Assessment of basin vulnerability to post-wildfire hydrologic and water quality effects through a multi-scale framework

by

Carli P. Brucker

B.S., University of Iowa, 2018

M.S., University of Colorado Boulder, 2021

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Civil, Environmental and Architectural Engineering

2023

Committee Members:

Dr. Ben Livneh, Chair

Dr. Fernando L. Rosario-Ortiz, Co-Chair

Dr. Balaji Rajagopalan

Dr. Stephanie K. Kampf

Dr. William C. Becker

Brucker, Carli (Ph.D., Civil, Environmental, and Architectural Engineering)

Assessment of basin vulnerability to post-wildfire hydrologic and water quality effects through a multi-scale framework

Thesis directed by Prof. Ben Livneh and Prof. Fernando Rosario-Ortiz

Wildfires can present challenges for water treatment plants and freshwater systems by increasing sediment, nutrient, and dissolved organic matter (DOM) loads in streamflow, as well as exacerbating flooding through high runoff rates. Although these effects are well known, high variability and data scarcity in post-wildfire *in situ* water quality observations have created challenges in analyses and predictive efforts. This dissertation attempts to increase knowledge of burn impacts on sediment, nutrients, DOM, and runoff through a multi-scale analysis—observing small-scale driving mechanisms as well as broad, large-scale response across multiple watersheds. This framework seeks to provide insights into key factors driving these responses, useful for assessments of watershed vulnerability to wildfire effects based on physiographic features. Underlying processes driving soil and water physical and chemical changes were observed on the small-scale using laboratory-scale wildfire and rainfall simulation experiment apparati, tested on 154 ~300 cm^2 soil samples. This framework observed burn effects both independently, as well as in the context of other key drivers—rainfall intensity and terrain slope—to simulate variable conditions in natural settings. However, the limited variability in a controlled laboratory environment allowed for isolation of each driver's effects. Runoff and sedimentation rates showed significant ($\alpha = 0.05$) monotonic increases from mild to severe burn intensities, while dissolved organic matter and nitrogen concentrations had significant inverse 'U' shaped trends, peaking at ~250°C with values 201-266% of unburned samples. A synthesis

of previous laboratory- and plot-scale wildfire simulation methods was also completed. Here, benefits and limitations of different techniques were discussed, as well as their usefulness in contributing information on burn effects on water quality and supply. Future studies were recommended to prioritize representation of natural processes, incorporation of multiple key drivers, analysis at multiple spatial scales, and uncertainty quantification based on their scale, scope, and subject matter. Next, broad changes in constituent responses after wildfires were assessed for 241 forested watersheds across the U.S. West. Using machine learning and statistical techniques, water quality data aggregated across all burned basins were assessed for significant responses across multiple post-fire years. Analyses applied to a set of 258 paired, unburned basins provided quantifications of baseline natural variability in water quality constituents, allowing for a comparison to burn responses. Inter-basin variability in post-fire responses was also characterized and attributed to physiographic watershed variables and wildfire characteristics. Significant responses were observed in the first 2-3 years post-wildfire for carbon, nitrogen, and phosphorus constituents, and up to six years for sediment constituents and turbidity, with forest cover highlighted as a driver of variability within response magnitudes. By analyzing both small-scale driving mechanisms, as well as broad effects across a large regional scale, this dissertation strives to provide a holistic understanding of wildfires' impacts on watersheds in the U.S. West. Key insights into the duration of elevated responses, as well as key factors exacerbating wildfire effects may help inform water managers' planning and mitigation efforts.

Dedication

To my Opa, Adrian Korpel, who taught me that equal parts of humility, humor, and whimsy, as well as hard work and "Hallelujah" moments, are what make up a whole scientist.

Acknowledgements

I would like to thank my advisors, Dr. Ben Livneh and Dr. Fernando Rosario-Ortiz, for their unending support and mentorship throughout my graduate career. Thank you for always believing in me and pushing me to achieve, even at times I did not believe in myself. With your encouragement and mentorship, I have become a confident, creative, and thoughtful scientist that a younger version of myself could only dream of. I would also like to thank my committee members, for their guidance, as well as the broad perspectives they have contributed to my research. Thank you to my incredibly supportive lab group members and graduate community. I have truly never experienced such a humble, yet passionate group of people, and am honored to call my cohort-mates colleagues and friends in my future career. I would like to additionally acknowledge Aaron Heldmyer and Ariel Retuta in their work on the initial simulation experiment apparati used in my dissertation research. Thank you to the many undergraduate students whose tireless work made this research possible—Cole Pragides, Abbey Turner, Rollin Jones, Yuexuan Meng, Alex Brunson, and Casey Bangs. Your patience and dedication to the completion of this research was inspiring. Special thank you to Claire Butler, for your initiative and leadership in the lab, and for always helping me keep my head on straight. Thank you to Dorothy Noble, Stefan Petersen, and James McCutchan for their diligence in managing the water quality

lab, as well as their tirelessness helping answer my many questions and troubleshooting issues.

Outside of my CU Boulder community, I would like to thank my family for their constant encouragement and support of my scientific endeavors from a young age. Special thanks to Amanda, for both the sisterly and statistical advice. Thank you to my partner, Henry, who has always been my biggest cheerleader, and has carried me through my toughest times. Thank you to my friends, both old and new, whose unwavering support was vital to the success of my graduate career. Thank you to my Day One, Melanie—our shared experiences have created the strongest bonds. Finally, thank you to my roommates, Kristen and Vinitha, who have grown with me as people and scientists throughout the last five years, all while having a silly goose time.

Finally, I gratefully acknowledge the support of the U.S. Environmental Protection Agency; the Cooperative Institute for Research in Environmental Sciences (CIRES) Graduate Student Research Award, part of the NOAA Cooperative Agreement with CIRES, Grant # NA17OAR4320101; the 2020 CIRES Innovative Research Program; and the Western Water Assessment NOAA Grant # NA21OAR4310309, 'Western Water Assessment: Building Resilience to Compound Hazards in the Inter-Mountain West'.

Contents

Appendix

Tables

Figures

Chapter 1

Introduction

1.1 Overview

Wildfires are a natural disturbance mechanism which support the long-term health of forested ecosystems (He et al., 2016), but also can degrade stream water quality and alter runoff generation mechanisms. Wildfire-driven increases in sediment, dissolved organic matter (DOM), nutrients, and heavy metals (Bladon et al., 2014; Hohner et al., 2019; Jian et al., 2018; Rhoades et al., 2019; Robichaud, 2005) can necessitate investments in infrastructure and altered methods of treatment in water treatment plants (Becker et al., 2018; Hohner et al., 2019; J. Raseman et al., 2017; Murphy et al., 2015; Writer et al., 2014), diminish reservoir storage due to sediment filling, and disturb freshwater ecosystems (Bladon et al., 2014; Moody and Martin, 2009). These impacts are illustrated in Figure 1.1. These effects occur immediately after a wildfire, but impacts can persist for up to 10 years

(Smith et al., 2011). Additionally, increased post-fire runoff generation can produce high peak flows, increasing flood risks (Brogan et al., 2017).

