4), 723-739. https://doi.org/10.1002/hyp.6128

Mote, P. W., Li, S., Lettenmaier, D. P., Xiao, M., & Engel, R. (2018). Dramatic declines in snowpack in the western US. *Npj Climate and Atmospheric Science*, 1(1), 1–6. https://doi.org/10.1038/s41612-018-0012-1

NASA. (2023). JPL GRACE and GRACE-FO Mascon Ocean, Ice, and Hydrology Equivalent Water Height Coastal Resolution Improvement (CRI) Filtered Release 06.3 Version 04 [Dataset]. Physical Oceanography Distributed Active Archive Center (PO.DAAC). Available at: https://podaac.jpl.nasa.gov/dataset/TELLUS_GRAC-GRFO_MASCON_CRI_GRID_RL06.3_V4 (Accessed 05/15/2023)

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

PRISM Climate Group. (2025). *Descriptions of PRISM Spatial Climate Datasets for the Conterminous United States* (p.33). Corvallis, OR: Oregon State University.https://prism.oregonstate.edu/documents/PRISM_datasets.pdf

Reichle, R. H., Liu, Q., Koster, R. D., Crow, W. T., De Lannoy, G. J. M., Kimball, J. S., Ardizzone, J. V., Bosch, D., Colliander, A., Cosh, M., Kolassa, J., Mahanama, S. P., Prueger, J., Starks, P., & Walker, J. P. (2019). Version 4 of the SMAP Level-4 soil moisture algorithm and data product. *Journal of Advances in Modeling Earth Systems*, 11(10), 3106–3130. https://doi.org/10.1029/2019MS001729

Rodell, M., & Famiglietti, J. S. (2001). An analysis of terrestrial water storage variations in Illinois with implications for the Gravity Recovery and Climate Experiment (GRACE). *Water Resources Research*, 37(5), 1327–1339. https://doi.org/10.1029/2000WR900306

Scanlon, B. R., Pool, D. R., Rateb, A., Conway, B., Sorensen, K., Udall, B., & Reedy, R. C. (2025). Multidecadal drought impacts on the Lower Colorado Basin with implications for future management. *Communications Earth & Environment*, 6(1), 214. https://doi.org/10.1038/s43247-025-02149-9

Schaperow, J. R., Li, D., Margulis, S. A., & Lettenmaier, D. P. (2021). A near-global, high resolution land surface parameter dataset for the variable infiltration capacity model. *Scientific Data*, 8(1), 216. https://doi.org/10.1038/s41597-021-00999-4

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

Shulgina, T., Gershunov, A., Hatchett, B. J., Guirguis, K., Subramanian, A. C., Margulis, S. A., Fang, Y., Cayan, D. R., Pierce, D. W., Dettinger, M., Anderson, M. L., & Ralph, F. M. (2023). Observed and projected changes in snow accumulation and snowline in California's snowy mountains. *Climate Dynamics*, 61(9), 4809–4824. https://doi.org/10.1007/s00382-023-06776-w

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

Strang, G. (1976). *Linear Algebra and Its Applications* (p.545). New York, NY: Academic Press.

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

Su, L., Cao, Q., Xiao, M., Mocko, D. M., Barlage, M., Li, D., Peters-Lidard, C. D., & Lettenmaier, D. P. (2021). Drought variability over the conterminous United States for the past century. *Journal of Hydrometeorology*, 22(5), 1153–1168. https://doi.org/10.1175/JHM-D-20-0158.1

Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C. (2004). The gravity recovery and climate experiment: Mission overview and early results. *Geophysical Research Letters*, 31(9), L09607. https://doi.org/10.1029/2004GL019920

The Colorado River Compact. (1922). *The Colorado River Compact* (p.8). Santa Fe, NM: Bureau of Reclamation. Available at: https://www.usbr.gov/lc/region/pao/pdfiles/crcompct.pdf

Udall, B., & Overpeck, J. (2017). The twenty-first century Colorado River hot drought and implications for the future. *Water Resources Research*, 53(3), 2404–2418. https://doi.org/10.1002/2016WR019638

USBR. (2024). Colorado River Basin Natural Flow and Salt Data. [Dataset]. U.S. Bureau of Reclamation. Available at: https://www.usbr.gov/lc/region/g4000/NaturalFlow/provisional.html (Accessed 01/05/2024)

Vano, J. A., Das, T., & Lettenmaier, D. P. (2012). Hydrologic sensitivities of Colorado River runoff to changes in precipitation and temperature. *Journal of Hydrometeorology*, 13(3), 932–949. https://doi.org/10.1175/JHM-D-11-069.1

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & Van Mulbregt, P. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature methods*, 17(3), 261-272. https://doi.org/10.1038/s41592-019-0686-2

Wang, Z., Vivoni, E. R., Whitney, K. M., Xiao, M., & Mascaro, G. (2024). On the sensitivity of future hydrology in the Colorado River to the selection of the precipitation partitioning method. *Water Resources Research*, 60(6), e2023WR035801. https://doi.org/10.1029/2023WR035801

Whitney, K. M., Vivoni, E. R., Bohn, T. J., Mascaro, G., Wang, Z., Xiao, M., Mahmoud, M. I., Cullom, C., & White, D. D. (2023a). Spatial attribution of declining Colorado River streamflow under future warming. *Journal of Hydrology*, 617, 129125. https://doi.org/10.1016/j.jhydrol.2023.129125

Whitney, K. M., Vivoni, E. R., Wang, Z., White, D. D., Quay, R., Mahmoud, M. I., & Templeton, N. P. (2023b). A stakeholder-engaged approach to anticipating forest disturbance impacts in the Colorado River basin under climate change. *Journal of Water Resources Planning and Management*, 149(7), 04023020. https://doi.org/10.1061/JWRMD5.WRENG-5905

Wiese, D. N., Landerer, F. W., & Watkins, M. M. (2016). Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. *Water Resources Research*, 52(9), 7490–7502. https://doi.org/10.1002/2016WR019344

Xiao, M., Mascaro, G., Wang, Z., Whitney, K. M., & Vivoni, E. R. (2022). On the value of satellite remote sensing to reduce uncertainties of regional simulations of the Colorado River. *Hydrology and Earth System Sciences*, 26(21), 5627–5646. https://doi.org/10.5194/hess-26-5627-2022

Xiao, M., Udall, B., & Lettenmaier, D. P. (2018). On the causes of declining Colorado river streamflows. *Water Resources Research*, 54(9), 6739–6756. https://doi.org/10.1029/2018WR023153

APPENDIX A

ACRONYMS

AMJ	April-May-June
CAP	Central Arizona Project
CBRFC	Colorado Basin River Forecasting Center
CLSM	Catchment Land Surface Model
COOP	Cooperative Observer Network
CRB	Colorado River Basin
GEOS-5 LDAS	Goddard Earth Observing System - Land Data Assimilation System
GRACE	Gravity Recovery and Climate Experiment
HUC	Hydrological Unit Codes
LCRB	Lower Colorado River Basin
MAF	Million-Acre-Feet
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NOHRSC	National Operational Hydrologic Remote Sensing Center
NRCS	Natural Resources Conservation Service
PODAAC	Physical Oceanography Distributed Active Archive Center
PRISM	Parameter-elevation Relationships on Independent Slopes Model
RMSE	Root Mean Squared Error
SCAN	Soil Climate Analysis Network
SMAP	Soil Moisture Active Passive
SNODAS	Snow Data Assimilation System
SNOTEL	Snow Telemetry
SWE	Snow Water Equivalent
TWSA	Terrestrial Water Storage Anomalies

UA	University of Arizona
UCRB	Upper Colorado River Basin
USBR	United States Bureau of Reclamation
USCRN	United States Climate Reference Network
VIC	Variable Infiltration Capacity

APPENDIX B

DATA REPOSITORY

This appendix provides detailed information about the datasets used in this study, including their sources, spatial and temporal characteristics, processing methods, and organization within the data repository. All datasets are stored in a centralized repository (https://www.dropbox.com/home/Swastik%20Ghimire/Data%20Repository) to facilitate reproducibility and future research.

B.1 Repository Structure

The data repository is organized into the following main directories:

- **VIC_outputs**: Contains model outputs from the VIC model
- **Analysis_Basin_Shapefiles**: Vector shapefiles defining major sub-basins within the CRB
- **README_files**: Documentation for dataset usage and processing workflows
- **Scripts**: Processing scripts for data manipulation and model execution
- **GRACE**: Terrestrial water storage anomaly data from GRACE
- **Naturalized_Streamflow**: Historical naturalized streamflow records from USBR
- **SMAP**: Soil moisture data from the SMAP. This folder contains surface *SM* data from individual SNOTEL *SM* stations as well.
- **SNODAS**: SWE data from SNODAS. This folder contains surface *SM* data from individual SNOTEL *SWE* stations as well.
- **SNOTEL**: Ground-based meteorological and snow observations from SNOTEL. This folder contains three files: metadata of all SNOTEL sites in the CRB, outputs from automated download containing *P, T* and *SWE* data and SWE data filtered for CRB.

- **PRISM**: Gridded climate data from PRISM merged with NLDAS wind speed data.

Table B1. Dataset Descriptions including format, spatial resolution, temporal coverage, variables and source.

