

DISSERTATION

CAPTURING THE VARIABILITY OF RAINFALL INTENSITY AND ITS IMPACTS ON
MOUNTAIN HYDROLOGY

Submitted by

Phoebe White

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2025

Doctoral Committee:

Advisor: Peter A. Nelson

Frances V. Davenport

Ryan R. Morrison

Russ S. Schumacher

Copyright by Phoebe White 2025

All Rights Reserved

ABSTRACT

CAPTURING THE VARIABILITY OF RAINFALL INTENSITY AND ITS IMPACTS ON MOUNTAIN HYDROLOGY

Storms in mountainous regions can develop quickly and cause significant flooding. Understanding how brief high intensity precipitation is influenced by terrain is important for evaluating the risk of flash floods and other natural hazards. The lack of precipitation gages in mountainous remote areas inhibits detailed monitoring of these hazardous events. As wildfires become more common in areas of complex terrain, increasing the likelihood of flash floods and debris flows, it is more important than ever to understand the potential for extreme precipitation. Precipitation estimates from remotely sensed data, such as radar and satellite, have improved in recent decades. Convection-permitting models also produce accurate precipitation estimates, that have outperformed interpolated gage datasets. In this dissertation, I examine the performance of these datasets at scales relevant to hazardous events in Colorado. Additionally, I investigate how considering the variability of natural rainfall observed in the mountains of Colorado influences runoff and erosion processes.

Radar estimates of precipitation can fill the gaps in areas where gages are sparse, but the signal can be blocked by mountains, depending on where the storm is relative to the radar site. Because the error of radar estimates of precipitation can change based on where the storm is located in relation to the surrounding terrain and location of the radar, the reliability of these precipitation estimates is variable, adding to the difficulty of monitoring storms in mountains. To address this uncertainty, I developed a novel method of identifying where and when the radar estimates of precipitation are reliable, based on attributes of the region, rainfall, and storm events. The model can assist in deciding when to trust radar estimates of precipitation and in determining where more gages or radar sites are necessary. The error model uses the Multi-Radar Multi-Sensor (MRMS)

product which incorporates radar, quantitative precipitation forecasts, and gage data at a high spatiotemporal resolution for the United States and southern Canada. For several time series samples of MRMS 15-minute intensity, various features related to the physical characteristics influencing MRMS performance are calculated from the topography, surrounding storms, and rainfall observed at the gage location. A gradient boosting regressor is trained and was used to predict a range of error throughout the mountains of Colorado during warm months. Mapping of this dataset by aggregating normalized RMSE over time reveals that areas further from radar sites in higher elevation terrain show consistently greater error. However, the model predicts larger performance variability in these regions compared to alternative error assessments.

Precipitation gages provide high temporal resolution data; however, terrain induces significant variability in precipitation across Colorado. As a result, interpolated datasets or frequency analyses based on simple linear regression of gage data may fail to capture critical extremes. Anomalous precipitation events and variation at subdaily time scales are likely omitted from gage-based datasets due to low station density. To explore this uncertainty, I use several remotely-sensed and model-derived hourly datasets and re-evaluate the influence of terrain on the magnitude of subdaily precipitation intensity throughout Colorado. Precipitation–elevation relationships differ among basins: the Missouri and Arkansas show decreasing precipitation with elevation—an effect stronger for hourly than daily totals—whereas the Colorado and Rio Grande exhibit increasing precipitation with elevation, with daily totals rising more steeply and significantly than hourly ones. Gage-based frequency studies, limited by sparse networks, miss the frequency of high intensity clusters along the Front Range and Pikes Peak shown by model data. Gage interpolation schemes might also fail to capture how particular terrain features, rather than elevation alone, affect precipitation development.

After evaluating the capability of various datasets to represent precipitation variability in Colorado’s mountains, I investigate how this variability impacts sediment mobility and runoff. Non-uniform rainfall profiles can result in significantly higher runoff and erosion rates compared to constant rainfall. In this study, I use a rainfall simulator capable of generating time varying inten-

sity profiles similar to natural rainfall observed by gages in the mountains of Colorado. I examine the effects of time-varying rainfall intensity on both mulched and bare soil on a steeply sloped flume. The effectiveness of mulch varies markedly between fluctuating rainfall intensities and a steady intensity. The wood mulch treatment failed to significantly reduce total runoff under time-varying rainfall, though it did under constant rainfall. Erosion rates were reduced with mulch for all rainfall events, despite the increasing intensity rain event causing significantly more erosion on bare soil. Accounting for the variability of rainfall directly impacts management solutions for post-wildfire recovery.

Rainfall variability in mountainous regions can be highly dynamic across both space and time, affecting surface processes in small plots and at larger scales. Accurately representing these fine-scale rainfall patterns remains challenging. Remotely sensed data and model outputs can complement or substitute gage data, especially when gages are sparse, enhancing the accuracy of rainfall estimates.

ACKNOWLEDGEMENTS

First, I'd like to thank my advisor, Peter Nelson. Throughout my graduate studies, he encouraged me to explore new ideas, trusted me to learn from failure, and ultimately helped me succeed—giving me invaluable confidence. He challenged me to grasp concepts at their most fundamental level, and his candid guidance on navigating graduate school and research projects has been a constant source of calm.

I'd also like to thank my committee members. I was fortunate to take at least one course with each of you, and every class exceeded my expectations. Each course left me with practical skills to bring to my research. They also exposed me to new ideas and levels of understanding that kept school fun and left me excited to try new things in research. I am grateful for the work you all have put into your courses, for breaking down complicated and confusing things and making them seem approachable, and for exposing me to new perspectives and fundamental ideas in an inclusive environment. All while having the patience to chat one on one about a topic or idea.

My former and present lab mates have kept me entertained, filled me with the motivation I need, and always made me feel like a part of a community. Thank you to Lindsey who got the rainfall simulator going and made me laugh. Danny, thank you for our passing conversations at the ERC. Your thoughtful advice and encouragement always lightened my much experiment frustrations. Thank you to Alex, I appreciate the time we've spent on the trail, your curiosity, and willingness to lend an ear. And thank you to Rahel for lending me a soil moisture probe and graciously teaching the basics of point measurements.

I've been very lucky to have most of my family and best friend in Fort Collins. They made me who I am and nurtured my curiosity. At every odd and unexpected turn I've chosen, whether it be West Point or pursuing a PhD, they've been in my corner believing in me and ready to catch me if I fall. Words are not enough to express my gratitude.

Finally, I'd like to thank the members of my home. Lou who taught me to be less fearful and introduced me to my neighbors in his relentless quest for more affection and dry cat food. Walden

and Julien who never take a walk in the woods for granted and always bound and howl with joy. And Matt, who makes hard stuff look easy. Thank you for being a constant sounding board and always supporting me along the way. Our shared love of learning has been an inspiration and guiding path.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	v
LIST OF TABLES	ix
LIST OF FIGURES	x
Chapter 1 Introduction	1
Chapter 2 Evaluation of sub-hourly MRMS quantitative precipitation estimates in mountainous terrain using machine learning	3
2.1 Summary	3
2.2 Introduction	4
2.3 Methods	7
2.3.1 Study area and data extents	7
2.3.2 Data sources and processing	9
2.3.3 Modeling	15
2.4 Results	17
2.4.1 MRMS performance trends	17
2.4.2 Features related to MRMS performance	19
2.4.3 Model performance	22
2.5 Discussion	23
2.5.1 When and where are sub-hourly MRMS precipitation estimates trustworthy in the mountains of Colorado?	23
2.5.2 What are the circumstances that lead to low error?	27
2.5.3 How well can error be predicted by the physical aspects of the region, rainfall patterns, and storm characteristics?	28
2.6 Conclusion	29
Chapter 3 Regional and temporal variability of orographic influence on subdaily precipitation intensity in Colorado	31
3.1 Summary	31
3.2 Introduction	32
3.3 Data	34
3.4 Methods	37
3.5 Results	39
3.5.1 Precipitation differences across datasets	39
3.5.2 Seasonal variability in precipitation	41
3.5.3 Comparison to precipitation frequency studies	42
3.5.4 Elevation-precipitation regression	44
3.6 Discussion	50
3.6.1 Regional precipitation intensity trends	50
3.6.2 Time dependence of orographic effect	54

3.7	Conclusion	55
Chapter 4	Effectiveness of mulching under varying rainfall intensity on steep slopes . .	57
4.1	Summary	57
4.2	Introduction	58
4.3	Methods	60
4.3.1	Rainfall simulator setup and soil	60
4.3.2	Mulch characteristics	63
4.3.3	Simulated rainfall	64
4.3.4	Experimental procedure	68
4.3.5	Statistical approach	70
4.4	Results	71
4.4.1	Sediment yield	71
4.4.2	Runoff response	73
4.5	Discussion	76
4.5.1	Effectiveness of mulch	76
4.5.2	Influence of rainfall variability	81
4.5.3	Implications for management	83
4.6	Conclusion	83
Chapter 5	Conclusion	86
Appendix A	Supporting information for Chapter 2	114

LIST OF TABLES

2.1	Feature descriptions	12
2.3	Cross fold validation and test results.	22
3.1	Comparison of dataset regression statistics.	47
3.2	Comparison of seasonal regression statistics.	47
3.3	Comparison of CONUS404 and precipitation frequency study regression statistics. . .	50
4.1	Soil texture and organic content.	63
4.2	Rainfall simulator performance.	65
4.3	Total erosion rate per unit area for grouped experiments.	74
4.4	Total erosion rate per unit area for individual runs.	74
4.5	Runoff ratio for grouped experiments.	77
4.6	Runoff ratio for individual runs.	77
4.7	Key parameters of wood mulch experiments used in this synthesis.	80
4.8	Rain–profile experiments and their key parameters.	82
A.1	Brief model descriptions for single output experiment	125
A.2	Model comparison test results	126
A.3	Hyperparameters tuned and values used	126
A.4	Test results for RMSE data selection experiments	126
A.5	Test results for nRMSE data selection experiments	126

LIST OF FIGURES

2.1	Terrain and instruments within the study area.	8
2.2	Kernel density estimations of the distributions of RMSE.	18
2.3	The median normalized RMSE for each coordinate.	19
2.4	Permutation feature importance.	20
2.5	Comparison of feature value distributions.	21
2.6	The target values as a function of the median predicted RMSE.	23
2.7	Comparison with RQI.	24
2.8	Influence of temporal inconsistency.	26
3.1	Map of study area.	38
3.2	Map of normalized median of annual maximum 1-hour and 24-hour precipitation accumulations for JJA covering 2016-2022 for three gridded datasets.	40
3.3	Map of median of annual maximum 1-hr and 24-hr precipitation accumulations for JJA for three gridded datasets for 2016-2022.	41
3.4	Normalized median of annual maximum 1-hr and 24-hr precipitation accumulations for CONUS404 for all seasons	42
3.5	Map of CONUS404, REPS, and Atlas 14 precipitation.	43
3.6	CONUS404 90th percentile of annual hourly and daily maxima and 10-year average recurrence interval for REPS and Atlas 14.	44
3.7	Linear regression for three gridded datasets.	46
3.8	Linear regression for all seasons.	48
3.9	Linear regression for CONUS404, REPS, and Atlas 14.	49
3.10	Absolute and mean relative error for CONUS404 and PRISM mean seasonal accumulation.	52
3.11	Comparison with lightning dataset.	53
4.1	Setup of rainfall simulator and key components of flume.	61
4.2	Rainfall simulator in indoor laboratory.	62
4.3	Example mulch coverage mask.	64
4.4	Drop size distributions	67
4.5	Example raindrop sample.	68
4.6	Rainfall profiles compared to gages.	69
4.7	Time series of erosion concentration for all experiments.	71
4.8	Example of rilling.	72
4.9	The distribution of total erosion per unit area for grouped (sequential) and individual runs.	73
4.10	Time series of runoff rate for all experiments.	75
4.11	The distribution of runoff ratio for grouped (sequential) and individual runs.	76
4.12	Runoff ratio and erosion rate from studies assessing the influence of wood mulch cover.	79
4.13	Runoff ratio and erosion rate from studies assessing the influence of variable rainfall intensity.	82

A.1	Distribution of RMSE for radar-only 15-minute intensity and multi-sensor corrected 15-minute intensity.	115
A.2	Spread of model RMSE predictions for the best performing model, a random forest regressor.	116
A.3	Comparison of feature value distributions.	116
A.4	Difference in error (computed directly from the gage) and mean RQI for each 8-hour sample across all years. For each year, error bars are separated by the latitude of the gage location.	117
A.5	Difference in error (computed directly from the gage) and mean RQI for each 8-hour sample across all years. For each year, error bars are separated by the longitude of the gage location.	117
A.6	Median normalized RMSE for each coordinate for the model trained with data prior to MRMS v12.0 (2018-2021).	118
A.7	Target values as a function of the median predicted RMSE for the model trained with data prior to MRMS v12.0 (2018-2021).	118
A.8	Kernel density estimations of the distributions of normalized RMSE for the model trained with data prior to MRMS v12.0 (2018-2021).	119
A.9	Difference in error (computed directly from the gage) and mean RQI for each 8-hour sample for all months used in the study. For each month, error bars are separated by the elevation of the gage location.	119
A.10	Difference in error (computed directly from the gage) and mean RQI for each 8-hour sample across all years. For each year, error bars are separated by the longitude of the gage location. Data only includes years 2021-2023, after MRMS version 12.0	119
A.11	The median normalized RMSE predictions, separated by month, from the original model for each coordinate, predicted by the $\alpha = 0.50$ model.	120
A.12	For the model trained with data prior to MRMS v12.0 and trained to predict mean error, kernel density estimations of the distributions of mean error are shown for all target error, the test target error, the test predicted error, and the statewide predicted error.	120
A.13	For the model trained with data prior to MRMS v12.0 and trained to predict mean error, the target values as a function of the median predicted mean error is shown for the test dataset.	121
A.14	Error maps For the model trained with data prior to MRMS v12.0. and trained to predict mean error.	121
A.15	Comparison of feature value distributions for the model trained with data prior to MRMS v12.0.	122
A.16	Comparison of feature value distributions for the model trained with data prior to MRMS v12.0 and trained to predict mean error.	122
A.17	Kernel density estimations of the distributions of nRMSE for the model trained with data prior to MRMS v12.0 (2018-2021) and trained to predict normalized RMSE (nRMSE).	123
A.18	Target values as a function of the median predicted nRMSE for the model trained with data prior to MRMS v12.0.	123
A.19	Maps of error For the model trained with data prior to MRMS v12.0 (2018-2021) and trained to predict normalized RMSE (nRMSE).	124

A.20	Permutation feature importance for the model trained with data prior to MRMS v12.0 (2018-2021) and trained to predict normalized RMSE (nRMSE).	124
A.21	Feature value distributions for the model trained with data prior to MRMS v12.0 (2018-2021) and trained to predict normalized RMSE (nRMSE).	125

Chapter 1

Introduction

Orographic uplift, moisture blocking, and solar exposure create distinct patterns in precipitation climatology in the mountains, yet their interaction with changing atmospheric conditions during individual storms makes event-scale precipitation highly variable (Roe, 2005). Despite the recognized spatiotemporal variability of mountain precipitation and its importance for hazards like flooding and landslides, accurately measuring precipitation in these areas and predicting its geomorphic and hydrologic impacts remains challenging (Lundquist et al., 2019; Partridge et al., 2024; Collar et al., 2025). In the absence of models that incorporate dynamic precipitation, disaster response often relies on static extreme precipitation thresholds for decision making that can result in high uncertainty in risk associated with flooding and geomorphic hazards (Mirus et al., 2024). Distributed hydrologic modeling of complex mountainous environments has improved (Ebel et al., 2023). Simultaneously, improvements in weather forecasting, climate modeling, observations, and remote sensing, along with progress in data-driven methods such as machine learning, offer numerous opportunities to apply higher quality precipitation data into these models. With high elevation terrain becoming more susceptible to wildfire, there is an increase in runoff and erosion potential from short duration intense rainfall, which makes accurate estimates of extreme precipitation at this scale more relevant (Higuera et al., 2021; Touma et al., 2022).

In Chapter 2, I address the need for additional high resolution precipitation data in gage-scarce regions by developing an error model for a radar-based precipitation estimate. Given the advancements of Multi-Radar Multi-Sensor (MRMS), I explore the performance of the 2-minute rate product in the mountains of Colorado. A machine learning model is developed to estimate the spatiotemporally varying error of MRMS and can be used to explore the physical conditions related to low error.

In Chapter 3, I explore the spatiotemporal variability in subdaily precipitation extremes that is missed by gage interpolation. Subdaily annual maxima from radar and model datasets are com-

pared to gage interpolations. Model-based precipitation appears to capture more terrain-driven spatiotemporal variability in precipitation intensity than gage-based precipitation.

In Chapter 4, I apply the natural variability of rainfall observed in the mountains of Colorado on a laboratory flume using a rainfall simulator. The experiments confirm the significance of rainfall variability in the runoff and erosion response. Experiments were conducted on bare soil and with a mulch cover, a commonly used post-wildfire recovery treatment. The effectiveness of the treatment depends on the variability of rainfall applied, showing that incorporating variable rainfall can alter management decisions.

Finally, in Chapter 5, I present a brief summary of key findings in Chapters 2 through 4.

Chapter 2

Evaluation of sub-hourly MRMS quantitative precipitation estimates in mountainous terrain using machine learning¹

2.1 Summary

The Multi-Radar Multi-Sensor (MRMS) product incorporates radar, quantitative precipitation forecasts, and gage data at a high spatiotemporal resolution for the United States and southern Canada. MRMS is subject to various sources of measurement error, especially in complex terrain. The goal of this study is to provide a framework for understanding the uncertainty of MRMS in mountainous areas with limited observations. We evaluate 8-hour time series samples of MRMS 15-minute intensity through a comparison to 204 gages located in the mountains of Colorado. This analysis shows that the MRMS surface precipitation rate product tends to overestimate rainfall with a median normalized root mean squared error (RMSE) of 42% of the maximum MRMS 15-minute intensity. For each time series sample, various features related to the physical characteristics influencing MRMS performance are calculated from the topography, surrounding storms, and rainfall observed at the gage location. A gradient-boosting regressor is trained on these features and is optimized with quantile loss, using the RMSE as a target, to model nonlinear patterns in the features that relate to a range of error. This model was used to predict a range of error throughout the mountains of Colorado during warm months, spanning 6 years, resulting in a spatiotemporally varying error model of MRMS for sub-hourly precipitation rates. Mapping of this dataset by aggregating normalized RMSE over time reveals that areas further from radar sites in higher elevation

¹A version of this chapter has been published in Water Resources Research: White, P., & Nelson, P. A. (2024). Evaluation of Sub-Hourly MRMS Quantitative Precipitation Estimates in Mountainous Terrain Using Machine Learning. *Water Resources Research*, 60, e2024WR037437. <https://doi.org/10.1029/2024WR037437>

terrain show consistently greater error. However, the model predicts larger performance variability in these regions compared to alternative error assessments.

2.2 Introduction

In mountainous regions, in-situ precipitation observations from sparse gage networks are inadequate for accurate hydrologic modeling and monitoring of hazardous events (Lundquist et al., 2019). Post wildfire debris flows and flash floods are often initialized by brief high-intensity rainfall (Cannon et al., 2008; Doswell et al., 1996) and therefore require high-resolution precipitation observations for accurate modeling and validation of forecasts (Moody et al., 2013; Sokol et al., 2021). Increasing atmospheric moisture tied to global warming has been shown to amplify the intensity of short-duration rainfall (Fowler et al., 2021). With wildfires becoming more widespread in the western United States (Higuera et al., 2021), these events are more likely to be coupled with intense rainfall, increasing the likelihood of flooding and debris flows (Touma et al., 2022). Several studies have shown that the magnitudes of slow-rise floods are also sensitive to spatiotemporal variability of precipitation (Syed et al., 2003; Nicótina et al., 2008; Zhu et al., 2018). Accurate assessment of hydrologic risk requires precipitation observations at the scale of the event, which may be increasingly seen at the sub-hourly and sub-basin scale.

Gages provide precipitation data at high temporal resolutions for point locations, but mountains can cause significant spatial variation in precipitation (Smith, 2019). Many interpolation schemes exist to create spatially distributed precipitation estimates from gages; for example, the Parameter-Elevation Regressions on Independent Slopes Model (PRISM) is a continuous dataset for the United States developed from interpolated gage data (Daly et al., 2008). However, the accuracy of these interpolated gage datasets declines where gages are sparse (Lundquist et al., 2019). And, the gage itself is subject to various sources of measurement error, such as recording errors and under-catching due to advection of hydrometeors and interception by the tree canopy (Sevruk, 2005; Sieck et al., 2007). Henn et al. (2018) compared the annual accumulation from six gage-

based gridded datasets across the Western United States and found that relative error varied from 5 to 60% with the largest error located in regions of high-elevation terrain.

Quantitative precipitation estimates (QPEs) are estimations of precipitation at locations or regions where ground or direct measurements do not exist. Satellite-based QPEs have been improving, but the resolution of QPEs derived from radar reflectivity is often superior (Derin et al., 2016). Radar-based precipitation estimates are high resolution and spatially continuous but are subject to several sources of uncertainty, made worse by mountainous terrain. Radar-based QPE accuracy has improved with the adoption of dual-polarization schemes and improved algorithms to correct for beam blockage (Zhang et al., 2016). The Multi-Radar Multi-Sensor (MRMS) system provides a mosaic of radar for the contiguous United States, improves ground clutter with dual-polarization and better conversion algorithms, and assimilates additional sources of data (Zhang et al., 2016, 2014). The MRMS products are used in mountainous operational forecasting settings for flash flood (Gourley et al., 2017) and debris flow monitoring (Force, 2005) and to validate satellite QPEs (Kirstetter et al., 2012; Sun et al., 2021). Despite the usefulness of this product, the imperfect conversion from reflectivity to rain rate, ground clutter, and beam blockage all contribute to significant and complex uncertainty (Berne and Krajewski, 2013; Bytheway et al., 2020). MRMS is especially uncertain in the mountainous West where the minimum height observed by radar is significantly higher in the atmosphere (Maddox et al., 2002).

Many studies have evaluated the error associated with QPEs; however, there is a lack of comprehensive error modeling, specifically at sub-hourly resolutions in mountainous terrain. Several case studies have evaluated MRMS at hourly and daily timescales over large regions of complex terrain by comparing precipitation estimates to gages. Moazami and Najafi (2021) compared hourly radar-only MRMS QPEs in Canada to gage data and found that MRMS tends to overestimate precipitation in the southern plains and underestimate precipitation in western and eastern parts of Canada. Bytheway et al. (2019) evaluated hourly gage-corrected and elevation-adjusted MRMS QPEs in the mountains of California and observed that MRMS failed to capture precipitation caused by small scale flow patterns and interactions with terrain. Additionally, rainfall

originating closer to the surface was often too low to be picked up by radar. Case studies reveal weather processes and precipitation characteristics related to performance of the QPE, but the error ranges are not generalizable as they can vary by storm and location.

Stream gages have been used to evaluate the spatial variability of error associated with stream flow predictions from well-calibrated hydrologic models (Liao and Barros, 2023; Moreno et al., 2012). The QPE skill can be evaluated throughout the basin, rather than where precipitation gages exist, but this requires accurate flow records and many mountainous watersheds are ungaged. Other models incorporate the spatiotemporal variability in uncertainty by modeling particular components of QPE error. The Radar Quality Index (RQI) is a temporally varying estimate of the reliability of MRMS based on known sources of radar sampling error (Zhang et al., 2012). Probabilistic models estimate the distribution of uncertainty using dense gage networks as reference (Ciach et al., 2007; Villarini et al., 2014). Kirstetter et al. (2015) created a probabilistic QPE for the continental US using radar-only MRMS data by modeling the range of possible precipitation values given the reflectivity and assuming the distribution of possible values. There is still a need for an error model that makes limited assumptions on the structure of error, can be applied in regions where in-situ observations are lacking, and provides spatiotemporally varying estimates of error. Osborne et al. (2023) use a deep learning approach to improve the hourly MRMS QPE product in areas of complex terrain, finding improvements where the physically based approach struggles with poor radar observations.

Machine learning excels in identifying trends in complex data with nonlinear and interconnected dependencies for which we have weak or incomplete theory and limited direct observations (Appling et al., 2022). Here we seek to develop a model for estimating error of sub-hourly MRMS rainfall estimates in a mountainous environment. The MRMS 15-minute intensity, an interval used in many applications such as flash flood and debris flow monitoring, is compared to gage observations throughout the mountains of Colorado. A gradient-boosting regressor is trained on sub-samples of time series at each gage location with associated storm and topographic features to predict a target root mean squared error (RMSE) calculated from the gage. The model learns what

the expected range of RMSE should be, based on the feature values and target RMSE, and can then predict the error of MRMS where we have no gage records. A statewide dataset, including the features used in training, is developed to predict the theoretical RMSE (if a ground truth existed) for the MRMS-estimated 15-minute intensity. This dataset can be used to understand the range of expected error for sub-hourly intensity estimates for a specific time window and location. Feature importance is calculated to understand what features the model deems most useful in determining QPE error.

Through the gage comparison, modeling of error, and interpolation of error throughout Colorado, we seek to address a series of questions. Firstly, when and where are sub-hourly MRMS precipitation estimates trustworthy in the mountains of Colorado? This information can help end-users understand when to trust MRMS. Also, trends in error can help identify where additional observations through gap-filling radar or more gages might be a priority. Secondly, what are the controls on performance or the circumstances that lead to low error? Lastly, how well can error be predicted by the physical aspects of the region, rainfall patterns, and storm characteristics? By examining these factors, we aim to gain a deeper understanding of the predictability of error in MRMS precipitation estimates.

2.3 Methods

2.3.1 Study area and data extents

The study area is confined to the state of Colorado, west of 104° longitude. This region includes several distinct mountain ranges with varying elevations and aspects. The topographic complexity of this region contributes to a spatially diverse precipitation climatology (Mahoney et al., 2015). The relief of the mountains in Colorado is significant enough to cause disturbances in lower atmospheric flow, which can cause significant spatiotemporal variability in precipitation (Smith, 2019). The area was selected to test the model's ability to learn the influence of terrain blockage on QPE performance with interactions between various characteristics of storms unique to mountain pre-

cipitation. The terrain of the study area is shown in Figure 2.1 with the locations of gages and radar for reference.

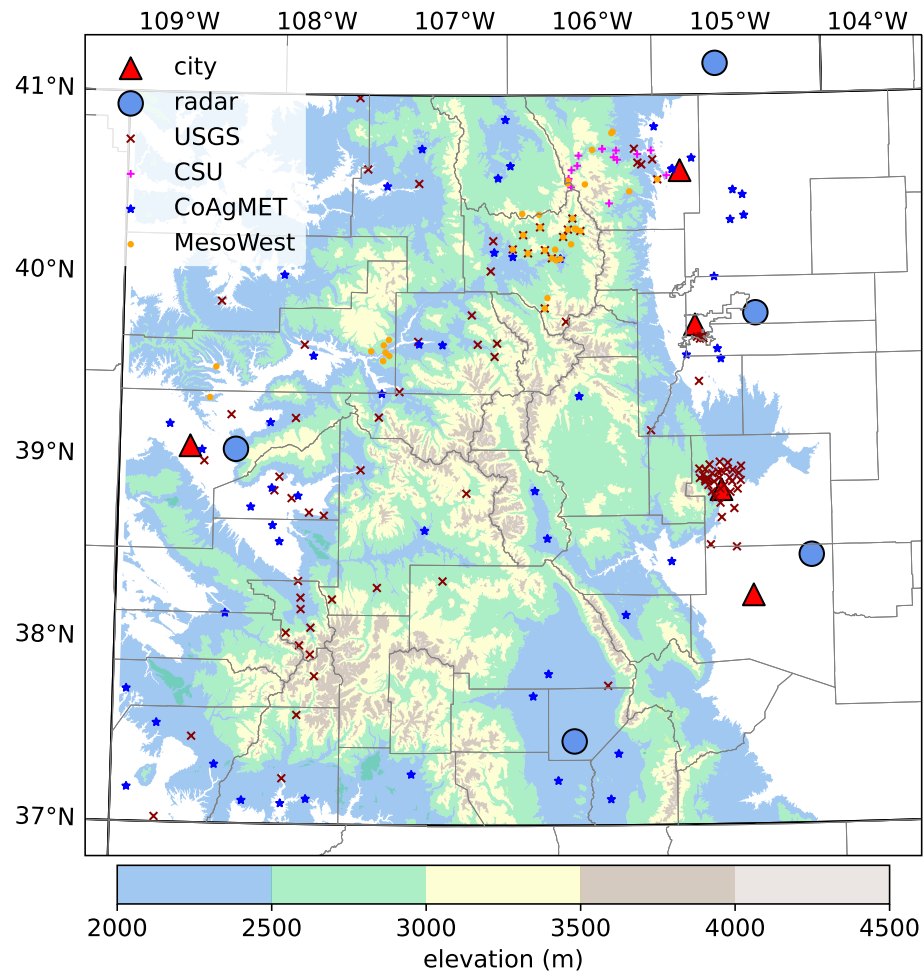


Figure 2.1: Terrain and instruments within the study area. Gages used in the study are labeled according to their source.

Both gages and radar-based QPEs are subject to several sources of error associated with freezing precipitation (Zhang et al., 2016), so the study period is constrained to warmer months (May through September) to limit the evaluation of MRMS with unreliable gage data. Based on availability of archived MRMS data, a total of 30 months were included in the modeling from years 2018 through 2023.

2.3.2 Data sources and processing

MRMS and gage data

MRMS includes a collection of products with a coverage of the United States and southern Canada. This study evaluates the surface precipitation rate product, a 2-minute precipitation rate estimate from a mosaic of the Weather Surveillance Radar-1988 Doppler (WSR-88D) radar network and out-of-network gap-filling radar with a spatial resolution of 1 km (Zhang et al., 2016, 2020). This product includes several quality controls and an evaporative correction from the Rapid Refresh (RAP) model data (NOAA, 2024). Missing MRMS data was filled with zero for calculations. All data were accessed through the Iowa Environmental Mesonet MRMS archived daily data (Mesonet, Accessed 2023). Each file's spatial coverage was resampled to the extents of Colorado using wgrib2 (NOAA/CPC, 2023). A 15-minute rate was computed from a rolling average of the 2-min rate estimate converted to accumulation.

We explored using the 1-hour multi-sensor QPE Pass 2 to improve the 2-minute precipitation rate accuracy by applying a correction factor. However, this did not significantly reduce the RMSE error, with a median normalized RMSE of 41%, similar to the radar-only rate of 42%. Since the multi-sensor product did not enhance sub-hourly estimates and is not available before 2021, we excluded it from our analysis (see Appendix A Figure A1 for error distributions and Text A1 for additional details).

Within the study period, several changes to the MRMS QPE have occurred, including algorithm updates and additional data inputs such as radar and gage sites. See <https://inside.nssl.noaa.gov/mrms/code-updates> for a detailed list of modifications (NSSL, 2024). Version 12.0, operational in February 2021, included substantial updates including a dual-polarization algorithm (Zhang et al., 2020), updates to the radar quality index and additional data assimilation for gap filling purposes (Martinaitis et al., 2020), and improvements in gage quality control (Martinaitis et al., 2021). With this update, Colorado gained a radar site in Alamosa. The influence of the MRMS data inconsistency on the training of the model and its predictions is addressed in the Discussion section.

The Radar Quality Index (RQI) was used to train the model and as a baseline to compare the performance of model estimated error. This product has a temporal resolution of 2 minutes and provides an estimate of uncertainty, represented by a floating-point number ranging from 0 to 1, that is related to terrain blockage, higher beam heights, and the beam position with respect to the freezing level (Zhang et al., 2012). An RQI of 1.0 suggests limited uncertainty associated with these problems. RQI is set to 0.0 when terrain blockage is greater than 50%.

The gage data was sourced from instruments monitored by Colorado State University (White and Nelson, 2024), the U.S. Geological Survey (USGS) (USGS, Accessed 2023; Rengers et al., 2023), CoAgMET (Colorado State University, Accessed 2023), and other networks aggregated by MesoWest (Mesowest, Accessed 2023). The locations of these stations are shown in Figure 2.1. The set of gages includes various models of tipping buckets and one disdrometer. Several gages record at regular time intervals (ranging from 1 minute to 5 minutes) and some gages record tips with varying record lengths. A total of 204 gages were used to develop the dataset. Gages were quality controlled by comparing the magnitude of RMSE across the study area. One gage, located in central Fort Collins, was removed from the dataset because it had significantly higher error, with a mean RMSE of 42.8 mm/hr compared to the overall dataset mean of 3.43 mm/hr.

Feature extraction and labeling

The labeled dataset is comprised of 37,215 samples, consisting of an 8-hour non-overlapping window of 15-minute intensity. Each sample has 28 features evaluated within the 8 hours and is labeled with a target RMSE calculated from the gage and MRMS intensity with the gage values assumed to be the ground truth. Features are input variables, associated with the error of MRMS, used by the model to make predictions. The sampling start times include 0800, 1600, and 0000 UTC. The purpose of the 8-hour window is to focus on periods of time impacted differently by the diurnal cycle. Sampling different parts of the diurnal cycle allows us to identify how well particular weather patterns impacted by the cycle, such as afternoon thunderstorms, are observed by MRMS. In the labeled dataset, features and error are evaluated in relation to the MRMS grid location closest to the gage. Only 8-hour samples with an MRMS total accumulation above 1 mm were

included to prevent sampling below the minimum accumulation that can be observed by a gage. The purpose of the model is to predict the error threshold of a non-zero time series of MRMS 15-minute intensity; the model cannot determine if the time series is a false negative because we did not include zero-value accumulations. The removal of precipitation below 1 mm effectively removed missing MRMS data, only 0.067% of the samples in the labeled dataset contain missing values.

Feature values were calculated for each 8-hour time series window, and include both point and storm attributes. In this study, a storm is defined as a continuous volume of precipitation in space and time. Storm attributes were averaged throughout the storm footprint. When several storms occur during an 8-hour window, storm values were averaged. Table 2.1 provides a description of all features with a comparison of the mean and standard deviation of feature values for the labeled and statewide datasets. Features were chosen to capture the variability and magnitude of the rainfall, the orographic influences, the radar QPE uncertainty, and space-time characteristics of the storms.

Table 2.1: Feature descriptions and associated mean and standard deviations for the labeled and the statewide Datasets

Feature	Mean labeled	Mean statewide	Std dev labeled	Std dev statewide
<i>Point</i>				
maximum 15-min intensity, mm/hr	8.68	7.41	13.02	11.91
median of nonzero intensity	1.41	1.25	2.16	2.05
std dev of non-zero intensity	2.40	2.03	3.73	3.44
variance of non-zero intensity (mm/hr) ²	19.69	16.00	71.92	68.71
mean of non-zero intensity	2.30	1.98	3.12	2.91
maximum accumulation, mm	0.47	0.39	0.70	0.64
total accumulation	5.09	4.52	6.42	5.62
median of non-zero accumulation	0.06	0.05	0.10	0.10
std dev of non-zero accumulation	0.12	0.10	0.19	0.17
variance of non-zero accumulation, mm ²	0.05	0.04	0.16	0.16
mean of non-zero accumulation	0.11	0.09	0.14	0.14
duration of non-zero accumulation values, min	234.33	242.20	134.50	133.23
month of 8-hour window	6.85	6.85	1.28	1.30
starting hour of 8-hour window	9.39	9.28	7.46	7.44
latitude of MRMS closest to gage	39.32	39.01	0.93	1.16
longitude of MRMS closest to gage	254.18	253.91	1.17	1.36
minimum Radar Quality Index	0.78	0.68	0.29	0.30
maximum Radar Quality Index	0.82	0.74	0.25	0.28
mean Radar Quality Index	0.81	0.72	0.26	0.28
median Radar Quality Index	0.82	0.73	0.26	0.28
std dev Radar Quality Index	0.01	0.01	0.03	0.04
elevation at MRMS coordinate, m	2161.07	2422.43	449.31	616.19
slope aspect at MRMS coordinate, deg	178.53	169.67	107.17	103.50
topographic slope at MRMS coordinate, deg	7.15	9.16	8.01	8.46

Continued on next page

Table 2.1 (continued) Feature descriptions (continued)

Feature	Mean labeled	Mean statewide	Std dev labeled	Std dev statewide
<i>Storm</i>				
mean variance of total accumulation between time steps (mm ²)	698.29	647.15	522.76	510.81
mean variance of accumulation between coordinates (mm ²)	5.48	5.08	4.67	4.45
mean storm-footprint area, km ²	91515.87	87971.39	62453.04	62948.86
mean storm-object velocity, m/s	13.94	13.56	4.63	4.76

Rainfall statistics were calculated from nonzero values in the 8-hour sample window. Elevation, aspect, and slope were extracted from the NASA Shuttle Radar Topography 1-arc second dataset (NASA Jet Propulsion Laboratory, 2022). Aspect and slope were calculated using ArcMap (Esri, 2020). Non-precipitation values with missing data were filled with NaN. All feature values were standardized for training of models by subtracting the mean and scaling it to the standard deviation of the feature values.

Storm regions were identified using the scikit-image label module after converting the MRMS precipitation rate product to a binary array, with a value of 1 assigned to values greater than zero (van der Walt et al., 2014). The label module assigns a unique storm-id to neighbors in the array with equal values. Diagonal neighbors were not considered connected to avoid stringing together objects. Objects are assigned storm ids throughout time, latitude, and longitude. Storm values were processed in monthly chunks to manage memory resources and values were not computed across chunks.

A statewide (west of 104° longitude) dataset was developed to visualize the predictions of the model and create a spatially continuous error estimate of MRMS where no gage exists. This dataset includes 598,310 8-hour time series of MRMS intensity with the 28 features sampled throughout the state where the total accumulation is greater than 1 mm. To decrease the size of the statewide dataset, the original MRMS dataset was sampled every 10 grid coordinates, resulting in a resolution of approximately 10 km. Data were sampled from May through September from 2018 through 2023.

MRMS is subject to several temporal sampling errors mainly caused by advection of hydrometeors (Fabry et al., 1994). Rather than directly compare the gage and MRMS time series, the time series was first upsampled by selecting the maximum 15-minute intensity from each 10 minute interval. A more complex approach that includes correcting for sampling error using the movement of the storm such as the method presented by Seo and Krajewski (2015), would likely decrease the error further.

2.3.3 Modeling

Model description

The dataset labeled with true RMSE values from gages is used to train a tree-based ensemble model, which is then used to predict the error of MRMS with the statewide dataset and understand how the extracted features are associated with performance. The output of the model is a range of RMSE values for the MRMS time series sample. To predict this range, a gradient-boosting regressor is trained with quantile loss, using the implementation from scikit-learn (Pedregosa et al., 2011). Gradient boosting, and other machine learning regressors, make predictions by identifying patterns during their training phase. They optimize these predictions using a suitable loss function and the target values. Once trained, these models can generate predictions for new data, without target predictions. Boosting methods in particular, combine or ensemble smaller decision trees to produce more powerful models, improving the balance of bias and variance in individual decision trees, with the aim of decreasing over-fitting and increasing accuracy (Géron, 2021). The gradient-boosting algorithm learns sequentially from the residual error of an ensemble of decision trees (Friedman, 2001). Each ensemble of decision trees is essentially attempting to improve on the residual error from its predecessor. An additional benefit of gradient-boosting models is that there are relatively few parameters to tune, making them relatively straightforward to use.