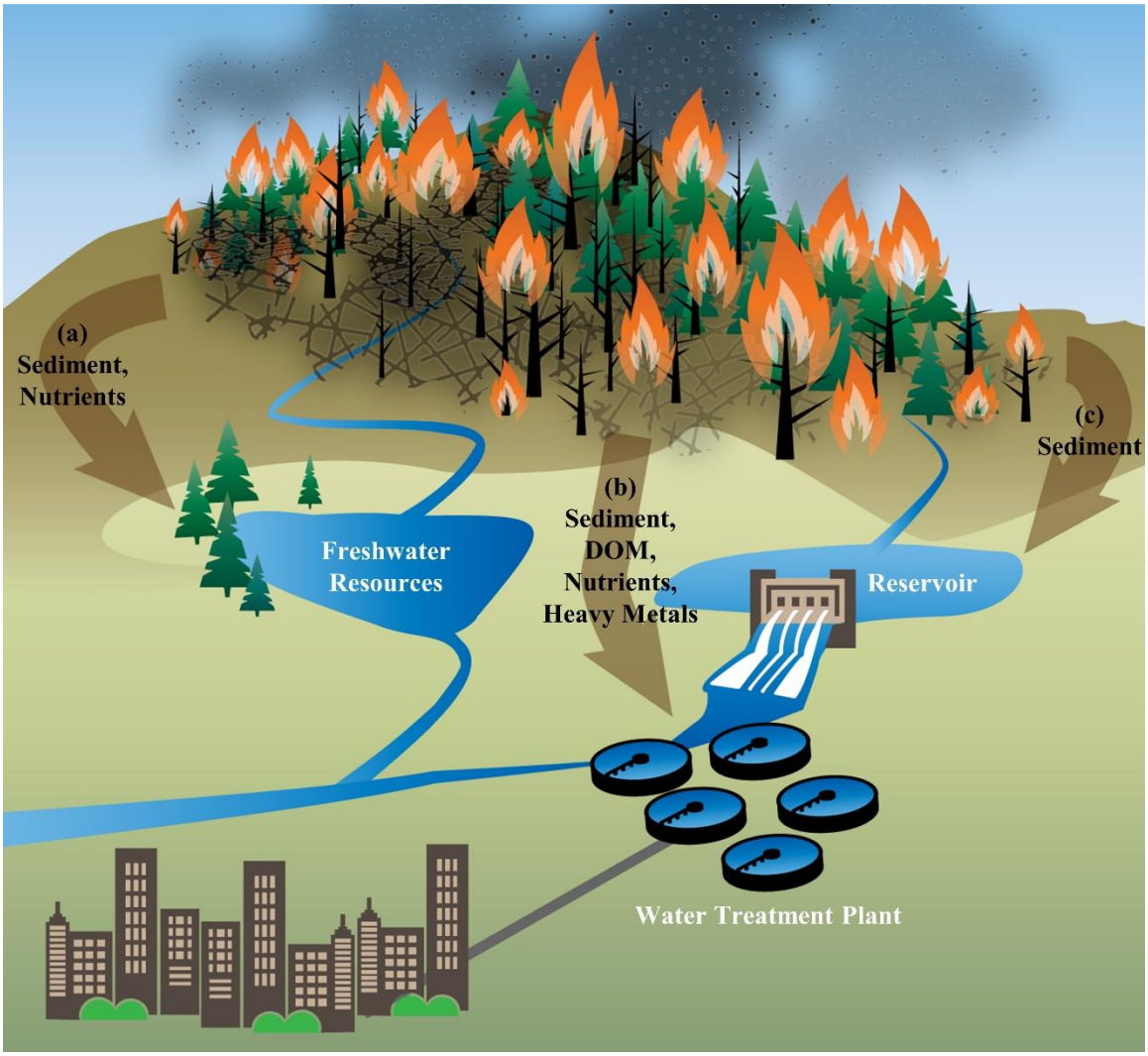

Figure 1.1: Conceptual illustration of the impacts of wildfire on water quality and supply. The constituent fluxes in streams are driven by burning effects, affecting human and natural systems. Increased sediment and nutrient concentrations: (a) drive eutrophication and disturb freshwater resources, increased sediment, DOM, nutrient, and heavy metal concentrations, (b) can exceed water treatment plant treatment capacities, and increased sediment loads, and (c) affect reservoir storage capacity.

In recent decades, an observed increase in wildfire size, frequency, and severity has been observed in certain forested regions—a trend predicted to continue (Edenhofer et al., 2015; Marlon et al., 2009; Sommerfeld et al., 2018; Spracklen et al., 2009). For example, the

mean annual burn area in the Western U.S. has doubled since 1984 and is projected to have a 24-169% increase in mean burn area by midcentury (Harvey, 2016; Liu et al., 2010; Spracklen et al., 2009; Yue et al., 2013). Current understanding of wildfire effects on water quality and supply is incomplete, such that research is critically needed to assist water managers in adaptation and mitigation strategies (Bladon et al., 2014; Murphy et al., 2015; Robichaud, 2005).

Hinderances in the collection of post-wildfire *in situ* data—unstable terrain and road closures immediately after wildfires, lack of comparable pre-burn control data, and high natural spatial and temporal variability—have contributed to a lack of knowledge of wildfire effects (Hohner et al., 2019; Murphy et al., 2015; Writer et al., 2014). Additionally, post-wildfire response can vary regionally due to differences in soil and vegetation regimes, as well as climate (Hogue and Inglett, 2012). This means that observations from post-wildfire settings often provide insights which are not directly transferable to other geographic areas. Little information exists about post-wildfire water quality response, in particular, with most studies primarily focusing on DOM and nutrients.

This dissertation takes a multi-pronged approach to assessing key hydrologic and water quality driving mechanisms in a post-wildfire environment, helping to fill this gap in knowledge and advance prediction efforts. Laboratory simulation experiments conducted on soil samples, as well as small- and large-scale statistical and physical modeling techniques will all be used as frameworks to observe burn effects—both independently and in the context of complex and varied post-wildfire systems. These different scales of analysis are shown in Figure 1.2. This holistic analysis will provide a unique perspective into wildfire effects at multiple physical and temporal scales, creating a robust understanding of linkages between scales.

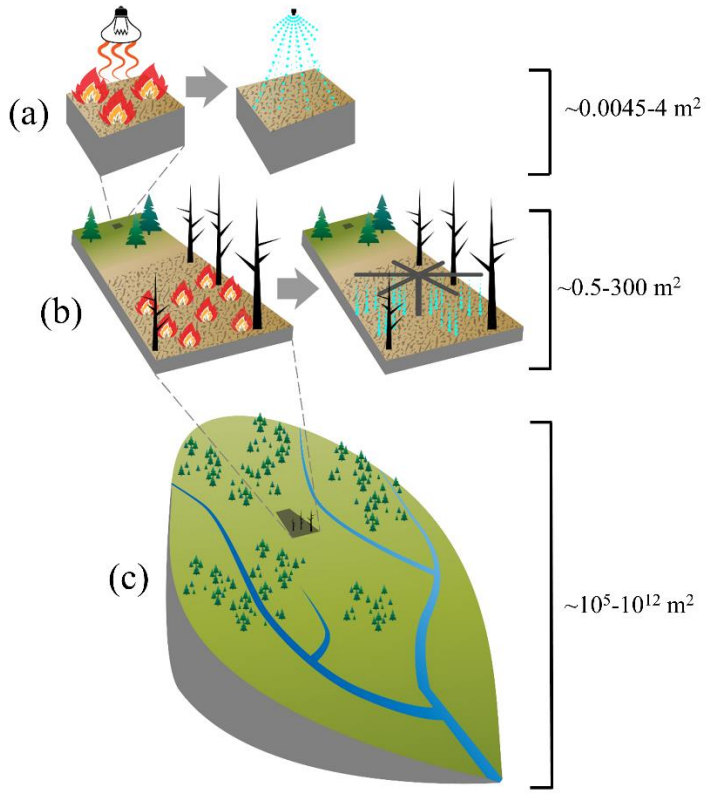

Figure 1.2: Illustration of laboratory-, plot-, and catchment-scale post-wildfire study areas, with (a) simulation techniques being used at the laboratory scale (~0.0045-4 m^2) for wildfire and rainfall simulation, respectively; (b) a prescribed burn and rainfall simulator being used for plot-scale (~0.5-300 m^2) wildfire and rainfall simulation, respectively; and (c) a catchment (~10^5-10^{12} m^2) affected by wildfire where physical and statistical models are applied.

The following sections discuss common terminology used throughout this dissertation relevant to the field of wildfire and wildfire simulation studies (*Section 1.2*), background information about known wildfire effects on water quality and hydrology and their implications for human and freshwater systems (*Section 1.3*), and finally research questions and hypotheses explored by this dissertation (*Section 1.4*).

1.2 Terminology

A variety of wildfire characterization terms are commonplace across studies, despite calls for standardization in recent decades (Bento-Gonçalves et al., 2012; Keeley, 2009; Lentile et al., 2006; Parson et al., 2010). In this dissertation, *wildfire* is used to describe fires which occur in a natural environment, though *bushfire* is an interchangeable term common in Australia. Following are the definitions for *wildfire intensity* and *wildfire severity* as used in this dissertation, as well as other relevant wildfire and rainfall simulation terms.