Dataset	Format	Spatial Resolution	Temporal Coverage	Variables	Source
VIC Model Outputs	NetCDF	6 km	Daily, 1984-2024	Precipitation (P), temperature (T), SWE, soil moisture (SM), ET, etc. See global file for full list.	Generated using calibrated VIC-5 model
Analysis Basin Shapefiles	ESRI Shapefiles	Vector boundaries	N/A	Sub-basin delineations for major sub-basins in the CRB	Whitney et al. (2023)
GRACE	NetCDF	300 km	Global, Monthly, 2002-2024	Terrestrial Water Storage Anomalies (TWSA)	NASA PODAAC
Naturalized Streamflow	Excel	Point locations (gauges)	Monthly, 1981-2020	Streamflow (acre-feet)	U.S. Bureau of Reclamation
SMAP	NetCDF, CSV	9 km	Daily, 2015-2024	Surface SM (0-5 cm), Root zone SM (0-100 cm)	NASA AppEEARS
SNODAS	NetCDF, CSV	1 km	Daily, 2004-2022	Snow Water Equivalent (SWE)	National Snow and Ice Data Center

Dataset	Format	Spatial Resolution	Temporal Coverage	Variables	Source
SNOTEL	CSV	Point locations (stations)	Daily, 1981-2024	SWE, Precipitation, Temperature, Soil Moisture	Natural Resources Conservation Service
PRISM	NetCDF	4 km	Daily, 1984-2024	Precipitation, Maximum Temperature, Minimum Temperature	PRISM Climate Group

NASA PODAAC website: https://podaac.jpl.nasa.gov/dataset/TELLUS_GRAC-GRFO_MASCON_CRI_GRID_RL06.3_V4

U.S. Bureau of Reclamation website: https://www.usbr.gov/lc/region/g4000/NaturalFlow/current.html

NASA AppEEARS platform: https://appeears.earthdatacloud.nasa.gov/

National Snow and Ice Data Center: https://nsidc.org/data/g02158

Natural Resources Conservation Service: https://www.nrcs.usda.gov/wps/portal/wcc/home/snowClimateMonitoring/snowmonitoring/

PRISM Climate Group: https://prism.oregonstate.edu/

APPENDIX C

DATA PROCESSING SCRIPTS

C.1 Processing PRISM data

A. Data Acquisition:

PRISM daily precipitation, maximum temperature, and minimum temperature data were downloaded from the Oregon State University PRISM Climate Group data server. Downloads were performed using automated scripts to ensure consistent data handling. Example terminal command for precipitation data:

```
wget -r -np -nH --cut-dirs=3 --reject "index.html*"
ftp://ftp.prism.oregonstate.edu/daily/ppt/2024
```

B. Data Extraction:

The downloaded data files were provided in compressed format and required extraction using the following Python code:

```
zip_dir = 'tmin'
output_dir = 'tmin'
os.makedirs(output_dir, exist_ok=True)

# Loop through all files in the zip directory
for filename in os.listdir(zip_dir):
    if filename.endswith('.zip'):
        zip_path = os.path.join(zip_dir, filename)
        # Open and extract each zip file
        with zipfile.ZipFile(zip_path, 'r') as zip_ref:
            zip_ref.extractall(output_dir)
```

C. File Format Conversion

PRISM data are provided in binary raster format (.bil), which required conversion to NetCDF format for compatibility with the VIC model. This conversion was performed using R:

```
library(raster)
library(ncdf4)

output_dir <- "/Users/sghimi14/ASU Dropbox/Swastik
Ghimire/Mac/Documents/Research/Transcend/PRISM/2024/ppt_nc"

input_dir <- file.path("/Users/sghimi14/ASU Dropbox/Swastik
Ghimire/Mac/Documents/Research/Transcend/PRISM/2024/ppt/")
setwd(input_dir)

f <- list.files(recursive = FALSE, full.names = FALSE, pattern = "\\.bil$")

for (i in 1:length(f)){

  rd <- raster(f[i])
  outputFile <- file.path(output_dir, paste(f[i], ".nc", sep=""))
  writeRaster(rd, filename=outputFile, format="CDF", overwrite=TRUE)
}
```

D. Temporal Dimension Addition

The NetCDF files created in the previous step lacked a time dimension necessary for VIC modeling. This was addressed using Python with the xarray library:

```
import xarray as xr
import geopandas as gpd
import os
from datetime import datetime
import pandas as pd

data_dir = 'tmin_nc/'
data_out = 'tmin_nc_with_time/'

if not os.path.exists(data_out):
   os.makedirs(data_out)

files = sorted(os.listdir(data_dir))
files = [file for file in files if file.endswith('.nc')]

for file in files:
   variable = file.split('.')[0].split('_')[1]
   var_name = file.split('.')[0]

   date = file.split('.')[0].split('_')[-2]
   date_nc = variable + '_'+  date + '.nc'
   year = int(date[:4])
   month = int(date[-4:-2])
   day = int (date[-2:])
   date_f = datetime(year, month, day)
   date_str = date_f.strftime('%Y-%m-%d')
   date_time_obj = pd.to_datetime(date_str)


   xr_file = xr.open_dataset(f'{data_dir}/{file}')
   xr_file = xr_file.expand_dims(dim = 'time').assign_coords({'time' :
[date_time_obj]})
   xr_file = xr_file.drop_vars('crs')

   xr_file = xr_file.rename({var_name: variable})
   file_path = os.path.join(data_out, date_nc)

   xr_file.to_netcdf(file_path)
```

E. Temporal Aggregation

Finally, daily files were combined into continuous time series using the Climate Data Operators (CDO) software package. The following command was applied to each meteorological variable:

```
cdo mergetime tmin_nc_with_time/*.nc tmin_combined.nc
```

C.2 Processing SNOTEL data

The 'snotelr' package in R provides a streamlined and easy interface for accessing the SNOTEL database. This method was preferred over manual downloads from the NRCS website due to its reproducibility and capacity for bulk processing multiple stations simultaneously.

```
install.packages("snotelr")
library(snotelr)
stations <- c(308,310)
snotelr::snotel_download(stations, path = 'Documents/Research/SNOTEL')
```

This initial step creates a file called snotel_data.csv in the specified path. The resulting dataset contains measurements for all requested stations, including snow water equivalent, precipitation, temperature, and other available variables. Additional processing was performed using Python to restructure the data for analysis:

```
import pandas as pd
import numpy as np

df = pd.read_csv('snotel_data.csv')
df = df.replace('NA', '')

pivoted = pd.pivot_table(df, values='snow_water_equivalent', index='date',
columns='site_id')

pivoted = pivoted.fillna('')

pivoted.to_csv('snotel_pivoted.csv')
```

C.3 Processing SNODAS data

SNODAS files are in a netcdf format, so processing the data is pretty straightforward using Python's standard netcdf processing package ‘xarray’. There is a minor issue with a few missing days in the downloaded dataset. The following code helps manage that error.

```
import xarray as xr
snodas_data = xr.open_mfdataset('*/nc')

SWE = snodas_data.SWE
SWE.rio.write_crs('epsg:4326', inplace = True)

start_date = '2004-01-01'
end_date = '2023-08-06'
new_time_values = pd.date_range(start=start_date, end=end_date, freq='D')
snodas_dates = SWE['Time'].astype(str).values.tolist()
snodas_dates = pd.to_datetime(snodas_dates)
missing_dates = set(new_time_values) - set(snodas_dates)
new_time_values = new_time_values[~new_time_values.isin(missing_dates)]
snodas_clipped = SWE.rio.clip(crb.geometry)
snodas_clipped['Time'] = new_time_values
```

C.4. Processing USBR naturalized streamflow

The naturalized streamflow data is in a excel workbook format, so it is necessary to identify the gage corresponding to the sub-basin while processing the data.

```
import pandas as pd
usbr = pd.read_excel('NaturalFlows1906-
2020_20221215.xlsx',sheet_name='TotalNaturalFlow',header= 1)

usbr = usbr[3:1385]
usbr.index = usbr['Corresponding USGS gauge number']
usbr.index.names = ['Date']
usbr = usbr.drop(['Corresponding USGS gauge number'], axis=1)

gage1 = usbr['09421500'] #Change this according to sub-basin
gage1 = gage1[(gage1.index >= '1980-01-01') & (gage1.index <= '2020-12-31')]
gage1= gage1 * 1233.48 #Conversion of acre-feet to cubic meters
gage1 = gage1.resample('Y').sum()*1e-9
```

C.5. Processing SMAP data

SMAP data is in the netcdf format. However, the date is in Julian format so it needs to be converted for ease of processing. The code below shows that steps for conversion and loading for further analysis.

```
import xarray as xr
from datetime import datetime  # Import the datetime class



def convert_to_datetime(julian_date):
    # Extract year, month, and day attributes
    year = julian_date.year
    month = julian_date.month
    day = julian_date.day

    # Create a Python datetime object
    return datetime(year, month, day)

smap_surface = xr.open_mfdataset('SPL4SMGP.007_9km_aid0001_surface.nc')
smap1 = smap_surface.mean('day_period')

convert_to_datetime_vec = np.vectorize(convert_to_datetime)
smap1['time'] = convert_to_datetime_vec(smap1['time'])
smap1.rio.write_crs('epsg:4326', inplace = True)

surface = smap1.Geophysical_Data_sm_surface
```

