DISSERTATION

ANALYZING MOUNTAIN HYDROLOGY IN POST-WILDFIRE LANDSCAPES

Submitted by

Megan Sears

Department of Ecosystem Science and Sustainability

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2026

Doctoral committee:

Advisor: Stephanie Kampf

David Barnard
Dan McGrath
Peter Nelson

ABSTRACT

ANALYZING MOUNTAIN HYDROLOGY IN POST-WILDFIRE LANDSCAPES

Estimating water balance components in mountainous areas is challenging due to high spatial and temporal variability. Wildfires exacerbate this challenge and are increasing in size and severity while burning at higher elevations in the western U.S. The region is also expected to experience more extreme rainfall events due to climate change. This dissertation aims to examine rainfall spatial variability, suitability of gridded climate products in hydrologic applications, and catchment characteristics that could be influencing post-fire streamflow responses.

Chapter 2 investigates the influence of elevation on precipitation patterns across the Front Range in northern Colorado and evaluates a gridded radar-based rainfall product compared to gauge data. Flash floods in the Colorado Rocky Mountains are triggered by intense rainfall during convective storms, and the likelihood and severity of flash flooding is increased further following wildfires. The most extreme floods in this region have occurred at lower to middle elevations (~1500 to 2500 m), within a landscape that extends above 4000 m, indicating that rainfall intensity may vary with elevation; however, the variability has not been quantified relative to orographic position (i.e., east or west of mountain ranges). This chapter combines Multi-Radar/Multi-Sensor System (MRMS), Parameter-elevation Regressions on Independent Slopes Model (PRISM), Gridded Surface Meteorological (gridMET), and tipping bucket precipitation datasets to evaluate how rainfall intensities vary with elevation along Colorado's northern Front Range mountains. The MRMS, PRISM, and gridMET datasets showed greater total warm season (i.e., June-September) precipitation east of the Colorado's northern Front Range mountains than west, but the magnitudes of rainfall differed between datasets. The MRMS dataset had greater warm season rainfall than gridMET and PRISM east

of the mountains but lower amounts to the west. The relationship between elevation and rainfall intensity varied between the three warm seasons evaluated, but the highest rainfall intensities were consistently observed east of the mountains. A network of stream stage sensors installed within the Cameron Peak and East Troublesome fires was used to evaluate the ability of MRMS and tipping bucket rainfall intensities to predict stream occurrence in burned areas using a logistic regression. The MRMS rainfall intensities predicted stream occurrence better east of the mountains, while tipping buckets performed better to the west; this difference likely results from poor radar coverage observed west of the mountains. The MRMS dataset may outperform tipping buckets for streamflow predictions in areas with high radar coverage, likely due to its broader spatial coverage than most rain gauge networks.

Chapter 3 evaluates how streamflow responses to rainfall vary between fires on the east and west side of the Front Range and between snow zones and year post-fire. Stream gauges installed at 26 burned and unburned catchments in the Cameron Peak and East Troublesome fires were used to quantify streamflow response as stage rises and lags to peak, and the MRMS data were used to quantify the 60-minute maximum rainfall intensities for each storm. Between the two fires, stage rises in the East Troublesome Fire streams were higher than those in Cameron Peak Fire. The differences in stage rise between fires were associated with higher clay content in soils and more valleys and hollows in the East Troublesome Fire compared to Cameron Peak. Across both fires, catchments in the intermittent snow zone (i.e., areas that do not hold a consistent winter snowpack) experienced higher rainfall intensities and have a greater proportion of their area at elevations above the 75th percentile of the catchment's elevation range, leading to greater stage rises and shorter lags to peak than catchments in the seasonal snow zone. The timing of intermittent snow zone streamflow responses was consistent with infiltration excess overland flow, indicating that these areas are at high risk for flash flooding. The results highlight that post-fire streamflow responses have high variability even

within a given region and that catchment morphology is also important to consider in post-fire studies.

Chapter 4 examined how gridded and station climate data sources affect hydrologic model calibration, performance, and behavior in a Colorado Front Range catchment. A small portion of this watershed was burned a year prior to the start of the study and was calibrated independently from the unburned portion. A streamflow model was run with a combination of the PRISM and Daymet products as the gridded climate inputs and a network of sensors as the station climate inputs. The two climate inputs were evaluated prior to modeling, and differences were likely due to elevation biases from the gridded climate data algorithms. The differences in climate inputs led to distinct parameter calibrations between the two datasets, and of the calibrations, the station climate inputs yielded better hydrologic model performance for streamflow at the catchment outlet. The gridded inputs resulted in greater snow accumulation and earlier, faster snowmelt due to lower snow interception and higher snowmelt parameter values compared to the station climate inputs. The burned area had different parameter constraints, including reduced interception, reduced maximum summer infiltration, and increased maximum depression storage, to reflect changes post-fire. These different parameter constraints for the burned area led to greater snow accumulation, higher soil saturation, and increased interflow and shallow groundwater. Overall, these results demonstrate that hydrologic model parameter optimizations can change depending on the climate inputs. This highlights the importance of accurate input data and illustrates the importance of not only calibrating to streamflow but also internal state variables.

ACKNOWLEDGEMENTS

I am deeply grateful for all those who have supported me throughout my graduate school journey. First, I want to thank Stephanie Kampf for her invaluable guidance, mentorship, and the many opportunities she provided. Her dedication to her students and countless hours reviewing and helping shape this work into a cohesive document are sincerely appreciated. I also thank Dave Barnard for the opportunity to work at the USDA ARS, for securing funding, and for generously sharing his expertise in statistical analyses. Additional thanks go to Dan McGrath, Peter Nelson, and John Hammond for their insight and support throughout my research.

Thank you to the USDA ARS team I've had the pleasure of working with over the past several years: Dave Barnard, Tim Green, Holm Kipka, Rob Erskine, Nathan Lighthart, and Kyle Mankin. I'm grateful for their support with field work, navigating Ages, and synthesizing results.

Thank you to the many graduate student peers in Stephanie's lab who I have overlapped with over the years. A special thanks to everyone who has shared time with me in A115 and patiently listened to my many ramblings. I'm also grateful to Matt Ross for his guidance, and to everyone in the ROSSyndicate for the many hours spent keeping me accountable.

Lastly, thank you to my family for their unwavering support and understanding of my dedication to graduate school. I am especially grateful to my husband, Tyler, who has listened to my talk about my research for countless hours and stood by me every step of the way.

TABLE OF CONTENTS

CHAPTER 1. Introduction

1.1 Hydrologic uncertainty in complex terrain

Mountain regions play an essential role in sustaining water supplies for downstream areas (Viviroli et al., 2007). As climate change alters snowpack dynamics, precipitation patterns, and evapotranspiration rates, understanding hydrological processes in mountain systems is critical for predicting future water availability. These hydrological processes not only sustain ecological communities within the mountains but also provide critical water resources for agriculture and municipal use. The western U.S. mountains experience dramatic topographic and vegetation changes as well as steep gradients between temperature and precipitation, leading to hydrologic regime changes with elevation. This region also has high interannual variability in climate (Frankson et al., 2022), which leads to challenges predicting and understanding hydrological processes. In this region, climate and hydrologic variables have high spatial variability across mountain gradients (Bales et al., 2006; Gentine et al., 2012).

Climate variables (e.g., precipitation, air temperature) can be measured on the ground at the point scale, but they are difficult to capture over large areas due to spatial heterogeneity. Bridging the gap between point-scale measurements and regional-scale processes is a major research challenge (Gentine et al., 2012). As technology has advanced, ground-based observations have been integrated into statistical and physical models to produce gridded datasets, and remote sensing techniques have been used to detect meteorological conditions over larger areas. However, gridded datasets share uncertainties related to both temporal and spatial resolution. Ground-based data to evaluate gridded products may be sparse or unavailable, especially in remote areas. These factors introduce considerable uncertainty into hydrological models and projections, complicating both scientific understanding and management decisions.

1.2 Post-fire hydrologic impacts

Beyond the inherent spatial complexity of mountain hydrology, natural disturbances such as wildfire add another layer of challenge. Wildfires dramatically alter vegetation cover and soil structure, causing shifts in energy and water fluxes that control runoff and infiltration. Additionally, post-fire areas have highly variable burn severity, creating a mosaic of altered hydrological responses that are difficult to predict. These post-fire effects amplify the spatial heterogeneity in an already complex environment.

Wildfire impacts can persist for years and up to decades (Ffolliott et al., 1996; Niemeyer et al., 2020). Interannual variability in climate adds challenges to parsing out hydrologic changes due to fire, climate, and the combination of the two. Increased sediment yields (Wagenbrenner and Robichaud, 2015; East et al., 2021), increased occurrence of debris flows (Nyman et al., 2011; McGuire et al., 2017), and altered snow accumulation and melt patterns (Burles & Boon, 2011; Kampf et al., 2022; McGrath et al., 2023) influence downstream water quantity and quality (Murphy et al, 2015; Murphy et al., 2018; Nunes et al., 2018; Paul et al., 2022). Climate change is expected to drive increases in wildfire activity in the western U.S. (Rocca et al., 2014; Abatzoglou and Williams, 2016; Abatzoglou et al., 2021; Higuera and Abatzoglou, 2021), and fires have increased in both size and severity (Dennison et al., 2014; Westerling, 2016) while also advancing upslope to higher elevations (Alizadeh et al., 2015; Kampf et al., 2022). As wildfires continue to impact watersheds in the western U.S., more source water areas for municipal drinking water supplies are likely to be impacted. Predicting which areas are likely to experience the most severe post-fire problems would help in allocating resources for post-fire mitigation.

1.3 Research objectives

The goal of this research is to explore mountain hydrology in three ways: rainfall, streamflow generation, and hydrologic modeling. Broadly, this dissertation evaluates gridded

meteorological products, their suitability for complex terrain and post-fire landscapes, and the catchment characteristics influencing post-fire streamflow generation. In Chapter 2, a gridded rainfall product is compared to rain gauge data and used to assess how rainfall intensity varies with elevation and across a mountain range in northern Colorado. The objectives of this chapter are to (1) investigate how elevation influences rainfall patterns on both sides of the Colorado Front Range, (2) evaluate accuracy of a gridded rainfall product compared to tipping buckets, and (3) assess the reliability of both gridded and observed rainfall data for capturing rainfall events that generate stream responses.

Along with the rainfall intensity patterns, snow accumulation, catchment morphology, and soil properties also vary across the northern Colorado mountains. Chapter 3 focuses on identifying whether fire, snow zone, and year post-fire are associated with different streamflow responses in the Cameron Peak and East Troublesome fires. The objectives of this chapter are to evaluate (1) how streamflow responses vary between fires, snow zone, and year and (2) what underlying factors explain these differences in post-fire stream response.

Finally, Chapter 4 brings together climate data and a hydrologic model to evaluate how gridded versus station datasets affect hydrologic partitioning for a partially burned Colorado watershed. The objectives are to (1) determine biases between gridded and station climate data, (2) assess differences in model parameters between calibration sets using gridded and station data, and (3) evaluate how the differences in climate input and calibrated parameters affect hydrologic partitioning and model performance. This chapter also examines how optimal model parameter values vary between burned and unburned parts of the watershed.

Together, these chapters bring together detailed analyses comparing station and gridded climate data and evaluating how climate relates to streamflow response in both burned and unburned Colorado watersheds.

CHAPTER 2. Spatial variability of rainfall in the Colorado Rocky Mountains: a comparison of radar-derived and rain gauge data

2.1 Introduction

Accurate rainfall data are essential for understanding hydrologic responses, forecasting floods, and managing water resources. In regions prone to flash floods, small errors in rainfall intensity and timing can lead to large differences in streamflow prediction (Vivoni et al., 2006; Collier, 2007). Rainfall can vary substantially over space and time, particularly in mountainous settings, making it difficult to accurately quantify. High spatial and temporal rainfall estimates are critical for detecting localized rainfall events and triggering timely flood warnings (Nikolopoulos et al., 2015). In addition, identifying spatial patterns in rainfall intensity is crucial for understanding which areas are most vulnerable to large hydrologic responses.

Flash floods in the Colorado Rocky Mountains are often triggered by intense rainfall. For example, in July 1976, as much as 305 mm of rain fell within a few hours in late July and produced the Big Thompson flood, which caused 139 deaths and an estimated $35 million in property damage (Jarrett, 1990). In this region the most extreme floods, such as the Big Thompson flood, have occurred at middle elevations (~1500 to 2500 m), leading researchers to speculate that rainfall intensity varies with elevation. One line of evidence for elevation dependence in high-intensity rainfall was a paleo-hydrological study in Colorado that found the highest area-normalized peak flows between 1500 and 2300 m, based on records ranging from 1844 to 1976 depending on the watershed (Jarrett, 1990). This led to the idea that more intense rains fell in this elevation range compared to higher elevations. These findings on peak flow elevation patterns were corroborated in another study on peak discharges (England et al., 2010), but when subsequent researchers examined rainfall data, they found that event-scale rainfall metrics only weakly depend on elevation in the Colorado Front Range (Mahoney et al., 2015; Bohne et al., 2020; Rossi et al. 2020; Yu et al., 2021). The subsequent researchers were

not examining paleoflood hydrology, and the contrasting findings could be related to a changing climate. Rossi et al. (2020) suggested that the greater peak flows at middle elevations were caused not only by high rainfall intensity but also by greater bedrock exposure at middle elevations, leading to more overland flow and flash flooding. Rainfall in northern Colorado also varies between the west and east sides of the Continental Divide, with larger storms on the east side of the mountains, where summer rainfall represents a greater percentage of the annual precipitation. It is likely the largest rainfall totals occur on the east side due to southerly and southeasterly wind regimes that bring moisture in from the Gulf of Mexico (Mahoney et al., 2015).

Understanding how rainfall intensity varies spatially and across elevation gradients is essential for assessing flood risk in northern Colorado, particularly during convective thunderstorms in summer, which tend to produce the most severe flash floods. Convective storms are typically a few kilometers in areal extent (Fosser et al., 2015) and may be brief, lasting only a few minutes to hours but with the potential for localized high-intensity rainfall. The high spatial and temporal variability of rainfall means that many gridded rainfall products, such as the Gridded Surface Meteorological (gridMET) and Parameter-elevation Regressions on Independent Slopes Model (PRISM), may not adequately represent these storms due to inadequate temporal (e.g., daily; Rossi et al., 2020) and/or spatial resolution (e.g., 4 km; Alvarez et al., 2014; McEvoy et al. 2014; Behnke et al., 2016; Henn et al., 2018; Walton and Hall, 2018; Serrano-Notivoli et al., 2021). Tipping bucket rain gauges can capture the fine temporal resolution of convective storms, but they lack sufficient spatial coverage (Michaud and Soroosian 1994; Vasiloff et al., 2009) unless a dense network is installed within each watershed. Additionally, tipping bucket rain gauges are known to underestimate high-intensity rainfall owing to undercatch, mechanical errors, and lack of accurate calibration (Ciach 2002; Molini et al., 2005; Lanza and Stagi, 2008). Tipping bucket rain gauge data are also susceptible to data logger failure and damage from wildlife or other environmental factors.

Radar data are increasingly used to monitor the spatial patterns of convective storms. For example, radar data are used in the Multi-Radar/Multi-Sensor System (MRMS) quantitative precipitation estimates (QPE) dataset (Zhang et al., 2016) by creating a 3D radar mosaic at 1 km^2 spatial resolution and 2-minute time intervals, which is then integrated with atmospheric data, satellite data, and lighting and rain gauge observations to create the Multi-Sensor QPE dataset. Ground-based radar is one of the highest spatial and temporal resolution sources of rainfall data available, but mountain peaks and other obstructions can interfere with the radar signals. Radar estimates can have errors associated with calibration biases, blockages, contamination from non-precipitation echoes, uncertainties in the relationship between precipitation rate and radar variables, and variability in the vertical profile of reflectivity (Zhang et al., 2016). A study in northern California compared MRMS QPEs with rain gauges and found that half of the events matched sufficiently (i.e., r2 and slope of the linear fit > 0.5; Bytheway et al., 2019). Over the mountains in Colorado, the MRMS overestimated rainfall rates, with a median normalized root mean square error (RMSE) of 42% for the maximum 15-minute rainfall intensity (White and Nelson, 2024). Studies comparing MRMS to rain gauges have reported RMSEs ranging from 0.5 to 10 mm in Texas for 15-minute temporal resolution and the southeastern U.S. and 0.17 to > 1 mm/hr in Canada (Bayabil et al., 2019; Moazami and Najafi, 2021; Rivera Giboyeaux and Weinbeck 2024). These examples demonstrate that although MRMS provides high spatial and temporal resolution for capturing rainfall, its accuracy varies by region and terrain, highlighting the need to understand the reliability of MRMS QPEs for streamflow prediction in different regions.

The importance of tracking high-intensity rainfall has increased in northern Colorado in recent years because of wildfires. Wildfires can cause runoff to increase by orders of magnitude compared to pre-fire conditions (Hallema et al., 2017), and runoff increases are most severe during high-intensity rainfall. This can be even more enhanced when soils are hydrophobic post-fire (Moody et al., 2013). Wildfires that have the greatest impact on water supply are often

located in mountainous terrain near headwater streams, where quantifying precipitation tends to be the most uncertain (Ouyang et al., 2021). Given the critical role of high-intensity rainfall in driving post-fire streamflow responses in northern Colorado, understanding the influence of elevation on rainfall patterns across the mountains is essential.

This goal of this study was to evaluate how rainfall intensity varies with elevation across the northern Colorado Front Range mountains using both tipping buckets and gridded precipitation data. The study leveraged existing stream sensors and rain gauges from the first three years post-fire at Cameron Peak Fire (CPF) and two years post-fire at East Troublesome Fire (ETF). Stream sensors were used to assess the capacity of the MRMS and tipping bucket data in detecting rainfall amounts and intensities that generate stream occurrences. By comparing these data sources, we addressed key uncertainties in point measurements and radar-informed rainfall and improved the understanding of warm season rainfall patterns (i.e., June through September) in the northern Colorado mountains. The specific objectives of this study were to (1) investigate how elevation influences rainfall intensity and totals on both sides of northern Colorado's Front Range mountains; (2) assess differences in warm season rainfall estimates between MRMS and those from gridMET, PRISM, and tipping buckets; and (3) assess the reliability of both MRMS and tipping bucket data for capturing rainfall events that generate stream occurrences.

2.2 Methods

2.2.1 Study area

The study area is located in the Southern Rockies Ecoregion (Omernik, 1987) in Colorado's northern Front Range mountains. The western and eastern portions of the study area ("western area" and "eastern area", respectively) are separated by the Medicine Bow Mountains and the Continental Divide in Rocky Mountain National Park and Indian Peaks Wilderness (Figure 2-1). The elevation in the western area ranges from 1977 to 4115 m, while

the eastern area ranges from 1474 to 4344 m. The study area encompasses the boundaries of the two largest wildfires in Colorado to date, CPF and ETF, which burned in northern Colorado in 2020 and both exceeded 715 km^2. The CPF elevation ranges from 1615 to 3618 m and ETF from 2507 to 3319 m. Post-fire burn severity mapping indicates that 44% of the CPF burned at low severity, 30% at moderate severity, and 6% at high severity. The burn severities for the ETF are 18% for low, 46% for moderate, and 15% for high (Monitoring Trends in Burn Severity [MTBS], 2017; Figure 2-2). The higher elevations of the CPF consist of subalpine spruce fir forests, mid-elevations consist of mixed-conifer forests, and lower elevations consist of ponderosa pines, prairie grasses, and shrubs (Addington et al., 2018). In contrast, ETF was largely composed of lodgepole pines interspersed with Engelmann spruce-subalpine fir and aspens (LANDFIRE, 2016).

Figure 2-1. The Cameron Peak (CPF) and East Troublesome (ETF) fire perimeters, including the delineation between the western and eastern study area. The inset map shows the study area location, denoted by a red star.

Across these two fires and in nearby unburned areas, 26 catchments were monitored (Table 2-1, Figure 2-2) with stream stage sensors. Tipping buckets were installed within or near all catchments. For the CPF, the study period of observed data includes three warm seasons post-fire (2021–2023) and for ETF two warm seasons post-fire (2022–2023). The catchments vary in contributing areas from 1 to 39 km^2. Five of the 26 catchments are unburned, and the remaining are burned (Figure 2-2, Table 2-1). In addition to the tipping buckets that were paired with the catchments, we utilized the Larimer County Early Flood Warning System, which includes real-time tipping bucket data in an online archive (Larimer County, 2024). In total, between our dataset and Larimer County, 62 tipping buckets were included in the study.

2.2.2 Data collection

The stream stage was measured at the outlets of most catchments using either capacitance rods (TruTrack WT-HR 1000) or pressure transducers (In-Situ Rugged TROLL and Onset HOBO Water Level Logger). The pressure transducers were unvented and paired with a barometric sensor to correct for atmospheric pressure. The stream stage was recorded at either 5- or 15-minute intervals. At each stream monitoring site, during site visits, the stream stage and bed position relative to the stage plate were recorded. The bed position measurements indicated whether the channel bed had aggraded or eroded, both of which are common changes after storms in burned watersheds. The bed change measurements were used to offset the stage data to account for bed changes after large flow events. For the lower elevation sites at ETF, the streams experienced large amounts of bed and channel changes post-fire during precipitation events, so it was not feasible to install stage sensors. Instead, time-lapse cameras were used, and stage changes during storms were estimated from repeated photos of the stream.

The majority of the CPF catchments had their own tipping buckets, except for the Washout and UM sites, which used nearby or adjoining catchments (Table 2-1). Tipping buckets at CPF were either Texas Instruments TE525-L tipping buckets or RainWise Rainew™ tipping buckets. At ETF, all tipping buckets were Onset RG3 tipping buckets. The catchments at ETF are located in clusters, and the tipping buckets were paired based on proximity to one of three clusters (Table 2-1, Figure 2-2), so not every catchment had a tipping bucket. The tipping bucket data were recorded using either Onset HOBO pendant event data loggers or Campbell Scientific CR1000 loggers. Each tip represented 0.3 mm of rainfall. A wind-correction factor was not applied to the tipping bucket data because wind speed data were not available. Error due to systematic wind field deformation above the tipping bucket orifice can be two to ten percent for rainfall (World Meteorological Organization, 2008). The tipping buckets were not dynamically calibrated to varying intensity rainfall, which may also affect gauge accuracy.

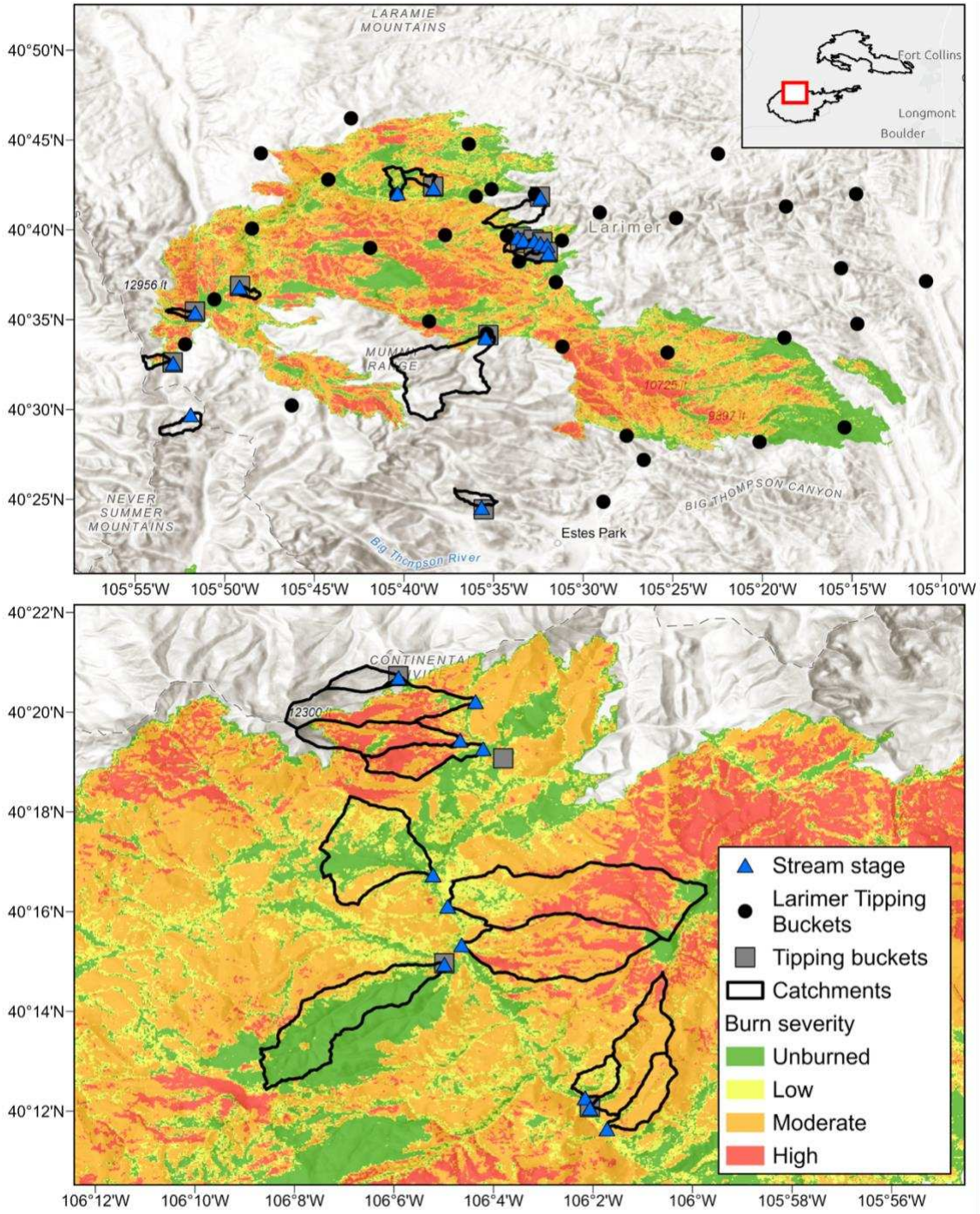

Figure 2-2. The Cameron Peak Fire (CPF; top) and East Troublesome Fire (ETF; bottom) study catchments with burn severity (MTBS, 2017) overlaid on hillshade (ESRI, 2017). The grey squares and black circles represent tipping bucket locations, and the blue triangles represent stream stage sensors or cameras. The red square shows the approximate location of the ETF study catchments within the fire perimeter.

Table 2-1. Study catchment characteristics. For rain gauge or stage sensor, TB is tipping bucket, cap. rod is a capacitance rod, and PT is a pressure transducer.

Catchment	Fire	Area Burned (%)	Area (km^2)	Elevation range (m)	Rain gauge	Stage sensor
ME	CPF	99	1.44	2397–2681	TB	Cap. rod
MM	CPF	99	0.72	2357–2640	TB	Cap. rod
MW	CPF	98	1.50	2382–2762	TB	Cap. rod
UE	CPF	100	0.61	2403–2650	TB	Cap. rod
UM	CPF	100	0.57	2459–2653	TB from UW	Cap. rod
UW	CPF	100	1.03	2457–2771	TB	Cap. rod
Bighorn	CPF	0.0	3.05	2587–3433	TB	PT
Washout	CPF	55	2.68	2357–3071	TB from Dry	PT (Larimer)
Dry	CPF	70	2.91	2317–2933	TB	PT
Michigan	CPF	0.0	3.97	3165–3840	TB	USGS
Montgomery	CPF	44	1.93	3059–3482	TB	Cap. rod
BL4	CPF	90	1.00	2907–3390	TB	Cap. rod
Aspen	CPF	73	1.13	2678–3041	TB	Cap. rod
Mt. Campus	CPF	3.4	38.93	2743–4080	TB	PT
Dadd	CPF	27	7.72	2145–2718	TB	Cap. rod
P1	ETF	0.0	1.20	2896–3456	TB	Cap. rod
P2	ETF	59	7.04	2768–3750	TB from P1	Cap. rod
HUM	ETF	85	4.03	2756–3724	TB	PT
HM	ETF	95	2.56	2729–3263	TB from HUM	PT
MUM	ETF	71	7.87	2652–3233	TB from MUB	PT
MPM	ETF	96	12.46	2637–3589	TB from MUB	PT
MUB	ETF	28	8.56	2620–3277	TB	PT
MM_ET	ETF	98	9.21	2627–3491	TB from MUB	Camera
LPM	ETF	98	4.46	2566–3328	TB	Camera
LM	ETF	100	2.23	2570–3007	TB from LPM	Camera
LUM	ETF	96	1.25	2565–2985	TB from LPM	Camera

2.2.3 Determining rainfall patterns based on elevation and location

The MRMS dataset is a radar-derived product with an approximate 1-km spatial resolution. In the study area, MRMS uses the S-band dual-polarization Weather Surveillance Radar-1998. The radar product is integrated with multisensor data (i.e., precipitation gauges) to

provide a more accurate product. Precipitation gauge data are used to correct the radar biases (Zhang et al., 2016). The MRMS versions utilized in this study were MRMS v12.1.0 (for 2021) and v12.2.0 (for 2022 and 2023). The hourly MRMS products include the Multi-Sensor QPE (MultiSensor_QPE_01H_Pass2) and radar-only QPE (RadarOnly_QPE_01H). The multi-sensor QPE product is terrain corrected. The surface precipitation rate (PrecipRate) product is a raw radar product available at 2-minute time steps and 0.1 mm precision. We used the surface precipitation rate and the relationships between surface precipitation rate and QPE to derive a correction factor, which was then applied to the surface precipitation rate to estimate precipitation at 2-minute intervals. The MRMS products were accessed through the Iowa Environmental Mesonet MRMS archived daily data (Iowa Environmental Mesonet, 2024) via an R script that downloads the rasters, clips the area of interest, and projects into a desired coordinate reference system using the R spatial packages *terra* and *stars* (Pebesma and Bivand, 2023; Hijmans, 2025). Prior to analyzing the MRMS data, they were converted from UTC to MST to match the time zone of the tipping bucket data. The 2-minute MRMS rainfall rasters were summed for each pixel to a 10-minute time step for easier computation when determining the maximum rainfall intensities.

To compare rainfall across the study area, we considered only warm season rainfall, defined as the total rainfall between 1 June and 30 September (2021-2023). During other times of the year, precipitation may sometimes fall as snow or ice, which can interfere with tipping bucket readings. Although our focus was on rainfall only, snowfall can occur in June and September; therefore, small amounts of snow may be captured in the measurements at higher elevations.

To characterize the spatial patterns of rainfall, we used three metrics: total rainfall, maximum 60-minute rainfall intensity (MI60), and total amount of high-intensity rainfall. MI60 was computed by finding the rolling sum over six time steps within each pixel and finding the maximum depth of rainfall over any 60-minute interval during each warm season. Other

durations were also considered (e.g., MI30), but since the maximum intensities at these durations are highly correlated with one another (Wilson et al., 2018), we focus only on MI60. To quantify the amount of rainfall that fell at high intensity, the rainfall was summed for each timestep that recorded MI60 greater than 10 mm/hr for each warm season (P>10) (Kampf et al. 2018; Wilson et al., 2018).

We examined how these rainfall metrics varied with elevation using a 10-meter USGS DEM (USGS, 2022a; USGS, 2022b; USGS, 2023a; USGS, 2023b) resampled to match the 1-km spatial resolution of MRMS. The elevation for each MRMS pixel was extracted and used to compare rainfall across elevation ranges in the western and eastern areas. To facilitate analysis, elevation values were separated into 150 m bins across both fires, and the median values were computed for each bin. For total rainfall, we also compared the MRMS-derived values to gridMET and PRISM for the warm seasons of 2021 through 2023 across both fires. The gridMET dataset contains daily surface meteorological variables at 4 km, derived from North American Land Data Assimilation System 2 (NLDAS-2) using climatically informed interpolation (Abatzoglou, 2013). The gridMET data were retrieved using the R packages *climateR* and *AOI* (Johnson, 2024a; Johnson, 2024b). The PRISM dataset contains daily surface meteorological variables at 800 m, derived mainly from gauge measurements and physiographic factors (Daly et al., 2008). The PRISM data were retrieved from the PRISM Group's (2025) online database using the R packages *httr* and *terra* (Wickham, 2023; Hijmans, 2025).

Finally, we evaluated how radar quality affected results using the radar quality index (RQI). The RQI has values from zero to one that indicate the uncertainty due to blockage and radar beam height with respect to freezing level. The radar beam blockage is determined using a static terrain map and assumes standard atmospheric refraction conditions. The radar beam height is not static and comes from the model freezing level height. A lower RQI indicates blockage and/or higher beam heights (Zhang et al., 2012). The RQI has the same spatial and

temporal resolution as the surface precipitation rate, and we extracted the mean RQI for each pixel over each warm season.

2.2.4 Comparing rainfall data sources

To compare the rainfall between the MRMS and tipping bucket data, the MRMS pixels containing each tipping bucket were identified. For these locations, each dataset was separated into discrete rainfall events. A new event was defined when more than 0.3 mm of rainfall fell after at least six hours with no rainfall. Only events with rainfall greater than 1 mm were included in the analysis (Miller et al., 2025). For each event in both tipping bucket and MRMS datasets, the total depth of rainfall (event rainfall), duration, and MI60 were calculated. For MRMS, the mean RQI for each event was also computed. Tipping bucket event metrics were calculated for the time periods of the MRMS events, with an hour buffer before the start and after the end of the tipping bucket rainfall. This buffer was applied since the event times were not always equal between the two datasets. As the event data were not normally distributed, the Spearman correlation coefficient was used to evaluate the correlations between the MRMS and tipping bucket rainfall metrics.

2.2.5 Evaluating MRMS and tipping buckets using stream occurrence

Tipping buckets are essentially point measurements of rainfall over a very small surface area (tipping bucket aperture), whereas each MRMS pixel covers 1 km^2. Both methods may have shortcomings in capturing rainfall at the catchment scale. The small area of the tipping bucket may miss some of the rainfall over a catchment or have biases related to equipment placement, wind, and maintenance, whereas MRMS is more likely to have inaccurate rainfall quantities. To evaluate how well these two data sources represent the rainfall events that cause streamflow, we used stage measurements from the catchments in each study area (Table 2-1). Binary stage occurrences were identified by a stage increase of 5 cm or greater above pre-event stream stage and these occurrences were filtered to exclude diurnal signals. For the

MRMS events, the weighted catchment average from pixels within or overlapping the catchment boundary was used to determine rainfall event metrics. Stream stage records were examined for dates when there was either a tipping bucket or MRMS rain event.

Rainfall events (MRMS and tipping bucket) that were not associated with a stream occurrence (i.e., a stage rise greater than 5 cm) were labeled with a zero, whereas those that were associated with a stream occurrence were labeled with a one. In some post-fire catchments, stream occurrence can be predicted by a rainfall intensity threshold (Wilson et al., 2018), but rainfall intensity alone is not a good predictor of streamflow occurrence in CPF because the initial wetness conditions of the catchment are also important (Miller et al., 2025). Therefore, we evaluated how well the rainfall metrics from tipping buckets and MRMS predict streamflow occurrences using logistic regressions with a rainfall metric (MI60 or total event rainfall), cumulative potential water deficit, and mean differenced normalized burn ratio (dNBR) as the predictors of streamflow occurrence (binary, occurrence/no occurrence). Potential water deficit (PWD) is defined as the difference between the daily rainfall and the daily reference evapotranspiration (Abatzoglou, 2013). For each catchment, we obtained the daily average PWD for each warm season from gridMET. Starting at the beginning of the respective water year, we computed the cumulative PWD by adding each day's PWD to the previous day's cumulative value. We used the cumulative PWD over a water year as a soil moisture proxy. The PWD value extracted for each rainfall event was the catchment average value for the day prior to the event.

The results of the logistic regression yielded stream occurrence probabilities from zero to one. We assumed that the regression predicts a stream occurrence when the probability is greater than 0.5 and no occurrence when the probability is <0.5. Cohen's kappa statistic was calculated to assess how well the logistic regression predicted binary stream occurrence, using either MRMS or tipping buckets for the rainfall inputs. Kappa values greater than 0.81 indicate near-perfect agreement, values between 0.61-0.80 indicate substantial agreement, and values

between 0.41-0.60 indicate moderate agreement (Viera and Garrett, 2005). In addition to the kappa statistic, the Kling-Gupta Efficiency (KGE) metric was used to evaluate the agreement between observed and predicted stream occurrence (i.e., the probability values), providing a continuous measure of model performance that balances correlation, bias, and variability (Gupta et al., 2009). The metrics were presented for burned sites only since burned areas are more affect by rainfall intensity and events that trigger a stream occurrence.

2.3 Results

2.3.1 Changes in rainfall with elevation

In the study area, total warm season rainfall was greater on average over the three years in the eastern area compared to the western area (202 and 162 mm, respectively, based on gridMET) and had considerable spatial variability. The gridded daily rainfall products, gridMET and PRISM, showed less spatial variability (based on the standard deviations across the study area) than MRMS and also recorded less warm season rainfall in the eastern area. In contrast, gridMET and PRISM reported greater total rainfall in the western area compared to MRMS (Figure 2-3). Although the warm season rainfall totals and spatial resolution are different between gridMET, PRISM, and MRMS, similar spatial patterns were observed. In Figure 2-3, for 2023, all the datasets show greater rainfall in the eastern area compared to the western area, but there is a difference in magnitude. For example, MRMS, gridMET, and PRISM show a hotspot of high rainfall on the eastern portion of the CPF (608 mm, 401 mm, and 417 mm, respectively), though with differing magnitudes. Among the three years, 2021 had the lowest median warm season rainfall and 2023 had the highest. In 2023, a significant portion of the study area experienced one of the wettest warm seasons on record (Colorado Climate Center, 2025). Over the entire study area, the overall maximum rainfall values were greater for MRMS than gridMET and PRISM. Conversely, the median values are generally greater for gridMET and PRISM (161 mm and 163 mm, respectively) compared to MRMS (106 mm).

Rainfall intensity also varied across the study area. The total high-intensity rainfall (P>10) was greater in the eastern area compared to the western area, but it exhibited much greater spatial variability than total warm season rainfall. Rainfall above the MI60 threshold of 10 mm/hr (Figure 2-3 and Figure 2-4) was highest in warm season 2023. In 2021, a year with lower rainfall, some areas experienced high-intensity rainfall at both high and low elevations. However, in 2022 and 2023, these areas shifted to primarily mid to lower elevations. Very little high-intensity rainfall was reported by the MRMS dataset in the western area.

Tipping bucket rain gauges within ETF (west) and CPF (east) boundaries with MI60 values separated by high and low elevations (2600 m threshold; Coe et al., 2014) also indicate some differences in rainfall intensity between the western and eastern areas (Figure 2-4). Although the median MI60 values for rain storms are similar across both sides of the mountain and elevation zones, the intensities in the eastern area reached greater magnitudes than those in the western area, and lower elevations of the CPF (east) experienced the highest maximum intensities.

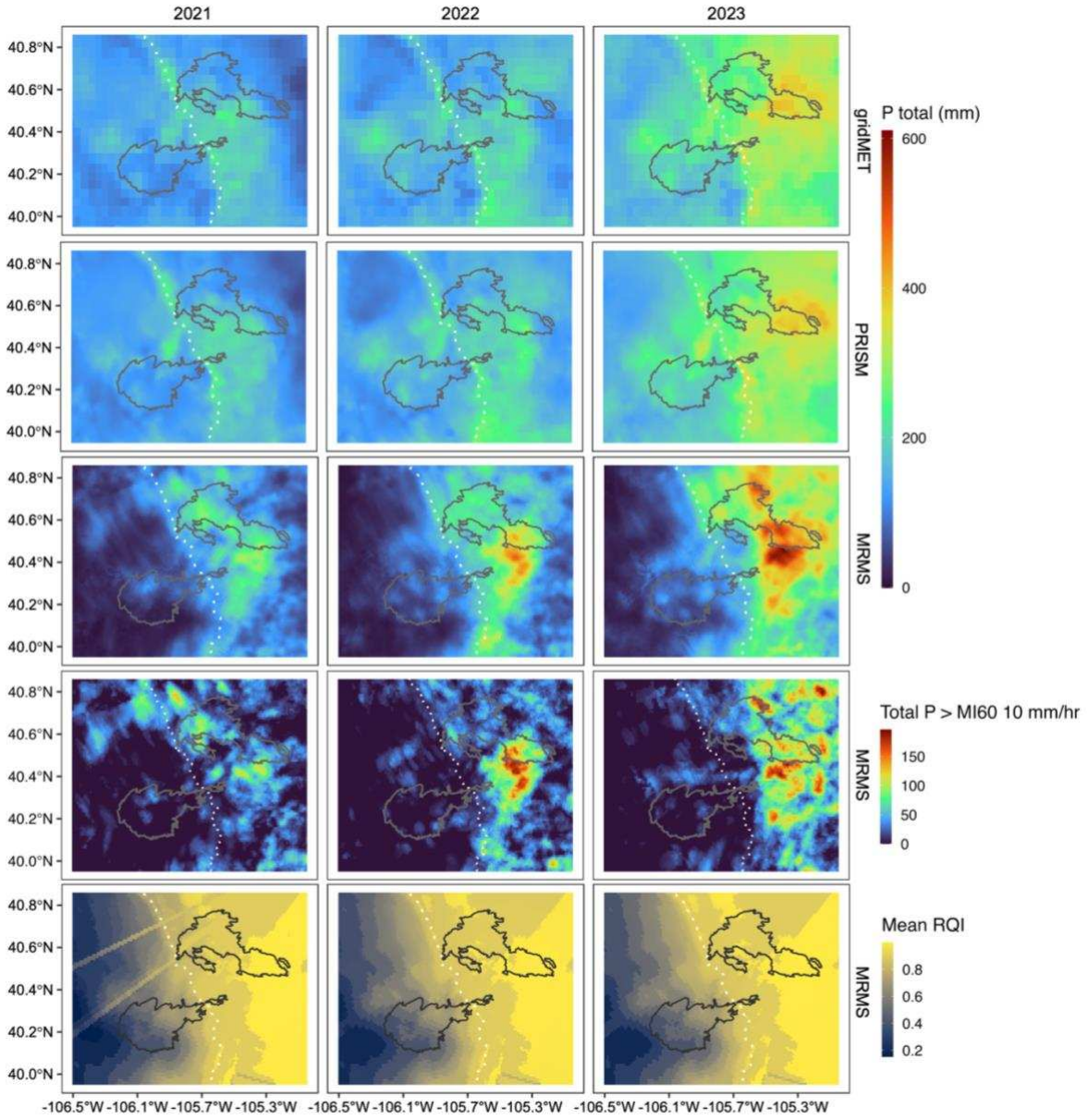

Figure 2-3. Warm season rainfall totals for 2021 to 2023 for gridMET, PRISM, and MRMS, total rainfall that exceeded MI60 10 mm/hr (P>10), and MRMS radar quality index (RQI). The grey and black outlines represent the CPF (east) and ETF (west), respectively. The white dotted line represents the divide between the western and eastern areas.