Wildfire Intensity is typically a quantitative characterization of energy, i.e., the amount or rate of fuel burned (Hogue and Inglett, 2012; Moreno and Oechel, 1989) or peak temperature (Blank et al., 1994; Cancelo-González et al., 2012a; Chandler et al., 1983; Keeley, 2009; Lentile et al., 2006; Stoof et al., 2011). This allows for explicit quantifications of mild to severe burn intensities, however the lack of a standardized scale across studies makes cross-comparisons difficult (Friedrich et al., 2018; Hogue and Inglett, 2012; Hohner et al., 2019; J.-J. Wang et al., 2015).

Wildfire Severity is typically a visual characterization of the response of an ecosystem (i.e. vegetation, soil, water systems, and atmosphere) to fire (Bento-Gonçalves et al., 2012; Jian et al., 2018; Parson et al., 2010), such as ash color or amount of biomass consumed (Keesstra et al., 2014; Nyman et al., 2014). Similar to *wildfire intensity*, a standard definition of a wildfire severity range from low to high does not exist across studies (Hardy, 2005; Keeley, 2009; Moody et al., 2013).

Laboratory-Scale Simulations of wildfire and rainfall are typically applied to smaller soil samples (~0.0045-4 m^2) (Hohner et al., 2019b; Kral et al., 2015; X. Wang et al., 2015) using experimental apparati. These analyses occur either inside a laboratory or in a

designated outdoor setting (Badía-Villas et al., 2014; Cancelo-González et al., 2015; Keesstra et al., 2014), limiting the size of samples to the dimensions of the equipment used.

Plot-Scale Simulations of wildfire and rainfall typically occur over a larger area of ground or hillslope (~0.5-300 m^2)—either undisturbed or where a natural wildfire has already occurred (Hester et al., 1997; Johansen et al., 2001; Wilson, 1999). If undisturbed, a prescribed burn is typically applied to the area, then rainfall simulators are used to generate runoff (Emmerich and Cox, 1992; Ferreira et al., 2008). These experiments are also referred to as *field-scale* or *hillslope-scale* experiments, but in this review only the term plot-scale is used.

Rainfall intensity is the ratio of the total rain depth to the duration of rainfall, typically in minutes, hours, or days.

Slope is the average topographical inclination or gradient across a terrain surface.

1.3 Background

Wildfires can drive a wide range of responses in runoff, sedimentation, dissolved organic matter, nutrients, and heavy metals (Bladon et al., 2014; Smith et al., 2011). A summary of these responses, as well as the unique mechanisms involved in producing them, as described in the following section.

1.3.1 Runoff Generation

Post-fire peak flows have been reported in ranges of 5—870 times greater than pre-fire flows (Bladon et al., 2014). Some studies point to burn induced increases in soil water repellency in post-wildfire environments as the main driver of this increased runoff (Benavides-Solorio and MacDonald, 2001; Doerr et al., 2006; Keesstra et al., 2014b; Lane et al., 2006; Larson-Nash et al., 2018). Water repellency has been observed 1-3 cm below the soil surface

after wildfires, likely due to the volatilization of certain organic compounds in litter and topsoil which infill lower pores, as well as the polymerization of organic molecules and melting and redistribution of waxes (Doerr et al., 2006; Larson-Nash et al., 2018). The thickness of the ash layer covering the soil surface can also affect post-fire runoff rates and is key in initial hydraulic processes (Ebel et al., 2012a; Ice et al., 2004; Stoof et al., 2010). Ash can absorb and store up to 99% of rainfall, thereby limiting initial runoff generation (Ebel et al., 2012b; Ebel and Moody, 2017). The combined effects of these complex mechanisms result in highly variable runoff generation with increasing burn severity, heavily dependent on storm type and duration (Doerr et al., 2006; Kampf et al., 2016a).

1.3.2 Sedimentation

Wildfire-driven increases in suspended sediment in streamflow have been reported from 1 to up to 1459 times pre-disturbance levels (Smith et al., 2011). Downstream impacts of elevated sediment loading include strain on water treatment plants, reductions in reservoir storage capacity, and disruption to freshwater ecosystems (Bladon et al., 2014; Moody and Martin, 2009; Writer et al., 2014). Increased sediment loads can require capacity increases in water treatment plants' infrastructure and monitoring (Bladon et al., 2014; Emelko et al., 2011; Writer et al., 2014). Sediment-driven reservoir filling reduces the storage capacity, limiting available water for municipalities, and is expensive to mitigate using dredging or tunnels (Fan Jiahua and Morris Gregory L., 1992; Minear and Kondolf, 2009; Moody and Martin, 2004). Finally, increased turbidity from wildfire-driven sediment limits sunlight needed for photosynthesizing organisms that produce oxygen and form the base of aquatic food chains, resulting in fish death (Kemp et al., 2011; O'Laughlin, 2005; Rieman and Clayton, 1997).

Sedimentation is enhanced by erosional effects due to loss of vegetation (both ground and canopy cover) and root structure (Kampf et al., 2016; Larson-Nash et al., 2018; Robichaud et al., 2016), as well as an accumulation of combusted vegetation and soil organic material (i.e., ash) (Lane et al., 2006). High runoff rates in post-fire settings driven by increased soil water repellency (Benavides-Solorio and MacDonald, 2001; Lane et al., 2006; Larson-Nash et al., 2018) can further exacerbate erosional effects and increase constituent transport. Rill erosion, for example, expands channel networks through high-flow rate runoff flowing through streamlets, the eroded soil mobilized into suspended sediment (Robichaud et al., 2016). Thought ash typically stores initial precipitation, subsequent or larger storms results in ash saturation and high runoff rates mobilize ash and soil particles downstream—contributing to high sedimentation rates (Ebel et al., 2012; Woods and Balfour, 2008).

1.3.3 Dissolved Organic Matter

Post-fire runoff DOM response varies widely across studies, from slight decreases in concentration to levels in the 95th percentile of pre-fire conditions (Bladon et al., 2014; Cawley et al., 2017; Hohner et al., 2016; Ice et al., 2004; Meixner, 2004; Murphy et al., 2015; Smith et al., 2011; X. Wang et al., 2015; Wilkerson and Rosario-Ortiz, 2021). Downstream, DOM is the is the main substrate in the formation of carcinogenic disinfection byproducts (DBPs) during the chlorination stage of water treatment (X. Wang et al., 2015). Elevated levels can require water treatment plants to implement expensive alternate disinfectants, precursor (i.e., DOM) and DBP removal strategies, or even force them to shut down (Hohner et al., 2016; Hua and Reckhow, 2007; X. Wang et al., 2015).

Alterations in the load and chemistry of DOM in post-fire settings is driven by thermal reactions during burning (Thurman et al., 2020; J.-J. Wang et al., 2015b). However, isolating wildfire-driven DOM response is challenging because of natural background sources of DOM, contributed by other hydrological, topographical, physicochemical, and microbiological processes (Hohner et al., 2019a; J.-J. Wang et al., 2015a). DOM levels monotonically decrease (Wieting et al., 2017) or remain the same (Badía-Villas et al., 2014) with increasing burn severity in some studies, while others report peak DOM concentrations under moderate burn severity conditions (Abraham et al., 2017; Cancelo-González et al., 2013; Cawley et al., 2017; Hohner et al., 2019b; Jian et al., 2018). Lower DOM concentrations are frequently reported at high burn intensities, likely due to DOM vaporization (transformation of organic material into carbon dioxide and water vapor) at extreme temperatures (Ice et al., 2004; Rhoades et al., 2019; Wieting et al., 2017).

1.3.4 Nutrients

Wildfire-driven increases in the nutrients phosphorous and nitrogen have been reported up to 250- and 400-times pre-burn conditions, respectively (Bladon et al., 2014; Rhoades et al., 2019; Smith et al., 2011). These dramatic increases in nutrients can lead to eutrophication in aquatic ecosystems (Conley et al., 2009). Eutrophication is a process where excess nutrients lead to accelerated growth of aquatic plants and benthic communities, but also algal blooms which produce toxins and deplete oxygen from the ecosystem (Bladon et al., 2014; Conley et al., 2009; Spencer et al., 2003). These effects, in addition to the impacts of sediment discussed earlier, result in the mortality of fish and other aquatic species, as well as loss of biodiversity (Conley et al., 2009; Meixner, 2004; Spencer et al., 2003).