C.6. Processing GRACE data

GRACE data is also in netcdf format. However, the coordinates span from 0 to 360 which makes it tricky when clipping using a shapefile that uses a geographic projection system. The conversion is done as shown below. Finally, the dataset is scaled using gain factors and is now ready for further analysis.

```
import xarray as xr
import geopandas as gpd
import matplotlib.pyplot as plt

# Open your dataset
df =
xr.open_dataset('GRCTellus.JPL.200204_202409.GLO.RL06.3M.MSCNv04CRI.nc')

# Adjust longitudes as needed
df['lon'] = ((df['lon'] + 180) % 360) - 180
df = df.sortby(df['lon'])

# Open scaling factors
scaling = xr.open_dataset('CLM4.SCALE_FACTOR.JPL.MSCNv03CRI.nc')
scaling.rio.set_crs('epsg:4326', inplace=True)

# Adjust longitudes and rename dimensions
scaling['lon'] = ((scaling['lon'] + 180) % 360) - 180
scaling = scaling.sortby(scaling['lon'])
scaling = scaling.rename({'lon': 'x', 'lat': 'y'})

# Process the anomalies data, setting CRS and renaming dimensions
anomalies = df.lwe_thickness
anomalies.rio.set_crs('epsg:4326', inplace=True)
anomalies = anomalies.rename({'lon': 'x', 'lat': 'y'})

# Calculate scaled anomalies
scaled_anomalies = anomalies * scaling
```

APPENDIX D

VIC SIMULATIONS

Table D1. Datasets/files used in running VIC, along with the location and description

File	Location	Description
Climate data (PRISM P,T and NLDAS wind speed)	/data/evivoni/forcings/PRISM_1982_2023/	Contains daily data from 1981-2024 (Sept). Files are aggregated monthly. These would be the input files to METSIM.
VIC forcings	/data/evivoni/sghimire/METSIM/outputs/	Contains hourly forcing data for the CRB obtained after running METSIM. These would be the input files to VIC, to be specified in the global file
Initial State (annual)	/data/evivoni/sghimire/VIC_calibrated/initial_files/annual_states	Contains initial file locations to be specified in the global file. These would be useful to start a simulation of a year from January 1 of any year.
Initial State (water year)	/data/evivoni/sghimire/VIC_calibrated/initial_files/water_year_states	Contains initial file locations to be specified in the global file. These would be useful to start a simulation of a year from October 1 of any year.
Soil and vegetation parameters	/data/evivoni/sghimire/VIC_calibrated/ para.Apr23_Calibrated.nc	Contains the calibrated VIC parameters. Location needs to be specified in the global file.
Domain	/data/evivoni/sghimire/VIC_calibrated/domain.CRB.0.0625_deg.nc	Contains domain information. Location needs to be specified in the global file.
Global	/data/evivoni/sghimire/VIC_calibrated/global_sample.txt	Contains parameters and locations of key inputs needs to run the model simulation.
Sbatch script	/data/evivoni/sghimire/VIC_calibrated/run_cmd_sample.sh	Contains a sbatch script to call and the global file in the Sol supercomputer
VIC Outputs	/data/evivoni/sghimire/VIC_calibrated/outputs	Contains all the VIC outputs using the calibrated parameters (1984-2024)

VIC requires different forcings, parameters, initialization, domain, global, and sbatch files to run. The location and description of these files are presented in Table D1. Following are the sample global and sbatch file. The bold text are the values that the user may need to change for running a simulation with adjustments (years, initialization with respect to start year, output location) as required. If the user requires other outputs other than what is listed, refer to https://vic.readthedocs.io/en/master/Documentation/OutputVarList/. It also contains information on the units of all the output files.

Global file:

Simulation Parameters
##
MODEL_STEPS_PER_DAY 24 # number of model time steps per day (set to 1 if FULL_ENERGY = FALSE, set to > 1 if FULL_ENERGY = TRUE)
SNOW_STEPS_PER_DAY 24 # number of time steps per day for which to solve the snow model (should = MODEL_STEPS_PER_DAY if MODEL_STEPS_PER_DAY > 1)
RUNOFF_STEPS_PER_DAY 24 # time step in hours for which to solve the runoff model (should be >= MODEL_STEPS_PER_DAY)
STARTYEAR **1982** # year model simulation starts
STARTMONTH **01** # month model simulation starts
STARTDAY **01** # day model simulation starts
ENDYEAR **2024** # year model simulation ends
ENDMONTH **09** # month model simulation ends
ENDDAY **30** # day model simulation ends
CALENDAR STANDARD

NODES 3 # number of soil thermal nodes
FULL_ENERGY TRUE # TRUE = calculate full energy balance; FALSE = compute water balance only
FROZEN_SOIL FALSE # TRUE = calculate frozen soils
AERO_RESIST_CANSNOW AR_406
SOIL_RARC 0
CLOSE_ENERGY FALSE
GRND_FLUX_TYPE GF_410
DEEP_ESOIL TRUE # TRUE = deep bare soil evaporation; FALSE = default depth of bare soil evaporation

```
#IRRIGATION TRUE
#CROPSPLIT TRUE
#IRR_FREE TRUE

DOMAIN        /scratch/sghimi14/domain/domain.CRB.0.0625_deg.nc
DOMAIN_TYPE    LAT     lat
DOMAIN_TYPE    LON     lon
DOMAIN_TYPE    MASK    mask
DOMAIN_TYPE    AREA    area
DOMAIN_TYPE    FRAC    frac
DOMAIN_TYPE    YDIM    lat
DOMAIN_TYPE    XDIM    lon

#######################################################################
# Forcing Files
#######################################################################
FORCING1            /data/evivoni/sghimire/METSIM/outputs/CRB_PRISM_  # Forcing
file path and prefix, ending in "_"
FORCE_TYPE          PREC     prec
FORCE_TYPE          AIR_TEMP temp
FORCE_TYPE          SWDOWN   shortwave
FORCE_TYPE          LWDOWN   longwave
FORCE_TYPE          PRESSURE air_pressure
FORCE_TYPE          VP       vapor_pressure
FORCE_TYPE          WIND     wind
WIND_H             10.0  # height of wind speed measurement (m)

#######################################################################
# Land Surface Parameters
#######################################################################
PARAMETERS      /data/evivoni/sghimire/VIC_calibrated/ para.Apr23_Calibrated.nc
NODES          3
SNOW_BAND       TRUE
#NEW_SNOW_ALB_SUPPLIED TRUE
#VEGLIB_IPRM     TRUE
#VEGPARAM_CSPFLG TRUE
#VEGPARAM_IFLAG  TRUE
#VEGPARAM_FCROP  TRUE
#VEGPARAM_FIRR   TRUE
LAI_SRC        FROM_VEGPARAM
FCAN_SRC        FROM_VEGPARAM
ALB_SRC        FROM_VEGPARAM
#FIMP_SRC       FROM_VEGPARAM
#FCROP_SRC      FROM_VEGPARAM
```

#FIRR_SRC FROM_VEGPARAM
SNOW_ALB_ACCUM_A 0.98 #def 0.98
SNOW_ALB_ACCUM_B 0.58 #def 0.58
SNOW_ALB_THAW_A 0.92 #def 0.92
SNOW_ALB_THAW_B 0.85 #def 0.85

###
Initial State
###
INIT_STATE
/data/evivoni/sghimire/VIC_calibrated/initial_files/annual_states/state.livneh.19820101_00000.nc

###
Save State
###
STATENAME **/scratch/sghimi14/state/state.Apr23_Calibrated_d2_G2_glc1_Ds0.4**
STATEYEAR **2024**
STATEMONTH **10**
STATEDAY **01**
STATESEC **00000**

###
Output Files and Parameters
###
RESULT_DIR
/scratch/sghimi14/output/Mar7/Apr23_Calibrated_d2_G2_glc1_Ds0.4
OUTFILE **CRB_PRISM_Apr23_Calibrated_d2_G2_glc1_Ds0.4**
COMPRESS FALSE
OUT_FORMAT NETCDF4
AGGFREQ NDAYS 1
HISTFREQ NYEARS 1
OUTVAR OUT_PREC
OUTVAR OUT_RAINF
OUTVAR OUT_EVAP
OUTVAR OUT_EVAP_CANOP
OUTVAR OUT_TRANSP_VEG
OUTVAR OUT_EVAP_BARE
OUTVAR OUT_RUNOFF
OUTVAR OUT_BASEFLOW
OUTVAR OUT_SWE
OUTVAR OUT_SOIL_MOIST
OUTVAR OUT_SNOW_MELT
OUTVAR OUT_BARESOILT

```
OUTVAR     OUT_AIR_TEMP
OUTVAR     OUT_SURF_TEMP
OUTVAR     OUT_SOIL_TEMP
OUTVAR     OUT_VEGT
OUTVAR     OUT_RAD_TEMP
OUTVAR     OUT_SNOW_SURF_TEMP
OUTVAR     OUT_SNOW_PACK_TEMP
OUTVAR     OUT_LATENT
```

Sbatch file:

```
#!/bin/bash

#SBATCH -N 1-2
#SBATCH -n 50          # number of cores
#SBATCH -t 16:00:00  # wall time
#SBATCH -q public #gmascaro  #gmascaro
#SBATCH -p general
##SBATCH --mem 128GB
#SBATCH -o ./log/testrun.%j.out # output file name
#SBATCH -e ./log/testrun.%j.err # output file name in which %j is replaced by the job ID
#SBATCH --mail-type=FAIL
#SBATCH --mail-user=sghimi14@asu.edu


# Script to run
module purge
module load gcc-11.2.0-gcc-11.2.0
module load openmpi/4.1.5
module load netcdf-c-4.8.1-gcc-11.2.0

vic_dir='/data/evivoni/tools/VIC-ASU-2.0'

prefix='Apr23_Calibrated_d2_G2_glc1_Ds0.4'
global_file='Apr23_Calibrated_d2_G2_glc1_Ds0.4.txt'
export MV2_ENABLE_AFFINITY=0

srun --mpi=pmix $vic_dir/vic/drivers/image/vic_image_sol_upd.exe -g
/scratch/sghimi14/global/Mar7/$global_file >& /scratch/sghimi14/log/log.$global_file
```

APPENDIX E

EVALUATION OF HUC10 WATERSHEDS

This analysis compares SNOTEL *SWE* measurements with corresponding HUC10 watersheds and collocated VIC pixels to determine appropriate boundaries for further analysis. For reference, the area of a VIC pixel is approximately 36 km^2, while the average area of a HUC10 watershed in the CRB is approximately 600 km^2. The methodology for comparing SNOTEL stations with their VIC pixel involved the following steps:

a. Time series of the 104 stations were extracted from VIC using geographic locations of the SNOTEL stations stored as points in a shapefile.
b. April 1 values were filtered out for all pixels.
c. April 1 mean for baseline (1991-2020) was evaluated for each pixel.
d. Relative values with respect to baseline were evaluated. To be clear, now each of the 104 pixels had a Relative April 1 *SWE* time series.
e. The median for each year was taken across all pixels.
f. The same procedure (b-e) was repeated for observation values, and the values were compared through a scatter plot.