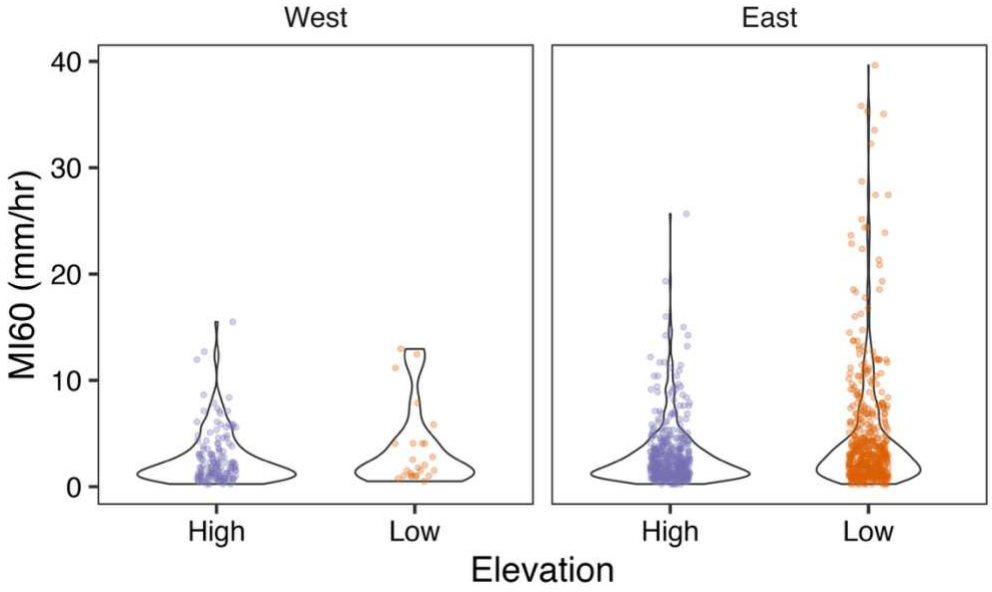

Figure 2-4. Rain event MI60 (mm/hr) values from the tipping bucket gauges located in the western (ETF) and eastern (CPF) areas separated by high and low elevations (2600 m threshold, Coe et al., 2014). The CPF tipping buckets include warm seasons 2021 through 2023 and ETF includes warm seasons 2022 and 2023.

To compare rainfall intensities across elevations, Figure 2-5 presents the median rainfall above a MI60 of 10 mm/hr for MRMS pixels within 150 m bins for the western and eastern areas. In 2021, in the eastern area, higher elevations had less high-intensity rainfall, whereas the opposite was true for the western area. In 2022, high-intensity rainfall increased with elevation until approximately 2500 m, then decreased at a similar magnitude with increasing elevation except at the highest elevation in the eastern area. For the western area in 2022, high-intensity rainfall increased with elevation, and no high-intensity rainfall was recorded at lower elevations (<2450 m). Warm season 2023 had the strongest relationship between high-intensity rainfall and elevation in the eastern area, where high-intensity rainfall was inversely related to elevation. When comparing differences of median P>10 within elevation bins, the majority between east and west are statistically significant (p-value < 0.05), except for a few elevations bins greater than 3500 m (Figure 2-4).

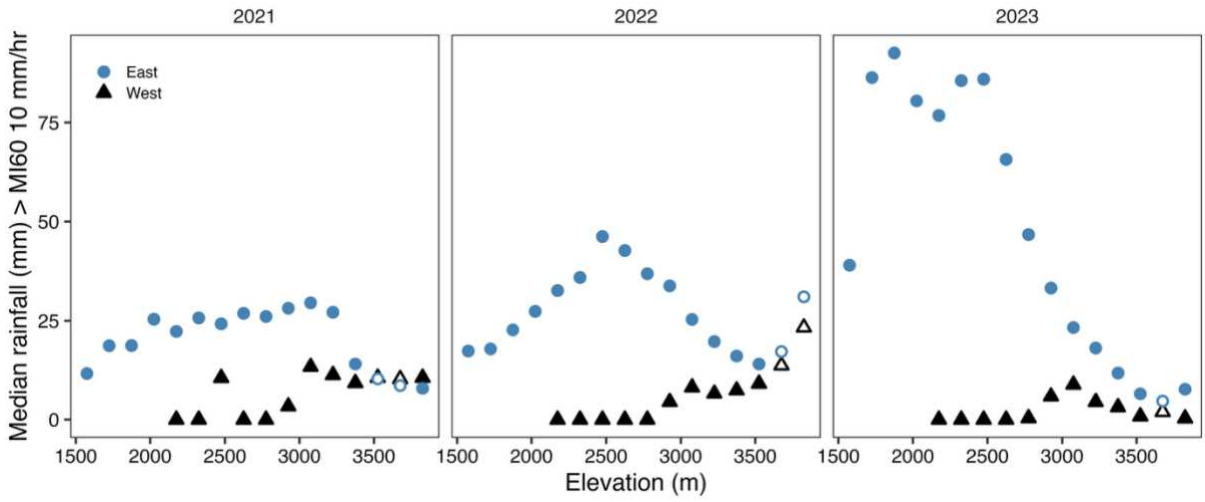

Figure 2-5. Median P>10 (rainfall (mm) over MI60 10 mm/hr) for 2021 to 2023 from MRMS. Elevation is grouped into 150 m bins, with the point shown at the median value for all pixels in the bin. Filled shapes indicate elevation bins where differences between east and west are statistically significant (p-value < 0.05), while unfilled shapes non-significant differences.

2.3.2 Comparing MRMS and rain gauges

The MRMS data and tipping bucket gauges showed moderate agreement for event rainfall and MI60 in both the western and eastern areas (Figure 2-6). For MI60, the Spearman correlations were approximately equal between the eastern and western areas (0.50 and 0.49, respectively). The MRMS data both underestimated and overestimated MI60 compared to tipping buckets for both areas. Event rainfall had slightly greater correlations (0.55 and 0.58) and similarly under and overpredicted. The RMSE values for the western area are slightly lower for MI60 and event rainfall (3.6 mm/hr and 4.7 mm) than the eastern area (5.3 mm/hr and 7.2 mm). The relationships between MRMS and tipping buckets show considerable scatter, with many instances in which tipping buckets are greater than MRMS and vice versa. Across the western and eastern areas, the highest density of points occurred at lower intensities (< 6 mm/hr) and lower event rainfalls (< 5 mm) (Figure 2-6).

We evaluated whether the accuracy of the MRMS relative to tipping buckets differed for pixels with varying radar quality by plotting the residuals (tipping bucket-MRMS) with RQI (Figure 2-7). No clear relationship between the RQI and MRMS performance is evident from the

point-to-grid analysis. The RQI values were much greater in the eastern area compared to the western area. The mean RQI was 0.92 with a standard deviation of 0.05 in the eastern area and 0.49 with a standard deviation of 0.09 in the western area.

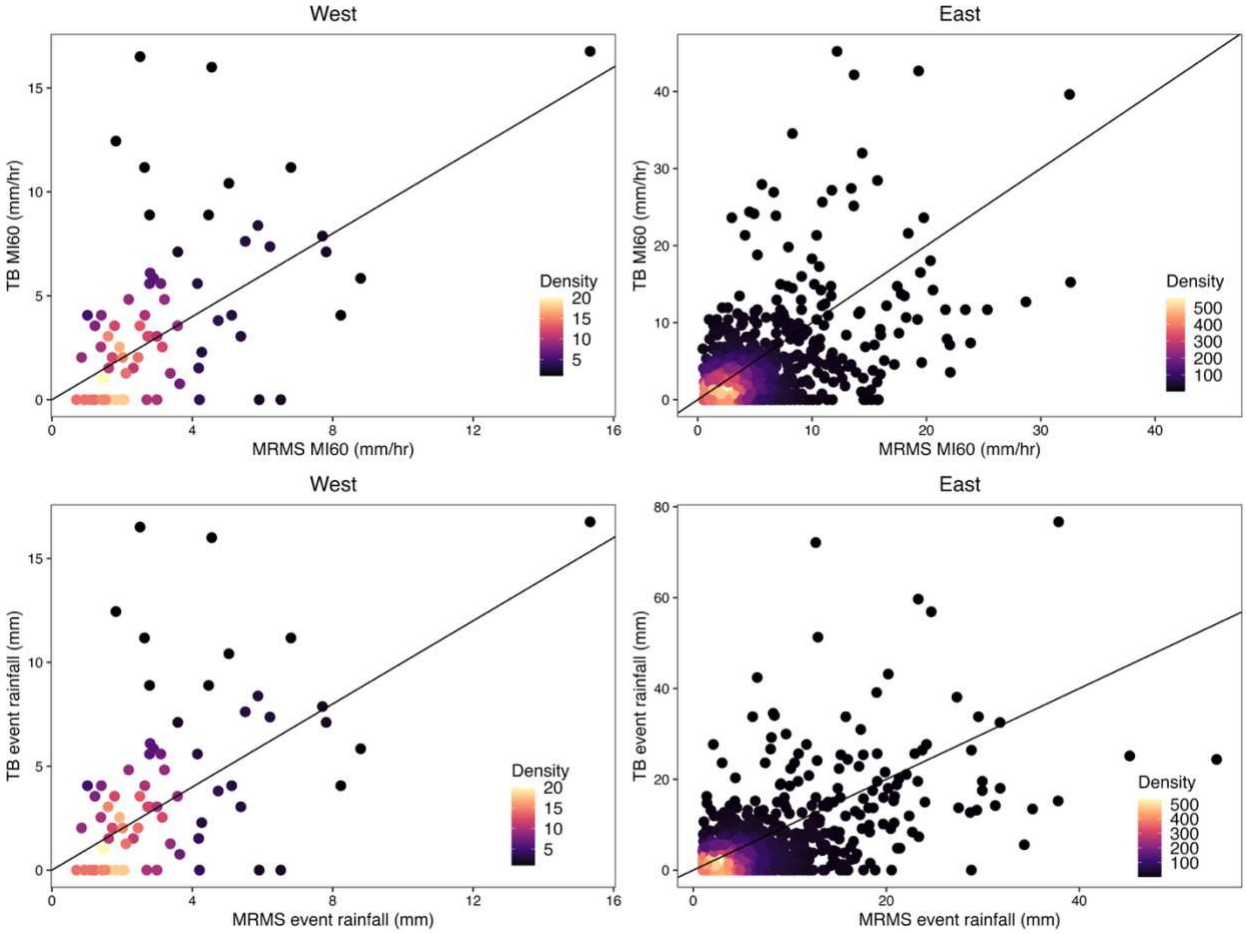

Figure 2-6. Comparison of tipping bucket (TB) measurements and corresponding MRMS pixel. The black solid line represents the one-to-one line, and points are colored by density. The western area sites include warm seasons 2022 and 2023 and the eastern area sites include warm seasons 2021 through 2023.

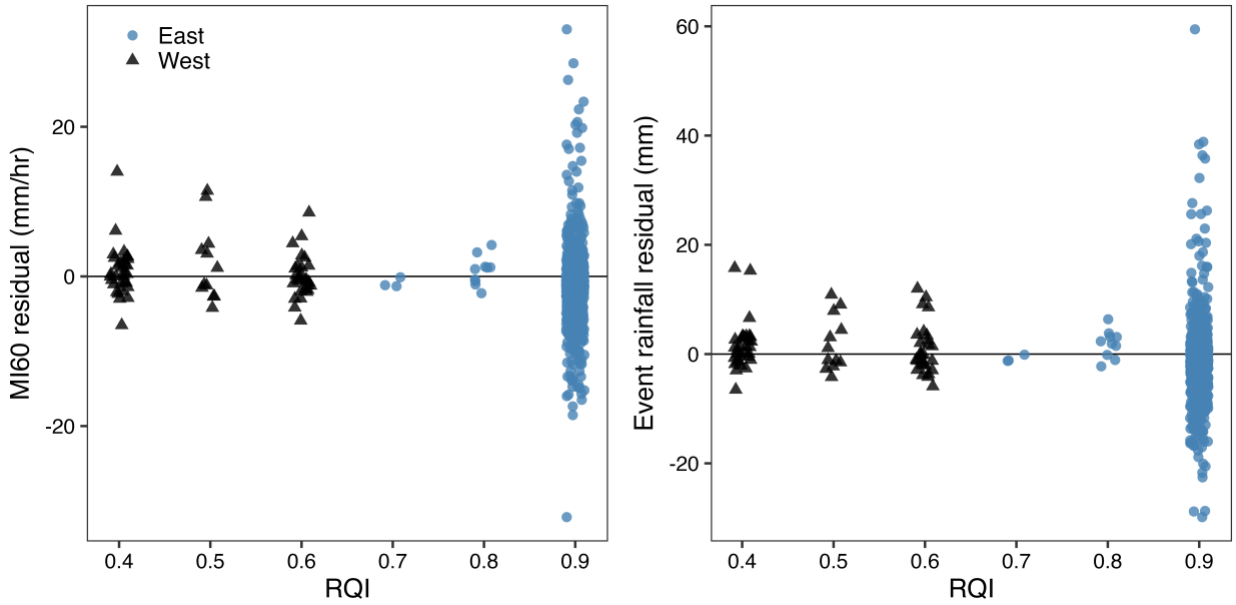

Figure 2-7. The MI60 and event rainfall residuals (tipping bucket - MRMS) plotted against radar quality index (RQI). The western area sites include warm seasons 2022 and 2023 and the eastern area sites include warm seasons 2021 through 2023.

2.3.3 Evaluating MRMS and tipping buckets using stream occurrence

Since areas affected by wildfire are more sensitive to rainfall intensity and rainfall events that trigger a stream occurrence, only burned sites were focused on when evaluating the datasets using stream occurrence. Neither MRMS nor tipping bucket data reliably predicted binary stream occurrence across the burned sites, and neither data source consistently outperformed the other (Figure 2-8). The mean MRMS kappa statistics from the logistic regression for predicting a binary stream occurrence were slightly greater than tipping buckets at the CPF (east) (0.65 and 0.48, respectively), and the opposite was seen at ETF (west) (0.40 and 0.48, respectively). The KGE values had a similar pattern across both fires. For the CPF, MRMS (0.61) exceeded tipping bucket (0.39), whereas for the ETF, tipping bucket (0.42) exceeded MRMS (0.16). The MI60 is shown in Figure 2-8 because it was better at predicting stream occurrence than event total rainfall (i.e., it had the most sites with kappa values > 0.6). To compare performance of MRMS and tipping buckets, the Wilcoxon signed-rank test was applied to each group in Figure 6a since the data were non-normal. For both fires, the kappa

statistics and KGE values for the tipping buckets and MRMS were not statistically different (p-value > 0.05; Figure 2-8a). Therefore, one dataset was not consistently better at predicting a stream occurrence. Although the ETF MRMS and tipping bucket kappa statistics and KGE values are not statistically different, the median tipping bucket kappa and KGE values were greater than MRMS (Figure 2-8a). Table A-1 provides more information on the comparison of streamflow predictions between MRMS and tipping buckets. Figure 2-8b shows the 1:1 kappa statistic (MRMS:TB) for each catchment. Sites above the 1:1 line had greater kappa statistics for tipping buckets and vice versa.

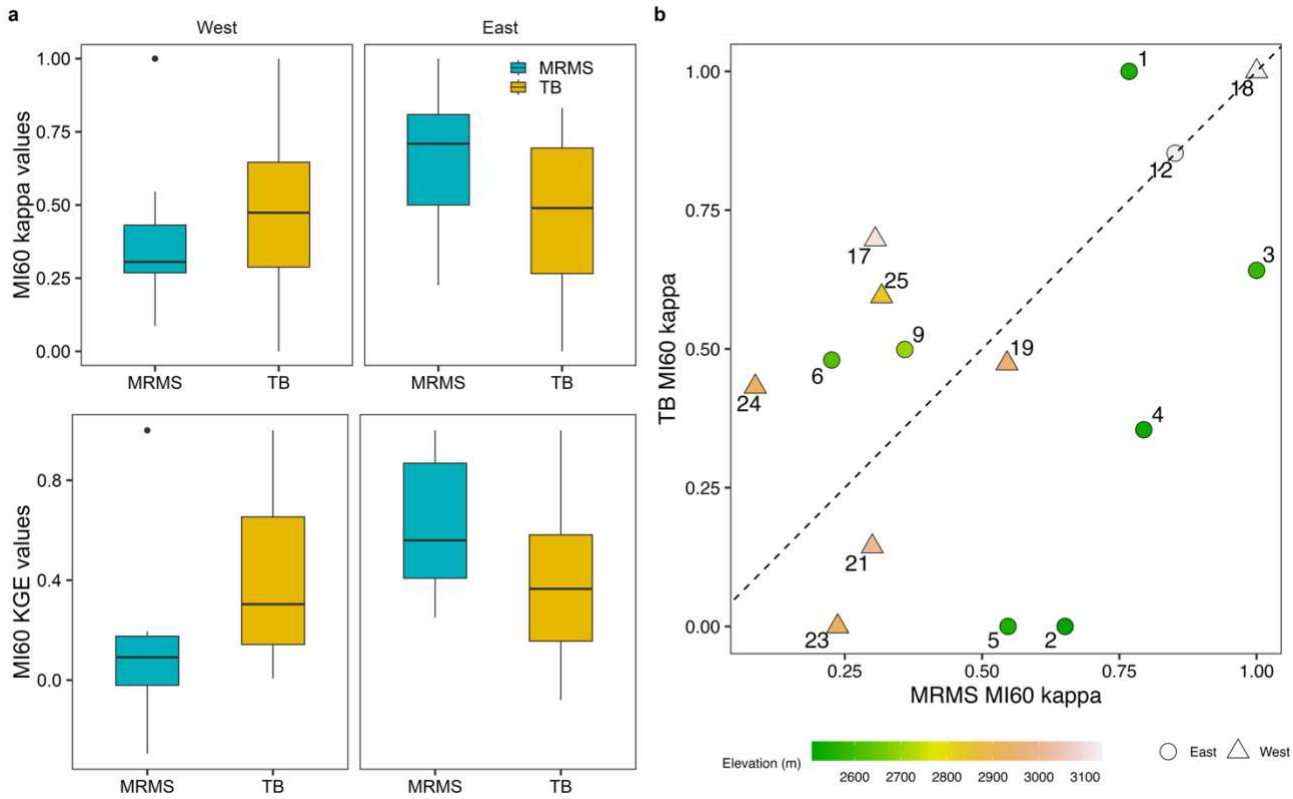

Figure 2-8. The MI60 kappa statistic from the logistic regression predicting binary stream occurrence at burned catchments by the western and eastern areas. (a) Kappa (top) and KGE (bottom) statistics for burned sites by rainfall data source. (b) The 1:1 comparison of the kappa statistics between MRMS and tipping bucket (TB).

The differences in MRMS and tipping bucket performance for predicting streamflow responses often related to the position of tipping buckets within a catchment. Figure 2-9 is an example of a convective storm at Dry Creek (CPF, mid-elevation, burned catchment) on 20 July

2021 over 70 minutes. This same storm system caused the catastrophic Black Hollow Creek flood less than 1000 m away, which resulted in loss of lives and homes. The MRMS rainfall started earlier (16:00 MST) than at the tipping bucket (16:40 MST). Although both reached maximum rainfall intensity at 16:50 MST, MRMS shows that the highest intensity was in the center of the catchment rather than at the tipping bucket location. This implies that the tipping bucket may have missed the highest intensities that influenced the stream occurrence. However, the MRMS values are lower than the tipping bucket values for the highest intensity part of the storm (62.5 mm/hr at the tipping bucket and 32.6 mm/hr in MRMS), indicating that MRMS may be somewhat inaccurate in the amounts of rainfall reported.

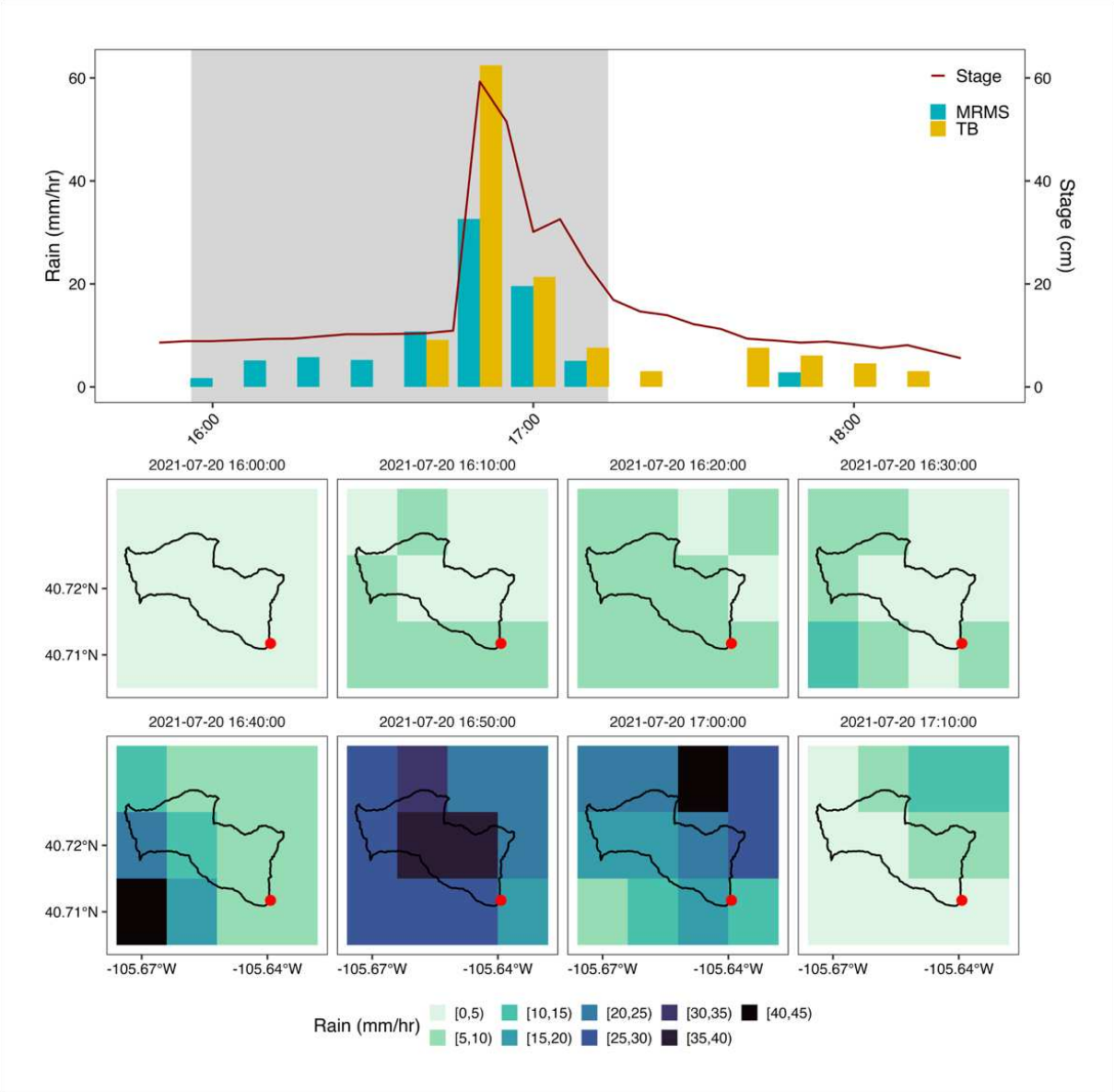

Figure 2-9. An example of a convective storm at Dry Creek on 2021-07-20 over 10-minute timesteps. The top plot is the rain intensity and stage over the storm for the TB and MRMS datasets. The bottom plot shows the MRMS rain intensity spatially across the catchment and the red point indicates the tipping bucket location.

2.4 Discussion

The first objective of this study was to evaluate the influence of elevation on rainfall along Colorado's northern Front Range mountains over three years. Overall, greater rainfall totals and greater rainfall intensities were observed in the eastern area compared to the western area (Figure 2-3, Figure 2-4, and Figure 2-5), consistent with climatological precipitation

patterns in northern Colorado for June through September (e.g., PRISM 30 year-normals). This was supported by four different datasets (tipping bucket gauges, MRMS, gridMET, and PRISM). In mountainous areas during the day, converging airflow on the leeward side, where warm upslope winds meet stronger winds from higher up, can lead to the formation of convective storms (Banta, 1984; Piekle, 2001). Both the western and eastern areas experience convective storms, but the eastern area experiences an east-to-west transport of moisture known as the mountain-plains solenoidal circulation (Toth and Johnson, 1985; Wolyn and McKee 1994; Douglas, 2024), primarily driven by broad daytime heating of the mountains with reversal at nighttime. This feedback loop could explain the higher-intensity rainfall observed in the eastern area compared to that observed in the western area, although additional research should be conducted to confirm this hypothesis.

Because of the high variability in both synoptic and local circulation patterns, we found little consistency in the relationships between rainfall intensity and elevation (Figure 2-5). For example, in 2023, rainfall intensity decreased with increasing elevation, but this pattern was not observed in either 2021 or 2022. The spatial patterns of high-intensity rainfall (Figure 2-2) show localized areas of high intensity that are not elevation-dependent and are not consistent between years. Based on the disagreement between years and the small sample size, we conclude that there is no clear relationship between high-intensity rainfall and elevation. The literature is also inconclusive with other studies finding weak or no relationships (Bohne et al., 2020; Mahoney et al., 2015; Rossi et al., 2020; Yu et al., 2021). The complexity of convection and mountain circulation warrants further study to determine which locations in the mountains are most or least conducive to developing high-intensity convective thunderstorms. This is especially important in areas that have experienced wildfires or are prone to debris flows (e.g., steep areas with loose, erodible material or sparse vegetation) since rainfall intensity is a critical driver of hydrologic hazards.

We also compared gridMET, PRISM, and MRMS warm season rainfall totals (Figure 2-3). GridMET and PRISM values are primarily derived from precipitation gauges. We found that MRMS total warm season rainfall values were greater in the eastern area than gridMET and PRISM, whereas gridMET and PRISM rainfall was greater than MRMS in the western area. The total warm season differences between datasets for the eastern area can be partially explained by the poor representation of convective storms in the gridMET and PRISM datasets. When there is a sparse rain gauge network used to develop the gridded products, they can miss some of the localized convective precipitation. In the western area, the difference between gridMET and PRISM compared to MRMS is likely due to poor radar coverage, potentially reducing the amount of rainfall detected by radar. White and Nelson (2024) found that areas further from radar sites in Colorado show consistently greater error, likely due to terrain blockages and sparse gauge networks. Therefore, MRMS should be used with caution in areas where the RQI is low, such as the western area of the study. When selecting a gridded precipitation dataset, it is important to consider the tradeoffs between spatial and temporal resolutions (Partridge et al., 2024). The gridMET and PRISM datasets may provide more representative rainfall estimates in the western area due to the low MRMS radar coverage, but their daily temporal resolution limits their usefulness for capturing short-duration, high-intensity events that often cause flash floods and debris flows. The PRISM dataset offers a finer spatial resolution than gridMET (800 m compared to 4 km), and it therefore would be a preferable option, provided enough rain gauge data are available to support the finer resolution. Additionally, the MRMS dataset begins in late 2014 and therefore has a shorter record length than many gridded products.

Our second objective was to evaluate the accuracy of MRMS compared to tipping bucket rain gauges. Comparing MRMS and tipping buckets at storm event time scales, we found that the two were significantly correlated, but with considerable variability in the relationships, with a mixture of underpredicting and overpredicting (Figure 2-6). The East Troublesome Fire (west) showed slightly better performance than Cameron Peak Fire (east)

based on point-to-grid RMSEs (5.3 and 3.6 mm/hr, respectively). This difference may reflect the higher rainfall intensities at Cameron Peak Fire than East Troublesome Fire, which can increase RMSE values and help explain the lack of a clear relationship between performance and RQI.

These RMSEs were slightly larger than RMSEs found in a study conducted in Canada where RMSEs were mainly less than 1 mm/hr during the summer (Moazami and Najafi, 2021). Other studies have found that MRMS overestimates rainfall (Bayabil et al., 2019; White and Nelson, 2024) or both over and underestimates dependent on location (Moazami and Najafi, 2021). Our results are aligned with those of Moazami and Najafi (2021), where biases were different based on regions (the eastern area tended toward overestimation, and the western area tended to underestimate). Many changes (e.g., algorithm updates, additional data inputs) have occurred to the MRMS QPEs since the prior studies were conducted, and each individual study area is likely to have different radar quality and density of rain gauges used to bias-correct the MRMS values. Tipping bucket rainfall data may be inaccurate due to undercatch and calibration errors; these errors can be present both in the rain gauges used to generate MRMS bias corrections and in the rain gauges we used to compare to MRMS values. Tipping buckets and the corresponding pixels may not be a perfect one-to-one comparison due to the size of the pixel (1 km) compared to the tipping bucket (point measurement) (Figure 2-9). Storms may occur within a pixel but not necessarily at the location of the gauge. For very short duration storms, the timing of the radar scans may be a source of uncertainty. For example, over a 10-minute period, exactly when the radar is sampling the specific area might affect the results.

Recognizing that both tipping buckets and radar-derived rainfall have different strengths, our third objective was to evaluate which of these two would be most useful for predicting binary streamflow occurrence. Although no significant difference was found between the two datasets, MRMS performed slightly better at CPF for predicting streamflow occurrence, while tipping buckets performed better at ETF (Figure 2-8). The radar data used to develop MRMS did not have high quality in the western area due to further distance from radar stations and more

terrain blockages compared to the eastern area, which is likely why they did not work as well for streamflow prediction. The better performance of MRMS for binary stream occurrence in the eastern area may be the result of more complete spatial coverage compared to tipping buckets. The eastern areas also generally had smaller size catchments than the western area. The predictive ability of the logistic regression may be affected by catchment size, as smaller catchments (eastern area) may only have a few MRMS pixels, whereas larger catchments (western area) average over more pixels. Since convective storms are on the order of 1 km the smaller catchments with fewer pixels could be more responsive and enhancing the predictive ability of the logistic regression.

Prior studies in the region have more reliably predicted occurrence of hillslope erosion using rainfall intensity from tipping buckets, with nearly half of the hillslopes showing substantial or better agreement most years post-fire (kappa value > 0.60; Wilson et al. 2018). This may partly be because the spatial scales of hillslopes are much smaller than the catchments we analyzed, so they would have experienced less rainfall variability. Miller et al. (2025) have also found that the streamflow occurrences in the CPF are highly variable between catchments and times of year. Stream dynamics are complex and can become even more variable post-fire. The logistic regressions we used to predict streamflow occurrence used a rainfall metric, a soil moisture proxy, and burn severity as inputs, but binary stream occurrence may depend on other factors as well. A low kappa statistic for both rainfall datasets likely indicates that the model fails to capture all the drivers of stream occurrence. If some of these other factors (e.g., observed soil moisture, soil type, vegetation regrowth) could be identified and incorporated into the model, it would help us better test which rainfall data source is more reliable for flow prediction.

Given the spatial variability in convective storm rainfall, the best source of data for streamflow prediction would still be a dense network of tipping buckets installed throughout areas of interest. However, given that such networks are often unavailable or logistically infeasible to install, MRMS provides a good representation of rainfall in places with good radar

quality. The assumption of uniform rainfall, which is often applied to small watersheds is clearly not valid for these convective storms. Without accurate, high-resolution rainfall data in areas that are susceptible to flash flooding or debris flows (e.g., post-fire landscapes), it is difficult to accurately estimate hydrologic occurrences, increasing the risk of underestimating flash flood or debris flow hazards.

2.5 Conclusions

Our study examined the influence of elevation on rainfall patterns along Colorado's northern Front Range mountains, evaluated MRMS compared to tipping bucket data, and evaluated how well these two products capture rainfall events that trigger a streamflow occurrence. Using radar-derived data, we found that the relationship between elevation and high-intensity rainfall varied across years, and the locations with the highest-intensity rainfall were not necessarily within consistent elevation bands. While some years showed higher-intensity rainfall at certain elevations, others did not, highlighting the complexity of orographic effects and local meteorological processes. Our results demonstrate that widely used gridded daily rainfall products can miss the spatial variability and intensity in warm season rainfall that radar products are able to detect.

We found that the different spatial scales between tipping buckets and MRMS led to inconsistent relationships between rainfall event metrics from the two data sources. Where radar quality (i.e., RQI) is high, MRMS may be better than a single tipping bucket per catchment for detecting storms that trigger a stream occurrence. Therefore, initial evaluation of MRMS RQI is recommended prior to using the dataset. The MRMS dataset offers excellent temporal coverage and finer spatial resolution than many other gridded rainfall products, and it captures the spatial patterns of convective storms that are missed in sparse rain gauge networks. However, its suitability may be limited in areas with complex terrain, insufficient ground data for

bias correction, significant distance from radar sites, or in studies that require a longer record length.

CHAPTER 3. Geographic variability in streamflow responses to wildfire

3.1 Introduction

After wildfires, landscapes face heightened risks of flash flooding (Moody and Martin 2001; Neale and Weir, 2015; Liu et al., 2022), increased erosion and sedimentation (Wagenbrenner and Robichaud, 2015; East et al., 2021), increased likelihood of debris flows (Nyman et al., 2011; McGuire et al., 2017), and mobilization of contaminants (Murphy et al, 2015; Murphy et al., 2018; Nunes et al., 2018; Paul et al., 2022). Such events can be life-threatening and economically damaging (Neary and Gottfried, 2002; Pausas et al., 2008). Fires can alter streamflow and water balance components for decades or more post-fire (Ffolliott et al., 1996; Niemeyer et al., 2020). Climate change is expected to drive increases in both wildfire activity (Rocca et al., 2014; Abatzoglou and Williams, 2016) and the intensity of extreme rainfall events (Prein et al., 2017, Fowler et al., 2021), which together can amplify post-fire hydrologic risks (Oakley, 2021; Touma et al., 2022). In the western U.S., wildfires have been increasing in size and severity (Dennison et al., 2014; Westerling, 2016) and advancing upslope to higher elevations (Alizadeh et al., 2015; Kampf et al., 2022). High elevation, snow-dominated mountains have historically had low fire frequency, so research on how fire affects streamflow has been limited in these areas.

Broadly, post-fire streamflow responses are affected by soils, burn severity, and topography (e.g., Staley et al., 2017 and Hallema et al., 2017). Fire can consume soil organic matter (Benavides-Solorio and MacDonald, 2005), increase soil hydrophobicity (Doerr et al., 2009), decrease soil surface roughness (Lavee et al., 1995), and reduce soil porosity (Ubeda and Outeiro, 2009). All these changes cause lower infiltration, which results in increased overland flow. Higher burn severity often leads to greater erosion rates because more soil is exposed, and more soil organic matter has been burned (Vieira et al., 2015). Steep slopes can

also be important in predicting post-fire debris flows (Cannon et al., 2010). These three factors (soils, burn severity, and topography) along with rainfall intensity are typically the primary variables considered in post-fire hydrologic (Kampf et al., 2020) and debris flow hazard models (Staley et al., 2016; Staley et al., 2017). However, field observations indicate that post-fire streamflow responses also vary geographically, with some regions tending to have much more extreme post-fire streamflow responses than others despite having similar burn severities and slopes.

This study focuses on identifying how and why streamflow responses to wildfire vary geographically. Our study area in the Colorado Rocky Mountains experiences a steep climate gradient from low snow in the plains (~1500 m) and foothills (up to ~2500 m) to high snowpack (up to ~4000 m) that lasts through the winter in the mountains, providing the primary source of streamflow in the spring and summer. The two largest fires in Colorado, Cameron Peak (CPF) and East Troublesome (ETF), both burned in 2020 through this elevation gradient, extending into higher elevations than fires previously studied in the region. Post-fire response may differ across the climate gradient from low snow to high snow zones due to variability in rainfall intensity, soil development and extent of bedrock exposure, and glaciation. In this region, flood peaks have historically been higher at lower elevations compared to high (Jarrett 1990; England et al., 2010), and some prior researchers attributed this to rainfall intensity. However, the lower elevations in this region also have less ground cover, less soil development, and greater bedrock exposure compared to high elevations (Rossi et al., 2020). These factors could also lead to greater overland flow and higher flood peaks. High elevations in this region have experienced glaciation in the past, which produced more uneven or hummocky terrain, likely resulting from frost heaving (Smith et al., 2012) or past glacial activity (Madole et al., 1998). These irregular landforms create natural depressions that disrupt continuous overland flow, leading to less direct runoff and a reduced stream response to light rainfall events (Le and Kumar, 2014). Variations in infiltration and overland flow between snow zones may also be

influenced by snow insulation and freeze-thaw dynamics. Repeated freeze-thaw cycles can decrease soil porosity and macropore connectivity (Leuther and Schulter, 2021). In contrast, deep seasonal snow cover acts as an insulator, reducing the likelihood of soil freezing at higher elevations compared to lower elevations (Harrison et al., 2021). Prior research in the CPF found that streamflow response post-fire varied between snow zones, with the higher elevation snow zone having greater total stormflow than the lower snow zone (Miller et al., 2025). However, it is unknown whether these snow zone differences occur in other fires.

In the ETF, initial post-fire floods and debris flow activity indicated that streams in this fire were much more responsive to rain than those in CPF. The two fires have different soil properties, which can influence infiltration and streamflow generation post-fire. The soils in CPF are generally sandy loams, while ETF is generally sandy clay loam and loam (Wieczorek, 2014) identified using an area- and depth-weighted averages from Natural Resources Conservations Service's (NRCS) Soil Survey Geographic database (SSURGO). By comparing the two fires, we can evaluate how both the climate gradient (snow zone) and soil affect postfire streamflow. We ask (1) how do streamflow responses to rainfall vary between fires, snow zones, and years, and (2) what underlying factors explain these differences in post-fire streamflow response? We hypothesize that the low elevation intermittent snow zone has quicker and greater streamflow response than the high elevation seasonal snow zone in both fires due to higher rainfall intensity. Further, we hypothesize that the ETF has greater streamflow responses to rainfall than CPF due to greater clay content in soils.

3.2 Methods

3.2.1 Study area

In 2020, CPF burned 844 km^2, and ETF burned 784 km^2 in northern Colorado. The CPF spanned elevations from 1615 to 3618 m, while ETF ranged from 2507 to 3319 m. The hydrologic regimes in the burned areas can be grouped based on seasonal (persistent and

transitional) or intermittent snow zones based by the fraction of time the snow is present on the ground (i.e., snow persistence). As initially defined, the persistent snow zone represents areas that maintain high snow cover during the winter (i.e., 75–100% snow persistence from January to July). The transitional snow zone represents snow persistence of 50–75%, and the intermittent snow zone is 25–50% (Moore et al., 2015; Richer et al., 2013). Both fires span the seasonal (persistent and transitional) and intermittent snow zones. Harrison et al. (2021) identified 60% as a threshold in this region where annual stream discharge increases greatly with increasing snow persistence, so in this study we use a threshold of 60% to separate the catchment into seasonal (snow persistence >60%) or intermittent (snow persistence <60%) snow zones.

Burn severity mapping shows that within CPF, 44% of the area burned at low severity, 30% at moderate severity, and 6% at high severity. For ETF, 18% burned at low severity, 46% moderate, and 15% high (Monitoring Trends in Burn Severity [MTBS], 2017; Figure 3-1). Remaining areas within the fire perimeters were unburned. Prior to burning, vegetation in the higher elevations of CPF primarily consisted of subalpine spruce-fir forests, with mixed-conifer forests at mid-elevations, and ponderosa pine forests and prairie ecosystems at lower elevations (Addington et al., 2018). In contrast, ETF was largely composed of lodgepole pines interspersed with Engelmann spruce-subalpine fir and aspens (LANDFIRE, 2016).