Many complex mechanisms contribute to changes in nutrients in post-fire settings (Bladon et al., 2014; Certini, 2005; Choromanska and DeLuca, 2002; Ice et al., 2004; Simon et al., 2016). However, studies have generally shown higher rates of nitrogen concentrations after burning due to deposition and changes in chemical structure (Bladon et al., 2014; Certini, 2005; Ice et al., 2004). Increases in phosphorous can result from constituent transport through increased post-wildfire sedimentation rates (Emelko et al., 2011).

1.3.5 Heavy Metals

Though heavily dependent on of the nature of the forest and climatic conditions, increases in dissolved heavy metal concentrations in runoff driven by wildfires can range from 2 to at least 2500 times pre-disturbance levels (Abraham et al., 2017; Ranalli and Stevens, 2004). These increases can exceed federal regulations, posing health hazards as some metals are carcinogenic or can cause anemia or heart failure (Chowdhury et al., 2016). Increases in Mg are of particular concern due to toxicity, volatility, and persistence in the environment (Huang et al., 2011; Wiedinmyer and Friedli, 2007). Fe, Mn, As, Cr, Al, Ba, and Pb have all been observed at statistically significantly higher concentrations in post-fire runoff (Smith et al., 2011). Moderate to high intensity fires can alter soil properties, releasing sequestered metals in organic matter which are transported downstream by high post-fire flowrates (Abraham et al., 2017; Aiken et al., 2011).

1.4 Research questions and objectives

This dissertation aims to assess basin vulnerability of post-wildfire hydrologic and water quality effects by providing in-depth insight into the interaction of key driving mechanisms. We analyze these hydrologic and water quality drivers in the context of laboratory-scale experimental simulations, basin-scale data-driven models, and physical

models run at both the small- and basin-scales. This allows for understanding of core burning, runoff transport, and soil hydrologic mechanisms interact on the smallest scale, as well as how these mechanisms are compounded at the basin-scale with larger-scale hydrologic processes and different regional characteristics. Following are the key research questions which I attempt to answer with each chapter, as well as specific objectives:

Chapter 2: What are the key characteristics in the design of wildfire and rainfall simulation experiments which future researchers should consider?

- Objective 1: Identify existing wildfire and rainfall simulation experiment literature and describe existing methodologies.
- Objective 2: Define the state-of-the-art of existing simulation techniques, highlighting strengths and weaknesses and providing recommendations for future wildfire simulation studies.

Chapter 3: Can unique laboratory-scale wildfire and rainfall simulation experiments be developed to quantify key hydrologic and water quality responses?

- Objective 1: Design and construct laboratory-scale wildfire and rainfall simulators to capture post-wildfire responses of runoff, sediment, dissolved organic matter, and nitrogen generation, as well as turbidity, through incorporation of three key drivers with multiple testing increments.
- Objective 2: Validate simulators' ability to observe small-scale burn effects on hydrologic and water quality response, by comparing experimental results to previous studies and characterizing uncertainty.

Chapter 4: Can machine learning techniques produce useful assessments of post-wildfire water quality response using *in situ* and satellite-derived data?

- Objective 1: Develop a data-driven modeling framework for wildfire-affected basins across the U.S. West, using burn severity and other geophysical data as predictor variables.
- Objective 2: Evaluate key geophysical variables driving post-wildfire response through characterizing inter-basins response variability with watershed and wildfire covariates.

Chapter 2

A review of simulation experiment techniques used to analyze wildfire effects on water quality and supply

The following chapter was published in Environmental Sciences: Processes & Impacts journal themed issue Wildfires – influence on air, soil and water:

Brucker, C.P., Livneh, B., Minear, J.T., Rosario, F., 2022. A review of simulation experiment techniques used to analyze wildfire effects on water quality and supply. Environmental Science: Processes & Impacts.

2.1 Overview

Laboratory- and plot-scale wildfire and rainfall simulation experiments offer an alternative analytical technique for estimating wildfire effects on water quality and supply (Cancelo-González et al., 2012a; Cotrufo et al., 2016; Hohner et al., 2019b, 2019b; Kampf et al., 2016; Robichaud, 2005; Wilkerson and Rosario-Ortiz, 2021). These studies primarily advance understanding of burn effects on small-scale soil and water physical and chemical properties in a controlled setting. Simulation experiments have many advantages including overcoming logistical challenges of collecting *in situ* wildfire data, reducing the high spatial variability observed in natural settings (i.e., the heterogeneity of burn intensity and the underlying vegetation and soil properties), and controlling the magnitude of key drivers of

wildfire impacts. In sum, simulation experiments allow for more direct attribution of water quality and quantity responses to specific drivers than experiments conducted *in situ*. However, this type of analysis also presents new limitations including the observation of only local-scale processes, the potential misrepresentation of natural settings (i.e., lack of spatial variability in vegetation, soil structure, burn intensity, etc.), uncertainty introduced through experimental error, and subsequent challenges in upscaling results to larger scales relevant for water management.

This dissertation chapter provides a comprehensive review which addresses the critical knowledge gap of wildfire simulation techniques used in assessing postfire impacts on water quality and quantity. Our assessment includes both laboratory and plot-scale techniques with burn and rainfall simulation components. Studies included focus on advancing understanding of changes in chemical and physical properties of soil, as well as subsequent runoff changes. This review focuses on simulation techniques in addition to simulated data, with the goal of providing a foundation of knowledge for the design of future simulation experiments.

2.2 Background

Existing reviews on wildfire effects on water quality and supply primarily summarize the state-of-the-art of *in situ* studies with observations of hydrology (DeBano, 2000; Ebel and Moody, 2017; Robinne et al., 2016; Shakesby and Doerr, 2006; Stavi, 2019; Williams et al., 2014; Wu et al., n.d.), sediment transport and erosion (Moody et al., 2013; Moody and Martin, 2009; Sankey et al., 2017; Shakesby, 2011; Smith et al., 2011), and streamflow concentrations of DOM and nutrients (i.e. nitrogen and phosphorous) (González-Pérez et al., 2004; Hohner et al., 2019a; Holden et al., 2012; Ice et al., 2004;

Martín and Vila, 2012; Olefeldt et al., 2013; Wang et al., 2012), as well as heavy metals (Abraham et al., 2017; Bladon et al., 2014). A novel contribution of this review is the focus on methodological techniques—specifically, simulation experiments—used to generate and collect wildfire response data, rather than the data themselves (Abraham et al., 2017; DeBano, 2000; Smith et al., 2011). We include an analysis of common hydrologic and water quality response data from reviewed simulation experiment studies—included in *Appendix A*. However, this is not the focus of this review due to studies' wide ranges of study goals, temporal and physical scales, and key experimental factors—making useful cross-study comparisons difficult.

One notable exception is Ferreira et al., 2008, which discusses limitations of common methods and techniques used to analyze hydrologic and erosional responses in post-wildfire settings, from laboratory to catchment scales. However, while Ferreira et al. (2008) does cover rainfall simulation methods, the present manuscript is the first to provide in-depth review of wildfire simulation techniques together with associated rainfall simulators. Where Ferreira et al. (2008) focus on hydrologic and sedimentation impacts of fire, this review also focuses on postfire water quality impacts including DOM, nutrients, and heavy metals. Finally, as over a decade has passed since the review of Ferreira et al. (2008), the time is ripe for a fresh perspective that considers more recent research on simulation experiment techniques, drawing upon a larger pool of studies.

The studies reviewed in this paper were compiled using the search strings of "wildfire experiments", "wildfire laboratory simulations", and variations of those in the Google Scholar search engine. These were further filtered by laboratory- and/or plot-scale studies, and those which had a hydrologic or water quality component, as opposed to studies examining post-fire air quality, wildfire behavior, ecosystem restoration, etc. From

these studies, only those focusing on burn effects of natural fuels, (i.e., litter and woody biomass) rather than on anthropogenic post-wildfire contaminants (i.e., plastics and metals) or the burning of human infrastructure were included. Additional studies were identified through the internal references across the initial set of publications. In total, 39 studies were included—23 had a wildfire simulation component, 27 had a rainfall simulation component, and 10 studies had both.