The analysis was subsequently conducted using selected HUC stations, utilizing watershed boundaries for both axes rather than comparing points with pixels.

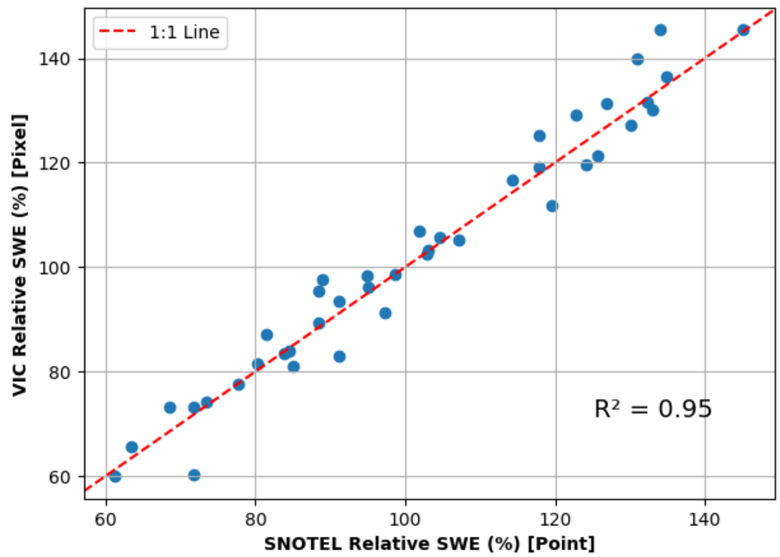


Figure E1. Scatter plot showing comparison of SNOTEL point data with corresponding VIC pixel

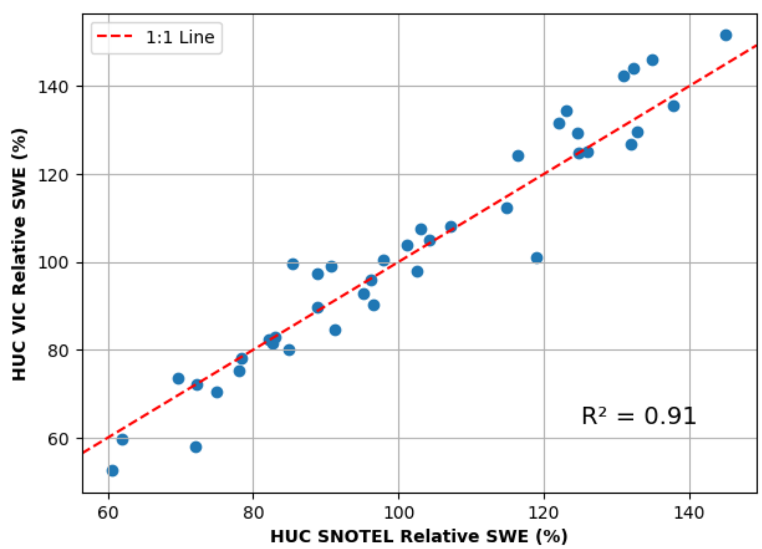


Figure E2. Scatter plot showing comparison of SNOTEL points within a HUC10 watershed with corresponding HUC10 value from VIC

Although the R^2 metric decreased, this is acceptable considering the area covered by HUC watersheds is approximately 20 times larger than individual VIC pixels. The HUC10 boundaries were selected over VIC pixels for two primary reasons::

i) HUC10 boundaries represent a widely recognized standard, making this approach more accessible and comprehensible to other researchers compared to using model-specific pixels.

ii) Larger HUC10 boundaries capture broader regional variability while representing the point SNOTEL stations well, whereas small areas like individual VIC pixels may not fully represent the overall hydrologic context.

The spatial patterns of biases in both comparisons were mapped to evaluate representation quality. The difference between Relative April 1 *SWE* from VIC and that from SNOTEL during the baseline period was calculated, and the median across all years was determined. Each point or watershed, therefore, represents the median difference in peak *SWE* between VIC and SNOTEL over the 30-year period.

Both analyses exhibit negative biases in certain points/watersheds, but the overall median bias approaches zero when averaged across spatial and temporal domains. VIC pixels demonstrate marginally better performance than HUC10 boundaries. However, HUC10 watersheds maintain adequate performance despite covering an area approximately 20 times larger, which justifies their selection for this analysis.