The residual error or training loss is calculated as mean quantile loss, otherwise known as mean pinball loss, shown in Equation 2.1:

$$L(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^{N-1} \alpha \cdot \max(y_i - \hat{y}_i, 0) + (1 - \alpha) \cdot \max(\hat{y}_i - y_i, 0) \quad (2.1)$$

Optimizing the loss involves minimizing the difference in the predicted quantile and a target group of size N . We essentially have three models with different α values, representing the quantile, to predict the median and 90% confidence interval. In Equation 2.1, when $\alpha = 0.50$ the loss is equal to half of the mean absolute error. The loss (L) is a function of the observed value (y) and the predicted value ($\hat{y}$).

Model performance was initially evaluated on a diverse set of single-output regression models to compare performance. The models included k -nearest neighbors, decision trees, various decision tree ensemble models, linear regression, multilayer perceptron, and support vector machines. Models were chosen based on common use and varying complexity. A brief description of these models can be found in the Appendix A Table A1. After hyperparameter tuning, the coefficient of determination and mean absolute error (MAE) was 0.80 and 0.96 mm/hr for the highest-performing model, random forest regression. The model is not robust to outliers and makes many predictions significantly higher and lower than the target value (Table A2, Figure A2 in Appendix A). The quantile regression model encapsulates these erroneous predictions by providing a range of error. The confidence interval can also account for heteroscedasticity in the data. Linear quantile regression was evaluated for comparison, the performance is shown in section 3.3.

Evaluation

Models were evaluated using 5-fold cross validation and a separate test set from the previously mentioned labeled dataset. The validation set is used to understand how each model generalizes to new data, compare model performance during training, select hyperparameter values without leaking knowledge from the test set, and to compare feature importance and model calibration to the test set. During k -fold cross validation, the model is trained on $k-1$ folds then validated with the remaining fold. 20% of the dataset was kept entirely separate to evaluate the final model performance.

Performance of spatiotemporally correlated data can easily be inflated if dependencies in the sample feature values are not accounted for when splitting the training and testing data (Roberts et al., 2017). To limit correlation between data folds and testing data, samples were grouped by location. Cross folds were manually set to keep samples with the same location in the same fold. To allow for an adequate number of samples for training, testing and validation, all six years of data were used.

The models tested in this study assume independence between features. Correlated features must be removed to interpret how features are used by the model. Several features listed in Table

2.1 are unsurprisingly highly correlated. The Spearman's rank correlation test was used to determine the correlation between features and Ward's linkage was used to sort clusters of correlated features. A threshold of 0.55 was used for the distance linkage, and only one feature was selected from each cluster below this threshold. Each permutation of features below the threshold was evaluated by the $\alpha = 0.50$ model to determine the best performing group, with a remaining correlation high of 0.61.

After feature selection, randomized search with cross fold validation was used to tune hyperparameters for each quantile separately. This method randomly samples parameter values from a specified range through multiple iterations then outputs the parameter values for the highest performing estimators, based on negative mean quantile loss. Hyperparameters were altered to increase performance while decreasing the tendency to over-fit. See Appendix A Table A3 for the hyperparameters tuned and the final values chosen.

The performance of the tuned models was evaluated with mean quantile loss in validation and testing. Additionally, the fraction of target values that fall outside of the predicted 90% confidence interval were calculated to get a sense of how well calibrated the predicted confidence interval is; this value should be close to 10%. Feature importance was evaluated by randomly permuting each feature for both the validation and test dataset and determines importance based on the mean increase in mean quantile loss when permuting the feature. The comparison of validation and test importance shows how the model might be over-fitting with the features.

2.4 Results

2.4.1 MRMS performance trends

Figure 2.2a compares the distributions of RMSE values for all target values, test target values, the test $\alpha = 0.50$ predictions, and the statewide $\alpha = 0.50$ predictions. The distributions peak at approximately 1.4 mm/hr. The test predictions closely follow the left tail of the target distributions, then fall below the right tail of the target distribution. This suggests that the model sometimes underestimates outliers.

Figure 2.2b shows the distribution of $\alpha = 0.50$ predictions for the state, compared to the distributions of $\alpha = 0.50$ state predictions grouped by coordinates with lower (median prediction below 0.1 quantile) and higher (median prediction above 0.9 quantile) errors. The distributions are aligned, with differences in the thickness of the right tail, suggesting that the spatial differences in predictions are heavily influenced by outliers.

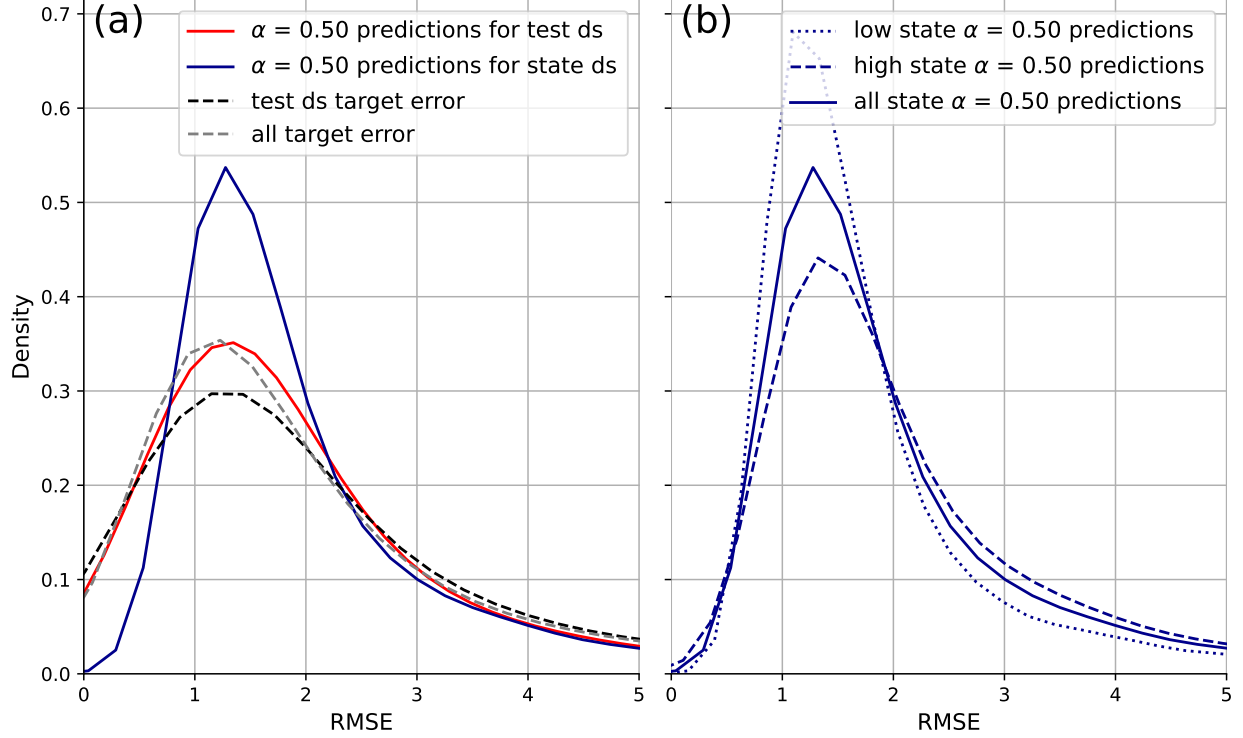


Figure 2.2: (a) Kernel density estimations of the distributions of RMSE for all target error, the test target error, the test predicted error, and the statewide predicted error. (b) The distribution of error for the coordinates in the state dataset with a median RMSE below the 0.1 quantile, above the 0.9 quantile, and all state predicted error values. The values only include the $\alpha = 0.50$ model predictions.

To understand the occurrence of false negatives the labeled dataset was constrained to a total accumulation of 1 mm for either the gage or MRMS (rather than just MRMS). 4.4% of the 8-hour samples had 0 mm MRMS total accumulation when the gage was above 1 mm (false negative). 24.2% showed greater than 1 mm MRMS total accumulation when the gage was 0 mm (false positive). 71% of the 8-hour samples had a higher total accumulation value for MRMS.

The spatial distribution of predicted median error is shown in Figure 2.3a. The values are normalized with the maximum MRMS 15-minute intensity of each 8-hour time series period. The MRMS grid locations with co-located gages are colored by the proportion of samples in the validation/testing labeled dataset. The lack of spatial uniformity in sampling for the labeled dataset does not appear to influence the spatial trends. There is a band of poor performance that arcs from the northwest corner of the state to the southwest corner that may be influenced by the terrain west of the continental divide and the distance from radar sites. However, areas on the map that show higher median nRMSE still have a distribution of error that includes lower nRMSE values, shown by the map of the first quartile of nRMSE in Figure 2.3b.

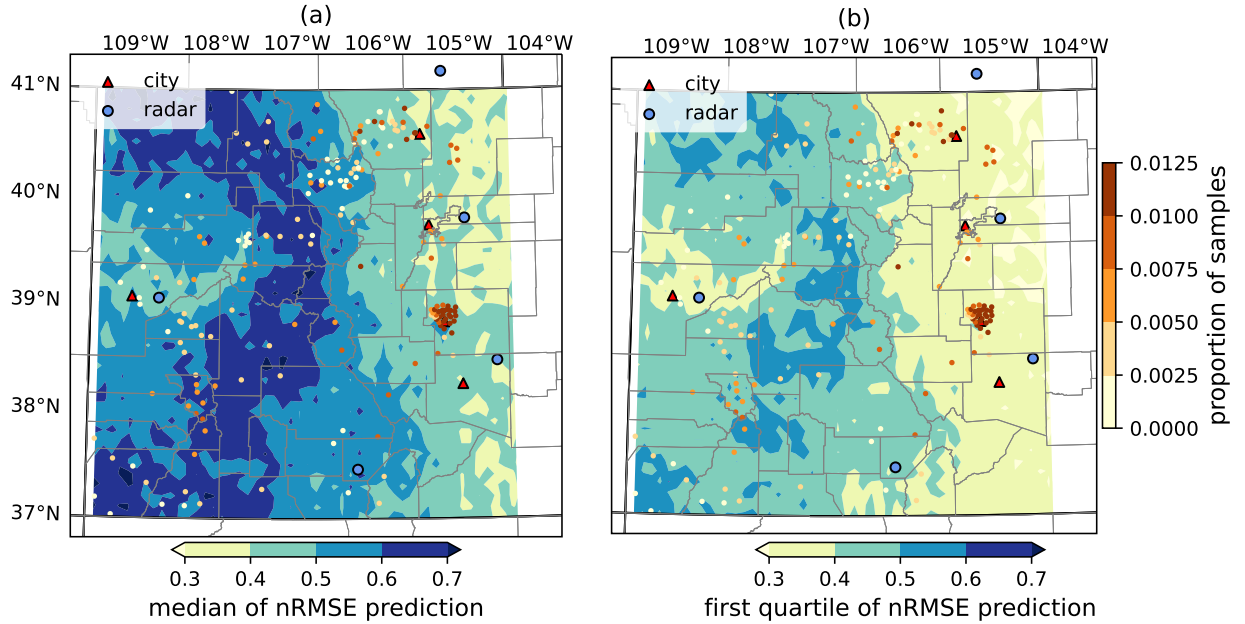


Figure 2.3: (a) The median normalized RMSE for each coordinate, predicted by the $\alpha = 0.50$ model. (b) The first quartile of normalized RMSE for each coordinate, predicted by the $\alpha = 0.50$ model. The gage locations are shown with colors representing the proportion of samples in the labeled dataset sourced from each location.

2.4.2 Features related to MRMS performance

The permutation feature importance for all quantile models for the validation and test dataset predictions is shown in Figure 2.4. Features used by the tuned model include: standard deviation

of intensity, median intensity, duration, latitude, elevation, aspect, area, velocity, minimum of RQI, and standard deviation of RQI. The error bars represent the standard deviation from 10 iterations of randomly permuting each feature. Unsurprisingly, the standard deviation of MRMS 15-minute intensity was most important. Linear regression, using only standard deviation of intensity, results in a coefficient of determination of 0.76 and MAE of 1.03 mm/hr. The standard deviation of intensity was removed from Figure 2.4 to understand if other features influence the model predictions or are used to over-fit in training.

There is some agreement between the feature importance in the testing and training data for the 0.50 and 0.95 quantile models. The 0.05 quantile model seems to rely entirely on the standard deviation and median of intensity. The discrepancy in the ranking of features between models and between the training and testing feature importance values within models is evidence that the models did over-fit. Although, these differences are marginal considering the magnitude of permutation importance. The RQI minimum and duration of positive values were both important for the 0.50 and 0.95 quantile models. This consistency suggests that these features are slightly useful in determining error.

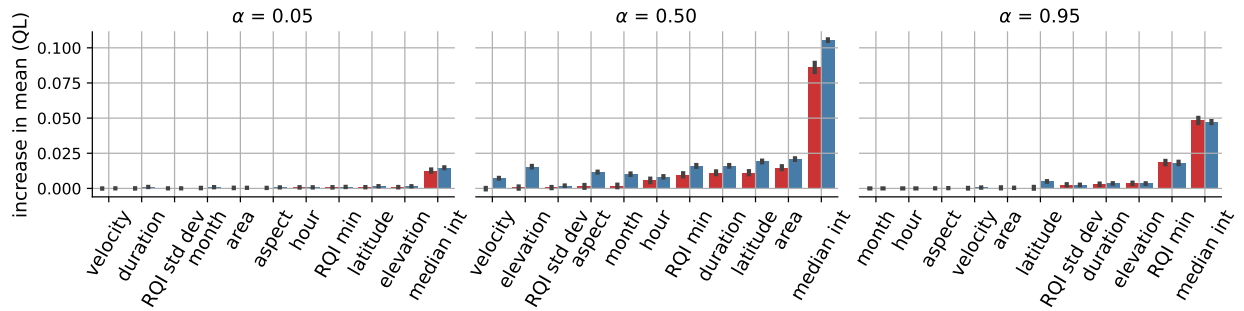


Figure 2.4: Permutation feature importance, defined by an increase in mean quantile loss (QL), across the three quantile models. The intensity standard deviation is excluded to focus on features not highly correlated with RMSE. The error bar represents the standard deviation between the 10 permutations. Importance is sorted from left to right by test importance (red). The training importance is shown in blue.

The difference in feature values for the statewide dataset for high and low predicted error extremes from the $\alpha = 0.50$ model are shown in Figure 2.5. Distributions of feature values are

separated by low (RMSE less than 0.95 mm/hr or the 0.10 quantile of predictions) and high (RMSE greater than 5.60 mm/hr or the 0.90 quantile of predictions) predictions. The hour and month features are excluded, because they are essentially categorical features.

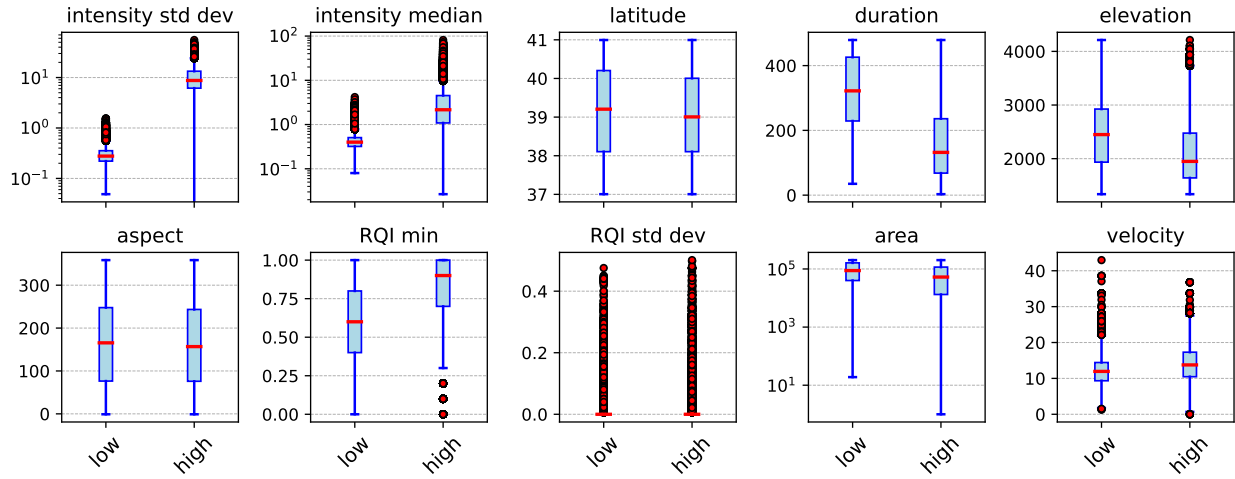


Figure 2.5: Comparison of feature value distributions for samples with low (0.95 mm/hr) and high (5.60 mm/hr) RMSE predictions for the $\alpha = 0.50$ statewide predictions. See Table 2.1 for feature units.

There is a clear difference in the inter-quartile range for standard deviation of intensity and median intensity. The inter-quartile range is lower for the low RMSE values for standard deviation and median intensity features, as expected. The inter-quartile range for duration is non-overlapping between groups, suggesting that MRMS is more accurate when longer duration rainfall occurs. The distributions for other features differ across performance groups, but are highly overlapping, meaning that there is no obvious decision boundary observed in these individual features. Surprisingly, the median of RQI minimum is lower (higher is better) for lower error. This might be explained by the more frequent observation of heavier precipitation, which will have a higher RMSE, in areas with better radar coverage and higher RQI. The median of RQI minimum becomes higher for lower error values, compared to higher error values, if RMSE is first normalized (see Appendix A Figure A3).

Table 2.3: Cross fold validation and test results ^a

	Validation				Test	
	Mean QL	Std QL	Mean FO	Std FO	QL	FO
GB $\alpha = 0.05$	0.100	0.013			0.104	
GB $\alpha = 0.50$	0.456	0.040	0.111	0.015	0.454	0.120
GB $\alpha = 0.95$	0.189	0.020			0.199	
L $\alpha = 0.05$	0.110	0.018			0.115	
L $\alpha = 0.50$	0.474	0.044	0.102	0.020	0.483	0.105
L $\alpha = 0.95$	0.208	0.018			0.215	

^aResults for quantile loss (QL) and fraction of outliers (FO) for both gradient-boosting (GB) and linear (L) regressor models. Validation results include the mean and standard deviation across folds.

2.4.3 Model performance

A summary of validation and test results is shown in Table 2.3 for both the gradient-boosting (GB) and linear (L) regressors. A perfect value for quantile loss (QL) is zero. A perfect value for the fraction of values outside of the 90% confidence interval (FO) is 0.10. The validation standard deviations for the quantile loss and outside of the 90% confidence interval show how the model performance varies across different data folds. Model performance on the test data is slightly lower, which is expected given that features were chosen and hyperparameters were adjusted based on the validation data. The linear regressor is slightly better calibrated, but the gradient-boosting regressor has lower variation in performance across folds and lower mean quantile loss for all models in both validation and testing.

Figure 2.6 shows the target RMSE as a function of the predicted median RMSE with the predicted 90% confidence interval the background for the test dataset. A 1:1 line is plotted to show how the median prediction either over or underestimates with growing RMSE magnitude. The confidence interval widens as the target RMSE values increase. The kernel density estimation for the target values as a function of the median predicted RMSE is shown in cyan contours. The kernel density contours show that the median predictions slightly underestimate the RMSE, consistent with Figure 2.2. The scatter points depict the target RMSE values, the red indicates values outside

and blue indicates values inside the 90% confidence interval. The RMSE truth values include many outliers, causing the model to be sensitive to these variations, which results in a noisy confidence interval.

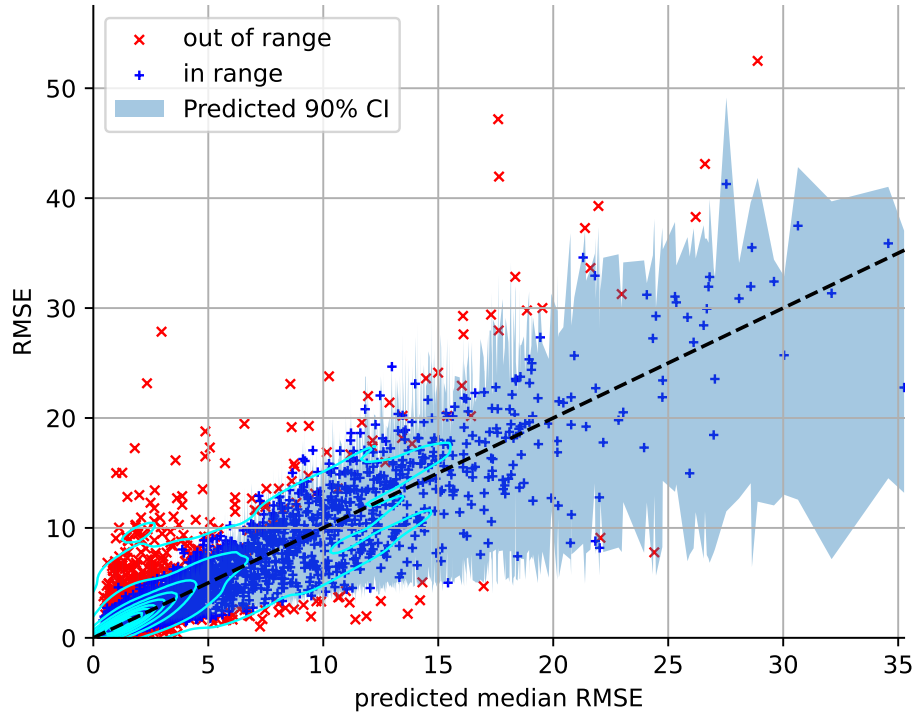


Figure 2.6: The target values as a function of the median predicted RMSE for the test dataset. Each data point represents an 8-hour sample RMSE between gage and MRMS plotted on the predicted. The predicted 90% confidence interval (CI) is plotted. Truth values in range are blue, those that fall outside the cutoff are red. The kernel density estimation for the target values as a function of the median predicted RMSE is shown in cyan contours.

2.5 Discussion

2.5.1 When and where are sub-hourly MRMS precipitation estimates trustworthy in the mountains of Colorado?

To better understand the quality of MRMS in the mountains of Colorado, sub-hourly MRMS precipitation estimates were compared to gage records from 2018 through 2023 during warm months. The gage record comparison shows a wide distribution of nRMSE with a median of 42%

and a standard deviation of 40% of the maximum MRMS 15-minute intensity. A gradient-boosting regressor was used to interpolate error estimates where no gages exist and provide a spatiotemporally varying record of error in Colorado. The model predictions of nRMSE and RQI aggregated through time at each coordinate show some spatial agreement for high error and uncertainty, shown in Figures 2.3a and 2.7a. According to our model results, these areas still see lower error values, perhaps when longer-duration storms occur higher in the atmosphere rather than during orographically enhanced or smaller convective disturbances. Figure 2.7b shows that areas with the lowest median RQI minimum (high uncertainty) remain low even in the upper quartile. This contrasts with the model results, shown in Figure 2.3b. The lower variability in RQI is caused by the essentially fixed beam blockage factor. The model shows that MRMS error in remote areas with poor radar coverage and complex terrain is not as consistently unreliable as suggested by RQI or the study by Maddox et al. (2002).

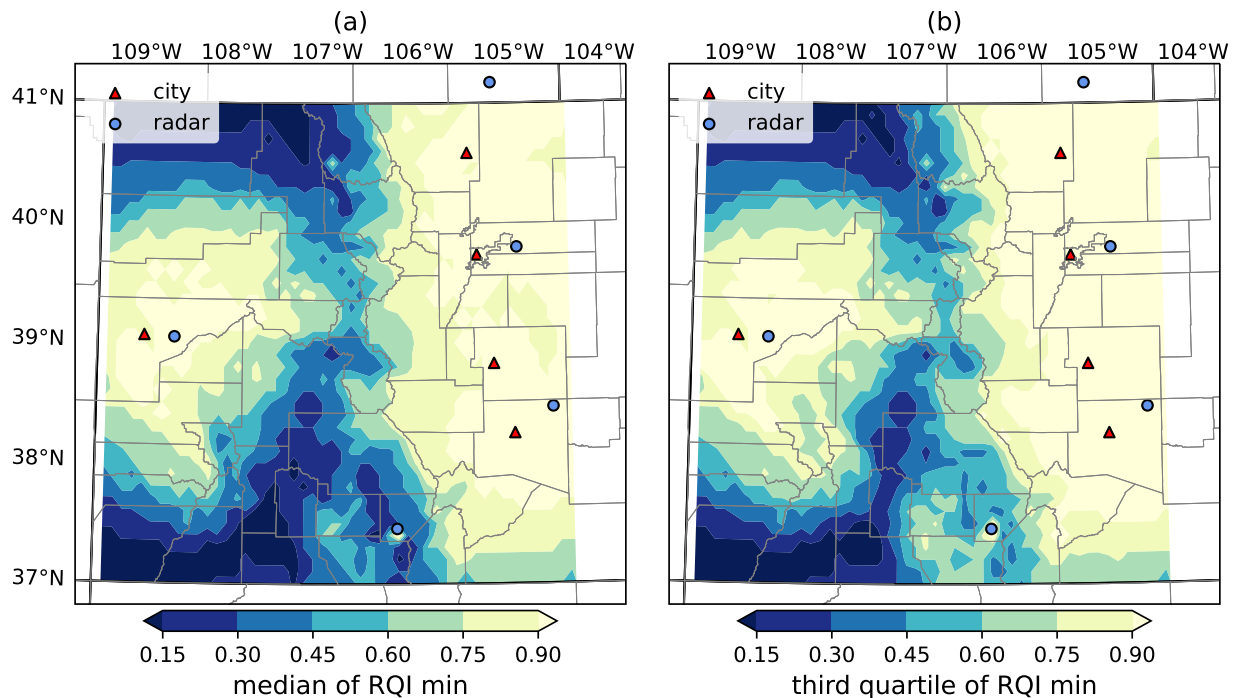


Figure 2.7: (a) Median of all sample RQI minima for each coordinate in the statewide dataset across the 6 years. (b) The third quartile of RQI minima.

To understand the impact of the changes to the algorithms and data inputs, described in Section 2.2.1, on the observed error (calculated from the gages) we first looked at the distribution of RMSE and normalized RMSE as well as RQI through the years. The difference in RQI, at the gage sites, before and after 2021 is clear at lower latitudes (see Appendix A Figure A4) and longitudes west of the Front Range (see Appendix A Figure A5). Changes to RMSE and normalized RMSE are less obvious, with the main difference after 2021 being that the inter-quartile range for RMSE and nRMSE appear to widen.

To understand the influence of the changes to the dataset on the model training and predictions, we re-trained the model using years 2018-2020 and then 2021-2023. The test skill, measured by mean quantile loss, improved (decreased) for the model trained with years 2018-2020 compared to the original model trained with all years (see Appendix A Table A4). We also experimented with adding year as feature to see how this would help the model implicitly learn the changes to the MRMS dataset, however this model performed worse than the original model. Spatial patterns between the original and pre-2021 model are similar (see Figure S6 and Figure 2.3a), with the pre-2021 model having slightly higher nRMSE (Figure S6, Figure 2.3a). The distribution of error for the pre-2021 model for the test and state predictions looks similar to the original model distribution output (see Appendix A Figure A8,A7 and Figure 2.2a, Figure 2.6).

In Figure 2.8 we compare the predictions and RQI values separated by the times before and after the major MRMS version change. The original model predictions are compared to the predictions from the models trained exclusively on pre-2021 and then post-2021 data. Predictions between the original model and models trained on subsets of data are different in magnitude, but the original model picks up on similar spatial patterns, including the area surrounding the Alamosa radar. The models make predictions consistent with spatial patterns of RQI values.

Freezing precipitation is not unlikely in May and September, especially in the mountains. This has an effect on the error and RQI, especially at higher elevations (see Appendix A Figure A9 and A10 for MRMS v12.0 data). We experimented with training the model on data from June through August; however, the performance was significantly worse (see Appendix A Table A4). The de-

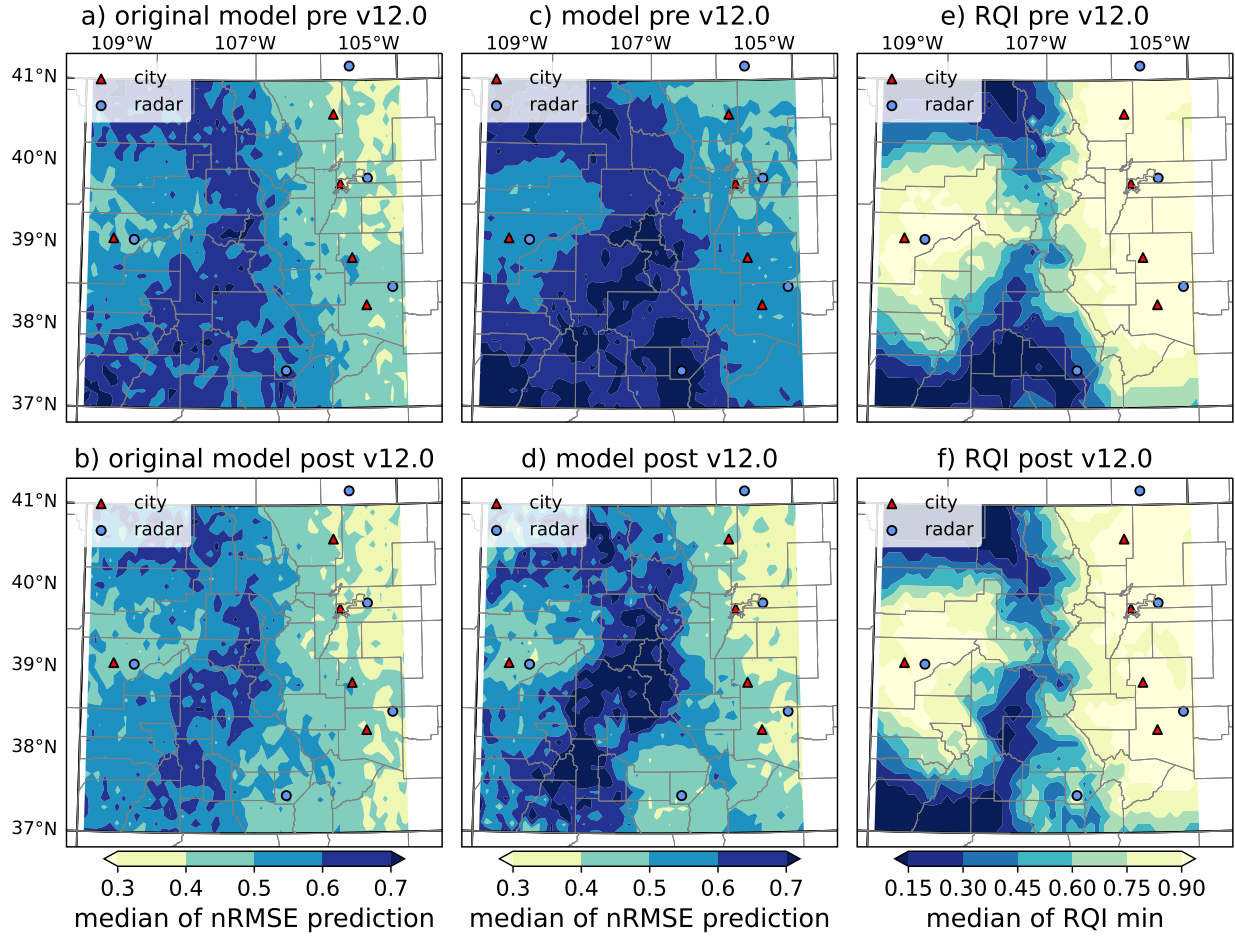


Figure 2.8: (a) Median normalized RMSE (nRMSE) $\alpha = 0.50$ predictions prior to MRMS v12.0 and (b) after MRMS v12.0. (c) Median nRMSE $\alpha = 0.50$ predictions from the model trained only on data prior to MRMS v12.0 and (d) from the model trained only on data after MRMS v12.0. (e) Median of the RQI minimum for data prior to MRMS v12.0 and (f) after MRMS v12.0.

crease in performance may be caused by insufficient training data. Separating the predictions by month shows that the model seems to accurately predict the differences in error at higher elevations in May and September (see Appendix A Figure A11). Still, including these months could introduce bias as the gage data we are training with may be inaccurate. Other MRMS variables could be used to filter out freezing precipitation; however, this additional data processing is outside the scope of our study.

We experimented with predicting the mean error (MRMS intensity minus gage intensity) for each 8-hour sample to capture the sign of the error. The distribution of these predictions, illustrated in Appendix A Figures A12 and A13, indicates that the model struggles to predict negative error. The map of median mean error predictions, shown in Figure A14, does not align with the spatial patterns of RQI, suggesting the model is unreliable.

2.5.2 What are the circumstances that lead to low error?

The standard deviation and median of intensity were unsurprisingly the most important predictors of RMSE. Although, the RQI minimum and duration of rainfall did add some predictive skill, see Figure 2.4. Separation of very low and very high RMSE predictions for the state dataset, shown in Figure 2.5, indicates that longer duration precipitation tends to have lower predicted error. This suggests that radar is more skilled at estimating precipitation for storms producing continuous rainfall, but could also be explained by correlated features of longer lasting rainfall not included in the model, such as the vertical extent of the storm. Permutation importance is lower for features associated with the terrain and spatiotemporal attributes of storms, meaning that the model did not find these features useful in determining the RMSE.

The model trained exclusively on data prior to the version 12.0 update shows similar feature value distributions for low and high RMSE predictions compared to the original model, with duration and RQI min being the most different (between high versus low RMSE) of the features not directly related to the magnitude of RMSE (see Appendix A Figure A15 and Figure 2.5). The model trained to predict mean error showed a similar pattern for duration (see Appendix A Figure

A16). These results support the conclusion that longer lasting rainfall is associated with lower MRMS error.

To evaluate feature importance without the impact of magnitude on RMSE, we trained a model to predict nRMSE. The distribution of these predictions, illustrated in Appendix A Figures A17 and A18, indicates that the model struggles to accurately identify outliers. To enhance performance, we attempted training with a subset of the data, specifically using only data prior to MRMS v12.0. This improved the skill (see Table S5), but the model still has difficulty predicting outliers, and the results, as mapped in Figure A19, appeared to be blurred. The model identified the maximum intensity and the duration of precipitation as the most important features (see Figures A20 and A21).

2.5.3 How well can error be predicted by the physical aspects of the region, rainfall patterns, and storm characteristics?

Given the features we developed, the model is well calibrated to the 90% confidence interval and the mean quantile error is low for the test results, shown in Table 2.3. Model predictions can be used to evaluate the reliability of historic sub-hourly MRMS time series and identify where more observations are needed. This method also provides an approach to estimate error without making assumptions on the sources of uncertainty that are most important or the distribution of error. The type of algorithm used to model the error does not appear to be critical, as evidenced by the marginal performance gains of the gradient-boosting regressor over the linear quantile regressor.

Model results are limited by the assumption that the gage network is a reliable ground truth for MRMS. The difference in spatial sampling scale (point versus 1 km² grid) as well as the estimation of surface precipitation from a high radar beam height is known to cause significant uncertainty in error estimates (Tang et al., 2018; Villarini et al., 2008; Ciach and Krajewski, 1999). Several studies have avoided the ground truth uncertainty by comparing multiple QPEs. For example, Bytheway (2022) created a probabilistic QPE by combining multiple high resolution precipitation

estimates, including MRMS. As satellite data and forecasts increase in temporal resolution this type of probabilistic precipitation estimate might be possible at sub-hourly resolutions.

The error model is designed to predict spatiotemporally varying errors for retrospective analysis. It is inadequate for real-time monitoring of events due to the delay in calculating dynamic storm features. The method could be used for such analysis by identifying features in a quantitative precipitation forecasts, such as the High-Resolution Rapid Refresh model (HRRR). Alternatively, the HRRR could be used to quantify error by monitoring when the duration of rainfall falls within the quantile range for low (or high) error.

2.6 Conclusion

In mountainous regions, high-resolution data is often lacking but necessary to quantify the unique hydrometeorology and its related hazards. In this study, we evaluate MRMS sub-hourly precipitation estimates in the mountains of Colorado to understand general trends in performance. We also evaluate the physical attributes that are most important in determining error and the characteristics that are associated with various error values. The performance and feature evaluation are enhanced by a novel method for modeling radar-based QPE error that can be used to estimate error where gages are sparse. The spatial patterns of performance generally align with RQI; however, gage comparisons and model predictions show that the error is highly variable throughout the state. The feature importance and predictions show that duration of rainfall is a useful indicator of the performance of MRMS.

The changes to the MRMS data processing schemes and inputs over time present a challenge for our analysis. Training the model with a more consistent dataset, by using only data prior to the version 12.0 update, did not significantly change the results. The additional analysis shows that the original model picks up on the temporal differences in the dataset implicitly, through its spatiotemporally varying input variables (or features). Likely, the original model uses differences in RQI and observed rainfall duration to track differences in the MRMS dataset and associated changes in error throughout time.

Given a sub-hourly and spatially continuous reference dataset, a logical next step would be to improve the QPE accuracy and reliability, rather than estimate error. Recent studies have successfully improved satellite and radar based QPEs using more complex machine learning methods such as a convolutional neural network (CNN). CNNs excel in identifying spatial dependencies by scanning parts of an image, with matrix weights, rather than ingesting all data as does a fully connected network (Géron, 2021). Osborne et al. (2023) used a CNN with several radar variables as input and a gage network as the target precipitation to obtain more accurate hourly precipitation. The CNN QPE showed some improvement over the radar-only QPE, a valuable result in operational forecasting where the latency of the multi-sensor QPE hinders its usefulness. Hilburn et al. (2021) used a CNN to create synthetic MRMS radar reflectivity images from GOES-R that preserves the spatial resolution of MRMS using a U-NET architecture. The approach outperformed traditional data assimilation techniques and has the potential to generate more precise reflectivity estimates from satellites in regions where ground radar blockage is notable, such as the western United States. Model architectures such as visual transformers may be used to leverage spatial as well as temporal context during learning. Our analysis shows that the error of MRMS is often high in several regions of Colorado, and likely unacceptable for many applications. It is worth exploring the potential of more advanced algorithms to enhance high resolution QPEs in areas of complex terrain in addition to improving observations.

Chapter 3

Regional and temporal variability of orographic influence on subdaily precipitation intensity in Colorado²

3.1 Summary

The influence of mountains on precipitation processes can produce distinct spatial patterns in climatology. These relationships are often used to inform interpolation between precipitation observations in mountainous areas. At the event scale, the control of orography on precipitation is more variable and less understood than long term rainfall patterns. Anomalous precipitation events and variation at subdaily time scales are likely omitted from gage-based datasets due to low station density. Here, we use several remotely-sensed and model-derived hourly datasets to explore the influence of terrain on the magnitude of subdaily precipitation intensity throughout Colorado. Terrain effects on precipitation vary sharply across Colorado and shift with season and accumulation period. Precipitation–elevation relationships differ among basins: the Missouri and Arkansas show decreasing precipitation with elevation—an effect stronger for hourly than daily totals—whereas the Colorado and Rio Grande exhibit increasing precipitation with elevation, with daily totals rising more steeply and significantly than hourly ones. Gage-based frequency studies, limited by sparse networks, miss the frequency of high intensity clusters along the Front Range and Pikes Peak shown by gridded model data.