2.3 Experimental Techniques to Observe and Quantify Post-Fire Impacts

The strengths of wildfire and rainfall simulation experiments are in the reduction of the logistical challenges associated with collecting *in situ* data, the ability to generate baseline pre-fire data as well as replicate samples, and to provide greater control over factors in post-fire systems (Cancelo-González et al., 2012; Keesstra et al., 2014c; Murphy et al., 2015; Smith et al., 2011). Greater control over the timescale and number of samples collected is conducive to isolating wildfire responses from background sources (Moody et al., 2013; Murphy et al., 2015). Limitations of simulation experiments include deviation from natural conditions, e.g. vegetation characteristics and lateral flow paths, as well as significant differences in hydrologic and chemical processes at different scales (Ferreira et al., 2008; Keesstra et al., 2014c; Robichaud, 2005a). Interpretation of results from laboratory- and plot-scale simulation experiments for catchment scale impacts should incorporate uncertainties, such as landscape heterogeneity, associated with the upscaling process. The following section will first discuss key differences in plot- vs. laboratory-scale simulation techniques, then present the strengths and weaknesses of specific methodologies for wildfire and rainfall simulation experiments.

2.3.1 Laboratory- vs. Plot-Scale Simulation Techniques

As seen in Table 2.1, different benefits and limitations exist for plot- and laboratory-scale simulation experiment techniques. Plot-scale experiments are conducted on *in situ* hillslopes or plots and are generally subject to higher spatial variability in soil properties and vegetation, as well as spatial and temporal variability in burning and rainfall, than laboratory-scale experiments (Balfour and Woods, 2008). Though high variability can hinder the attribution of responses to drivers, this type of analysis also minimizes area-to-edge ratios (i.e. limiting edge effects) and allows for larger-scale vegetation, such as trees, and hydrologic processes to be captured (Ferreira et al., 2008; Robichaud et al., 2016). Larger-scale hydrologic processes, i.e. rill formation and other erosional processes, can be key drivers of post-wildfire sedimentation and runoff responses (Robichaud et al., 2016), and thus important to consider in post-wildfire analyses. A wide range of replicates are typically tested in plot-scale analyses, from 0 to approximately 15 (Robichaud et al., 2016; Rosso et al., 2007).

Table 2.1: Comparison of pros and cons of plot- and laboratory-scale simulation experiments based on the following factors: spatial and temporal variability, control over environmental factors, representation of the natural environment, edge effects, and subsurface flows. An "N/A" entry indicates that no significant pro or con exists for that category.

	Plot-Scale (~0.5-300 m²)		**Laboratory-Scale (~0.0045-4 m²)**	
	Pros	*Cons*	*Pros*	*Cons*
Spatial and temporal variability	Higher variability in climatic, burn, rainfall, and soil and vegetation factors similar to the natural environment	Higher variability can hinder attribution of responses to drivers	Lower variability allows for greater attribution of responses to drivers	Lower variability not as representative of the natural environment
Control over environment and experimental factors	N/A	Little control over climatic conditions and other factors in experimental environment	Precise control over climatic, burn, and rainfall factors, as well as soil and vegetation characteristics	N/A

Representation	*Wildfire intensity*	Able to simulate high intensity fires	Trade-off between scale and fire intensity (i.e., high intensity typically applied on smaller scales)	N/A	Typically, only low to moderate intensity fires simulated
	Hydrologic and erosional processes	Allows for some larger-scale erosional processes (e.g., rill formation)	Does not capture basin-scale erosional processes such as streambed erosion	N/A	Larger-scale hydrologic processes typically not represented
	Vegetation	Larger-scale vegetation such as trees represented	N/A	N/A	Typically only small-scale vegetation represented
	Soil structure	Soil minimally disturbed or not at all	N/A	N/A	Soil structure may be compromised during excavation or purposefully homogenized
Edge effects		Minimized area-to-edge ratio	N/A	N/A	Greater edge effects
Subsurface flows		N/A	Typically does not allow for measurements of subsurface flows	Allows for the analysis of infiltration and lateral subsurface flows	N/A

Laboratory-scale experiments typically use excavated, intact soil cores or homogenized samples which have low spatial variability, due in part to their smaller size (Busse et al., 2010; Ferreira et al., 2008; Robichaud et al., 2016; Stoof et al., 2011; Wieting et al., 2017). Samples are typically excavated by hammering lysimeter boxes or PVC cylinders into soil (Cancelo-González et al., 2013; Wieting et al., 2017) or collected from loose soil and litter on the ground—often homogenized to minimize variability (Hohner et al., 2019; Kibet et al., 2014; Rosso et al., 2007; Stoof et al., 2011). Laboratory-scale samples are often less representative of natural conditions in terms of wildfire intensity, erosional processes, vegetation type, and soil structure, but allow for precise control over experimental factors. This is conducive to more direct attribution of the impacts of burning and other factors in the system (e.g. vegetation, rainfall intensity, etc.) to observed responses (Keesstra et al., 2014c). Another key feature of laboratory-scale analyses is they allow for the measurement of infiltration through the bottom of samples and subsurface

later flows, which is difficult or impossible in plot-scale *in* situ analyses (Cancelo-González et al., 2013; Keesstra et al., 2014c; Kibet et al., 2014). The number of replicate samples in laboratory-scale analyses typically range from 2-5 (Klopatek et al., 1988; Wang et al., 2015a).

2.3.2 Wildfire Simulation

Prescribed and slash burns, propane torches and heat lamps, litter burns, and muffle furnaces have all been used to study the effects of burning on soil and runoff physical and chemical properties (Busse et al., 2010; Cancelo-González et al., 2013; Cawley et al., 2017; Ferreira et al., 2005; Hogue and Inglett, 2012; Hohner et al., 2019b; Stoof et al., 2011). The key features which differ across these techniques are the range of simulated wildfire intensities, methods of burn characterization, and spatial variability of the combustion and burn properties. These studies are tabulated in Table 2.2 and are described in the following sub-sections.

Table 2.2: Summary of major wildfire simulation studies included in this review, listed in alphabetical order. The characteristics included are scale, size and shape of samples, wildfire simulation technique, peak burning temperature, duration of the burn, and a summary of key results from each study—reported with respect to unburned conditions unless otherwise stated. An "N/A" entry denotes an unreported study characteristic and "WR" is "Water Repellency".

Scale	Size and shape of samples	Simulation Technique	Peak burn temp	Duration of burn	Hydrologic and Water quality results	Source
Laboratory	0.25 x 0.25 x 0.15 m	Propane Torch or Heat Lamps	500 ºC	~ 220 min	WR, TOC, pyrogenic carbon significantly decreased in O horizon (98, 47, 74%, respectively) WR, pyrogenic carbon significantly decreased in upper cm of Ah horizon (74, 38%, respectively)	Badía-Villas et al., 2014
Laboratory	N/A	Muffle Furnace	Time-temperature matrix: 100, 150, 200, 250, 300, 350, 400, 450, and 500 ºC	5, 10, 15, and 30 min	KCl-extractable organic anion concentrations peaked between 150 and 350 ºC, also significantly affected by duration exposed to heating	Blank et al., 1994