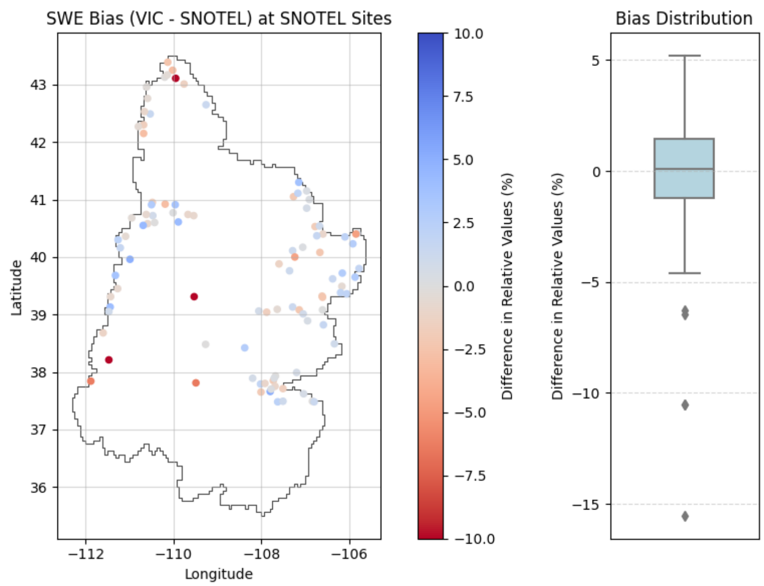


Figure E3. Biases between SNOTEL (point) and corresponding VIC grid cells

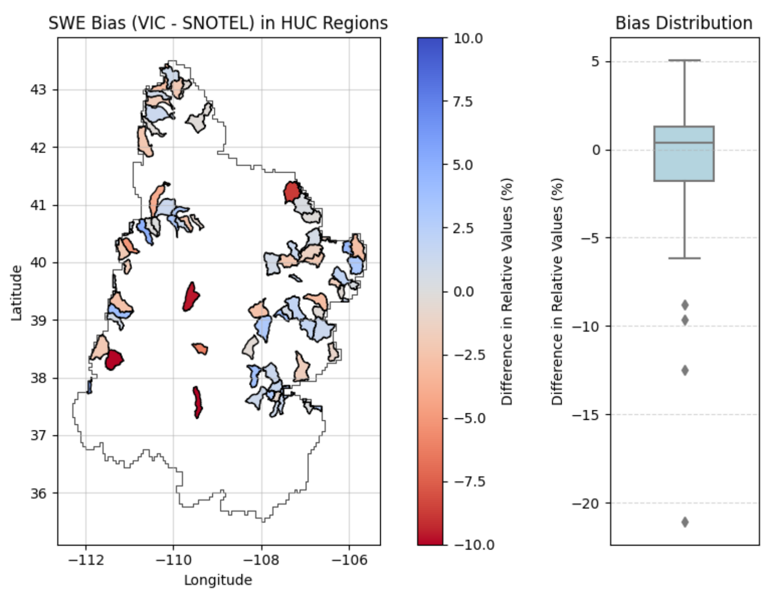


Figure E4. Biases between SNOTEL (point) and corresponding HUC10 watershed

APPENDIX F

SNOW CALIBRATION RESULTS

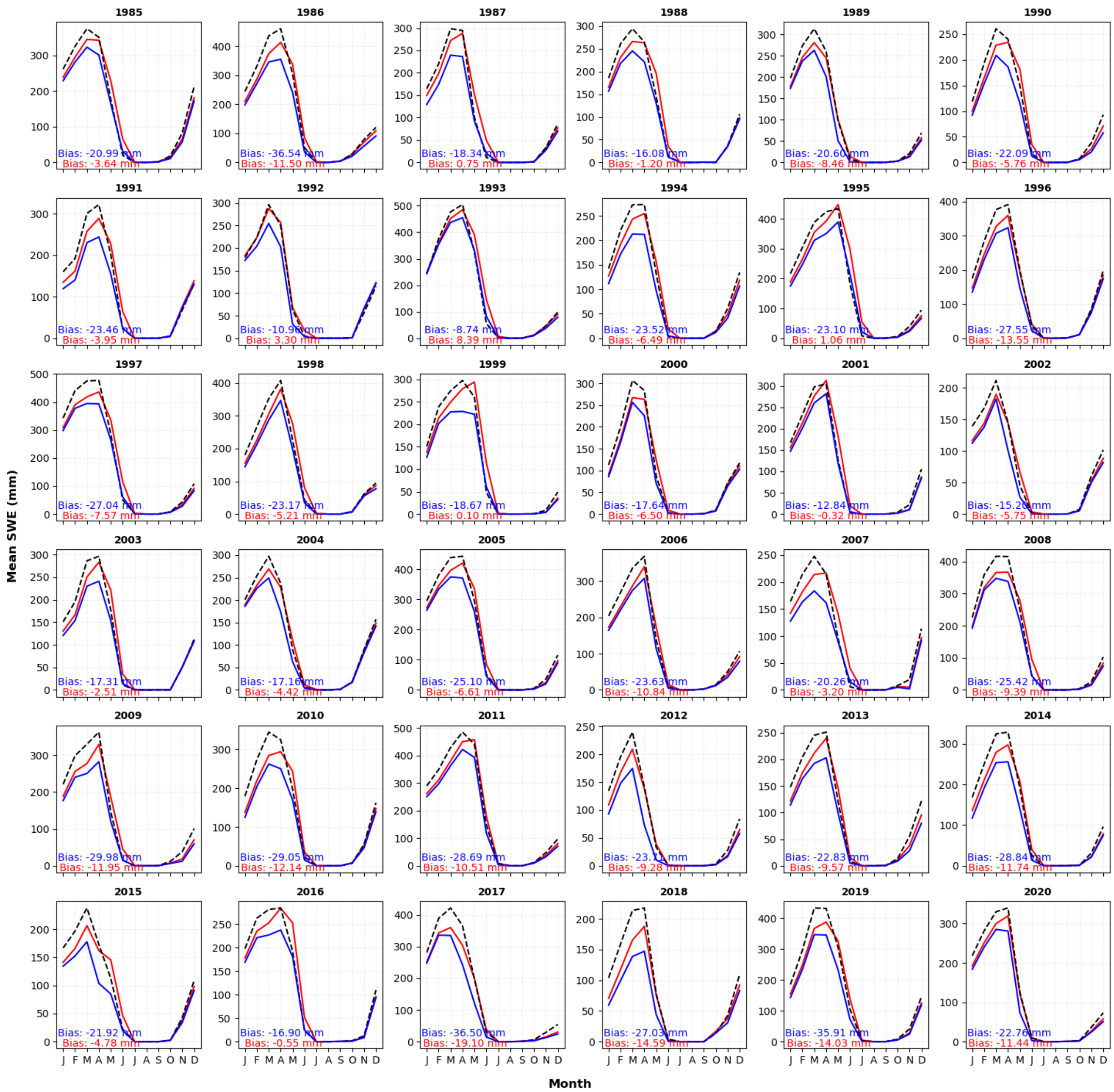


Figure F1. Comparison of daily *SWE* across SNOTEL observations, and VIC baseline and calibrated versions for all WY, along with biases for peak SWE. Dotted black line represents SNOTEL values, blue line represents baseline (uncalibrated) VIC outputs and red line represents calibrated VIC *SWE* outputs. Values are averaged across stations (SNOTEL) or corresponding pixels (VIC)

APPENDIX G

VIC OUTPUTS

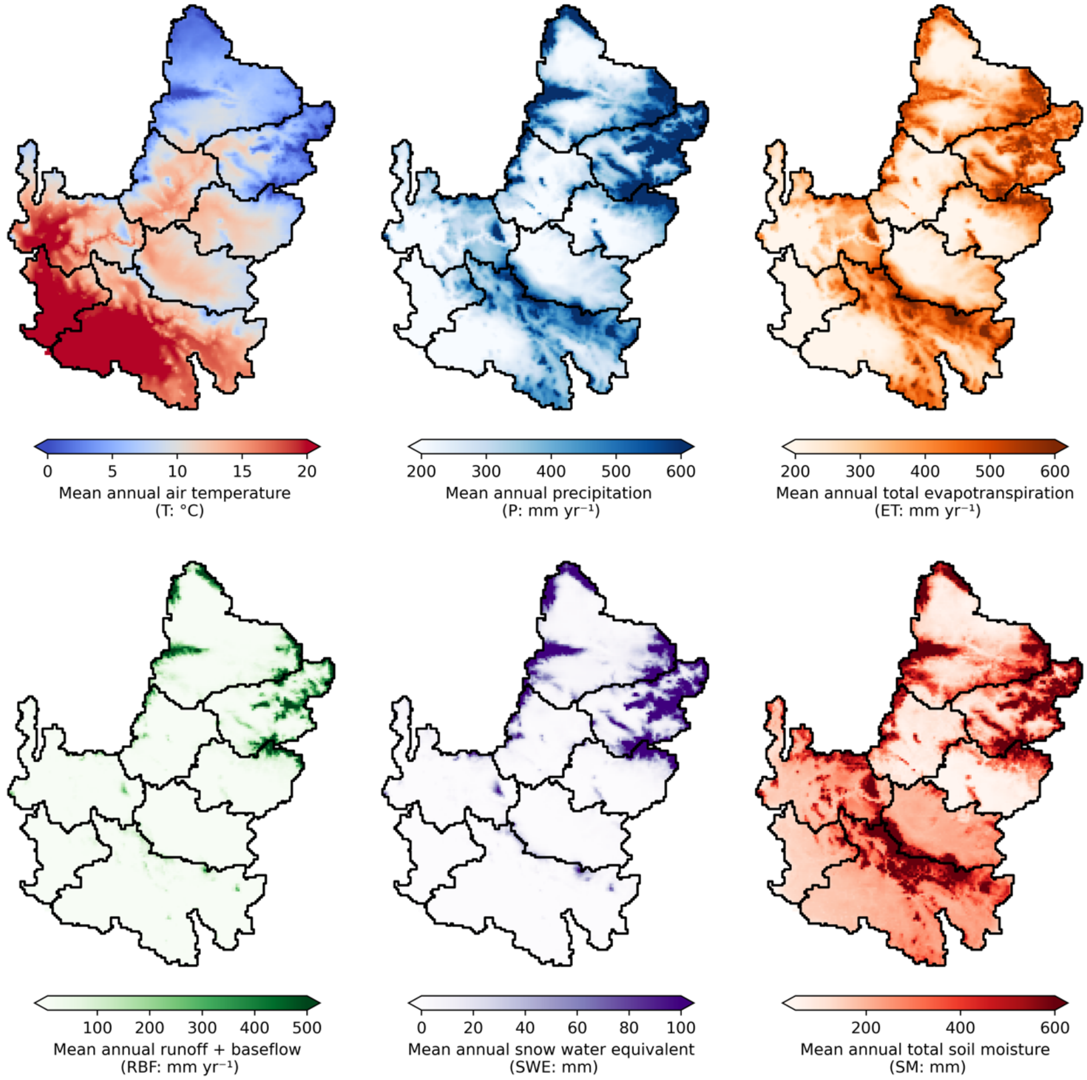


Figure G1. Spatial means of the annual values observed across variables in the baseline period (1984-2023)

Table G1. Spatially averaged (1984-2023) mean annual values for variables across sub-basins.

Basin/ **Variable**	**P(mm /yr)**	**SWE (mm)**	**Melt (mm/yr)**	**ET (mm/yr)**	**SM (mm)**	**R (mm/yr)**	**BF (mm/yr)**
CRB	348.7	17.1	148.9	316.4	274.9	18.8	17.8
Upper Basin	398.1	34	216.5	332.5	260.7	36.7	32.8
Lower Basin	306.2	2.6	90.6	302.6	287.2	3.3	4.9
Green	402.7	37.9	228.4	334.7	285.9	40.2	32
Upper Colorado	542.5	61.5	336.4	418	390.3	56.1	73.8
MiddleColo	290.1	3.7	113.9	286.3	283.6	2.7	5
San Juan	345.4	18.9	152.1	300.9	161.6	36.2	10.8
Glen Canyon	277.9	10.4	122.4	262.8	164.8	8	10.6
Little Colorado	291.6	3	102.9	290.8	282.7	1.1	3.6
Grand Canyon	288.9	4.4	122.7	282.7	284.3	4	6.2
Lower Colorado	197.9	0.1	39.7	195	173.3	4.8	1
Gila	352.3	2.1	81.8	348.7	322.5	3.5	5.8