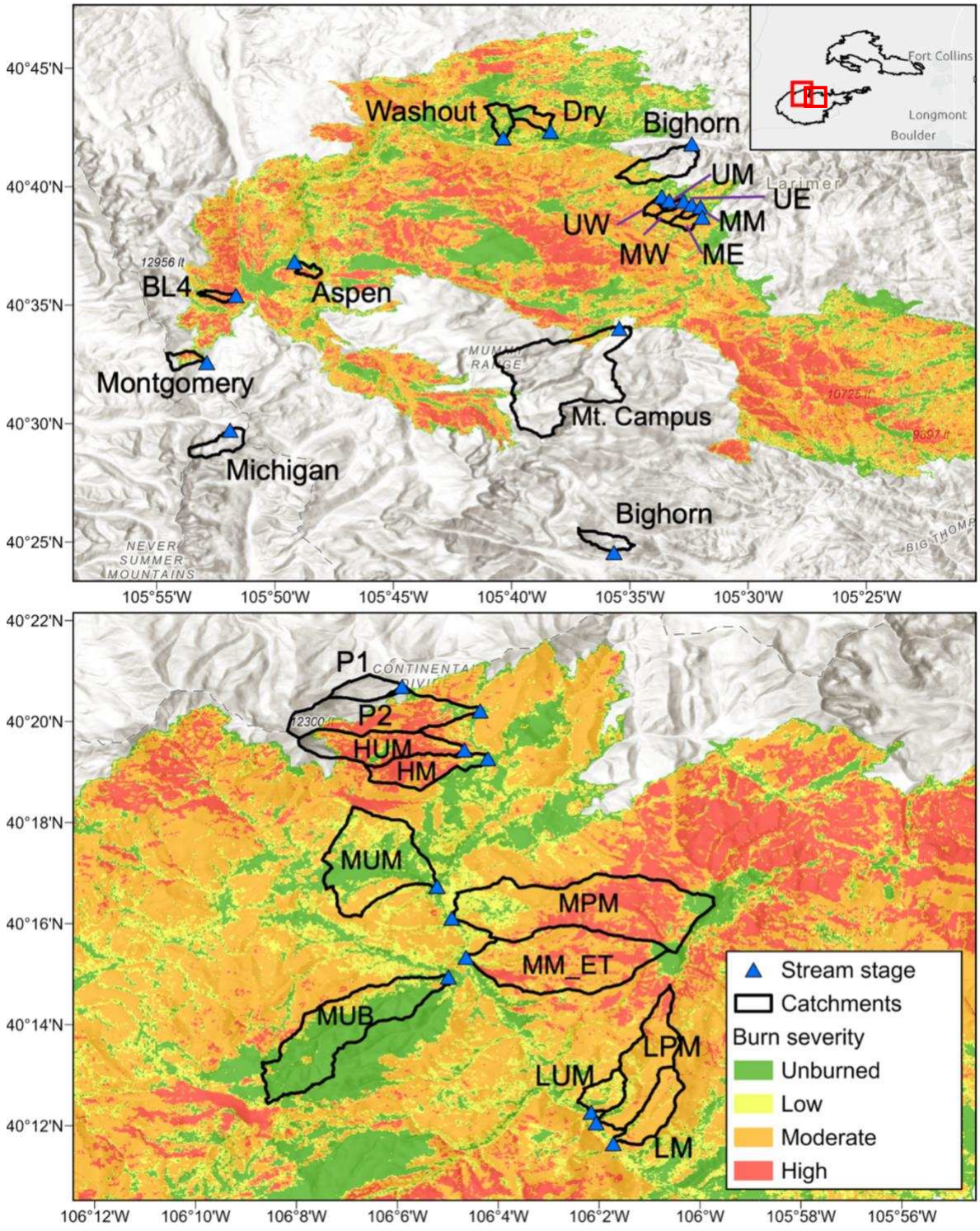

Figure 3-1. The Cameron Peak Fire (CPF, top) and East Troublesome Fire (ETF, bottom) study catchments with burn severity (MTBS, 2017) overlaid on a hillshade (ESRI, 2017). The inset map in the upper right shows the outline of CPF and ETF in respect to Boulder and Fort Collins, CO. The red square in the inset map shows the approximate location of the ETF catchments within the fire perimeter.

3.2.2 Data collection

Stream stage sensors were used to monitor 26 catchments across the two fire-affected areas and nearby unburned catchments (Table 3-1, Figure 3-1). Data collection spanned three summers post-fire (2021-2023) for CPF and two summers (2022-2023) for ETF. Catchment sizes ranged from 1 to 39 km² (Table 3-1). Stream stage was measured at the outlets of most catchments using either capacitance rods (TruTrack WT-HR 1000) or pressure transducers (In-Situ Rugged TROLL and Onset HOBO Water Level Logger). The pressure transducers were unvented and paired with barometric sensors to correct for atmospheric pressure. The stage sensors recorded on 5- or 15-minute intervals. During site visits, stream stage and the position of the channel bed relative to the stage plate were recorded to capture post-fire aggradation or erosion, and these measurements were used to adjust stage data following major stream response events. For ETF lower elevation sites, extensive channel changes made installing sensors impractical. Instead, time-lapse cameras were deployed, and stage changes during storms were estimated from photographs.

Table 3-1. Study catchment characteristics. For stage sensors, CR is capacitance rod and PT is pressure transducer. The mean snow persistence was determined from Hammond (2020), the percent area burned was determined using MTBS (2017), and elevation range from a USGS 1-m LiDAR DEM.

Catchment	Fire	Percent area burned	Area (km^2)	Elevation range (m)	Mean snow persistence	Snow zone	Stage sensor
me	CPF	99	1.44	2397–2681	52	Intermittent	CR
mm	CPF	99	0.72	2357–2640	50	Intermittent	CR
mw	CPF	98	1.50	2382–2762	53	Intermittent	CR
ue	CPF	100	0.61	2403–2650	47	Intermittent	CR
um	CPF	100	0.57	2459–2653	53	Intermittent	CR
uw	CPF	100	1.03	2457–2771	47	Intermittent	CR
bighorn	CPF	0.0	3.05	2587–3433	60	Intermittent	PT
washout	CPF	55	2.68	2357–3071	51	Intermittent	PT - Larimer
dry	CPF	70	2.91	2317–2933	49	Intermittent	PT
michigan	CPF	0.0	3.97	3165–3840	91	Seasonal	USGS
montgomery	CPF	44	1.93	3059–3482	89	Seasonal	CR
bl4	CPF	90	1.00	2907–3390	81	Seasonal	CR
aspen	CPF	73	1.13	2678–3041	66	Seasonal	CR
mtcampus	CPF	3.4	38.93	2743–4080	83	Seasonal	PT
dadd	CPF	27	7.72	2145–2718	47	Intermittent	CR
p1	ETF	0.0	1.20	2896–3456	76	Seasonal	PT
p2	ETF	59	7.04	2768–3750	77	Seasonal	PT
hum	ETF	85	4.03	2756–3724	77	Seasonal	PT
hm	ETF	95	2.56	2729–3263	70	Seasonal	PT
mum	ETF	71	7.87	2652–3233	63	Seasonal	PT
mpm	ETF	96	12.46	2637–3589	71	Seasonal	PT
mub	ETF	28	8.56	2620–3277	63	Seasonal	PT
mm_et	ETF	98	9.21	2627–3491	67	Seasonal	Camera
lpm	ETF	98	4.46	2566–3328	63	Seasonal	Camera
lm	ETF	100	2.23	2570–3007	59	Intermittent	Camera
lum	ETF	96	1.25	2565–2985	56	Intermittent	Camera

3.2.3 Rainfall and stage events

The Multi-Radar Multi-Sensor System (MRMS) product, based on S-band dual-polarization WSR-88D radar and corrected with precipitation gauge data, provides precipitation estimates at 1 km resolution (Zhang et al., 2016). We used the Surface Precipitation Rate (SPR) product, available at two-minute intervals, to estimate precipitation by applying correction factors

derived from the relationship between radar only quantitative precipitation estimate (QPE) and the Multi-Sensor QPE. The MRMS data, with 0.1 mm precision, were accessed through the Iowa Environmental Mesonet (2024). The 2-minute rainfall rasters were summed to 10-minute timestep for faster computing. For the MRMS rainfall events, the catchment average from pixels within or overlapping the catchment boundary were used to determine rainfall event metrics. A new event was initiated when more than 0.3 mm of rainfall fell after at least six hours with no precipitation. Only events with greater than 1 mm of precipitation were included in the analysis. For each rainfall event, total depth of precipitation (event rainfall), duration, and maximum rainfall intensity for 60 minutes (MI60) was calculated. The MI60 was selected over other rainfall metrics based on Miller et al. (2025), who found that MI60 had highest correlation with stream responses at some of the study sites.

The two streamflow variables we analyzed are stage rise and lag to peak. Stream stage was used instead of stream discharge to identify stream responses to rain events due to the difficulty of developing rating curves in burned streams. Many of the catchments often experienced channel changes during rainfall events, which led to large uncertainty in the rating curves developed. Stream responses were summarized by their peak and minimum stage values, which were used to calculate the stage rise. Only events with a total stage rise of 2 cm or greater were retained. The stage and rainfall event datasets were joined, and lag to peak was determined as the lag from the time when half of the rainfall had fallen to the time of maximum stage. The analysis was limited to the June through September period to exclude the majority of snowmelt events. For high elevation sites, some events identified in June still had diurnal snowmelt responses, and these were removed from the dataset.

3.2.4 *Question 1: how do streamflow responses to rainfall vary between fires, snow zones, and years?*

To address the first research question, we compared the streamflow response variables, stage rise and lag to peak, between fires and snow zones. This was done by using a two-

sample t-test or Wilcoxon-rank sum test, depending on whether the data were normally distributed. Stage rise was not normally distributed, so we used the Wilcoxon-rank sum test to compare the medians. Lag to peak was normally distributed, so we used two-sample t-tests to compare the means. When comparing stage rise and lag to peak across years, we used a one-way ANOVA (for lag to peak) or Kruskal-Wallis test (for stage rise) since there were multiple years. The one-way ANOVA compared the means between groups, whereas the Kruskal-Wallis evaluates whether the distribution of stage rise differs between years by ranking the combined data and comparing the sum of ranks between groups. Next, since the streamflow responses are likely tied to rainfall characteristics, we compared the relationship between stage rise and MI60 between fires.

The basic comparison of streamflow response variables is insufficient on its own to verify whether streamflow responses are significantly different between fires and snow zones. Rainfall post-fire varies over time, and the responses also vary through time and space as vegetation regrows on the burned landscape. Since each fire and snow zone had a different sequence of rain events and time-varying post-fire responses, we needed a model to represent the interactions between rain, snow zone, fire, and time post-fire. We used a Bayesian generalized linear mixed-effects models to represent how the streamflow response variables (stage rise and lag to peak) interact with rainfall intensity (MI60), fire (CPF and ETF), snow persistence zone (seasonal and intermittent), and year post-fire (2021-2023). The MI60, fire, snow zone, and year were represented as fixed effects (i.e., predictors whose influence was assumed to be constant across all sites and years). To account for hierarchical structure of the data, random intercepts were included for site and year to capture heterogeneity among sites and interannual variability. Year was included as both a fixed effect and a random intercept to capture both a general trend over time and year-specific deviations due to interannual differences. A random slope for MI60 was specified by site to allow the rainfall-stage relationship to vary across sites, recognizing that local hydrologic conditions can influence stream response (Table 3-2).

Table 3-2. Fixed effects, random intercepts, and random slope for each Bayesian model used in the study.

Model	Fixed effects	Random intercepts	Random slope
Fire/snow zone/year	MI60, fire, snow zone, year	site, year	MI60 by site
Normalized difference vegetation index (NDVI)	MI60 x year x NDVI, fire, snow zone	site, year	-
PWD	MI60 x PWD, fire, snow zone, year	site, year	-
Principal component analysis (PCA)	MI60 x PC1, MI60 x PC2, year	site, year	-

Based on the structure of the data, the model used a Gamma distribution and log link for both stage rise and lag to peak. The independent variables were scaled (mean = 0, standard deviation = 1) prior to modeling to allow direct comparison of coefficients across variables and to improve model convergence. A total of five Markov chain Monte Carlo chains were run for 5,000 iterations each. Convergence was assessed using the potential scale reduction factor (equal to 1.0) and the effective sample size diagnostics (greater than 20% of the post-warm up iterations). Posterior distributions of fixed effect coefficients were visualized using the mean and credible intervals (66%, 95%, and 100%). The posterior distributions represent the effects on the log of expected stage rise or lag to peak per one standard deviation increase in the independent variable. This approach allowed us to test whether the stage rise and lag to peak were different across fires, snow zones, and year post-fire. The R package *rstanarm* was used to conduct the Bayesian modeling (Goodrich et al., 2024). Model performance and potential overfitting was evaluated using approximate leave-one-out cross validation (Vehtari et al., 2017) using the *loo* R package. All Pareto *k* estimates were ≤ 0.7 after targeted refitting, indicating reliable leave-one-out cross-validation diagnostics.

3.2.5 Question 2: what underlying factors explain differences in streamflow response?

Some of the differences in streamflow response may be caused by time-varying conditions, like the rate of vegetation regrowth and the antecedent soil moisture prior to each rain event. To evaluate the influences of these factors, we again used Bayesian generalized linear mixed-effect models to assess how vegetation recovery (peak annual normalized differenced vegetation index [NDVI]) and potential water deficit (PWD) influenced the relationship between MI60 and stream response. These were done in two separate models since we were limited on model terms due to the relatively small sample size (n=247). The NDVI is the ratio between near-infrared and red bands, and it gives an indication of vegetation greenness. Peak summer NDVI values were determined for the pixels within a catchment for each year in the study. The NDVI calculations were completed using Landsat 8 imagery within Google Earth Engine. For the NDVI-specific model, a three-way interaction among MI60, year, and NDVI was included to assess whether vegetation and interannual variability jointly modified rainfall-runoff responses. Fire and snow zone were also included as fixed effects. Random intercepts for site and year were included to capture the spatial and temporal variation not explained by the fixed effects (Table 3-2).

To represent time-varying soil moisture conditions, we used the PWD from gridMET (Abatzoglou, 2011), which represents daily precipitation minus reference evapotranspiration. To approximate wetting and drying conditions, we computed cumulative PWD for each water year starting on 1 October. For the PWD-specific model, antecedent moisture conditions represented by PWD were included as an interaction with MI60 to test whether drier or wetter conditions modulated stream response to rainfall. Fire, snow zone, and year were included as additional fixed effects. Random intercepts for site and year were incorporated (Table 3-2). The remaining model setup for both the NDVI and PWD-specific models were consistent with the previous model above.

Differences in streamflow responses may also relate to inherent differences between the study catchments. These differences include burn severity, soil properties, and catchment morphology (Table B-2).

Table 3-3All the catchments were delineated using a U.S. Geological Survey 1 m LiDAR DEM (USGS, 2024), and each catchment variable was averaged across the drainage area. The burn severity variable identified was mean differenced normalized burn ratio (dNBR), which is defined by the difference of the normalized burn ratio pre-fire and post-fire. The normalized burn ratio is the ratio between near-infrared and shortwave infrared bands (MTBS, 2017).

The soil variables identified were percent clay and percent sand, which relate to the infiltration capacity of the soil. For each soil variable we used the depth-weighted averages of the USDA NRCS's SSURGO variables from the Wieczorek (2014) dataset. This dataset provides summaries of soil properties weighted by both areal extent and depth within the soil profile (0-200 cm). The dataset spatial resolution is 30 m but the underlying soil map units are from SSURGO polygons, which are often much coarser and irregular in shape. Once soil variables were extracted for each catchment, an area weighted average value across the polygons within each catchment was determined for each soil variable.

Catchment morphology variables included fraction of areas containing valleys and hollows, fraction of area greater than 75% of the elevation range, and maximum flow path length. These were computed using a U.S. Geological Survey 1 m LiDAR DEM. Catchments can be described based on the landforms they contained using geomorphons, which identify flat, peak, ridge, shoulder, spur, slope, hollow, footslope, valley, and depression landforms (Jasiewicz and Stepinski, 2013). We computed the fraction of area in each catchment represented by valleys and hollows, assuming that these would be the landforms that would slow or detain flow before it reaches the catchment outlet, which would reduce streamflow magnitude and extend the lag to peak. Another way to describe the catchment shape is through elevation-area distributions, or hypsometric analysis. We computed hypsometric curves for each

catchment and computed the fraction of area above 75% of the elevation range of the catchment. The fraction of catchment area at different elevations may affect how quickly water is routed to streams. The length of the catchment streams was determined using the maximum flow path length from the D8 flow direction algorithm. Longer flow paths should extend the time it takes for runoff to reach the outlet and allow for more infiltration leading to a smaller amount of streamflow. The hypsometric analysis, fraction of areas that are valleys and hollows, and maximum flow path length, were determined using the *WhiteboxTools* R package (Wu and Brown, 2022).

To reduce the number of predictor variables among these catchment attributes, a principal component analysis (PCA) was performed to identify the most informative predictors. This approach was necessary because testing each catchment variable in the Bayesian models was not feasible due to the relatively small sample size (n=247), which would have limited statistical power and increased the risk of overfitting (Babyak, 2004). The six variables contributing most to the cumulative variance explained were prioritized, and the stability of the first two principal component loadings were assessed using bootstrapped singular value decomposition (Fisher et al., 2014). Following the PCA, the first two principal components (PC1 and PC2) were standardized and included in a Bayesian generalized linear mixed-effect model with a Gamma distribution and log link. The model included interactions between MI60 and each principal component, as well as a term for year. Random intercepts for site and year were included to account for site and temporal variability (Table 3-2). The remaining model setup was consistent with the first model explained. This PCA-based approach allowed us to test the combined effects of multiple catchment variables simultaneously while maintaining statical power in the Bayesian models.

Table 3-3. Catchment variables used in the study.

Type	Name	Variable	Unit	Source
Climate	60-minute maximum rainfall intensity	MI60	mm/hr	MRMS
Climate	Potential water deficit	PWD	mm	gridMET
Climate	Mean snow persistence	SP	%	Hammond, 2020
Soil	Percent clay	Clay	%	SSURGO, Wieczorek 2014
Terrain morphology	Fraction of area above 75% elevation	Area > 75% elevation	-	USGS 1 m LiDAR, WhiteboxTools
Terrain morphology	Fraction of area that are valleys and hollows	Valleys & hollows	-	USGS 1 m LiDAR, WhiteboxTools
Terrain morphology	Maximum flow path length	Max flow length	m	USGS 1 m LiDAR, WhiteboxTools
Land cover	Mean dNBR	Mean dNBR	-	MTBS, 2017
Land cover	NDVI	NDVI	-	Landsat 7/8, Google Earth Engine
-	Year post-fire	Year	-	-
-	Site number	Site number	-	-

3.3 Results

3.3.1 Rainfall events and stream response

The first question was how streamflow response varies between fires, snow zones, and years. When we examined this question using group comparisons (e.g., Wilcoxon rank sum-test), the intermittent snow zone had greater median stage rise than the seasonal snow zone, but the difference was not statistically significant (p-value = 0.08). The ETF had greater stage rise compared to CPF (p-value < 0.001; Figure 3-2). Across the study period, the median stage rise in the CPF intermittent snow zone was 6.9 cm, while the ETF intermittent snow zone had a median stage rise of 30.0 cm. In the seasonal snow zone, CPF had a relatively small median stage rise (4.0 cm) compared to ETF (23.5 cm). The CPF intermittent, ETF intermittent, and ETF seasonal snow zones all displayed large interquartile ranges (29.2 cm, 50.0, and 40.5 cm, respectively), indicating considerable variability. At CPF, median stage rise in the intermittent snow zone increased from 5.1 cm in 2021 to 20.6 cm in 2022, followed by a decrease to 6.2 cm

in 2023. However, these changes were not statistically significant (p-value = 0.07). In the CPF seasonal snow zone, median stage rise remained relatively consistent each summer. For ETF, data were only available for 2022 and 2023, but both the intermittent and seasonal snow zones showed decreases in median stage rise between those years. The ETF intermittent snow zone declined from 45.0 cm to 30.0 cm, and the seasonal snow zone from 31.3 cm to 11.3 cm. While both ETF snow zones showed decreases between 2022 and 2023, only the seasonal snow zone decrease was significant (p-value = 0.002).

Overall, the lag to peak at the intermittent snow zone was shorter than at the seasonal snow zone (p-value < 0.001; Figure 3-2). At CPF and ETF, the intermittent snow zone had a mean lag to peak of 1.1 hours. In contrast, the seasonal snow zones showed longer mean lag to peak (2.9 hours). At the CPF intermittent snow zone, the mean lag to peak values were 1.4 hours in 2021, 0.8 hours in 2022, and 0.7 hours in 2023 (p-value > 0.05). At the ETF intermittent snow zone, mean lag to peak decreased from 2.1 hours in 2022 to 0.4 hours in 2023, and this decrease was statistically significant (p-value = 0.03). For the CPF seasonal snow zone, the mean lag to peak slightly increased (4.0, 4.6, and 4.9 hours) over the study period but was not statistically significant (p-value = 0.74). The ETF seasonal snow zone also showed an increase in mean lag to peak, rising from 1.3 hours in 2022 to 1.8 hours in 2023 but was not statistically significant (p-value = 0.3).

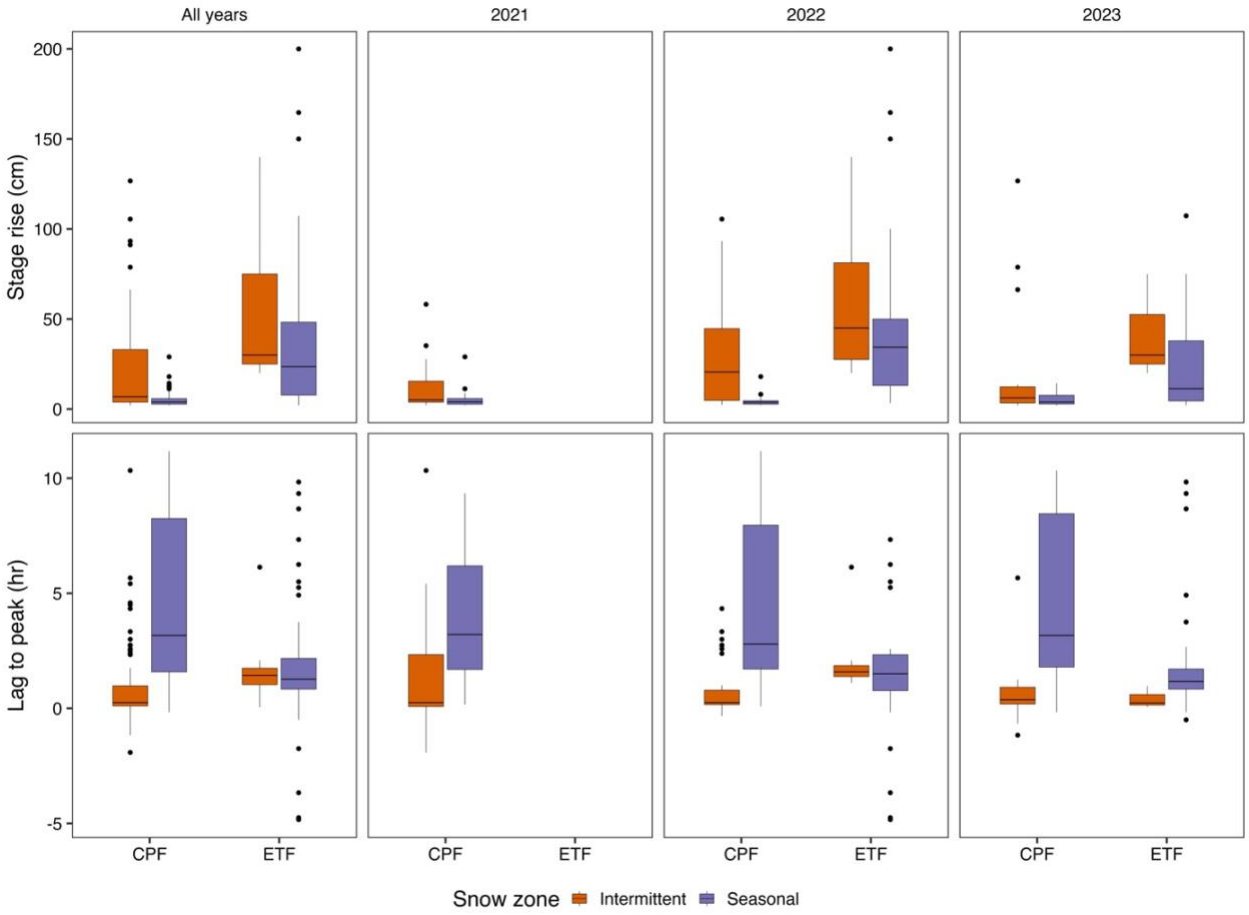

Figure 3-2. Stage rise and lag to peak for each fire and snow zone for the study period (2021, 2022, and 2023). The ETF stream sensors were not installed until 2022.

For rainfall events that produced a stream response, the ETF catchments exhibited larger stage responses at lower rainfall intensities compared to CPF (Figure 3-3). The CPF intermittent snow zone had the highest average MI60 (10.5 mm/hr), followed by the CPF seasonal snow zone (7.4 mm/hr). In contrast, the ETF intermittent and seasonal snow zones had similar MI60 values, but both were lower on average (4.2 and 4.4 mm/hr, respectively) than CPF. Despite experiencing lower rainfall intensities, the ETF catchments were more responsive than those at CPF. For example, the largest ETF stage response was 200 cm in response to 4.1 mm/hr, while the largest at CPF was 127 cm in response to 12.7 mm/hr. The ETF intermittent snow zone responses were much higher than CPF intermittent snow zone for rainfall intensities less than 5 mm/hr. The ETF seasonal snow zone had much higher responses across all

intensities compared to the CPF seasonal snow zone, with little evident relationships between intensity and stage rise.

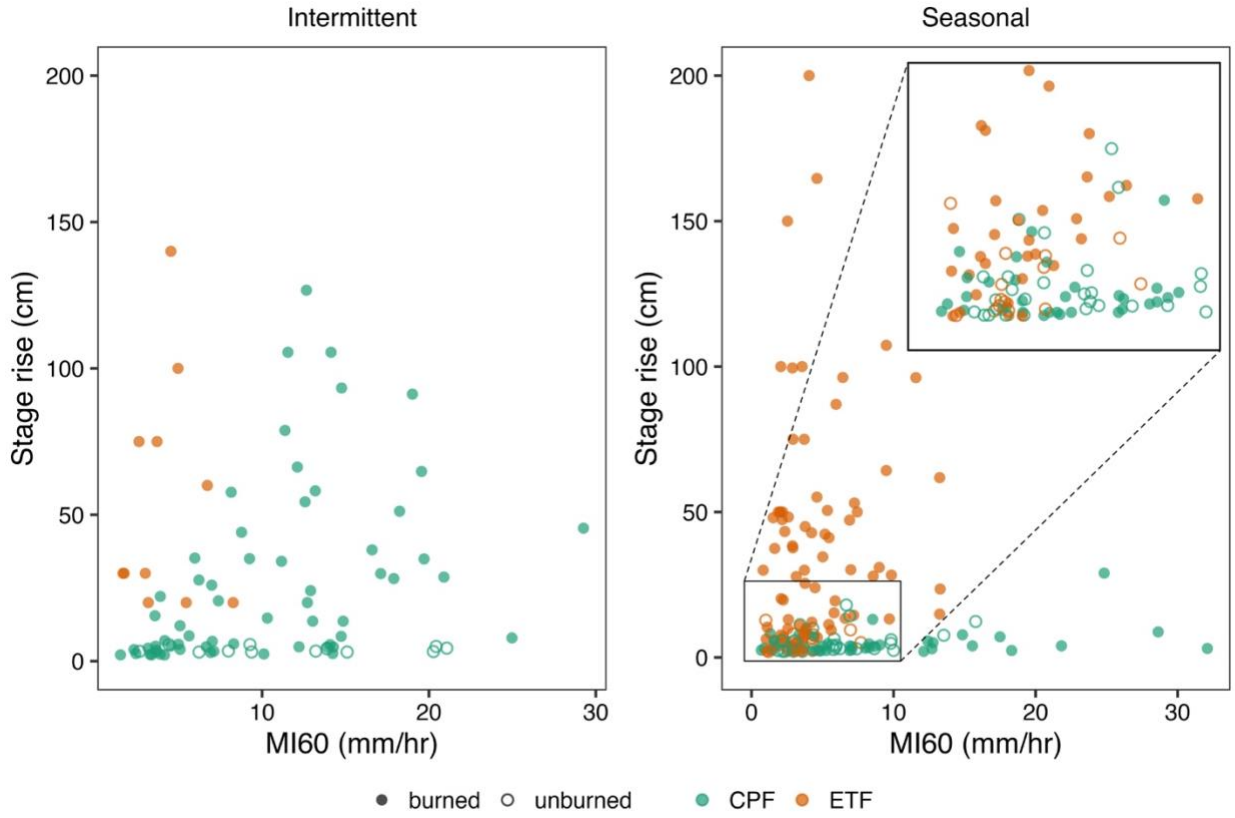

Figure 3-3. The 60-minute maximum intensity (MI60) and stage rise during the study period for the intermittent and seasonal snow zones at CPF and ETF. Open circles represent the unburned sites and filled circles burned sites.

3.3.2 Fire, snow zone, and year effect on stream response

The Bayesian model results indicate that stage rise was strongly influenced by rainfall intensity (MI60), fire (ETF vs. CPF), and snow zone (intermittent vs. seasonal) (Figure 3-4). Each predictor (y-axis in Figure 3-4) has a probability distribution of coefficient values on a log-scale. Exponentiating the mean scaled coefficient (β), gives the multiplicative effect on stage rise per standard deviation increase in the predictor. Where the coefficient values are positive, this means that increase in the predictor increases the stage rise. Greater MI60 increased stage rise (β = 0.45, 95% CI [0.26, 0.65]), corresponding to approximately 57% greater stage rise per one standard deviation increase. The ETF sites exhibited substantially larger stage rise than

CPF (β = 1.50, 95% CI [0.79, 2.18], 4.5 times greater). The intermittent sites tended to have larger stage rise (β = 0.68, 95% CI [-0.03, 1.35]; 97% greater) than the seasonal snow zone. There was no clear evidence for a temporal trend in stage rise (β = -0.07, 95% CI [-1.35, 1.12]), as the probability distribution of the coefficients is centered near zero.

Lag to peak was influenced strongly by MI60 and snow zone (intermittent vs. seasonal) (Figure 3-4). Greater MI60 was associated with slightly shorter lag to peak (β = -0.13, 95% CI [-0.32, 0.06]), corresponding to 12% shorter lag to peak per standard deviation increase. Fire did not have a strong effect on lag to peak (β = -0.15, 95% CI [-0.87, 0.60]). Sites in the intermittent snow zone exhibited shorter lag to peak than those in the seasonal snow zone (β = -0.91, 95% CI [-1.62, -0.17]; 60% shorter). There was no clear temporal trend in lag to peak (β = -0.08, 95% CI [-0.74, 0.57]).

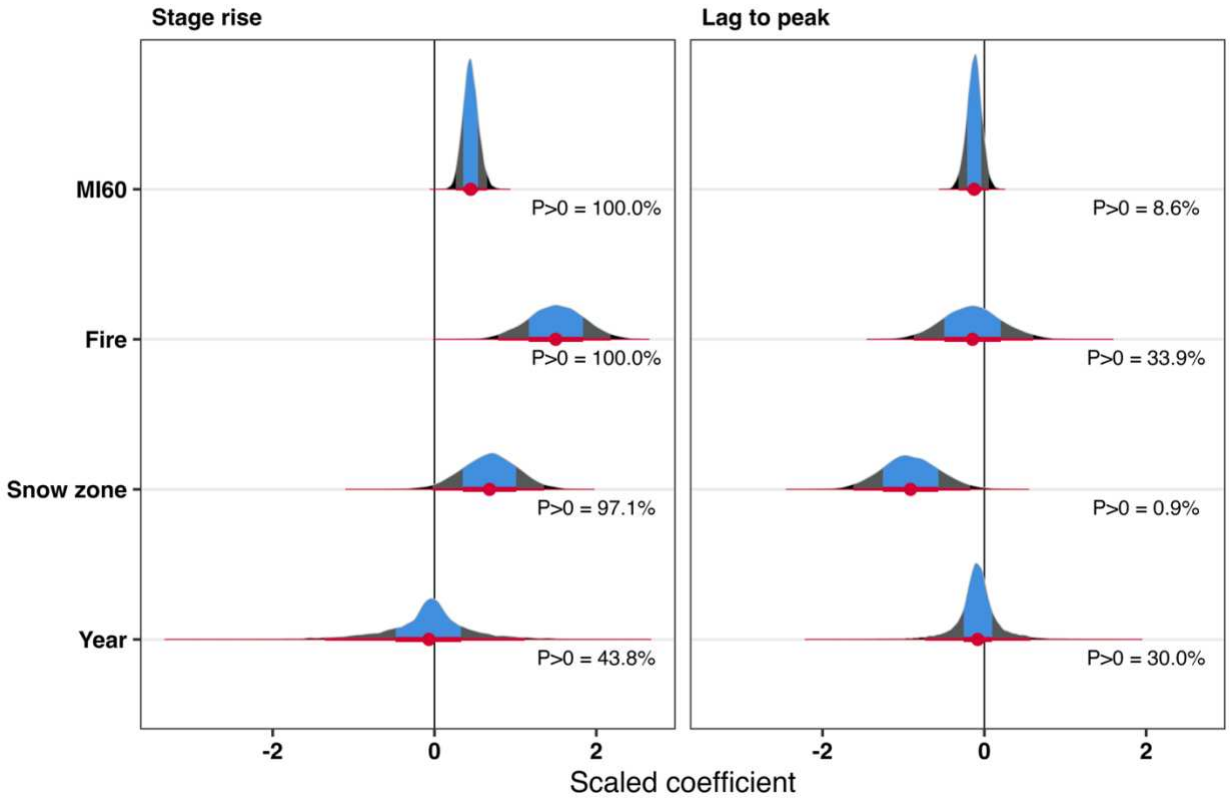

Figure 3-4. Stage rise and lag to peak scaled coefficients from a Bayesian generalized linear mixed-effects model. The point represents the mean (β) of the posterior distribution. The blue is the 66% credible interval, grey 95% credible interval, and the black is the 100% credible interval. The link thickness also indicates 66, 95, and 100% credible intervals with the thickest

being 66% and the thinnest 100%. The P>0 represents the proportion of posterior samples where the coefficient was greater than 0.

3.3.3 NDVI and PWD effect on stream response

The burned sites across both fire and snow zone showed vegetation recovery, reflected in increasing NDVI, each summer post-fire (Figure 3-5). The unburned sites had relatively consistent NDVI across all three summers. The CPF burned intermittent sites mean NDVI increased from 0.30 in 2021 to 0.47 in 2023. This is almost the same NDVI as the CPF unburned intermittent site in 2023 (0.53). The CPF burned seasonal sites mean NDVI increased from 0.32 in 2021 to 0.44 in 2023. The CPF unburned seasonal sites mean NDVI in 2023 was 0.41. For ETF, the intermittent burned sites mean NDVI increased from 0.27 in 2021 to 0.48 in 2023. The seasonal burned sites mean NDVI increased from 0.30 in 2021 to 0.39 in 2023. The ETF unburned seasonal sites mean NDVI ranged from 0.58 to 0.60.

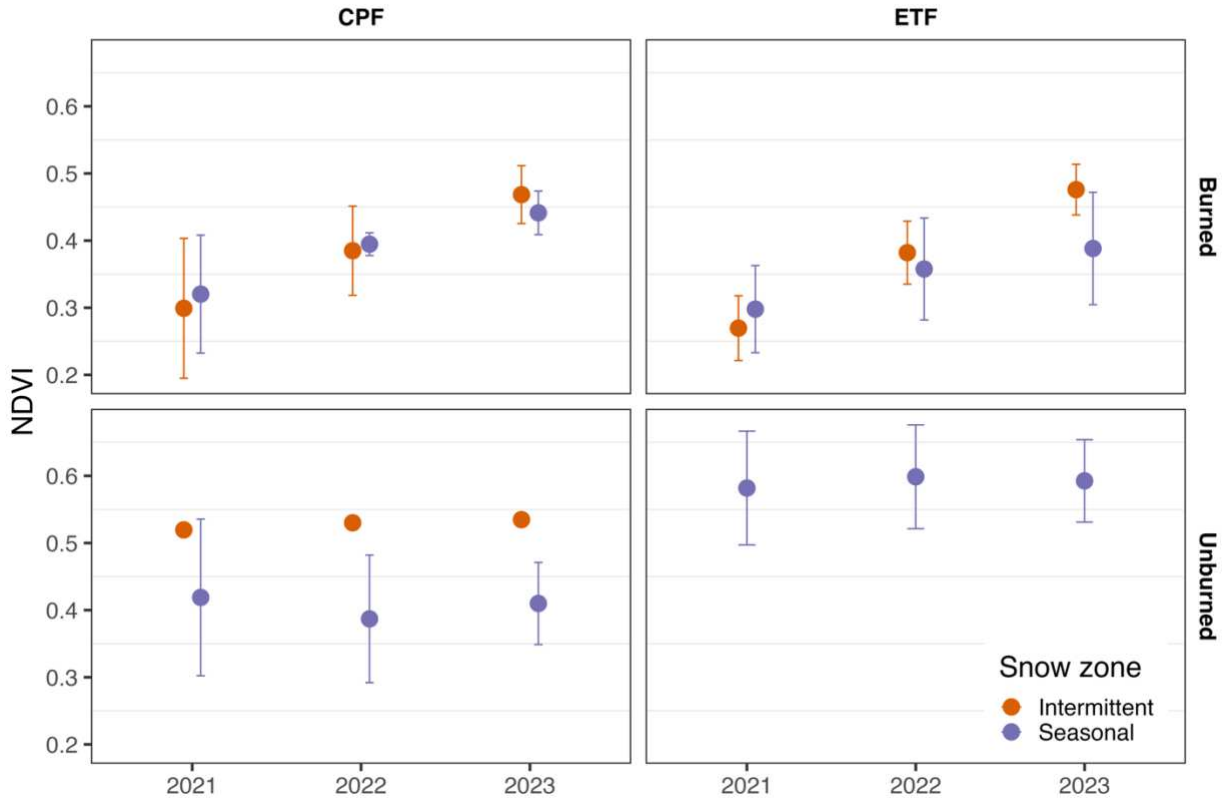

Figure 3-5. The NDVI for each summer post-fire for CPF and ETF within the intermittent (orange) and seasonal (purple) snow zones. The points represent the mean and the error bars one standard deviation from the mean. The CPF unburned intermittent sites do not have error

bars since there is only one site within that category. There are no unburned intermittent sites at ETF.

In the NDVI model, greater NDVI decreased stage rise (β = -0.25, 95% CI [-0.50, 0.05]), corresponding to a 22% reduction in stage rise per one standard deviation increase of NDVI. The effect of MI60 was slightly reduced at higher NDVI (MI60 × NDVI: β = -0.17, 95% CI [-0.32, -0.02]; 10% smaller per one standard deviation NDVI), so the effect of MI60 on stage rise is slightly weaker when NDVI is greater. Later years with higher NDVI tended to decrease stage rise (year*NDVI: β = -0.15, 95% CI [-0.33, 0.02]; 15% lower per standard deviation NDVI per standard deviation year), so in areas with higher NDVI, stage rise may not increase as much over time. The three-way interaction (MI60*year*NDVI: β = 0.07, 95% CI [-0.10, 0.25]; 8% increase) suggests that the combined effect of MI60, year, and NDVI slightly increases stage rise, but is uncertain.

For lag to peak, the model indicates that higher NDVI tended to increase lag to peak (β = 0.31, 95% CI [0.06, 0.56]; 36% longer per standard deviation increase). The MI60-NDVI interaction term tended to slightly increase lag to peak (β = 0.15, 95% CI [0.00, 0.32]; 17% longer), so sites with greater NDVI tend to have slightly longer lag to peaks per one standard deviation increase. The year-NDVI interaction term had a negligible effect (β = 0.05, 95% CI [-0.15, 0.25]; 5% longer), indicating little evidence of a combined effect, meaning changes over time in lag to peak do not meaningfully depend on NDVI. The three-way interaction (MI60*yearCont*NDVI: β = 0.2, 95% CI [0.03, 0.41]; 24% longer) suggests a potential small, combined effect of rainfall, year, and NDVI on lag to peak. In other words, in later years, high MI60 in areas with higher NDVI could be associated with slightly longer lag to peaks.