²A version of this chapter has been submitted to Water Resources Research.

3.2 Introduction

Orographic uplift, moisture blocking, and solar exposure create distinct patterns in precipitation climatology in the mountains, yet their interaction with atmospheric conditions during individual storms makes event-scale precipitation highly variable (Roe, 2005). At the same time, in-situ observations are sparse in remote mountainous areas and insufficient to capture this variability (Lundquist et al., 2019). Characterizing the potential for extreme precipitation at the event scale is important, as this drives hazards such as debris flows and flash floods (Moody et al., 2013; Cannon et al., 2008; Doswell et al., 1996). With high elevation terrain becoming more susceptible to wildfire, there is an increase in runoff and erosion potential from short duration intense rainfall, which makes accurate estimates of extreme precipitation at this scale more relevant (Higuera et al., 2021; Touma et al., 2022).

Interpolation schemes that account for orographic effects on precipitation—such as the Parameter-elevation Regressions on Independent Slopes Model (PRISM)—use daily observations to model the climatological gradient across stations (Daly et al., 2008). Because the datasets are informed by climatology, higher temporal resolutions (even daily) are known to be less accurate in the mountainous western US than flatter regions east of the Rockies (Daly et al., 2021). Events that occur between stations or processes that occur at time scales shorter than the input data duration will not be reflected in the interpolation scheme (Marra et al., 2021). Henn et al. (2018) found that the average difference in annual accumulation between several gage-based gridded datasets exceeds 200 mm/year in some areas in the Western US and that these datasets drastically underestimate precipitation in high elevation basins, showing less annual total precipitation than streamflow. The uncertainty appears to be worse in winter, higher elevation terrain, and in areas of lower gage density (Gutmann et al., 2012; Jing et al., 2017).

The increase in annual precipitation on the windward slope caused by orographic uplift is well documented and understood; however, at shorter durations orographically induced turbulence may weaken rainfall intensity (Dallan et al., 2023). Resolving differences in subdaily orographic effects has an impact on extreme value estimates of precipitation. Allamano et al. (2009) and Avanzi et al.

(2015) describe a reverse orographic effect for hourly intensities of annual maximum events in the Italian Alps. For shorter windows (1-3 hours) intensity decreases with elevation, but at longer durations intensity increases with elevation. Marra et al. (2021) and Formetta et al. (2022) explored sub-hourly windows further, finding that the reverse orographic effect persists for sub-hourly mean annual maximum intensities, but its influence on 100-year events diminishes as the intensity window shortens. Terrain-driven turbulence disrupts updrafts and lowers the rainfall intensity captured by a brief accumulation period; over longer accumulation periods, the organization of convective cells by terrain is captured and increases the intensity (Formetta et al., 2022). Differences in intensity trends at sub-daily durations is not captured by interpolated datasets.

This study focuses on the mountains of Colorado, an area with distinct and complex terrain morphology as well as precipitation climatology. The complexity of this region likely contributes to analogous conclusions from several studies on the terrain-precipitation relationship. Jarrett and Costa (1988) found that in the Front Range the largest floods occur from orographically enhanced convection at lower elevations. Jarrett (1993) used paleohydrologic methods to determine a significant flood elevation limit of 2300 m throughout the Colorado Rockies, suggesting a corresponding limit for extreme precipitation, confirmed by other regional studies (England et al., 2010; Javier et al., 2007). However, Mahoney et al. (2015), using daily maximum accumulations from gage records, found that extreme precipitation events occur at all elevations in Colorado. Rossi et al. (2020) reconciled these findings, observing a weak negative correlation between elevation and rainfall in the Front Range, while noting that other land surface variables more clearly explain the flood-elevation relationship. Recent improvements in convection-resolving models and methods to correct for radar beam blockage offer appealing opportunities to better capture the probability of extreme event-scale precipitation in mountainous terrain (Lundquist et al., 2019; Zhang et al., 2016).

Here, we compare patterns of precipitation across mountainous portions of Colorado produced by several gridded datasets. Our objectives are to (1) compare datasets; (2) characterize relationships between precipitation and elevation across regions and for different accumulation windows;

(3) evaluate seasonality of these trends; and (4) compare these to gage-based precipitation frequency studies. We focused on the median and 90th percentile of annual maximum hourly and daily accumulations in several gridded datasets. Regional and temporal variability of the orographic effect were evaluated both by the spatial overlap of relatively high annual max hourly and daily accumulations with prominent terrain features and with linear regression. Relationships vary significantly by region and duration. Comparing results from precipitation frequency studies that rely on interpolated gage data with a model-derived precipitation dataset revealed marked differences.

3.3 Data

Three datasets with hourly resolution were used to evaluate the influence of terrain on rainfall intensity and to understand the uncertainty of the findings. These sources include the four-kilometer long-term Regional hydroclimate reanalysis over the conterminous United States (CONUS404) (Rasmussen et al., 2023b), NOAA Analysis of Record for Calibration (AORC) (NOAA National Weather Service, 2024), and Multi-Radar Multi-Sensor (MRMS)(Zhang et al., 2016, 2020). The methods used to produce the datasets, the spatial resolutions, temporal consistencies, biases, and data input types vary significantly.

CONUS404 has a 4-km resolution and was developed as a reanalysis type product for water years 1980-2022, meaning that the methods used to develop the dataset are temporally consistent. We removed the first five years of the dataset for our analysis (only including water years 1986-2022) as Rasmussen et al. (2023a) identified this as their spin up period, or the time required for the model to become physically consistent given the initial boundary conditions. The dataset was developed by downscaling the European Center for Medium-Range Weather Forecasts (ECMWF) version five of the atmospheric reanalysis of the global climate dataset (ERA5) using a convection-permitting Weather Research and Forecasting (WRF) mesoscale model. Rasmussen et al. (2023a) evaluate the resulting precipitation variable by its ability to capture climatology, the diurnal cycle, various mesoscale convective systems (MCS) events, and hourly distributions of precipitation

rates at various locations. In general the performance is favorable to benchmarks, with one notable exception being its ability to capture the upper tail distribution of hourly rates in high elevation areas (AORC seemed to perform slightly better in these locations). The authors suggest that the spatial resolution fails to capture the influence of the highest topography. For winter precipitation, CONUS404 performed well in comparison to SNOTEL sites throughout the west, achieving a correlation coefficient of 0.914 for annual precipitation from water years 1986 through 2020 (Rasmussen et al., 2023a). The dataset was downloaded from the THREDDS server (Rasmussen et al., 2023c).

Analysis of Record for Calibration (AORC) has a spatial resolution of 1 km and was developed for a similar period as CONUS404, 1979-2023. However, data inputs vary significantly through time. The most recent period from 2016 through 2023 includes disaggregation of daily precipitation from the NOAA Climate Prediction Center (CPC) gage data and a Next Generation Weather Radar (NEXRAD) Stage IV quantitative precipitation estimate (QPE) based on a radar mosaic (Stage IV includes gage corrections and quality control) using mainly hourly Stage IV data (and other radar, gage, and satellite data when Stage IV is unavailable). The lack of climatology data to constrain the product during this period causes seasonal and regional error (Fall et al., 2023). At the same time, significant improvements in radar-based quantitative precipitation estimates (QPE) were made during this time period. This improves performance but adds temporal inconsistency. The most important change for the region of interest is likely the gap-filling radar added to the network in 2021 in Alamosa, CO (NSSL, 2024). Because the radar-based Stage IV is the primary data input, AORC is subject to imperfect conversion from reflectivity to rain rate, ground clutter, and beam blockage (a significant issue in Colorado's mountains) (Berne and Krajewski, 2013; Maddox et al., 2002). The dataset was accessed through the Registry of Open Data on AWS (NOAA, 2024).

Multi-Radar Multi-Sensor (MRMS) provides coverage of the United States and southern Canada and includes several datasets. This study evaluates the radar-only 1-hr quantitative precipitation estimate (QPE) developed using the surface precipitation rate product, a 2-minute precipitation rate estimate from a mosaic of the Weather Surveillance Radar-1988 Doppler (WSR-88D) radar

network and out-of-network gap-filling radar with a spatial resolution of 1-km (Zhang et al., 2016, 2020). The product includes several quality controls and an evaporative correction from the Rapid Refresh (RAP) model data (NOAA, 2024). All data were accessed through the Iowa Environmental Mesonet MRMS archived daily data (Mesonet, Accessed 2023). MRMS faces the same radar coverage limitations in mountainous areas mentioned earlier. Similar to AORC, this product is also temporally inconsistent with several algorithm updates and additional data inputs such as radar and gage sites. See <https://inside.nssl.noaa.gov/mrms/code-updates> for a detailed list of modifications (NSSL, 2024). Version 12.0, operational in February 2021, introduced substantial updates including a dual-polarization algorithm (Zhang et al., 2020), updates to the radar quality index, additional data assimilation for gap filling purposes (Martinaitis et al., 2020). This version update also included the Alamosa, CO radar site referenced above.

We limited comparison of results across datasets to the period 2016-2022, when AORC started using Stage IV QPE as the primary input and when MRMS data is available. CONUS404 was regridded from a projected grid to a rectilinear 4-km grid using bilinear interpolation. Datasets were regridded to the 4-km CONUS404 grid using conservative interpolation which is recommended when decreasing resolution to maintain local maxima (xESMF Developers, 2024). Radar reflectivity conversions perform worse for freezing precipitation, so our analysis was limited to June, July, and August for this comparison (Zhang et al., 2016). To examine long-term trends and seasonal variability, the 1986–2022 CONUS404 dataset was analyzed separately.

CONUS404 90th percentile of annual hourly and daily maxima are also compared to two precipitation frequency studies, Atlas 14 (Perica et al., 2013) and the Regional Extreme Precipitation Study (REPS) (NOAA Earth Systems Research Laboratory, 2018). REPS was developed for Colorado and New Mexico to update probable maximum precipitation for dam safety needs. Both studies use the NWS Cooperative Observer Program’s (COOP) and SNOTEL station data as the primary observation sources. Additionally, both studies use PRISM to produce a continuous grid. Based on the record length of CONUS404, we limited comparison to the 10-year average recurrence interval, the lowest return interval for REPS. Although the percentile approach does not

match the magnitudes produced by precipitation-frequency analyses that rely on extreme-value distributions, it still captures comparable spatial patterns of intensity. The REPS dataset was accessed from RTI International (2024) and the Atlas 14 dataset was accessed from NOAA (2024). These datasets were regridded to the CONUS404 grid using conservative interpolation.

3.4 Methods

The study area includes the mountainous area of Colorado, west of 104° longitude. To distinguish the effects of various hydroclimatic and topographic factors on annual maximum intensities, the results were separated by river basins using hydrologic unit map level 2 (HUC2) codes, henceforth referred to as regions (Seaber et al., 1987). Figure 3.1 shows the separate regions and varying topography within each region. Prominent topography is highlighted to assist in qualitative assessment of where intensity increases within regions around particular terrain features.

Orographic precipitation is highly sensitive to wind direction and the geometry of mountains (Picard and Mass, 2017). The regions may include morphology with drastically different orographic effects such as windward and leeward facing slopes in Region 14. The intent of separating regions is not to perfectly delineate orographic influences, but to sample the varying orographic effects throughout the vastly complex state.

To compare spatial patterns throughout the state, we calculated the median of the annual maxima of hourly and daily precipitation accumulations for each grid cell for summer months. CONUS404 was evaluated during all seasons from 1986 through 2022, separated by December through February (DJF/winter), March through May (MAM/spring), June through August (JJA/summer) and September through November (SON/fall). The 90th percentile of the annual maxima (for the entire calendar year) of CONUS404 was calculated at each grid cell for comparison with the 10-year recurrence interval values from Atlas 14 and REPS.

The influence of topography on rainfall at different temporal scales was evaluated at hourly and daily accumulation windows. Changing the accumulation window shows how the orographic effect is sensitive to the temporal scale. Linear regression was used to explore general trends

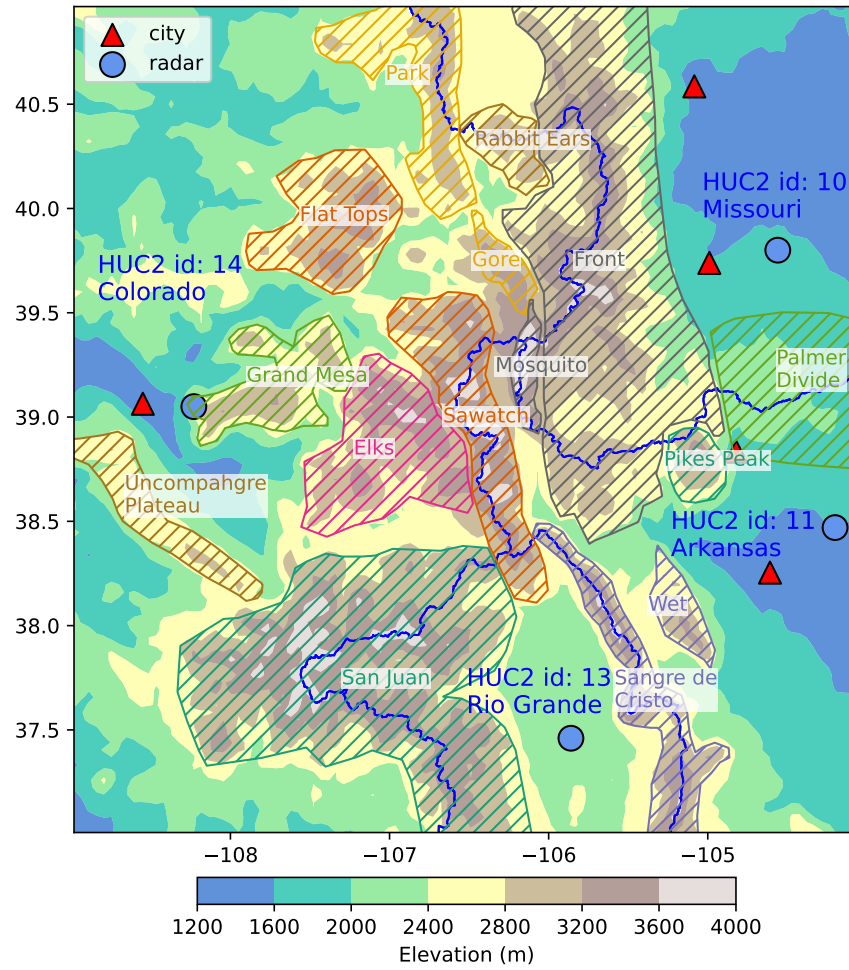


Figure 3.1: Terrain within the study area with relevant features highlighted. HUC2 boundaries used to evaluate regional differences are shown with black lines.

between elevation and intensity. Elevation data was sourced from the NASA Shuttle Radar Topography 1-arc second dataset (NASA Jet Propulsion Laboratory, 2022). The dataset was regridded to CONUS404 using bilinear interpolation to produce the area weighted mean elevation for each grid cell.

3.5 Results

3.5.1 Precipitation differences across datasets

Median annual maximum hourly and daily accumulation values are mapped to observe spatial patterns and compare differences across accumulation time windows, datasets, seasons, and precipitation frequency studies. Normalized median maximum accumulations for the three gridded datasets during JJA for 2016 through 2022 are mapped in Figure 3.2. Median accumulations for each grid cell are normalized by minimum and maximum median values for each dataset and HUC2 river basin. Normalization shows where precipitation is relatively high for the region, according to each dataset, and allows for comparison across hourly and daily accumulations. Hourly and daily accumulations are compared to understand how the location of relatively high precipitation shifts spatially in the region with a longer time window. Prominent terrain is outlined in white dashed lines to show which features overlap with relatively higher precipitation.

Several differences between datasets are evident in Figure 3.2. CONUS404 relatively high daily accumulations align with most terrain features. The Front Range in particular shows a clear increase in relatively high precipitation at higher elevations for daily accumulations. MRMS values in higher terrain, especially in Region 10 and 14, are clearly impacted by the distance from the radar sites and the beam blockage caused by mountains. AORC appears to be influenced by the quality of radar coverage as well, but to a lesser extent for daily accumulations, with higher values in areas like the Front Range and San Juans.

Dots represent the normalized median annual maximum hourly and daily accumulation amounts for the CoAgMET gages, colored with the same scale as the map (Colorado State University, Accessed 2023). Only gages that recorded during years 2016-2022 were used; this included a total

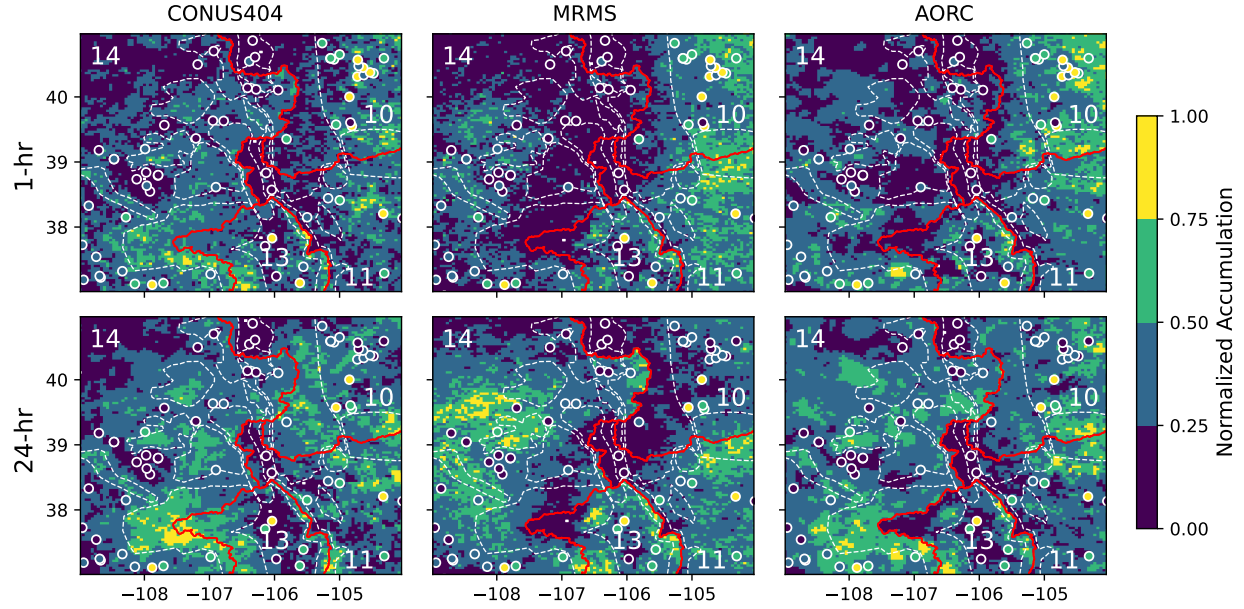


Figure 3.2: Normalized median annual maximum 1-hour and 24-hour precipitation accumulations for JJA covering 2016-2022. Accumulation amounts are normalized using the minimum and maximum median accumulations for each dataset and HUC2 region. Scatter points represent the normalized CoAgMET gage values for the same season and time period, with gages also normalized by HUC2 region (outlined in red). Terrain features are outlined by dashed white lines.

of 66 gages. There is some consistency between gage values and the gridded data. However, considerable variability exists between gages in similar locations—see the hourly values in the cluster in the northeast part of the study area. This variability may be due to the small sample size within each region. Another limitation in this comparison is that the locations of gages are limited to areas of low terrain. The mean elevation of this sample of gages is 1896 m, whereas the mean elevation of the study area is 2387 m.

The non-normalized median accumulation values for the three datasets are mapped in Figure 3.3 to compare magnitudes across datasets. Figure 3.3 shows that, relative to the other datasets, CONUS404 underestimates hourly accumulations east of the Continental Divide, especially at lower elevations and around the Palmer Divide. Higher elevation parts of the Front Range are similar for non-normalized values across datasets, corroborating the relatively higher values at higher elevations observed by CONUS404. Compared to CONUS404, AORC and MRMS underestimate annual maximums west of Continental Divide for hourly and daily accumulations, especially at

higher elevations. Artifacts of the River Forecast Center quality control of NEXRAD Stage IV data are evident in AORC, resulting in abrupt changes in magnitude along HUC2 region boundaries, particularly near the Continental Divide.

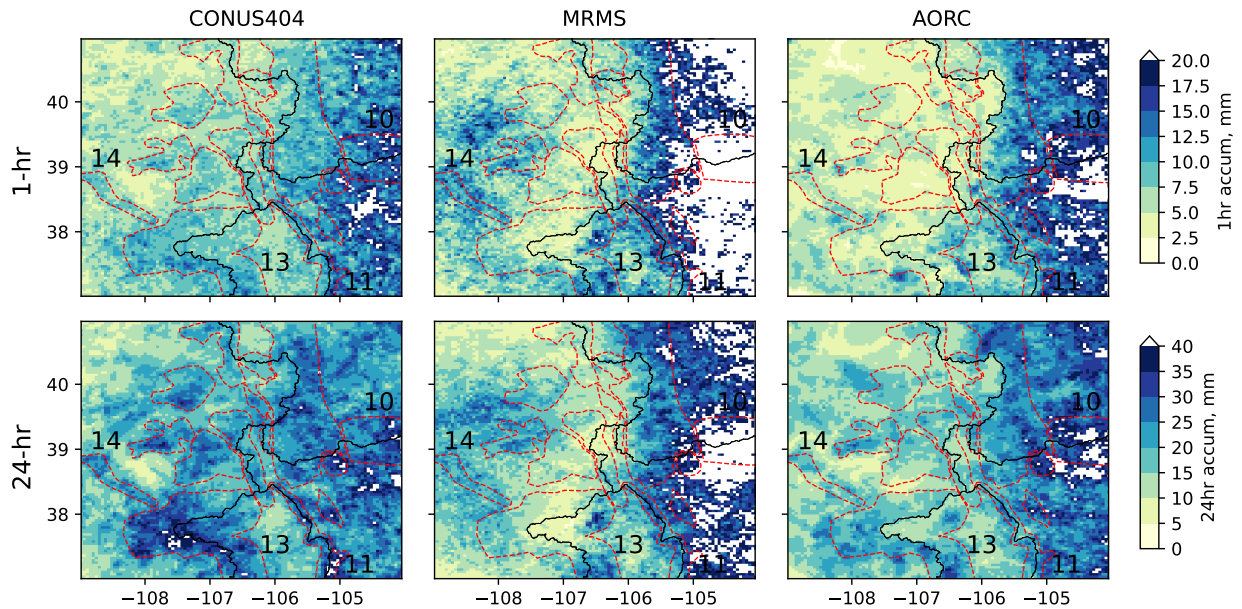


Figure 3.3: Comparison of median annual maximum 1-hr and 24-hr precipitation accumulations for JJA for three gridded datasets for 2016-2022. Terrain features are outlined by dashed red lines. Values above 20 mm for the hourly accumulation or 40 mm for daily are shown in white.

3.5.2 Seasonal variability in precipitation

Figure 3.4 shows the median annual maximum hourly and daily accumulations for CONUS404 separated by season, and normalized by region. The analysis uses CONUS404 data from years 1986-2022 to assess differences in annual maxima across seasons. The spatial patterns observed for JJA during 2016-2022 (see Figure 3.2) are similar to the JJA map from the longer record, indicating that the 2016-2022 period provides a sufficient sample to capture spatial patterns in median annual maxima. Although the higher relative values for the longer record are more spatially distributed for hourly accumulations, and more concentrated at higher elevations for daily accumulations, especially throughout the Front Range.

The distribution of normalized annual maxima varies notably by season. Clear terrain-related patterns emerge depending on the accumulation window and season: for example, higher relative values shift towards higher terrain in the Front Range (including the Pikes Peak region) for daily accumulations during MAM and JJA. The Wet Mountains show a strong daily signal in MAM and SON. And the Sawatch Range (along with the Park Range) sees higher relative daily values in DJF compared to other seasons. Terrain exerts a clear influence on annual maximum precipitation accumulations during all seasons, with the effect being more pronounced on daily timescales.

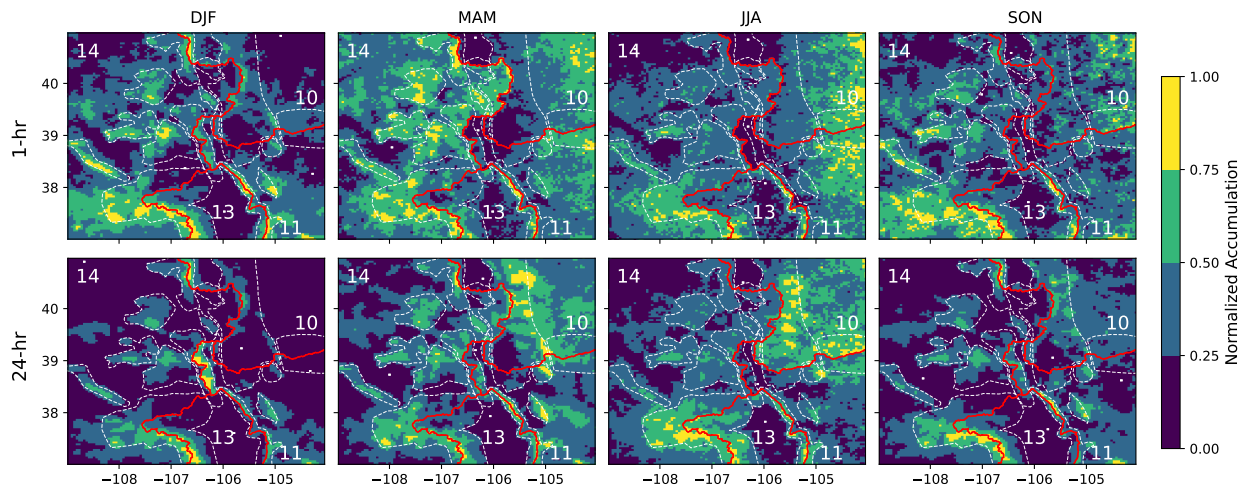


Figure 3.4: Normalized median of annual maximum 1-hr and 24-hr precipitation accumulations for CONUS404 for all seasons for longer record (1986-2022). Accumulation amounts are normalized with the minimum and maximum median accumulations for each season and HUC2 region (outlined in red). Terrain features are outlined by dashed white lines.

3.5.3 Comparison to precipitation frequency studies

Figure 3.5 compares the 90th percentile of annual maximum hourly and daily accumulations for years 1986-2022 of the CONUS404 dataset to the 10-year average recurrence interval hourly and daily accumulation values from REPS and Atlas 14. Values are normalized by HUC2 region.

Although the precipitation frequency studies are spatially smoother due to the methods used in producing gridded products from gage observations (see NOAA Earth Systems Research Laboratory (2018) and Perica et al. (2013)) there are similarities in the locations of relatively higher values.

The REPS hourly values appear to be limited by elevation, similar to CONUS404, especially along the Front Range and Sawatch. Contrastingly, Atlas 14 hourly values are relatively higher in higher terrain features. REPS shows a significantly higher area in Region 13 on the eastern side of the San Juans, this may be caused by lower station density and interpolation methods.

Daily values in Regions 13 and 14 are similar for all datasets, with Atlas 14 appearing to be more smoothed out in higher parts of Region 14, with the exception of the San Juans. The clusters of higher values along the Front Range and Pikes Peak in CONUS404, also shown in median values for MAM and JJA in Figure 3.4, do not appear in REPS or Atlas 14. The daily values in the Wet mountains and around the Palmer Divide appear to be higher for REPS and Atlas 14.

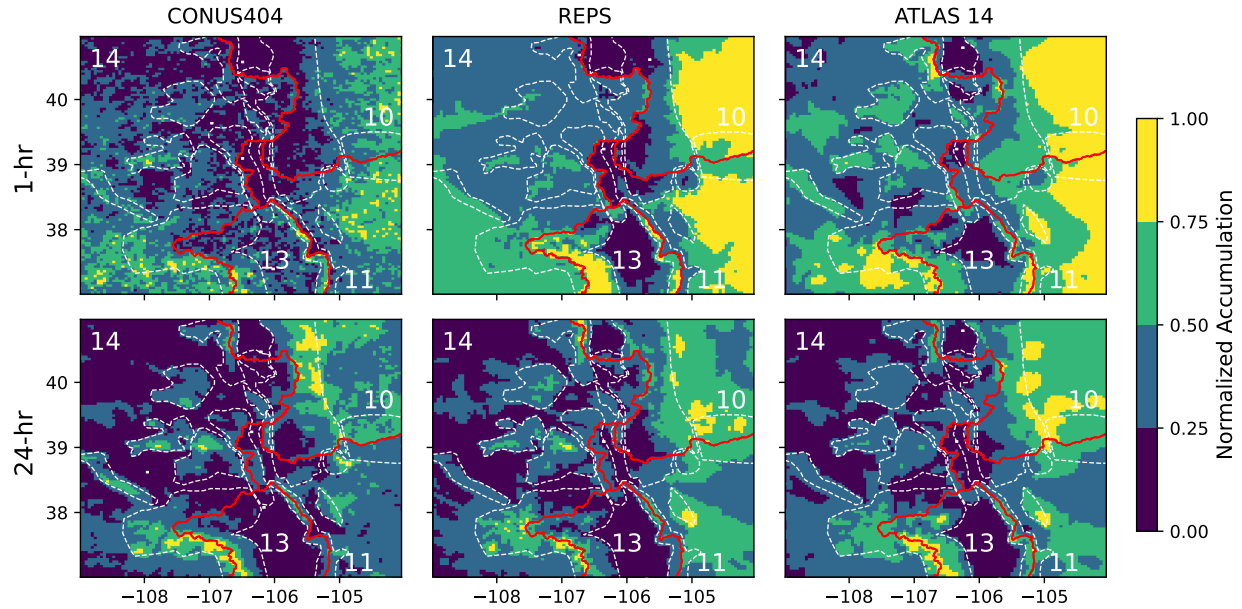


Figure 3.5: Normalized CONUS404 90th percentile of annual maxima and 10-year average recurrence interval for REPS and Atlas 14 for hourly and daily accumulations. Accumulation values are normalized with the minimum and maximum amounts for each dataset and HUC2 region (outlined in red). Terrain features are outlined by dashed white lines.

The non-normalized values for the 90th percentile of annual max hourly and daily accumulations for CONUS404 and the 10 year recurrence interval for REPS and Atlas 14 are shown in Figure 3.6. The non-normalized values help distinguish where differences occur between datasets irrespective of large outliers in the region. REPS and Atlas 14 values are clearly lower than the

higher intensity areas that CONUS404 shows along the higher elevation areas on the Front Range (including Pikes Peak) for daily accumulations. The lower values in Figure 3.5 along the Wets and Palmer Divide for daily CONUS404 values appear to be more similar to REPS and Atlas 14 in Figure 3.6. CONUS404 shows lower hourly values compared to both REPS and Atlas 14, which may be due to its use of hourly accumulated precipitation based on fixed, non-overlapping time windows.

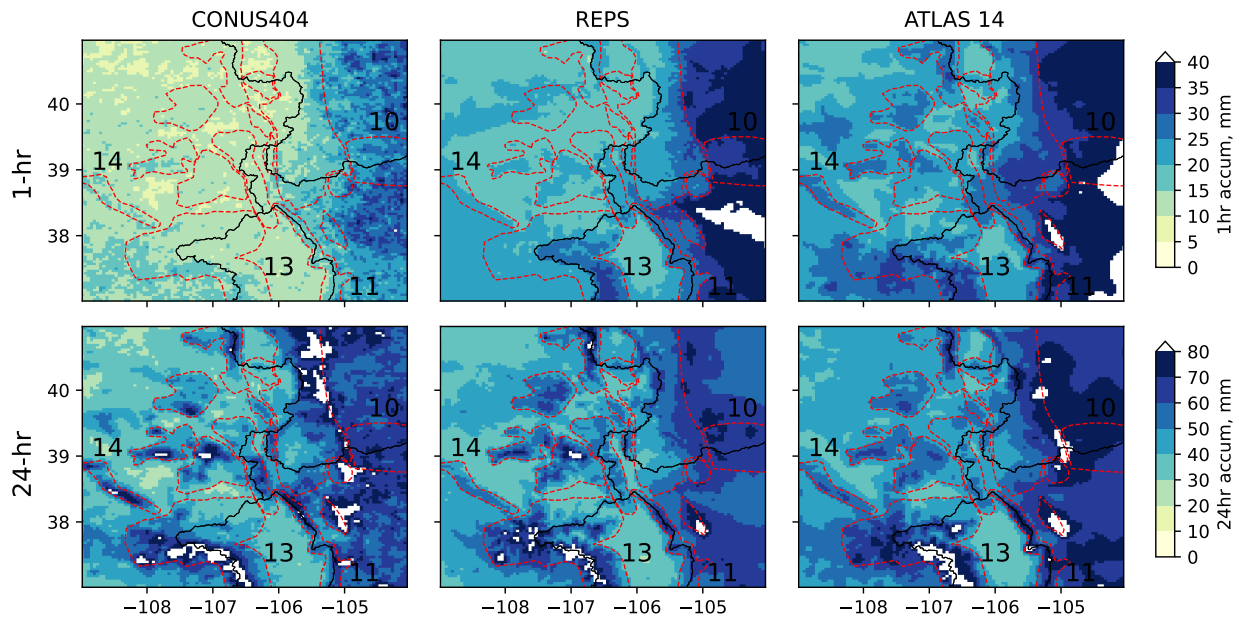


Figure 3.6: CONUS404 90th percentile of annual hourly and daily maxima and 10-year average recurrence interval for REPS and Atlas 14. Terrain features are outlined by dashed red lines. Values above 40 mm for the hourly accumulation or 80 mm for daily are shown in white.

3.5.4 Elevation-precipitation regression

Linear regression is used to explicitly examine the relationship between precipitation intensity and elevation. The slope resulting from this regression is shown in Tables 3.1 through 3.3. To compare values across time windows, datasets, and regions the normalized slope is also shown. Normalized slopes are bold and underlined where the daily slope is lower than the hourly slope, to highlight where the relationship deviates from the general trend of a stronger orographic effect

with a longer (daily) timestep. The p-value associated with the slope is marked with an asterisk when it meets the 0.05 significance threshold, based on a hypothesis test where the null hypothesis states that the slope is zero.

Figure 3.7 displays the linear regression for each gridded dataset for JJA. The elevation at each grid cell is regressed with the median annual maximum hourly and daily accumulations. Generally, slopes are negative in Regions 10 and 11 and positive in 13 and 14. The magnitude and trends between datasets are significantly different. MRMS and AORC slopes are lower or more negative for all regions, which is likely caused by beam blockage at higher elevations. For comparison, the longer CONUS404 record for JJA is also plotted, showing a similar slope to the shorter 2016–2022 period. The R^2 values (shown in Table 3.1) are generally low, which is expected given the variability in the data. The p-value tends to be small even when the regression appears to be a poor fit, likely due to the large sample size. Error bars show the inter-quartile range of median accumulation for elevation in 100 m bins. The error bars, especially for Regions 10 and 11 and non-CONUS404 data, indicate that elevation and accumulation are not linearly related; the map corroborates this, showing clusters of higher intensity rather than a smooth gradient with changing elevation.

Figure 3.8 shows the trends in median annual hourly and daily accumulations with elevation for CONUS404 separated by season. Again, the elevation at each grid cell is regressed with the median max annual hourly and daily accumulation at each grid cell. For all seasons and regions, the slopes shown in Table 3.2 become less negative or increase in magnitude for the longer accumulation window, with the exception of winter in Region 10 and 13. The slopes are more variable in Regions 10 and 11 across time windows and seasons. Regions 10 and 11 exhibit lower R^2 values than Regions 13 and 14, and their R^2 declines when moving from hourly to daily durations; in contrast, in Regions 13 and 14 R^2 values increase from hourly to daily durations.

Figure 3.9 compares the linear relationship between annual precipitation extremes and elevation for CONUS404 and the precipitation frequency studies. The elevation at each grid cell is regressed with the 90th percentile of the annual maximum hourly and daily accumulations or the

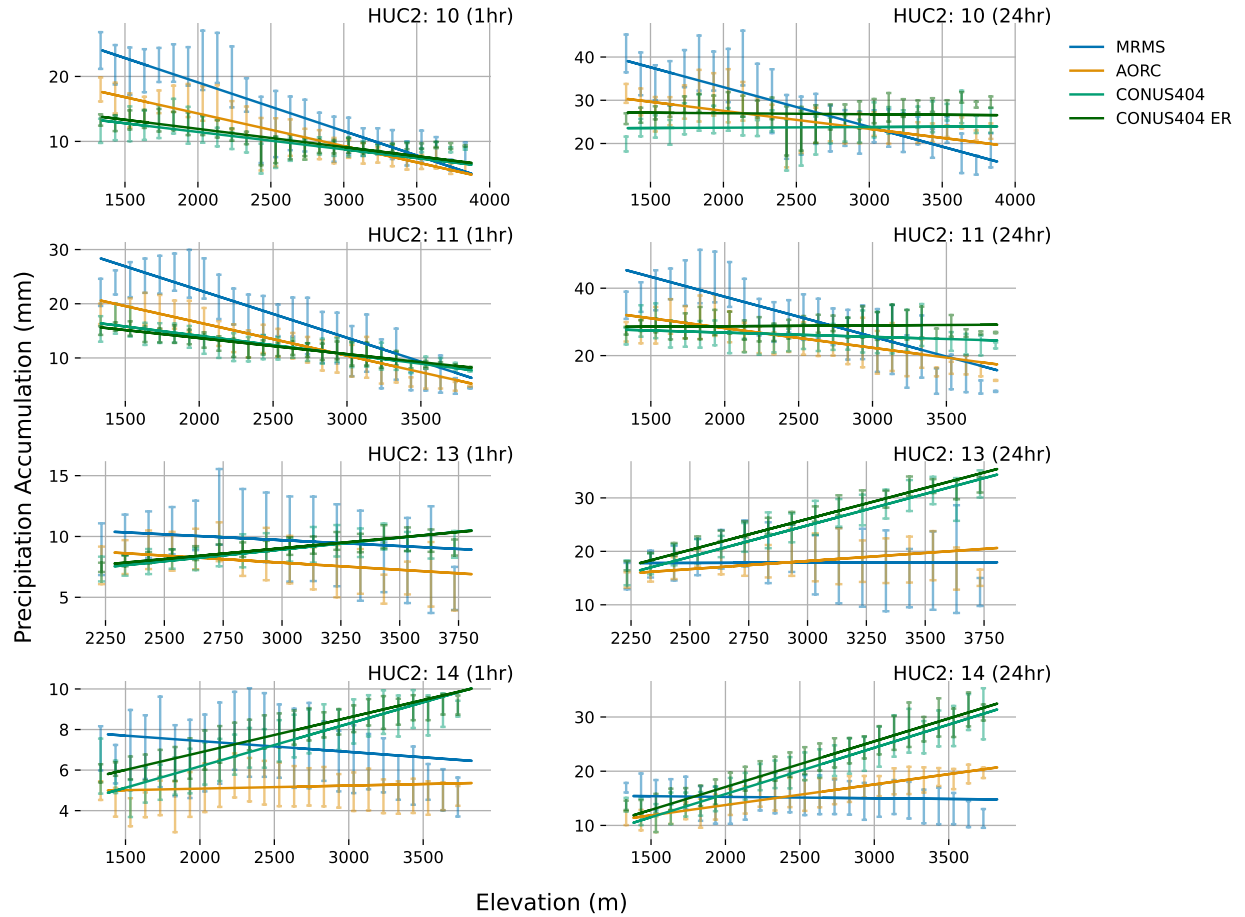


Figure 3.7: Linear regression between median annual maximum hourly and daily accumulation for JJA and elevation for the period 2016-2022 for each gridded dataset. The longer record (ER, years 1986-2022) for CONUS404 for JJA is also plotted for comparison. Error bars show the inter-quartile range for the median of annual maximum 1-hr and 24-hr precipitation for grid cells within 100 m elevation bins.