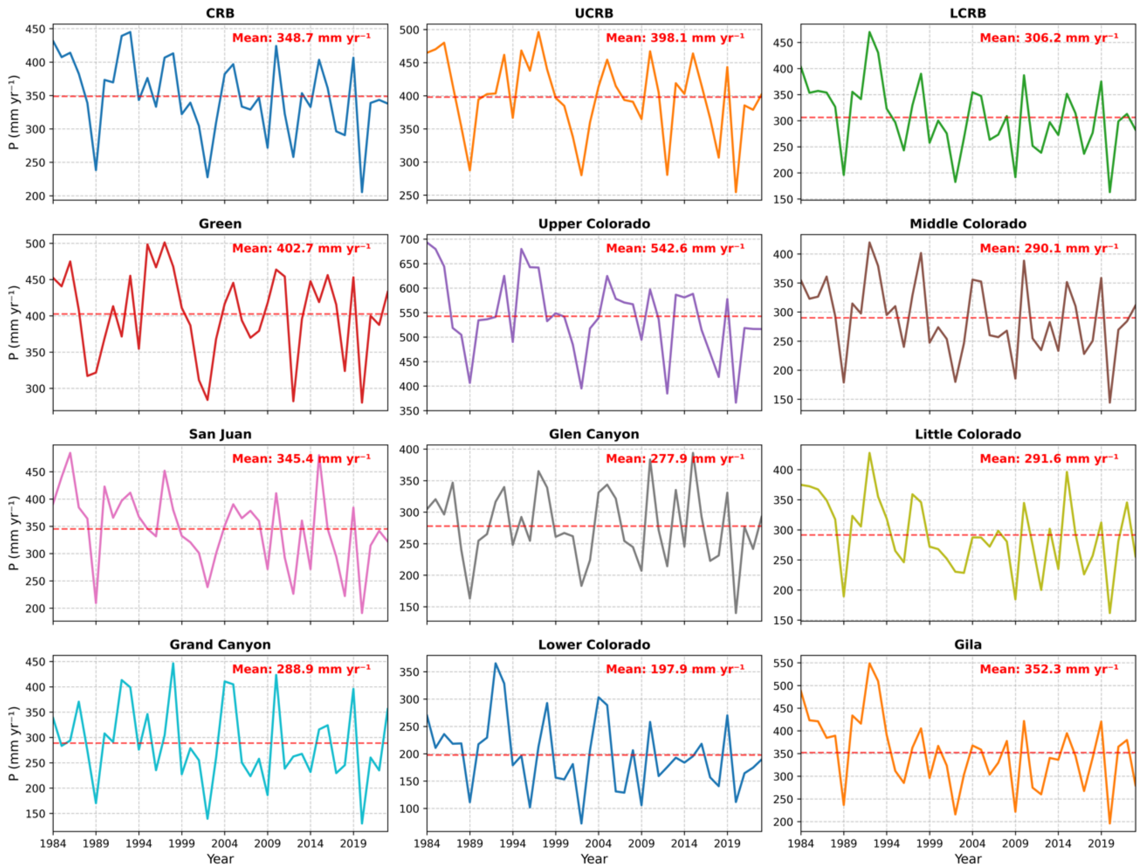


Figure G2. Time series of the annual precipitation (mm/year) values averaged across major sub-basins in the baseline period (1984-2023)

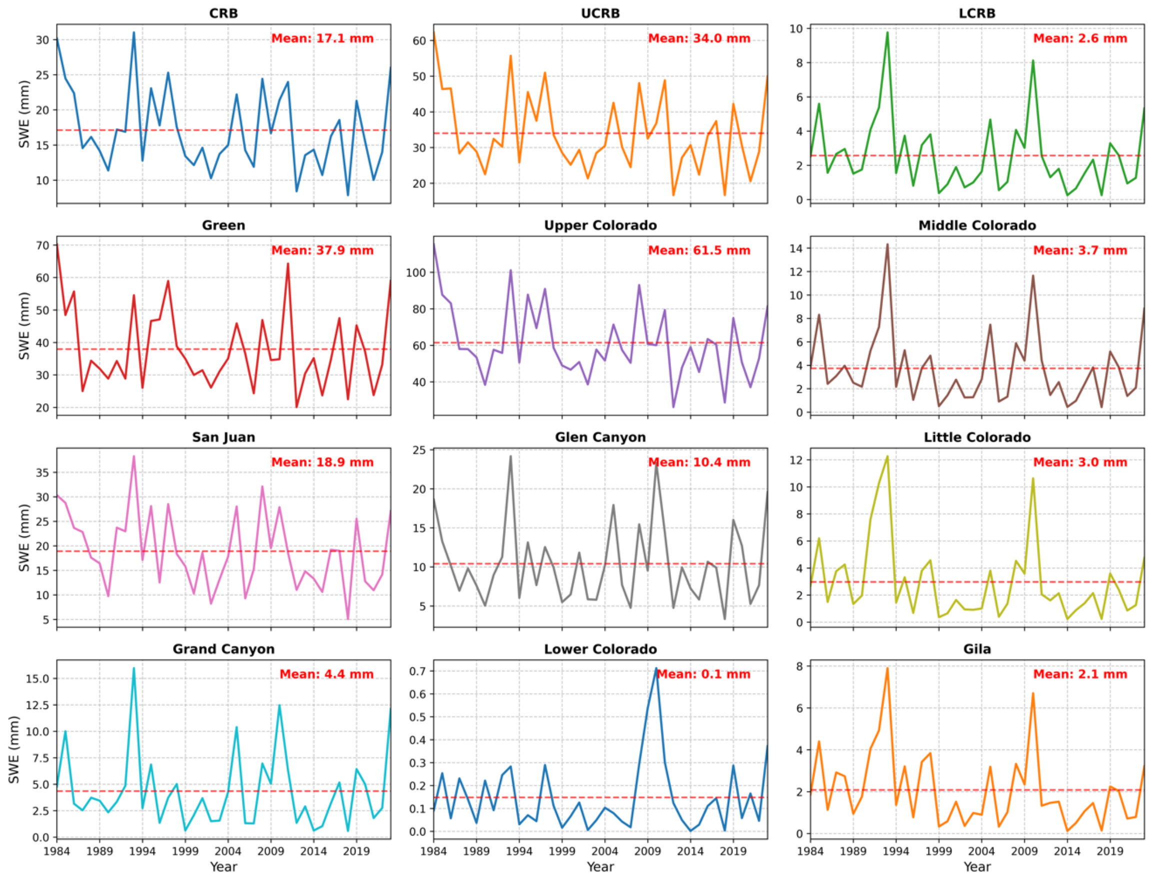


Figure G3. Time series of the annual mean *SWE* (mm) values averaged across major sub-basins in the baseline period (1984-2023)

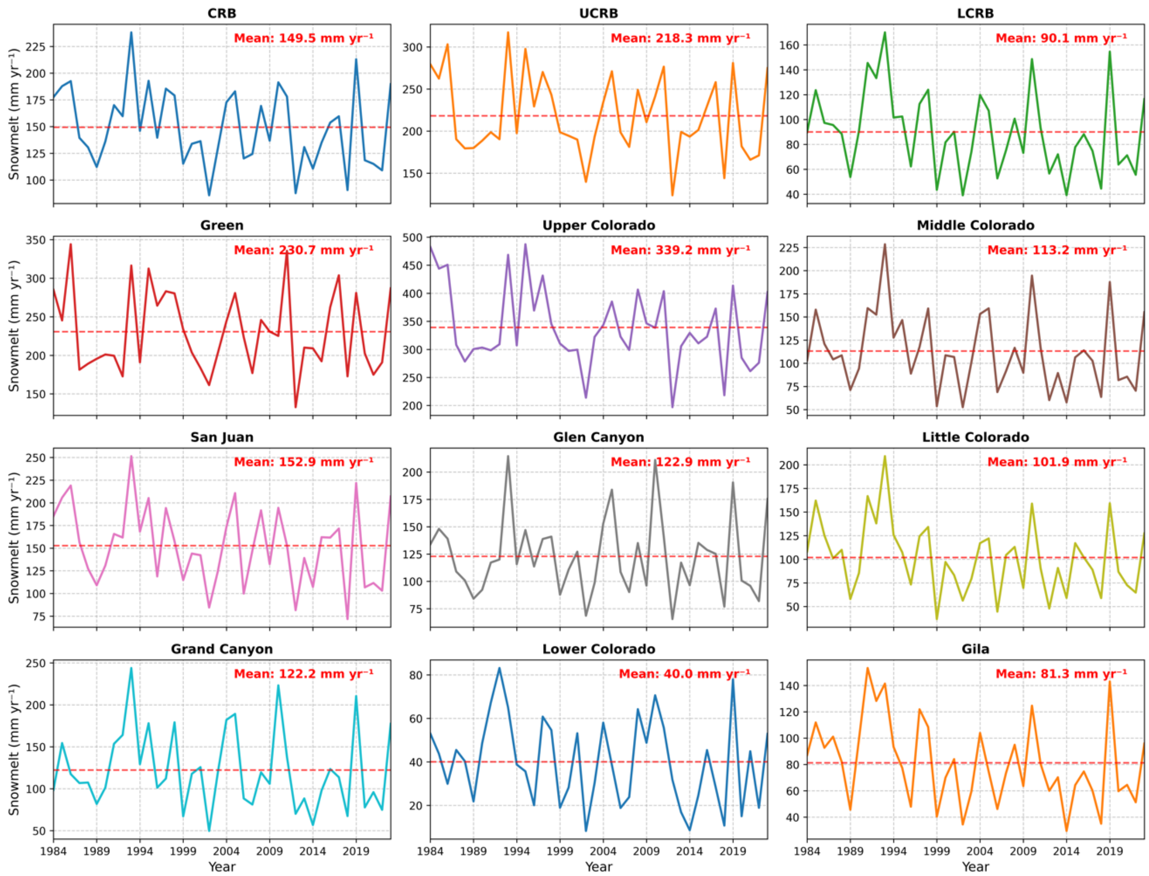


Figure G4. Time series of the annual snowmelt (mm/year) values averaged across major sub-basins in the baseline period (1984-2023)