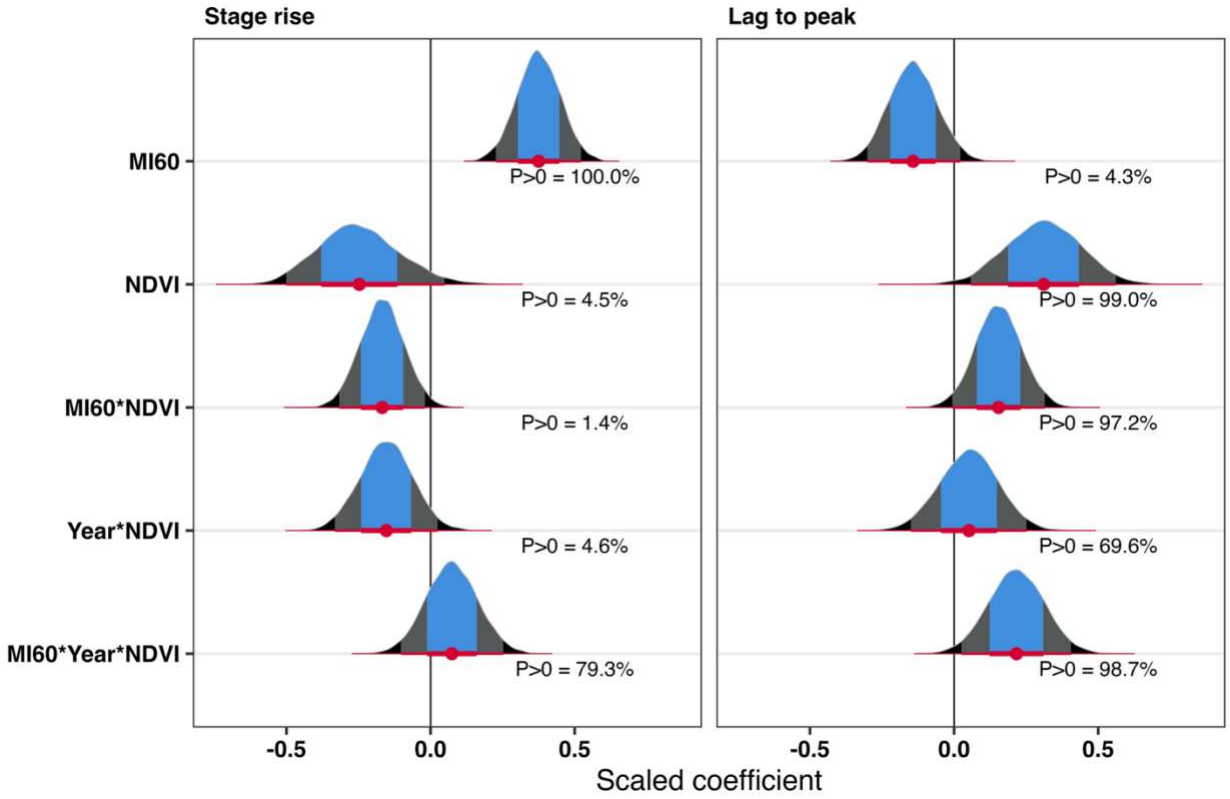

Figure 3-6. Stage rise and lag to peak scaled coefficients from the NDVI model. The point represents the mean (β) of the posterior distribution. The blue is the 66% credible interval, grey 95% credible interval, and the black is the 100% credible interval. The link thickness also indicates 66, 95, and 100% credible intervals with the thickest being 66% and the thinnest 100%. The P>0 represents the proportion of posterior samples where the coefficient was greater than 0.

The PWD and the interaction between PWD and MI60 had smaller effects (negative and positive, respectively) on stage rise than MI60 (Figure 3-7). The PWD had a β of -0.14 and a 95% credible interval of [-0.40, 0.12]. The interaction between MI60 and PWD was small and uncertain (β=0.03, 95% CI [-0.09, 0.15], suggesting limited evidence of a combined effect of rainfall and antecedent moisture on stage rise. Similar to stage rise, the PWD and interaction between PWD and MI60 had a smaller effect on lag to peak (Figure 3-7). The PWD had a β of 0.05 and a 95% credible interval of [−0.35, 0.24], indicating weak evidence of an effect. The interaction between MI60 and PWD was small but less uncertain (β=0.07, 95% CI [-0.07, 0.22].

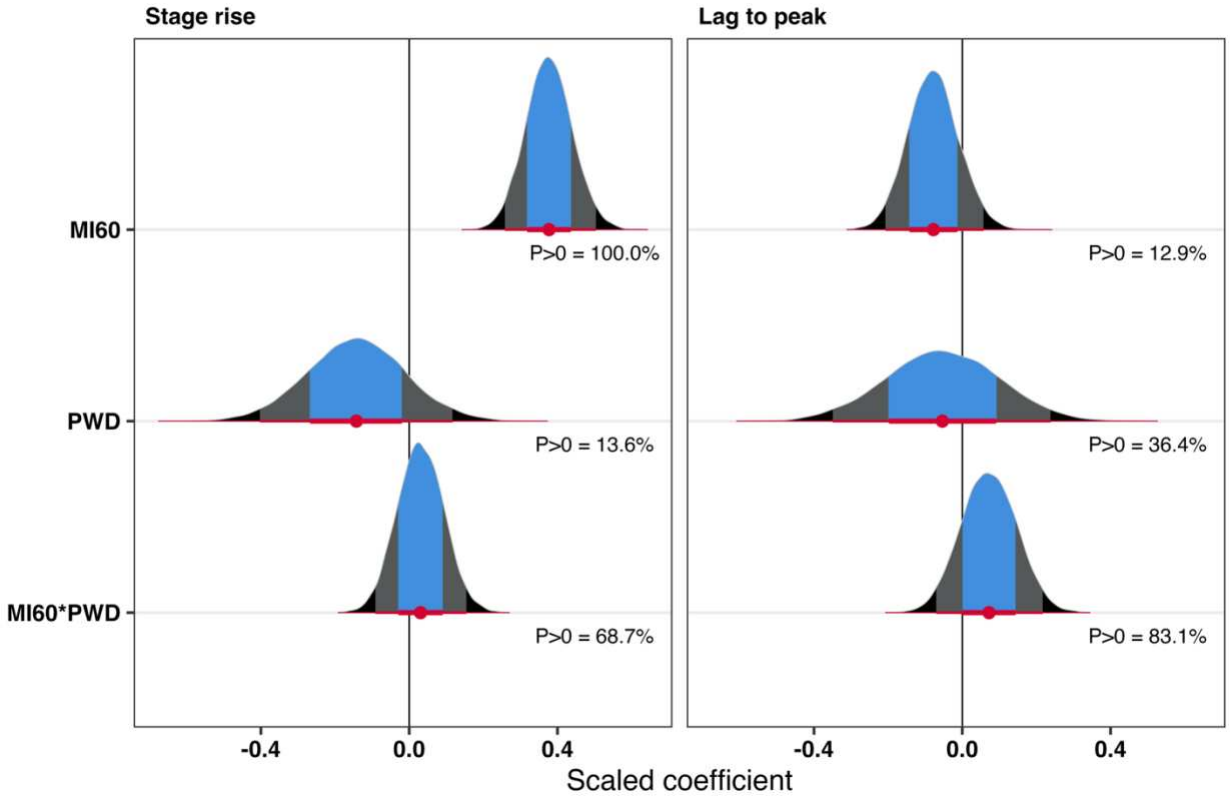

Figure 3-7. Stage rise and lag to peak scaled coefficients from the PWD model. The point represents the mean (β) of the posterior distribution. The blue is the 66% credible interval, grey 95% credible interval, and the black is the 100% credible interval. The link thickness also indicates 66, 95, and 100% credible intervals with the thickest being 66% and the thinnest 100%. The P>0 represents the proportion of posterior samples where the coefficient was greater than 0.

3.3.4 Catchment variables

The percent clay, percent sand, and fraction of valleys and hollows varied by fire. The CPF intermittent and seasonal snow zones had average clay contents of 12 and 10% and average sand contents of 68 and 65%. In contrast, the ETF intermittent and seasonal snow zones had higher average clay contents of 21 and 17% and lower average sand contents (51 and 57%). The average fraction of valleys and hollows was greater in CPF (average of 0.15 [intermittent] and 0.16 [seasonal]) compared to ETF (average of 0.11 [intermittent] and 0.12 [seasonal]). In both fires, the average proportion of area above 75% elevation was consistently higher in the intermittent snow zone (CPF = 0.174; ETF = 0.155) compared to the seasonal snow zone (average: CPF = 0.060; ETF = 0.063). Maximum flow path lengths varied widely,

especially in the CPF seasonal snow zone, which had the highest average value (1152 m). In contrast, the intermittent snow zones showed similar average maximum flow path lengths, with 579 m at CPF and 573 m at ETF. The mean dNBR was similar across the CPF intermittent snow zone and ETF intermittent and seasonal snow zones (348, 335, and 359, respectively). The CPF seasonal snow zone mean dNBR was 195.

Using the stepwise PCA approach, percent clay, percent sand, maximum flow length, dNBR, fraction of area containing valleys and hollows, and fraction of area greater than 75% elevation were selected (Figure 3-8). Percent clay, fraction of area containing valleys and hollows, and sand had the largest loadings for PC1, with clay having a negative loading and the others positive. For PC2, dNBR, fraction of area greater than 75%, and maximum flow length had the largest loadings with flow length negative and the others positive. PC1 represented approximately 42% of the cumulative variance and PC2 was approximately 27%.

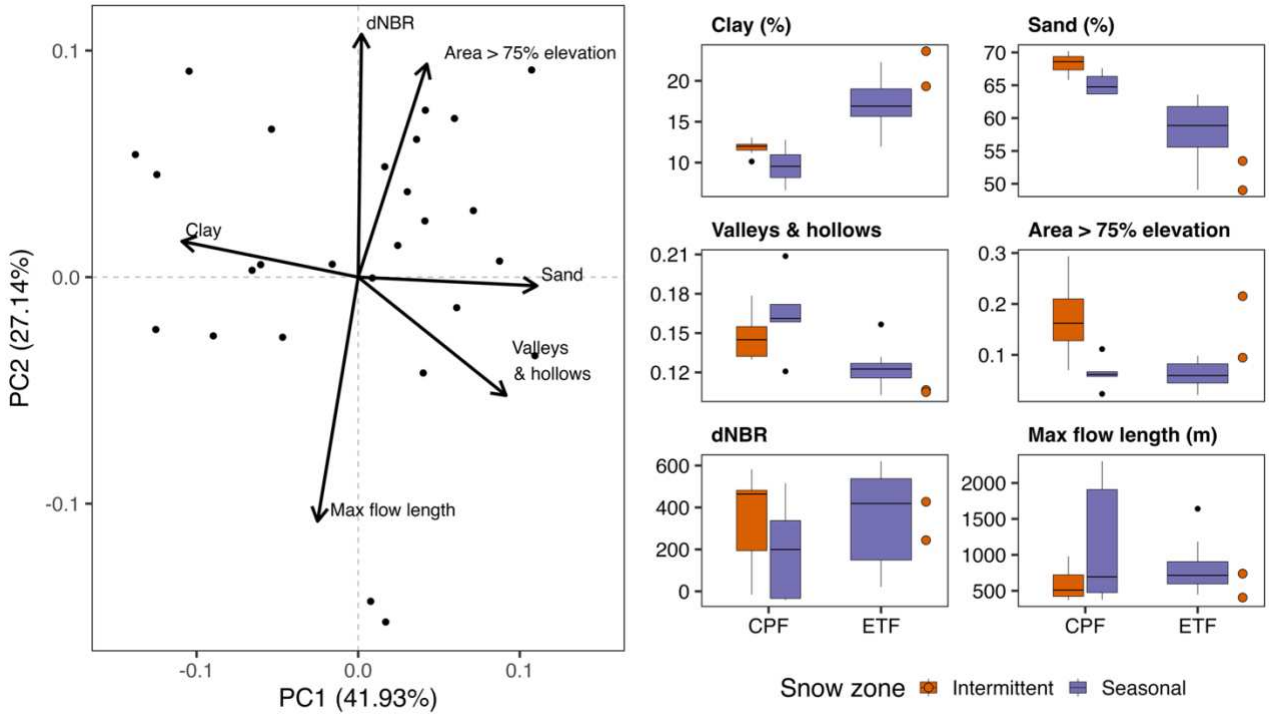

Figure 3-8. The biplot for the stepwise PCA including PC1 and PC2. Catchment variables for the intermittent and seasonal snow zones across both fires that were selected based on the PCA. The ETF intermittent snow zone had two sites and are represented by points instead of a boxplot.

In the PCA-based model, PC1 (catchments with more sand and valleys and hollows) had a negative effect on stage rise, and PC2 (catchments with greater burn severity and area greater than 75% elevation) had a positive effect (Figure 3-9). An increase in PC1 would lead to a decrease in stage rise (β = -0.65, 95% CI [-0.92, -0.37]; 48% decrease per standard deviation increase in PC1). This suggests that catchments with sandier soils and more valleys and hollows tend to have smaller stage rise. The interaction between MI60 and PC1 was negative (β = -0.18, 95% CI [-0.32, -0.04]; 17% decrease), indicating that the effect of MI60 on stage rise was weaker in catchments with sandier soils and more valleys and hollows.

PC2 increased stage rise (β = 0.61, 95% CI [0.30, 0.93]), equivalent to an estimated 85% increase in stage rise per standard deviation increase in PC2. In this analysis, PC2 reflects catchments with higher burn severity and more area greater than 75% of elevation, which suggests that catchments with those properties tend to have greater stage rise. The interaction between MI60 and PC2 was positive (β = 0.13, 95% CI [0.00, 0.25]; 14% increase), suggesting a possible enhancement of the MI60-stage rise relationship in areas with greater burn severity and larger area greater than 75% of elevation.

The PC2 and interaction between MI60 and PC2 (catchments with greater burn severity and area greater than 75% elevation) had the strongest effect on lag to peak. Higher PC1 values were associated with a decrease in lag to peak (β = -0.17, 95% CI [-0.50, 0.16]). The interaction of PC1 and MI60 had a very small negative effect but was uncertain. The PC2 had a negative effect on stage rise (β = -0.42, 95% CI [-0.81, -0.04]), indicating catchments with greater burn severity and area greater than 75% elevation tended to have shorter lag to peak. The interaction between PC2 and MI60 was negative (β = -0.31, 95% CI [-0.50, 0.01]), indicating that the effect of MI60 on lag to peak was weaker in catchments with greater burn severity and area greater than 75% elevation.

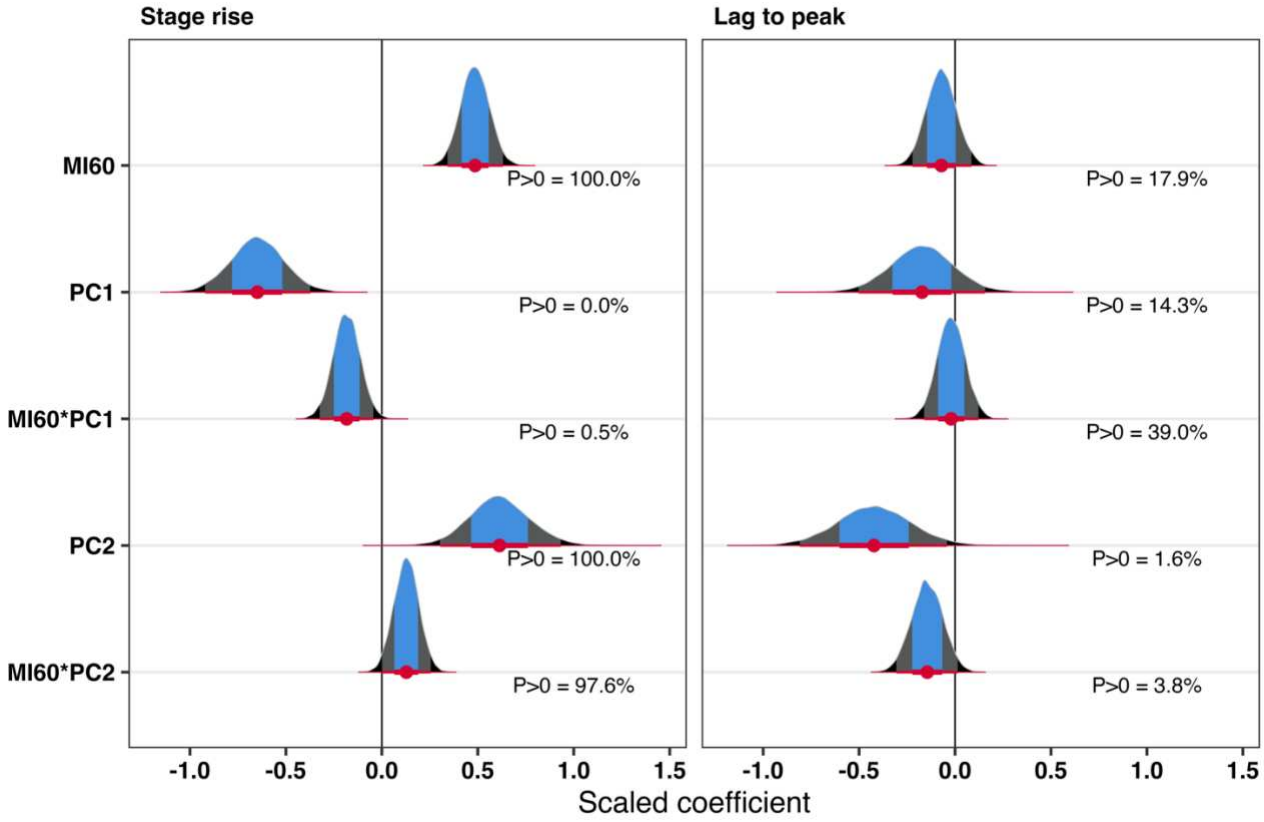

Figure 3-9. Stage rise and lag to peak coefficients from the PCA-based model. The blue is the 66% credible interval, grey 95% credible interval, and the black is the 100% credible interval. The link thickness also indicates 66, 95, and 100% credible intervals with the thickest being 66% and the thinnest 100%. The P>0 represents the proportion of posterior samples where the coefficient was greater than 0.

3.4 Discussion

Rainfall intensity emerged as the dominant driver of both stage rise and lag to peak in both fires, but fire location, snow zone, vegetation, and catchment morphology also affected post-fire hydrologic response in northern Colorado. Overall, the study highlights the importance of considering additional catchment characteristics beyond rainfall intensity, burn severity, and slope in post-fire modeling.

3.4.1 Does stream response vary by fire?

The stream response differences between CPF and ETF catchments were expected due to their contrasting rainfall intensities, soil properties, and catchment morphology (Figures 3-3 and 3-8). The ETF sites had greater clay content, lower fractions of valleys and hollows, but

lower rainfall intensities overall. Despite these lower rainfall intensities, the ETF catchments had more than a fourfold greater stage rise compared to CPF catchments. The differences in rainfall intensities between ETF and CPF are likely attributed to the mountain-plains feedback cycles that enhance convection on the east side of the Rockies. The higher stage rises at ETF catchments despite lower intensities (Figure 3-3) could be caused by lower infiltration into the soils due to their higher clay content. This could lead to infiltration excess overland and more runoff reaching the channel. Additionally, the smaller fraction of valleys and hollows in ETF indicates the catchments have fewer places where water can be detained. Fast infiltration excess overland flow reaching streams would cause the large spikes in stage rise. However, lag to peak was not substantially different between the two fires even with their contrasting soil types and terrain morphology (Figure 3-2 and Figure 3-4), which means the role of valleys and hollows could have been to cause more infiltration of water that did not reach the stream outlet. If the valleys and hollows had just slowed the flow, we would have expected slower lags to peak in CPF.

3.4.2 Does stream response vary by snow zone?

The differences between intermittent and seasonal snow zone stream response highlighted the role of hydroclimatic regimes shaping stream response post-fire. The intermittent snow zones had larger stage rise and shorter lag to peak times (Figure 3-2 and Figure 3-4). These response differences between the snow zones were expected since prior research has shown an abrupt shift in hydrologic regime between the two snow zones in this region (Hammond et al., 2018; Harrison et al., 2021). Although there was a clear difference in stream response between the snow zones, the ETF seasonal snow zone had much greater stage rise than the CPF seasonal snow (Figure 3-2 and Figure 3-3). Again, this could be attributed to the higher clay content of soils. Intermittent snow zone catchments in CPF had greater average proportion of area above 75% elevation compared to seasonal snow zone catchments. Since

the highest elevation portions of catchments tend to have steep slopes, this could lead to greater streamflow generation over a larger area of steeply sloped terrain. This could cause greater stage rise and shorter lag to peak time in the intermittent snow zones. In this study, the mean slope for each catchment was considered but was not used, as it had very weak correlation with stream response. We therefore assumed that proportion of area with slopes greater than 30 degrees was more predictive of the role of slope in streamflow response. Based on the importance of the proportion of area above 75% elevation, the location of steeper slopes within a catchment may be a better metric to represent slope. The CPF rainfall intensity varied by snow zone with the intermittent snow zone having greater rainfall intensities, but this was not the case at ETF. Therefore, the response differences between snow zones include more than the hydroclimatic regime (e.g., snowpack, soil moisture), but also the catchment soil properties and morphology.

3.4.3 Does stream response vary by year?

Several studies have found that post-fire hydrologic impacts are most prevalent within the first two years after fire (e.g., Hallema et al., 2017; Saxe et al., 2018; Ebel et al., 2020) but can persist for a decade or longer (Ffolliot et al., 1996; Niemeyer et al., 2020). We found mixed results on temporal trends after fire. Trends by year were not evident for either stage rise or lag to peak, but the change in vegetation cover between years did affect both stream response variables. Higher NDVI (i.e., more vegetation cover) was associated with reduced stage rise and longer lag to peak (Figure 3-6), both of which align with expected vegetation-stream response interaction (Dunne and Leopold, 1978). With increased vegetation, we would expect enhanced interception and increased soil infiltration capacity, which would reduce runoff and slow water movement. The NDVI increased over the three years post-fire at all burned catchments (Figure 3-5). The MI60-NDVI interaction term showed a weakened rainfall effect at higher NDVI, reinforcing the assumption that vegetation is acting as a buffer resulting in slower and smaller

streamflow responses (Figure 3-6). The fact that NDVI did affect the stream response variables, whereas year did not may reflect the role of interannual climate variability and the limited information provided by year alone. NDVI values reflect the vegetation conditions during each year, whereas year on its own provides less useful information.

3.4.4 What factors influence stream response?

The previous sections discussed how rainfall intensity, soil properties, terrain morphology, and vegetation regrowth influence stream response. In contrast, contrary to hydrological expectations, there was weaker and uncertain evidence that antecedent moisture conditions influence stream response (Figure 3-7). For stage rise, PWD had a negative effect, which was contrary to expectations. However, the majority of the larger rainfall events occurred later in the summer season when soil drying and ET were the greatest. This may explain why the model identifies a negative relationship between MI60 and PWD. Additionally, the PWD may not adequately capture conditions due to the coarse spatial resolution and using a spatial average over the catchment. The PWD also does not represent soil moisture directly; it uses precipitation and reference evapotranspiration. The reference ET value does not account for differences in vegetation types or in how transpiration varies with soil drying. Other methods to account for soil moisture, including remote sensing and land-surface models (e.g., Soil Moisture Active Passive and North American Land Data Assimilation System), have demonstrated good agreement with in-situ sensors. (Ford and Quiring, 2019). However, their coarse spatial resolution (> 9 km) limits their applicability at the catchment scale for this study, as a single pixel would exceed the size of most catchments in this study. Future studies could test alternate ways of characterizing antecedent moisture (e.g., standardized precipitation evapotranspiration index and Palmer drought severity index).

The PCA analysis revealed how the many catchment variables interact to affect streamflow response (Figure 3-9). Sandier catchments with more valleys and hollows (PC1)

produced smaller stage rise indicating that post-fire stream responses are more muted in permeable soils and landscapes that slow or detain water. Sandier soils increase infiltration capacity, while valley and hollows act as localized storage zones that slow hillslope to channel water movement. These features can reduce runoff even under high rainfall intensities. Effects on lag to peak were weaker, suggesting that these catchment features influence magnitude more than timing. For PC2, severely burned catchments with limited proportion of area above 75% elevation and short flow paths generate rapid runoff. To summarize, PC1 buffers stream response and PC2 amplifies it.

Other studies have found dramatic post-fire response differences across the western U.S., but they did not identify the primary reasons for those differences (Moody and Martin, 2009; Saxe et al. 2018). Based on the complex results found between fires, it is important to continue evaluating drivers of post-fire stream response across different fires.

3.4.5 Limitations

In this study, there are several limitations that should be considered. First, there was an imbalance of ETF intermittent sites (two) compared to CPF intermittent sites (ten), which could skew the statistical results more heavily towards trends seen at the CPF intermittent snow zone when comparing by snow zone. Additionally, there were more seasonal snow zone catchments (nine) at ETF than intermittent (two). Several of the ETF catchments were just above the 60% snow zone threshold. Based on field observations alone, it was assumed that a few of the ETF catchments that were grouped as seasonal were intermittent catchments suggesting that categorical division by snow zone may be limiting. When comparing across fires, CPF included three summers post-fire, and ETF only included two. The amount of burned versus unburned catchments also was not balanced within fires or snow zones.

Second, we were not able to use stream discharge as a response variable due to frequent channel changes at several burned sites. While stage rise and stream discharge tend to be

correlated, the stage rise depends on the channel geometry. Two catchments with equal stage rise values could have completely different stream discharge depending on channel width and depth. We experimented with strategies for adjusting stage rise by channel geometry (e.g., estimated channel width using LiDAR), and these tests gave similar findings to those using stage rise alone. We also tried normalizing stage rise by catchment area and those tests results in unrealistic findings. Additionally, we estimated the stage offset based on bed elevation change to the best of our ability, but there is still uncertainty. For example, if the bed elevation changes, then the amount of flow response indicated by a given stage rise could change. We tried to minimize this uncertainty by removing any apparent stage changes caused by bed elevation change. There were channel changes (in addition to bed elevation) that occurred throughout the study at the burned catchments, which could introduce bias in the modeling results.

Finally, the MRMS rainfall data introduces uncertainty in rainfall timing and magnitude, especially at ETF where the radar quality is reduced (Chapter 2). We chose to use MRMS for this study so that the precipitation data source would be consistent across all sites, but future work could repeat this analysis using tipping bucket data to examine how rain data source affects results.

3.5 Conclusions

This study explored how post-fire streamflow responses, both magnitude and timing, differed by fire, snow zone, and year post-fire and identified factors that cause these differences. The CPF catchments have less clay and more valleys and hollows compared to ETF, leading to more infiltration and more potential for water detention within catchments. The result is less responsive streams in CPF compared to ETF. Because multiple factors vary between these two fires, we cannot distinguish which are more responsible for the greater responsiveness of ETF

streams to rainfall. Likely the combination of both created the notable differences in streamflow responses for the two fires.

When comparing between snow zones, the intermittent snow zone had higher rainfall intensities, greater stage rise, and shorter lag to peak than the seasonal snow zone. The effect of MI60 on stage rise was greater in the intermittent snow zone. The shorter lag to peaks in the intermittent snow zone is consistent with infiltration excess overland flow. Locations with infiltration excess overland flow are most likely to experience rapid and potentially very large flash floods, and these are key locations to identify for post-fire flood risks.

With the expected increase in wildfire fire activity and extreme rainfall events and in combination with fires increasing in size, severity, and advancing upslope, it is important to understand how different hydrologic regimes will respond. In this study, we found clear differences between streamflow responses and physical catchment attributes across fires and snow zones. Common post-fire debris flow or erosion models often only consider burn severity, soil, slope, vegetation cover, and rainfall. However, these results suggest catchment geology and morphology also play a role in stream response.

CHAPTER 4. Gridded versus station datasets yield divergent hydrologic partitioning in simulations of a partially burned Colorado watershed

4.1 Introduction

In mountainous portions of the western US, streamflow from snowmelt produces 70% of the total runoff (Li et al., 2017). In this region, streamflow is highly sensitive to spatial differences in meteorological variables, soil properties, and land cover, challenging the ability to forecast streamflow accurately. Streamflow is also highly sensitive to land disturbances, such as wildfires. Hydrologic response to wildfires is complex and can result in either a decrease or increase in water yield (Goeking and Tarboton, 2020). Increased water yield may be caused by greater water input due to reduced canopy interception (Williams et al., 2019), reduced infiltration due to burned soil organic matter (DeBano, 2000; Moody and Ebel, 2012), and reduced transpiration caused by loss of vegetation (Nolan et al., 2013). Reduced water yield is also possible due to increasing soil evaporation (Poulos et al., 2021), greater mid-winter snowpack ablation, and earlier snow free dates (Burles and Boon, 2011; Harpold et al., 2014; Giovando and Niemann, 2022; Kampf et al., 2022; McGrath et al., 2023).

One approach to studying the complex interactions between climate, disturbance, and streamflow is through hydrologic modeling. Hydrologic models use meteorological input data combined with information on vegetation and soils to simulate streamflow. A distributed hydrologic model represents spatial variability in the climate and land properties, allowing it to simulate how disturbances in different parts of a watershed will impact streamflow. However, in mountain regions, the spatial patterns of climate may be difficult to represent accurately.

Input data for distributed hydrologic models can come directly from in-situ sensors and gauges, which should be the most accurate values of precipitation at or near the gauge. However, station data are spatially limited, offering only point measurements at a few locations.

The spatial coverage of in-situ sensors can be especially limiting in mountainous terrain where access is difficult, and varying microclimates are common. Regular maintenance and calibration of sensor networks is essential, requiring site access and funding for staff to maintain the sensors. Another common limitation of station data is sensor or datalogger failure, which can cause gaps in records, as well as sensor errors that lead to inaccurate measurements. Errors can arise from environmental factors (e.g., undercatch) or technological difficulties (e.g., bad wire). Bridging the gap between point-scale measurement and regional scale processes is an ongoing research challenge (Gentine et al., 2012), and many researchers have recommended increases in sensor networks (e.g., Partridge et al., 2024).

An alternate source of input to distributed models is gridded meteorological datasets. These have been derived from station data but extrapolated across the landscape in a grid for ease of use, uniform spatial coverage, and full temporal coverage (Mankin et al., 2025). The Parameter-elevation Regressions on Independent Slopes Model (PRISM) and Daymet datasets are widely used daily gridded meteorological datasets in hydrologic studies. PRISM calculates daily precipitation, air temperature, dew point temperature, and vapor pressure deficit using climate-elevation regression equations. The stations used in the regressions are weighted based on physiographic factors (e.g., aspect, topographic barriers, coastal effects). The PRISM dataset has a spatial resolution of 4 km, and a higher-resolution 800 m product has more recently become available at no cost (Daly et al., 2008). Daymet uses a weighted linear regression approach to account for elevation, using a truncated Gaussian filter that adjusts weights based on station density. The daily Daymet dataset includes day length, precipitation, incident shortwave radiation flux density, air temperature, snow water equivalent (SWE) and water vapor pressure. The Daymet dataset spatial resolution is 1 km (Thornton et al., 1997).

While gridded datasets are ideal for inputs in spatially distributed hydrologic models, they can introduce systematic biases, especially in mountainous terrain where they may not capture spatial variability and extremes (Alvarez et al. 2014; McEvoy et al. 2014; Behnke et al.,

2016; Henn et al., 2018; Walton and Hall, 2018; Serrano-Notivoli et al., 2021; Mankin et al., 2025). For example, large precipitation differences, up to several hundred millimeters per year (accumulated November to May and cumulative annual), have been noted in the Colorado Rocky Mountains (regional scale [Gutmann et al., 2012] and pixel-based [Livneh et al., 2014]) and across the western United States (Henn et al., 2018) when using the PRISM dataset compared to observed data, the Weather Research and Forecasting model (Skamarock et al., 2019), and other gridded datasets. Since gridded meteorological datasets are often informed by station data, areas with sparse station coverage, especially those at higher elevation, often have greater uncertainties due to extrapolating values from lower elevations (Henn et al., 2018; Serrano-Notivoli and Tejedor, 2021). Precipitation gauges often underestimate actual precipitation due to undercatch (Rasmussen et al., 2012), which can propagate additional error into gridded precipitation products. For air temperature gridded products, errors may result from spatial and temporal variability in temperature-elevation lapse rates. Air temperature typically decreases with increasing elevation, but the rate of the decrease (i.e., lapse rate) varies both between locations and over time (Rolland, 2003; Lundquist and Cayan, 2007; Blandford et al., 2008; Shen et al., 2016; Navarro-Serrano et al., 2018; Lute & Abatzoglou, 2020; Collados-Lara et al., 2021). Cold air drainage, which is common in mountains, can reverse the air temperature-elevation relationship (Navarro-Serrano et al., 2018; Collados-Lara et al., 2021), and these inversions may not be accurately captured in gridded products (McEvoy et al., 2014). PRISM and Daymet have daily temporal resolutions, so they do not capture event-based measurements such as rainfall intensity.

Understanding the biases and uncertainties in meteorological input data is important in hydrologic modeling since these propagate into simulated variables like evapotranspiration (ET), snowmelt, and streamflow (Mizukami et al., 2014; Tercek et al., 2021). The values of calibrated model parameters can also be significantly impacted by the choice of climate datasets dataset (Hossain and Anagnostou, 2005; Maggioni et al., 2012; Elsner et al., 2014; Mizukami et al.,

2014). Consequently, a model calibrated using inaccurate input data may be representing hydrologic processes incorrectly as well. Little guidance exists on which meteorological data sources and types are optimal for distributed hydrologic modeling in mountain regions. Few studies have evaluated how meteorological input data and their inherent biases influence changes in model behavior and performance.

The Fourmile Creek watershed, located in the Front Range foothills in Colorado, is a snowmelt-dominated headwater stream that is ideal for testing how different climate data sources affect streamflow simulations. Fourmile Creek has a high-density network of meteorological sensors across an elevation gradient. The watershed was partially burned and thus offers an opportunity to calibrate model parameters for both burned and unburned conditions. Last, streamflow measurements are available over a ten-year study period that includes large interannual climate variability (e.g., extreme flood, drought). Leveraging the high density of meteorological sensors, partial burn status, and the streamflow dataset with strong interannual variability, we aim to (1) determine biases between gridded and station meteorological data, (2) assess differences in model parameters between calibration sets using gridded and station data, (3) use a leave-one-out resampling technique for observed streamflow to assess calibration uniqueness and parameter uncertainty, and (4) evaluate hydrologic responses and quantify model performance, including wildfire effects of heterogeneous burned and unburned areas.

4.2 Site information

The Fourmile Creek watershed ranges in elevation from 1750 to 3510 m (Figure 4-1) and has a contributing area of 63 km^2. The mean annual precipitation in the watershed ranges spatially from 500 to 1000 mm, with the greatest amounts being at the highest elevations. Approximately 23% of the watershed burned in the Fourmile Canyon fire in 2010 (Figure 4-1; Murphy et al., 2015). The fire perimeter covered 24 km^2, of which 25% burned at low severity,

33% at moderate, and 28% at high (MTBS, 2017). The pre-fire land cover was Foothill Pseudotsuga-Pinus ponderosa forest (77-79%), shrubland/grassland (16-19%), and developed land (<1%; Ebel et al., 2012; Murphy et al., 2015). The vegetation across the watershed is aspect dependent. Ponderosa pine (Pinus ponderosa P. Lawson & C. Lawson) and Rocky Mountain juniper (Juniperus scopulorum Sarg.) are on the south-facing slopes, and aspen (Populus tremuloides Michx.), Rocky Mountain Douglas-fir [Pseudotsuga menziesii (Mirb.) Franco var. glauca (Beissn.) Franco], and limber pine (Pinus flexilis E. James) are on north-facing slopes (FEST, 2010). The watershed geology consists of Precambrian metamorphic and granitic rocks that were intruded with stocks and dikes during the Tertiary period (Murphy et al., 2003). Soils are mainly coarse textured, sandy soils that are poorly developed, shallow, and well-drained (Beganskas, 2012).

Fourmile Creek has runoff hydrographs driven primarily by snowmelt amount and timing from different elevations. Baseflow is generally from October to February, and peak flows (snowmelt runoff and mixed rain-snow events) are April to June (Murphy et al. 2015). The watershed spans three snow zones. The persistent snow zone represents areas that maintain high snow cover during the winter (i.e., 75–100% snow persistence January to July). The transitional snow zone represents snow persistence of 50–75%, and intermittent snow zone is 25–50% (Moore et al., 2015). A total of 11% of the watershed is located in the persistent snow zone, 24% in transitional, and 65% in intermittent (Hammond, 2020). Summer rainfall events (June-September) also drive streamflow variability and peaks.

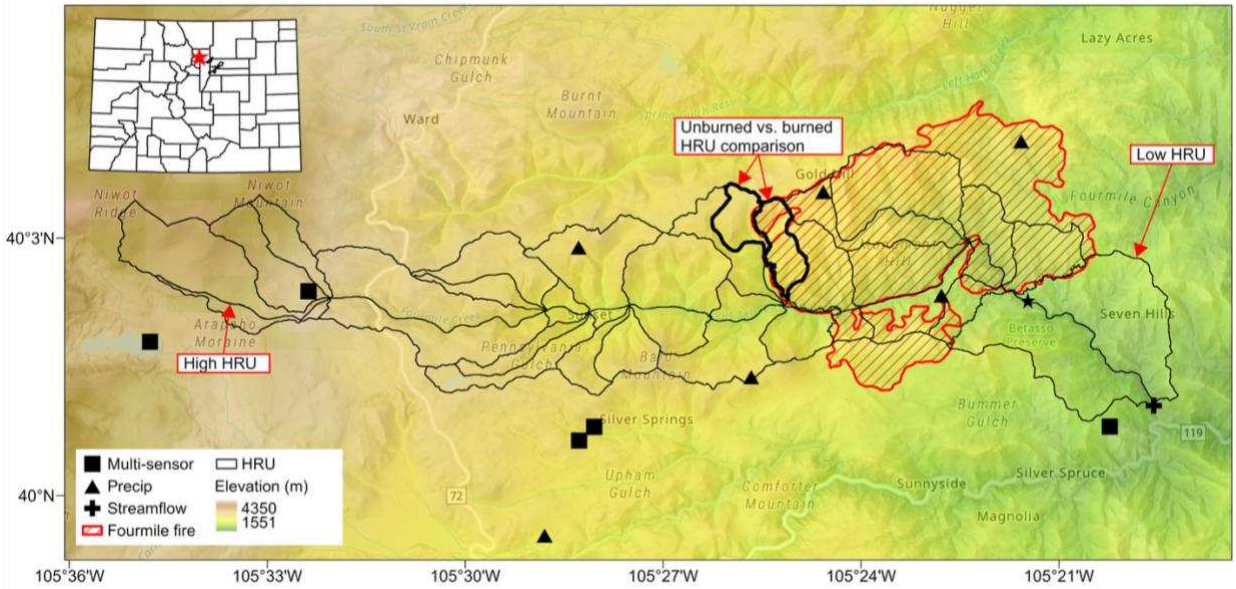

Figure 4-1. Fourmile Creek watershed with hydrologic response units (HRUs) used in the Ages model. The sensor locations are represented by black shapes. The SNOTEL sites are the two black squares on the most western portion of the figure. The inset map represents the location in Colorado Fourmile Creek is located. The black star is the location is the location of the diversion.

4.3 Methods

4.3.1 Daily meteorology: Gridded and station data

For the study period (2011-2020), the USGS 06727500 Fourmile Creek at Orodell, CO gauge is used in the calibrations and to assess the model performance in simulating streamflow. The gauge collects daily streamflow data from April through September. For meteorological sensor inputs, 11 precipitation gauges, 5 air temperature sensors, 4 relative humidity (RH) sensors, 4 incoming shortwave radiation, and 4 wind speed sensors were incorporated (Table 4-1, Figure 4-1). The 11 proximal precipitation gauges (or four directly within the watershed boundary) represent a station density of 175 (or 95) per 1000 km^2, greatly exceeding the minimum observational network recommendation of 1 to 4 per 1000 km^2 (Mankin et al., 2025). Most of the precipitation gauges are tipping buckets, which measure only rain. The only sensor data for snow precipitation was from the SNOTEL stations located at high elevations. To supplement the winter precipitation at across the elevation range, PRISM was substituted for station/sensor data for November through May.

The gridded datasets included PRISM daily data for air temperature, precipitation, and RH and Daymet for incoming shortwave radiation. Seven PRISM (4 km) and Daymet (1 km) pixels are located within or nearby the watershed. In this study, we used PRISM meteorological variables as the primary gridded dataset since it has a more sophisticated interpolation scheme. The PRISM dataset does not include incoming shortwave radiation so Daymet was supplemented for this variable. A Daymet day is defined based on local time from midnight to midnight, whereas PRISM defines a day from 12:00 UTC to 12:00 UTC. Because of this, the PRISM precipitation, air temperature, and RH data were shifted one day earlier. All seven pixels were incorporated into the Ages spatially distributed daily hydrologic model (Ascough et al., 2012; Green et al., 2015). The Ages model interpolates meteorological data and assigns values to each hydrologic response unit (HRU). To assess spatial differences or biases, the gridded and station meteorological variables at each HRU were differenced (HRU station minus HRU gridded value) for mean values (temperature, incoming shortwave radiation, and RH) and annual cumulative values (summer precipitation). When evaluating precipitation differences, we only focus on summer precipitation (May through September) because most of the precipitation gauges did not measure snow precipitation.

Table 4-1. Station data used for Ages meteorological inputs. NRCS is Natural Resources Conservation Service, UC Boulder is University of Colorado Boulder, UDFCD is Urban Drainage and Flood Control District, CZ Net is the Critical Zone Collaborative Network, and NADP is the National Atmospheric Deposition Program.