Table 3.1: Comparison of dataset regression statistics for 1-hour and 24-hour windows by HUC2. Asterisks indicate p-values < 0.05 . Normalized slopes are bold and underlined where the daily slope is lower than the hourly slope.

HUC2	Dataset	Slope (mm/km)		Norm Slope		R ²		p-value	
		1hr	24hr	1hr	24hr	1hr	24hr	1hr	24hr
10	MRMS	-7.50	-9.17	-0.19	-0.11	0.46	0.33	0.000*	0.000*
	AORC	-5.02	-4.17	-0.20	-0.10	0.43	0.16	0.000*	0.000*
	CONUS404	-2.68	0.15	-0.12	0.01	0.23	0.00	0.000*	0.352
	CONUS ER	-2.80	-0.24	-0.21	-0.01	0.44	0.00	0.000*	0.058
11	MRMS	-8.78	-11.83	-0.20	-0.16	0.47	0.34	0.000*	0.000*
	AORC	-6.10	-5.84	-0.17	-0.12	0.44	0.20	0.000*	0.000*
	CONUS404	-3.44	-1.26	-0.12	-0.03	0.27	0.01	0.000*	0.000*
	CONUS ER	-2.95	0.25	-0.21	0.01	0.45	0.00	0.000*	0.128
13	MRMS	-0.96	0.09	-0.04	0.00	0.01	0.00	0.000*	0.844
	AORC	-1.17	3.03	-0.08	0.11	0.04	0.08	0.000*	0.000*
	CONUS404	1.94	11.79	0.17	0.36	0.22	0.66	0.000*	0.000*
	CONUS ER	1.78	11.59	0.27	0.47	0.42	0.87	0.000*	0.000*
14	MRMS	-0.53	-0.25	-0.02	-0.01	0.01	0.00	0.000*	0.035*
	AORC	0.15	3.80	0.01	0.13	0.00	0.24	0.000*	0.000*
	CONUS404	2.11	8.55	0.15	0.21	0.33	0.54	0.000*	0.000*
	CONUS ER	1.72	8.42	0.18	0.23	0.38	0.61	0.000*	0.000*

Table 3.2: Comparison of seasonal regression statistics for 1-hour and 24-hour windows by HUC2. Asterisks indicate p-values < 0.05 . Normalized slopes are bold and underlined where the daily slope is lower than the hourly slope.

HUC2	Season	Slope (mm/km)		Norm Slope		R ²		p-value	
		1hr	24hr	1hr	24hr	1hr	24hr	1hr	24hr
10	DJF	0.48	4.90	0.14	<u>0.11</u>	0.38	0.33	0.000*	0.000*
	JJA	-2.80	-0.24	-0.21	-0.01	0.44	0.00	0.000*	0.058
	MAM	-1.35	-2.26	-0.21	-0.06	0.55	0.04	0.000*	0.000*
	SON	-0.31	1.13	-0.08	0.04	0.09	0.05	0.000*	0.000*
11	DJF	0.40	4.98	0.12	0.19	0.19	0.42	0.000*	0.000*
	JJA	-2.95	0.25	-0.21	0.01	0.45	0.00	0.000*	0.128
	MAM	-0.62	1.09	-0.11	0.03	0.19	0.01	0.000*	0.000*
	SON	-0.13	2.70	-0.02	0.10	0.01	0.18	0.000*	0.000*
13	DJF	1.69	20.04	0.43	<u>0.35</u>	0.69	0.57	0.000*	0.000*
	JJA	1.78	11.59	0.27	0.47	0.42	0.87	0.000*	0.000*
	MAM	1.00	12.71	0.29	0.42	0.49	0.72	0.000*	0.000*
	SON	1.36	15.13	0.31	0.38	0.46	0.68	0.000*	0.000*
14	DJF	0.63	10.61	0.16	0.18	0.20	0.39	0.000*	0.000*
	JJA	1.72	8.42	0.18	0.23	0.38	0.61	0.000*	0.000*
	MAM	0.86	9.35	0.23	0.25	0.52	0.69	0.000*	0.000*
	SON	0.58	8.04	0.12	0.18	0.11	0.39	0.000*	0.000*

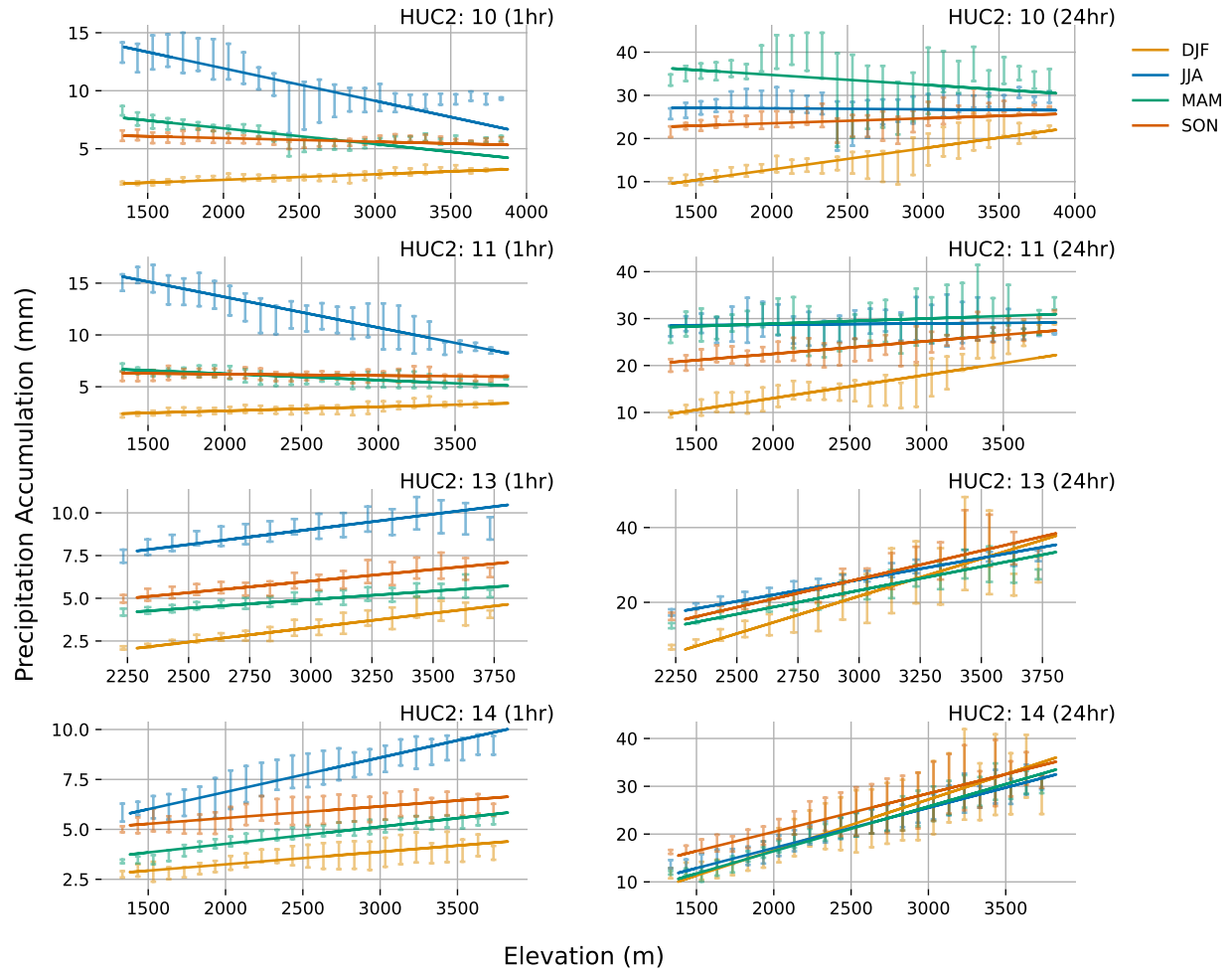


Figure 3.8: Linear regression between median annual maximum hourly and daily accumulation and elevation, separated by season, for CONUS404 for years 1986-2022.

10 year recurrence value at each grid cell. Trends are similar to those shown in Figure 3.8. For all datasets, slopes shown in Table 3.3 are negative across Regions 10 and 11, but positive in Regions 13 and 14. In Regions 10 and 11, the slopes approach zero and values are more variable for daily accumulations. In Regions 13 and 14 the slopes become more positive and values are more variable for daily accumulations. Atlas 14 shows a smaller hourly-to-daily slope gap than CONUS404 or REPS, consistent with its higher hourly values in Figure 3.5.

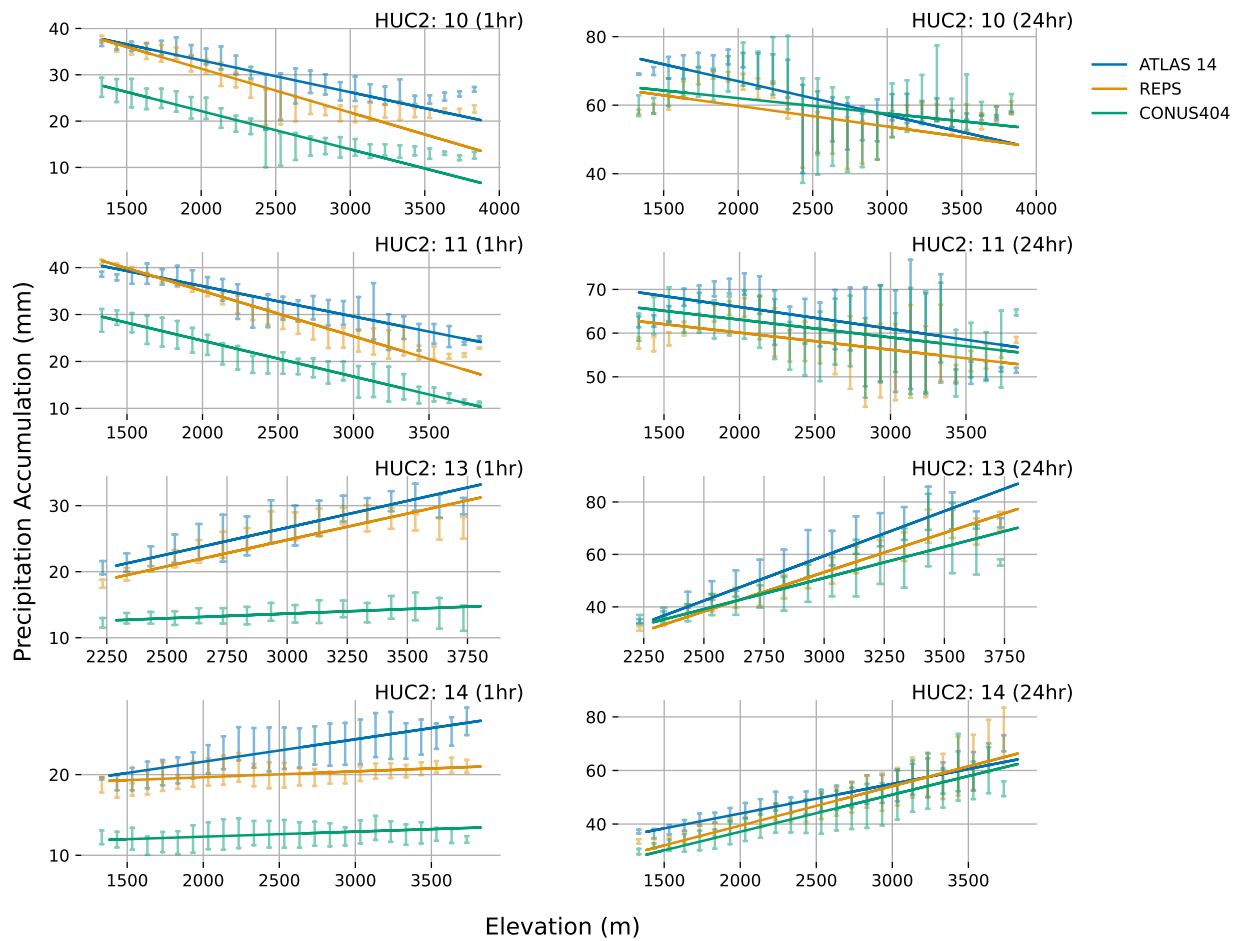


Figure 3.9: The linear relationship between the 90th percentile of the annual maximum hourly and daily accumulation and 10-year average recurrence interval versus elevation for CONUS404 for years 1986-2022, REPS, and Atlas 14.

Table 3.3: Comparison of CONUS404 and precipitation frequency study regression statistics for 1-hour and 24-hour windows by HUC2. Asterisks indicate p-values < 0.05 . Normalized slopes are bold and underlined where the daily slope is lower than the hourly slope.

HUC2	Dataset	Slope (mm/km)		Norm Slope		R ²		p-value	
		1hr	24hr	1hr	24hr	1hr	24hr	1hr	24hr
10	ATLAS 14	-6.89	-9.87	-0.27	-0.18	0.59	0.38	0.000*	0.000*
	REPS	-9.45	-6.06	-0.42	-0.13	0.81	0.21	0.000*	0.000*
	CONUS404	-8.26	-4.49	-0.25	-0.07	0.64	0.06	0.000*	0.000*
11	ATLAS 14	-6.46	-5.00	-0.22	-0.08	0.50	0.10	0.000*	0.000*
	REPS	-9.67	-3.90	-0.42	-0.07	0.89	0.08	0.000*	0.000*
	CONUS404	-7.65	-4.06	-0.25	-0.05	0.61	0.05	0.000*	0.000*
13	ATLAS 14	8.10	34.15	0.34	0.48	0.56	0.80	0.000*	0.000*
	REPS	7.98	29.94	0.52	<u>0.37</u>	0.68	0.81	0.000*	0.000*
	CONUS404	1.40	23.75	0.13	0.35	0.11	0.57	0.000*	0.000*
14	ATLAS 14	2.78	11.07	0.13	0.16	0.18	0.42	0.000*	0.000*
	REPS	0.72	14.74	0.04	0.19	0.04	0.54	0.000*	0.000*
	CONUS404	0.62	13.86	0.04	0.18	0.02	0.38	0.000*	0.000*

3.6 Discussion

3.6.1 Regional precipitation intensity trends

There are distinct trends in precipitation magnitude with elevation across the four river basins. In the Colorado (Region 14) and Rio Grande (Region 13) river basins, CONUS404 shows relatively higher precipitation at higher elevations for all seasons for both hourly and daily durations (see Figure 3.4) and a positive slope for the regression with elevation (see Figure 3.8). The Missouri (Region 10) and Arkansas (Region 11) river basins are more variable. With the exception of winter, slopes in these regions are negative or close to zero.

For JJA, MRMS and AORC show similar patterns to CONUS404 in Regions 10, 11, and 13 with some differences at the highest elevations likely caused by beam blockage (see Figure 3.2). Despite the well-known limitations of radar-based precipitation estimates in complex terrain, several mountainous areas have median annual maxima comparable to CONUS404, such as the Front Range, and the lower Arkansas including Pikes Peak and the Wets (see non-normalized map, Figure 3.3). MRMS values are clearly impeded by complex terrain in Region 14, while AORC seems to match CONUS404 more closely in this region.

Linear regression slopes are similar in sign and magnitude between CONUS404, REPS, and Atlas 14 (see Figure 3.9). However, there are important differences in the location of higher precipitation around specific terrain shown in the map (Figure 3.6). The Front Range and Pikes Peak in particular show several clusters of higher precipitation in higher elevations in daily CONUS404 that appear to be missed by REPS and Atlas 14. Daily MRMS and AORC align more closely with these areas of higher intensity (see Figure 3.3). The potential underestimation of daily extremes by precipitation frequency studies in this area is particularly consequential due to its proximity to dense population areas.

Local orographic enhancement has been shown in other studies. Banta and Schaaf (1987) used satellite data to identify the locations of thunderstorm initiation throughout Colorado, they found thunderstorms tend to initiate in clusters around the state. Using higher resolution satellite data, Schaaf et al. (1988) found specific terrain features and flow conditions that led to thunderstorms in the Sangre de Cristo mountains. Mazzoglio et al. (2022) found local terrain morphology more useful as a regressive predictor of annual rainfall extremes in Italy than elevation alone.

Other studies have found spatial error in longer accumulations across interpolated datasets (Henn et al., 2018). Figure 3.10 shows the difference in the mean seasonal accumulation for years 1986 to 2022 between PRISM and CONUS404. The dataset was downloaded from the PRISM Climate Group FTP system (PRISM Climate Group, 2025). The top row shows the absolute difference and the bottom row shows the mean relative error (with respect to CONUS404). The difference between datasets is significant, with relative error larger than 25% of the mean seasonal accumulation of CONUS404 for large areas of the state for all seasons. Seasonal accumulation of CONUS404 shows similar clusters of heavy precipitation along the Front Range that were shown in daily accumulations. The values in these clusters are significantly higher than PRISM estimates. A comparison of two convection permitting WRF simulations with important differences in model setup (including boundary conditions, run time periods, and physical schemes) to PRISM seasonal accumulations resulted in similar bias in higher elevations (Jing et al., 2017). This suggests that the uncertainty in seasonal accumulation is more likely caused by interpolation errors and/or gage

under-catch than from the modeled precipitation. The models used in this study have the same spatial resolution as CONUS404 and one has the same boundary conditions.

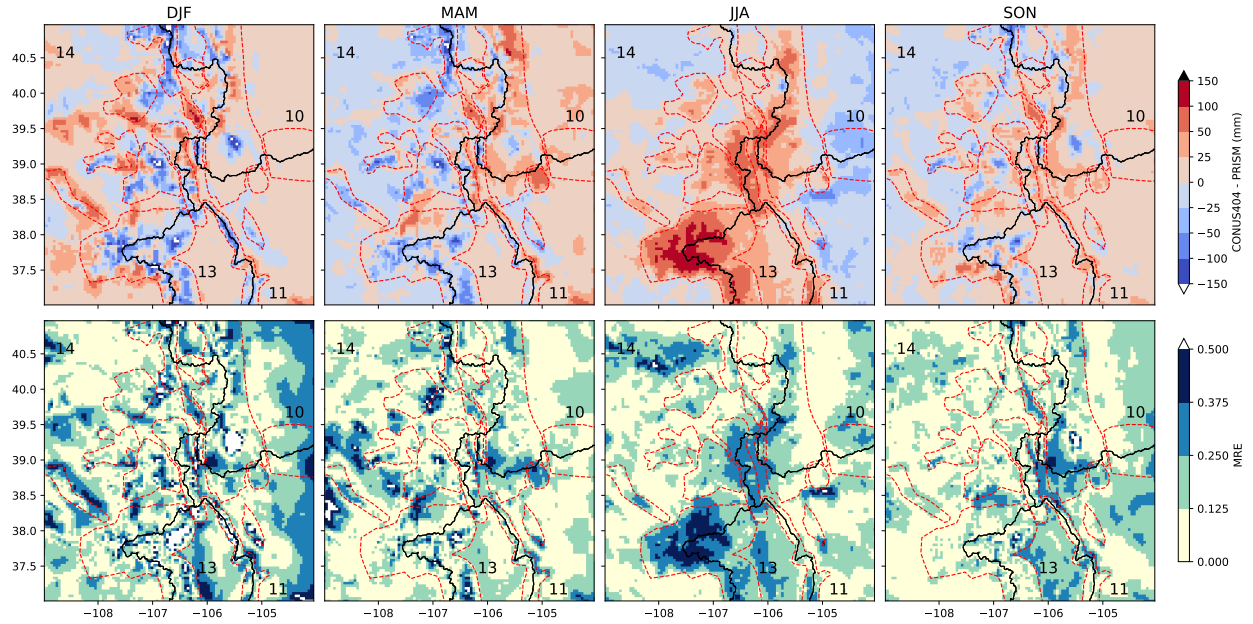


Figure 3.10: Absolute error (top) and mean relative error (bottom, $\text{CONUS404-PRISM}/\text{CONUS404}$) for the mean seasonal accumulation for years 1986-2022. HUC2 regions are outlined in black and terrain features are outlined in red.

Spatial patterns of relatively high median annual max daily precipitation shown by CONUS404 during summer months are compared to lightning frequency in Figure 3.11. Although ground lightning sensors have several sources of error (Zhu et al., 2020), and the relationship between rainfall and lightning is highly variable (Wu et al., 2017), the data provides a useful comparison to verify locations of higher rainfall intensity. The World Wide Lightning Location Network (WWLLN) monthly thunder hour dataset was used to map the median accumulated summer thunder hours for 2013 through 2023 (Virts and Holzworth, 2025). A thunder hour is defined as two sensors observing ground strokes within 15 km from the grid cell. The dataset is half degree resolution, and was regridded to the CONUS404 4-km grid with conservative interpolation. The daily JJA CONUS404 median accumulation map (Figure 3.4) matches lightning frequency more closely than hourly accumulation, suggesting that there is convective organization occurring throughout the study area.

Lightning is more frequent at the higher elevations of the Front Range and Pikes Peak, supporting the high-intensity patterns highlighted by CONUS404.

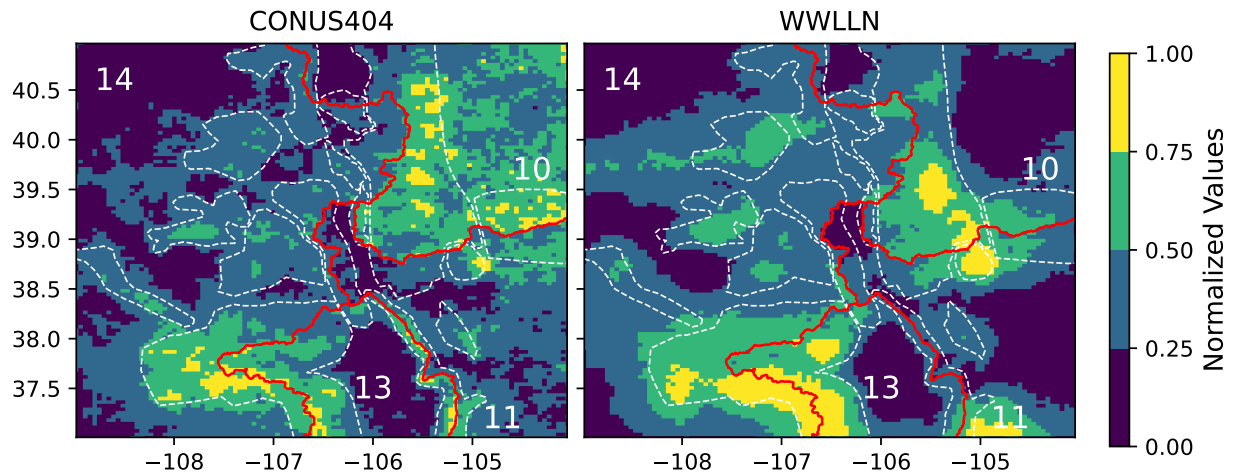


Figure 3.11: Normalized median of annual maximum daily precipitation accumulations for CONUS404 during JJA for 1986-2022 (left). And normalized thunder hours during JJA for 2013-2023 (right). Accumulation and thunder hours are normalized with the minimum and maximum median values for each HUC2 region (outlined in red). Terrain features are outlined by dashed white lines.

Hodanish et al. (2019) created a climatology of lightning density from cloud to ground stroke observations from years 1996 to 2016 throughout Colorado using the National Lightning Detection Network (NLDN). They also observed higher lightning density at higher elevation areas in the Front Range and Pikes Peak. Generally, they found that stroke density increases with elevation but is limited when surrounding mountains block moisture. For example, moisture in the upper Arkansas river basin is blocked by the San Juans to the west and the Wets to the east, explaining the lower accumulations in higher elevation areas in Region 11. Hodanish et al. (2019) also provide explanations for patterns in lightning density based on known interactions between atmospheric conditions and terrain. Most notably, additional instability during warmer months combined with the increased moisture from the North American Monsoon fuels the summertime spike in lightning activity, especially along the southern San Juans.

Differences in the relationship between annual precipitation extremes and elevation across regions suggest that there is not a clear elevation limit to extreme rainfall in Colorado that can explain

differences in flood history. Rossi et al. (2020) gave three explanations for the flood elevation limit in Colorado: lower rainfall intensity at higher elevations, higher likelihood of freezing precipitation at higher elevations, and variability in land surface variables. Using gage data, they found a non-significant downward trend in storm intensity with elevation in the South Platte. CONUS404 shows that the relationship between intensity and elevation is not uniform throughout the state and, depending on the region, may have a significant effect on flood potential.

3.6.2 Time dependence of orographic effect

The relationship between elevation and annual max intensity is dependent on the duration of the accumulation window. This is shown across datasets, seasons, and river basins (see Figures 3.7 through 3.9). In Regions 10 and 11 the relationship, shown by regression, is negative for hourly durations then shifts to a more variable and non-linear relationship for daily durations. In Regions 13 and 14 the relationship is less positive and more variable for hourly durations then increases for daily durations. This shift is clear in the regression for CONUS404 across seasons (except for winter in Regions 10 and 13), REPS, and Atlas 14 (except for REPS Region 13). The regression for MRMS and AORC consistently shows a negative relationship with elevation due to radar blockage.

Regions 13 and 14 appear to have a more gradual shift of relatively higher precipitation to areas of higher terrain for daily durations, whereas Regions 10 and 11 show higher precipitation at specific terrain features for daily durations. This illustrates why the linear regression performance changes across regions, likely reflecting differences in flow regimes, moisture availability, and terrain morphology.

Javier et al. (2007) observed the influence of terrain on the temporal organization of precipitation for JJA in Region 11. They evaluated 66 storms with radar precipitation estimates, and found that convective storms tend to initiate at higher elevations (most often around Pikes Peak and the Wets) then propagate eastward where storms become more organized. This general trend aligns with the median annual max daily CONUS404 values as well as REPS and Atlas 14 daily values (see Figure 3.6). However, analyses of annual maximum hourly and daily rainfall in other

regions suggest that patterns in the location of storm initiation and direction of propagation relative to terrain varies across the state.

The hourly values for Atlas 14 are higher than REPS and CONUS404 at higher elevations. This suggests Atlas 14 does not capture the temporal variability caused by terrain. Convection permitting models have also been shown to underestimate the reverse orographic effect, overestimating shorter duration extremes (Dallan et al., 2023). Additional analysis is required to understand the extent to which CONUS404 resolves turbulence and terrain features related to the decrease in the orographic effect at shorter durations.

3.7 Conclusion

A recently developed high resolution convection permitting reanalysis dataset (CONUS404) is used to re-evaluate the orographic impact on annual maximum precipitation intensity throughout Colorado. CONUS404 is temporally consistent, spans several decades, and seems to perform well in all seasons (Rasmussen et al., 2023b). Results were compared to radar-based precipitation estimates in the four major river basins in Colorado as well as a recently developed precipitation frequency study (REPS) and Atlas 14. There is a downward trend with elevation in the Missouri and Arkansas River Basins that is more significant for hourly than daily durations. And there is an upward trend with elevation in the Colorado and Rio Grande River Basins that is more significant and more positive for daily than for hourly durations. The radar-based datasets struggle to capture this due to beam blockage, especially in the Colorado River Basin. Precipitation frequency studies miss the high-accumulation clusters along the Front Range and Pikes Peak seen in CONUS404 daily data, likely due to sparse gages and terrain complexity. Compared to other datasets, Atlas 14 appears to overestimate hourly accumulations in higher terrain, suggesting it does not represent the reverse orographic effect adequately.

While CONUS404 generally aligns with lightning climatology and precipitation frequency studies, its capacity to capture the true spatiotemporal variability of precipitation in complex terrain is not clear. Uncertainty can be quantified by comparing convection permitting models (Adhikari

et al., 2024; Jing et al., 2017). Several studies have estimated true error indirectly with stream gages through highly calibrated hydrologic models (Moreno et al., 2012; Liao and Barros, 2023). These methods can be used to understand the accuracy of model-based precipitation estimates in complex terrain where limited observations exist.

These findings caution against the use of interpolated datasets or precipitation frequency studies based on interpolation for specific regions and shorter timescales in mountainous areas. The influence of terrain on precipitation varies considerably across Colorado, and also varies by season and timescale. Linear regression methods that overlook these nuances may fail to capture the full impact of terrain on precipitation extremes. Changes to the climate of Colorado could have significant implications for runoff and the potential for flooding (Bolinger et al., 2023). Accurately capturing precipitation extremes at the event scale is essential for characterizing changes to future flooding in the mountains.

Chapter 4

Effectiveness of mulching under varying rainfall intensity on steep slopes³

4.1 Summary

Wildfires alter the runoff and erosion potential of watersheds through the removal of vegetation and alteration of soil physical and chemical properties. Various soil and water conservation practices have been used to attempt to mitigate these post-fire hazards. Field and lab experiments have shown that organic cover material, such as mulch, can be effective at specific application rates. However, results vary significantly based on environmental factors, many of which have not been thoroughly tested. The substantial cost of mulching over large burn areas has driven continued interest in developing a fully quantitative approach to planning mulching operations. Although rainfall is the primary driver of post-wildfire runoff and erosion, its influence on the effectiveness of cover treatments has yet to be thoroughly evaluated. Field studies typically assess the influence of precipitation using long-term accumulations, while laboratory rainfall simulations often rely on uniform intensities. Non-uniform rainfall profiles can result in significantly higher runoff and erosion rates compared to uniform intensities. In this study, we use a rainfall simulator capable of generating time varying intensity profiles similar to natural rainfall observed by gages in the mountains of Colorado. We examine the effects of time-varying rainfall intensity on both mulched and bare soil on a steeply sloped flume. The effectiveness of mulch varies markedly between fluctuating rainfall intensities and a steady intensity. The wood mulch treatment failed to significantly reduce total runoff under time-varying rainfall, though it did under constant rainfall. Erosion rates were reduced with mulch for all rainfall events, despite the increasing intensity rain event causing significantly more erosion on bare soil.

³A version of this chapter has been submitted to Earth Surface Processes and Landforms.

4.2 Introduction

Wildfire-induced vegetation loss and soil changes can rapidly alter a watershed's runoff and erosion dynamics (Shakesby and Doerr, 2006). These changes vary with the intensity of the fire, along with the existing soil and vegetation types, and can influence several factors affecting soil erodibility and infiltrability. Specific changes to the soil include aggregate instability, soil sealing, and water repellency (DeBano, 2000; Varela et al., 2010; Ebel et al., 2022), while the vegetation removal dramatically impacts interception and increases splash erosion (Vieira et al., 2018). Changes to the soil composition and the vegetative cover removal can result in debris flows from relatively common storms and flooding that exceeds historic records prior to the burn (Moody et al., 2013; Cannon et al., 2008).

Post-fire hazards have been mitigated with various soil and water conservation practices. Cover treatments and barriers have been used in many contexts to reduce runoff and erosion including in construction sites, agricultural lands, rangelands, and burn scars (Fan et al., 2023). Several post-wildfire studies show that cover treatments, specifically wood and straw mulching, most effectively limit soil erosion and runoff (Girona-García et al., 2021). As a result, mulching has become a widely adopted practice following wildfires (Robichaud et al., 2013a).

Mulching decreases runoff and erosion through several complementary mechanisms. The material itself is absorptive and provides additional storage for rainfall, decreasing runoff (Prats et al., 2017). The surface cover also prevents direct impact of raindrops or splash erosion (Faucette et al., 2004; Foltz and Wagenbrenner, 2010). The protection from raindrops can prevent soil sealing, which is common in fire affected soils and can significantly reduce infiltration (Faucette et al., 2004; Larsen et al., 2009). At the hillslope and watershed scale, mulch increases roughness and diverts overland flow, dissipates its energy, and lowers sediment transport capacity (Foltz and Wagenbrenner, 2010; Fernández et al., 2020). Pieces of mulch can create barriers to sediment movement and cause localized deposition (Foltz and Copeland, 2009). By impeding overland flow and introducing small surface barriers, mulch promotes shallow ponding; the resulting rise in soil pressure head enhances infiltration (Montenegro et al., 2013; Prats et al., 2017). Over time, mulch can

promote faster vegetation regrowth in arid regions by retaining soil moisture (Fernández and Vega, 2016).

In mountainous terrain, straw mulch is easily blown down slope and may introduce non-native seeds, whereas wood mulch is more wind-stable and seed-free (Robichaud et al., 2013b; Copeland et al., 2009). In some locations, wood is more readily available (Dooley, 2003). Additionally, several studies have shown that wood mulch reduces runoff and erosion as effectively as straw mulch when applied at the same surface cover percentage (Robichaud et al., 2013b; Dooley, 2003; Girona-García et al., 2021). There is some evidence that wood mulch is more effective on fine-grained soils (Yanosek et al., 2006). For these reasons, wood mulch is typically favored for post-wildfire recovery, but it has been studied less than straw mulch because straw sees broader use in agriculture and construction (Girona-García et al., 2021). This study uses wood mulch on a steep slope to further explore its effectiveness for wildfire applications in a mountainous environment.

Rainfall is the primary cause of runoff and erosion after a wildfire, but it is difficult to accurately measure precipitation in the mountains at high resolution, and hydrologic models commonly used in post-fire studies fail to fully integrate spatiotemporally varying rainfall (Lundquist et al., 2019; Ebel et al., 2023; Collar et al., 2025). Rainfall variability remains a large source of uncertainty in the effectiveness of mulch (Vieira et al., 2018; Robichaud et al., 2013a; Girona-García et al., 2021). Rainfall simulations conducted on small field plots or laboratory flumes offer the potential to vary rainfall profiles and observe their complete impact on runoff response and sediment yield (Dunkerley, 2021). The sequence of a storm's intensity can drive peak and total runoff and erosion amounts. If the high intensity rainfall occurs earlier in an event, the soil is drier and absorbs more rainfall; if high intensity occurs later, infiltration is lower because soil has been wetted by preceding rainfall and more runoff occurs (Dunkerley, 2021). Early low-intensity rain can break down aggregates, fill depressions, and prime the surface for connected flow paths, so later high-intensity pulses trigger more ponding, rilling, and erosion (Dunkerley, 2012, 2019, 2021).

The influence of rainfall variability on bare soils has been studied outside of mulching, although there are conflicting results with different soils and moisture conditions (Dunkerley, 2012; Parsons

and Stone, 2006). Montenegro et al. (2013) applied different intensity profiles on mulched soils using a rainfall simulator, finding no significant difference between varying and constant rainfall for total runoff. However, because these events were sequential, the impact of the later pulse was likely influenced by antecedent conditions. Moreover, by using straw mulch on a gentle 10% slope, the study's findings have limited applicability to steep, burned mountain terrain.

In this study, we evaluate the effectiveness of wood mulch on a steep slope under varying rainfall. Rainfall simulations on small plots and laboratory flumes cannot capture every hillslope nuance, but they do allow precise control over rainfall intensity, rainfall timing, soil type, and slope. Our objectives are to (1) compare the influence of constant, decreasing, and increasing rainfall profiles on runoff and erosion; (2) compare the influence of different rainfall profiles under dry (stand-alone event) conditions and in sequence to understand the influence of antecedent moisture on runoff and erosion; and (3) directly compare results to past studies for wood mulch and rainfall profiles. Because large-scale mulching operations are expensive, there is an interest in developing more quantitative methods for determining the effective mulch application rate (Girona-García et al., 2023). The temporal variability of rainfall has been shown to be more important than time since fire, and incorporating this factor will likely improve estimates of the required mulch application (Ebel, 2020).

4.3 Methods

4.3.1 Rainfall simulator setup and soil

A schematic of the rainfall simulator setup is shown in Figure 4.1. The flume bed used for experiments is 92 cm wide and 278 cm long with a depth of 20 cm. Subsurface drainage flows freely through a grate covered by geotextile fabric in the bottom 40 cm of the flume and surface runoff is directed to a tapered funnel at the bottom of the flume to collect samples. The flume is connected to the frame by a center pivot, allowing for slope adjustment. The slope is set to 38%, a steep slope to represent mountainous conditions where mulch will be used. Similar rainfall

simulator setups have been used in several mulch effectiveness studies (Abrantes et al., 2018; De Lima et al., 2019; Prats et al., 2018, 2017; Montenegro et al., 2013).

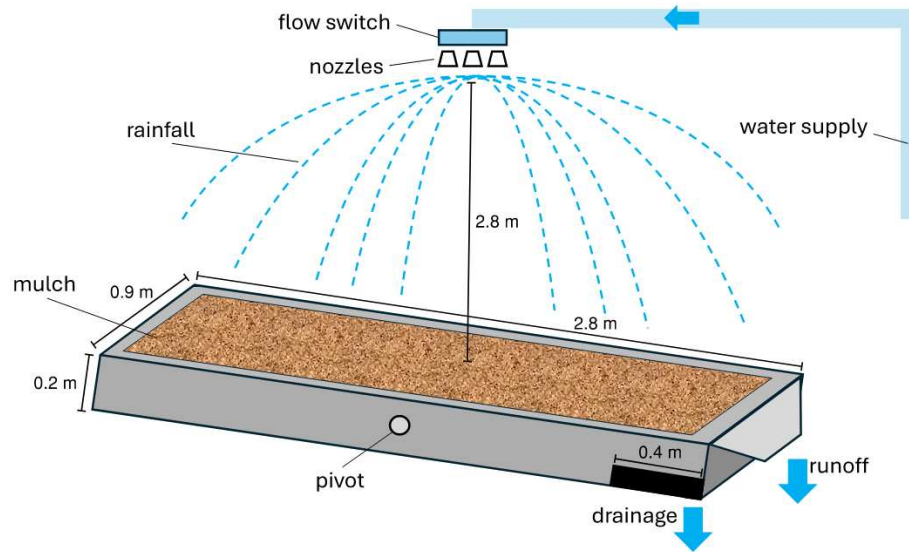


Figure 4.1: Setup of rainfall simulator and key components of flume.

The image in Figure 4.2 shows the simulator in the hydraulics laboratory at Colorado State University. Three nozzles (see section 2.3 for more detail on nozzles and rainfall) positioned downward above the center of the flume are connected to a three way hose splitter with a valve shutoff and flow regulator that sets the desired intensity. Nozzles are set to a height of 2.8 m from the center of the flume to expose the flume to the maximum possible spray coverage (reached at 2.1 m according to Spraying Systems Co.). The metal gutter shown in the image is used to collect samples. Nozzle performance was sensitive to particulates in the water supply, so a spin down sediment filter on the water supply line was required to maintain constant flow through the nozzles.