Laboratory	0.3 m in diam., 0.2 m in height	Litter Burn	230 - 867 ºC	1 - 9.9 hrs	Soil moisture significantly affected downward heat transfer (> 20% volumetric moisture quenched heating lower than 2.5 cm)	Busse, et al., 2010
Laboratory	0.2 x 0.4 x 0.15 m	Propane Torch or Heat Lamps	200 and 400 ºC	Until temp reached, cooled afterwards	Degree hours (ºC·hrs measured during burning) were positively linearly related (R^2=0.966) to cation concentrations (sum of Na, K, Ca, and Mg)	Cancelo-Gonzalez, et al., 2012
Laboratory	0.45 x 0.2 x 0.15 m	Propane Torch or Heat Lamps	200 and 400 ºC (at 1 cm depth)	Until temp reached, cooled afterwards	Leached cations increased significantly (114% and 183% in moderate and severe burns, respectively)	Cancelo-Gonzalez, et al., 2013
Laboratory	0.2 x 0.4 x 0.15 m	Propane Torch or Heat Lamps	200 and 400 ºC (at 1 cm depth)	Until temp reached, cooled afterwards	Al and DOC increased 19 and 21%, respectively, for moderate burns and 44 and 679% for severe burns Fe increased only at severe burns (66%)	Cancelo-Gonzalez, et al., 2015
Laboratory	N/A	Muffle Furnace	225, 350, and 500 ºC	2 hrs	DOC peaked at mild-moderate burns (84% increase) and was lowest for severe burns (99% decrease)	Cawley et al., 2017
Laboratory	N/A	Muffle Furnace	149, 204, 260, 316, 371, 427, and 482 ºC	5, 10, 15, and 20 min	Non-wettability did not increase for mild burns, increased after 10-20 mins for mild-moderate burns Impenetrability reached after 5-10 mins for moderate burns, was destroyed after 10 mins for severe burns	Debano and Krammes, 1966
Plot	3 x 10.7 m	Prescribed or Slash Burn	Low intensity	N/A	Runoff ratio and sediment yield increases (21% and 40%, respectively) were not significantly greater than the natural variability for the locations or seasons	Emmerich and Cox, 1992
Plot	0.5 m²	Prescribed or Slash Burn	N/A	N/A	Infiltration significantly decreased in bunchgrass, shortgrass, and oak-dominated sites (25-36%) Sediment yield significantly increased for all sites (344-225,000%)	Hester et al., 1997
Laboratory	N/A	Muffle Furnace/ Litter Burn	Muffle: 200, 250, 300, 350, 400, 450, and 550 ºC; Flame: 520 (high), 460 (moderate), 350 (low) ºC	Muffle: 2 hrs; Flame: 200 - 700 sec (until 5g, 10g, and 25g of fuel was completely incinerated)	Muffle furnace: total C, P changed monotonically with burn intensity (total 89% decrease and 1381% increase, respectively); total N peaked at moderate burns (385% increase), decreased to no response at severe burns Litter burning: results showed no significant trends with burn intensity	Hogue and Inglett, 2012
Laboratory	0.5 x 0.5 x 0.05 m (+ litter layer)	Muffle Furnace	225, 350, and 500 ºC	2 hrs	DOC decreased significantly for mild-moderate burns (52%), DOC:DON showed a non-significant 27% decrease	Hohner et al., 2019
Laboratory	0.48 x 0.28 x 0.05 m	Litter Burn	250 - 300 ºC	Until litter was completely burned	WR was moderate, peak runoff rates increased up to 5525%, infiltration rates decreased up to 86%, and sedimentation rates increased	Keesstra, et al., 2014
Laboratory	0.2 m diam., 0.35 m height	Propane Torch or Heat Lamps	40 - 94 ºC	15 min	Positive correlation (R^2=0.90) between vesicular-arbuscular mycorrhizal colonization and soil temperature	Klopatek et al., 1988
Plot	N/A	Prescribed or Slash Burn	N/A	N/A	Infiltration rates and sediment yield decreased 2% and 3%, respectively, though neither change was statistically significant	Knight et al., 1983

Laboratory /Plot	2 x 2 m (burn box), 1.2 x 2.4 m (burn table)	Litter Burn/ Propane Torch or Heat Lamps	650 ºC (prescribed); 600 ºC (burn box); 400 ºC (burn table); 400 ºC (propane prong)	Until fuel burned (prescribed , burn box, burn table); 60 sec (propane prong)	Burn box, burn table, and prescribed fires had similar time-temperature profiles; propane prong fires had longer durations near the max temperature, but most similar total heat dosages to prescribed fires	Kral et al., 2015
Plot	1 m²	Prescribed or Slash Burn	35 - 563 ºC	Until done burning	Runoff ratios and sediment yields increased 275% and 775%, respectively	Marcos et al., 2000
Laboratory /Plot	0.1 x 0.1 x 0.1 m	Litter Burn/ Propane Torch or Heat Lamps	Low intensity	N/A	Root mat soils with moisture contents of 93% and 145% had a 50% and 10% probability, respectively, of sustained smouldering	Reardon et al., 2007
Laboratory	0.305 m diam.	Propane Torch or Heat Lamps	100 - 150 ºC (low), 250 - 300 ºC (mod), 400 - 500 ºC (high)	N/A	WR peaked for mild burns (WDPT tests > 60 s), then decreased at moderate and severe burns (WDPT tests 5-60 and < 5 s, respectively)	Robichaud and Hungerford, 2000
Plot	0.83 m²	Prescribed or Slash Burn	N/A	N/A	Infiltration rates, for field capacity and dry conditions, respectively, significantly decreased (~11%) and did not change Sediment yields significantly increased ~141% for all conditions	Roundy et al., 1978
Laboratory /Plot	70 x 70 m (plot)	Prescribed or Slash Burn/Muffle Furnace	370, 470, and 570 ºC (muffle furnace)	155, 105, and 65 sec	Pyrogenic carbon mass decreased 7-15% for experimental fire and 1-15%, 28-63%, and 26-37% for increasing muffle burn temperatures	Santín et al., 2013
Laboratory	0.103 m diam., 0.1 m height	Propane Torch or Heat Lamps	600 ºC	5 min	Soil moisture significantly reduced peak temperatures and duration of heating above 60 and 175 ºC; rock cover also reduced peak temperatures, but increased heating durations	Stoof, et al., 2011
Laboratory	0.2 m diam., 0.1 m height	Propane Torch or Heat Lamps	200 - 250 ºC (low), 450 - 500 ºC (high)	~40 min	Hydraulic conductivity significantly decreased for moderate burns (~62%), increased for severe burns (~229%) TOC significantly increased for moderate burns (~60%), decreased for severe burns (~11%)	Wieting et al., 2017

Prescribed and Slash Burns

Prescribed and slash burns are used as wildfire simulation techniques for plot-scale analysis of erosional, hydrologic, and sedimentation response to burning (Fernandes and Botelho, 2003; Hester et al., 1997; Robichaud et al., 2016). They involve the application of an incendiary device to *in situ* accumulated fuels under conditions conducive to the control of a fire, i.e., low winds and high soil moisture (Emmerich and Cox, 1992; Fernandes and Botelho, 2003; Hester et al., 1997; Marcos et al., 2000; Robichaud and Waldrop, 1994; West,

1917). Either aerial or ground-based incendiary devices are used to start these fires, such as a Plastic Sphere Dispenser ("Wildland Fire," 2020) or a drip torch (Arkle and Pilliod, 2010; Santín et al., 2013), respectively. Prescribed burning, also referred to as control or experimental burns (Chapman, 1947; Santín et al., 2013), typically span a 40-200 ha (100-500 acre) area and most often burn at a low intensity, only consuming fuels with a small diameter (e.g., pine needles and small branches) (Robinson et al., 2008). Slash burns, also referred to as pile burns (Robinson et al., 2008), involve first collecting surplus woody debris (i.e., trees or brush) into a concentrated area, which typically burns at a high intensity (Fornwalt and Rhoades, 2011; Robinson et al., 2008; Shahlaee et al., 1991). Prescribed and slash burns are most often qualitatively characterized by visual characteristics, such as biomass consumed (i.e., wildfire severity) (Arkle and Pilliod, 2010; Carter and Darwin Foster, 2004).

Propane Torches and Heat Lamps

Propane torches and heat lamps are common wildfire simulation techniques used in laboratory-scale analyses of burn impacts on soil and runoff chemical composition, as well as sedimentation response and changes to soil structure (Badía-Villas et al., 2014; Cancelo-González et al., 2013; Klopatek et al., 1988). These techniques involve a steady heat flux concentrated at a point (~50-100 mm) on the soil surface (Badía-Villas et al., 2014; Klopatek et al., 1988; Kral et al., 2015; Reardon et al., 2007; Robichaud, 2000; Stoof et al., 2011). Propane torches typically produce peak soil surface temperatures of 500-600°C (Badía-Villas et al., 2014; Stoof et al., 2011) whereas heat lamps produce peak temperatures of 200-400°C (Cancelo-González et al., 2013) from 5 to up to 220 minutes of exposure (Badía-Villas et al., 2014; Stoof et al., 2011), representing a range in burn

intensities. The temperature gradient extending down through the soil surface in these experiments is typically measured using thermocouples placed 0-15 cm below the soil surface (Badía-Villas et al., 2014; Busse et al., 2010; Cancelo-González et al., 2012a; Hogue and Inglett, 2012; Keesstra et al., 2014; Stoof et al., 2011).

Litter Burns

Litter burns are a laboratory-scale wildfire simulation technique used to study burn effects on soil and runoff composition, as well as the effects of soil composition on heating profiles (Busse et al., 2010; Keesstra et al., 2014; Kral et al., 2015; Reardon et al., 2007). The method involves igniting small amounts of litter spread evenly on top of soil samples, reaching temperatures of 230-867°C (Busse et al., 2010; Keesstra et al., 2014; Reardon et al., 2007). This measured amount of fuel allows for a direct measurement of burn intensity (Hogue and Inglett, 2012; Keesstra et al., 2014). Additionally, similar to propane torches and heat lamps, thermocouples placed 0-15 cm below the soil surface are used to measure heating profiles during the burn process (Busse et al., 2010; Keesstra et al., 2014; Lentile et al., 2006).