Station name	**Elevation (m)**	**Managed by**	**Variables**	**Years**
University	3139	NRCS	P, T	2010-2020
Niwot_Snotel/Niwot	3021	NRCS/UC Boulder	P, T, RH, SWin, U	2011-2020
ST	2648	UDFCD	P	2011-2020
SP	2603	UDFCD	P	2011-2020
Gordon_north	2568	CZ Net	P, RH, T, SWin, U	2012-2020
Gordon_south	2545	CZ Net	P, RH, T, SWin, U	2012-2020
GH	2525	UDFCD	P	2011-2020
Sugarloaf	2524	NADP	P	2011-2015
SE	2295	UDFCD	P	2011-2015
LM	2097	UDFCD	P	2011-2015
Betasso	1951	CZ Net	P, RH, T, SWin, U	2011-2015

4.3.2 Watershed model description, calibration, and evaluation

Streamflow modeling was completed using the Ages (version 1.0) model, a semi-distributed, process-based hydrologic model (Ascough et al., 2012; Green et al., 2015). The watershed was split into 38 HRUs based on topography using a 10 m DEM from the USGS National Map database (USGS, 2023a; USGS, 2023b) The DEM was used to derive slope, elevation, aspect, and contributing area layers. Using the centroid of each HRU, gridded and station daily meteorological data were interpolated using inverse distance weighting. The catchment areas delineation tool (Cadel) is used to create topological routing files (Kipka et al.,

2019). The Cadel tools allow for multiple flow directions and interactions between adjacent HRUs.

The model uses climate data to partition precipitation for each HRU into snow, rain, or a mixture based on air temperature. A portion of precipitation is intercepted by the canopy, which is determined by the leaf area index (LAI) and land cover. Once the maximum canopy storage is filled, excess precipitation is throughfall and reaches the land surface. Contributions from throughfall, snowmelt, and upgradient HRUs are added to surface storage, infiltrated up to a maximum daily amount, evaporated, or routed to downslope HRUs and/or adjacent channel reaches. In Ages, the infiltration was modeled using the Eagleson method (1970), which is calculated based on the seasonal soil infiltrability parameters (summer, winter, or snow), soil saturation of the uppermost layer, and field-saturated hydraulic conductivity. The USDA NRCS's Soil Survey Geographic Database (SSURGO) was used to identify soil depth, field capacity, bulk density, and other soil variables. The excess water is routed to depression storage and surface runoff when that is exceeded. Surface runoff can include infiltration-excess and saturation-excess in the model. Once the water infiltrates the soil, it is distributed between medium and large pore storage (MPS and LPS), driving vertical and lateral flows depending on soil properties and slope. Soil storage capacities are represented with a field capacity parameter, where LPS is porosity minus field capacity, and MPS is field capacity minus dead capacity (immobile water). Evaporation and transpiration remove water from soil storage, with quantities determined by the Penman-Monteith potential evapotranspiration (PET) equation. Albedo and aerodynamic resistance variables from the Penman-Monteith potential ET equation are derived from land use variables determined from the USGS National Land Cover dataset. The simulated actual ET is derived from PET and the degree of saturation of the medium pore storage. The HRU slope and pedological parameters are used to determine interflow and deep drainage. Groundwater is split into two components, shallow and deep, and the rates of storage and outflow are based on storage and recession constants (Wells et al., 2024). Water flow can

reach the channels via surface runoff, interflow, and lateral groundwater flow. The channel routing is based on the kinematic wave approach and Manning-Strickler equation (Wells et al., 2024).

The calibration of the Ages model was completed using Luca, a multiple-objective, stepwise, automated calibration procedure based on the Shuffled Complex Evolution method (Hay and Umemoto, 2006). The calibrations were completed for the entire study period even though the observed streamflow is only available for a portion of each year (April through September). The objective function was based on Kling-Gupta efficiency (KGE) (Gupta et al., 2009) calculated from simulated and observed streamflow at the outlet. In the Fourmile Creek watershed, the streamflow is impacted by a diversion to Pine Brook. The Colorado Division of Water Resources reports weekly total diversions from Fourmile Creek for the study period. For the weeks when water was removed from Fourmile Creek, the weekly totals were evenly distributed to create daily water removed and incorporated into the Ages model. A total of 28 global parameters and 6 spatially variable parameters (Table 4-3, Table 4-4, and Table C-1) were calibrated (Table 4-3). For parameters that were selected to vary spatially, values were determined separately for the unburned area and burned areas. All burned HRUs had the same calibrated values. A total of 22 calibrations were run for this study, 11 calibration runs for each type of meteorological input data (gridded or station), following a leave-one-out approach. In this approach, one year of the observed streamflow is removed for each of the 10 calibrations per meteorological input data source, and a final calibration for each meteorological input data source was conducted with no years removed (complete calibration scenario). We used this approach as an alternative to separating the time series into calibration and evaluation segments. The time series simulated was only ten years, so the leave-one-out approach allowed us to increase sample size when evaluating parameters, hydrologic responses, and performance between model runs that include gridded and station meteorological input data.

Once the model was run forward with each calibration set, daily KGE, Nash-Sutcliffe Efficiency (NSE), and bias (%) values were determined to assess the model performance. The KGE combines correlation, variability, and bias components, and a value of one indicates perfect agreement between simulated and observed streamflow (Gupta et al., 2009). The NSE measures the model skill based on the mean of observed data, and a value of one indicates perfect agreement between simulated and observed. An NSE value less than zero indicated that the mean of the observed data is a better estimator than the model (Nash and Sutcliffe, 1970). Bias quantifies the average tendency of the simulated values to be greater or less than the observed data. For bias, a value of zero indicates perfect agreement; negative values indicate an underestimation, and positive values indicate an overestimation. Using the criteria from Moriasi et al. (2007), the NSE and bias values for all model runs were separated into performance categories. The performance categories included very good ($0.75 < NSE \leq 1.00$; bias $< \pm 10$), good ($0.65 < NSE \leq 0.75$; $\pm 10 \leq$ bias $\leq \pm 15$), satisfactory ($0.50 < NSE \leq 0.65$; $\pm 15 \leq$ bias $\leq \pm 25$), and unsatisfactory ($NSE \leq 0.5$; bias $\geq \pm 25$). Additionally, residuals were plotted for the model runs where no observed streamflow data were removed (complete calibration scenarios where no years were left out).

To compare the two meteorological input datasets in a standardized way, we selected the highest-performing calibration set for each meteorological input dataset based on the KGE values. Each calibration set selected was then run with the gridded and station meteorological inputs, while holding all other model parameters constant. In other words, we applied the best calibration set derived from each meteorological input type and then ran it forward with the alternate meteorological input dataset to assess simulated output differences between gridded versus station meteorological inputs.

4.3.3 Assessing model responses to differences in parameters

A Halton sequence sensitivity analysis (Halton, 1960) was conducted based on KGE values to determine which parameters most affected simulated discharge. A Halton sequence uses quasi-random sampling to explore parameter sensitivity. Two sensitivity analyses were performed using the full 10-year daily climate sequences, one for the gridded meteorological inputs and the other for the station meteorological inputs. These sensitivities were ranked for both meteorological input scenarios, and the ranks were averaged. The nine top average-ranking parameters (Table 4-3) were selected to be shown in the results section, and the remaining are presented in the appendix (Table C-1).

Using the selected parameters, we evaluated whether the parameter ranges differed between those in the station vs. the gridded meteorological input scenarios. Because the parameters across the calibration sets (22 total) were not normally distributed, we used the non-parametric Wilcoxon rank-sum test to assess differences in medians and determine p-values. Additionally, six parameters that were calibrated separately for burned and unburned areas (Table 4-4) were evaluated for differences between meteorological input data and burn status. A two-way ANOVA was used to evaluate the effects of burn status and input data type on parameter values. Tukey's HSD post hoc test was applied to examine pairwise difference among the groups.

4.3.4 Assessing hydrologic responses to different meteorological data sources

To understand how differences between input data and model parameters propagate through the model, we assessed flow partitioning from the two complete calibration scenarios. Similar to the assessment of the spatially varying parameters, a two-way ANOVA was used to evaluate the effects of input data and flow type (surface runoff, interflow, fast groundwater, and slow groundwater) on streamflow at the outlet. Next, Tukey's HSD post hoc test was applied to examine pairwise differences. An example year was included to show how flow partitioning

cumulative values differ between the model runs. Here, cumulative flow for each type was calculated over 2016, which was selected based on having the best average model performance for the two meteorological input sources for the complete calibration scenarios. In addition to evaluating the 2016 daily streamflow, we evaluated spatial patterns of daily mean air temperature, cumulative precipitation, and cumulative SWE.

After determining differences between parameters based on burn status, we selected a burned HRU and an adjacent unburned HRU to evaluate how these differences propagate through the model. Again, we selected 2016 and evaluated this only for the station complete calibration scenario. Variables that are directly impacted by the parameters that were calibrated separately for burn status included cumulative precipitation, interception storage, SWE, soil saturation, interflow, and shallow groundwater storage.

4.4 Results

Fourmile Creek has a strong seasonal pattern in the hydrograph, with lower flows during August through April and higher flows associated with snowmelt in May through July (Figure 4-2). The average streamflow during the study period (April through September 2011-2020) was 0.34 cms. A flood event on 12 September 2013 was removed from the observed dataset prior to modeling since the gauge was damaged during the event. Murphy et al. (2015) estimated the peak streamflow during that event reached 71 cms, nearly an order of magnitude higher than any other peak flows. The years 2015 and 2017 had higher peak streamflow during the snowmelt season than other years (8.04 and 3.57 cms, respectively; Figure 4-2). A drought in 2012 led to low streamflow, with an average of 0.10 cms.

The monthly climate data across the watershed had similar seasonal patterns but different magnitudes for the station and gridded datasets. The gridded data generally had greater summer precipitation, warmer air temperature, greater incoming shortwave radiation, and lower RH values (Figure 4-2).

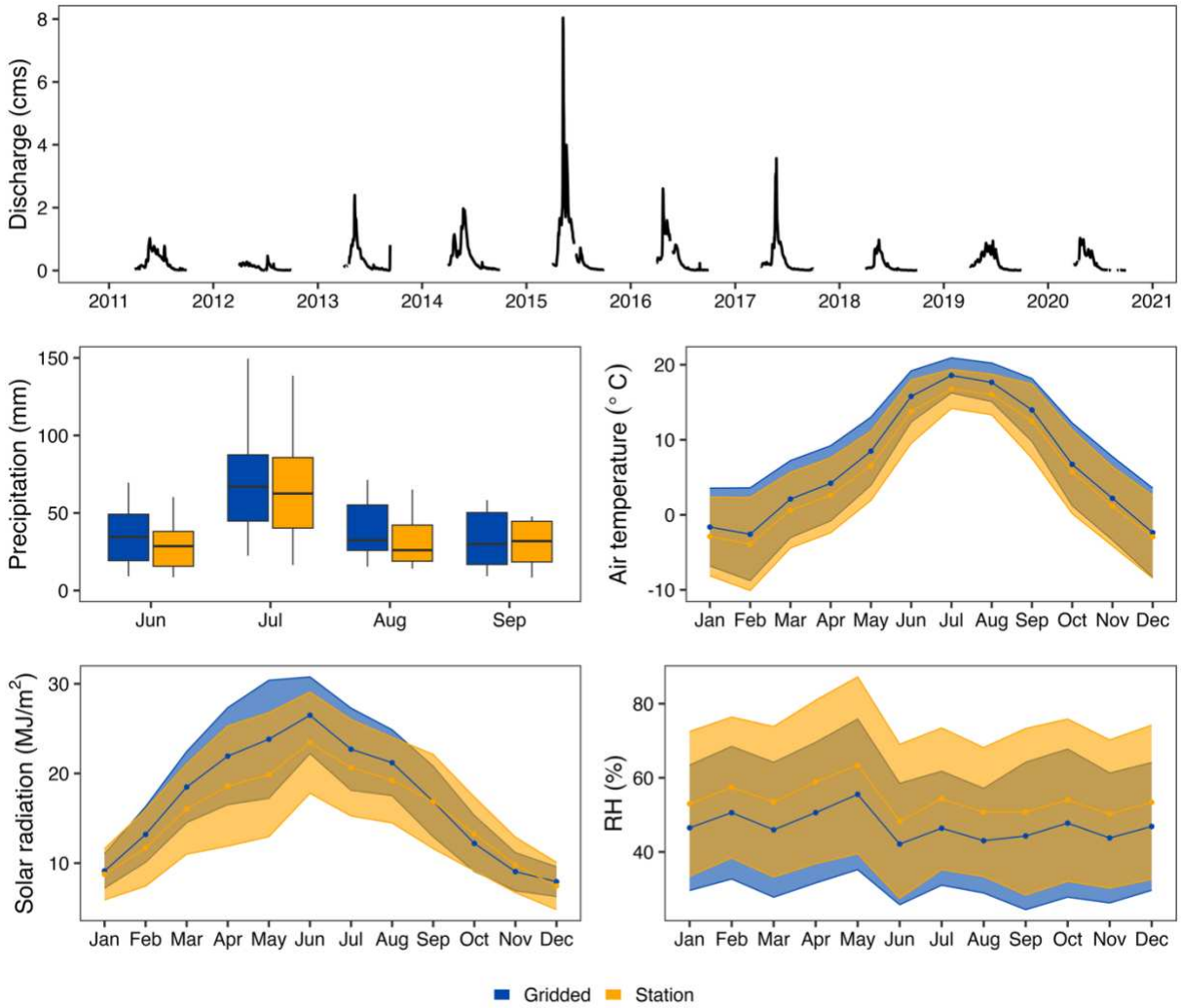

Figure 4-2. Measured daily streamflow (USGS 06727500) at the Fourmile Creek outlet, and watershed-average monthly precipitation, air temperature, incoming shortwave radiation, and relative humidity (RH). The shaded areas show one standard deviation for the respective month over the 10-year study period. In the precipitation plot, two outliers were removed from September (314 mm and 278 mm for gridded and station, respectively) related to the 2013 flood.

4.4.1 Meteorological HRU differences

Spatial patterns of station and gridded data across the watershed reveal systematic differences in summer precipitation, incoming shortwave radiation, air temperature, and RH (Figure 4-3). These differences often vary with elevation and highlight potential biases in gridded data when compared to station data. In 29 of 38 HRUs, the differenced summer cumulative precipitation values were negative (-75 to -6.0 mm), meaning the gridded dataset had greater precipitation. In the HRUs that had positive differenced summer cumulative precipitation, the

values ranged from 1.0 to 53 mm. Incoming shortwave radiation, air temperature, and RH biases generally followed the elevation in the watershed. At higher elevations, gridded data had greater incoming shortwave radiation than station data, and the difference decreased with decreasing elevation. Despite the smaller differences at lower elevation, gridded incoming shortwave radiation remained greater than station data across the catchment (-3.3 to -0.93 MJ/m^2). Conversely, gridded and station RH are more similar at higher and mid elevations than the lower elevation. The RH differences ranged from 4.6 to 8.7% and were only positive, meaning station RH was greater than gridded across the catchment. For air temperature, high elevations showed more similar gridded and station values. At lower HRUs, mean gridded air temperature was greater than station. Mean air temperature differences ranged from -3.8 to 1.6 °C, with only eight HRUs being positive.

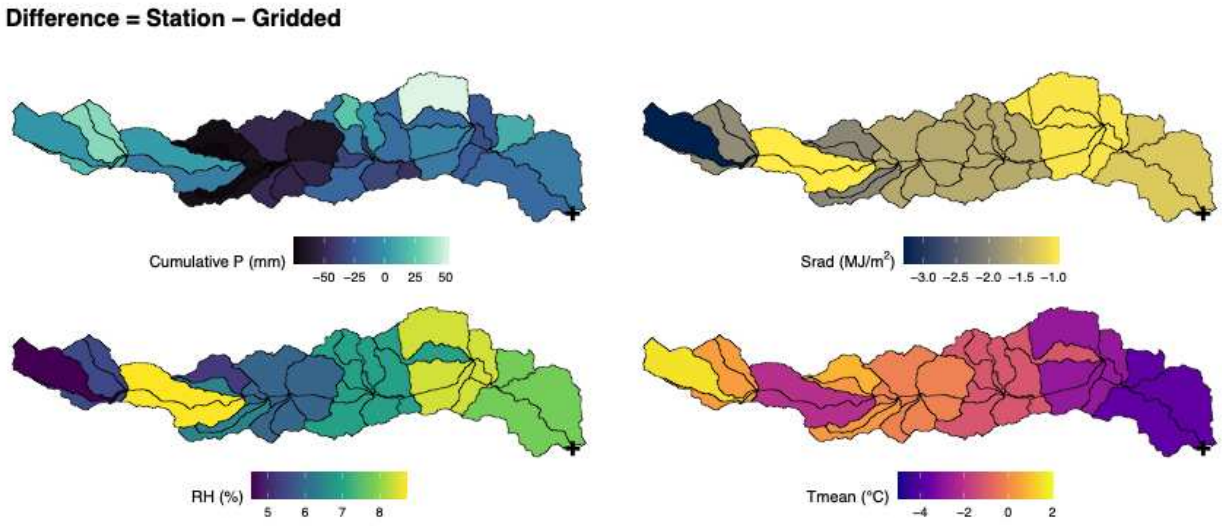

Figure 4-3. Differences between station and gridded inputs over the summer period (June through September) for cumulative precipitation (P) and temporal averages of incoming shortwave radiation (Srad), air temperature (Tmean), and relative humidity (RH) by HRU. The point represents the streamflow gauge at the outlet of the watershed. A negative number indicates a larger gridded value.

4.4.2 Model performance

To evaluate model performance, we computed the KGE, bias, and NSE (Figure 4-4) for simulated versus observed daily streamflow. These metrics are presented for all 22 model runs

(y-axis) across all years in the study period (x-axis) for simulations that included gridded (left) and station meteorological inputs (right). Over the entire study period (year = all) for all model runs, KGE values from models using gridded data ranged from -0.23 to 0.45, while the station model runs ranged from 0.70 to 0.84 (Figure 4-4a). The models using station meteorological inputs were generally acceptable (i.e., greater than 0.40), except for years 2011 and 2012. For the model runs using gridded meteorological inputs, performance was consistently poor for 2012 and 2013, and generally marginal in 2014. Year 2012 greatly reduced the overall average KGE for the entire study period (year = all).

The bias values fell within the very good range for both gridded and station data when evaluating the entire study period, but performance varied when looking at individual years (Figure 4-4b). The bias over the entire study period (year = all) using gridded meteorological inputs ranged from -9.8% to 5.7% over the study period, and the calibration sets using station data ranged from -8.7% to 4.3%. The model runs using gridded meteorological inputs generally estimated for half of the years in the study and underestimated the other half. For station models runs, seven of the years were underestimating and three overestimating.

The NSE values for gridded data over the entire study period were often unsatisfactory, while those using station data were generally satisfactory or better (Figure 4-4c). The NSE over the entire study period (year = all) for gridded data were all unsatisfactory (-2.62 to -0.28), whereas the model runs for the entire study period using station inputs were satisfactory or good (0.61 to 0.71). All model runs using the station meteorological inputs were unsatisfactory in 2011 and 2012, with a maximum NSE of 0.48. Similarly, 70% of model runs for the gridded meteorological inputs had unsatisfactory performance from 2011 through 2014, which greatly affected the NSE values for the entire study period (year = all). This period of unsatisfactory performance was likely when the greatest wildfire impacts occurred. In 2013, station data generally performed in the satisfactory to very good category across all three metrics, whereas gridded data was generally in the unsatisfactory category.

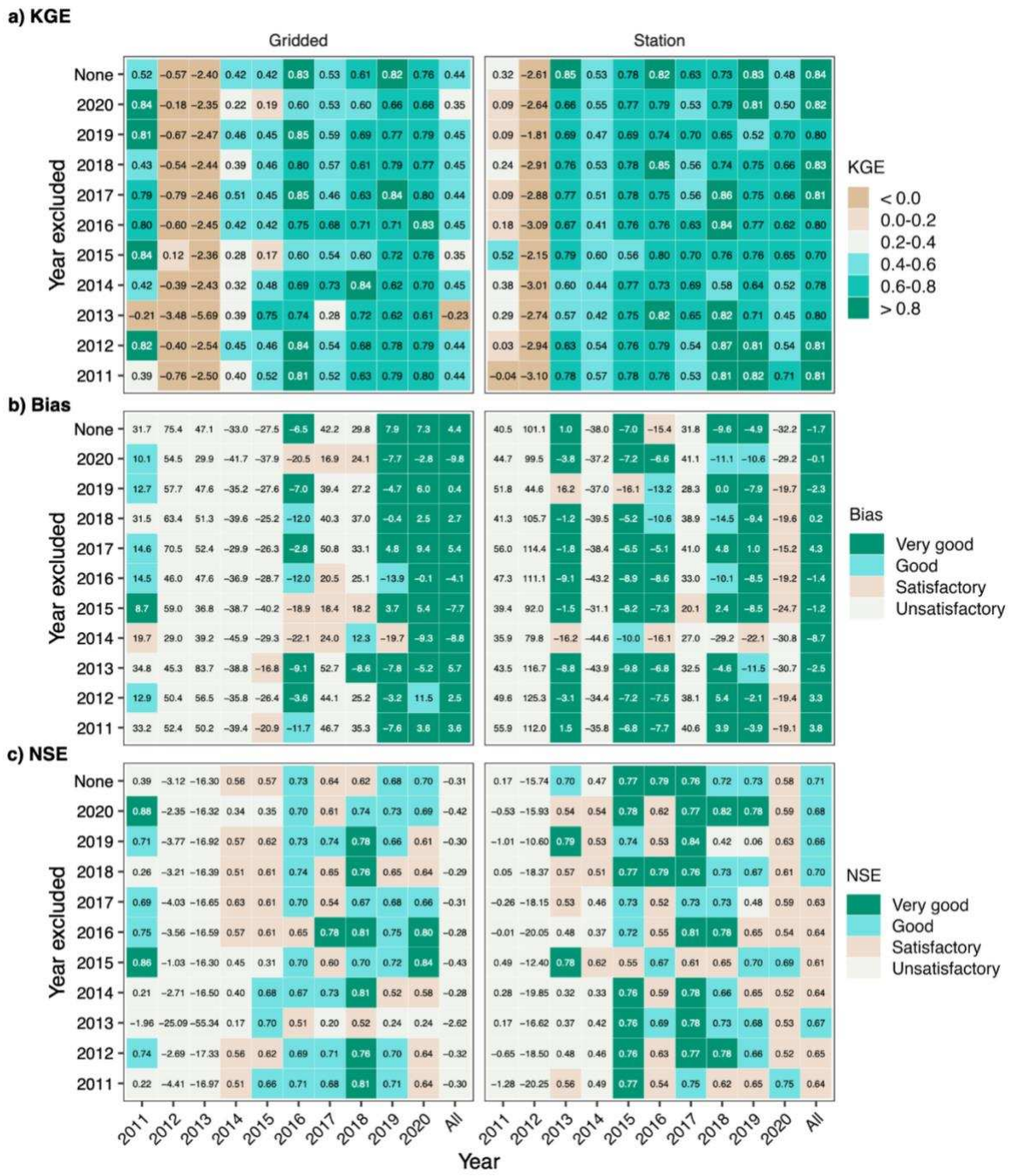

Figure 4-4. Kling-Gupta Efficiency (KGE), bias (%), and Nash-Sutcliffe Efficiency (NSE) values for all model runs using gridded and station meteorological inputs over each year in the study period. The bias and NSE performance categories are based on Moriasi et al. (2017).

Since each of the calibration runs had a different set of parameter values, we selected examples for comparing effects of using gridded vs. station inputs. The four calibration

scenarios that performed best using gridded and station inputs were selected (Table 4-2). The selected calibration sets are the one with 2019 left out for gridded inputs and the one with no years left out for station inputs. Across both selected calibration sets, KGE values were higher for station vs. gridded meteorological inputs (0.84 vs. 0.09 and 0.66 vs. 0.45, respectively) indicating more reliable model performance using station data. Station inputs outperformed gridded inputs even for best gridded scenario (calibration excluding 2019 streamflow data) for both KGE and NSE; however, the station inputs produced a large bias (-19.3%) whereas gridded inputs produced negligible bias 0.4%).

Table 4-2. Model runs where parameters are consistent across the two meteorological inputs. The calibration sets were selected from the highest KGE values for gridded and station each. The bold numbers in NSE and bias columns represent values that are considered satisfactory or above (Moriasi et al., 2017).

Calibration scenario	**Meteorological inputs**	**KGE**	**Bias (%)**	**NSE**
Gridded, excluding 2019	Gridded	0.45	**0.4**	-0.30
Gridded, excluding 2019	Station	0.66	**-19.3**	0.40
Station, excluding none	Gridded	0.09	**34.3**	**0.05**
Station, excluding none	Station	0.84	**-1.7**	**0.71**

4.4.3 Calibration parameters

The parameters selected from the Ages sensitivity analysis are presented in Table 4-3, and the range of values obtained from calibration sets are shown in Figure 4-5. A Wilcoxon rank-sum test for each parameter revealed significant differences between median parameter values for calibration sets that used gridded and station meteorological inputs (Figure 4-5). The points within the boxplots are filled or unfilled to represent model performance determined from Moriasi et al. (2007) for NSE and bias. The coefficient used to partition precipitation phase, snow_trs, showed no significant differences between calibration sets using gridded and station meteorological inputs (p-value > 0.05). Two snowmelt parameters, baseTemp and t_factor,

which influence the temperature at which snow melts and the melt rate, respectively, did differ significantly between gridded and station calibration sets (p-value < 0.0001), with both parameters having higher values for gridded (higher threshold temperature before melting starts and higher melt rate). Both rainfall interception and snow interception differed significantly between the gridded and station calibration sets (p-values <0.001 and < 0.0001, respectively), with rainfall interception being greater when using gridded meteorological inputs and snow interception being greater when using station meteorological inputs. Two soil-related parameters (soilLatVertLPS and soilOutLPS) differed significantly between gridded and station calibration sets (p-values < 0.001 and < 0.05, respectively), with station exhibiting higher values for both soil variables. The lag surface runoff parameter was significantly greater using gridded meteorological inputs compared to station (p-value < 0.0001), while the lag interflow parameter did differ significantly but not as significantly (p-value < 0.05). All the model runs using gridded meteorological inputs performed satisfactorily or above based on bias but performed unsatisfactorily based on NSE.

Table 4-3. Parameters selected based on the sensitivity analysis. The lower and upper limit represents the model calibration range. The parameters in bold indicate a significant difference between calibration sets using gridded and station meteorological inputs (p-value < 0.05).

Parameter	**Unit**	**Description**	**Lower limit**	**Upper limit**
snow_trs	-	Coefficient used to partition precipitation into snow and rain	-5	5
baseTemp	ºC	Base melting temperature for snow	-5	5
t_factor	-	Temperature factor for snow melt calculation	0	5
a_rain	mm	Maximum storage capacity per LAI for rain	0.1	10
a_snow	mm	Maximum storage capacity per LAI for snow in mm of SWE	0.1	10
lagSurfaceRunoff	-	Surface runoff lag factor	1	1.5
soilLatVertLPS	-	Coefficient for distribution of the LPS outflow on the lateral and vertical component	0.001	10
soilOutLPS	-	Coefficient for definition of LPS outflow	0.01	10
lagInterflow	-	Interflow lag factor	1	2

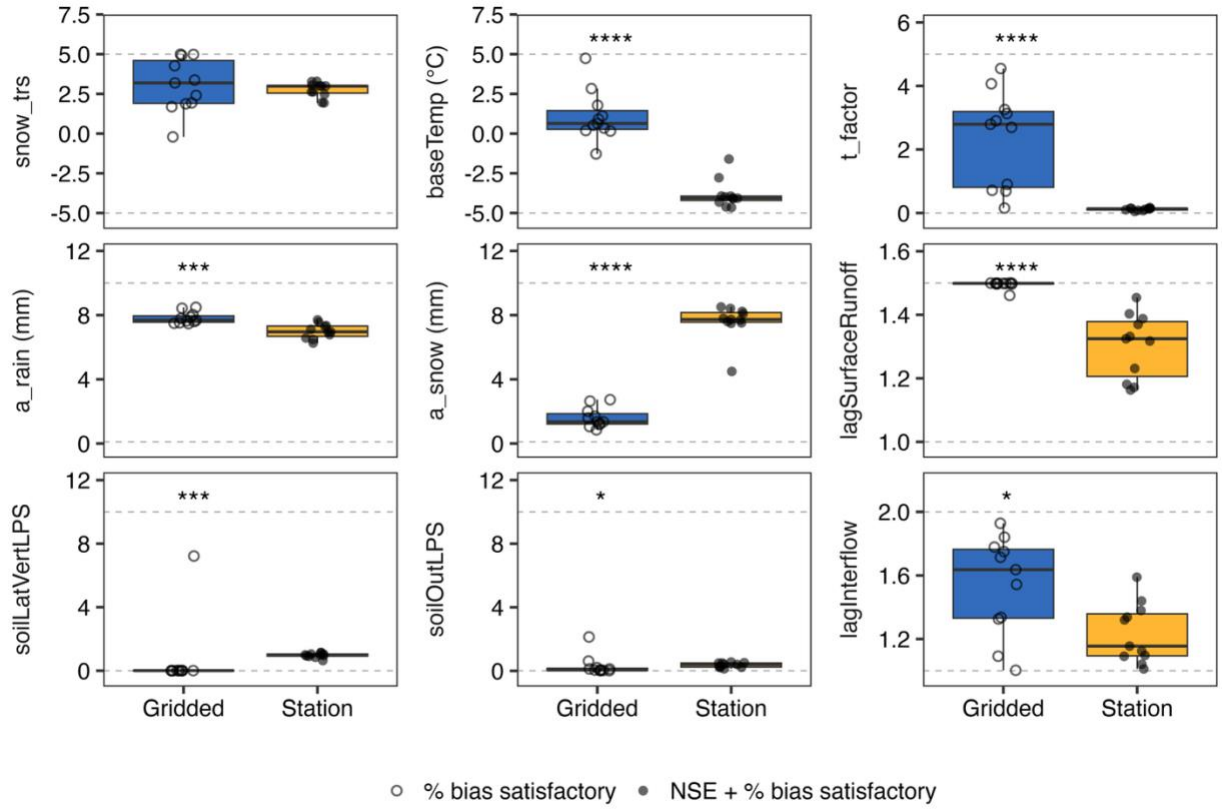

Figure 4-5. Parameter variability between the calibration sets using gridded and station meteorological inputs. The asterisks represent significant difference determined by the Wilcoxon-rank sum test (**** indicates p-value < 0.0001; *** indicates p-value < 0.001; * indicates p-value < 0.05). The points are based on performance categories from Moriasi et al. (2007), indicating where NSE and bias were satisfactory or better (filled circle) or only bias was satisfactory or better (open circle). The dashed grey lines represent the lower and upper limits used for each parameter during calibration. Parameters are defined in Table 4-3.

4.4.4 Flow partitioning

For the model runs using gridded meteorological inputs for the complete calibration scenario, most of the flow was classified as fast groundwater (on average 95%), while the station meteorological inputs for the complete calibration scenario resulted in a split between interflow (on average 44%) and fast groundwater (on average 54%) (Figure 4-6). For surface runoff, the values from the 2013 flood were removed from the boxplot since they were such large outliers (about 25 mm surface runoff in a single event for station meteorological inputs and 86 mm for gridded). In model runs from the complete calibration scenario, surface runoff and slow groundwater contributed a small proportion of the total annual flow (Figure 4-6). There

were no statistically significant differences between the estimated means of surface runoff and slow groundwater for both gridded and station complete calibration scenarios. However, the gridded complete scenario had surface runoff values from nearly 0.0 to 2.2 mm and station from nearly 0.0 to 3.1 mm. For slow groundwater, gridded had values from 0.06 to 0.12 mm and station 0.06 to 0.16 mm. Additionally, the estimated mean for gridded interflow, which ranged from 0.62 to 2.7 mm, was comparable to those of surface runoff and slow groundwater.

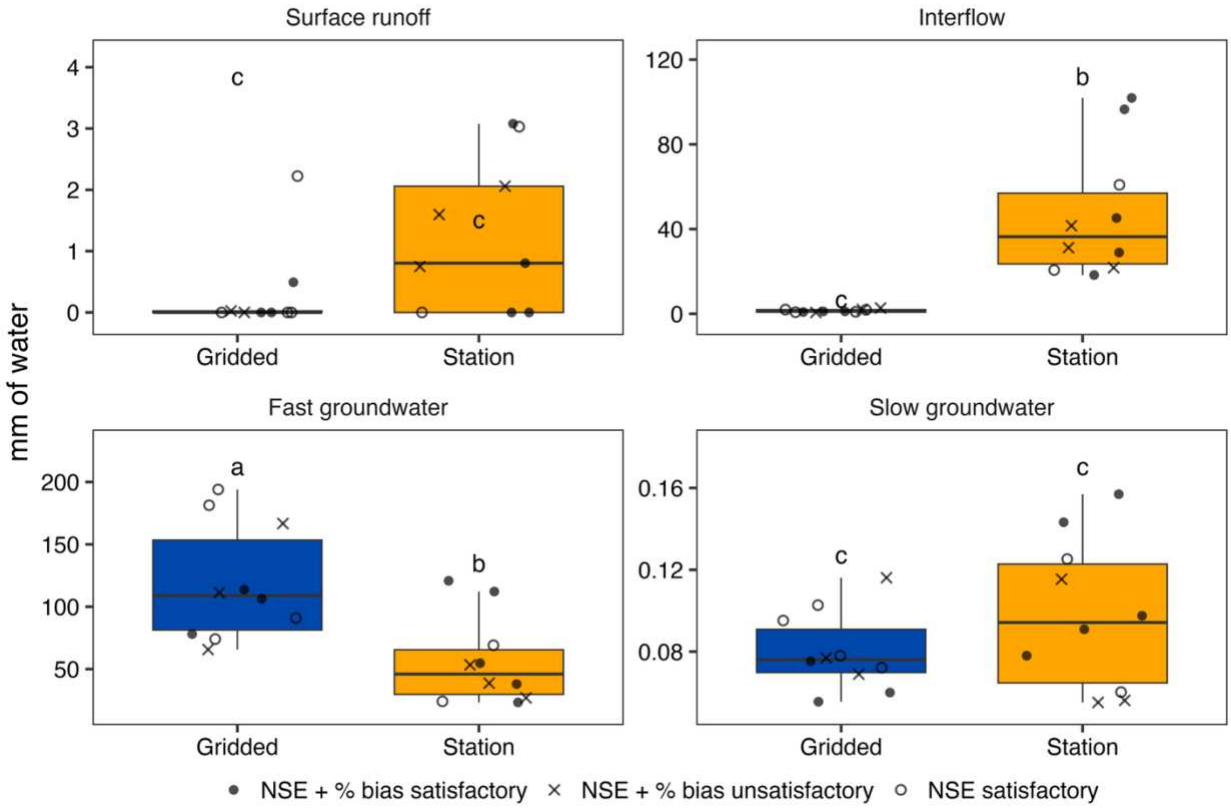

Figure 4-6. Flow partition amounts for model runs calibrated using the complete calibration scenario for each year in the study period. The letters represent statistically significant difference ($p < 0.05$) between flow partition types based on Tukey's HSD post hoc test. The points are based on performance categories from Moriasi et al. (2007) indicating where NSE and bias were satisfactory (filled circle), NSE and bias were unsatisfactory (x), and only NSE was satisfactory (open circle).

4.4.5 An example year: 2016

To further evaluate how different meteorological input datasets can affect calibration, and thus state variables and streamflow, we selected the highest performing year (2016) from the complete calibration scenarios (Figures 4-4, 4-5, 4-6). Since the models particularly struggled to

simulate the magnitude and timing of snowmelt accurately (Figure C-2), we evaluated air temperature, precipitation, and SWE during this period in 2016 (Figure 4-7). For this comparison, we selected two HRUs with mean elevations of 1956 m (low) and 3316 m (high). Low elevation temperatures (April to July) were mostly above freezing for both gridded and station (mean air temperature of 12.9 versus 8.9 °C, respectively), and little snow accumulated (maximum for gridded was 49 mm and station was 23 mm). At high elevations, gridded air temperatures were lower than station (mean of 4.3 versus 5.3 °C, respectively), but both data sources had low enough temperatures to allow snow accumulation. The temporal patterns of snow accumulation were similar for both scenarios until peak SWE (410 mm for gridded and 254 mm for station) due to similar air temperatures and precipitation. Although the accumulation patterns were similar, gridded had greater SWE than station due to differences in the snow interception parameter (a_snow). However, despite the higher elevation HRU with station inputs having greater mean air temperature, the melt out date for both input datasets was 8 June. The melt rate was more gradual for station and this was attributed to the calibrated melt rate parameter (t_factor) being lower than gridded.

As an example of the differences in simulated streamflow between model scenarios using gridded versus station data as input, Figure 4-8 shows the 2016 hydrographs from the gridded and station complete calibration scenarios. The bias for the models runs from these scenarios were negative for both gridded and station (-6.5% and -15.4%, respectively), indicating an underestimation. The gridded and station scenarios were both higher than the observed streamflow at the beginning of April, and the station-calibrated simulation became more similar to the observed streamflow mid-April. After snowmelt, the gridded and station scenarios were mostly lower than the observed streamflow for the rest of the flow season. In this example, the gridded scenario had a more gradually varying hydrograph than the station scenario, probably because more water was partitioned into groundwater (Figure 4-8). In contrast, the model run using station meteorological inputs partitioned more of the water into

interflow. Although 2016 is an example year, these flow type patterns (i.e., gridded scenario having more groundwater and station having more interflow) remain for each year in the study period.

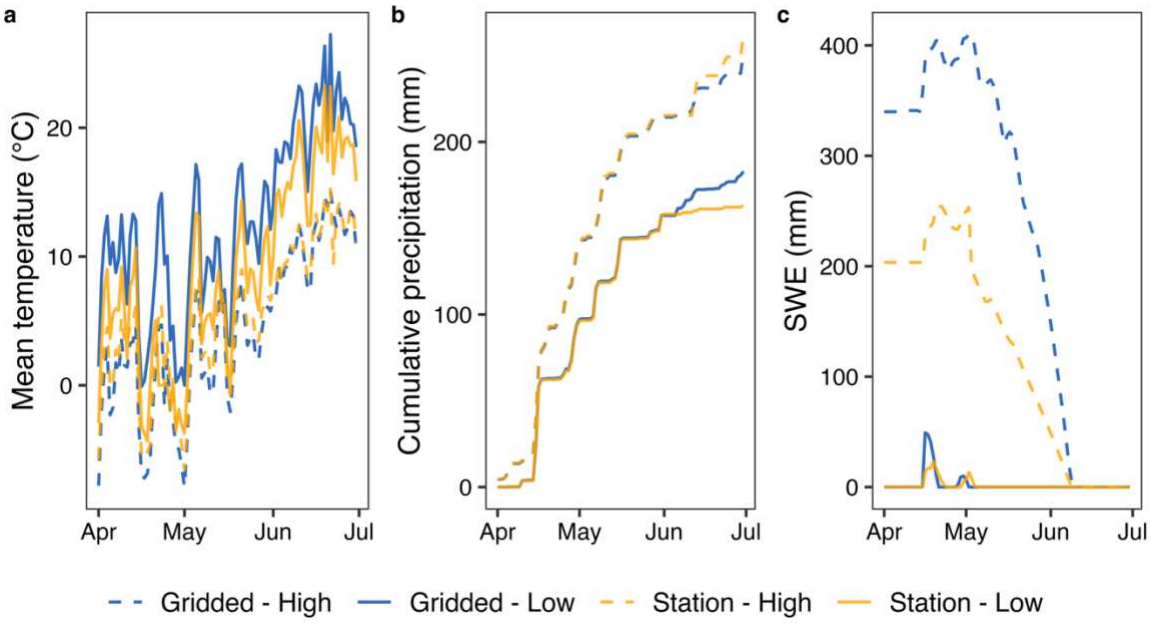

Figure 4-7. Daily mean air temperature (a), cumulative precipitation (b), and snow water equivalent (SWE) (c) during the 2016 snowmelt period for a low (1956 m) and high (3316 m) elevation HRU from the complete calibration scenarios using gridded and station meteorological inputs.