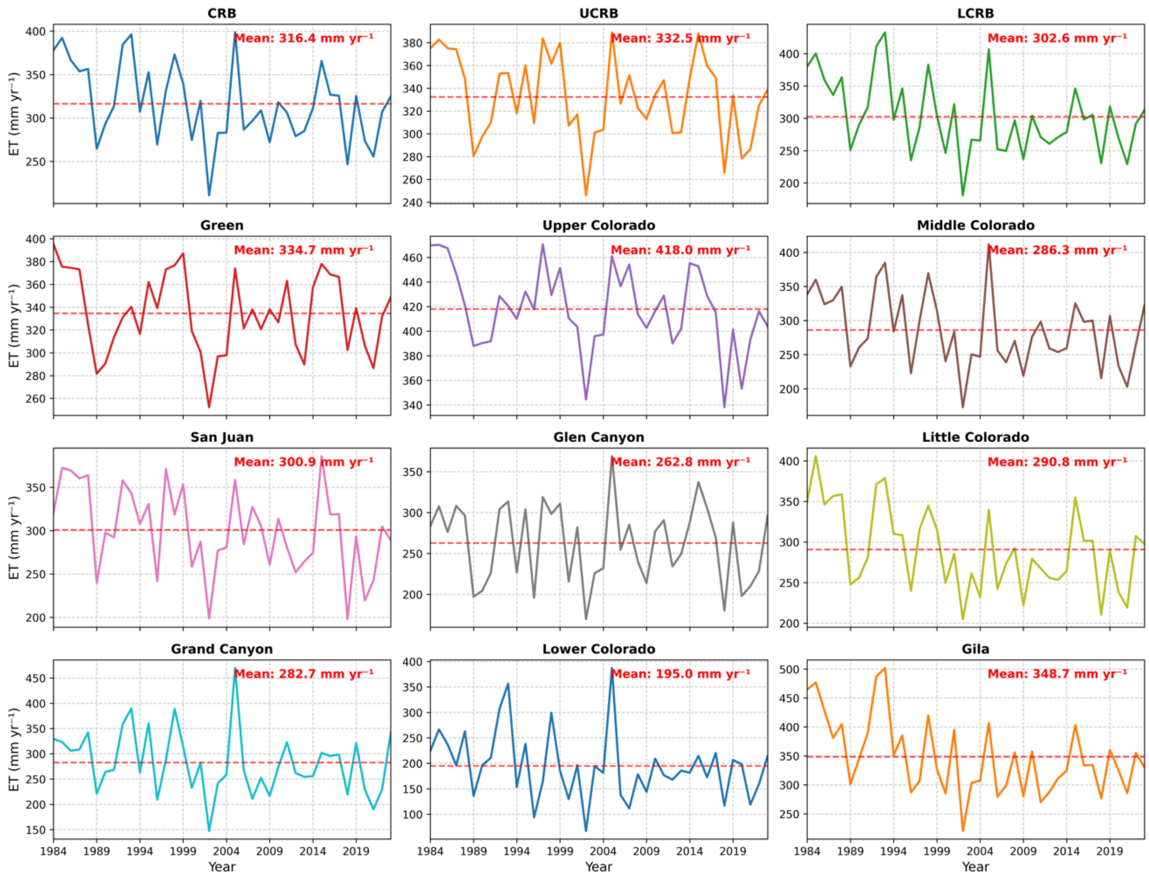


Figure G5. Time series of the annual *ET* (mm/year) values averaged across major sub-basins in the baseline period (1984-2023)

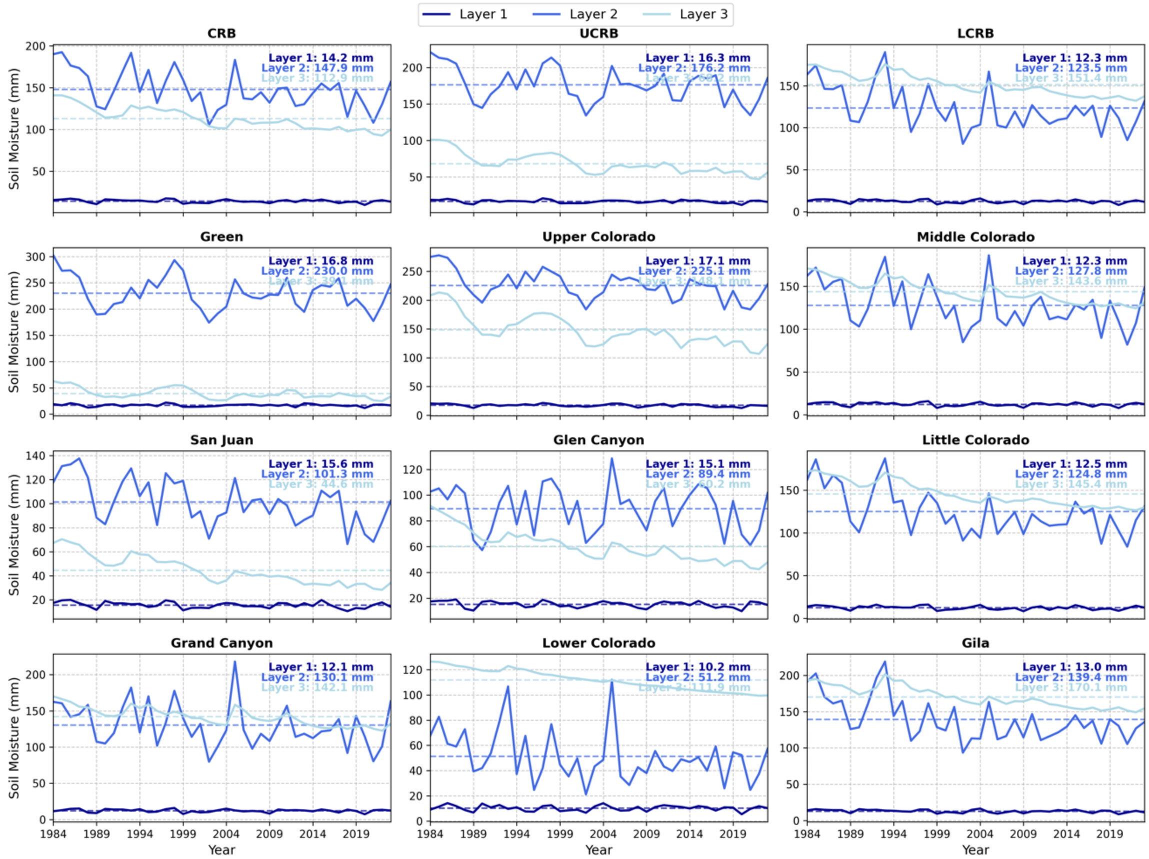


Figure G6. Time series of the annual mean *SM* (mm) values averaged across major sub-basins in the baseline period (1984-2023). Layer 1 represents surface layer, Layer 2 represents rootzone layer and Layer 3 represents deep layer. Values shown in subplots are mean values across the time period for the sub-basin.

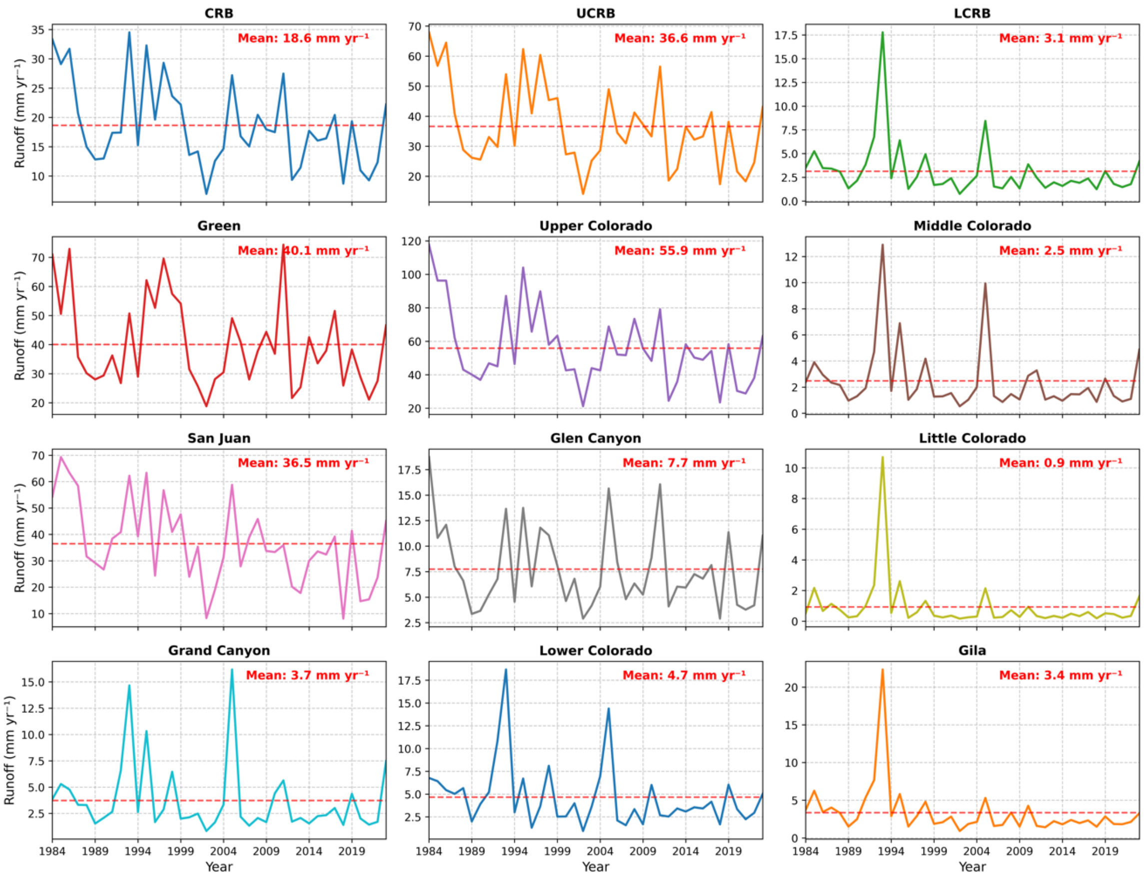


Figure G7. Time series of the annual runoff (mm/year) values averaged across major sub-basins in the baseline period (1984-2023)