Muffle Furnaces

Muffle furnaces are a laboratory-scale wildfire simulation technique typically used to analyze burn impacts on soil chemical composition (Blank et al., 1994; Debano and Krammes, 1966; Hohner et al., 2019b; Santín et al., 2013). Samples are placed in an oven which is typically raised to temperatures between 100-570°C for as little as 65 seconds up to 2 hours to simulate a range of burn intensities (Blank et al., 1994; Cawley et al., 2017; Hohner et al., 2019b; Santín et al., 2013).

2.3.3 Rainfall Simulation

Nozzle-based and drip-based rainfall simulators, water drop penetration time (WDPT) tests, and leaching are typically used in conjunction with one of the wildfire simulation techniques mentioned above or they are implemented *in situ*, over an area already burned by a wildfire (Benavides-Solorio and MacDonald, 2001; Blake et al., 2010; Chevone et al., 1984; Ferreira et al., 2005; Keesstra et al., 2014; Kibet et al., 2014; Norris P. Swanson, 1965). The effects of consecutive rainfall events, with drying periods anywhere from 30 minutes to 1 year, are examined by some rainfall simulation studies, as antecedent moisture content and weathering over time can greatly influence post-wildfire hydrologic response (Ebel et al., 2012; Johansen et al., 2001; Murphy et al., 2015). Runoff collection chambers in these simulations are typically located at the lower end of a sloped plot or sample with a guard to deflect the simulated rainfall, sampling frequencies ranging from 20 seconds to 20 minutes (Johansen et al., 2001; Keesstra et al., 2014; Kibet et al., 2014; Robichaud et al., 2016). To facilitate discussion, various rainfall simulation techniques are divided into four categories: fixed nozzle-based simulators, dynamic nozzle-based simulators, drip-style rainfall simulators, and water drop penetration time (WDPT) tests and leaching. The key features which vary between these simulation techniques are range of rainfall intensities, precision of the droplet size and kinetic energy, and spatial distribution. Published wildfire studies which employ these techniques are described in the following subsections. These studies are tabulated in Table 2.3 and described in the following subsections.

Table 2.3: Summary of major rainfall simulation studies included in this review, listed in alphabetical order. The characteristics included are study subject, scale, size and shape of samples, rainfall simulation technique, rainfall intensity, duration and scheduling of the simulated rainfall, and a summary of key results from each study—reported with respect to

unburned conditions unless otherwise stated. An "N/A" entry denotes an unreported study characteristic and "WR" is "Water Repellency".

Scale	Size and shape of samples	Simulation Technique	Rate of rainfall	Duration and schedule of rainfall and drying	Hydrologic and water quality results	Source
Plot	1 m^2 plots	Dynamic Nozzle-Based	79 mm/hr	1 hr	Runoff ratios peaked immediately post-fire (5% and 20% increases for moderate and severe burns, respectively), muted effects from older fires Sediment yield peaked for 1-yr old fires (485% and 2392% increases for moderate and severe burns, respectively)	Benavides-Solorio and MacDonald, 2001
Plot	N/A	Dynamic Nozzle-Based	N/A	N/A	N/A	Bertrand, 1961
Plot	3 x 3 m	Drip-Style	5.08 - 83.82 mm/hr	N/A	N/A	Blackburn et al., 1974
Plot	1 x 0.5 m	Drip-Style	20, 40, and 60 mm/h	Until three temporally discrete runoff samples had been collected	Total phosphorus increased 20% and 49% for moderate and severe burns, respectively Suspended sediment increased 323% and 717% for moderate and severe burns, respectively	Blake et al., 2010
Laboratory	0.2 x 0.4 x 0.15 m	Fixed Nozzle-Based	N/A	N/A	Degree hours (°C·hrs measured during burning) were positively linearly related (R^2=0.966) to cation concentrations (sum of Na, K, Ca, and Mg)	Cancelo-Gonzalez, et al., 2012
Laboratory	0.45 x 0.2 x 15 m	Fixed Nozzle-Based	75 mm/hr	Runoff collected in 300 mL aliquots. 3 sessions, 2 hrs long, 15 days apart	Leached cations increased significantly (114% and 183% in moderate and severe burns, respectively)	Cancelo-Gonzalez, et al., 2013
Laboratory	0.2 x 0.4 x 0.15 m	Fixed Nozzle-Based	75 mm/hr	Runoff collected in 300-mL aliquots. 300 mm per session, 2 sessions, 2 hrs long, 15 days apart	Al and DOC increased 19 and 21%, respectively, for moderate burns and 44 and 679% for severe burns Fe increased only at severe burns (66%)	Cancelo-Gonzalez, et al., 2015
Plot	0.24 m^2	Fixed Nozzle-Based	54.6 mm/hr	N/A	N/A	Cerda et al., 1997
Laboratory	Potted plants	Drip-Style	6.5, 8.5, and 9.9 mm/hr	1 hr	N/A	Chevone et al., 1984
Plot	3 x 10.7 m	Dynamic Nozzle-Based	55 and 110 mm/hr	45 and 15 min	Runoff ratio and sediment yield increases (21% and 40%, respectively) were not significantly greater than the natural variability for the locations or seasons	Emmerich and Cox, 1992

Plot	0.24 m² (microplot), 8 x 2m (plot), and < 1.5 km² (catchment)	Fixed Nozzle-Based	50.5 mm/h	45 - 60 min	Runoff ratios increased 343%, 614%, and 1746% for mild, moderate, and severe burns, respectively, for micro-plot scale; increased 14% and 16,471% for mild and severe burns, respectively, for plot scale	Ferreira et al., 2005
Plot	0.5 m²	Drip-Style	203 mm/hr	50 min	Infiltration significantly decreased in bunchgrass, shortgrass, and oak-dominated sites (25-36%) Sediment yield significantly increased for all sites (344-225,000%)	Hester et al., 1997
Plot	N/A	Fixed Nozzle-Based	N/A	N/A	N/A	Holland, 1969
Plot	3.03 x 10.7 m	Dynamic Nozzle-Based	60 mm/h	1 hr event, 24 hr drying period, 0.5 hr event, 0.5 hr drying period, 0.5 hr event	Runoff ratios and sediment yields increased 103% and 2240%, respectively	Johansen, et al., 2001
Laboratory	0.48 x 0.28 x 0.05 m	Dynamic Nozzle-Based/WDPT Tests and Leaching	33 mm/h	4 hr event, drying period in oven, then second 2 hr wetting phase	WR was moderate, peak runoff rates increased up to 5525%, infiltration rates decreased up to 86%, and sedimentation rates increased	Keesstra et al., 2014
Laboratory	1 x 0.2 x 0.075 m	Fixed Nozzle-Based	31.7 mm/h	40 min	N/A	Kibet, et al., 2014
Plot	N/A	Drip-Style	203 mm/hr	30 min	Infiltration rates and sediment yield decreased 2% and 3%, respectively, though neither change was statistically significant	Knight et al., 1983
Plot	1 m²	Fixed Nozzle-Based	180 mm/hr	5 min	Runoff ratios and sediment yields increased 275% and 775%, respectively	Marcos et al., 2000
Plot	0.75 x 0.75 m	Dynamic Nozzle-Based	100 mm/hr	1 hr (three events, each 1 year apart)	Runoff ratios increased 21% immediately post-fire, lesser values in years 1-5 Infiltration decreased 30% immediately post-fire, greater values in years 1-5 Sediment yield peaked 1 yr post-fire (7227% increase), lesser values in years 0, 2, and 5	Robichaud et al., 2016
Plot	4 x 8 m	Fixed Nozzle-Based	76 mm/hr	1 hr (three events, the second a couple days after first event, then the third 2 hrs after the second)	Runoff ratio and sedimentation rates increased ~4000% and ~17,000%, respectively, less for higher antecedant moisture contents	Rosso et al., 2007
Plot	0.83 m²	Drip-Style	83.8 mm/hr	1 hr	Infiltration rates, for field capacity and dry conditions, respectively, significantly decreased (~11%) and did not change Sediment yields significantly increased ~141% for all conditions	Roundy et al., 1978