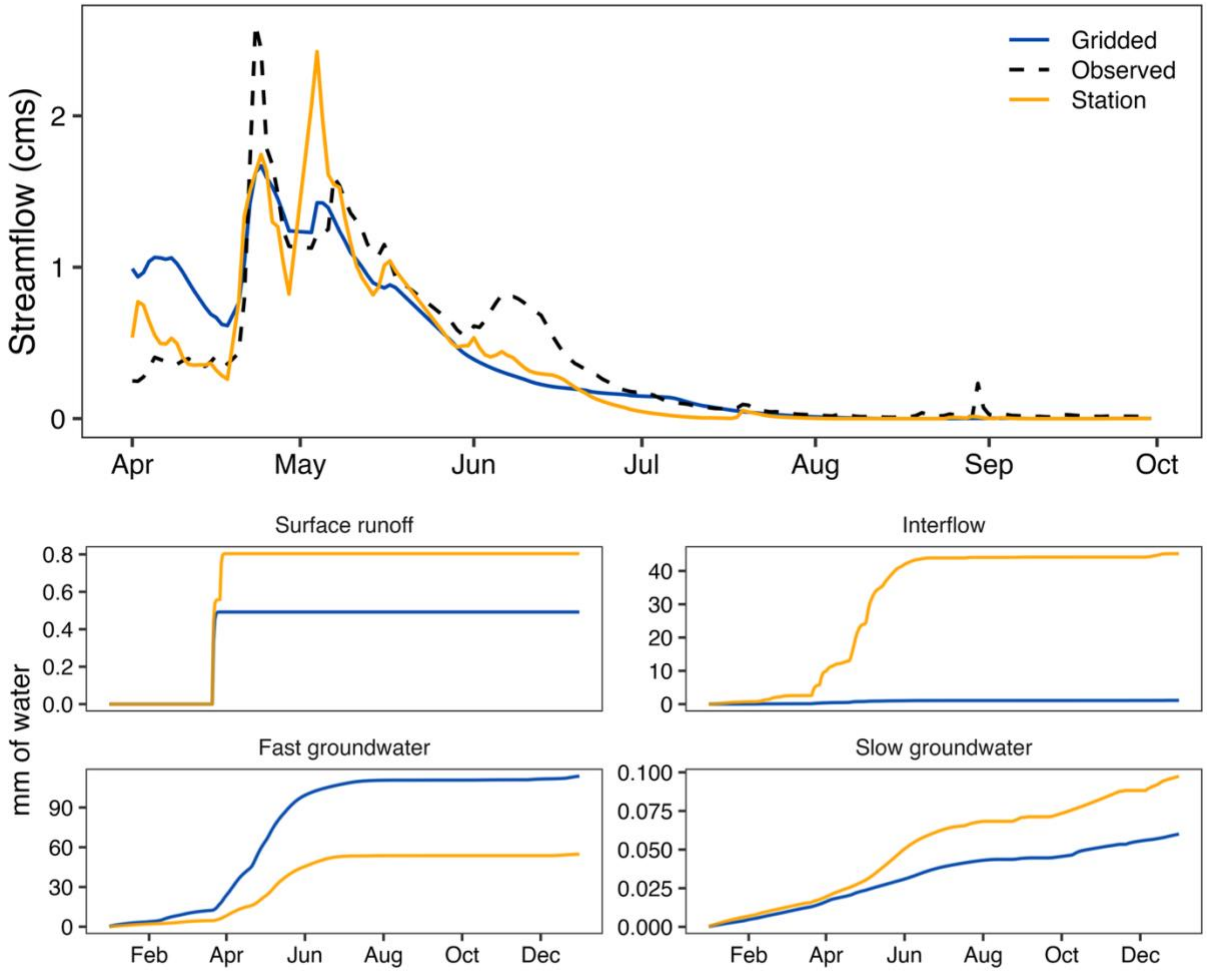

Figure 4-8. Simulated daily streamflow (cms) and cumulative flow partition amounts (mm) from the complete calibration scenario using gridded and station meteorological inputs and observed streamflow in 2016.

4.4.6 Wildfire effects

For the parameters that were calibrated separately for burned and unburned areas (Table 4-3), the values varied with burn status (i.e., burned or unburned) and input source (i.e., gridded or station; Figure 4-9, Table C-2). While some parameters exhibited consistent differences between burned and unburned areas or between gridded and station meteorological inputs, others showed no significant variation. For rainfall interception (a_rain), the unburned gridded and station calibration sets had the greatest values, with gridded having an estimated mean value 1.9 mm greater than station. Snow interception (a_snow) differed greatly between unburned gridded and station (estimated mean difference = 8.3 mm). There was not a

significant difference between simulated burned and unburned gridded or burned gridded and station snow interception. Note, when LAI equals zero, a_rain and a_snow are not used.

The maximum summer infiltration was similar between burned gridded, unburned gridded, and unburned station parameters and similar between burned gridded, burned station, and unburned station. For soilMaxDPS, the unburned station and gridded parameters values were lower than the burned, meaning the burned areas had higher maximum depression storage. Here, depression storage is the average depth over a HRU and includes both microtopography and any sub-HRU depressions. The unburned gridded had an estimated mean difference 2.3 mm greater than unburned station. Burned and unburned gridded lag surface runoff values had much less variability than station values. Additionally, burned and unburned gridded lag surface runoff values did not differ significantly, and the same was observed with unburned and burned station. The soilDistMPSLPS parameter (a coefficient for distribution of infiltration to soil storages) show no significant differences between burned station and gridded and burned station and unburned gridded. Increasing soilDistMPSLPS causes a greater fraction of infiltration to go into MPS (smaller pores) and less into large pore storage (LPS).

Table 4-4. Parameters that were calibrated separately for burned areas. The lower and upper limits represent the model calibration range. If the lower and upper limits were different for the burned areas, the burned limits are shown in parentheses.

Parameter	**Unit**	**Description**	**Lower limit**	**Upper limit**
a_rain	mm	Maximum storage capacity per LAI for rainfall	0.1 (0.1)	10 (5)
a_snow	mm	Maximum storage capacity er LAI for snow	0.1 (0.1)	10 (5)
soilMaxInfSummer	mm/day	Coefficient used to calculate maximum infiltration in summer	0 (0)	150 (100)
soilMaxDPS	mm	Maximum depression storage capacity	0 (0)	7 (10)
lagSurfaceRunoff	-	Surface runoff lag factor	1	1.5
soilDistMPSLPS	-	Coefficient for distribution of infiltration to the LPS and MPS soil storages	0	7

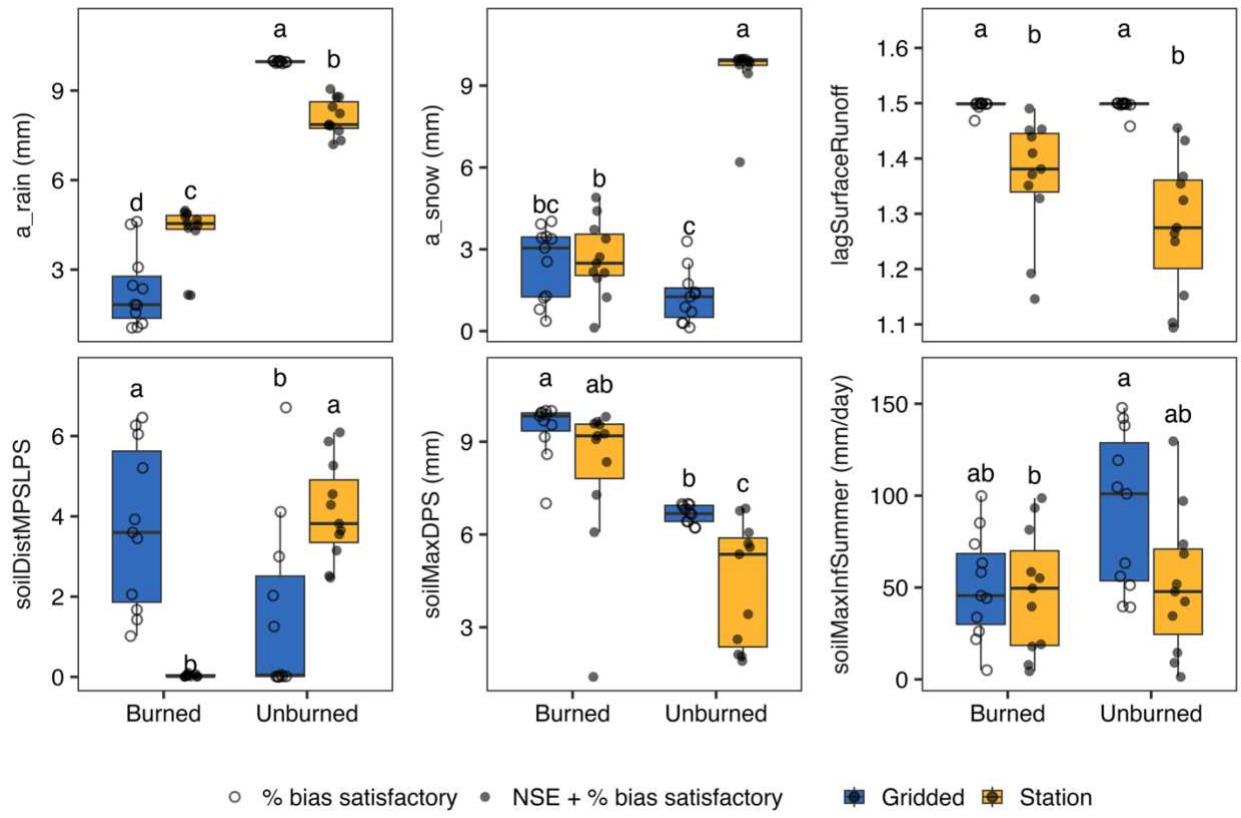

Figure 4-9. Burned and unburned parameters across all calibration sets using gridded and station meteorological inputs. The letters denote significant differences between means according to the Tukey HSD post-hoc test. The points are based on performance categories from Moriasi et al. (2007), indicating where NSE and bias were satisfactory or better (filled circle) or only bias was satisfactory or better (open circle). Parameters are defined in Table 4-4.

A burned and adjacent unburned HRU were selected to compare how calibrations for specific parameters (Table 4-4) affected hydrologic partitioning (Figure 4-10) in 2016 using the station complete calibration scenario. The precipitation values are slightly different since Ages interpolates precipitation for each HRU, and they have only a 3 mm difference for the cumulative value. The average interception at the burned HRU was 1.1 mm compared to 9.2 mm at the unburned HRU. Because of lower interception, SWE accumulation was higher for the burned HRU (peak SWE of 52 mm compared to 33 mm for the unburned HRU). Because of greater snowmelt input, soil saturation is higher during snowmelt for the burned HRU. The average daily ET is slightly higher for the burned HRU compared to the unburned HRU (0.35 and 0.12 mm, respectively). The burned HRU has a lower value of soilDistMPSLPS (meaning

water is moving faster to large pore storage and thus interflow and percolation), leading to more interflow and shallow groundwater than the unburned HRU (104 and 126 L, respectively for the burned HRU and 35 and 36 L for unburned HRU).

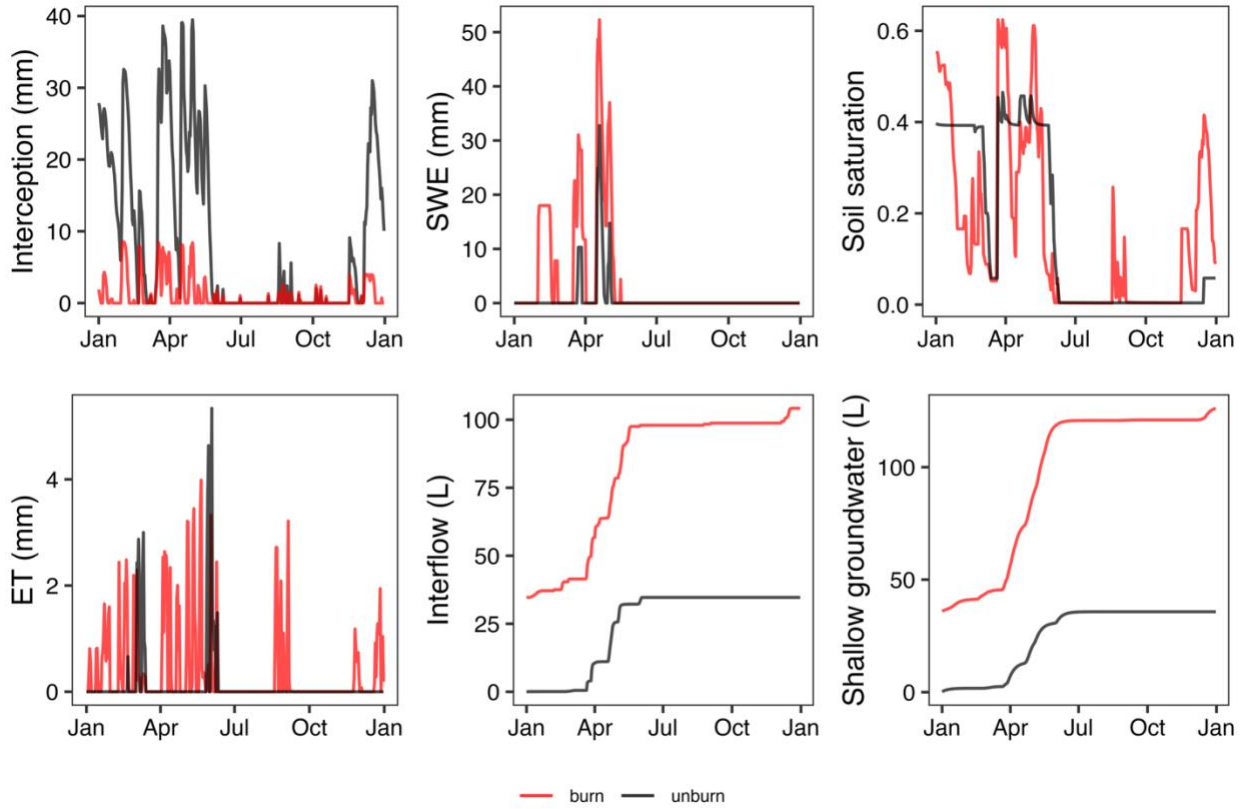

Figure 4-10. The 2016 cumulative precipitation, interception, snow water equivalent (SWE), soil saturation, cumulative interflow, and cumulative shallow groundwater for adjacent burned and unburned HRUs from the complete calibration scenario.

4.5 Discussion

4.5.1 Biases between gridded and station datasets

This study assessed differences in meteorological input data, compared calibrated parameters between the gridded and station inputs, and analyzed resulting differences in hydrologic responses. We found that station meteorological inputs resulted in better model performance overall, as indicated by KGE and NSE values for daily streamflow (Table 4-2). This may suggest that the station data are more accurate than the gridded datasets. The gridded datasets were developed using large gauge networks like SNOTEL, but they do not incorporate

data from research weather stations. Besides the two SNOTEL sites within and nearby the watershed, the sensors used in the station analysis were research stations. The broad spatial coverage of weather stations in this watershed led to improved representation of meteorological inputs compared to the nationwide sensor networks that were used to generate PRISM and Daymet.

While station meteorological inputs may have led to improved model accuracy, the high density of stations in this watershed (11 precipitation gauges, 5 air temperature sensors, and 4 incoming shortwave radiation and RH sensors) and their elevation coverage is unusual; few places have as many weather stations in a small watershed. Where gauges are sparse, a gridded dataset may perform as well or even better than station data (Gao et al., 2017; Shuai et al., 2022).

Across the watershed, gridded meteorological data generally showed more summer precipitation compared to station data (Figure 4-2 and Figure 4-3). Differences in gridded precipitation datasets are well documented in the western U.S. (Gutmann et al., 2012; Livneh et al., 2014; Henn et al., 2018; Mankin et al., 2025) and are often attributed to sparse gauge coverage, complex terrain, and localized convective storms. In the Colorado Front Range, summer precipitation differences may be due to elevation variability and heterogeneity in summer storm patterns (Chapter 2).

The air temperature differences between station and gridded data varied with elevation. High-elevation HRUs showed better agreement between gridded and station data, while low elevation HRUs had higher gridded air temperatures. The low elevation HRUs having greater gridded air temperature is potentially related to the gridded dataset failing to capture cold air drainage (McEvoy et al., 2014). Although the PRISM dataset does use a locally weighted regression of air temperature against elevation and incorporates terrain orientation and topographic position (Daly et al., 2008), its reliance on coarse resolution station data and DEMs limit its accuracy in mountainous terrain regardless of interpolation complexity. Additionally, the

station data informing PRISM are typically from valley-situated ground stations and ridgetop locations, and because of this mountain slopes are poorly represented (Strachan and Daly, 2017).

Incoming shortwave radiation in the Daymet dataset is estimated using a multi-step process that begins with top-of-atmosphere radiation (based on latitude, day of year, and solar declination angle), followed by a clear-sky radiation calculation (adjusted for solar zenith angle, terrain shading, and sun angle), and then corrected for atmospheric transmissivity using diurnal temperature range as a proxy for cloud cover (Thornton & Running, 1999). Unlike PRISM variables, these estimates are not directly informed by observed incoming shortwave radiation measurements but instead rely on empirical relationships. The higher incoming shortwave radiation gridded data observed in this study may reflect underestimation of cloud cover or shading effects in complex terrain that are not captured by a coarser resolution DEM. Additionally, inaccuracies in air temperature can propagate into the transmissivity correction, further affecting incoming shortwave radiation estimates. Field-based measurements using radiometers are also subject to error related to sensor specifications, calibration, maintenance, and environmental interference (Konings and Habte, 2015). For example, snow can accumulate on a sensor and significantly impact recorded incoming shortwave radiation values. These limitations should be considered when comparing modeled radiation data to observations.

Similar to incoming shortwave radiation, RH is not directly measured but derived from other variables in the PRISM dataset. The RH was calculated using the Magnus formula as specified in Alduchov and Eskridge (1996), which includes dewpoint temperature and air temperature, both from the PRISM dataset. As with incoming shortwave radiation, the accuracy of RH estimates is influenced by the accuracy of the underlying temperature data. Terrain complexity and sub-grid scale variability in moisture availability can also introduce uncertainty, especially in mountainous regions. Unlike direct RH sensor measurements, modeled RH cannot capture localized effects such as canopy interception or cold air drainage in valley bottoms. In

this study, station RH was greater than gridded, and the difference was largest at lower elevations, which also had the largest air temperature differences. The biases in the gridded dataset seemed to have been strongly influenced by elevation. Few studies have assessed gridded RH and incoming shortwave radiation datasets to station data (Mankin et al., 2025).

4.5.2 Parameter differences between input datasets

Differences between gridded and station meteorological inputs led to the calibration algorithm selecting different optimal calibration parameter values, which affected how water was partitioned in the simulations. The differences created in the simulations start with the magnitude and timing of snow accumulation and melt (Figure 4-7). Although winter precipitation totals were similar, the model run with gridded inputs produced greater SWE than station during the accumulation period, likely due to the smaller canopy snow interception parameter (Figure 4-5). The station model run intercepted more snow, which led to less SWE accumulation. During snowmelt, higher values of a melt rate parameter (t_factor) for the gridded model run led to earlier and more rapid melt (Figure 4-7), which led to an overestimation of streamflow in April (Figure 4-8). Two soil parameters related to interflow (soilLatVertLPS and soilOutLPS) were calibrated to higher values in the station model runs (Figure 4-5), resulting in greater interflow compared to the gridded model runs (Figure 4-6). In contrast, the gridded models routed more water to vertical percolation, increasing the fast groundwater flow (Figure 4-6). In summary, the gridded runs had less interception, more SWE, earlier/faster snow melt, and slower release of infiltrated water to streamflow (95% fast groundwater), compared to station runs with more interception, less SWE, later/slower snowmelt, and faster subsurface discharge to streamflow (46% interflow, 50% fast groundwater). These findings highlight that meteorological input source data propagate through the processes simulated and result in differences in calibrated parameters and, consequently, the simulated water balance. These differences in hydrologic processes due to differences in meteorological input datasets have been found in other studies

(Elsner et al., 2014; Radcliffe and Mukudan, 2017; Tercek et al., 2021; Shuai et al., 2022). Since the model was only calibrated to streamflow, it was not constrained to particular values and timing of SWE, and there are no measurements of interflow and groundwater to inform the related parameters. The model can simulate streamflow at the outlet in many ways where multiple parameter combinations can yield acceptable streamflow results, known as equifinality (Beven, 2006). For example, the model can produce similar streamflow outputs through different combinations of parameters that affect SWE accumulation and melt (e.g., snow_trs, baseTemp, t_factor) and water partitioning between surface and subsurface zones (e.g., soilLatVertLPS, soilOutLPS, lagInterflow). Without constraints from physical measurements, many different calibration solutions can produce similar streamflow simulations.

4.5.3 Post-fire-specific parameter calibration

Because a portion of the watershed was burned, we calibrated heterogeneous parameters likely to be affected by wildfire. Infiltration-excess overland flow is the most common reason for surface runoff post-fire. This can be a result of vegetation canopy declines, changes in ground cover, and soil infiltration reduction (Moody et al., 2013). When the canopy is reduced, more snow and rainfall reach the surface (Stoof et al., 2012; Harpold et al., 2014). In this study, we further constrained the snow and rainfall interception parameters (a_rain and a_snow) to a maximum value of 5 instead of 10 mm (Table 4-4). The burned areas showed a decrease in the snow and rainfall interception parameter values in gridded and station calibration scenarios except for the gridded snow interception parameter (Figure 4-9). Maximum interception values should be independent of the meteorological input source and may be better informed by values reported in literature, remote sensing methods, or field measurements rather than calibrated (Ebel et al., 2023). A study conducted in the subalpine region of the watershed (an unburned area) estimated snow interception using several methods (e.g., time-lapse imagery with machine learning, eddy covariance measurements) and found the combination of time-lapse

images and machine learning performed best at estimating snow interception, which could be used in future studies in place of calibrating maximum interception (Harvey et al., 2025). Even so, scaling up such estimates from measurements to the HRU and watershed scales remains challenging. In the station complete calibration scenario, a burned HRU had greater snow accumulation than in an adjacent unburned HRU, reflecting the differences in snow interception. Also, on the individual storm scale we can see greater snow accumulation in burned HRUs than unburned, which is consistent with field observations in burned sites (Harpold et al., 2014).

In this region, fire reduces infiltration due to soil-water repellency, surface seal formation, loss of ground cover, development of hyper-dry conditions, and changes in soil structure (Ebel, 2020). Although the Ages model does not have a parameter for each of these processes, there was a reduction in maximum summer infiltration (soilMaxInfSummer) and an increase in maximum depression storage (soilMaxDPS) in the burned areas (Figure 4-9). A burned HRU had higher soil saturation than an adjacent unburned HRU during snowmelt likely due to the greater snowpack in the burned HRU. Initially, we suspected the higher soil saturation was also related to reduced transpiration, but the mean ET values were greater for the burned than unburned areas. The burned HRUs are located in the intermittent snow zone, so they do not have a consistent snowpack (Figure 4-10). Because of the higher soil saturation in the burned HRU and more rapid water movement into deeper soil layers (controlled by soilDistMPSLPS) compared to unburned areas, interflow and shallow groundwater contributions to streamflow were higher in the burned HRU (Figure 4-10). These findings demonstrate how post-fire changes can alter water partitioning and highlight the importance of explicitly representing burned areas in hydrologic models, even if they account for a small portion of the watershed. In this study, streamflow data upstream of the burned area were not available to isolate the burn effects on hydrology, but the spatially explicit model appears to have simulated the heterogeneous responses in a physically realistic manner.

4.5.4 Leave-one-out resampling

The model calibration approach of leave-one-out for observed streamflow during calibration is not a common approach in hydrologic modeling due to the high cost of model calibration. Often hydrologic studies instead incorporate a training and testing split to avoid overfitting and evaluate model performance. The leave-one-out method was used to increase the amount of calibration scenarios as a way of expanding the comparisons between gridded and station parameters and help account for equifinality when comparing station and gridded parameter differences. The increased calibration scenarios enabled statistical comparisons between parameters obtained from different input data sources, which was an advantage of the approach. Since only one year was left out at a time, the model performance and parameters selected were fairly consistent across each set of gridded or station calibration scenarios (Figure 4-4, years = all). However, this could be due to performance metrics being less sensitive to poor performing years that had lower streamflow. For example, when evaluating station KGE and NSE on a year-by-year basis, poorly performing years such as 2011 and 2012, which had less streamflow, have only a small impact on the metrics calculated over the full period (Figure 4-3). The overall KGE and NSE values are less affected because the data exhibit large variance across the study period.

Multiple performance metrics were used to evaluate model performance (Figure 4-4). Since the model was calibrated to KGE, we incorporated bias and NSE to provide a more robust assessment. Based on a comparison of metrics for the “All Years” model runs, covering the full study period (Figure 4-11), bias had no relationship to KGE even though bias is one of three components of the KGE metric. When evaluating the full calibration set, this pattern held true except when KGE values were below -2, which generally corresponded to large positive bias values. In contrast, NSE was directly correlated with KGE. This pattern held when evaluating the full calibration set. The NSE values were only positive for the station calibration sets when KGE was greater than 0.75. The model runs with KGE values above 0.7 had NSE values

greater than 0.5. This indicates that the model performance could be considered better or worse depending on which metric is evaluated, but they generally follow the same trend. For example, while a KGE value around 0.5 may be considered acceptable, the associated NSE value is below 0.0, indicating poor model performance and that the mean of the observed data is a better estimator than the model.

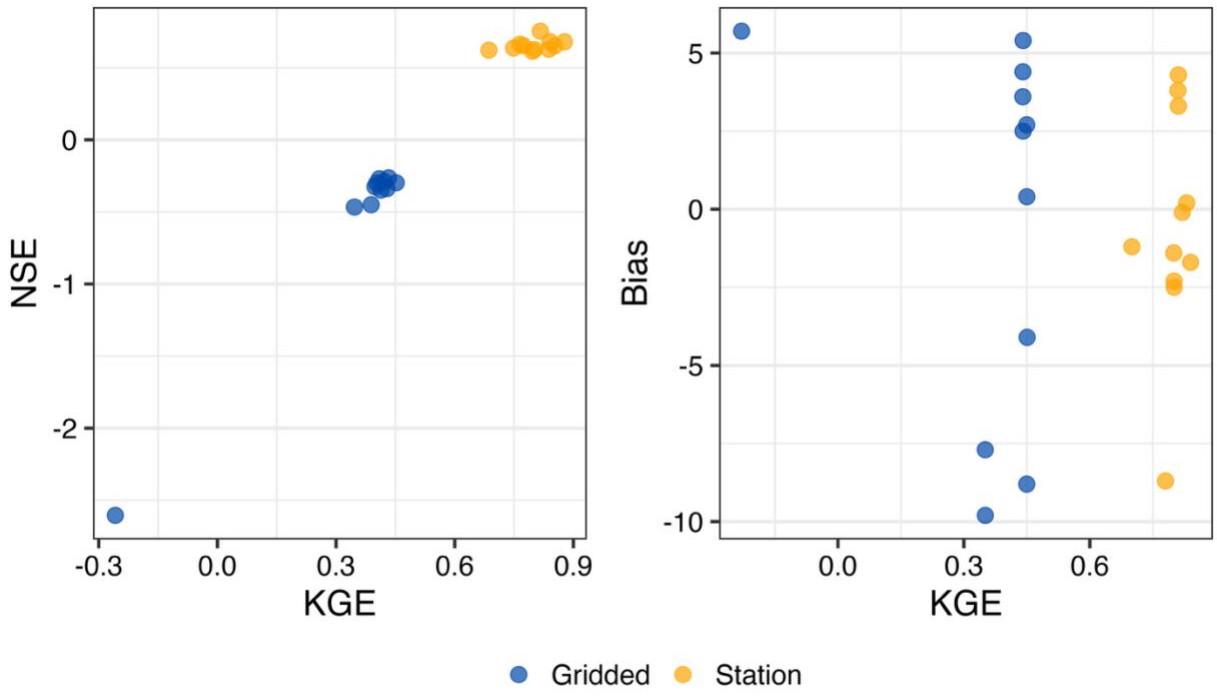

Figure 4-11. The KGE compared to NSE and bias (%) for all calibration sets for the full study period.

4.5.5 Study limitations

Several limitations related to this study should be acknowledged. First, the model calibration was constrained to available streamflow observations, which only existed between April and September. This likely contributed to the poor performance during the snowmelt season (Figure 4-8 and C-2). Additionally, the model did not include other hydrologic observations such as SWE or soil moisture for the calibrations, which could help improve parameter constraints. Another limitation is that the burned area parameters were held constant throughout the 10-year study period. However, infiltration processes likely recovered within 5 to

7 years post-fire and exhibited the most dramatic changes during the first two years, based on a study completed within the Fourmile Canyon Fire (Ebel, 2020). Our findings are specific to a single watershed and may not generalize to regions with different hydrologic regimes. Future work could address these limitations by incorporating additional observational datasets, and this is especially true in burned areas.

4.6 Conclusion

This study aimed to examine how the source of meteorological data affects model calibration, performance, and behavior for a catchment in the Colorado Front Range. The station meteorological data resulted in better model performance likely because the high number of stations available better represented magnitudes and spatial variability of climate variables in this watershed. The differences between gridded and station data appeared to be influenced by elevation, indicating limitations in the elevation-dependent functions used by gridded meteorological data products to estimate values of air temperature, RH, and precipitation. The differences in meteorological input datasets led to distinct parameter calibrations. The gridded inputs resulted in greater snow accumulation, earlier, faster snowmelt due to lower snow interception and higher snowmelt parameter values, and slower release of infiltrated water to streamflow (95% fast groundwater). In contrast, station inputs led to more snow interception, less SWE, and greater interflow driven by higher interception and soil parameter values. These results demonstrate that hydrologic model parameter optimizations change with differences in meteorological input data source. To improve the accuracy of hydrologic partitioning, the model could be calibrated to other hydrologic variables (e.g., SWE, soil moisture). A burned HRUs had greater interception, leading to greater SWE accumulation, and increased soil saturation compared to an adjacent unburned HRU. These results show that wildfire impacts, such as reduced canopy cover and altered soil structure, can meaningfully influence hydrologic partitioning. Calibrating burned and unburned areas heterogeneously and

constraining parameters related to post-fire effects resulted in greater snow accumulation, higher soil saturation, and increased interflow and shallow groundwater contributions. The findings highlight the importance of parameterizing burned areas separately and suggest that some parameters, like interception, may be better informed by literature or remote sensing rather than calibration alone.

CHAPTER 5. Conclusions

This dissertation explored mountain hydrology in three ways: (1) assessing rainfall variability and evaluating the suitability of different rainfall data sources; (2) investigating streamflow response to rainfall across two fires, snow zones, and years, including key drivers influencing those responses; and (3) examining biases between gridded and observed climate datasets and synthesizing how these biases impact hydrologic modeling results. A central theme throughout was how spatial variability in climate, soils, and catchment attributes shape hydrologic processes in mountainous regions. These effects become even more pronounced in post-fire landscapes, where changes to vegetation and soil increase the already complex interactions in hydrologic responses. Together, these findings highlight the importance of carefully selecting climate data sources and accounting for spatial heterogeneity in post-fire landscapes to improve hydrologic modeling and our understanding of the complex interactions in mountain hydrology.

Chapter 2 examined how elevation influences rainfall patterns across the Rocky Mountains in northern Colorado. Based on prior research, we expected higher rainfall intensities at lower elevations, but instead there was no clear elevation-intensity relationship. The east side of the mountains has experienced the highest rainfall intensities, but the location of these high rainfall intensities varied between years. There is an important tradeoff to consider when choosing between using gridded climate products or installing a local gauge network. While gridded datasets offer broad spatial coverage and a complete time series, they may not accurately capture spatial variability, especially in complex terrain where there could be differences within a square kilometer. Conversely, a gauge network can provide accurate measurements, but if the network is not expansive enough, it may fail to capture the spatial patterns of convective storms. Additionally, purchasing and maintaining a network of gauges can

be costly and requires access to sites for maintenance. Both approaches have limitations, and careful consideration is needed to balance spatial and temporal resolutions.

Chapter 3 evaluated how stream response varies between two fires, snow zones, and year in northern Colorado. Additionally, catchment attributes and soil characteristics were used to evaluate how these variables and burn characteristics influence post-fire streamflow response. We hypothesized that the intermittent snow zone would have greater and faster streamflow responses because of higher rainfall intensities. Second, we hypothesized that East Troublesome fire (ETF) would have greater streamflow response than Cameron Peak fire (CPF) because of the higher clay content at ETF. The results were consistent with our hypotheses; however, valleys and hollows and proportion of area greater than 75% of catchment elevation were important catchment characteristics in shaping stream response. Because of the higher rainfall intensities and steeper terrain, the intermittent snow zone in CPF is more vulnerable to large flash floods because of infiltration excess overland flow. At ETF, few catchments were in the intermittent snow zone, and the rainfall intensities were lower than in CPF. Despite these lower intensities, stream responses were greater due to differences in soil properties and catchment shape. This demonstrates that the underlying soil and catchment properties are a critical component to consider in future post-fire risk assessments.

Chapter 4 examined how different climate data sources (station versus gridded) affect hydrologic model calibration, performance, and behavior in a Colorado Front Range catchment. Because the study catchment had a high density of stations, the station climate inputs yielded better hydrologic model performance at the catchment outlet than the gridded inputs. Water resource managers should consider climate input data carefully prior to modeling since small climate changes could lead to substantial shifts in parameter values resulting in altered water balance components. To improve model accuracy, it is recommended to calibrate not just to streamflow but also internal state variables (e.g., SWE, soil moisture) when measurements are possible. Post-fire impacts, such as reduced canopy cover and altered soil properties, affected

hydrologic processes in burned areas enough that different calibration parameter values were required for burned areas. This highlights the need for ongoing development of approaches to accurately account for post-fire physical processes and to identify reliable parameter values. This could be a combination of field data and remote sensing rather than relying only on user experience to constrain parameters.

Together, these chapters show that hydrologic responses in mountainous, post-fire landscapes are strongly shaped from the interaction between spatially variable climate and landscape, rather than from a single or separate controlling factors. Differences in precipitation patterns, soil properties, vegetation recovery, and catchment morphology combine to produce distinct streamflow responses and recovery, even among fires that are geographically close. At their closest points, the CPF and ETF burned within 16 km of one another, but their post-fire hydrologic behavior diverged due to rainfall and its interactions with soil texture and catchment morphology. The ETF soil had more clay content and less valleys and hollows than at CPF, which likely led to the greater stream responsiveness at ETF. This underscores that post-fire risk inferences would be challenging grouped in a regional context alone (e.g., ecoregions). The proportion of area above 75% elevation did not differ between the two fires but did vary by snow zone, with higher values in the intermittent snow zone contributing to a faster and greater stream response. This likely led to greater streamflow generation because of the larger area of steeply sloped terrain. Therefore, the configuration of slopes within a catchment is also important to consider. This work has highlighted that variables related to catchment morphology also play a role in stream response and researchers should consider more than the common variables (e.g., slope, soil properties) used in post-fire debris flow or erosion models.

The findings demonstrate that while advances in gridded climate datasets and hydrologic models provide powerful tools, they also come with limitations that require knowledgeable consideration. Future research should involve more measurements (e.g., rainfall, streamflow, soil properties) across different hydrologic regimes to account for not only spatial variation within

a region but also differences across regions. For example, more fine spatial scales of soil texture measurements than what SSURGO offers could better inform how soil properties are mediating streamflow response. Additionally, more work is needed to better represent post-fire effects into hydrologic models, including improved parameter constraints and calibrating state variables to observed measurements. This could be a combination of field measurements (e.g., SWE, subsurface flow) and remote sensing (e.g., LAI). Overall, this dissertation shows that effective post-fire risk assessment and hydrologic modeling in mountain environments require an interdisciplinary, spatially explicitly framework that includes climate, hydrologic regimes, and catchment attributes.

REFERENCES

Abatzoglou, J. T. (2013), Development of gridded surface meteorological data for ecological applications and modelling. Int. J. Climatol., 33: 121–131.

Abatzoglou, J. T., & Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. Proceedings of the National Academy of Sciences, 113(42), 11770–11775. https://doi.org/10.1073/pnas.1607171113

Abatzoglou, J.T., Battisti, D.S., Williams, A.P. et al. (2021). Projected increases in western US forest fire despite growing fuel constraints. Commun Earth Environ 2, 227. https://doi.org/10.1038/s43247-021-00299-0

Addington, R. N., G. H. Aplet, M. A. Battaglia, et al. (2018). Principles and practices for the restoration of ponderosa pine and dry mixed-conifer forests of the Colorado Front Range.

Alduchov, O. A., & Eskridge, R. E. (1996). Improved Magnus form approximation of saturation vapor pressure. Journal of Applied Meteorology (1988-2005), 601-609.

Alizadeh, M. R., Abatzoglou, J. T., Luce, C. H., Adamowski, J. F., Farid, A., & Sadegh, M. (2021). Warming enabled upslope advance in western US forest fires. Proceedings of the National Academy of Sciences, 118(22), e2009717118.

Alvarez, O., Guo, Q., Klinger, R. C., Li, W., & Doherty, P. (2014). Comparison of elevation and remote sensing derived products as auxiliary data for climate surface interpolation. International Journal of Climatology, 34(7), 2258–2268. https://doi.org/10.1002/joc.3835

Ascough, J.C., David, O., Krause, P., Heathman, G.C., Kralisch, S., Larose, M., Ahuja, L.R., and Kipka, H., 2012. Development and application of a modular watershed-scale hydrologic model using the Object Modeling System: Runoff response evaluation. Transactions of the ASABE, 55(1), p.117-135.

Babyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. Biopsychosocial Science and Medicine, 66(3), 411-421.

Bales, R. C., Molotch, N. P., Painter, T. H., Dettinger, M. D., Rice, R., & Dozier, J. (2006). Mountain hydrology of the western United States. Water Resources Research, 42(8).

Banta, R. M. (1984). Daytime boundary-layer evolution over mountainous terrain. Part 1: Observations of the dry circulations. Monthly Weather Review, 112(2), 340-356.

Bayabil, H. K., Fares, A., Sharif, H. O., Ghebreyesus, D. T., & Moreno, H. A. (2019). Effects of spatial and temporal data aggregation on the performance of the Multi-Radar Multi-Sensor system. JAWRA Journal of the American Water Resources Association, 55(6), 1492–1504. https://doi.org/10.1111/1752-1688.12799

Beganskas, S. (n.d.). The Geochemical Impact of Wildfire and Mining on the Fourmile Creek Watershed, Colorado.

Behnke, R., Vavrus, S., Allstadt, A., Albright, T., Thogmartin, W. E., & Radeloff, V. C. (2016). Evaluation of downscaled, gridded climate data for the conterminous United States. Ecological Applications, 26(5), 1338–1351. https://doi.org/10.1002/15-1061

Benavides-Solorio, J.D., MacDonald, L.H., 2005. Measurement and prediction of post-fire erosion at the hillslope scale, Colorado Front Range. Int. J. Wildland Fire 14 (4), 457–474. https://doi.org/10.1071/WF05042.

Beven, K. (2006). A manifesto for the equifinality thesis. Journal of hydrology, 320(1-2), 18-36.

Blandford, T. R., Humes, K. S., Harshburger, B. J., Moore, B. C., Walden, V. P., & Ye, H. (2008). Seasonal and synoptic variations in near-surface air temperature lapse rates in a mountainous basin. Journal of Applied Meteorology and Climatology, 47(1), 249–261. https://doi.org/10.1175/2007JAMC1565.1

Bohne, L., Strong, C., & Steenburgh, W. J. (2020). Climatology of orographic precipitation gradients in the contiguous Western United States. Journal of Hydrometeorology, 21(8), 1723–1740. https://doi.org/10.1175/jhm-d-19-0229.1

Burles, K., & Boon, S. (2011). Snowmelt energy balance in a burned forest plot, Crowsnest Pass, Alberta, Canada. Hydrological processes, 25(19), 3012-3029.

Bytheway, J. L., Abel, M. R., Mahoney, K., & Cifelli, R. (2019). A multiscale evaluation of multisensor quantitative precipitation estimates in the Russian River Basin. Journal of Hydrometeorology, 20(3), 447–466. https://doi.org/10.1175/jhm-d-18-0142.1

Cannon, S. H., Boldt, E. M., Laber, J. L., Kean, J. W., & Staley, D. M. (2011). Rainfall intensity–duration thresholds for postfire debris-flow emergency-response planning. Natural Hazards, 59(1), 209–236. https://doi.org/10.1007/s11069-011-9747-2

Ciach, G. J. (2003). Local Random errors in Tipping-Bucket rain gauge measurements. Journal of Atmospheric and Oceanic Technology, 20(5), 752–759. https://doi.org/10.1175/1520-0426(2003)20

Coe, J. A., Kean, J. W., Godt, J. W., Baum, R. L., Jones, E. S., Gochis, D. J., & Anderson, G. S. (2014). New insights into debris-flow hazards from an extraordinary event in the Colorado Front Range. GSA Today, 24(10), 4-10.

Collados-Lara, A. J., Fassnacht, S. R., Pardo-Igúzquiza, E., & Pulido-Velazquez, D. (2021). Assessment of high resolution air temperature fields at rocky mountain national park by combining scarce point measurements with elevation and remote sensing data. Remote Sensing, 13(1), 1–26. https://doi.org/10.3390/rs13010113

Collier, C. (2007). Flash flood forecasting: what are the limits of predictability? Q. J. Roy. Meteor. Soc. 133, 3–23.

Colorado Climate Center, Colorado State University. (2025). Climate at a Glance: Rank Maps [Webpage]. Retrieved October 23, 2025, from https://climate.colostate.edu/co_cag/rank_maps_prism.html.

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H. et al. (2008) Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28, 2031–2064.

DeBano, L. F. (2000). The role of fire and soil heating on water repellency in wildland environments: A review. Journal of Hydrology, 231–232, 195–206. https://doi.org/10.1016/S0022-1694(00)00194-3

Dennison, P. E., Brewer, S. C., Arnold, J. D., & Moritz, M. A. (2014). Large wildfire trends in the western United States, 1984-2011. Geophysical Research Letters, 41(8), 2928–2933. https://doi.org/10.1002/2014gl059576

Doerr, S.H., Shakesby, R.A., MacDonald, L.H., 2009. Soil water repellency: A key factor in post-fire erosion. In: Cerd´a, A., Robichaud, P.R. (Eds.), Fire Effects on Soils and Restoration Strategies. Science Publishers Inc., Enfield, pp. 197–223.

Douglas, Z. A. (2024). Topographic and Diurnal Influences on Storms Associated With Heavy Rainfall in Northern Colorado (Master's thesis, Colorado State University).

Eagleson, P. S. (1970). Dynamic hydrology.

East, A. E., Logan, J. B., Dartnell, P., Lieber-Kotz, O., Cavagnaro, D. B., McCoy, S. W., & Lindsay, D. N. (2021). Watershed sediment yield following the 2018 Carr Fire, Whiskeytown National Recreation Area, northern California. Earth and Space Science, 8(9), e2021EA001828.https://doi.org/10.1029/2021EA001828

Ebel, B. A., Moody, J. A., & Martin, D. A. (2012). Hydrologic conditions controlling runoff generation immediately after wildfire. Water Resources Research, 48(3).