The flume was filled to a depth of 20 cm with clay loam topsoil sourced from a local landscape company. The geographic patterns of Colorado's soil are diverse, with many regions underlain by clay loam topsoil (Johnson and Cline, 1965). Table 4.1 summarizes the soil type, the soil texture was determined with a hydrometer and the organic matter content was determined by loss



Figure 4.2: Rainfall simulator in indoor laboratory.

on ignition. The soil was initially air dried and hand tamped to create a uniform density throughout the flume.

Table 4.1: Soil texture and organic content.

organic matter (%)	6.1
sand (%)	41
silt (%)	26
clay (%)	33

4.3.2 Mulch characteristics

The treated flume was covered with 410 g/m^2 of mulch which corresponds to a mean coverage 70% with a standard deviation of 2.7%. This is at the lower end of effective application rate for wood mulch according to Girona-García et al. (2021), but matches recommendations for cover rate percentage (Foltz and Wagenbrenner, 2010; Fan et al., 2023; Yanosek et al., 2006). The mulch is a very heterogeneous blend of shreds, pieces range in length from 1 mm to greater than 10 cm. The mixture was dried indoors for several months before use in experiments.

Mulch coverage was estimated using several image segmentation methods in scikit-image (van der Walt et al., 2014). Gaussian blurring was used to quantify colors in the image as well as lines and texture. Additionally, a limited set of labeled samples (11 examples of bare soil and 17 examples mulch from a single image) were used to train a random forest classifier to label each pixel as either mulch or background based on the input of local features from gaussian blurring. The method follows scikit-image developers (2025), which shows the use of an unsupervised method of image classification with minimal supervised learning. The classifier was applied to 15 pictures for each experiment, an example mask is shown in Figure 4.3. The mulch fraction is calculated by dividing the number of pixels classified as mulch by the total pixel count for the image.



Figure 4.3: Example mulch coverage mask.

4.3.3 Simulated rainfall

Variable rainfall profiles were achieved using three unique nozzles (relative intensity in parenthesis): 1/8GG-4.3W (low), 1/8GG-8W (medium), and 1/4GG-14W (high) from Spraying Systems Co. Target rainfall rates were achieved with the medium and high nozzles individually and by running the low-plus-high and medium-plus-high nozzle combinations simultaneously. Wide nozzles were chosen to achieve a high spatial uniformity, calculated with the Christiansen's uniformity coefficient (CU) (Christiansen, 1942). A pressure regulator was used to maintain the desired intensity, and pressure was calibrated to maximize CU. 68 kPa was used for the high nozzle, and 82 kPa was used for all other nozzle combinations. Table 4.2 shows the mean and standard deviation of the intensity and CU for the nozzle combinations used. The CU achieved here is similar to other studies using rainfall simulators (Salem and Meselhy, 2021; Prats et al., 2017). The mean intensity was estimated using 20-cm diameter plastic bowls that acted as gages. Each nozzle combination was tested twice with 36 gages spread evenly throughout the flume. Rainfall intensity and CU were calculated on a horizontal flume instead of the 40% slope used during the experiments, which could have affected the actual rainfall applied.

Table 4.2: Mean rainfall intensity and Christiansen uniformity (CU) for the different nozzle combinations with standard deviations shown in parentheses.

Nozzle	Intensity (mm hr ⁻¹)	CU (%)
medium	17.0 (2.6)	86.0 (0.3)
high	29.4 (3.5)	90.0 (0.5)
low & high	39.5 (6.5)	86.9 (0.1)
medium & high	59.2 (6.8)	90.2 (0.1)

Nozzles were selected to produce high intensity events with a 1-2 year average recurrence interval based on NOAA Atlas-14 values for western Colorado (Perica et al., 2013). These values are used to simulate convective events that are likely to occur and can also cause flash flooding and erosion after a wildfire. Intensities are generally lower than other rainfall simulation experiments that use unnaturally high rainfall intensities (Dunkerley, 2021). These profiles are similar to the design storm (24 mm/hr 15-minute intensity) used by the USGS to develop post-wildfire debris flow likelihood maps for the western US (U.S. Geological Survey, 2024).

The drop-size distributions for the various intensities were estimated using water sensitive paper. The paper was exposed to the given intensity for approximately 1 second. Droplet detection employed an RGB threshold derived by applying k-means clustering (Pedregosa et al., 2011) to pixel colors and then manually selecting the cluster corresponding to droplets. The scikit-image measure module was used to label droplets and calculate the area of each droplet (van der Walt et al., 2014). Figure 4.4 shows the distribution of droplets and D_{50} on the paper sample (colors show different paper samples, 2 samples were taken for each intensity value and 3 were taken for the highest intensity value). An example paper sample, from the 29 mm/hr intensity (sample 1), is shown in Figure 4.5. The D_{50} for all intensity values falls within light (0.9 +/- 0.3 mm) to moderate (1.1 +/- 0.4 mm) rain (Gatlin and Petersen, 2018). The samples are compared to the Marshall-Palmer distribution (Equation 1) to get a sense of how realistic the rainfall distributions are (Marshall and Palmer, 1948):

$$N(D) = N_0 e^{-\Lambda D}, \quad N_0 = 8\,000 \text{ m}^{-3} \text{ mm}^{-1}, \quad \Lambda = 4.1 R^{-0.21} \quad (1)$$

Where $N(D)$ is the drop-size distribution, i.e. the number of raindrops per cubic meter per millimeter diameter interval of size D ; $N_0 = 8.0 \times 10^3$ is the intercept parameter of the Marshall-Palmer distribution; $\Lambda = 4.1 R^{-0.21}$ is the slope parameter; and R denotes the rainfall rate.

The concentration of drops per unit volume was converted to a flux expected on the water sensitive paper by assuming a terminal fall speed relationship from Best (1950) for the sample of droplets (Equation 2 and 3). The nozzles produce a higher number of drops than expected according to the Marshall-Palmer distribution, although the D_{50} is within close range for all intensities. This discrepancy may be due to the paper being exposed for longer than 1 second. The nozzles also produce more of a bimodal distribution, but the skew towards lower diameter drops is somewhat realistic.

$$v(D) = 9.65 - 10.3 e^{-0.6D} \quad (2)$$

$$N_{\text{paper},i} = N(D_i) v(D_i) A \quad (3)$$

In Equation 2, $v(D)$ is the terminal fall speed of a raindrop of diameter D ; and in Equation 3, $N_{\text{paper},i}$ is the expected number of drops in diameter bin i striking the water-sensitive card; $N(D_i)$ is the drop-size distribution evaluated at the diameter D_i ; $v(D_i)$ is the corresponding fall speed from Equation 2; and A is the active collection area of the card.

The rainfall profiles tested included a uniform, decreasing, and increasing intensity event. Each profile delivered 9.9 mm of rain over 15 minutes, with the increasing and decreasing events being symmetrical. The uniform intensity was set to 39.5 mm/hr. The decreasing and increasing profiles were meant to loosely follow natural rainfall profiles for the mountains of CO. Figure 4.6 shows a comparison between the rainfall profiles used in the study and CoAgMET gages in higher terrain (above 2000 m) during summer months (June-August) (Colorado State University, Accessed 2023). The comparison is limited to events that share the same maximum intensity; this included

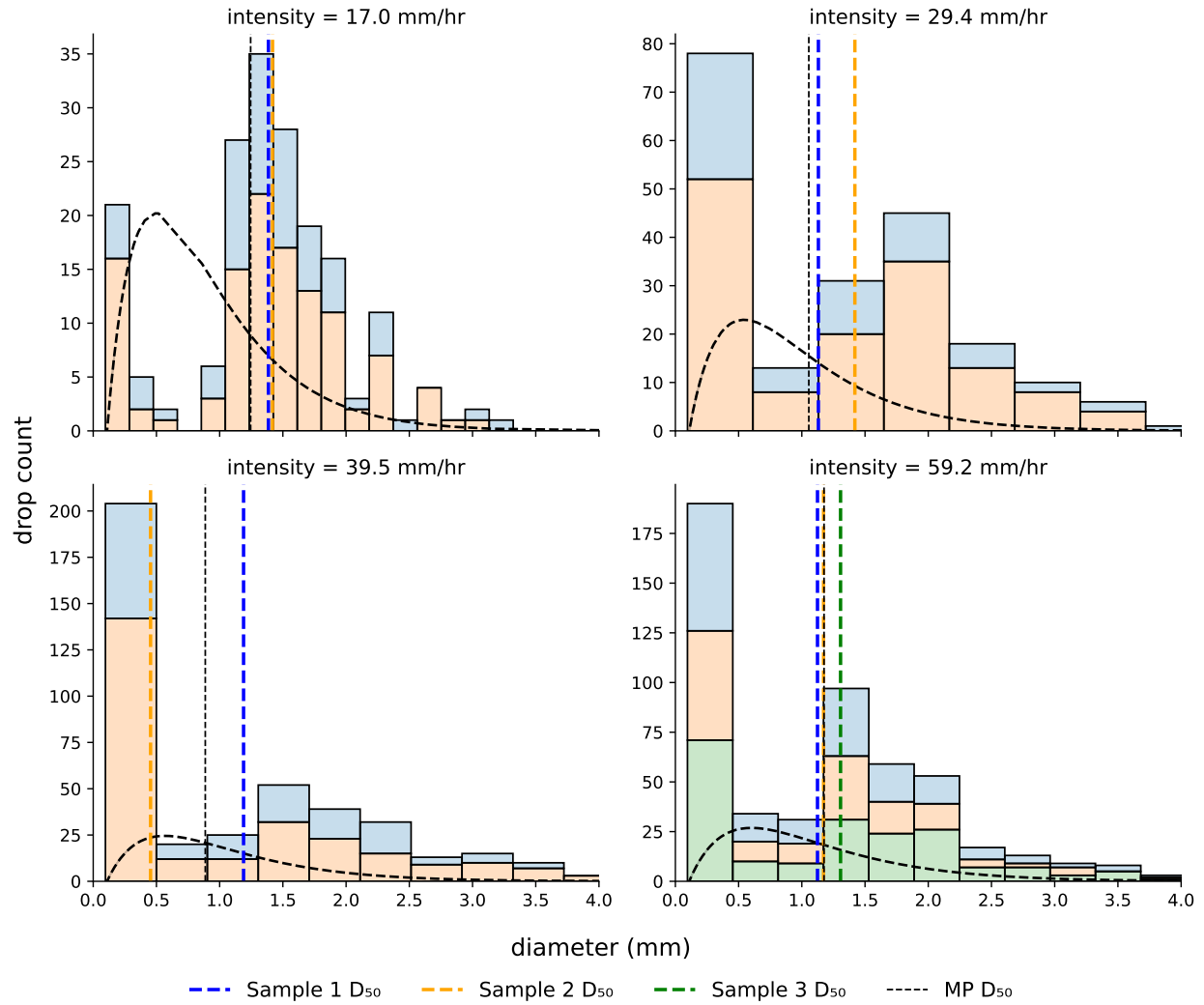


Figure 4.4: Drop size distribution of samples compared to the Marshall-Palmer distribution. The distribution represents the flux of drops through the water sensitive paper. Colors differentiate samples.

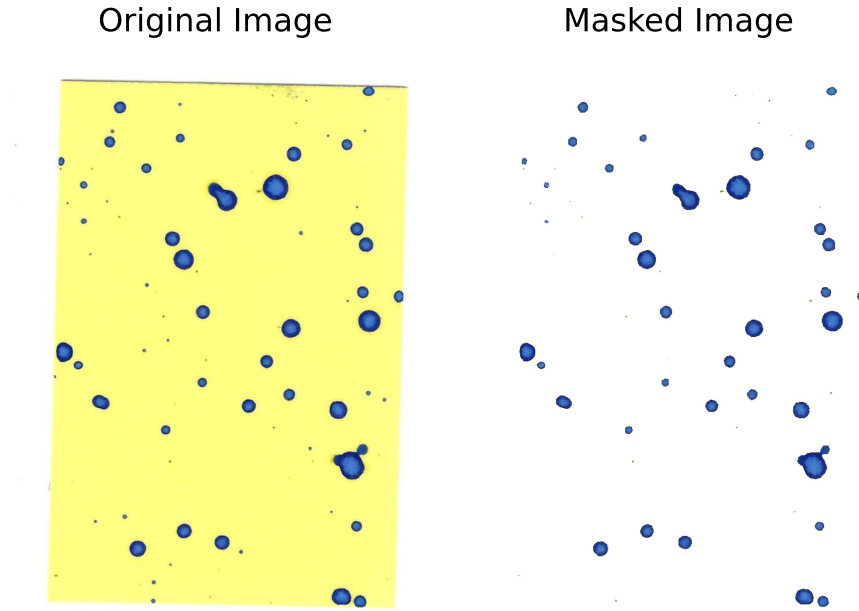


Figure 4.5: Example raindrop sample on the water sensitive paper for the 29.4 mm/hr intensity (sample 1) and the mask for drops.

8 gages for decreasing profiles and 5 gages for increasing profiles. The temporal resolution of CoAgMET gages is 5 minutes; so we are assuming temporal variability is self-similar.

4.3.4 Experimental procedure

The uniform, decreasing, and increasing rainfall profiles were tested both sequentially (in stated order) and individually with at least 24 hours to allow the soil to dry. The sequential events were tested with and without mulch with three replicates. The decreasing and increasing rainfall profiles were also tested with and without mulch, individually, with three replicates. A total of 30 rain events and 18 separate experiments were conducted. Rainfall events were grouped and tested separately to evaluate how antecedent moisture influences results and to compare to past studies that only test grouped profiles, such as Montenegro et al. (2013). During multiple event experiments the flume was left to dry for 45 minutes between events to allow for runoff recession.

The base soil was reused for the experiments. This approach follows the procedure outlined by De Lima et al. (2019) and Parhizkar et al. (2021). Soil moisture was monitored to start each exper-

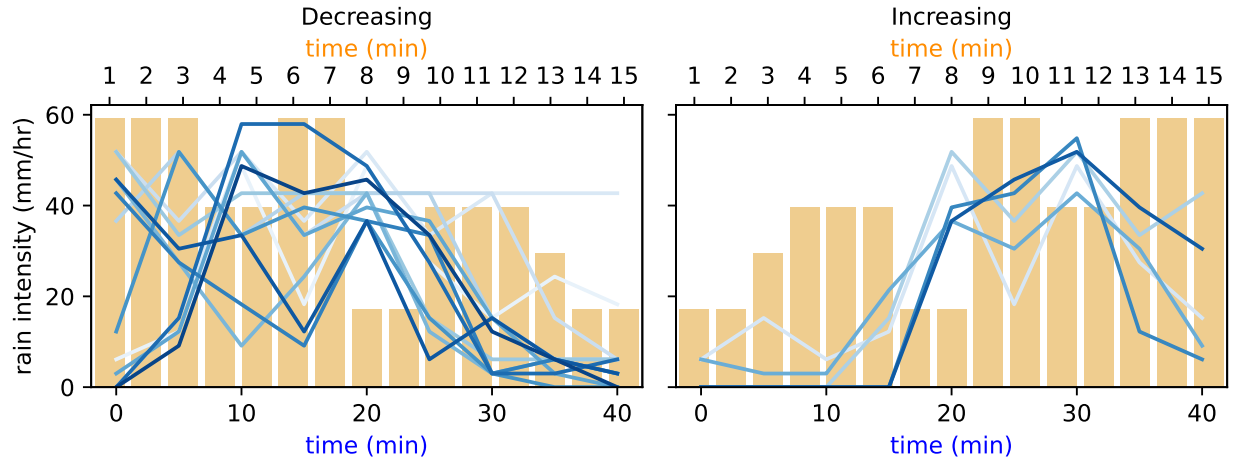


Figure 4.6: Rainfall profiles used in experiments are shown in orange bar plots. Blue lines show events from CoAgMET gages. A longer sampling time (40 minutes versus 15 minutes) was used for the gages because the time resolution is 5-minutes, so a second axis is shown.

iment with a consistent initial level using a TEROS 12 capacitance 3 prong probe. Soil moisture stayed relatively high, with a mean initial soil moisture of $0.313 \text{ m}^3/\text{m}^3$ (standard deviation was $0.014 \text{ m}^3/\text{m}^3$). The soil moisture of the flume was recorded in 14 locations before each experiment and had a mean CU of 94.8%.

Prior to each experiment, a thin layer of soil was removed (less than 1 cm) and replaced to ensure the distribution of fine sediment was consistent for each experiment. The new layer was mixed with the top layer of the remaining soil and the soil was screeded twice to create a flat plane and maintain a consistent density. To ensure the new layer had a consistent soil moisture the soil was misted with the low nozzle (9.8 mm/hr) for 15 minutes and left to drain for 24 hours prior to the next experiment. After each mulch experiment the wet mulch was removed (mulch was not reused).

Runoff was sampled at the end of every minute for 5 seconds. After weighing the sample bottles, samples were filtered with paper filters with a pore diameter of approximately 20 micrometers. Filters were dried for 24 hours at 105°C to weigh the sediment. A small amount of fines passed through the filters. Four samples that passed through the filter were fully dried. The mass of fines

in the filtered sample was 0.23 g with a standard deviation of 0.08 g, and it was assumed that the filter was still retaining the majority of fine material in the sample.

4.3.5 Statistical approach

ANOVA and pairwise post-hoc tests were used to evaluate the effectiveness of the mulch treatment for the runoff ratio and total erosion rate. Additionally, each complex profile was compared to the constant-rate profile for both mulched and bare runs as well as sequential and individual runs. The grouped and independent events were tested separately. The constant intensity runs were considered grouped and independent because they were conducted at the start of the grouped runs.

A two-way mixed model ANOVA with repeated measures was used to determine the significance of the effect of mulch (the between subject factor) and rainfall profile type (the within subject factor) on the runoff ratio and erosion rate for the grouped events. For the independent experiments, a two-way ANOVA was used with the mulch treatment and profile type as the main factors. The Shapiro-Wilk test was used to test the normality of the residuals of the linear model, the Mauchly test was used to test for sphericity for the grouped run, and the Levene test was used to assess homoscedasticity of the treatment groups (Vallat, 2018; Seabold and Perktold, 2010). The data passed all assumption tests, although the logarithm of the erosion rate was used to satisfy the homoscedasticity and normality assumptions. Where there is a significantly different group, a pairwise t-test is used to determine which groups are significantly different across mulch treatments and which rainfall profiles produce significantly different results from the uniform rain profile. All statistical tests were performed at a 0.05 significance level.

The order of the experiments and the initial soil moisture were both tested as covariates (Seabold and Perktold, 2010). Neither factor was significant in the linear model for runoff or erosion in the grouped or individual runs. The Pearson correlation coefficient between initial soil moisture and runoff and erosion was evaluated for the individual runs and was not significant.

4.4 Results

4.4.1 Sediment yield

A time series of the sediment concentration for all experiments is shown in Figure 4.7. Erosion generally follows the increases in intensity. The significant increase in concentration in the beginning of the experiment is likely caused by fine particles on the sample outlet tray that were removed with very small amounts of runoff. The individual decreasing event shows slightly higher concentration, compared to sequential runs, especially at the intensity peaks. The individual increasing event without mulch shows significantly higher erosion, compared to the sequential run, during the high-intensity peaks at the end of the event. Mulch appears to suppress sediment concentration for all experiments.

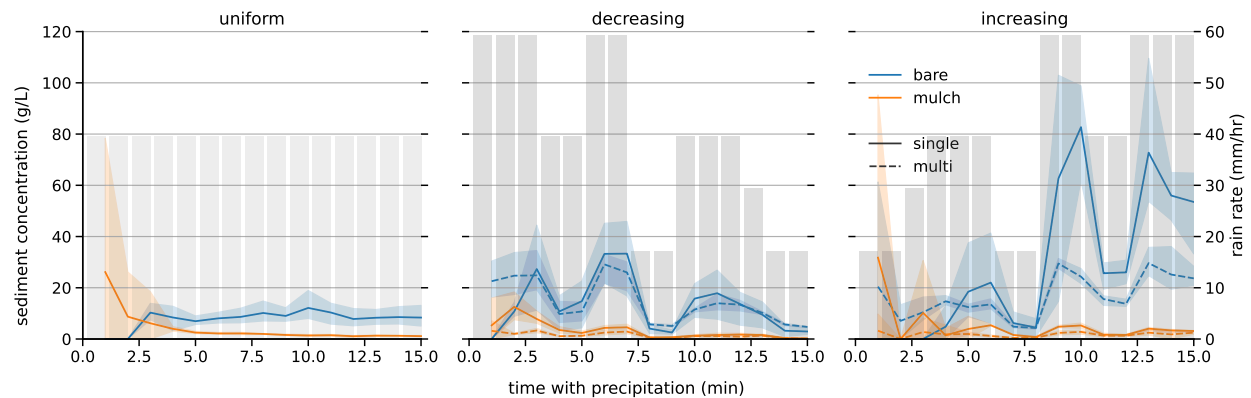


Figure 4.7: Time series of erosion concentration for all experiments. Mulched runs are plotted in orange and bare-soil runs in blue, with dashed lines indicating sequential runs and solid lines indicating individual runs. The shaded band represents the 95% confidence interval from the replicates.

Figure 4.8 gives an example of the rilling observed after a stand-alone increasing event, all of which developed similar rilling patterns. Only one stand-alone decreasing event showed rilling and one grouped run. No rills were observed for events with the mulch treatment.

Figure 4.9 shows the mean total erosion per unit area for all experiments, the error bars show the standard deviation for each group. Tables 4.3 and 4.4 show the individual experiment values with pairwise significance. Only individual runs had a significant F-score for the interaction between



Figure 4.8: Example rilling observed after an increasing intensity profile experiment.

mulch and rainfall profile. Mulch and rainfall profile were significant independent factors. Pairwise t-tests show a significant difference between all mulched rainfall profiles, showing that mulch does effectively reduce erosion. Grouped runs with the decreasing profile differed significantly from those with the constant profile, but the interaction term was non-significant, so the result is not considered meaningful. For both bare soil and mulched conditions, individual increasing-profile runs differed significantly from the constant-profile runs; this contrast is especially evident for the bare soil in Figure 4.9.

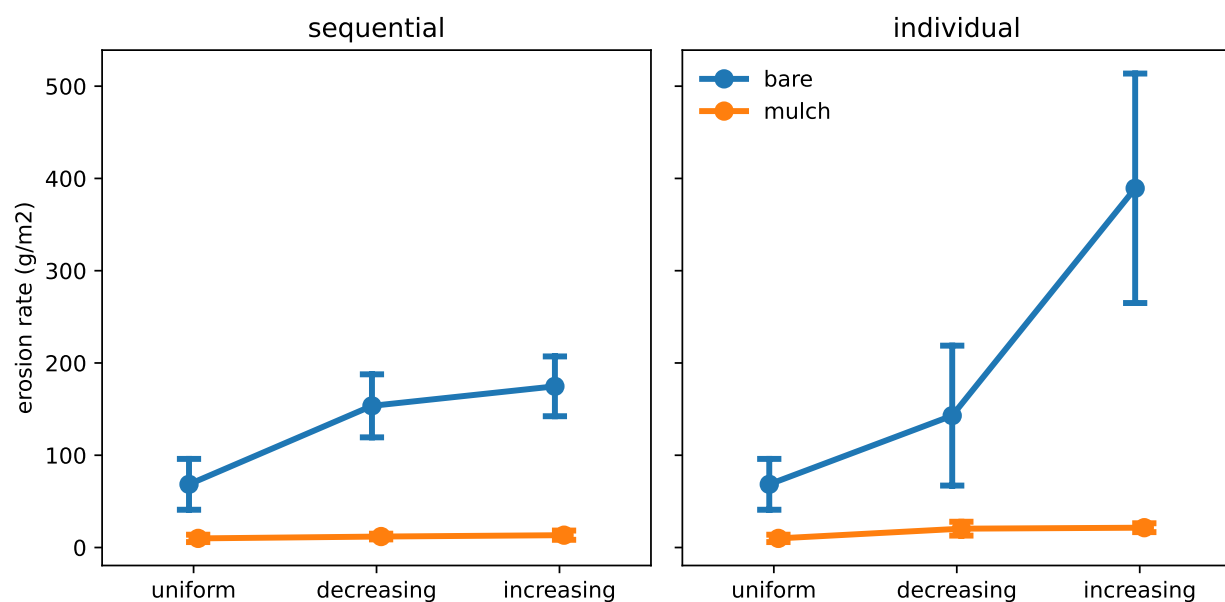


Figure 4.9: The distribution of total erosion per unit area for grouped (sequential) and individual runs. The dot represents the mean and the error bars represent the standard deviation.

4.4.2 Runoff response

Figure 4.10 shows the time series of runoff rate for all experiments. All mulched runs show a delay in time to runoff of approximately 1 minute. Starting the experiment with drier soil (individual runs) also adds an additional minute to the runoff time. Mulch appears to decrease the uniform intensity at a constant rate. Mulch under the complex profiles shows suppression of the peak runoff rates, but after a peak the runoff matches the bare soil rate. Mulch seems to delay runoff, perhaps

Table 4.3: Total erosion rate per unit area for grouped experiments. Rows 1–3 list individual replicates; the last two rows give the sample mean and SD across replicates. Bold mean values represent significant mulch treatments. Mean values are underlined where the complex profile was significantly different from the constant-rate run.

Replicate	Erosion rate (g/m^2)					
	Uniform		Decreasing		Increasing	
	No mulch	Mulch	No mulch	Mulch	No mulch	Mulch
1	99.1	7.5	190.3	11.6	155.8	14.1
2	45.6	14.5	147.4	15.1	212.1	18.0
3	60.8	7.7	122.9	8.9	156.1	8.0
Mean	68.5	9.9	<u>153.5</u>	11.9	174.7	13.4
SD	27.6	4.0	34.2	3.1	32.4	5.0

Table 4.4: Total erosion rate per unit area for individual runs. Rows 1–3 list individual replicates; the last two rows give the sample mean and SD across replicates. Bold mean values represent significant mulch treatments. Mean values are underlined where the complex profile was significantly different from the constant-rate run.

Replicate	Erosion rate (g/m^2)			
	Decreasing		Increasing	
	No mulch	Mulch	No mulch	Mulch
1	166.0	14.3	378.0	15.9
2	58.3	28.8	518.9	24.5
3	204.6	18.2	271.0	24.0
Mean	142.9	20.4	<u>389.3</u>	21.5
SD	75.8	7.5	124.4	4.8

by slowly releasing retained water or by lengthening the flow pathways. The decreasing rainfall profile extends the runoff delay in line with the added intensity, resulting in higher post-peak runoff rates under mulch than with bare soil. The time series for grouped versus individual events are similar after an initial period of saturation.

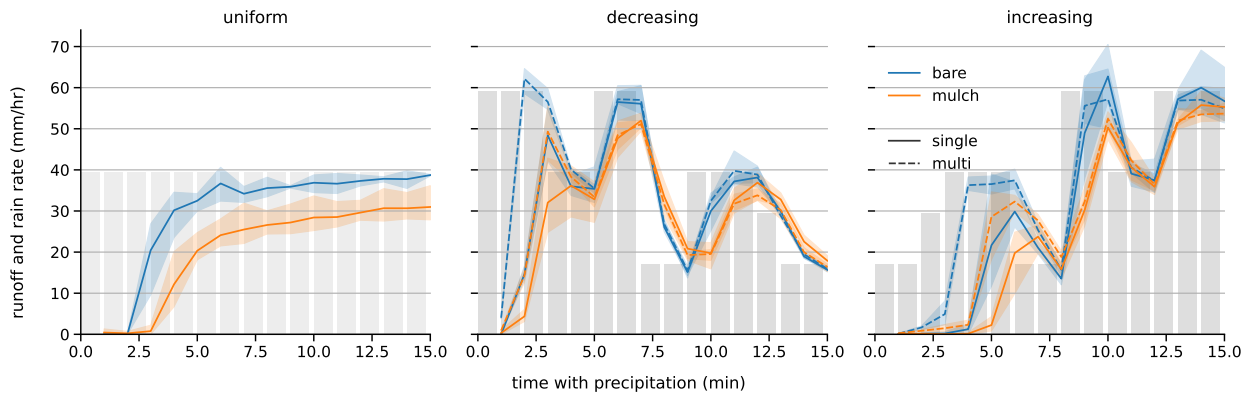


Figure 4.10: Time series of runoff rate for all experiments. Mulched runs are plotted in orange and bare-soil runs in blue, with dashed lines indicating sequential runs and solid lines indicating individual runs. The shaded band represents the 95% confidence interval from the replicates.

Figure 4.11 shows the mean runoff ratio for the different groups of experiments, the error bar shows the standard deviation of each group. Tables 4.5 and 4.6 show the individual experiment values with pairwise significance. While mulch and rainfall profile were significant factors for grouped runs, only mulch was significant for individual runs, and the interaction term was not significant for either. Pairwise comparison shows significant differences in runs with mulch for grouped runs, but not for individual runs, suggesting that under drier conditions mulching is not effective at reducing runoff during varying rainfall events. The individual mulched decreasing profile yields higher values than the constant-intensity and increasing profile runs, and the pairwise test confirms this difference as significant (although profile was not a significant factor based on ANOVA). The initial high intensity likely oversaturates the mulch, diminishing its effectiveness.

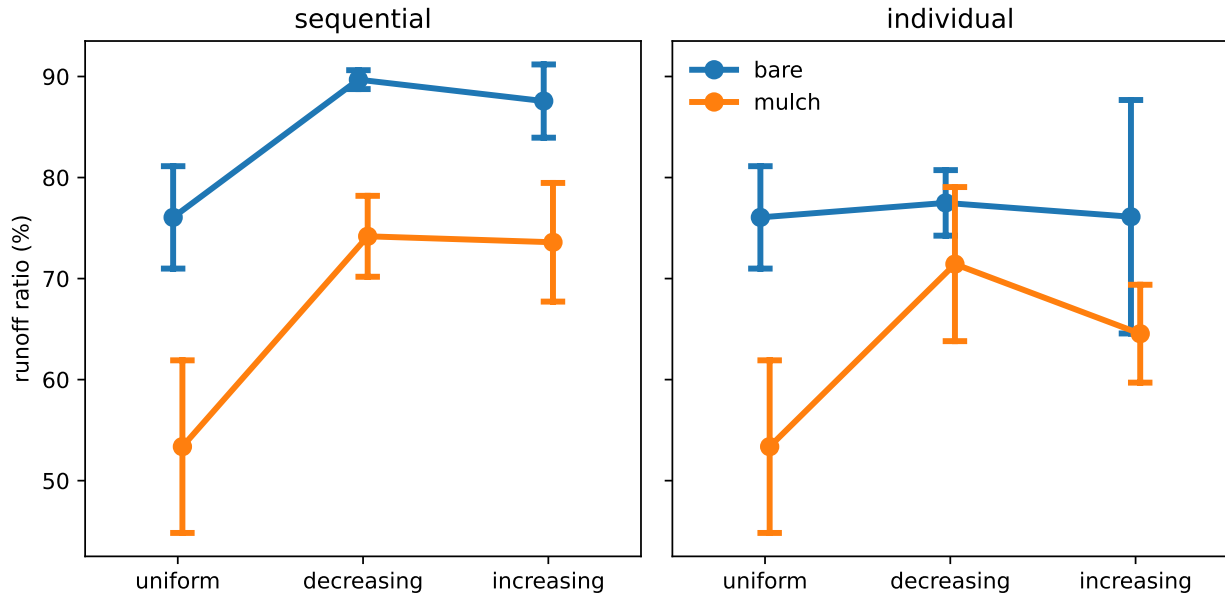


Figure 4.11: The distribution of runoff ratio for grouped (sequential) and individual runs. The dot represents the mean and the error bars represent the standard deviation.

4.5 Discussion

4.5.1 Effectiveness of mulch

Figure 4.9 shows significantly higher erosion for the stand-alone experiments using an increasing rainfall rate, and extensive rilling was also observed throughout the flume for this type of event (see Figure 4.8). Mulch was effective in reducing this significant increase in sediment yield, in addition to the other rainfall profiles. After mulching, both variable rainfall profiles still produced twice the amount of sediment as the constant-rate profile, although only the increasing mulched profile was significantly different from the constant rate. Complex profiles under drier conditions did not produce significantly different runoff totals from constant rainfall (see Figure 4.11); however, mulch was not effective in reducing the complex profile runoff ratio. Constant profiles appear to exaggerate the effectiveness of mulch in reducing runoff, as does grouping events together. The time series for runoff rate (Figure 4.10) shows that mulch does slightly reduce runoff rate peaks, which was also observed by Montenegro et al. (2013).

Table 4.5: Runoff ratio (%) for grouped experiments. Rows 1–3 list individual replicates; the last two rows give the sample mean and SD across replicates. Bold mean values represent significant mulch treatments. Mean values are underlined where the complex profile was significantly different from the constant-rate run.

Replicate	Runoff ratio (%)					
	Uniform		Decreasing		Increasing	
	No mulch	Mulch	No mulch	Mulch	No mulch	Mulch
1	79.7	48.3	90.4	69.8	87.3	67.0
2	70.3	63.2	90.0	77.7	91.3	78.3
3	78.2	48.6	88.6	75.1	84.1	75.4
Mean	76.1	53.4	<u>89.7</u>	74.2	87.6	73.6
SD	5.1	8.5	0.9	4.0	3.6	5.9

Table 4.6: Runoff ratio (%) for individual runs. Rows 1–3 list individual replicates; the last two rows give the sample mean and SD across replicates. Bold mean values represent significant mulch treatments. Mean values are underlined where the complex profile was significantly different from the constant-rate run.

Replicate	Runoff ratio (%)			
	Decreasing		Increasing	
	No mulch	Mulch	No mulch	Mulch
1	76.3	62.7	70.4	59.0
2	75.0	76.6	89.4	66.4
3	81.2	75.0	68.6	68.2
Mean	77.5	<u>71.4</u>	76.1	64.5
SD	3.2	7.6	11.6	4.8

Figure 4.12 compares the effectiveness of mulch in these experiments to past rainfall simulation studies that applied wood mulch. Points below the red dashed line show applications of mulch that reduced the runoff ratio or erosion rate by 50% or more. Only studies that stated the rainfall intensity, used bare soil as the control, and reported total amounts for runoff and erosion were included. Montenegro et al. (2013) was included to compare the effect of variable rainfall profile, although it is the only study included that used straw mulch. Montenegro et al. (2013) observed that the rainfall profile has no effect on total runoff (different clusters are caused by an increase in mulch cover; see Table 4.12 for a description of groups labeled in the scatterplot). Although Montenegro et al. (2013) observed less erosion for bare soils under complex profiles (likely because the complex

profiles were at the end of the experiment and there were fewer fine grains available to erode) they did observe nearly twice the amount of erosion in complex profiles, compared to constant intensity, when mulched.

In general, the comparison shows that the effectiveness of wood mulch in reducing total runoff is more variable than the effectiveness of mulch in reducing total erosion. Faucette et al. (2004) and G. D. Jennings and A. R. Jarrett (1985) show that higher cover percentage improves the reduction of total runoff. Only Dooley (2003), Foltz and Copeland (2009), and Foltz and Wagenbrenner (2010) show a greater than 50% reduction for runoff ratio. Intensity in these experiments is lower and all experiments were conducted with dry soil, which likely explains the high effectiveness (see Table 4.12). Foltz and Wagenbrenner (2010) exposed the flume to concentrated flow and found a decrease in mulch effectiveness for sediment reduction, and to some extent runoff, on steeper slopes. Foltz and Copeland (2009) also found a decrease in effectiveness under concentrated flow for runoff with coarser soil, although increasing the cover amount kept the effectiveness high.

Prats et al. (2017) show an unexplained increase in runoff with wood treatment. Prats et al. (2017) experimented with covering a fraction of the flume. For runoff, covering 2/3 or 3/3 of the flume was effective if the percentage covered was at least 70%. The coverage percent was less important than the fraction of the flume covered for soil erosion when exposed to concentrated flow. Most studies conclude that the cover percentage matters more than the material used (Cerdá and Doerr, 2005; Foltz and Copeland, 2009; Wagenbrenner et al., 2006; Yanosek et al., 2006). However, wood mulch is generally more dense than straw, making application more costly. Methods to decrease the mass required to reach higher coverage, such as applying mulch in strips (Prats et al., 2017) or removing small bits (Foltz and Wagenbrenner, 2010) are potentially viable methods to make wood mulch more cost effective.

Field study findings are similar to those from laboratory studies. Robichaud et al. (2013b) looked at mulch effectiveness in small plots within a burn scar and found sediment reduction for both post-fire years 0 and 1. Other studies have found that the effect of mulch is most significant immediately after the fire in severely burned sites (Girona-García et al., 2021; Vieira et al., 2018).

Meta-analysis reviews (that do not include rainfall variability) have more optimistic results for runoff reduction from mulch. Fan et al. (2023) found that wood and straw mulch is effective at reducing runoff (average reduction 47%) and erosion (average reduction 76%) and recommends 60% for wood and straw mulch, with the caveat that mulch is more effective on medium texture soils and slopes less than 60%. Girona-García et al. (2021) also found wood mulch to be effective at reducing runoff and erosion, although the effective application rate was highly variable (230-1750 g/m^2).

Complex rainfall patterns can reduce runoff to roughly the same extent as increasing slope steepness or adding more surface cover. Within a given study, runoff mainly responds to the cover percentage, whereas differences among studies are shaped by rainfall intensity, antecedent soil moisture, and infiltration capacity. A more rigorous quantitative synthesis is needed to isolate the relative influence of these factors both within and across experiments.

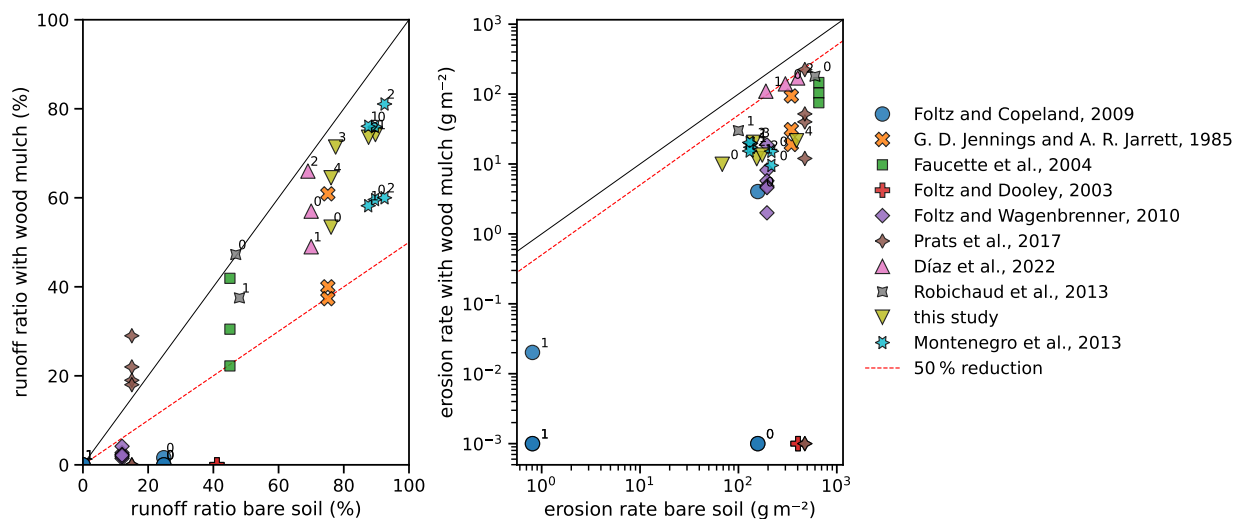


Figure 4.12: Runoff ratio and erosion rate from studies assessing the influence of wood mulch cover. Numbers adjacent to scatter points represent different experiment groups.