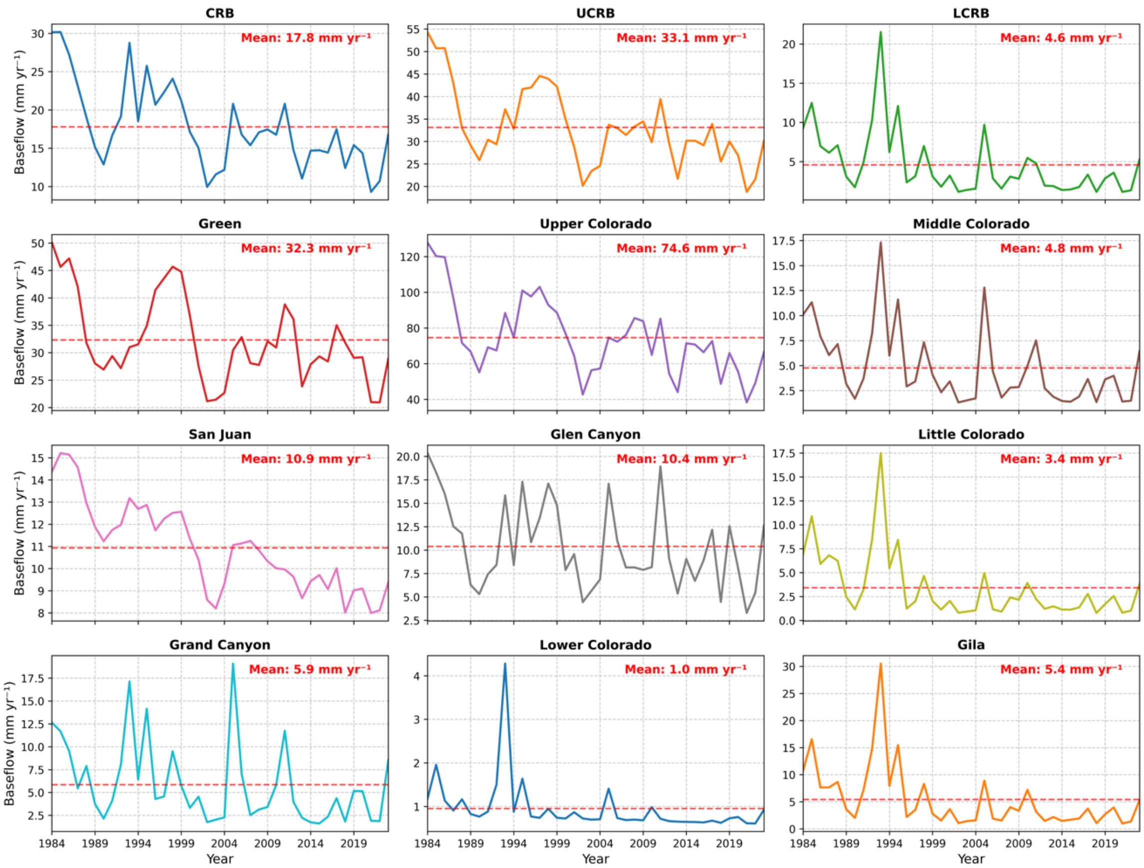


Figure G8. Time series of the annual baseflow (mm/year) values averaged across major sub-basins in the baseline period (1984-2023)

APPENDIX H

DISAGGREGATION OF GRACE COMPONENTS

Calibrated outputs from the VIC model were utilized to disaggregate the components of GRACE. Figure J1 illustrates the temporal variation of individual water storage components across the UCRB and LCRB from 2002 to 2024. The four panels represent anomalies (in km^3) for *SWE*, *SM*, surface water (*SW*) from Lake Mead and Powell, and groundwater (*GW)*, with the latter calculated as the residual component after removing *SWE, SM*, and *SW* from the total GRACE signal. A 100 cm equivalent depth was used for the SM calculations by combining data from layers 1 and 2 of the VIC model. This approach was followed to maintain a consistent depth of soil layer and follows an approach from other studies that use *SM* from products like GLDAS (Castle et al., 2014). The *SW* component was estimated using water level changes in Lakes Mead and Powell, with Lake Powell changes distributed across the Upper Basin and Lake Mead changes across the Lower Basin. This methodology is consistent with previous studies analyzing GRACE data in the CRB (Castle et al., 2014).

The *SWE* and *SM* components derived from the VIC model show characteristic seasonal patterns, with SWE exhibiting pronounced winter peaks primarily in the UCRB. SW storage, represented by gauged volumes of Lakes Mead and Powell, shows a long-term declining trend, particularly after 2010, with both regions showing substantial negative anomalies by 2020-2023. The GW component, obtained as a residual by subtracting modeled snow, *SM*, and measured surface water from the GRACE total signal, reveals interesting dynamics, particularly in the post-2020 period. During this recent period, the UCRB has shown notably strong positive groundwater anomalies, with several peaks exceeding 20 km^3, therefore suggesting subsurface recovery. This is supported by a recent study (Scanlon et al., 2025), which shows near-stable/increasing groundwater trends in the UCRB (2002-2024).

However, these positive signals might also represent error propagation from the residual calculation method, potentially due to a large decline in *SW* components. This limitation has been acknowledged in previous GRACE decomposition studies, where Castle et al. (2014) and Scanlon et al. (2015) discuss challenges in accurately separating groundwater from other water storage components. In contrast, the LCRB *GW* shows more variable but generally negative anomalies in the post-2020 period, aligning with findings from Scanlon et al. (2025) documenting continuing groundwater depletion in this region during recent drought conditions. The difference between UCRB and LCRB *GW* trends highlights regional variations in water management approaches, as suggested by several basin studies (Castle et al., 2014; Scanlon et al., 2025). However, additional research would be needed to quantify precise causes of these regional differences and their relationship to specific water management practices.

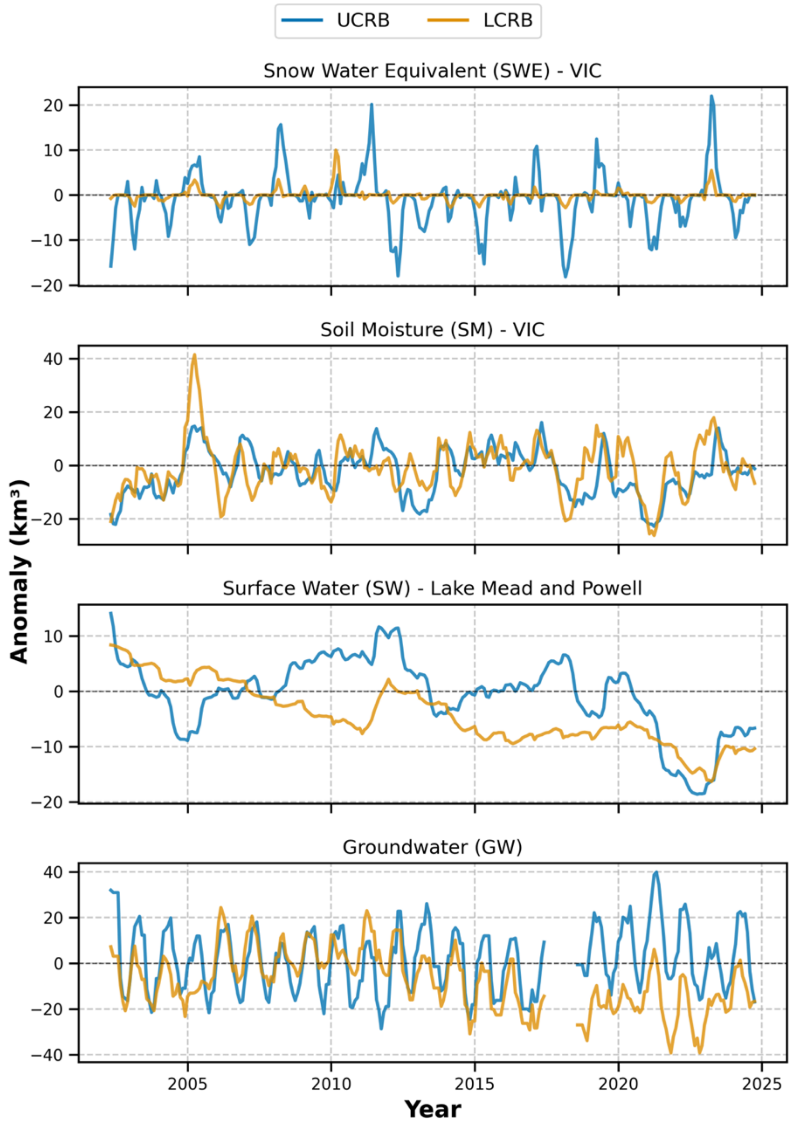


Figure H1. Time series showing trends of each component of GRACE: *SWE* and *SM* component from VIC, surface water component from Lake Mead and Powell storages, and groundwater by removing snow, *SM*, and surface water from the GRACE signal.