Ebel, B. A. (2020). Temporal evolution of measured and simulated infiltration following wildfire in the Colorado Front Range, USA: Shifting thresholds of runoff generation and hydrologic hazards. Journal of Hydrology, 585, 124765.

Ebel, B. A., Shephard, Z. M., Walvoord, M. A., Murphy, S. F., Partridge, T. F., & Perkins, K. S. (2023). Modeling post-wildfire hydrologic response: Review and future directions for applications of physically based distributed simulation. Earth's Future, 11(2), e2022EF003038.

Elsner, M. M., Gangopadhyay, S., Pruitt, T., Brekke, L. D., Mizukami, N., & Clark, M. P. (2014). How does the choice of distributed meteorological data affect hydrologic model calibration and streamflow simulations?. Journal of Hydrometeorology, 15(4), 1384-1403.

England, J., Godaire, J., Klinger, R., Bauer, T., & Julien, P. (2010). Paleohydrologic bounds and extreme flood frequency of the Upper Arkansas River, Colorado, USA. Geomorphology, 124(1–2), 1–16. https://doi.org/10.1016/j.geomorph.2010.07.021

Esri. (2017). Hillshade basemap accessed via ArcGIS.

FEST (Fourmile Emergency Stabilization Team) (2010), Fourmile Emergency Stabilization Burned Area Report, Boulder County, CO, 14 pp.

Ffolliott, P. F., DeBano, L. F., Baker, M. B., Gottfried, G. J., Solis-Garza, G., Edminster, C. B., Neary, D. G., Allen, L. S., & Hamre, R. H. (1996). Effects of fire on Madrean Province ecosystems: A symposium proceedings; March 11–15, 1996; Tucson, AZ. https://doi.org/10.2737/RM-GTR-289

Fisher, A., Caffo, B., and Zipunnikov, V. (2014). Fast, Exact Bootstrap Principal Component Analysis for p>1 million. http://arxiv.org/abs/1405.0922

Fosser, G., Khodayar, S., & Berg, P. (2015). Benefit of convection permitting climate model simulations in the representation of convective precipitation. Climate Dynamics, 44(1–2), 45–60. https://doi.org/10.1007/s00382-014-2242-1

Ford, T. W., & Quiring, S. M. (2019). Comparison of contemporary in situ, model, and satellite remote sensing soil moisture with a focus on drought monitoring. Water Resources Research, 55(2), 1565-1582.

Fowler, H. J., Lenderink, G., Prein, A. F., Westra, S., Allan, R. P., Ban, N., Barbero, R., Berg, P., Blenkinsop, S., Do, H. X., Guerreiro, S., Haerter, J. O., Kendon, E. J., Lewis, E., Schaer, C., Sharma, A., Villarini, G., Wasko, C., & Zhang, X. (2021). Anthropogenic

intensification of short-duration rainfall extremes. Nature Reviews Earth & Environment, 2(2), 107–122. https://doi.org/10.1038/s43017-020-00128-6

Frankson, R., K.E. Kunkel, L.E. Stevens, D.R. Easterling, N.A. Umphlett, C.J. Stiles, R. Schumacher, and P.E. Goble. (2022): Colorado State Climate Summary 2022. NOAA Technical Report NESDIS 150-CO. NOAA/NESDIS, Silver Spring, MD, 5 pp.

Gao, J., Sheshukov, A. Y., Yen, H., & White, M. J. (2017). Impacts of alternative climate information on hydrologic processes with SWAT: A comparison of NCDC, PRISM and NEXRAD datasets. Catena, 156, 353-364.

Gentine, P., Troy, T. J., Lintner, B. R., & Findell, K. L. (2012). Scaling in surface hydrology: Progress and challenges. Journal of Contemporary Water research & education, 147(1), 28-40.

Giovando, J., & Niemann, J. D. (2022). Wildfire impacts on snowpack phenology in a changing climate within the western US. Water Resources Research, 58(8), e2021WR031569.

Green, T.R., R.H. Erskine, M.L. Coleman, O. David, J.C. Ascough, H. Kipka. 2015. The AgroEcoSystem (AgES) response-function model simulates layered soil-water dynamics in semiarid Colorado: Sensitivity and calibration. Vadose Zone J. 14(8). https://doi.org/10.2136/vzj2014.09.0119

Goeking, S. A., & Tarboton, D. G. (2020). Forests and water yield: A synthesis of disturbance effects on streamflow and snowpack in western coniferous forests. Journal of Forestry, 118(2), 172-192.

Goodrich B, Gabry J, Ali I & Brilleman S. (2024). rstanarm: Bayesian applied regression modeling via Stan. R package version 2.32.1 https://mc-stan.org/rstanarm.

Google Earth Engine. (n.d.). LANDSAT/LC08/C02/T1_L2: Landsat 8 Collection 2 Tier 1 Level-2 Surface Reflectance [Data set]. Google Earth Engine. https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C02_T1_L2 (Accessed September 30, 2024)

Google Earth Engine. (n.d.). LANDSAT/LE07/C02/T1_L2: Landsat 7 Collection 2 Tier 1 Level-2 Surface Reflectance [Data set]. Google Earth Engine. https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LE07_C02_T1_L2 (Accessed September 30, 2024)

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of hydrology, 377(1-2), 80-91.

Gutmann, E. D., Rasmussen, R. M., Liu, C., Ikeda, K., Gochis, D. J., Clark, M. P., Dudhia, J., & Thompson, G. (2012). A comparison of statistical and dynamical downscaling of winter precipitation over complex terrain. Journal of Climate, 25(1), 262–281. https://doi.org/10.1175/2011JCLI4109.1

Hallema, D. W., Sun, G., Caldwell, P. V., Norman, S. P., Cohen, E. C., Liu, Y., Ward, E. J., & McNulty, S. G. (2017). Assessment of wildland fire impacts on watershed annual water yield: Analytical framework and case studies in the United States. Ecohydrology, 10(2). https://doi.org/10.1002/eco.1794

Halton, J. H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Mathematik, 2(1), 84-90.

Hammond, J.C., 2020, Contiguous U.S. annual snow persistence and trends from 2001-2020: U.S. Geological Survey data release, https://doi.org/10.5066/P9U7U5FP.

Harpold, A. A., Biederman, J. A., Condon, K., Merino, M., Korgaonkar, Y., Nan, T., ... & Brooks, P. D. (2014). Changes in snow accumulation and ablation following the Las Conchas Forest Fire, New Mexico, USA. Ecohydrology, 7(2), 440-452.

Harvey, N., Burns, S. P., Musselman, K. N., Barnard, H., & Blanken, P. D. (2025). Identifying canopy snow in subalpine forests: A comparative study of methods. Water Resources Research, 61, e2023WR036996. https://doi.org/10.1029/ 2023WR036996

Hay, L. E, Umemoto, M. (2006). Multiple-objective stepwise calibration using Luca. U.S. Geological Survey Open-File Report 2006–1323. https://pubs.er.usgs.gov/publication/ofr20061323

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

Higuera, P.E. and Abatzoglou, J.T. (2021). Record-setting climate enabled the extraordinary 2020 fire season in the western United States. Glob. Change Biol., 27: 1-2. https://doi.org/10.1111/gcb.15388

Hijmans R., (2025). terra: Spatial Data Analysis. doi:10.32614/CRAN.package.terra <https://doi.org/10.32614/CRAN.package.terra>, R package version 1.8-50, <https://CRAN.R-project.org/package=terra>.

Hossain, F., and E. N. Anagnostou, 2005: Numerical investigation of the impact of uncertainties in satellite rainfall estimation and land surface model parameters on simulation of soil moisture. Adv. Water Resour., 28, 1336–1350, doi:10.1016/j.advwatres.2005.03.013.

Iowa Environmental Mesonet. (2024). MRMS data archive. Iowa State University. Retrieved from https://mtarchive.geol.iastate.edu/

Jacobucci R. (2023). regsem: Regularized Structural Equation Modeling., R package version 1.9.5. doi:10.32614/CRAN.package.regsem https://doi.org/10.32614/CRAN.package.regsem

Jarrett, R. D. (1990). Paleohydrologic techniques used to define the spatial occurrence of floods. Geomorphology, 3(2), 181–195. https://doi.org/10.1016/0169-555x(90)90044-q

Jasiewicz, J., and Stepinski, T. F. (2013). Geomorphons — a pattern recognition approach to classification and mapping of landforms. Geomorphology, 182, 147-156.

Johnson M (2024a). climateR, R package version 0.3.3, https://github.com/mikejohnson51/climateR.

Johnson M (2024b). AOI, R package version 0.3.0, https://github.com/mikejohnson51/AOI/.

Kampf, S. K., Faulconer, J., Shaw, J. R., Lefsky, M., Wagenbrenner, J. W., & Cooper, D. J. (2018). Rainfall thresholds for flow generation in desert ephemeral streams. Water Resources Research, 54(12), 9935–9950. https://doi.org/10.1029/2018wr023714

Kampf, S. K., Gannon, B. M., Wilson, C., Saavedra, F. A., Miller, M. E., Heldmyer, A., et al. (2020). Pemip: Post-fire erosion model inter-comparison project. Journal of Environmental Management, 268, 110704. https://doi.org/10.1016/j.jenvman.2020.110704

Kampf, S. K., McGrath, D., Sears, M. G., Fassnacht, S. R., Kiewiet, L., & Hammond, J. C. (2022). Increasing wildfire impacts on snowpack in the western US. Proceedings of the National Academy of Sciences, 119(39), e2200333119.

Kipka, H. Lighthart, N., David, O., Green, T. R., Patterson, D., & Arabi, M. (2019). The enhanced Catchment areas delineation (Cadel) tool for watershed models with spatially explicit routing between simulated areas. Proc. Hydrology Days, March 27–29, 2019, Fort Collins, CO.

Konings, J., & Habte, A. (2016). Uncertainty evaluation of measurements with pyranometers and pyrheliometers (No. NREL/CP-5D00-68394). National Renewable Energy Lab.(NREL), Golden, CO (United States).

LANDFIRE, 2016, Existing Vegetation Type Layer, LANDFIRE 2.0.0, U.S. Department of the Interior, Geological Survey, and U.S. Department of Agriculture. Accessed 15 January 2025 at http://www.landfire/viewer.

Lanza, L. G., & Stagi, L. (2008). Certified accuracy of rainfall data as a standard requirement in scientific investigations. Advances in Geosciences, 16, 43–48. https://doi.org/10.5194/adgeo-16-43-2008

Larimer County. (2024). Novastar systems map. Retrieved from https://larimerco-ns5.trilynx-novastar.systems/novastar/operator/map

Lavee, H., Kutiel, P., Segev, M., Benyamini, Y., 1995. Effect of surface roughness on runoff and erosion in a mediterranean ecosystem: the role of fire. Geomorphology 11 (3), 227–2384. https://doi.org/10.1016/0169-555X(94)00059-Z.

Le, P. V., & Kumar, P. (2014). Power law scaling of topographic depressions and their hydrologic connectivity. Geophysical Research Letters, 41(5), 1553-1559.

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.

Liu, T., McGuire, L. A., Oakley, N., & Cannon, F. (2022). Temporal changes in rainfall intensity–duration thresholds for post-wildfire flash floods in southern California. Natural Hazards and Earth System Sciences, 22(2), 361–376. https://doi.org/10.5194/nhess-22-361-2022

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

Lundquist, J. D., & Cayan, D. R. (2007). Surface temperature patterns in complex terrain: Daily variations and long-term change in the central Sierra Nevada, California. Journal of Geophysical Research Atmospheres, 112(11), 1–15. https://doi.org/10.1029/2006JD007561

Lute, A. C., & Abatzoglou, J. T. (2020). Best practices for estimating near-surface air temperature lapse rates. International Journal of Climatology, November 2019, 1–16. https://doi.org/10.1002/joc.6668

Maggioni, V., Anagnostou, E. N., & Reichle, R. H. (2012). The impact of model and rainfall forcing errors on characterizing soil moisture uncertainty in land surface modeling. Hydrology and Earth System Sciences, 16(10), 3499-3515.

Mahoney, K., Ralph, F. M., Wolter, K., Doesken, N., Dettinger, M., Gottas, D., Coleman, T., & White, A. (2015). Climatology of extreme daily precipitation in Colorado and its diverse spatial and seasonal variability. Journal of Hydrometeorology, 16(2), 781–792. https://doi.org/10.1175/jhm-d-14-0112.1

Mankin, K. R., Mehan, S., Green, T. R., & Barnard, D. M. (2025). Review of gridded climate products and their use in hydrological analyses reveals overlaps, gaps, and the need for a more objective approach to selecting model forcing datasets. Hydrology and Earth System Sciences, 29(1), 85-108. https://doi.org/10.5194/hess-2024-58

McEvoy, D. J., Mejia, J. F., & Huntington, J. L. (2014). Use of an observation network in the Great Basin to evaluate gridded climate data. Journal of Hydrometeorology, 15(5), 1913–1931. https://doi.org/10.1175/jhm-d-14-0015.1

McGrath, D., Zeller, L., Bonnell, R., Reis, W., Kampf, S., Williams, K., ... & Rittger, K. (2023). Declines in peak snow water equivalent and elevated snowmelt rates following the 2020 Cameron Peak wildfire in northern Colorado. Geophysical Research Letters, 50(6), e2022GL101294.

McMillan, H. K., Westerberg, I. K., & Krueger, T. (2018). Hydrological data uncertainty and its implications. Wiley Interdisciplinary Reviews: Water, 5(6), e1319.

McGuire, L., Rengers, F. K., Kean, J. W., & Staley, D. M. (2017). Debris flow initiation by runoff in a recently burned basin: Is grain-by-grain sediment bulking or en masse failure to blame? Geophysical Research Letters, 44(14), 7310–7319. https://doi.org/10.1002/2017GL074243

Michaud, J. D., & Sorooshian, S. (1994). Effect of rainfall-sampling errors on simulations of desert flash floods. Water Resources Research, 30(10), 2765–2775. https://doi.org/10.1029/94wr01273

Miller, Q., Barnard, D., Sears, M., Hammond, J., & Kampf, S. (2025). Variability in Hydrologic Response to Wildfire between Snow Zones in Forested Headwaters. Manuscript in review at Hydrological Processes.

Mizukami, N., P. Clark, M., G. Slater, A., D. Brekke, L., M. Elsner, M., R. Arnold, J., & Gangopadhyay, S. (2014). Hydrologic implications of different large-scale meteorological model forcing datasets in mountainous regions. Journal of Hydrometeorology, 15(1), 474-488.

Moazami, S., & Najafi. (2021). A comprehensive evaluation of GPM-IMERG V06 and MRMS with hourly ground-based precipitation observations across Canada. Journal of Hydrology, 594, 125929. https://doi.org/10.1016/j.jhydrol.2020.125929

Molini, A., Lanza, L. G., & La Barbera, P. (2005). The impact of tipping-bucket rain gauge measurement errors on design rainfall for urban-scale applications. Hydrological Processes, 19(5), 1073–1088. https://doi.org/10.1002/hyp.5646

Moody, J. A., & Martin, D. A. (2001). Initial hydrologic and geomorphic response following a wildfire in the Colorado front range. Earth Surface Processes and Landforms, 26(10), 1049–1070. https://doi.org/10.1002/esp.253

Moody John A., Martin Deborah A. (2009) Synthesis of sediment yields after wildland fire in different rainfall regimes in the western United States. International Journal of Wildland Fire 18, 96-115.

Moody, J. A., & Ebel, B. A. (2012). Hyper-dry conditions provide new insights into the cause of extreme floods after wildfire. CATENA, 93, 58–63. https://doi.org/10.1016/j.catena.2012.01.006

Moody, J. A., Shakesby, R. A., Robichaud, P. R., Cannon, S. H., & Martin, D. A. (2013). Current research issues related to post-wildfire runoff and erosion processes. Earth-Science Reviews, 122, 10–37. https://doi.org/10.1016/j.earscirev.2013.03.004

Moody, J. A., Ebel, B. A., Nyman, P., Martin, D. A., Stoof, C., & McKinley, R. (2016). Relations between soil hydraulic properties and burn severity. International Journal of Wildland Fire, 25(3), 279. https://doi.org/10.1071/WF14062

Moore, C., Kampf, S., Stone, B., & Richer, E. (2015). A GIS-based method for defining snow zones: application to the western United States. Geocarto International, 30(1), 62-81.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, 50(3), 885-900.

MTBS Data Access: Fire Level Geospatial Data. (2017, July - last revised). MTBS Project (USDA Forest Service/U.S. Geological Survey). Available online: http://mtbs.gov/direct-download

Murphy, S. F., Verplanck, P. L., & Barber, L. B. (2003). Comprehensive water quality of the Boulder Creek Watershed, Colorado, during high-flow and low-flow conditions, 2000 (Vol. 3, No. 4054). US Department of the Interior, US Geological Survey.

Murphy, S. F., Writer, J. H., McCleskey, R. B., & Martin, D. A. (2015). The role of precipitation type, intensity, and spatial distribution in source water quality after wildfire. Environmental Research Letters, 10(8). https://doi.org/10.1088/1748-9326/10/8/084007

Murphy, S. F., McCleskey, R. B., Martin, D. A., Writer, J. H., & Ebel, B. A. (2018). Fire, flood, and drought: Extreme climate events alter flow paths and stream chemistry. Journal of Geophysical Research: Biogeosciences, 123(8), 2513–2526. https://doi.org/10.1029/2017JG004349

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. Journal of hydrology, 10(3), 282-290.

Navarro-Serrano, F., López-Moreno, J. I., Azorín-Molina, C., Alonso-González, E., Tomás-Burguera, M., Sanmiguel-Vallelado, A., ... & Vicente Serrano, S. M. (2018). Estimation of near-surface air temperature lapse rates over continental Spain and its mountain areas.

Neale, T., & Weir, J. K. (2015). Navigating scientific uncertainty in wildfire and flood risk mitigation: A qualitative review. International Journal of Disaster Risk Reduction, 13, 255–265. https://doi.org/10.1016/j.ijdrr.2015.06.010

Neary, D.G., Gottfried, G.J., 2002. Fires and floods: post-fire watershed responses. In: Viegas, V.R. (Ed.), Forest Fire Research & Wildland Fire Safety. Millpress, Rotterdam, pp. 1-9.

Niemeyer, R. J., Bladon, K. D., & Woodsmith, R. D. (2020). Long-term hydrologic recovery after wildfire and post-fire forest management in the interior Pacific Northwest. Hydrological Processes, 34(5), 1182–1197. https://doi.org/10.1002/hyp.13665

Nikolopoulos, E. I., Bartsotas, N. S., Anagnostou, E. N., & Kallos, G. (2015). Using high-resolution numerical weather forecasts to improve remotely sensed rainfall estimates: The case of the 2013 Colorado flash flood. Journal of Hydrometeorology, 16(4), 1742-1751.

Nolan, R. H., Lane, P. N., Benyon, R. G., Bradstock, R. A., & Mitchell, P. J. (2014). Changes in evapotranspiration following wildfire in resprouting eucalypt forests. Ecohydrology, 7(5), 1363-1377.

Nunes, J. P., Doerr, S. H., Sheridan, G., Neris, J., Santín, C., Emelko, M. B., Silins, U., Robichaud, P. R., Elliot, W. J., & Keizer, J. (2018). Assessing water contamination risk from vegetation fires: Challenges, opportunities and a framework for progress. Hydrological Processes, 32(5), 687–694. https://doi.org/10.1002/hyp.11434

Nyman, P., Sheridan, G. J., Smith, H. G., & Lane, P. N. J. (2011). Evidence of debris flow occurrence after wildfire in upland catchments of south-east Australia. Geomorphology, 125(3), 383–401. https://doi.org/10.1016/j.geomorph.2010.10.016

Oakley, N. S. (2021). A warming climate adds complexity to post-fire hydrologic hazard planning. Earth's Futures, 9(7), e2021EF002149. https://doi.org/10.1029/2021EF002149

Omernik, J.M. 1987. Ecoregions of the conterminous United States. Map (scale 1:7,500,000). Annals of the Association of American Geographers 77(1):118-125.

Ouyang, L., Lu, H., Yang, K., Leung, L. R., Wang, Y., Zhao, L., Zhou, X., Lazhu, N., Chen, Y., Jiang, Y., & Yao, X. (2021). Characterizing uncertainties in ground “Truth” of precipitation over complex terrain through High-Resolution numerical modeling. Geophysical Research Letters, 48(10). https://doi.org/10.1029/2020gl091950

Partridge, T. F., Johnson, Z. C., Sleeter, R. R., Qi, S. L., Walvoord, M. A., Murphy, S. F., ... & Ebel, B. A. (2024). Opportunities and challenges for precipitation forcing data in post-wildfire hydrologic modeling applications. Wiley Interdisciplinary Reviews: Water, 11(5), e1728.

Paul, M. J., LeDuc, S. D., Lassiter, M. G., Moorhead, L. C., Noyes, P. D., & Leibowitz, S. G. (2022). Wildfire induces changes in receiving waters: A review with considerations for water quality management. Water Resources Research, 58(9), e2021WR030699. https://doi.org/10.1029/2021WR030699

Pausas, J.G., Llovet, J., Rodrigo, A., Vallejo, R., 2008. Are wildfires a disaster in the Mediterranean basin?—a review. International Journal of Wildland Fire 17, 713–723.

Pebesma E., Bivand R. (2023). Spatial Data Science: With applications in R. Chapman and Hall/CRC, London. doi:10.1201/9780429459016 <https://doi.org/10.1201/9780429459016>, <https://r-spatial.org/book/>.

Pielke Sr, R. A., Pitman, A., Niyogi, D., Mahmood, R., McAlpine, C., Hossain, F., et al. (2011). Land use/land cover changes and climate: modeling analysis and observational evidence. Wiley Interdisciplinary Reviews: Climate Change, 2(6), 828-850.

Poulos, H. M., Barton, A. M., Koch, G. W., Kolb, T. E., & Thode, A. E. (2021). Wildfire severity and vegetation recovery drive post-fire evapotranspiration in a southwestern pine-oak forest, Arizona, USA. Remote Sensing in Ecology and Conservation, 7(4), 579-591.

Prein, A. F., Rasmussen, R. M., Ikeda, K., Liu, C., Clark, M. P., & Holland, G. J. (2017). The future intensification of hourly precipitation extremes. Nature Climate Change, 7(1), 48–52. https://doi.org/10.1038/nclimate3168

Price, K., Purucker, S. T., Kraemer, S. R., Babendreier, J. E., & Knightes, C. D. (2014). Comparison of radar and gauge precipitation data in watershed models across varying spatial and temporal scales. Hydrological Processes, 28(9), 3505-3520.

PRISM Group, Oregon State University. (2025). PRISM 30-year normals (monthly), 1991–2020 [800 m]. Oregon State University. https://prism.oregonstate.edu/normals/

Radcliffe, D. E., & Mukundan, R. (2017). PRISM vs. CFSR precipitation data effects on calibration and validation of SWAT models. JAWRA Journal of the American Water Resources Association, 53(1), 89-100.

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., Black, J., Theriault, J.M., Kucera, P., Gochis, D., Smith, C., Nitu, R., Hall, M., Ikeda, K., & Gutmann, E. (2012). How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed. Bulletin of the American Meteorological Society, 93(6), 811-829.

Rivera-Giboyeaux, A. M., & Weinbeck, S. (2024). Evaluation of WSR-88D Level III and MRMS rainfall estimates against rain gauge observations at the Savannah River site. Journal of Operational Meteorology, 93–106. https://doi.org/10.15191/nwajom.2024.1207

Rocca, M. E., Brown, P. M., MacDonald, L. H., & Carrico, C. M. (2014). Climate change impacts on fire regimes and key ecosystem services in Rocky Mountain forests. Forest Ecology and Management, 327, 290-305. http://dx.doi.org/10.1016/j.foreco.2014.04.005

Rolland, C. (2003). Spatial and seasonal variations of air temperature lapse rates in alpine regions. Journal of Climate, 16(7), 1032–1046. https://doi.org/10.1175/1520-0442(2003)016<1032:SASVOA>2.0.CO;2

Rossi, M. W., Anderson, R. S., Anderson, S. P., & Tucker, G. E. (2019). Orographic controls on subdaily rainfall statistics and flood frequency in the Colorado Front Range, USA. Geophysical Research Letters, 47(4). https://doi.org/10.1029/2019gl085086

Saxe, S., Hogue, T. S., & Hay, L. (2018). Characterization and evaluation of controls on post-fire streamflow response across western US watersheds. Hydrology and Earth System Sciences, 22(2), 1221-1237.

Serrano-Notivoli, R., & Tejedor, E. (2021). From rain to data: A review of the creation of monthly and daily station-based gridded precipitation datasets. Wiley Interdisciplinary Reviews Water, 8(6). https://doi.org/10.1002/wat2.1555

Shen, Y. J., Shen, Y., Goetz, J., & Brenning, A. (2016). Spatial-temporal variation of near-surface temperature lapse rates over the Tianshan mountains, central Asia. Journal of Geophysical Research, 121(23), 14,006-14,017. https://doi.org/10.1002/2016JD025711

Shuai, P., Chen, X., Mital, U., Coon, E. T., & Dwivedi, D. (2021). The effects of spatial and temporal resolution of gridded meteorological forcing on watershed hydrological responses. Hydrology and Earth System Sciences Discussions, 2021, 1-43.

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Liu, Z., Berner, J., Wang, W., Powers, J.G., Duda, M.G., Barker, D.M. and Huang, X.Y., 2019. A description of the advanced research WRF model version 4. National Center for Atmospheric Research: Boulder, CO, USA, 145(145), p.550.

Staley, D. M., Negri, J. A., Kean, J. W., Laber, J. L., Tillery, A. C., & Youberg, A. M. (2016). Updated logistic regression equations for the calculation of post-fire debris-flow likelihood in the western United States [Report] (2016–1106). (Open-File Report, Issue). U.S.G. Survey. https://pubs.usgs.gov/publication/ofr20161106

Staley, D. M., Negri, J. A., Kean, J. W., Laber, J. L., Tillery, A. C., & Youberg, A. M. (2017). Prediction of spatially explicit rainfall intensity duration thresholds for post-fire debris-flow generation in the western United States. Geomorphology, 278, 149–162. https://doi.org/10.1016/j.geomorph.2016.10.019

Stoof, C. R., Vervoort, R. W., Iwema, J., Van Den Elsen, E., Ferreira, A. J. D., & Ritsema, C. J. (2012). Hydrological response of a small catchment burned by experimental fire. Hydrology and Earth System Sciences, 16(2), 267-285.

Strachan, S., & Daly, C. (2017). Testing the daily PRISM air temperature model on semiarid mountain slopes. Journal of Geophysical Research: Atmospheres, 122(11), 5697-5715.

Thornton, P. E., Running, S. W., & White, M. A. (1997). Generating surfaces of daily meteorological variables over large regions of complex terrain. Journal of hydrology, 190(3-4), 214-251.

Thornton, P. E., & Running, S. W. (1999). An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. Agricultural and forest meteorology, 93(4), 211-228.

Tercek, M. T., Rodman, A., Woolfolk, S., Wilson, Z., Thoma, D., & Gross, J. (2021). Correctly applying lapse rates in ecological studies: comparing temperature observations and gridded data in Yellowstone. Ecosphere, 12(3). https://doi.org/10.1002/ecs2.3451

Toth, J. J., & Johnson, R. H. (1985). Summer surface flow characteristics over northeast Colorado. Monthly weather review, 113(9), 1458-1469.

Touma, D., Stevenson, S., Swain, D. L., Singh, D., Kalashnikov, D. A., & Huang, X. (2022). Climate change increases risk of extreme rainfall following wildfire in the western United States. Science Advances, 8(13), eabm0320. https://doi.org/10.1126/sciadv.abm0320

Úbeda, X., Outeiro, L., 2009. Physical and chemical effects of fire on soil. In: Cerd´a, A., Robichaud, P.R. (Eds.), Fire Effects on Soils and Restoration Strategies. Science Publishers Inc., Enfield, NH, pp. 105–132.

U.S. Geological Survey. 2024b. USGS 3D Elevation Program 1--meter LiDAR Digital Elevation Model. https:// elevation.nationalmap.gov/arc-gis/rest/services/3DEPElevation/Image Server.

U.S. Geological Survey. (2022a). USGS 1/3 Arc Second n40w107 20220216: U.S. Geological Survey.

U.S. Geological Survey. (2022b). USGS 1/3 Arc Second n41w106 20220216: U.S. Geological Survey.

U.S. Geological Survey. (2023a). USGS 1/3 Arc Second n40w106 20230602: U.S. Geological Survey.

U.S. Geological Survey. (2023b). USGS 1/3 Arc Second n41w106 20230314: U.S. Geological Survey.

U.S. Geological Survey. (n.d.). National Geologic Map Database. https://ngmdb.usgs.gov/ (Accessed February 1, 2025)

U.S. Geological Survey. (2023). NLCD 2021 Land Cover, CONUS [Data set]. Multi-Resolution Land Characteristics (MRLC) Consortium. https://www.mrlc.gov/data

Vasiloff, S. V., Howard, K. W., & Zhang, J. (2009). Difficulties with Correcting Radar Rainfall Estimates Based on Rain Gauge Data: A Case Study of Severe Weather in Montana on 16–17 June 2007. Weather and Forecasting, 24(5), 1334–1344. https://doi.org/10.1175/2009waf2222154.1

Vehtari, A., Gelman, A. & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput 27, 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. PubMed, 37(5), 360–363. https://pubmed.ncbi.nlm.nih.gov/15883903

Vieira, D. C. S., Serpa, D., Nunes, J. P. C., Prats, S. A., Neves, R., & Keizer, J. J. (2018). Predicting the effectiveness of different mulching techniques in reducing post-fire runoff and erosion at plot scale with the RUSLE, MMF and PESERA models. Environmental Research, 165, 365–378. https://doi.org/10.1016/j.envres.2018.04.029

Viviroli, D., Dürr, H. H., Messerli, B., Meybeck, M., & Weingartner, R. (2007). Mountains of the world, water towers for humanity: Typology, mapping, and global significance. Water resources research, 43(7).

Vivoni, E.R., Entekhabi, D., Bras, R.L., Ivanov, V.Y., Van Horne, M.P., Grassotti, C., Hoffman, R.N., 2006. Extending the predictability of hydrometeorological flood events using radar rainfall nowcasting. J. Hydrometeorol. 7, 660–677.

Wagenbrenner, J. W., & Robichaud, P. R. (2014). Post-fire bedload sediment delivery across spatial scales in the interior western United States. Earth Surface Processes and Landforms, 39(7), 865–876. https://doi.org/10.1002/esp.3488

Walton, D., & Hall, A. (2018). An Assessment of High-Resolution Gridded Temperature Datasets over California. Journal of Climate, 31(10), 3789–3810. https://doi.org/10.1175/jcli-d-17-0410.1

Wells, R., Mankin, K. R., Niemann, J. D., Kipka, H., Green, T. R., & Barnard, D. M. (2024). Estimating changes in streamflow attributable to wildfire in multiple watersheds using a semi-distributed watershed model. Ecohydrology. https://doi.org/10.1002/eco.2697

Westerling, A. L., Hidalgo, H. G., Cayan, D. R., & Swetnam, T. W. (2006). Warming and earlier spring increase western US forest wildfire activity. Science, 313(5789), 940–943. https://doi.org/10.1126/science.1128834

White, P., & Nelson, P. A. (2024). Evaluation of Sub-Hourly MRMS quantitative precipitation estimates in mountainous terrain using machine learning. Water Resources Research, 60(12). https://doi.org/10.1029/2024wr037437

Wickham, H. (2023). httr: Tools for Working with URLs and HTTP_. doi:10.32614/CRAN.package.httr <https://doi.org/10.32614/CRAN.package.httr>, R package version 1.4.7, <https://CRAN.R-project.org/package=httr>.

Wieczorek, M.E., 2014, Area- and Depth-Weighted Averages of Selected SSURGO Variables for the Conterminous United States and District of Columbia: U.S. Geological Survey data release, https://doi.org/10.5066/P92JJ6UJ.

Williams, C. H., Silins, U., Spencer, S. A., Wagner, M. J., Stone, M., & Emelko, M. B. (2019). Net precipitation in burned and unburned subalpine forest stands after wildfire in the northern Rocky Mountains. International Journal of Wildland Fire, 28(10), 750-760.

Wilson, C., Kampf, S. K., Wagenbrenner, J. W., & MacDonald, L. H. (2018). Rainfall thresholds for post-fire runoff and sediment delivery from plot to watershed scales. Forest Ecology and Management, 430, 346–356. https://doi.org/10.1016/j.foreco.2018.08.025

Wolyn, P. G., & McKee, T. B. (1994). The mountain-plains circulation east of a 2-km-high north–south barrier. Monthly weather review, 122(7), 1490-1508.

Wu Q, Brown A (2022). 'whitebox': 'WhiteboxTools' R Frontend. R package version 2.2.0, https://CRAN.R-project.org/package=whitebox.

Yu, G., Wright, D. B., & Holman, K. D. (2021). Connecting hydrometeorological processes to low-probability floods in the mountainous Colorado front range. Water Resources Research, 57(4), e2021WR029768.

Zhang, J., Qi, Y., Langston, C., & Kaney, B. (2012). Radar quality index (RQI) – A combined measure for beam blockage and VPR effects in a national network. Weather Radar and Hydrology, 351, 388–393.

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., Grams, H., Wang, Y., Cocks, S., Martinaitis, S., Arthur, A., Cooper, K., Brogden, J., & Kitzmiller, D. (2016). Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: initial operating capabilities. Bulletin of the American Meteorological Society, 97(4), 621–638. https://doi.org/10.1175/bams-d-14-00174.1

Zou H., Hastie, T. (2005). Regularization and Variable Selection Via the Elastic Net, Journal of the Royal Statistical Society Series B: Statistical Methodology, Volume 67, Issue 2, Pages 301–320, https://doi.org/10.1111/j.1467-9868.2005.00503.x

APPENDIX A: Chapter 2 supplementary information

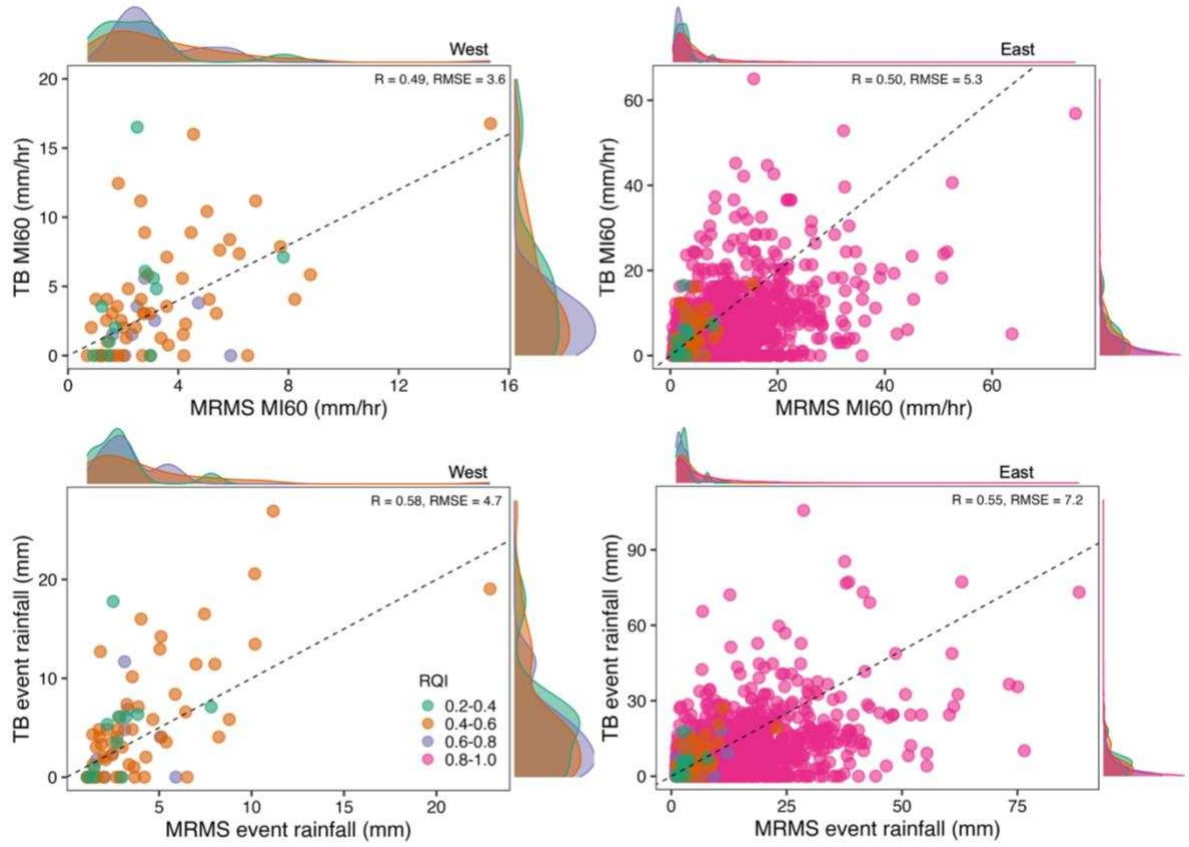

Figure A-1. Tipping bucket compared to corresponding MRMS pixel for MI60 (mm/hr) and event rainfall (mm). The black dotted line is the one-to-one line. The points are colored by the average RQI over the event.

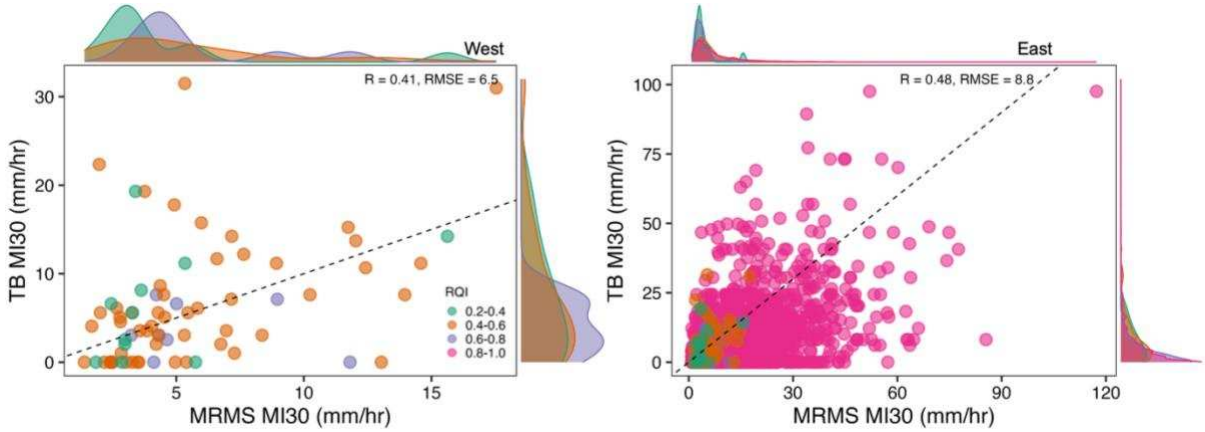

Figure A-2. The 30-minute maximum intensity (MI30) tipping bucket compared to corresponding Multi-Radar/Multi-Sensory System (MRMS) pixel. The black dotted line is the one-to-one line. The points are colored by the average radar quality index over the event.

Table A-1. Kappa statistic for 60-minute maximum intensity (MI60), MI30, and total event rainfall from the logistic regressions predicting stream response. Counts of positive and negative stream responses for each precipitation source are included. Values in bold indicate substantial agreement or better ((Viera and Garrett, 2005).