Table 4.7: Key parameters of wood mulch experiments used in this synthesis.

Reference	Setting	Group	Soil type	Max Dur (min) int (mm/hr)		Pre-wet?	Slope (%)	Cover (% or class)
Díaz et al. 2022	wood field	0	sandy loam	320.0	5	no	39.1	48.35
		1	sandy loam	320.0	5	no	30.1	48.35
		2	sandy loam	320.0	5	no	48.1	48.35
Faucette et al. 2004	wood lab	0	sandy clay loam	160.0	60	no	10.0	coarse, fine, med
Foltz & Copeland 2009	wood lab	0	sandy loam	50.0	15	no	30.0	30,50,70
		1	gravelly sand	50.0	15	no	30.0	30,50,70
Foltz & Dooley 2003	wood lab	0	gravelly sand	50.0	15	no	30.0	70
Foltz & Wagenbrenner 2010	wood lab	0	burned sandy loam	50.0	15	no	40.0	50,70
Jennings & Jarrett 1985	wood lab	0	silty clay loam	135.0	10	no	2.0	40,85,98
Montenegro et al. 2013	wood lab	0	sandy loam	112.0	15	yes	10.0	63.1,80.3
		1	sandy loam	112.0	25	yes	10.0	63.1,80.3
		2	sandy loam	112.0	25	yes	10.0	63.1,80.3
Prats et al. 2017	wood lab	0	sandy loam	56.0	20	no	40.0	1-3m strips at 50/70%
Robichaud et al. 2013	wood field	0	loam	80.0	20	no	44.0	71
		1	loam	80.0	20	no	44.0	73
<i>This study</i>	wood lab	0	clay loam	37.6	15	no	38.0	70
		1	clay loam	58.9	30	yes	38.0	70
		2	clay loam	58.9	45	yes	38.0	70
		3	clay loam	58.9	15	no	38.0	70
		4	clay loam	58.9	15	no	38.0	70

4.5.2 Influence of rainfall variability

This study shows that when complex profiles are run individually over bare soil, the difference in total runoff is not significant. Total erosion over bare soil increases significantly for increasing profiles. The influence of rainfall variability on runoff and erosion on bare soils is contextualized by past studies that experiment with increasing and decreasing intensity profiles compared to constant intensity.

Figure 4.13 shows that rainfall profile types appear to have a more variable influence on erosion, compared to total runoff. Individual runs of complex profiles in this study are most similar to the the D. C. Flanagan et al. (1988) dry run (see Table 4.8 for parameter values of groups labeled in the scatter plot), although they found significantly higher runoff peaks for increasing profiles. D. C. Flanagan et al. (1988) also ran the experiments wet and found decreasing profiles to cause significantly higher total and peak runoff amounts. Parsons and Stone (2006) tested complex profiles on three soil types, and found no difference for runoff between soils and rainfall profiles (values from this study are not plotted because they did not report total rainfall). Mohamadi and Kavian (2015) collected runoff from large field plots and determined the rainfall profile from a gage, and found no difference across profiles (results not plotted because they did not include constant rainfall as a control). Dunkerley (2012) found all complex profiles to have higher runoff than constant intensity. This unique result may have been caused by soil sealing and the lower rainfall intensity.

The impact of the rainfall profile on sediment yield is highly variable. D. C. Flanagan et al. (1988) shows sediment yield is higher for an increasing profile (same as this study) when dry, but under wet conditions the decreasing profile causes equal erosion compared to constant intensity and the increasing profile causes significantly lower erosion. Montenegro et al. (2013) also found less erosion on bare soil with complex profiles, likely because the fine grains had already been removed from previous events. Parsons and Stone (2006) found that the rainfall profile effects varied with soil type. Römken et al. (2002) found more erosion from a decreasing profile compared to increasing, and reasoned that lower initial intensities caused soil sealing which was more resistant

to erosion (not plotted because they did not include constant rainfall as a control). Mohamadi and Kavian (2015) found increasing profiles caused the highest erosion.

Rainfall profiles appear to influence the erosion response more than runoff, but their effect is modulated by intensity, slope, soil moisture, and soil type. Runoff differences may depend on whether the rain drop impacts cause soil sealing, the soil infiltrability, and the initial soil moisture.

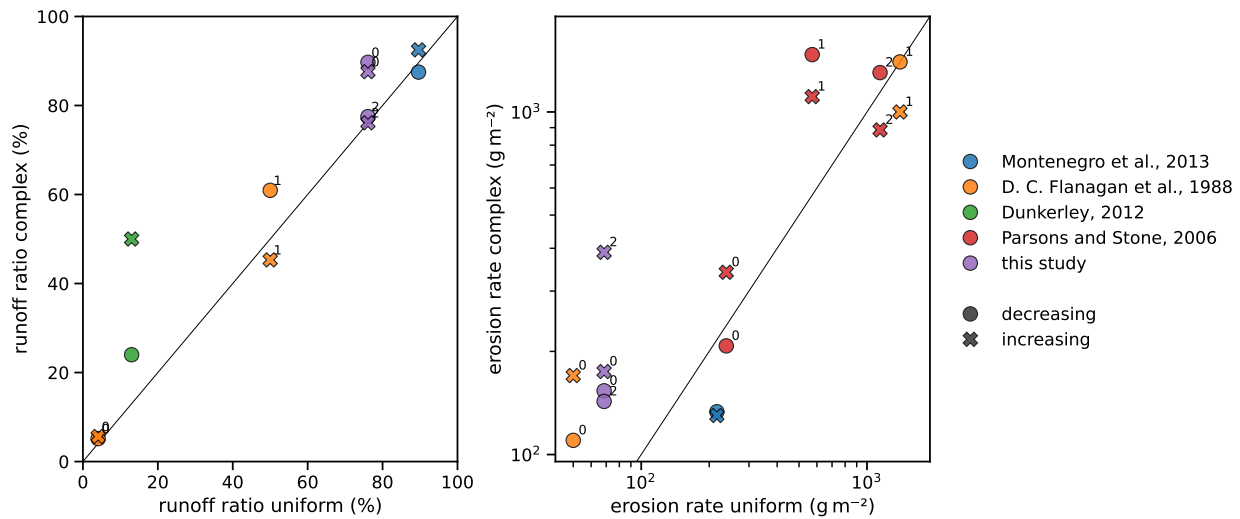


Figure 4.13: Runoff ratio and erosion rate from studies assessing the influence of variable rainfall intensity. Numbers adjacent to scatter points represent different experiment groups.

Table 4.8: Rain–profile experiments and their key parameters.

Reference	Group	Soil type	Max int (mm/hr)	Mean int (mm/hr)	Dur (min)	Pre-wet?	Slope (%)
D. C. Flanagan et al., 1988	0	silt loam	250.0	60.0	60	no	6.8
	1	silt loam	250.0	60.0	60	yes	6.8
Dunkerley, 2012	0	sandy loam	30.0	10.0	90	no	3.5
Montenegro et al., 2013	0	sandy loam	112.0	70.8	25	yes	10.0
Parsons & Stone, 2006	0	clay loam	170.8	93.9	30	no	17.6
	1	sandy loam	170.8	93.9	30	no	17.6
	2	sandy soil	170.8	93.9	30	no	17.6
<i>This study</i>	0	clay loam	58.9	37.6	45	yes	38.0
	2	clay loam	58.9	37.6	15	no	38.0

4.5.3 Implications for management

The impact of the rainfall profile on runoff and erosion is highly variable in our experiments, shifting with both mulch presence and antecedent soil moisture. A 70% cover was not effective in reducing the runoff ratio for varying intensity events under dry conditions. With other soil types and moisture levels, the mulch cover may have also failed to significantly reduce erosion. Several field studies have made broad recommendations for the amount of mulch cover necessary to reduce runoff and erosion (Girona-García et al., 2021; Fan et al., 2023), and empirical models have attempted to quantify a more precise amount of cover required given particular environmental conditions (Schmeer et al., 2018). The performance of these models would likely benefit from considering the effect of rainfall variability.

A relevant consideration in modeling rainfall variability is the frequency of particular profile sequences in a given geographic area. An increasing profile causes more erosion, based on the results of this study, but it is less common for the Colorado mountains that motivate this work. The CoAgMET gages were also used to explore the percentage of early (decreasing) versus late (increasing) peak events for this area. Events were separated by dry periods of 6 hours or greater. This included 26 gages, above 2000 m during summer months. Early events are defined as those where 50% of the total accumulation occurs in the 1st quarter of the event duration, whereas during late events 50% of the total accumulation occurs in the last quarter of the event duration. The mean frequency of 1st quantile events was 0.62 with a standard deviation of 0.05. The frequency of 4th quantile events was much lower, with a mean frequency of 0.09 with a standard deviation of 0.03. Including a quantile descriptor of rainfall events could be included in empirical models. Collar et al. (2025) presents a method to characterize rainfall regime for a region in terms of magnitude and decay rate that captures duration and intermittent intensity.

4.6 Conclusion

The effectiveness of wood mulch was evaluated on a steep slope under variable rainfall. Total runoff was not significantly different across rainfall profiles for bare soil. Mulch was not effective

in reducing total runoff for complex, more natural, rainfall profiles. Erosion was significantly different on bare soils for an increasing event, but mulch was effective at reducing erosion for all run types. Experiments were also conducted in sequence, and show different results. Comparison with other wood mulch simulations shows contrasting conclusions for runoff (depending on initial soil moisture, rainfall intensity, slope, and soil type) but all studies show a significant reduction in erosion. Other rainfall simulations on bare soil show that rainfall variability does not influence rainfall totals, but generally does significantly impact erosion (compared to constant intensity). The experiments from this study and comparisons with past studies show that the impact of rainfall profiles is highly variable, with non-uniform rain often causing significant differences in runoff and sediment yields relative to constant rainfall. The simple setup described in this study produces somewhat natural rainfall, is relatively inexpensive, and can easily be included in future rainfall simulations to account for these differences.

The effectiveness of mulch under variable rainfall should be explored further with additional initial moisture conditions and soils with different erodibility and infiltrability. A non-burnt soil was used here as bare soil has been shown to produce similar runoff and erosion to burnt soil after the ash layer has eroded (Benavides-Solorio and MacDonald, 2005; Larsen et al., 2009). Erosion and runoff on surfaces with an ash layer will likely react differently to rainfall variability based on the additional absorptive capacity of the ash layer (Robichaud et al., 2013a; Ebel et al., 2012; Bodí et al., 2012).

The influence of rainfall variability with concentrated flow has not been studied, although it clearly impacts mulch effectiveness under constant rainfall intensity (Prats et al., 2017; Foltz and Wagenbrenner, 2010). The direction of storm movement matters in a watershed (Syed et al., 2003) and even in a flume (de Lima and Singh, 2003). Many additional complications exist outside the flume or plot scale. Several studies have accounted for scaling differences when using a rainfall simulator (Smets et al., 2008; Poesen et al., 1999; Sadeghi et al., 2015; Prats et al., 2016). Including these factors, in addition to rainfall variability, would likely improve the modeling of mulch effectiveness.

Organic cover has been shown to be a very effective, albeit costly, tool for mitigating the effects of wildfires. When the effective dose is applied, the ecological benefits and erosion prevention make it cost effective. Determining the coverage required for effectiveness is challenging, as even in a controlled lab or small plot there is considerable variability in the erosive and runoff response caused by soil moisture, soil type, slope, and cover amount. This study shows that the rainfall variability is also a significant factor, and can easily be included in rainfall simulations to better understand the effectiveness of mulch under realistic conditions.

Chapter 5

Conclusion

Dynamic changes to precipitation (Fowler et al., 2021) and surface processes (Sharma et al., 2018), combined with the compounding risks caused by wildfires (Touma et al., 2022) highlight the importance of accurately understanding extreme precipitation patterns. Simultaneously, advancements in observations, remote sensing technologies, and modeling approaches have created new opportunities to do so (Lundquist et al., 2019). This dissertation explores the potential for these precipitation data sources to improve our understanding of extreme precipitation events in mountainous regions and the associated hazards.

In Chapter 2, I introduce a spatiotemporally dynamic error model for sub-hourly radar QPEs developed specifically for the mountains of Colorado. The model uses attributes of the region, rainfall, and storm events to diagnose the factors influencing QPE performance and explain why it performs as it does under different conditions. The accuracy of radar QPEs varies significantly, particularly in remote areas with complex terrain where they are often considered unreliable. While high radar beam heights are typically associated with limited coverage, radar can still be valuable in certain conditions—for instance, longer duration precipitation events tend to exhibit lower QPE errors.

In Chapter 3, A recently developed high resolution convection permitting reanalysis dataset (CONUS404) is used to re-evaluate the orographic impact on annual maximum subdaily precipitation intensity throughout Colorado. Several remotely-sensed (via radar and lightning sensors) datasets are also used to explore the influence of terrain on the magnitude of subdaily precipitation intensity throughout Colorado. Results show that in the Missouri and Arkansas river basins, there is a clear decrease in maximum annual hourly accumulation with elevation, especially during warm months. Elevation alone does not have a significant effect on daily accumulations in these regions, although specific terrain features do. The Colorado and Rio Grande river basins experience increasing maximum annual accumulation with elevation that is enhanced with longer

(daily) time windows. With few exceptions, the elevation–accumulation relationship becomes more positive with a longer accumulation window. Model-based precipitation appears to capture more terrain-driven spatiotemporal variability in precipitation intensity than gage-based precipitation. Precipitation frequency studies miss the high-accumulation clusters along the Front Range and Pikes Peak seen in CONUS404 daily data, likely due to sparse gages and terrain complexity.

In Chapter 4, The effectiveness of wood mulch was evaluated on a steep slope under variable rainfall. Total runoff was not significantly different across rainfall profiles for bare soil. However, mulch was not effective in reducing total runoff for complex, more natural, rainfall profiles. Erosion was significantly different on bare soils for an increasing event, but mulch was effective at reducing erosion for all run types. Experiments were also conducted in sequence, and show different results. Comparison with other wood mulch simulations shows contrasting conclusions for runoff (depending on the initial soil moisture, rainfall intensity, slope, and soil type) but all studies show a significant reduction in erosion. Other rainfall simulations on bare soil show that rainfall variability does not influence rainfall totals, but generally does significantly impact erosion (compared to constant intensity). The experiments from this study and comparisons with past studies show that the impact of rainfall profiles is highly variable, with non-uniform rain often causing significant differences in runoff and sediment yields relative to constant rainfall. Experiments that use constant rainfall exaggerate effectiveness of mulch cover. The effectiveness of mulch under variable rainfall should be explored further with additional initial moisture conditions and soils with different erodibility and infiltrability.

Bibliography

- João R.C.B. Abrantes, Sérgio A. Prats, J. Jacob Keizer, and João L.M.P. De Lima. Effectiveness of the application of rice straw mulching strips in reducing runoff and soil loss: Laboratory soil flume experiments under simulated rainfall. *Soil and Tillage Research*, 180:238–249, August 2018. ISSN 01671987. doi: 10.1016/j.still.2018.03.015. URL <https://linkinghub.elsevier.com/retrieve/pii/S0167198718302423>.
- P. Adhikari, B. Geerts, S. Rahimi-Esfarjani, K. Smith, B. N. Shuman, and T. L. Schneider. Evaluation of the mountain hydroclimate across the western united states in dynamically downscaled climate models. *Journal of Hydrometeorology*, 25(12):1877–1894, 2024. doi: 10.1175/JHM-D-24-0063.1.
- P. Allamano, P. Claps, F. Laio, and C. Thea. A data-based assessment of the dependence of short-duration precipitation on elevation. *Physics and Chemistry of the Earth, Parts A/B/C*, 34(10): 635–641, January 2009. ISSN 1474-7065. doi: 10.1016/j.pce.2009.01.001. URL <https://www.sciencedirect.com/science/article/pii/S1474706509000035>.
- Alison P. Appling, Samantha K. Oliver, Jordan S. Read, Jeffrey M. Sadler, and Jacob A. Zwart. Machine Learning for Understanding Inland Water Quantity, Quality, and Ecology. In *Encyclopedia of Inland Waters*, pages 585–606. Elsevier, 2022. ISBN 978-0-12-822041-2. doi: 10.1016/B978-0-12-819166-8.00121-3. URL <https://linkinghub.elsevier.com/retrieve/pii/B9780128191668001213>.
- Francesco Avanzi, Carlo De Michele, Salvatore Gabriele, Antonio Ghezzi, and Renzo Rosso. Orographic Signature on Extreme Precipitation of Short Durations. *Journal of Hydrometeorology*, 16(1):278–294, February 2015. ISSN 1525-755X, 1525-7541. doi: 10.1175/JHM-D-14-0063.1. URL <http://journals.ametsoc.org/doi/10.1175/JHM-D-14-0063.1>.

- R. M. Banta and C. B. Schaaf. Thunderstorm genesis zones in the colorado rocky mountains as determined by traceback of geosynchronous satellite images. *Monthly Weather Review*, 115: 463–476, 1987.
- Juan de Dios Benavides-Solorio and Lee H. MacDonald. Measurement and prediction of post-fire erosion at the hillslope scale, Colorado Front Range. *International Journal of Wildland Fire*, 14 (4):457, 2005. ISSN 1049-8001. doi: 10.1071/WF05042. URL <http://www.publish.csiro.au/?paper=WF05042>.
- A. Berne and W.F. Krajewski. Radar for hydrology: Unfulfilled promise or unrecognized potential? *Advances in Water Resources*, 51:357–366, January 2013. ISSN 03091708. doi: 10.1016/j.advwatres.2012.05.005. URL <https://linkinghub.elsevier.com/retrieve/pii/S0309170812001157>.
- A. C. Best. The size distribution of raindrops. *Quarterly Journal of the Royal Meteorological Society*, 76(327):16–36, 1950. doi: 10.1002/qj.49707632704.
- Merche B. Bodí, Stefan H. Doerr, Artemi Cerdà, and Jorge Mataix-Solera. Hydrological effects of a layer of vegetation ash on underlying wettable and water repellent soil. *Geoderma*, 191: 14–23, December 2012. ISSN 00167061. doi: 10.1016/j.geoderma.2012.01.006. URL <https://linkinghub.elsevier.com/retrieve/pii/S0016706112000286>.
- Rebecca Bolinger, Jeff Lukas, Russ Schumacher, and Peter Goble. Climate change in Colorado. 2023. doi: 10.25675/10217/237323. URL <https://mountainscholar.org/items/99896af1-0564-4531-9628-be1e13dbc4cd>. Medium: born digital,reports Publisher: [object Object].
- Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.
- Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- Leo Breiman, JH Friedman, RA Olshen, and CJ Stone. Classification and regression trees. *Wadsworth, Belmont, CA*, 37(15):237–251, 1984.

- Janice L. Bytheway. Demonstrating a Probabilistic Quantitative Precipitation Estimate for Evaluating Precipitation Forecasts in Complex Terrain. *Weather and Forecasting*, 37(1):45–64, January 2022. ISSN 0882-8156, 1520-0434. doi: 10.1175/WAF-D-21-0074.1. URL <https://journals.ametsoc.org/view/journals/wefo/37/1/WAF-D-21-0074.1.xml>.
- Janice L. Bytheway, Mimi Hughes, Kelly Mahoney, and Robert Cifelli. A Multiscale Evaluation of Multisensor Quantitative Precipitation Estimates in the Russian River Basin. *Journal of Hydrometeorology*, 20(3):447–466, March 2019. ISSN 1525-755X, 1525-7541. doi: 10.1175/JHM-D-18-0142.1. URL <http://journals.ametsoc.org/doi/10.1175/JHM-D-18-0142.1>.
- Janice L. Bytheway, Mimi Hughes, Kelly Mahoney, and Rob Cifelli. On the Uncertainty of High-Resolution Hourly Quantitative Precipitation Estimates in California. *Journal of Hydrometeorology*, 21(5):865–879, May 2020. ISSN 1525-755X, 1525-7541. doi: 10.1175/JHM-D-19-0160.1. URL <https://journals.ametsoc.org/view/journals/hydr/21/5/jhm-d-19-0160.1.xml>.
- Susan H. Cannon, Joseph E. Gartner, Raymond C. Wilson, James C. Bowers, and Jayme L. Laber. Storm rainfall conditions for floods and debris flows from recently burned areas in southwestern Colorado and southern California. *Geomorphology*, 96(3-4):250–269, April 2008. ISSN 0169555X. doi: 10.1016/j.geomorph.2007.03.019. URL <https://linkinghub.elsevier.com/retrieve/pii/S0169555X07001778>.
- Artemi Cerdá and Stefan H. Doerr. Influence of vegetation recovery on soil hydrology and erodibility following fire: an 11-year investigation. *International Journal of Wildland Fire*, 14(4):423, 2005. ISSN 1049-8001. doi: 10.1071/WF05044. URL <http://www.publish.csiro.au/?paper=WF05044>.
- J.E. Christiansen. *Irrigation by Sprinkling*. California Agricultural Experiment Station Bulletin 670. University of California, Berkeley, CA, USA, 1942.
- Grzegorz J. Ciach and Witold F. Krajewski. On the estimation of radar rainfall error variance. *Advances in Water Resources*, 22(6):585–595, February 1999. ISSN 03091708.

doi: 10.1016/S0309-1708(98)00043-8. URL <https://linkinghub.elsevier.com/retrieve/pii/S0309170898000438>.

Grzegorz J. Ciach, Witold F. Krajewski, and Gabriele Villarini. Product-Error-Driven Uncertainty Model for Probabilistic Quantitative Precipitation Estimation with NEXRAD Data. *Journal of Hydrometeorology*, 8(6):1325–1347, December 2007. ISSN 1525-7541, 1525-755X. doi: 10.1175/2007JHM814.1. URL <http://journals.ametsoc.org/doi/10.1175/2007JHM814.1>.

Natalie M. Collar, John A. Moody, and Brian A. Ebel. Rainfall as a driver of post-wildfire flooding and debris flows: A review and synthesis. *Earth-Science Reviews*, 260:104990, January 2025. ISSN 00128252. doi: 10.1016/j.earscirev.2024.104990. URL <https://linkinghub.elsevier.com/retrieve/pii/S0012825224003180>.

CSU Colorado State University. Colorado agricultural meteorological network, Accessed 2023. URL <https://coagmet.colostate.edu/>.

S. Copeland, N. S. Sharratt, B. Q. Wu, J. B. Foltz, R. and H. Dooley, J. A wood-strand material for wind erosion control: Effects on total sediment loss, PM₁₀ vertical flux, and PM₁₀ loss. *Journal of Environmental Quality*, 38(1):139–148, 2009. doi: 10.2134/jeq2008.0115.

Thomas M. Cover and Peter E. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

D. C. Flanagan, G. R. Foster, and W. C. Moldenhauer. Storm Pattern Effect on Infiltration, Runoff, and Erosion. *Transactions of the ASAE*, 31(2):0414–0420, 1988. ISSN 2151-0059. doi: 10.13031/2013.30724. URL <http://elibrary.asabe.org/abstract.asp??JID=3&AID=30724&CID=t1988&v=31&i=2&T=1>.

E. Dallan, F. Marra, G. Fosser, M. Marani, G. Formetta, C. Schär, and M. Borga. How well does a convection-permitting regional climate model represent the reverse orographic effect of extreme hourly precipitation? *Hydrology and Earth System Sciences*, 27(5):1133–1149, 2023. doi: 10.5194/hess-27-1133-2023.

- Christopher Daly, Michael Halbleib, Joseph I. Smith, Wayne P. Gibson, Matthew K. Doggett, George H. Taylor, Jan Curtis, and Phillip P. Pasteris. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, 28(15):2031–2064, December 2008. ISSN 0899-8418, 1097-0088. doi: 10.1002/joc.1688. URL <https://rmets.onlinelibrary.wiley.com/doi/10.1002/joc.1688>.
- Christopher Daly, Melissa K. Doggett, Joseph I. Smith, Kimberly V. Olson, Mark D. Halbleib, Zoran Dimcovic, Daniel Keon, Ronald A. Loiselle, Brian Steinberg, Andrew D. Ryan, Carla M. Pancake, and Erica M. Kaspar. Challenges in observation-based mapping of daily precipitation across the conterminous united states. *Journal of Atmospheric and Oceanic Technology*, 38(11): 1979–1992, 2021. doi: 10.1175/JTECH-D-21-0054.1.
- J. L. M. P. de Lima and V. P. Singh. Laboratory experiments on the influence of storm movement on overland flow. *Physics and Chemistry of the Earth, Parts A/B/C*, 28(6):277–282, January 2003. ISSN 1474-7065. doi: 10.1016/S1474-7065(03)00038-X. URL <https://www.sciencedirect.com/science/article/pii/S147470650300038X>.
- João L. M. P. De Lima, Lara Santos, Babar Mujtaba, and M. Isabel P. De Lima. Laboratory assessment of the influence of rice straw mulch size on soil loss. *Advances in Geosciences*, 48: 11–18, April 2019. ISSN 1680-7359. doi: 10.5194/adgeo-48-11-2019. URL <https://adgeo.copernicus.org/articles/48/11/2019/>.
- L.F DeBano. The role of fire and soil heating on water repellency in wildland environments: a review. *Journal of Hydrology*, 231-232:195–206, May 2000. ISSN 00221694. doi: 10.1016/S0022-1694(00)00194-3. URL <https://linkinghub.elsevier.com/retrieve/pii/S0022169400001943>.
- Yagmur Derin, Emmanouil Anagnostou, Alexis Berne, Marco Borga, Brice Boudevillain, Wouter Buytaert, Che-Hao Chang, Guy Delrieu, Yang Hong, Yung Chia Hsu, Waldo Lavado-Casimiro, Bastian Manz, Semu Moges, Efthymios I. Nikolopoulos, Dejene Sahlu, Franco Salerno, Juan-Pablo Rodríguez-Sánchez, Humberto J. Vergara, and Koray K. Yilmaz. Multiregional Satellite

Precipitation Products Evaluation over Complex Terrain. *Journal of Hydrometeorology*, 17(6): 1817–1836, June 2016. ISSN 1525-755X, 1525-7541. doi: 10.1175/JHM-D-15-0197.1. URL <http://journals.ametsoc.org/doi/10.1175/JHM-D-15-0197.1>.

Foltz and Dooley. COMPARISON OF EROSION REDUCTION BETWEEN WOOD STRANDS AND AGRICULTURAL STRAW. *Transactions of the ASAE*, 46(5), 2003. ISSN 2151-0059. doi: 10.13031/2013.15450. URL <http://elibrary.asabe.org/abstract.asp??JID=3&AID=15450&CID=t2003&v=46&i=5&T=1>.

Charles A. Doswell, Harold E. Brooks, and Robert A. Maddox. Flash Flood Forecasting: An Ingredients-Based Methodology. *Weather and Forecasting*, 11(4):560–581, December 1996. ISSN 0882-8156, 1520-0434. doi: 10.1175/1520-0434(1996)011<0560:FFFAIB>2.0.CO;2. URL [http://journals.ametsoc.org/doi/10.1175/1520-0434\(1996\)011<0560:FFFAIB>2.0.CO;2](http://journals.ametsoc.org/doi/10.1175/1520-0434(1996)011<0560:FFFAIB>2.0.CO;2).

David Dunkerley. Effects of rainfall intensity fluctuations on infiltration and runoff: rainfall simulation on dryland soils, Fowlers Gap, Australia. *Hydrological Processes*, 26(15):2211–2224, July 2012. ISSN 0885-6087, 1099-1085. doi: 10.1002/hyp.8317. URL <https://onlinelibrary.wiley.com/doi/10.1002/hyp.8317>.

David Dunkerley. The importance of incorporating rain intensity profiles in rainfall simulation studies of infiltration, runoff production, soil erosion, and related landsurface processes. *Journal of Hydrology*, 603:126834, December 2021. ISSN 00221694. doi: 10.1016/j.jhydrol.2021.126834. URL <https://linkinghub.elsevier.com/retrieve/pii/S0022169421008842>.

David L. Dunkerley. Rainfall intensity bursts and the erosion of soils: an analysis highlighting the need for high temporal resolution rainfall data for research under current and future climates. *Earth Surface Dynamics*, 7(2):345–360, April 2019. ISSN 2196-6311. doi: 10.5194/esurf-7-345-2019. URL <https://esurf.copernicus.org/articles/7/345/2019/>. Publisher: Copernicus GmbH.

- Brian A. Ebel. 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, June 2020. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2020.124765. URL <https://www.sciencedirect.com/science/article/pii/S0022169420302250>.
- Brian A. Ebel, John A. Moody, and Deborah A. Martin. Hydrologic conditions controlling runoff generation immediately after wildfire. *Water Resources Research*, 48(3), 2012. ISSN 1944-7973. doi: 10.1029/2011WR011470. URL <https://onlinelibrary.wiley.com/doi/abs/10.1029/2011WR011470>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2011WR011470>.
- Brian A. Ebel, John A. Moody, and Deborah A. Martin. Post-fire temporal trends in soil-physical and -hydraulic properties and simulated runoff generation: Insights from different burn severities in the 2013 Black Forest Fire, CO, USA. *Science of The Total Environment*, 802:149847, January 2022. ISSN 00489697. doi: 10.1016/j.scitotenv.2021.149847. URL <https://linkinghub.elsevier.com/retrieve/pii/S0048969721049226>.
- Brian A. Ebel, Zachary M. Shephard, Michelle A. Walvoord, Sheila F. Murphy, Trevor F. Partridge, and Kim S. Perkins. Modeling Post-Wildfire Hydrologic Response: Review and Future Directions for Applications of Physically Based Distributed Simulation. *Earth's Future*, 11(2), February 2023. ISSN 2328-4277, 2328-4277. doi: 10.1029/2022ef003038. URL <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022EF003038>. Publisher: American Geophysical Union (AGU).
- J.F. England, J.E. Godaire, R.E. Klinger, T.R. Bauer, and P.Y. Julien. Paleohydrologic bounds and extreme flood frequency of the Upper Arkansas River, Colorado, USA. *Geomorphology*, 124(1-2):1–16, December 2010. ISSN 0169555X. doi: 10.1016/j.geomorph.2010.07.021. URL <https://linkinghub.elsevier.com/retrieve/pii/S0169555X1000334X>.
- Esri. ArcGIS Desktop (version 10.8.1), 2020. URL <https://desktop.arcgis.com/en/>.

- Frederic Fabry, Aldo Bellon, Mike R. Duncan, and Geoffrey L. Austin. High resolution rainfall measurements by radar for very small basins: the sampling problem reexamined. *Journal of Hydrology*, 161(1-4):415–428, September 1994. ISSN 00221694. doi: 10.1016/0022-1694(94)90138-4. URL <https://linkinghub.elsevier.com/retrieve/pii/0022169494901384>.
- Greg Fall, David Kitzmiller, Sandra Pavlovic, Ziya Zhang, Nathan Patrick, Michael St. Laurent, Carl Trypaluk, Wanru Wu, and Dennis Miller. The Office of Water Prediction’s Analysis of Record for Calibration, version 1.1: Dataset description and precipitation evaluation. *JAWRA Journal of the American Water Resources Association*, 59(6):1246–1272, December 2023. ISSN 1093-474X, 1752-1688. doi: 10.1111/1752-1688.13143. URL <https://onlinelibrary.wiley.com/doi/10.1111/1752-1688.13143>.
- Dengxing Fan, Guodong Jia, Yangyang Wang, and Xinxiao Yu. The effectiveness of mulching practices on water erosion control: A global meta-analysis. *Geoderma*, 438:116643, October 2023. ISSN 0016-7061. doi: 10.1016/j.geoderma.2023.116643. URL <https://www.sciencedirect.com/science/article/pii/S0016706123003208>.
- L B Faucette, L M Risse, M A Nearing, J W Gaskin, and L T West. Runoff, erosion, and nutrient losses from compost and mulch blankets under simulated rainfall. 2004.
- C. Fernández and A. Vega, J. Effects of mulching and post-fire salvage logging on soil erosion and vegetative regrowth in NW Spain. *Forest Ecology and Management*, 375:46–54, 2016. doi: 10.1016/j.foreco.2016.05.024.
- Cristina Fernández, José Ma Fernández-Alonso, and José A. Vega. Exploring the effect of hydrological connectivity and soil burn severity on sediment yield after wildfire and mulching. *Land Degradation & Development*, 31(13):1611–1621, 2020. ISSN 1099-145X. doi: 10.1002/ldr.3539. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/ldr.3539>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/ldr.3539>.

- Randy B. Foltz and Natalie S. Copeland. Evaluating the efficacy of wood shreds for mitigating erosion. *Journal of Environmental Management*, 90(2):779–785, February 2009. ISSN 03014797. doi: 10.1016/j.jenvman.2008.01.006. URL <https://linkinghub.elsevier.com/retrieve/pii/S0301479708000212>.
- R.B. Foltz and N.S. Wagenbrenner. An evaluation of three wood shred blends for post-fire erosion control using indoor simulated rain events on small plots. *CATENA*, 80(2):86–94, February 2010. ISSN 03418162. doi: 10.1016/j.catena.2009.09.003. URL <https://linkinghub.elsevier.com/retrieve/pii/S0341816209001787>.
- NOAA-USGS Debris Flow Task Force. Noaa-usgs debris-flow warning system - final report. Report 1283, USGS Publications Warehouse, Reston, VA, 2005. URL <https://pubs.usgs.gov/publication/cir1283>. Accessed: January 19, 2024.
- Giuseppe Formetta, Francesco Marra, Eleonora Dallan, Mattia Zaramella, and Marco Borga. Differential orographic impact on sub-hourly, hourly, and daily extreme precipitation. *Advances in Water Resources*, 159:104085, January 2022. ISSN 03091708. doi: 10.1016/j.advwatres.2021.104085. URL <https://linkinghub.elsevier.com/retrieve/pii/S0309170821002372>.
- Hayley J. Fowler, Geert Lenderink, Andreas F. Prein, Seth Westra, Richard P. Allan, Nikolina Ban, Renaud Barbero, Peter Berg, Stephen Blenkinsop, Hong X. Do, Selma Guerreiro, Jan O. Haerter, Elizabeth J. Kendon, Elizabeth Lewis, Christoph Schaer, Ashish Sharma, Gabriele Villarini, Conrad Wasko, and Xuebin Zhang. Anthropogenic intensification of short-duration rainfall extremes. *Nature Reviews Earth & Environment*, 2(2):107–122, January 2021. ISSN 2662-138X. doi: 10.1038/s43017-020-00128-6. URL <https://www.nature.com/articles/s43017-020-00128-6>.
- Yoav Freund and Robert E. Schapire. Experiments with a new boosting algorithm. *Machine Learning: Proceedings of the Thirteenth International Conference*, 96:148–156, 1996.

- Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), October 2001. ISSN 0090-5364. doi: 10.1214/aos/1013203451. URL <https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-function-approximation-A-gradient-boosting-machine/10.1214/aos/1013203451.full>.
- G. D. Jennings and A. R. Jarrett. Laboratory Evaluation of Mulches in Reducing Erosion. *Transactions of the ASAE*, 28(5):1466–1470, 1985. ISSN 2151-0059. doi: 10.13031/2013.32461. URL <http://elibrary.asabe.org/abstract.asp??JID=3&AID=32461&CID=t1985&v=28&i=5&T=1>.
- P. N. Gatlin and W. A. Petersen. A rain taxonomy for degraded visual environment mitigation. NASA Technical Memorandum NASA/TM-2018-219882, National Aeronautics and Space Administration, Marshall Space Flight Center, Huntsville, Alabama, March 2018. URL <https://ntrs.nasa.gov/citations/20180002207>. NASA Technical Memorandum.
- Antonio Girona-García, Diana C.S. Vieira, Joana Silva, Cristina Fernández, Peter R. Robichaud, and J. Jacob Keizer. Effectiveness of post-fire soil erosion mitigation treatments: A systematic review and meta-analysis. *Earth-Science Reviews*, 217:103611, June 2021. ISSN 00128252. doi: 10.1016/j.earscirev.2021.103611. URL <https://linkinghub.elsevier.com/retrieve/pii/S0012825221001112>.
- Antonio Girona-García, Carola Cretella, Cristina Fernández, Peter R. Robichaud, Diana C.S. Vieira, and Jan Jacob Keizer. How much does it cost to mitigate soil erosion after wildfires? *Journal of Environmental Management*, 334:117478, May 2023. ISSN 03014797. doi: 10.1016/j.jenvman.2023.117478. URL <https://linkinghub.elsevier.com/retrieve/pii/S0301479723002669>.
- Jonathan J. Gourley, Zachary L. Flamig, Humberto Vergara, Pierre-Emmanuel Kirstetter, Robert A. Clark, Elizabeth Argyle, Ami Arthur, Steven Martinaitis, Galateia Terti, Jessica M. Erlingis, Yang Hong, and Kenneth W. Howard. The FLASH Project: Improving the Tools for Flash Flood Monitoring and Prediction across the United States. *Bulletin of the American Meteorological Society*, 98(2):361–372, February 2017. ISSN 0003-0007, 1520-0477. doi: 10.1175/BAMS-D-15-00247.1. URL <https://journals.ametsoc.org/doi/10.1175/BAMS-D-15-00247.1>.

- Ethan D. Gutmann, Roy M. Rasmussen, C. Liu, K. Ikeda, D. J. Gochis, Martyn P. Clark, Jimmy Dudhia, and G. Thompson. A comparison of statistical and dynamical downscaling of winter precipitation over complex terrain. *Journal of Climate*, 25(1):262–281, 2012. doi: 10.1175/2011JCLI4109.1.
- Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O’Reilly Media, 3rd edition, 2021. ISBN 978-1492032649.
- Brian Henn, Andrew J. Newman, Ben Livneh, Christopher Daly, and Jessica D. Lundquist. An assessment of differences in gridded precipitation datasets in complex terrain. *Journal of Hydrology*, 556:1205–1219, January 2018. ISSN 00221694. doi: 10.1016/j.jhydrol.2017.03.008. URL <https://linkinghub.elsevier.com/retrieve/pii/S0022169417301452>.
- Philip E. Higuera, Bryan N. Shuman, and Kyra D. Wolf. Rocky Mountain subalpine forests now burning more than any time in recent millennia. *Proceedings of the National Academy of Sciences*, 118(25):e2103135118, June 2021. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2103135118. URL <http://www.pnas.org/lookup/doi/10.1073/pnas.2103135118>.
- Kyle A. Hilburn, Imme Ebert-Uphoff, and Steven D. Miller. Development and Interpretation of a Neural-Network-Based Synthetic Radar Reflectivity Estimator Using GOES-R Satellite Observations. *Journal of Applied Meteorology and Climatology*, 60(1):3–21, January 2021. ISSN 1558-8424, 1558-8432. doi: 10.1175/JAMC-D-20-0084.1. URL <https://journals.ametsoc.org/view/journals/apme/60/1/jamc-d-20-0084.1.xml>.
- Stephen Hodanish, Brandon Vogt, and Paul Wolyn. Colorado Lightning Climatology. *Journal of Operational Meteorology*, pages 45–60, June 2019. ISSN 23256184. doi: 10.1519/nwajom.2019.0704. URL <http://nwafiles.nwas.org/jom/articles/2019/2019-JOM4/2019-JOM4.pdf>.

- R. D. Jarrett and J. E. Costa. Evaluation of the flood hydrology of the colorado front range using precipitation, streamflow and paleoflood data. Technical Report Water-Resources Investigations Report 87-4117, U.S. Geological Survey, Denver, Colorado, 1988.
- Robert D. Jarrett. Flood Elevation Limits in the Rocky Mountains. *Engineering Hydrology*, pages 180–185, January 1993. MAG ID: 75798858 S2ID: 0eb3069c228a490527cdc112096d4daf2d3aba87.
- Julie Rose N. Javier, James A. Smith, John England, Mary Lynn Baeck, Matthias Steiner, and Alexandros A. Ntelekos. Climatology of extreme rainfall and flooding from orographic thunderstorm systems in the upper Arkansas River basin. *Water Resources Research*, 43(10): 2006WR005093, October 2007. ISSN 0043-1397, 1944-7973. doi: 10.1029/2006WR005093. URL <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2006WR005093>.
- X. Jing, B. Geerts, Y. Wang, and C. Liu. Evaluating seasonal orographic precipitation in the interior western united states using gauge data, gridded precipitation estimates, and a regional climate simulation. *Journal of Hydrometeorology*, 18(9):2541–2558, 2017. doi: 10.1175/JHM-D-17-0056.1.
- D. D. Johnson and A. J. Cline. Colorado mountain soils. *Advances in Agronomy*, 17:233–281, 1965. doi: 10.1016/S0065-2113(08)60415-0.
- Pierre-Emmanuel Kirstetter, Y. Hong, J. J. Gourley, S. Chen, Z. Flamig, J. Zhang, M. Schwaller, W. Petersen, and E. Amitai. Toward a Framework for Systematic Error Modeling of Spaceborne Precipitation Radar with NOAA/NSSL Ground Radar–Based National Mosaic QPE. *Journal of Hydrometeorology*, 13(4):1285–1300, August 2012. ISSN 1525-755X, 1525-7541. doi: 10.1175/JHM-D-11-0139.1. URL <http://journals.ametsoc.org/doi/10.1175/JHM-D-11-0139.1>.
- Pierre-Emmanuel Kirstetter, Jonathan J. Gourley, Yang Hong, Jian Zhang, Saber Moazami-goodarzi, Carrie Langston, and Ami Arthur. Probabilistic precipitation rate estimates with ground-based radar networks. *Water Resources Research*, 51(3):1422–1442, 2015. doi:

<https://doi.org/10.1002/2014WR015672>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014WR015672>.

Isaac J. Larsen, Lee H. MacDonald, Ethan Brown, Daniella Rough, Matthew J. Welsh, Joseph H. Pietraszek, Zamir Libohova, Juan de Dios Benavides-Solorio, and Keelin Schaffrath. Causes of post-fire runoff and erosion: Water repellency, cover, or soil sealing? *Soil Science Society of America Journal*, 73(4):1393–1407, 2009. doi: 10.2136/sssaj2007.0432.

Mochi Liao and Ana P. Barros. Toward optimal rainfall for flood prediction in headwater basins: Orographic QPE error modeling using machine learning. preprint, Preprints, January 2023. URL <https://doi.org/10.22541/essoar.168867019.12345678/v1>.

Jessica Lundquist, Mimi Hughes, Ethan Gutmann, and Sarah Kapnick. Our Skill in Modeling Mountain Rain and Snow is Bypassing the Skill of Our Observational Networks. *Bulletin of the American Meteorological Society*, 100(12):2473–2490, December 2019. ISSN 0003-0007, 1520-0477. doi: 10.1175/BAMS-D-19-0001.1. URL <https://journals.ametsoc.org/view/journals/bams/100/12/bams-d-19-0001.1.xml>.

Robert A. Maddox, Jian Zhang, Jonathan J. Gourley, and Kenneth W. Howard. Weather Radar Coverage over the Contiguous United States. *Weather and Forecasting*, 17(4):927–934, August 2002. ISSN 0882-8156, 1520-0434. doi: 10.1175/1520-0434(2002)017<0927:WRCOTC>2.0.CO;2. URL [http://journals.ametsoc.org/doi/10.1175/1520-0434\(2002\)017<0927:WRCOTC>2.0.CO;2](http://journals.ametsoc.org/doi/10.1175/1520-0434(2002)017<0927:WRCOTC>2.0.CO;2).

Kelly Mahoney, F. Martin Ralph, Klaus Wolter, Nolan Doesken, Michael Dettinger, Daniel Gottas, Timothy Coleman, and Allen White. Climatology of Extreme Daily Precipitation in Colorado and Its Diverse Spatial and Seasonal Variability. *Journal of Hydrometeorology*, 16(2):781–792, April 2015. ISSN 1525-755X, 1525-7541. doi: 10.1175/JHM-D-14-0112.1. URL <http://journals.ametsoc.org/doi/10.1175/JHM-D-14-0112.1>.

- Francesco Marra, Moshe Armon, Marco Borga, and Efrat Morin. Orographic Effect on Extreme Precipitation Statistics Peaks at Hourly Time Scales. *Geophysical Research Letters*, 48(5), March 2021. ISSN 0094-8276, 1944-8007. doi: 10.1029/2020GL091498. URL <https://onlinelibrary.wiley.com/doi/10.1029/2020GL091498>.
- J. S. Marshall and W. McK. Palmer. The distribution of raindrops with size. *Journal of Meteorology*, 5:165–166, 1948. doi: 10.1175/1520-0469(1948)005<0165:TDORWS>2.0.CO;2.
- Steven M. Martinaitis, Andrew P. Osborne, Micheal J. Simpson, Jian Zhang, Kenneth W. Howard, Stephen B. Cocks, Ami Arthur, Carrie Langston, and Brian T. Kaney. A Physically Based Multisensor Quantitative Precipitation Estimation Approach for Gap-Filling Radar Coverage. *Journal of Hydrometeorology*, 21(7):1485–1511, July 2020. ISSN 1525-755X, 1525-7541. doi: 10.1175/JHM-D-19-0264.1. URL <https://journals.ametsoc.org/view/journals/hydr/21/7/jhmD190264.xml>.
- Steven M. Martinaitis, Stephen B. Cocks, Micheal J. Simpson, Andrew P. Osborne, Sebastian S. Harkema, Heather M. Grams, Jian Zhang, and Kenneth W. Howard. Advancements and Characteristics of Gauge Ingest and Quality Control within the Multi-Radar Multi-Sensor System. *Journal of Hydrometeorology*, March 2021. ISSN 1525-755X, 1525-7541. doi: 10.1175/JHM-D-20-0234.1. URL <https://journals.ametsoc.org/view/journals/hydr/aop/JHM-D-20-0234.1/JHM-D-20-0234.1.xml>.
- Paola Mazzoglio, Ilaria Butera, Massimiliano Alvioli, and Pierluigi Claps. The role of morphology in the spatial distribution of short-duration rainfall extremes in Italy. *Hydrology and Earth System Sciences*, 26(6):1659–1672, March 2022. ISSN 1607-7938. doi: 10.5194/hess-26-1659-2022. URL <https://hess.copernicus.org/articles/26/1659/2022/>.
- Iowa Environmental Mesonet. Mtarchive daily selected files, Accessed 2023. URL <https://mtarchive.geol.iastate.edu/>.

- Mesowest. Mesowest - Weather and Climate Data. Mesowest, Accessed 2023. URL {<https://mesowest.utah.edu/cgi-bin/droman/mesomap.cgi?state=CO>}.
- Benjamin B. Mirus, Thom A. Bogaard, Roberto Greco, and Manfred Stähli. Invited Perspectives: Integrating hydrologic information into the next generation of landslide early warning systems, May 2024. URL <https://egusphere.copernicus.org/preprints/2024/egusphere-2024-1219/>.
- S. Moazami and M.R. Najafi. A comprehensive evaluation of GPM-IMERG V06 and MRMS with hourly ground-based precipitation observations across Canada. *Journal of Hydrology*, 594: 125929, March 2021. ISSN 00221694. doi: 10.1016/j.jhydrol.2020.125929. URL <https://linkinghub.elsevier.com/retrieve/pii/S0022169420313901>.
- Mohamad Ayob Mohamadi and Ataollah Kavian. Effects of rainfall patterns on runoff and soil erosion in field plots. *International Soil and Water Conservation Research*, 3(4):273–281, December 2015. ISSN 20956339. doi: 10.1016/j.iswcr.2015.10.001. URL <https://linkinghub.elsevier.com/retrieve/pii/S209563391530071X>.
- A.A.A. Montenegro, J.R.C.B. Abrantes, J.L.M.P. De Lima, V.P. Singh, and T.E.M. Santos. Impact of mulching on soil and water dynamics under intermittent simulated rainfall. *CATENA*, 109: 139–149, October 2013. ISSN 03418162. doi: 10.1016/j.catena.2013.03.018. URL <https://linkinghub.elsevier.com/retrieve/pii/S0341816213000866>.
- John A. Moody, Richard A. Shakesby, Peter R. Robichaud, Susan H. Cannon, and Deborah A. Martin. Current research issues related to post-wildfire runoff and erosion processes. *Earth-Science Reviews*, 122:10–37, July 2013. ISSN 00128252. doi: 10.1016/j.earscirev.2013.03.004. URL <https://linkinghub.elsevier.com/retrieve/pii/S0012825213000536>.
- Hernan A. Moreno, Enrique R. Vivoni, and David J. Gochis. Utility of Quantitative Precipitation Estimates for high resolution hydrologic forecasts in mountain watersheds of the Colorado Front Range. *Journal of Hydrology*, 438-439:66–83, May 2012. ISSN 00221694. doi: 10.1016/j.jhydrol.2012.03.019. URL <https://linkinghub.elsevier.com/retrieve/pii/S002216941200217X>.

NASA Jet Propulsion Laboratory. NASA Shuttle Radar Topography Mission 1-arc second Global Land Data. NASA LP DAAC, 2022. <https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1>.

L. Nicótina, E. Alessi Celegon, A. Rinaldo, and M. Marani. On the impact of rainfall patterns on the hydrologic response: RAINFALL PATTERNS AND HYDROLOGIC RESPONSE. *Water Resources Research*, 44(12), December 2008. ISSN 00431397. doi: 10.1029/2007WR006654. URL <http://doi.wiley.com/10.1029/2007WR006654>.

NOAA. Analysis of record for calibration (aorc) dataset, 2024. URL <https://registry.opendata.aws/noaa-nws-aorc>. Accessed 2024-08-22.

NOAA. Noaa atlas 14 precipitation frequency data server (pfds): Gis data, 2024. URL https://hdsc.nws.noaa.gov/pfds/pfds_gis.html. Accessed 2024-12-09.

NOAA. Virtual lab. <https://vlab.noaa.gov/web/wdtd/>, 2024. Accessed: January 19, 2024.

NOAA Earth Systems Research Laboratory. Colorado – new mexico regional extreme precipitation study: Volume iv - application of dynamical model approaches using the noaa high resolution rapid refresh and weather research and forecast models. Technical report, NOAA Earth Systems Research Laboratory, Boulder, CO, November 2018.

NOAA National Weather Service. Analysis of Record for Calibration (AORC) Dataset, 2024. URL <https://registry.opendata.aws/noaa-nws-aorc>. Accessed: 2024-08-22.

NOAA/CPC. Wgrib2. <https://www.cpc.ncep.noaa.gov/products/wesley/wgrib2/>, 2023.

NOAA National Severe Storms Laboratory. NSSL. Code updates – mrms qpe, 2024. URL <https://inside.nssl.noaa.gov/mrms/code-updates/>. Accessed: 2024-06-10.

Andrew P. Osborne, Jian Zhang, Micheal J. Simpson, Kenneth W. Howard, and Stephen B. Cocks. Application of Machine Learning Techniques to Improve Multi-Radar Multi-Sensor (MRMS)

- Precipitation Estimates in the Western United States. *Artificial Intelligence for the Earth Systems*, 2(2):220053, April 2023. ISSN 2769-7525. doi: 10.1175/AIES-D-22-0053.1. URL <https://journals.ametsoc.org/view/journals/aies/2/2/AIES-D-22-0053.1.xml>.
- Misagh Parhizkar, Mahmood Shabanpour, Manuel Esteban Lucas-Borja, Demetrio Antonio Zema, Siyue Li, Nobuaki Tanaka, and Artemio Cerdà. Effects of length and application rate of rice straw mulch on surface runoff and soil loss under laboratory simulated rainfall. *International Journal of Sediment Research*, 36(4):468–478, August 2021. ISSN 10016279. doi: 10.1016/j.ijsrc.2020.12.002. URL <https://linkinghub.elsevier.com/retrieve/pii/S100162792030127X>.
- A.J. Parsons and P.M. Stone. Effects of intra-storm variations in rainfall intensity on interrill runoff and erosion. *CATENA*, 67(1):68–78, August 2006. ISSN 03418162. doi: 10.1016/j.catena.2006.03.002. URL <https://linkinghub.elsevier.com/retrieve/pii/S034181620600052X>.
- Trevor F. Partridge, Zachary C. Johnson, Rachel R. Sleeter, Sharon L. Qi, Michelle A. Walvoord, Sheila F. Murphy, Cara L. Peterman-Phipps, and Brian A. Ebel. Opportunities and challenges for precipitation forcing data in post-wildfire hydrologic modeling applications. *WIREs Water*, 11(5):e1728, 2024. doi: 10.1002/wat2.1728.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- S. Perica, D. Martin, S. Pavlovic, I. Roy, M. S. Laurent, C. Trypaluk, et al. *Precipitation-Frequency Atlas of the United States, Volume 8, Version 2*, volume 8. Dept. of Commerce, NOAA, NWS, United States, 2 edition, 2013. Version 2.
- Lee Picard and Clifford Mass. The Sensitivity of Orographic Precipitation to Flow Direction: An Idealized Modeling Approach. *Journal of Hydrometeorology*, 18(6):1673–1688, June 2017.

ISSN 1525-755X, 1525-7541. doi: 10.1175/JHM-D-16-0209.1. URL <http://journals.ametsoc.org/doi/10.1175/JHM-D-16-0209.1>.

J Poesen, E De Luna, A Franca, J Nachtergaele, and G Govers. Concentrated flow erosion rates as affected by rock fragment cover and initial soil moisture content. *CATENA*, 36(4):315–329, August 1999. ISSN 03418162. doi: 10.1016/S0341-8162(99)00044-2. URL <https://linkinghub.elsevier.com/retrieve/pii/S0341816299000442>.

Sergio Alegre Prats, Joseph W. Wagenbrenner, Martinho António Santos Martins, Maruxa Cortizo Malvar, and Jan Jacob Keizer. Mid-term and scaling effects of forest residue mulching on post-fire runoff and soil erosion. *Science of The Total Environment*, 573:1242–1254, December 2016. ISSN 00489697. doi: 10.1016/j.scitotenv.2016.04.064. URL <https://linkinghub.elsevier.com/retrieve/pii/S0048969716307458>.

Sergio Alegre Prats, João Rafael Abrantes, Isabela P. Crema, Jacob J. Keizer, and João L.M.P. De Lima. Runoff and soil erosion mitigation with sieved forest residue mulch strips under controlled laboratory conditions. *Forest Ecology and Management*, 396:102–112, July 2017. ISSN 03781127. doi: 10.1016/j.foreco.2017.04.019. URL <https://linkinghub.elsevier.com/retrieve/pii/S0378112716311902>.

Sergio Alegre Prats, João Rafael Cardoso De Brito Abrantes, Celeste De Oliveira Alves Coelho, Jan Jacob Keizer, and João Luis Mendes Pedroso De Lima. Comparing topsoil charcoal, ash, and stone cover effects on the postfire hydrologic and erosive response under laboratory conditions. *Land Degradation & Development*, 29(7):2102–2111, July 2018. ISSN 1085-3278, 1099-145X. doi: 10.1002/ldr.2884. URL <https://onlinelibrary.wiley.com/doi/10.1002/ldr.2884>.

PRISM Climate Group. PRISM Gridded Climate Data: Daily precipitation, 2025. URL <https://ftp.prism.oregonstate.edu/daily/ppt/>. Accessed 2025-01-25.

R. M. Rasmussen, F. Chen, C.H. Liu, K. Ikeda, A. Prein, J. Kim, T. Schneider, A. Dai, D. Gochis, A. Dugger, Y. Zhang, A. Jaye, J. Dudhia, C. He, M. Harrold, L. Xue, S. Chen, A. New-

- man, E. Dougherty, R. Abolafia-Rosenzweig, N. D. Lybarger, R. Viger, D. Lesmes, K. Skalak, J. Brakebill, D. Cline, K. Dunne, K. Rasmussen, and G. Miguez-Macho. CONUS404: The NCAR–USGS 4-km Long-Term Regional Hydroclimate Reanalysis over the CONUS. *Bulletin of the American Meteorological Society*, 104(8):E1382–E1408, August 2023a. ISSN 0003-0007, 1520-0477. doi: 10.1175/BAMS-D-21-0326.1. URL <https://journals.ametsoc.org/view/journals/bams/104/8/BAMS-D-21-0326.1.xml>.
- R.M. Rasmussen, F. Chen, C. Liu, K. Ikeda, A. Prein, J. Kim, T. Schneider, A. Dai, D. Gochis, A. Dugger, Y. Zhang, A. Jaye, J. Dudhia, C. He, M. Harrold, L. Xue, S. Chen, A. Newman, E. Dougherty, R. Abolafia-Rozenzweig, N. Lybarger, R. Viger, K. Dunne, K. Rasmussen, and G. Miguez-Macho. CONUS404: Four-kilometer long-term regional hydroclimate reanalysis over the conterminous United States (ver. 2.0, December 2023), 2023b. URL <https://doi.org/10.5066/P9PHPK4F>.
- Roy M. Rasmussen, Changhai Liu, Kyoko Ikeda, Fei Chen, Ju-Hye Kim, Timothy L. Schneider, David J. Gochis, Aubrey L. Dugger, and Roland Viger. Four-kilometer long-term regional hydroclimate reanalysis over the conterminous united states (conus), 2023c. URL <https://thredds.rda.ucar.edu/thredds/dodsC/files/g/d559000/>. Accessed 2025-01-27.
- F.K. Rengers, S.J. Bower, A. Knapp, J.W. Kean, and D.M. Staley. Debris flow, precipitation, and volume measurements in the grizzly creek burn perimeter june 2021-september 2022, glenwood canyon, colorado. U.S. Geological Survey data release, October 2023. URL <https://doi.org/10.5066/P9Z7RROL>.
- David R. Roberts, Volker Bahn, Simone Ciuti, Mark S. Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, José J. Lahoz-Monfort, Boris Schröder, Wilfried Thuiller, David I. Warton, Brendan A. Wintle, Florian Hartig, and Carsten F. Dormann. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8):913–929, August 2017. ISSN 09067590. doi: 10.1111/ecog.02881. URL <https://onlinelibrary.wiley.com/doi/10.1111/ecog.02881>.

- Peter R. Robichaud, Sarah A. Lewis, Joseph W. Wagenbrenner, Louise E. Ashmun, and Robert E. Brown. Post-fire mulching for runoff and erosion mitigation. *CATENA*, 105:75–92, June 2013a. ISSN 03418162. doi: 10.1016/j.catena.2012.11.015. URL <https://linkinghub.elsevier.com/retrieve/pii/S0341816212002524>.
- P.R. Robichaud, P. Jordan, S.A. Lewis, L.E. Ashmun, S.A. Covert, and R.E. Brown. Evaluating the effectiveness of wood shred and agricultural straw mulches as a treatment to reduce post-wildfire hillslope erosion in southern British Columbia, Canada. *Geomorphology*, 197: 21–33, September 2013b. ISSN 0169555X. doi: 10.1016/j.geomorph.2013.04.024. URL <https://linkinghub.elsevier.com/retrieve/pii/S0169555X13002316>.
- Gerard H. Roe. OROGRAPHIC PRECIPITATION. *Annual Review of Earth and Planetary Sciences*, 33(1):645–671, May 2005. ISSN 0084-6597, 1545-4495. doi: 10.1146/annurev.earth.33.092203.122541. URL <https://www.annualreviews.org/doi/10.1146/annurev.earth.33.092203.122541>.
- Matthew W. Rossi, Robert S. Anderson, Suzanne P. Anderson, and Gregory E. Tucker. Orographic Controls on Subdaily Rainfall Statistics and Flood Frequency in the Colorado Front Range, USA. *Geophysical Research Letters*, 47(4), February 2020. ISSN 0094-8276, 1944-8007. doi: 10.1029/2019GL085086. URL <https://onlinelibrary.wiley.com/doi/10.1029/2019GL085086>.
- RTI International. Metportal: Colorado–new mexico region extreme precipitation metrics, 2024. URL https://rti-metportal.shinyapps.io/conm_region/. Accessed 2024-12-09.
- M.J.M Römken, K Helming, and S.N Prasad. Soil erosion under different rainfall intensities, surface roughness, and soil water regimes. *CATENA*, 46(2-3):103–123, January 2002. ISSN 03418162. doi: 10.1016/S0341-8162(01)00161-8. URL <https://linkinghub.elsevier.com/retrieve/pii/S0341816201001618>.
- S. H. R. Sadeghi, L. Gholami, E. Sharifi, A. Khaledi Darvishan, and M. Homaei. Scale effect on runoff and soil loss control using rice straw mulch under laboratory conditions. *Solid Earth*, 6

(1):1–8, January 2015. ISSN 1869-9529. doi: 10.5194/se-6-1-2015. URL <https://se.copernicus.org/articles/6/1/2015/>.

Haytham M. Salem and Adil A. Meselhy. A portable rainfall simulator to evaluate the factors affecting soil erosion in the northwestern coastal zone of Egypt. *Natural Hazards*, 105(3):2937–2955, February 2021. ISSN 0921-030X, 1573-0840. doi: 10.1007/s11069-020-04432-8. URL <http://link.springer.com/10.1007/s11069-020-04432-8>.

Crystal B. Schaaf, Joshua Wurman, and Robert M. Banta. Thunderstorm-producing terrain features. *Bulletin of the American Meteorological Society*, 69(3):272–277, 1988. doi: 10.1175/1520-0477(1988)069<0272:TPTF>2.0.CO;2. URL https://journals.ametsoc.org/view/journals/bams/69/3/1520-0477_1988_069_0272_tptf_2_0_co_2.xml. Accessed: 2025-01-10.

Sarah R. Schmeer, Stephanie K. Kampf, Lee H. MacDonald, Josh Hewitt, and Codie Wilson. Empirical models of annual post-fire erosion on mulched and unmulched hillslopes. *CATENA*, 163:276–287, April 2018. ISSN 03418162. doi: 10.1016/j.catena.2017.12.029. URL <https://linkinghub.elsevier.com/retrieve/pii/S0341816217304319>.

scikit-image developers. Trainable segmentation using local features and random forests, 2025. URL https://scikit-image.org/docs/0.25.x/auto_examples/segmentation/plot_trainable_segmentation.html. Accessed 21May2025.

P. R. Seaber, F. P. Kapinos, and G. L. Knapp. *Hydrologic Unit Maps*, volume 2294 of *Water-Supply Paper*. U.S. Geological Survey, 1987.

Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 57–61, Austin, Texas, 2010. URL <https://statsmodels.org>. SciPy 2010.

Bong-Chul Seo and Witold F. Krajewski. Correcting temporal sampling error in radar-rainfall: Effect of advection parameters and rain storm characteristics on the correction accuracy. *Journal*

- of Hydrology*, 531:272–283, December 2015. ISSN 00221694. doi: 10.1016/j.jhydrol.2015.04.018. URL <https://linkinghub.elsevier.com/retrieve/pii/S002216941500267X>.
- Boris Sevruk. Rainfall Measurement: Gauges. In Malcolm G. Anderson and Jeffrey J. McDonnell, editors, *Encyclopedia of Hydrological Sciences*, page hsa038. John Wiley & Sons, Ltd, Chichester, UK, October 2005. ISBN 978-0-471-49103-3 978-0-470-84894-4. doi: 10.1002/0470848944.hsa038. URL <https://onlinelibrary.wiley.com/doi/10.1002/0470848944.hsa038>.
- R Shakesby and S Doerr. Wildfire as a hydrological and geomorphological agent. *Earth-Science Reviews*, 74(3-4):269–307, February 2006. ISSN 00128252. doi: 10.1016/j.earscirev.2005.10.006. URL <https://linkinghub.elsevier.com/retrieve/pii/S0012825205001467>.
- Aman Sharma, Conrad Wasko, and Dennis P. Lettenmaier. If precipitation extremes are increasing, why aren't floods? *Water Resources Research*, 54(11):8545–8551, 2018. doi: 10.1029/2018WR023749.
- Lisa C. Sieck, Stephen J. Burges, and Matthias Steiner. Challenges in obtaining reliable measurements of point rainfall: RELIABLE MEASUREMENTS OF POINT RAINFALL. *Water Resources Research*, 43(1), January 2007. ISSN 00431397. doi: 10.1029/2005WR004519. URL <http://doi.wiley.com/10.1029/2005WR004519>.
- T. Smets, J. Poesen, and A. Knapen. Spatial scale effects on the effectiveness of organic mulches in reducing soil erosion by water. *Earth-Science Reviews*, 89(1-2):1–12, July 2008. ISSN 00128252. doi: 10.1016/j.earscirev.2008.04.001. URL <https://linkinghub.elsevier.com/retrieve/pii/S001282520800041X>.
- Ronald B. Smith. 100 Years of Progress on Mountain Meteorology Research. *Meteorological Monographs*, 59:20.1–20.73, January 2019. ISSN 0065-9401. doi: 10.1175/AMSMONOGRAPHS-D-18-0022.1. URL <http://journals.ametsoc.org/doi/10.1175/AMSMONOGRAPHS-D-18-0022.1>.

- Zbyněk Sokol, Jan Szturc, Johanna Orellana-Alvear, Jana Popová, Anna Jurczyk, and Rolando Céleri. The Role of Weather Radar in Rainfall Estimation and Its Application in Meteorological and Hydrological Modelling—A Review. *Remote Sensing*, 13(3):351, January 2021. ISSN 2072-4292. doi: 10.3390/rs13030351. URL <https://www.mdpi.com/2072-4292/13/3/351>.
- Luyao Sun, Haonan Chen, Zhe Li, and Lei Han. Cross Validation of GOES-16 and NOAA Multi-Radar Multi-Sensor (MRMS) QPE over the Continental United States. *Remote Sensing*, 13(20): 4030, October 2021. ISSN 2072-4292. doi: 10.3390/rs13204030. URL <https://www.mdpi.com/2072-4292/13/20/4030>.
- Kamran H. Syed, David C. Goodrich, Donald E. Myers, and Soroosh Sorooshian. Spatial characteristics of thunderstorm rainfall fields and their relation to runoff. *Journal of Hydrology*, 271(1-4):1–21, February 2003. ISSN 00221694. doi: 10.1016/S0022-1694(02)00311-6. URL <https://linkinghub.elsevier.com/retrieve/pii/S0022169402003116>.
- Guoqiang Tang, Ali Behrangi, Di Long, Changming Li, and Yang Hong. Accounting for spatiotemporal errors of gauges: A critical step to evaluate gridded precipitation products. *Journal of Hydrology*, 559:294–306, April 2018. doi: 10.1016/j.jhydrol.2018.02.057. MAG ID: 2793745286 S2ID: 2ccc8ef53cbcede6c79d54ae70ce548c2e959e7b.
- Danielle Touma, Samantha Stevenson, Daniel L. Swain, Deepti Singh, Dmitri A. Kalashnikov, and Xingying Huang. Climate change increases risk of extreme rainfall following wildfire in the western United States. *Science Advances*, 8(13):eabm0320, April 2022. doi: 10.1126/sciadv.abm0320. URL <https://www.science.org/doi/full/10.1126/sciadv.abm0320>. Publisher: American Association for the Advancement of Science.
- U.S. Geological Survey. Post-fire debris flow hazards - frequently asked questions. https://landslides.usgs.gov/hazards/postfire_debrisflow/faq.php, 2024. URL https://landslides.usgs.gov/hazards/postfire_debrisflow/faq.php. Accessed: 2024-11-08.

- USGS. Usgs national water information system, Accessed 2023. URL <https://nwis.waterservices.usgs.gov/>.
- Raphaël Vallat. Pingouin: Statistics in python. *Journal of Open Source Software*, 3(31):1026, 2018. doi: 10.21105/joss.01026. URL <https://doi.org/10.21105/joss.01026>.
- Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu, and the scikit-image contributors. scikit-image: image processing in Python. *PeerJ*, 2:e453, 6 2014. ISSN 2167-8359. doi: 10.7717/peerj.453. URL <https://doi.org/10.7717/peerj.453>.
- M.E. Varela, E. Benito, and J.J. Keizer. Effects of wildfire and laboratory heating on soil aggregate stability of pine forests in Galicia: The role of lithology, soil organic matter content and water repellency. *CATENA*, 83(2-3):127–134, November 2010. ISSN 03418162. doi: 10.1016/j.catena.2010.08.001. URL <https://linkinghub.elsevier.com/retrieve/pii/S0341816210001219>.
- D.C.S. Vieira, M.C. Malvar, M.A.S. Martins, D. Serpa, and J.J. Keizer. Key factors controlling the post-fire hydrological and erosive response at micro-plot scale in a recently burned Mediterranean forest. *Geomorphology*, 319:161–173, October 2018. ISSN 0169555X. doi: 10.1016/j.geomorph.2018.07.014. URL <https://linkinghub.elsevier.com/retrieve/pii/S0169555X18302745>.
- Gabriele Villarini, Pradeep V. Mandapaka, Witold F. Krajewski, and Robert J. Moore. Rainfall and sampling uncertainties: A rain gauge perspective. *Journal of Geophysical Research*, 113(D11): D11102, June 2008. ISSN 0148-0227. doi: 10.1029/2007JD009214. URL <http://doi.wiley.com/10.1029/2007JD009214>.
- Gabriele Villarini, Bong-Chul Seo, Francesco Serinaldi, and Witold F. Krajewski. Spatial and temporal modeling of radar rainfall uncertainties. *Atmospheric Research*, 135-136:91–101, January 2014. ISSN 01698095. doi: 10.1016/j.atmosres.2013.09.007. URL <https://linkinghub.elsevier.com/retrieve/pii/S0169809513002482>.

- Katrina R. Virts and Robert H. Holzworth. World wide lightning location network (wwlln) monthly thunder hour data, 2025. URL <https://doi.org/10.5067/WWLLN/DATA101>. Data set.
- J. W. Wagenbrenner, L. H. MacDonald, and D. Rough. Effectiveness of three post-fire rehabilitation treatments in the Colorado Front Range. *Hydrological Processes*, 20(14):2989–3006, September 2006. ISSN 0885-6087, 1099-1085. doi: 10.1002/hyp.6146. URL <https://onlinelibrary.wiley.com/doi/10.1002/hyp.6146>.
- P. White and P. A. Nelson. Evaluation of sub-hourly mrms quantitative precipitation estimates in mountainous terrain using machine learning. *Water Resources Research*, 60:e2024WR037437, 2024. doi: 10.1029/2024WR037437.
- F. Wu, X. Cui, D.-L. Zhang, and L. Qiao. The relationship of lightning activity and short-duration rainfall events during warm seasons over the beijing metropolitan region. *Atmospheric Research*, 195:31–43, 2017. doi: 10.1016/j.atmosres.2017.04.032.
- xESMF Developers. Comparison of six regridding algorithms — xesmf 0.7.0 documentation, 2024. URL https://xesmf.readthedocs.io/en/stable/notebooks/Compare_algorithms.html. Accessed: 2025-05-01.
- K A Yanosek, R B Foltz, and J H Dooley. Performance assessment of wood strand erosion control materials among varying slopes, soil textures, and cover amounts. 2006.
- J. Zhang, Y. Qi, C. Langston, and B. Kaney. Radar quality index (rqi) – a combined measure for beam blockage and vpr effects in a national network. *Weather Radar and Hydrology*, 351:388–393, 2012.
- Jian Zhang, Youcun Qi, Carrie Langston, Brian Kaney, and Kenneth Howard. A Real-Time Algorithm for Merging Radar QPEs with Rain Gauge Observations and Orographic Precipitation Climatology. *Journal of Hydrometeorology*, 15(5):1794–1809, October 2014. ISSN 1525-755X, 1525-7541. doi: 10.1175/JHM-D-13-0163.1. URL <http://journals.ametsoc.org/doi/10.1175/JHM-D-13-0163.1>.

Jian Zhang, Kenneth Howard, Carrie Langston, Brian Kaney, Youcun Qi, Lin Tang, Heather Grams, Yadong Wang, Stephen Cocks, Steven Martinaitis, Ami Arthur, Karen Cooper, Jeff Brogden, and David Kitzmiller. Multi-Radar Multi-Sensor (MRMS) Quantitative Precipitation Estimation: Initial Operating Capabilities. *Bulletin of the American Meteorological Society*, 97(4):621–638, April 2016. ISSN 0003-0007, 1520-0477. doi: 10.1175/BAMS-D-14-00174.1. URL <https://journals.ametsoc.org/doi/10.1175/BAMS-D-14-00174.1>.

Jian Zhang, Lin Tang, Stephen Cocks, Pengfei Zhang, Alexander Ryzhkov, Kenneth Howard, Carrie Langston, and Brian Kaney. A Dual-Polarization Radar Synthetic QPE for Operations. *Journal of Hydrometeorology*, 21(11):2507–2521, November 2020. ISSN 1525-755X, 1525-7541. doi: 10.1175/JHM-D-19-0194.1. URL <https://journals.ametsoc.org/view/journals/hydr/21/11/JHM-D-19-0194.1.xml>.

Y. Zhu, W. Lyu, J. Cramer, V. Rakov, P. Bitzer, and Z. Ding. Analysis of location errors of the u.s. national lightning detection network using lightning strikes to towers. *Journal of Geophysical Research: Atmospheres*, 125(9):e2020JD032530, 2020. doi: 10.1029/2020JD032530.

Zhihua Zhu, Daniel B. Wright, and Guo Yu. The Impact of Rainfall Space-Time Structure in Flood Frequency Analysis. *Water Resources Research*, 54(11):8983–8998, 2018. ISSN 1944-7973. doi: 10.1029/2018WR023550. URL <https://onlinelibrary.wiley.com/doi/abs/10.1029/2018WR023550>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018WR023550>.

Appendix A

Supporting information for Chapter 2

Description: This Supporting Information document includes figures and tables describing the multi-sensor correction method, single-output model experiment, and hyperparameters tuned.

Text A1. Initially, we explored how the 1-hour multi-sensor QPE Pass 2 might be used to improve the 2-minute precipitation rate accuracy. The multi-sensor QPE Pass 2 includes additional information from several numerical weather prediction (NWP) models, various gage networks, and climatological data such as elevation corrections from the PRISM dataset. The inputs vary depending on radar coverage, the type of precipitation, and local topography (NOAA, 2024). A simple multiplicative correction factor, using a ratio of the 1-hr multi-sensor QPE to the 1-hr radar-only QPE, was applied to the 2-minute precipitation rate product to decrease error. A factor of 1 was used where the 1-hr radar-only QPE was equal to zero.

The corrected and uncorrected radar-only 2-minute precipitation rate products were compared to the gage-observed 15-minute intensity for each 8-hour time series. The multi-sensor correction did not significantly improve the RMSE error, with a median normalized RMSE of 41% of maximum MRMS 15-minute intensity for years 2021 through 2022 (compared to 42% for the radar only rate), see Figure S1 for a comparison of the overall distributions of errors. Because the multi-sensor product did not significantly improve the sub-hourly precipitation estimates, and the multi-sensor product is not archived before 2021, we did not include the correction in this analysis.

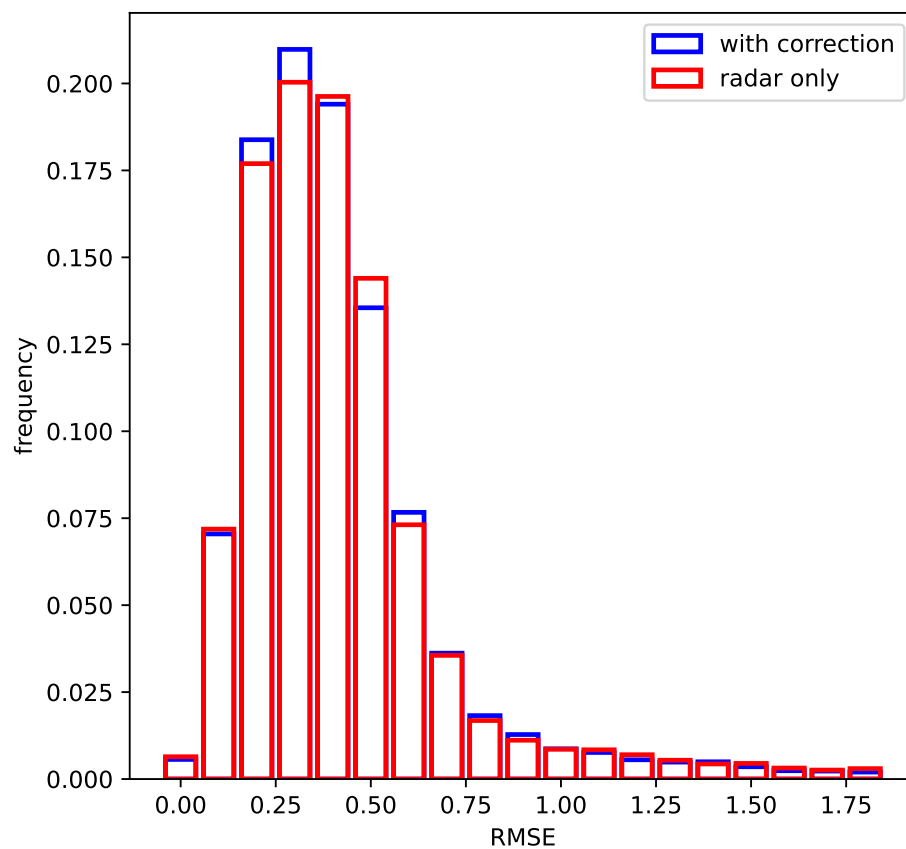


Figure A.1: Distribution of RMSE for radar-only 15-minute intensity and multi-sensor corrected 15-minute intensity.