Site	MI60 (mm/hr)		MI30 (mm/hr)		Event rainfall total (mm)		Positive response	Total	Positive response	Total
	TB	MRMS	TB	MRMS	TB	MRMS	TB		MRMS	
Bighorn	0.39	0.00	0.39	0.00	-0.01	0.49	3	134	3	101
BL4	**0.85**	**0.85**	0.56	**0.85**	0.56	**1.00**	4	112	4	91
Dry	0.50	0.36	0.50	0.27	0.16	0.16	9	105	9	115
ME	**1.00**	**0.77**	**1.00**	**0.77**	**0.87**	**0.62**	5	36	5	36
MM	0.00	**0.65**	0.00	**1.00**	**0.65**	**0.65**	2	28	2	28
Montgomery	0.10	0.14	0.10	0.14	0.50	0.37	12	77	12	93
Mt. Campus	0.00	-0.02	0.00	-0.02	0.17	0.33	8	99	8	81
MW	**0.64**	**1.00**	**0.64**	**1.00**	**0.64**	**0.63**	2	19	2	15
UE	0.35	**0.79**	0.48	**0.79**	0.35	0.38	3	34	3	92
UM	0.00	0.55	0.00	0.55	0.00	0.30	4	57	4	62
UW	0.48	0.23	0.00	0.23	-0.02	0.42	6	78	6	97
Washout	0.00	0.00	0.00	0.00	0.00	0.00	3	66	3	68
hm	0.47	0.55	0.58	0.35	**0.74**	0.40	6	20	6	9
hum	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	4	20	4	9
lm	0.59	0.32	0.59	-0.14	**1.00**	0.32	5	25	5	14
lpm	0.43	0.09	**0.60**	-0.04	**0.65**	0.09	6	25	6	14
lum	**0.60**	0.45	**0.90**	0.18	**0.80**	0.45	7	25	7	15
mm_et	0.00	0.23	0.30	0.19	-0.03	0.19	10	65	11	43
mpm	0.14	0.30	0.08	0.30	0.04	-0.05	21	70	21	43
mub	**0.88**	0.00	**0.78**	0.00	**0.89**	0.30	5	69	5	34
mum	0.52	**0.83**	0.25	**0.74**	0.46	**0.64**	8	70	8	38
p1	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	1	55	1	23
p2	**0.70**	0.31	**0.70**	0.41	0.50	0.12	6	55	6	25

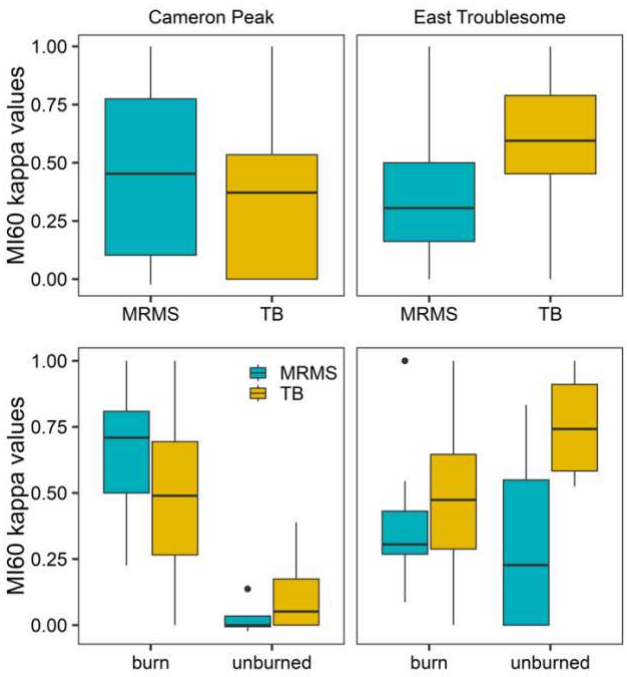

Figure A-3. MI60 kappa statistic from the logistic regression predicting stream response by fire.

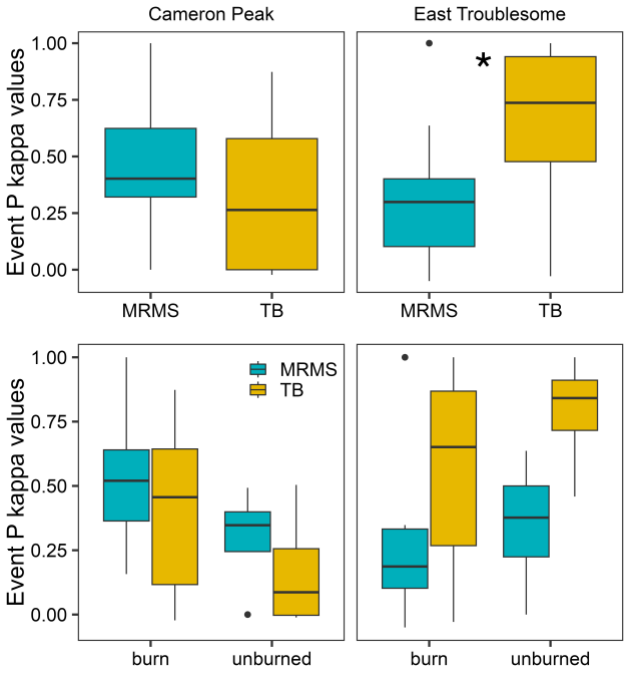

Figure A-4. Event rainfall (P) kappa statistic from the logistic regression predicting stream response by fire. The asterisk represents significant difference determined by the Wilcoxon-rank sum test (* indicates p-value < 0.05).

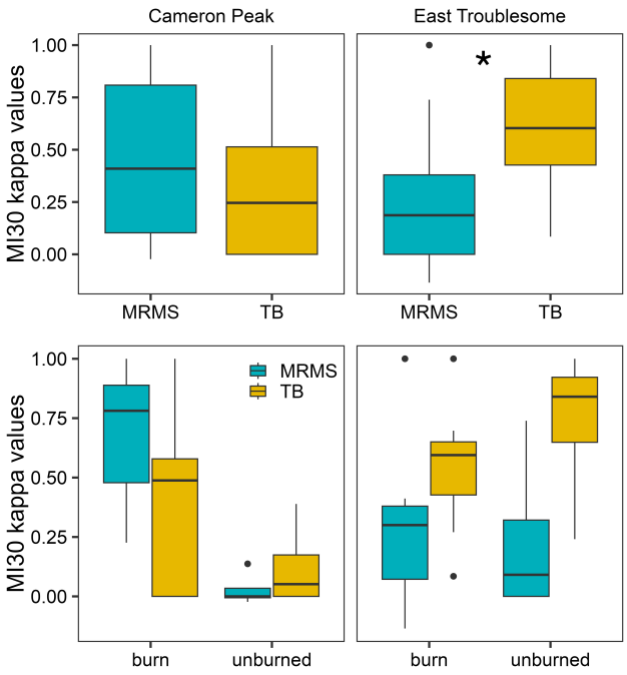

Figure A-5. The MI30 kappa statistic from the logistic regression predicting stream response by fire. The left plots represent the kappa statistics by precipitation source. The asterisk represents significant difference determined by the Wilcoxon-rank sum test (* indicates p-value < 0.05).

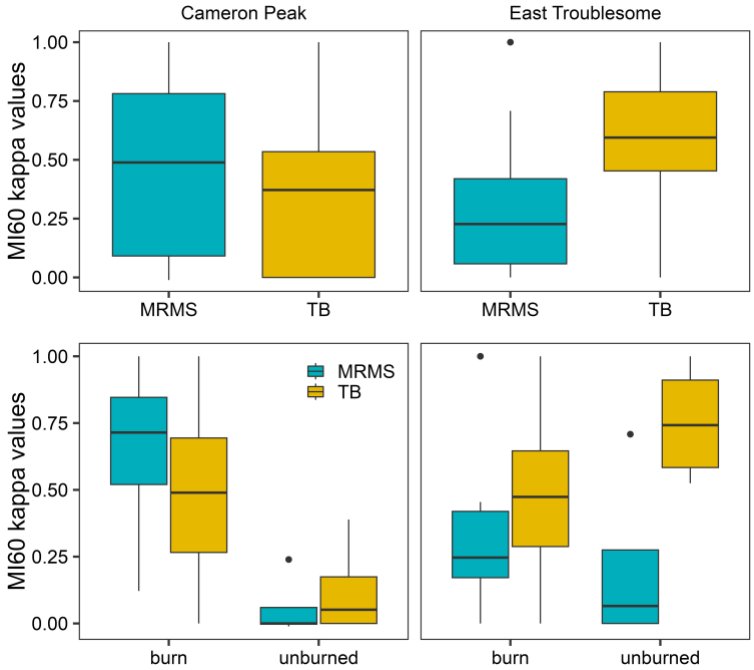

Figure A-6. The MI60 kappa statistic when using maximum MRMS value from the logistic regression predicting stream response by fire. The left plots represent the kappa statistics by precipitation source.

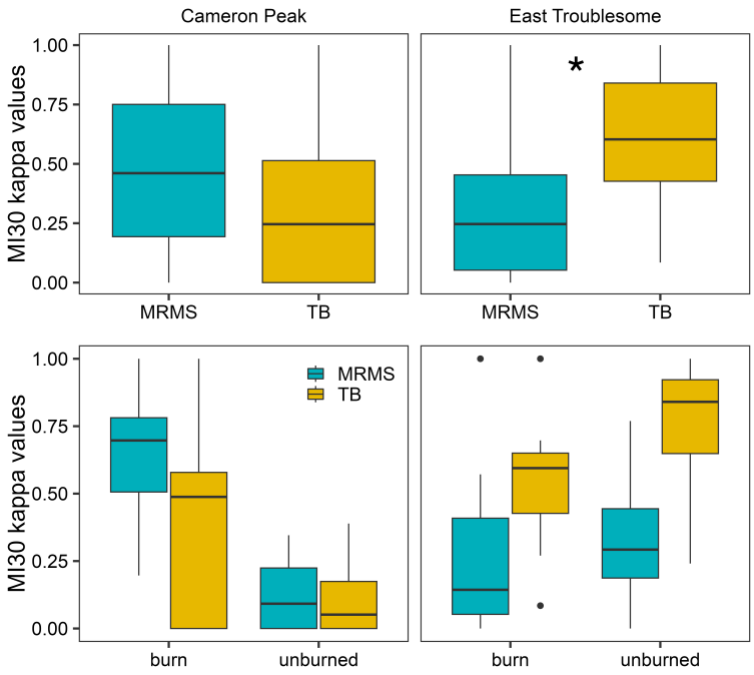

Figure A-7. The MI30 kappa statistic when using maximum MRMS value from the logistic regression predicting stream response by fire. The left plots represent the kappa statistics by precipitation source. The asterisk represents significant difference determined by the Wilcoxon-rank sum test (* indicates p-value < 0.05).

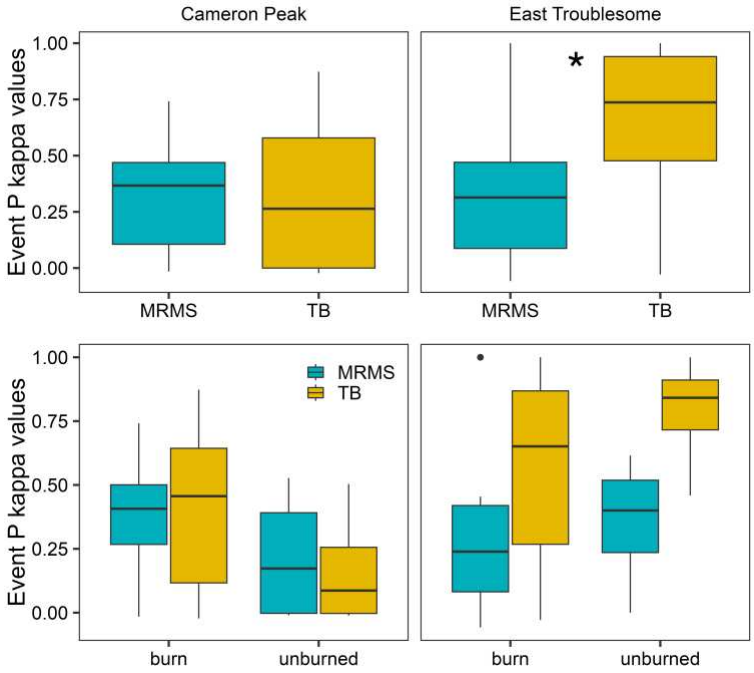

Figure A-8. Event rainfall kappa statistic when using maximum MRMS value from the logistic regression predicting stream response by fire. The left plots represent the kappa statistics by precipitation source. The asterisk represents significant difference determined by the Wilcoxon-rank sum test (* indicates p-value < 0.05).

APPENDIX B: Chapter 3 supplementary information

Table B-1. Catchment variables considered but not used in the study due to correlation with other catchment variables. The LiDAR datasets were downloaded from the USGS National Map at 1 m spatial resolution.

Name	Unit	Source
Percent silt	%	SSURGO
Available water storage	cm water / cm soil	SSURGO
Minimum bedrock depth	cm	SSURGO
Related circumscribing circle	-	Derived from LiDAR using Whitebox Tools
Hydraulic function connectivity	-	Moody 2012
Hypsometric integral	-	Derived from LiDAR using Whitebox Tools
Hypsometric 50%	-	Derived from LiDAR using Whitebox Tools
Hypsometric 25%	-	Derived from LiDAR using Whitebox Tools
Fraction of area containing glacial deposits	-	USGS National Geologic Map Database
Elongation ratio	-	Derived from LiDAR using Whitebox Tools
Fraction of area with slopes > 30	-	Derived from LiDAR
Index of connectivity	-	USGS 1 m LiDAR using WhiteboxTools, Borselli et al., 2008
Fraction of catchment containing recharge areas	-	USGS 1 m LiDAR using Whitebox Tools

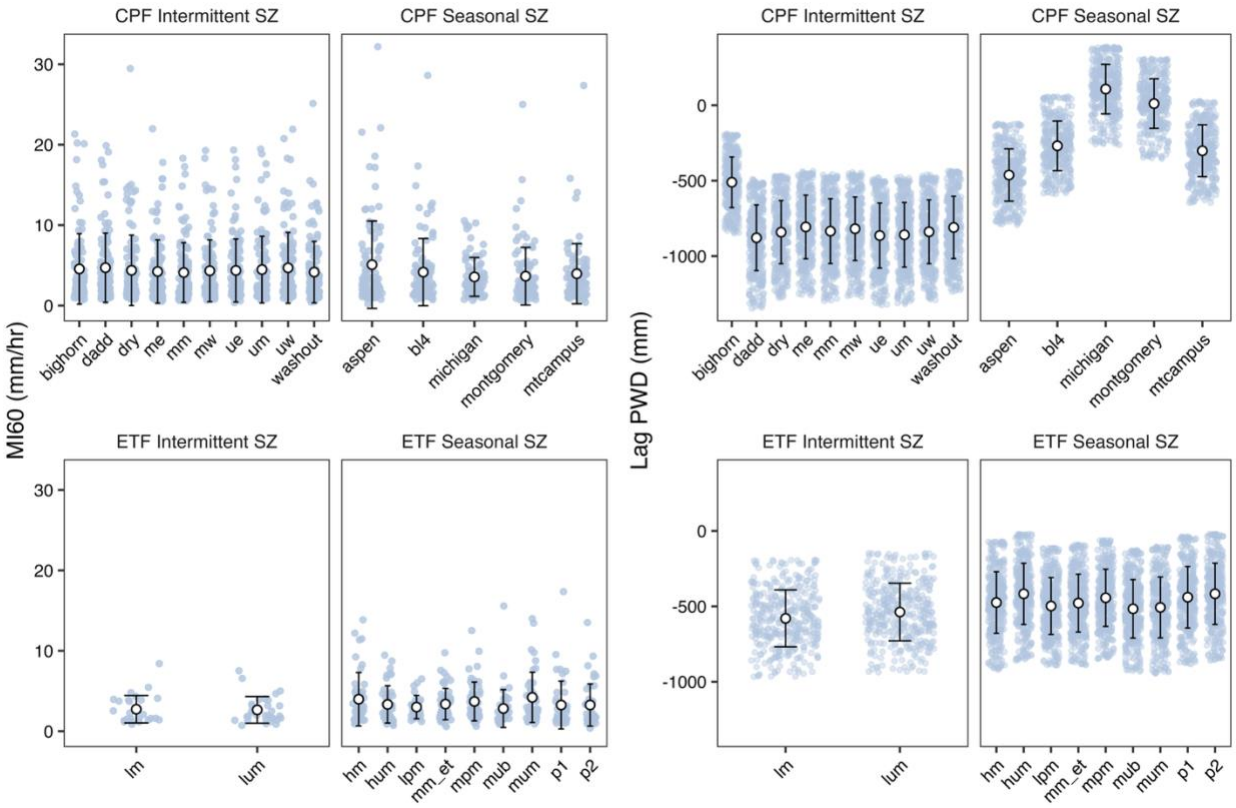

Figure B-1. The 60-minute maximum intensity (MI60) and lag potential water deficit (PWD) for each site used in the study. The Cameron Peak Fire (CPF) sites include summers 2021 through 2023 and East Troublesome (ETF) sites summers 2022 and 2023.

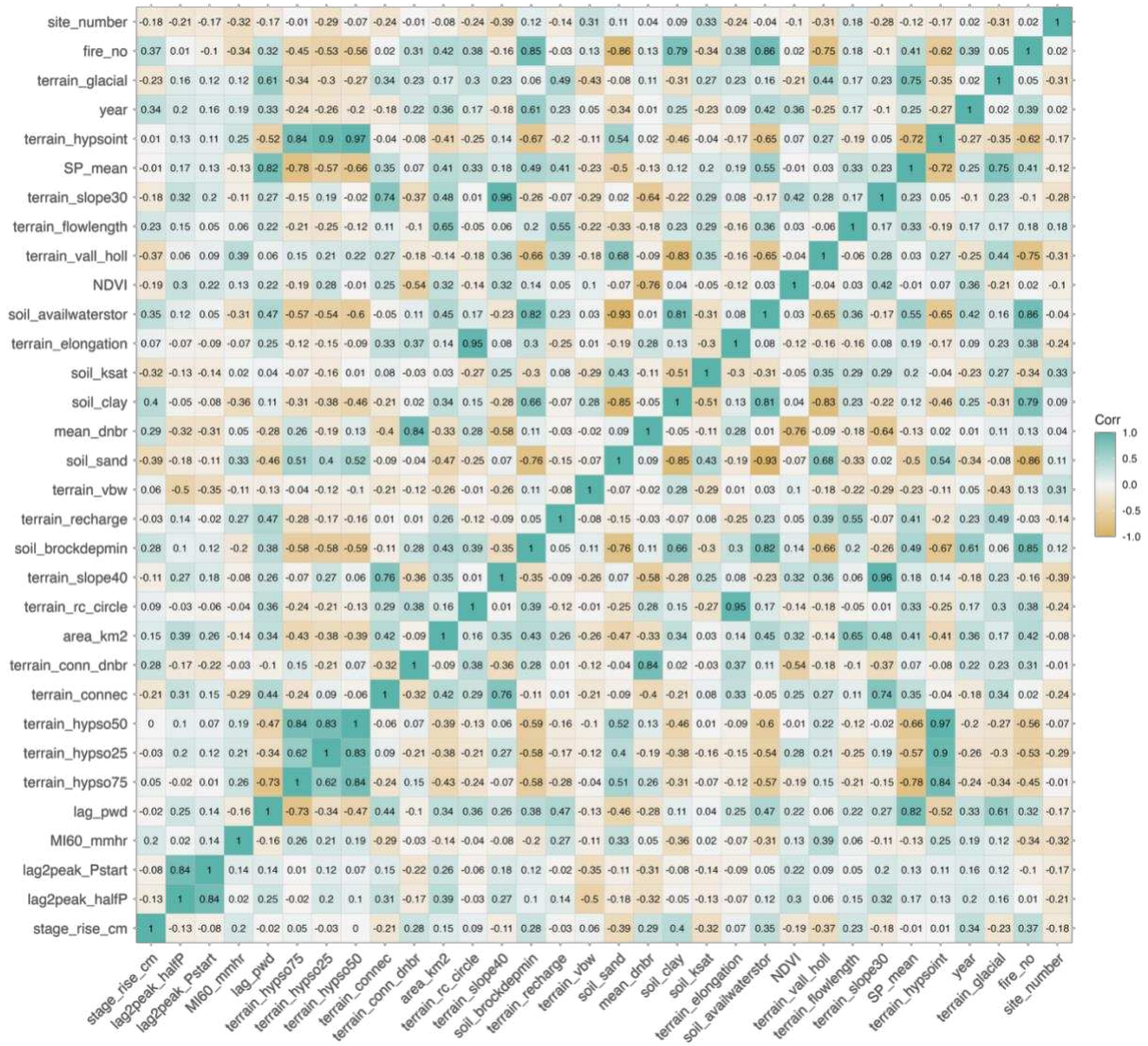

Figure B-2. Spearman correlations for all variables considered in the study.

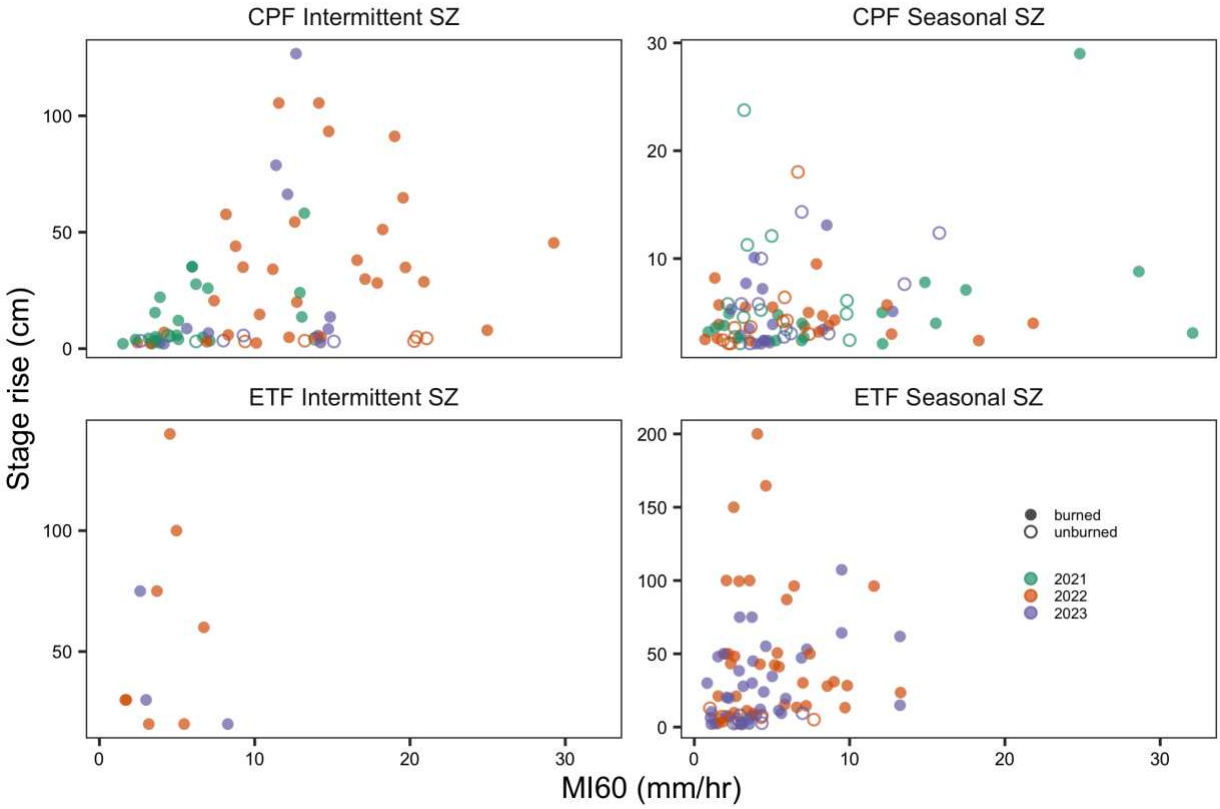

Figure B-3. Stage rise for each fire and snow zone colored by year. The open circles represent burned sites.

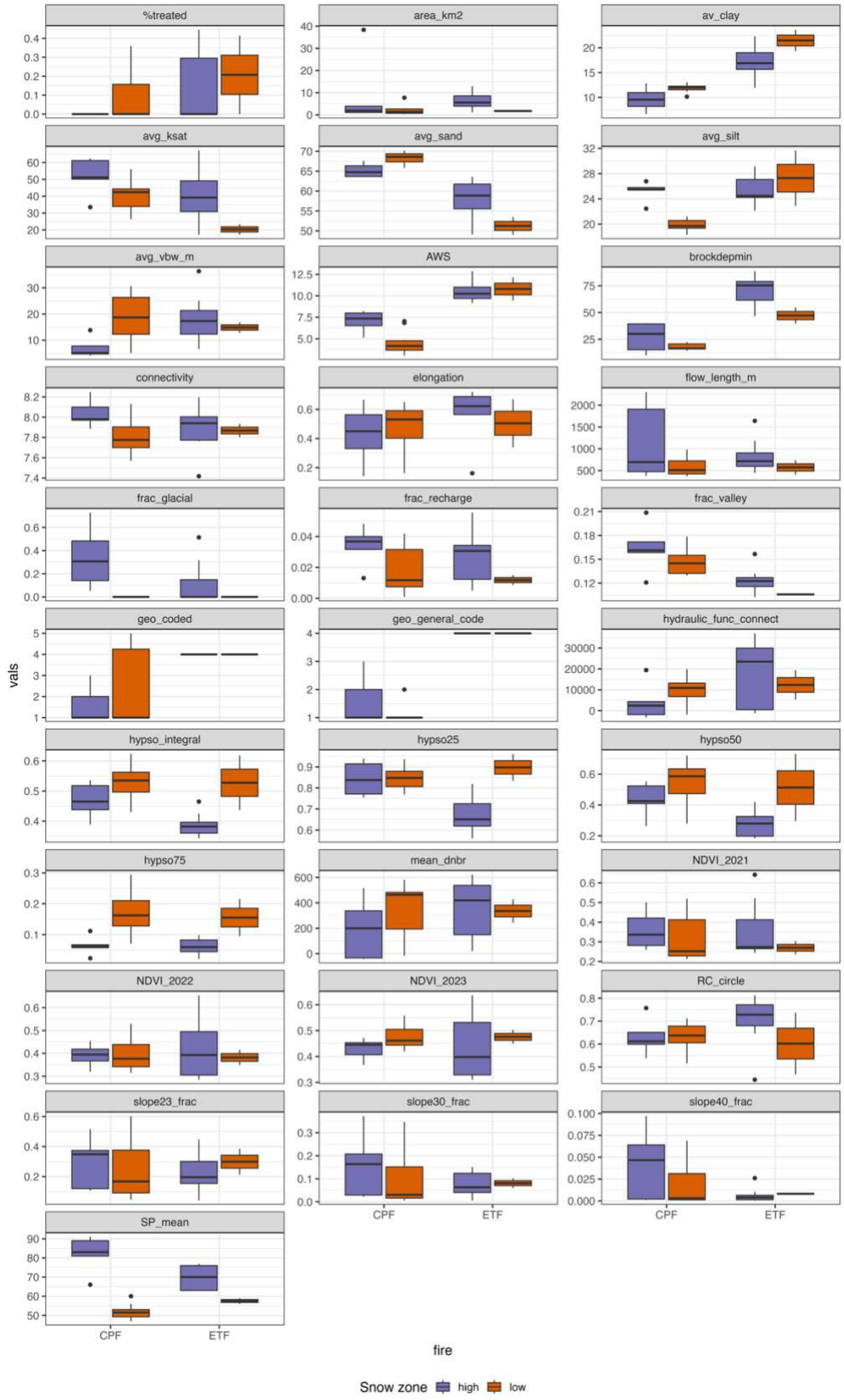

Figure B-4. Boxplots by fire and snow zone of variables considered in the study.

APPENDIX C: Chapter 4 supplementary information

Table C-1. The remaining parameters used in Ages that were not discussed in Chapter 4.

Parameter	Unit	Range	Description
kdiff_layer	-	1.0-50	Coefficient used to calculate MPS water flux between soil layers
BetaW	-	0.1-15	Coefficient used to calculated transpiration in soil layers
soilMaxInfSnow	mm/day	0.0-100	Coefficient used to calculate maximum infiltration for snow covered areas
soilMaxInfWinter	mm/day	0.0-100	Coefficient used to calculate maximum infiltration in winter
g_factor	-	0.0-5.0	Soil heat factor for snow melt calculation
flowRouteTA	-	1.0-1000	Flood routing coefficient
gwCapRise	-	0.1-0.6	Capillary rise coefficient
gwRG1Fact	-	1.0-20	Fast groundwater outflow coefficient
gwRG2Fact	-	1.0-80	Slower groundwater outflow coefficient
gwRG1RG2dist	-	0.1-0.9	RG1 to RG2 gradient coefficient
geoMaxPerc	mm/day	0.01-10	Maximum percolation rate to groundwater
soilDiffMPSLPS	-	0.1-10	Coefficient for the diffusion of LPS storage in relation to MPS
snow_trans	ºC	0.0-2.5	Coefficient for temperature-dependent fractions precipitation phases
r_factor	-	0.0-5.0	Rain factor for snow melt calculation
snowCritDens	kg/m3	0.2-0.8	Snowpack density beyond which free water is released
ACAdaptation	-	0.9-1.2	Snow temperature lag factor
ccf_factor	-	0.0001-0.1	Cold content snow factor for snow
DCAdaptation	m3/m3	0.8-1.2	Dead capacity
FCAdaptation	m3/m3	0.8-1.3	Field capacity
kf_fact	ha MJ mm	0.01-2.0	Soil erodibility factor
soilPolRed	-	0.1-50	Polynomial evapotranspiration reduction factor

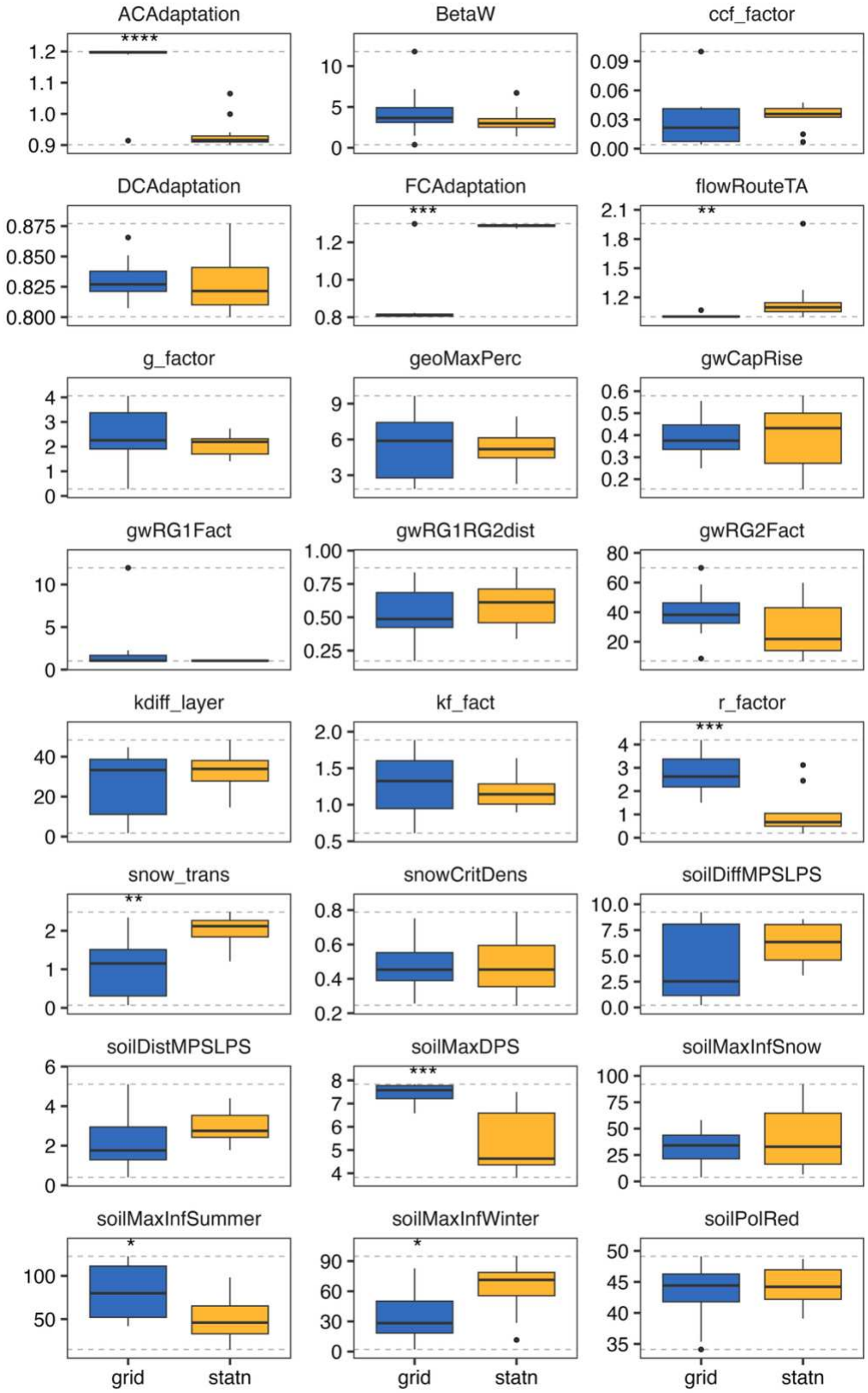

Figure C-1. Parameter variability for gridded (grid) and station (statn) model runs that were not discussed in the manuscript. Parameters are defined in Table C-1. The asterisks represent

significant difference determined by the Wilcoxon-rank sum test (**** indicates p-value < 0.0001; *** indicates p-value < 0.001; ** indicates p-value < 0.01; * indicates p-value < 0.05). The dashed grey lines represent the lower and upper limits used for each parameter during calibration.

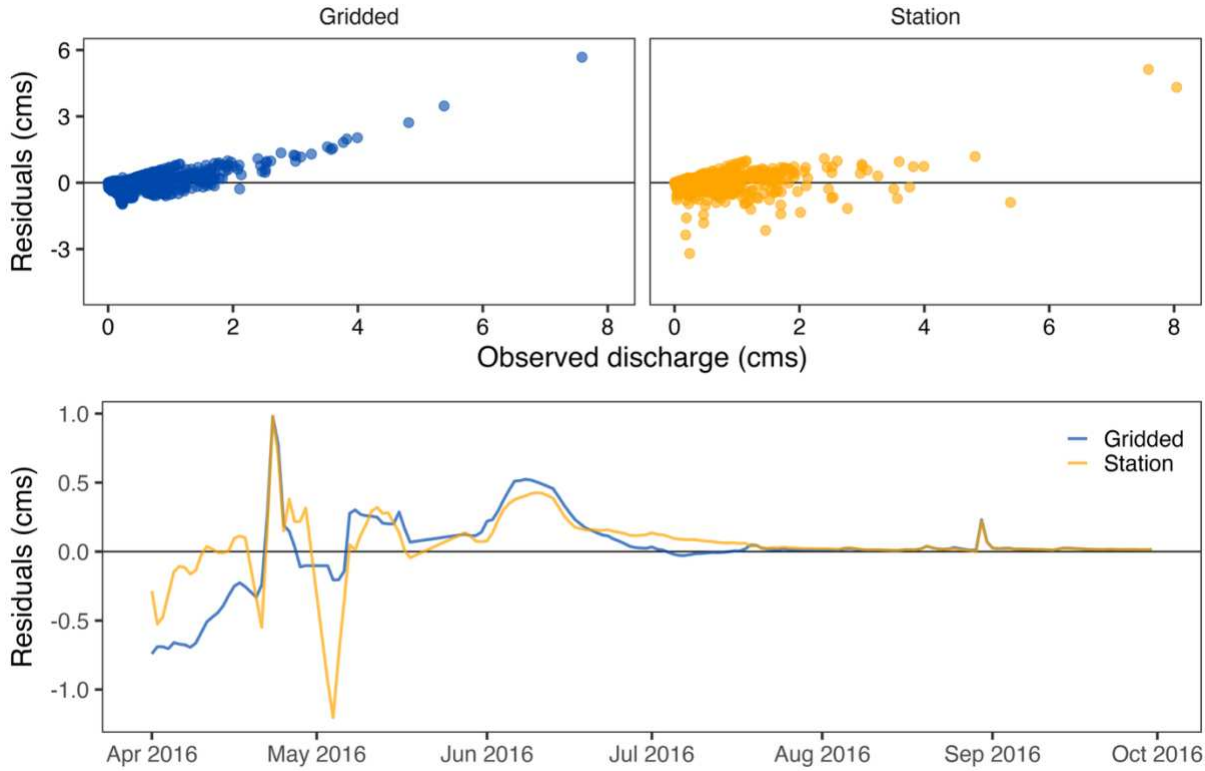

Figure C-2. Daily streamflow residuals and 2016 daily hydrograph for gridded and station (no years removed) model runs.

Table C-2. Comparisons between parameters from unburned (U) and burned (B) areas using gridded (G) and station (S) climate inputs means using a two-way ANOVA and Tukey HSD post-hoc analysis. U:G is unburned:gridded, B:G is burned:gridded, B:S is burned:station, U:S is unburned:station.

Parameter	Comparison	Estimated difference in means	95% CI (Lower, Upper)	p-value
a_snow	U:G vs. B:G	-1.237	[-2.624, 0.150]	0.095
	B:S vs. B:G	0.158	[-1.229, 1.545]	0.99
	U:S vs. B:G	7.022	[5.635, 8.409]	<0.001
	B:S vs. U:G	1.395	[0.008, 2.782]	0.048
	U:S vs. U:G	8.259	[6.873, 9.646]	<0.001
	U:S vs. B:S	6.865	[5.478, 8.251]	<0.001
a_rain	U:G vs. B:G	7.644	[6.641, 8.647]	<0.001
	B:S vs. B:G	1.884	[0.880, 2.887]	<0.001
	U:S - B:G	5.771	[4.768, 6.775]	<0.001

Parameter	Comparison	Estimated difference in means	95% CI (Lower, Upper)	p-value
	B:S - U:G	-5.760	[-6.764, -4.757]	<0.001
	U:S vs. U:G	-1.873	[-2.876, -0.869]	<0.001
	U:S vs. B:S	3.888	[2.884, 4.891]	<0.001
soilMaxInfSummer	U:G vs. B:G	40.533	[-0.863, 81.929]	0.057
	B:S vs. B:G	-2.844	[-44.240, 38.552]	0.998
	U:S vs. B:G	1.207	[-40.189, 42.603]	1
	B:S vs. U:G	-43.376	[-84.772, -1.981]	0.037
	U:S vs. U:G	-39.326	[-80.722, 2.070]	0.068
	U:S vs. B:S	4.05	[-37.345, 45.446]	0.994
soilMaxDPS	U:G vs. B:G	-2.751	[-4.658, -0.844]	0.002
	B:S vs. B:G	-1.304	[-3.211, 0.603]	0.273
	U:S vs. B:G	-5.009	[-6.916, -3.102]	<0.001
	B:S vs. U:G	1.447	[-0.460, 3.354]	0.193
	U:S vs. U:G	-2.258	[-4.165, -0.351]	0.015
	U:S vs. B:S	-3.705	[-5.612, -1.798]	<0.001
lagSurfaceRunoff	U:G vs. B:G	-0.001	[-0.095, 0.094]	1
	B:S vs. B:G	-0.131	[-0.225, -0.037]	0.003
	U:S vs. B:G	-0.216	[-0.311, -0.122]	<0.001
	B:S vs. U:G	-0.130	[-0.225, -0.036]	0.003
	U:S vs. U:G	-0.216	[-0.310, -0.121]	<0.001
	U:S vs. B:S	-0.085	[-0.180, 0.009]	0.088
soilDistMPSLPS	U:G vs. B:G	-2.175	[-4.039, -0.312]	0.017
	B:S vs. B:G	-3.706	[-5.569, -1.842]	<0.001
	U:S vs. B:G	0.373	[-1.491, 2.236]	0.95
	B:S vs. U:G	-1.531	[-3.394, 0.333]	0.14
	U:S vs. U:G	2.548	[0.684, 4.411]	0.004
	U:S vs. B:S	4.078	[2.215, 5.942]	<0.001

Table C-3. Comparisons between mean annual cumulative flow for four flow types (RD1 [surface runoff], RD2 [interflow], RG1 [fast groundwater], RG2 [slow groundwater]) using gridded (G) and station (S) climate data using a two-way ANOVA and Tukey HSD post-hoc analysis.

Comparison	Estimated difference in means	95% CI (Lower, Upper)	p-value
S:RD1 vs. G:RD1	-5.18	[-40.67, 30.30]	0.9998
G:RD2 vs. G:RD1	-7.73	[-43.22, 27.76]	0.9973
S:RD2 vs. G:RD1	37.57	[2.08, 73.06]	0.0304
G:RG1 vs. G:RD1	109.12	[73.63, 144.60]	<0.001
S:RG1 vs. G:RD1	47	[11.52, 82.49]	0.0023

Comparison	Estimated difference in means	95% CI (Lower, Upper)	p-value
G:RG2 vs. G:RD1	-9.05	[-44.53, 26.44]	0.9928
S:RG2 vs. G:RD1	-9.03	[-44.52, 26.46]	0.9929
G:RD2 vs. S:RD1	-2.55	[-38.03, 32.94]	1
S:RD2 vs. S:RD1	42.75	[7.27, 78.24]	0.0078
G:RG1 vs. S:RD1	114.3	[78.81, 149.79]	<0.001
S:RG1 vs. S:RD1	52.19	[16.70, 87.67]	<0.001
G:RG2 vs. S:RD1	-3.86	[-39.35, 31.63]	1
S:RG2 vs. S:RD1	-3.84	[-39.33, 31.64]	1
S:RD2 vs. G:RD2	45.3	[9.81, 80.79]	0.0038
G:RG1 vs. G:RD2	116.85	[81.36, 152.33]	<0.001
S:RG1 vs. G:RD2	54.73	[19.24, 90.22]	<0.001
G:RG2 vs. G:RD2	-1.32	[-36.80, 34.17]	1
S:RG2 vs. G:RD2	-1.3	[-36.79, 34.19]	1
G:RG1 vs. S:RD2	71.55	[36.06, 107.04]	<0.001
S:RG1 vs. S:RD2	9.43	[-26.05, 44.92]	0.9908
G:RG2 vs. S:RD2	-46.61	[-82.10, -11.13]	0.0026
S:RG2 vs. S:RD2	-46.6	[-82.08, -11.11]	0.0026
S:RG1 vs. G:RG1	-62.12	[-97.60, -26.63]	<0.001
G:RG2 vs. G:RG1	-118.16	[-153.65, -82.68]	<0.001
S:RG2 vs. G:RG1	-118.15	[-153.63, -82.66]	<0.001
G:RG2 vs. S:RG1	-56.05	[-91.54, -20.56]	<0.001
S:RG2 vs. S:RG1	-56.03	[-91.52, -20.54]	<0.001
S:RG2 vs. G:RG2	0.02	[-35.47, 35.51]	1