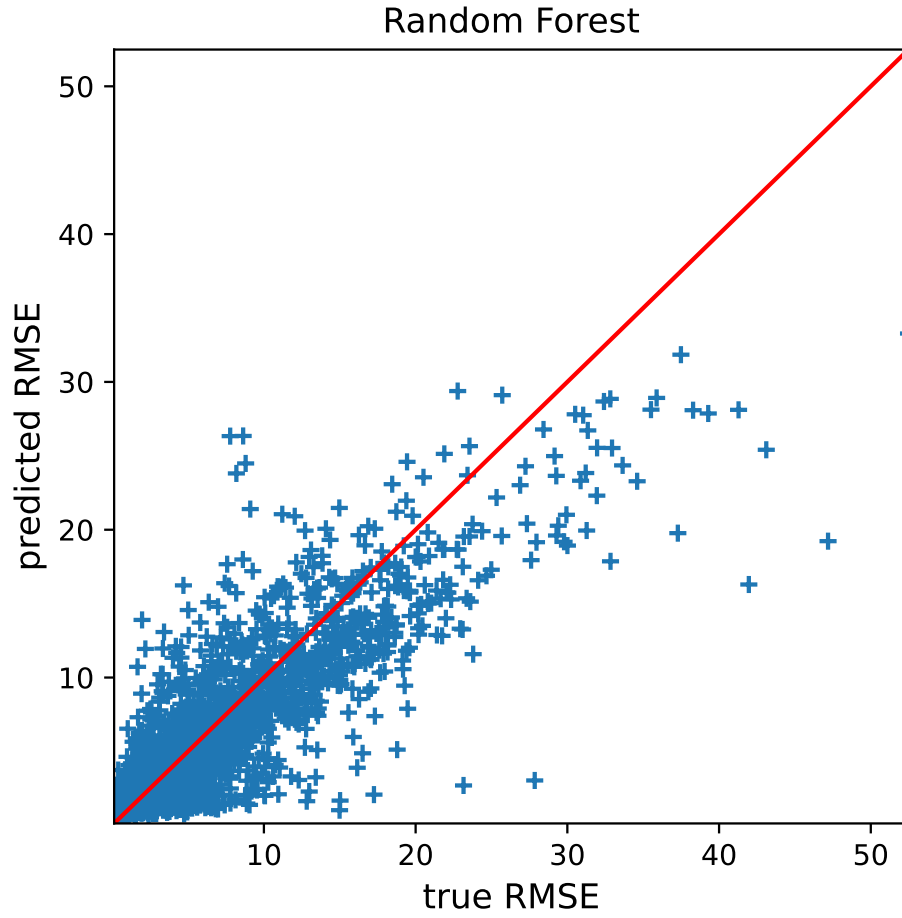


Figure A.2: Spread of model RMSE predictions for the best performing model, a random forest regressor.

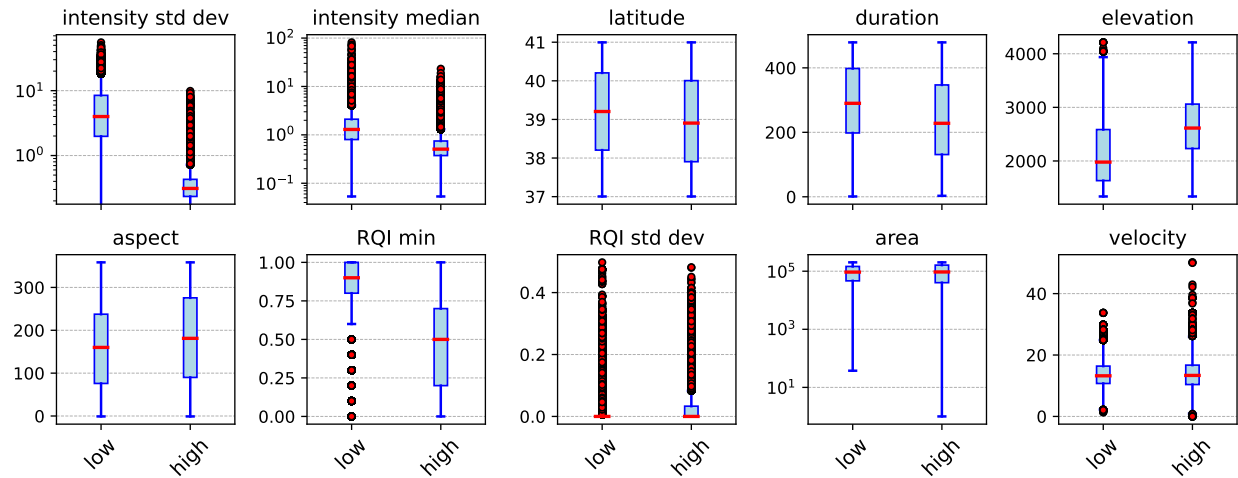


Figure A.3: Comparison of feature value distributions for samples with low and high normalized RMSE predictions for the $\alpha = 0.50$ model.

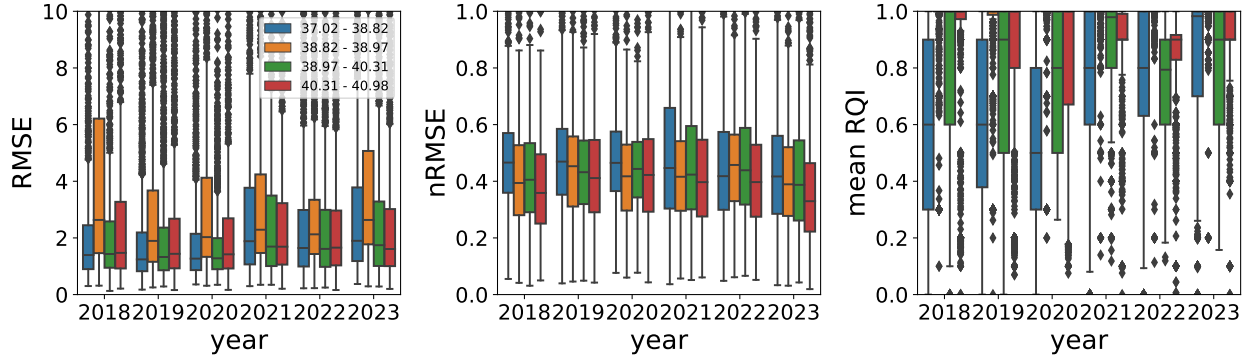


Figure A.4: Difference in error (computed directly from the gage) and mean RQI for each 8-hour sample across all years. For each year, error bars are separated by the latitude of the gage location.

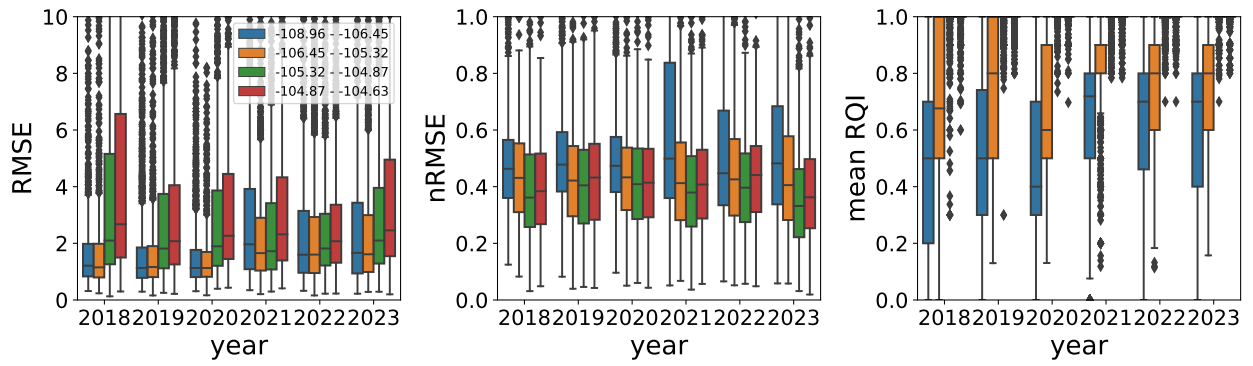


Figure A.5: Difference in error (computed directly from the gage) and mean RQI for each 8-hour sample across all years. For each year, error bars are separated by the longitude of the gage location.

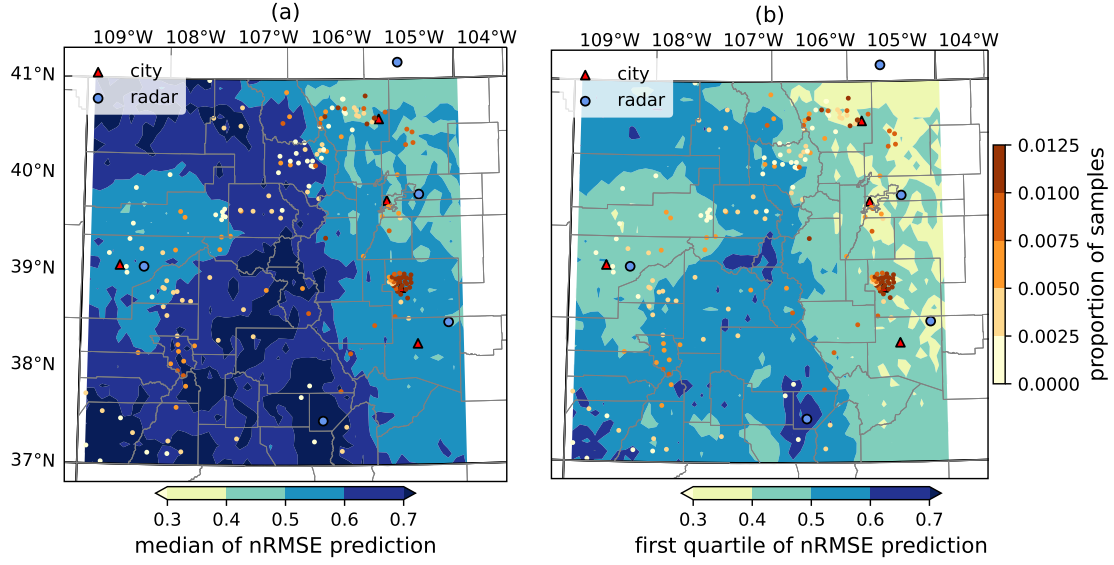


Figure A.6: For the model trained with data prior to MRMS v12.0 (2018-2021), (a) The median normalized RMSE for each coordinate, predicted by the $\alpha = 0.50$ model. (b) The first quartile of normalized RMSE for each coordinate, predicted by the $\alpha = 0.50$ model. The gage locations are shown with colors representing the proportion of samples in the labeled dataset sourced from each location.

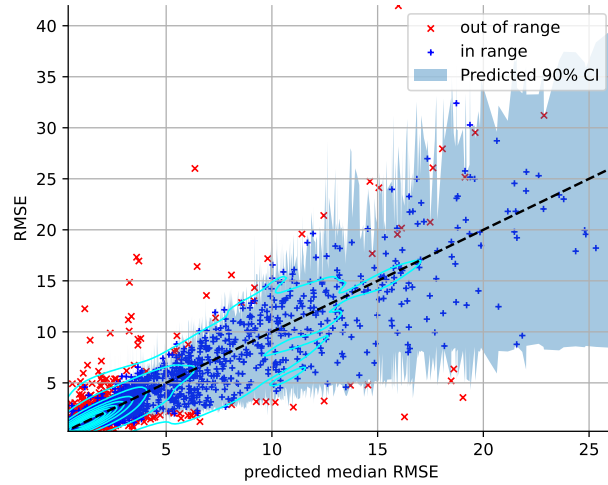


Figure A.7: For the model trained with data prior to MRMS v12.0 (2018-2021), the target values as a function of the median predicted RMSE is shown for the test dataset. Each data point represents an 8-hour sample RMSE between gage and MRMS plotted on the predicted. The predicted 90% confidence interval (CI) is plotted. Truth values in range are blue, those that fall outside the cutoff are red. The kernel density estimation for the target values as a function of the median predicted RMSE is shown in cyan contours.

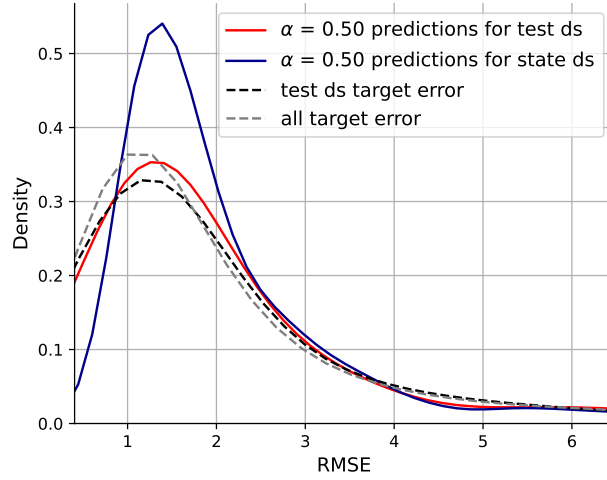


Figure A.8: For the model trained with data prior to MRMS v12.0 (2018-2021), kernel density estimations of the distributions of normalized RMSE (nRMSE) are shown for all target error, the test target error, the test predicted error, and the statewide predicted error.

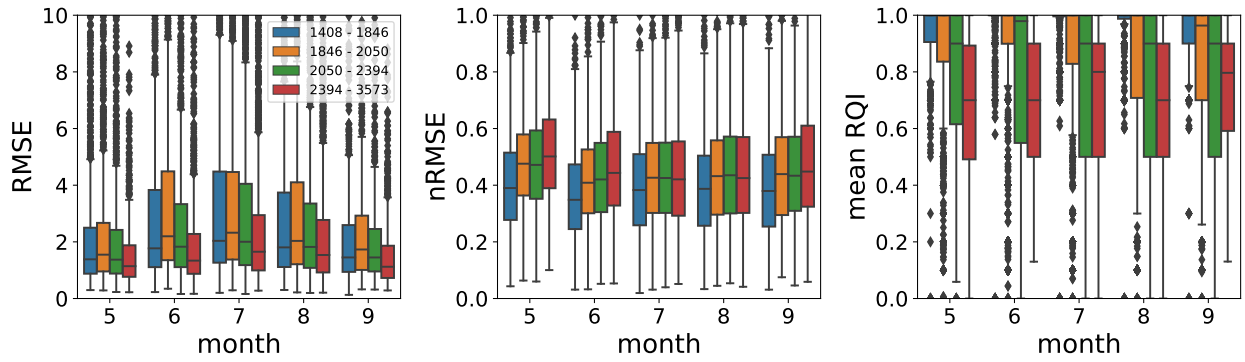


Figure A.9: Difference in error (computed directly from the gage) and mean RQI for each 8-hour sample for all months used in the study. For each month, error bars are separated by the elevation of the gage location.

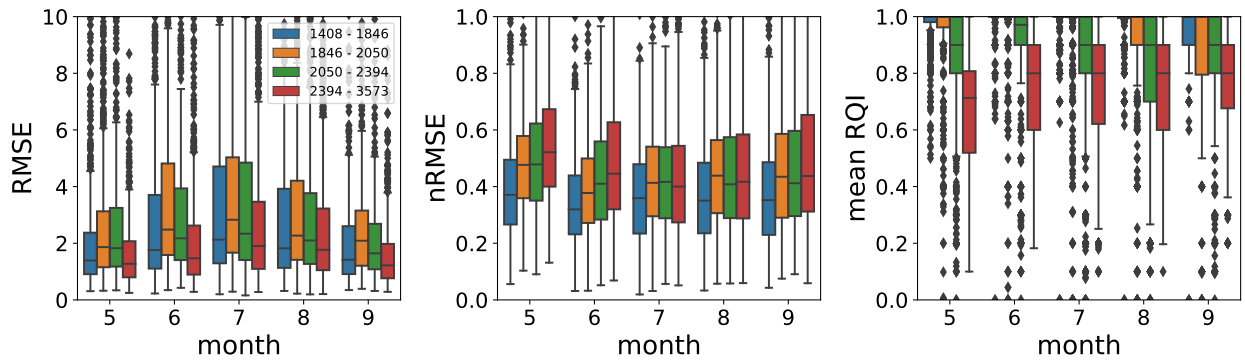


Figure A.10: Difference in error (computed directly from the gage) and mean RQI for each 8-hour sample across all years. For each year, error bars are separated by the longitude of the gage location. Data only includes years 2021-2023, after MRMS version 12.0

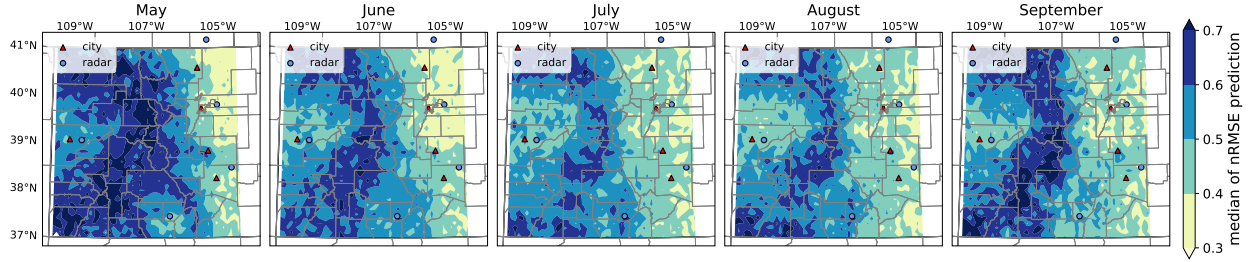


Figure A.11: The median normalized RMSE predictions, separated by month, from the original model for each coordinate, predicted by the $\alpha = 0.50$ model.

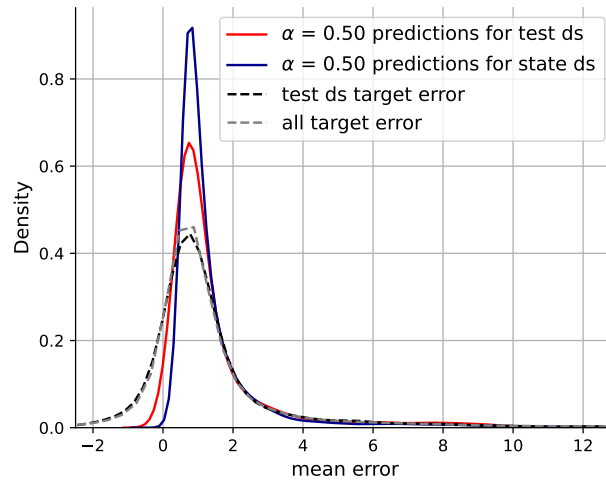


Figure A.12: For the model trained with data prior to MRMS v12.0 and trained to predict mean error, kernel density estimations of the distributions of mean error are shown for all target error, the test target error, the test predicted error, and the statewide predicted error.

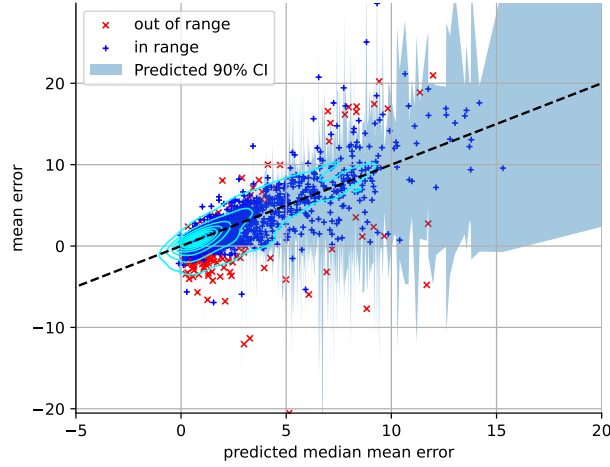


Figure A.13: For the model trained with data prior to MRMS v12.0 (2018-2021) and trained to predict mean error, the target values as a function of the median predicted mean error is shown for the test dataset. Each data point represents an 8-hour sample mean error between gage and MRMS plotted on the predicted. The predicted 90% confidence interval (CI) is plotted. Truth values in range are blue, those that fall outside the cutoff are red. The kernel density estimation for the target values as a function of the median predicted mean error is shown in cyan contours.

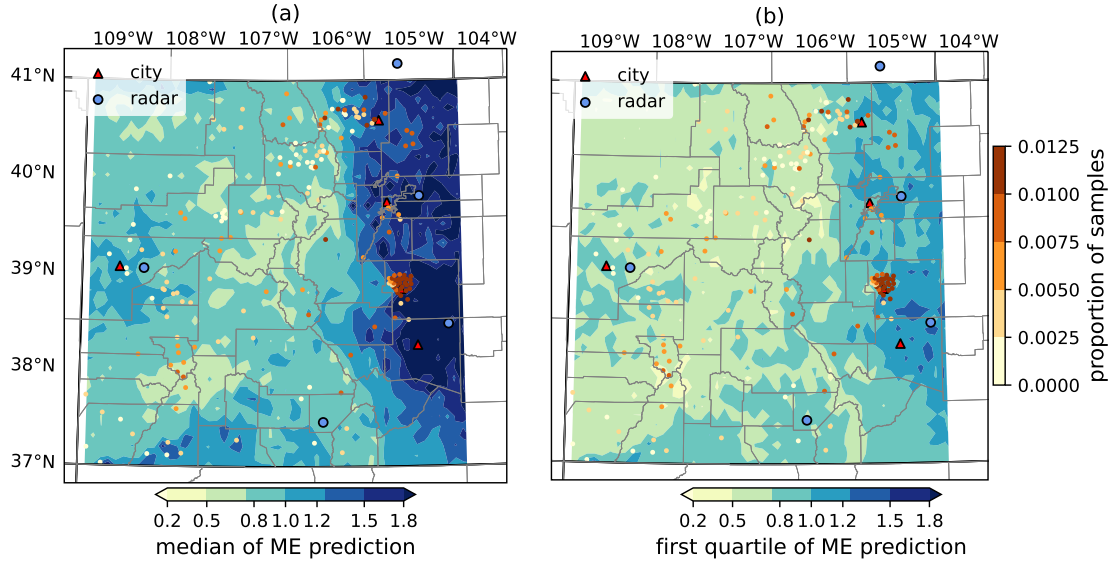


Figure A.14: For the model trained with data prior to MRMS v12.0 (2018-2021) and trained to predict mean error, (a) The median mean error for each coordinate, predicted by the $\alpha = 0.50$ model. (b) The first quartile of mean error for each coordinate, predicted by the $\alpha = 0.50$ model. The gage locations are shown with colors representing the proportion of samples in the labeled dataset sourced from each location.

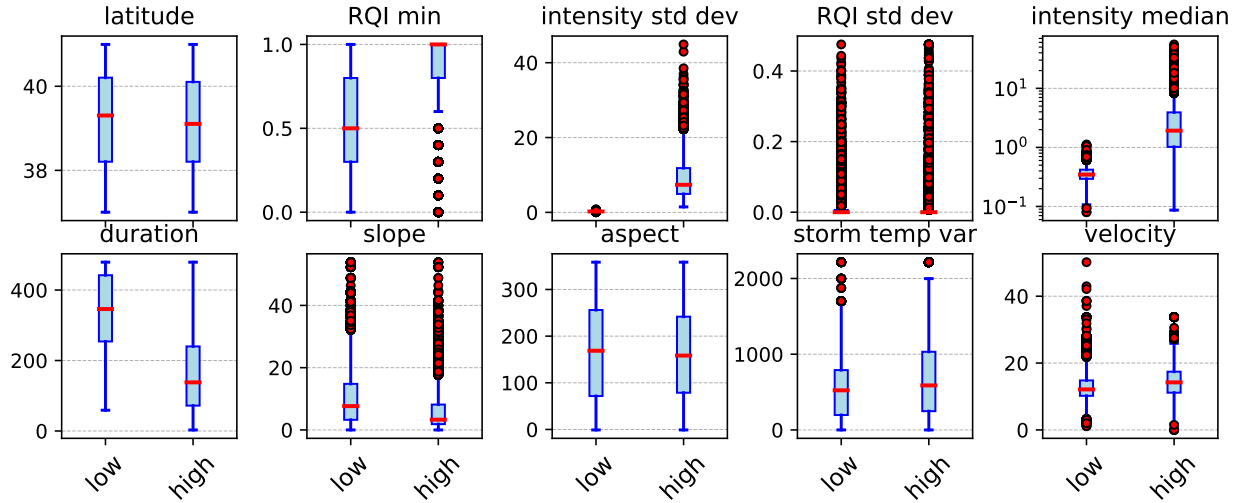


Figure A.15: For the model trained with data prior to MRMS v12.0 (2018-2021), a comparison of feature value distributions is shown for samples with low (1.01 mm/hr) and high (5.43 mm/hr) RMSE predictions for the $\alpha = 0.50$ statewide predictions.

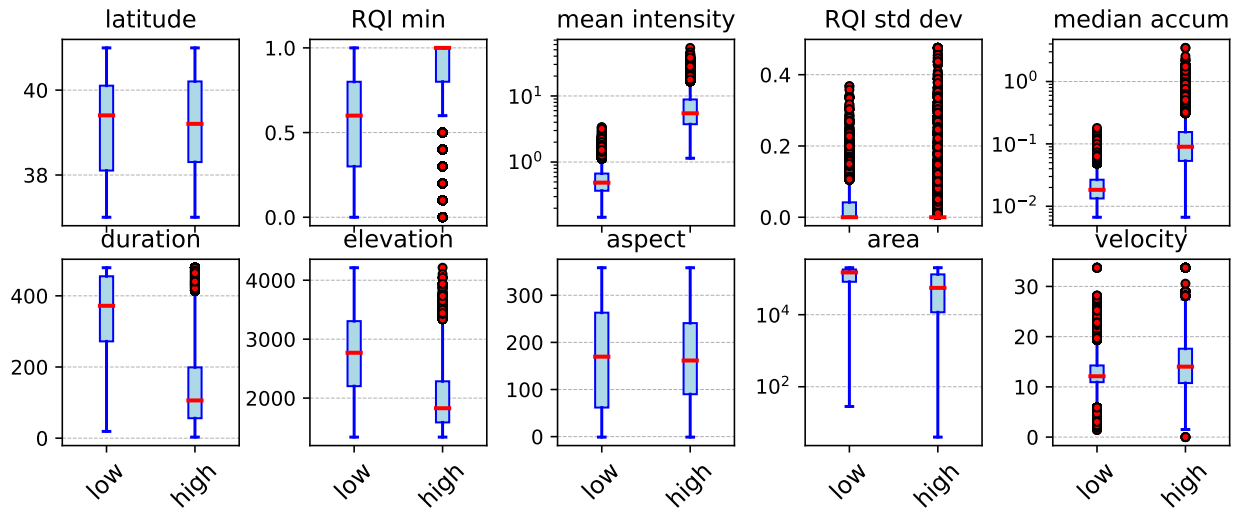


Figure A.16: For the model trained with data prior to MRMS v12.0 (2018-2021) and trained to predict mean error, a comparison of feature value distributions is shown for samples with low (0.55 mm/hr) and high (2.73 mm/hr) mean error predictions for the $\alpha = 0.50$ statewide predictions.

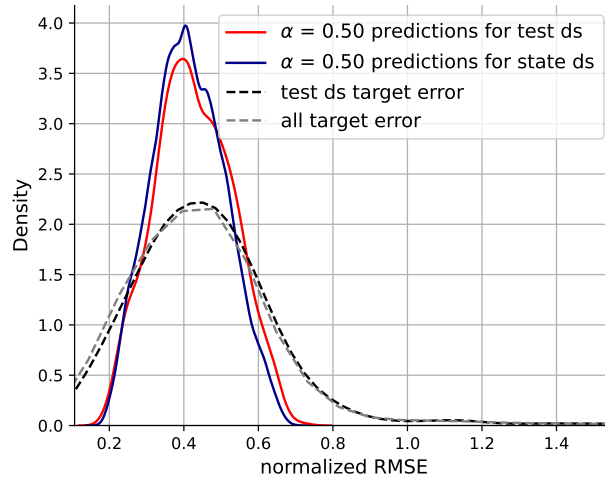


Figure A.17: For the model trained with data prior to MRMS v12.0 (2018-2021) and trained to predict normalized RMSE (nRMSE), kernel density estimations of the distributions of nRMSE are shown for all target error, the test target error, the test predicted error, and the statewide predicted error.

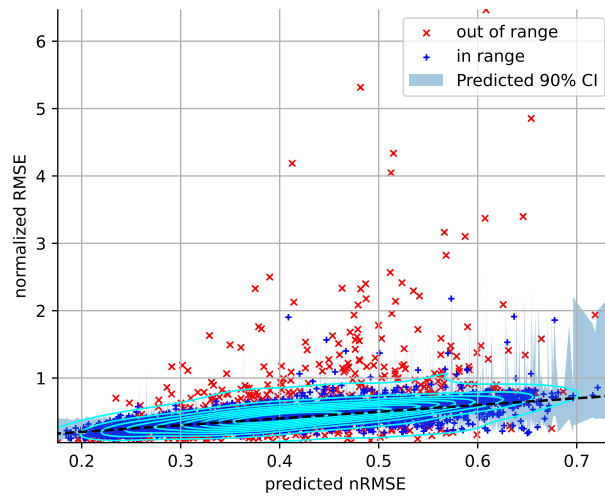


Figure A.18: For the model trained with data prior to MRMS v12.0 (2018-2021) and trained to predict normalized RMSE (nRMSE), the target values as a function of the median predicted nRMSE is shown for the test dataset. Each data point represents an 8-hour sample nRMSE between gage and MRMS plotted on the predicted. The predicted 90% confidence interval (CI) is plotted. Truth values in range are blue, those that fall outside the cutoff are red. The kernel density estimation for the target values as a function of the median predicted nRMSE is shown in cyan contours.

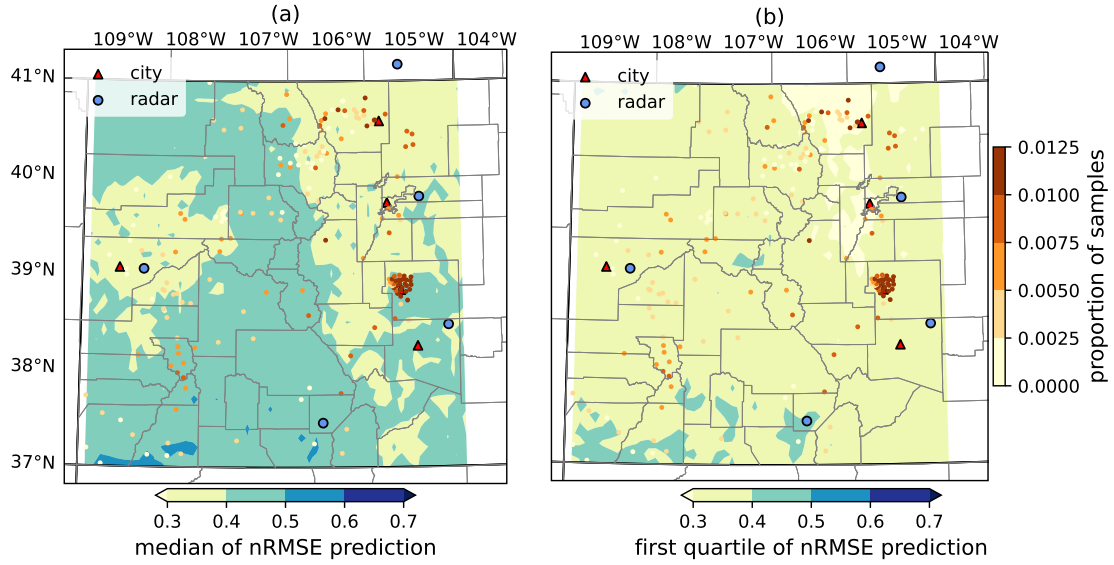


Figure A.19: For the model trained with data prior to MRMS v12.0 (2018-2021) and trained to predict normalized RMSE (nRMSE), (a) The median normalized RMSE for each coordinate, predicted by the $\alpha = 0.50$ model. (b) The first quartile of normalized RMSE for each coordinate, predicted by the $\alpha = 0.50$ model. The gage locations are shown with colors representing the proportion of samples in the labeled dataset sourced from each location.

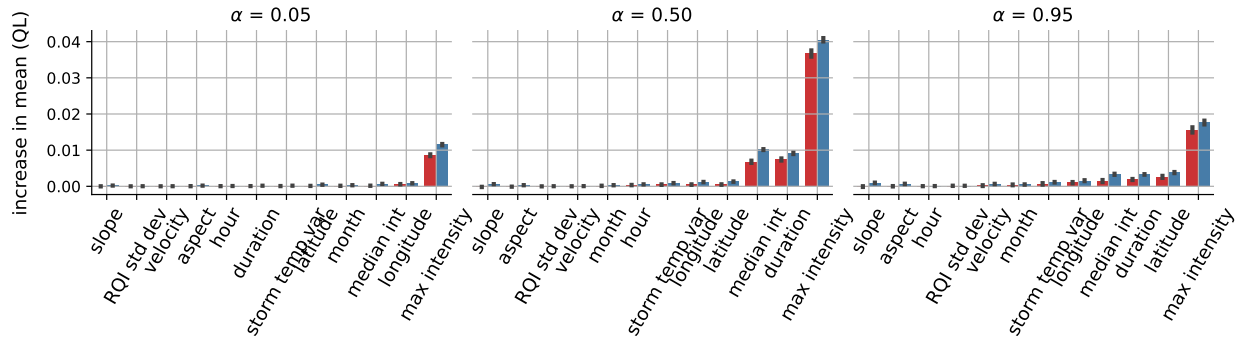


Figure A.20: For the model trained with data prior to MRMS v12.0 (2018-2021) and trained to predict normalized RMSE (nRMSE), permutation feature importance is shown, defined by an increase in mean quantile loss (QL), across the three quantile models. The error bar represents the standard deviation between the 10 permutations. Importance is sorted from left to right by test importance (red). The training importance is shown in blue.

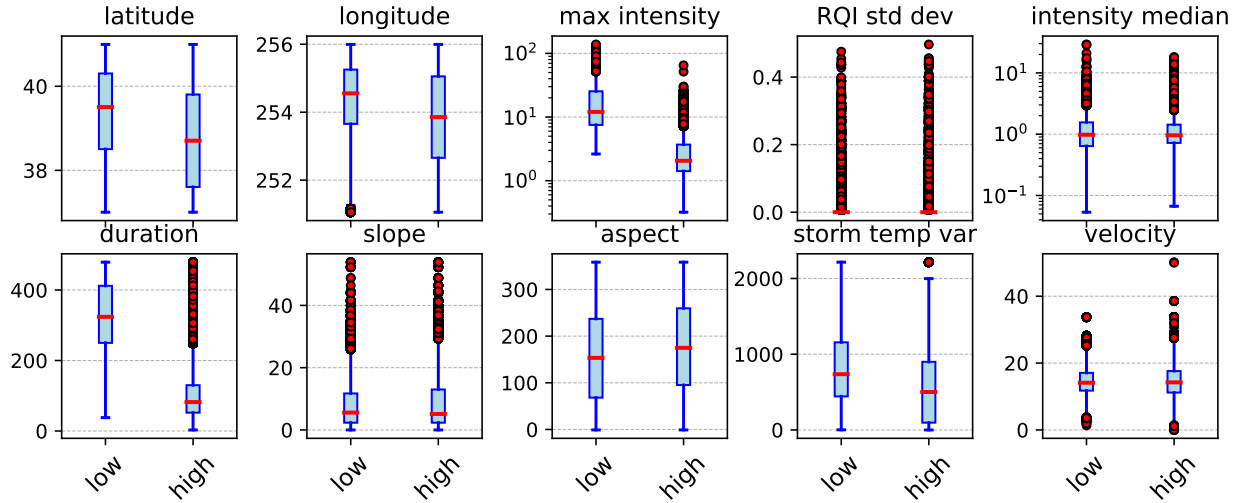


Figure A.21: For the model trained with data prior to MRMS v12.0 (2018-2021) and trained to predict normalized RMSE (nRMSE), a comparison of feature value distributions is shown for samples with low (0.28) and high (0.54) nRMSE predictions for the $\alpha = 0.50$ statewide predictions.

Table A.1: Brief model descriptions for single output experiment

Model	Description
Nearest Neighbors	Samples are classified based on proximity in feature space (Cover and Hart, 1967)
Decision Tree	A hierarchical series of questions to optimally split the classes based on feature values (Breiman et al., 1984)
Random Forest	Decision trees are fit to samples of features and predictions are averaged to improve performance and limit overfitting (Breiman, 2001)
Bagged Tree	A decision tree is trained on a sample of data and predictions are averaged (Breiman, 1996)
Multilayer Perceptron	Fully connected neural network with one or more hidden layer (Pedregosa et al., 2011)
AdaBoost	Decision trees are repeatedly fit to samples of dataset and classification is made based on a weighted average. Weights are increased with each iteration if the learner predicts correctly, and decreased otherwise (Freund and Schapire, 1996)
Logistic Regression	Estimates the probability of a class while minimizing log loss (Géron, 2021)
xgboost	Similar to gradient boosting with various optimizing features (Friedman, 2001)
Gradient Boosting	Decision trees are repeatedly fit to samples and classification is made based on an average. New learner fits to residual error of previous learner (Friedman, 2001)
Support vector	Determines the maximum distance between samples in feature space and assigns classes based on location relative to this distance (Géron, 2021)

Table A.2: Model comparison test results ^a

Model	R2	MAE	MSE
Nearest Neighbors	0.78	0.99	3.97
Decision Tree	0.69	1.15	5.53
Random Forest	0.8	0.96	3.54
Bagged Tree	0.78	1.2	3.94
Neural Net	0.69	1.44	5.51
AdaBoost	0.71	1.43	5.06
linear	0.8	0.97	3.61
xgboost	0.78	1.0	3.86
Gradient Boosting	0.68	1.3	5.69
SVR	0.76	1.03	4.22

^a Acronyms: mean absolute error (MAE), mean squared error (MSE), .
multilayer perceptron (MLP), support vector regressor (SVR)

Table A.3: Hyperparameters tuned and values used

Hyperparameter	Range Tested	$\alpha = 0.05$	$\alpha = 0.50$	$\alpha = 0.95$
learning_rate	0.01 - 0.91	0.17	0.26	0.056
max_depth	2 - 4	2	4	2
min_samples_leaf	1 - 19	4	17	1
min_samples_split	2 - 19	19	18	2
n_estimators	50 - 250	153	221	196
random_state	0 - 999	387	219	637

Table A.4: Test results for RMSE data selection experiments

Experiment	$\alpha = 0.05$ QL	$\alpha = 0.50$ QL	$\alpha = 0.95$ QL	FO
pre_2021	0.088	0.368	0.142	0.109
post_2021	0.113	0.570	0.269	0.112
add_year_as_feature	0.106	0.518	0.224	0.182
post_2021_jja	0.142	0.694	0.293	0.185
jja	0.127	0.605	0.274	0.196
pre_2021_jja	0.141	0.760	0.434	0.433

^a Acronyms: quantile loss (QL), fraction of outliers (FO)

Table A.5: Test results for nRMSE data selection experiments

Experiment	$\alpha = 0.05$ QL	$\alpha = 0.50$ QL	$\alpha = 0.95$ QL	FO
nr_rmse_pre	0.013	0.058	0.034	0.093
nr_rmse	0.015	0.075	0.050	0.128
nr_rmse_post	0.019	0.114	0.067	0.155

^a Acronyms: quantile loss (QL), fraction of outliers (FO)