Plot	10.7 x 3.05 m	Dynamic Nozzle-Based	N/A	1 hr, then 30 min 24 hrs later, then 30 min immediately	Runoff ratios and sediment yields increased 432% and 845%, respectively, immediately post-fire, 391% and 690% 1-yr post-fire for severe burns; little substantial changes for mild burns	Simanton et al., 1986
Plot	N/A	Dynamic Nozzle-Based	63.5 and 127 mm/hr	N/A	N/A	Swanson, 1965
Laboratory	0.2 m diam, 0.1 m height	Fixed Nozzle-Based/WDPT Tests and Leaching	25 mm/hr	1 hr	Hydraulic conductivity significantly decreased for moderate burns (~62%), increased for severe burns (~229%) TOC significantly increased for moderate burns (~60%), decreased for severe burns (~11%)	Wieting et al., 2017
Plot	1 m^2	Fixed Nozzle-Based	69 - 202.5 mm/hr	N/A	N/A	Wilcox et al., 1986
Plot	15 x 20 m	Fixed Nozzle-Based	35, 75, and 150 mm/hr	10 - 49 min	Peak runoff rates and infiltration had no substantial change Sediment yield increased 730-353%	Wilson, 1999
Plot	0.5 m2 plots	Dynamic Nozzle-Based	75 mm/hr	1 hr events	With respect to removal of ash: Runoff ratios and infiltration increased 180% and decreased 32% 1 mo post-burn, respectively; almost no change 12 mo post-burn	Woods and Balfour, 2008

Fixed Nozzle-Based Rainfall Simulators

Fixed nozzle-based rainfall simulators have been used in both plot- and laboratory-scale analyses of wildfire impacts on soil and runoff physical and chemical changes (Cancelo-González et al., 2013; Cerdà et al., 1997; Dunne et al., 1980; Kibet et al., 2014; Wilcox et al., 1986; Wilson, 1999). Most of these simulators use a single, stationary nozzle ~2 m above the ground which points downward, covering areas of 0.08-1 m^2 (Cancelo-González et al., 2015; Marcos et al., 2000). One exception is the plot-scale Field Efficient Colorado State Rainfall Simulator which has ten 3 m risers covering 300 m^2, each with 1-2 nozzles pointing upwards (Holland, 1969; Wilson, 1999). Nozzles are rated to produce a droplet size and kinetic energy similar to natural rainfall at a specified distance beneath

the nozzle, as well as an even distribution of rainfall intensity (e.g., the FullJet® and VeeJet© nozzles produced by the Spraying Systems Company) (Cerdà et al., 1997; Kibet et al., 2014; Wilcox et al., 1986). Droplet sizes range from 0.8-4.0 mm and kinetic energies from 0.1-28 J m^{-2} mm^{-1}, with rainfall intensities ranging from 5-203 mm/hr and durations from 5 min to 2 hrs (Cancelo-González et al., 2013; Cerdà et al., 1997; Dunne et al., 1980; "FullJet Full Cone Spray Nozzles," n.d.; "VeeJet Flat Spray Nozzles," n.d.; Marcos et al., 2000; Wilcox et al., 1986).

Dynamic Nozzle-Based Rainfall Simulators

Although dynamic nozzle-based rainfall simulators have been used to analyze similar wildfire effects as fixed nozzle simulators, their use has been more common in plot-scale analyses (Balfour and Woods, 2008; Emmerich and Cox, 1992; Johansen et al., 2001; Norris P. Swanson, 1965; Robichaud et al., 2016; Simanton et al., 1990). These simulators incorporate horizontal rotation (Keesstra et al., 2014; Norris P. Swanson, 1965) or sweeping motions ~3 m above the plot or samples (Balfour and Woods, 2008; Benavides-Solorio and MacDonald, 2001; Bertrand, 1961; Robichaud et al., 2016), covering large areas of up to ~30 m^2 (Emmerich and Cox, 1992). The nozzles used are rated to produce natural rainfall kinetic energy and droplet size, same as the fixed nozzle simulators, and produce rainfall intensities ranging from 33-127 mm/hr for 15 min up to 4 hours (Norris P. Swanson, 1965).

Drip-Style Rainfall Simulators

Drip-style rainfall simulators have been used to study the same processes as nozzle-style simulators, typically at a plot-scale (Blackburn et al., 1974; Blake et al., 2010; Hester et al., 1997). In these simulators, water is channeled to a large number (~168 to 2209) of fine tubes or needles which periodically release droplets due to gravitational forces

(Blackburn et al., 1974; Chevone et al., 1984), covering areas from 0.4-9 m^2 (Blackburn et al., 1974; Knight et al., 1983; Roundy et al., 1978). This technique produces droplet sizes ranging from 2.6-3.3 mm in diameter and rainfall intensities ranging from 20-203 mm/hr (Blake et al., 2010; Hester et al., 1997) for durations of 0.5-1 hours (Knight et al., 1983; Simanton et al., 1990).

Water Drop Penetration Time Tests and Leaching

WDPT tests and leaching do not simulate the mechanics of rainfall, but are important laboratory-scale techniques used to assess wildfire impacts on soil water repellency and changes in chemical composition, respectively (Badía-Villas et al., 2014; Blank et al., 1994; Cawley et al., 2017; Debano and Krammes, 1966; Wieting et al., 2017). WDPTs involve placing droplets of water or a water-ethanol mixture on burned soil and recording the duration of time for each drop to infiltrate as a measure of soil water repellency (Badía-Villas et al., 2014; DeBano, 1981; Robichaud and Hungerford, 2000). Alternatively, leaching involves dissolving water-soluble chemical constituents in burned soil or litter into water, analyzing the water for chemical composition (Hohner et al., 2019b).

2.4 Simulation Technique Comparisons

Each of the wildfire and rainfall simulation experiment techniques described above has benefits and limitations which future researchers must take into consideration in their own designs. The following sections describes the pros and cons of each technique.

2.4.1 Wildfire Simulation

The following paragraphs discuss common *pros* of wildfire simulation experiments. A summary is provided in Table A.1 in *Appendix A*.

Heterogeneous combustion is a key strength of prescribed fires and slash burns, as well as litter burns. The variability in spatial distribution of heating and volatilization created by these types of burns produce variable combustion residues analogous to natural wildfires. This variability is due to spatial heterogeneity of fuel types and amounts, as well as variable wind speed, direction, and air temperature across the burn area in plot-scale experiments (Hogue and Inglett, 2012).

Similar intensity and duration to a natural wildfire is a strength of prescribed fires and slash burns. As a comparable amount of fuel is available in a prescribed burn as a natural wildfire, prescribed fires typically match the peak temperatures and duration of a low intensity wildfire (Arkle and Pilliod, 2010). Slash burns can reach the extreme temperatures reached by severe wildfires, sometimes as high as ~2200°C—often unachievable by laboratory-scale simulation techniques (Shahlaee et al., 1991).

Precise control over burn intensity and spatial distribution is a strength of propane torches and heat lamps, as well as litter burns (Robichaud and Hungerford, 2000). This technique allows for more direct attribution of burning effects to specific intensities, given the controlled range and spatial distribution of burn intensities (Keesstra et al., 2014).

Low variability in heating is a benefit of propane torches and heat lamps, as well as muffle furnaces. Low spatial and temporal variability created by the consistent heating distribution of these methods is conducive to quantitative analysis, as it limits variability-driven uncertainty in responses (Santín et al., 2013).

Allowance for measurement of heating profiles is a key aspect of propane torches and heat lamps, litter burns, and muffle furnace methods. Propane torches and heat lamp methods, as well as litter burns, achieve this using thermocouples, and muffle furnaces

achieve this through digital temperature readings. This allows for consistent characterization of burn intensity, conducive to quantitative post-wildfire analyses.

Control over duration of heating is an attribute of propane torches and heat lamps, as well as muffle furnaces. This allows for further control and precision of burn intensity—sometimes characterized by burn duration (Cancelo-González et al., 2012a).

Incremental control over burn intensity is a benefit unique to muffle furnaces. This feature allows for analysis of changing burn effects over a range of burn intensities—conducive to analyzing response trends with increasing burn intensity (Santín et al., 2013).

The following paragraphs discuss common *cons* of wildfire simulation experiments:

Qualitative burn severity characterization is a limitation for prescribed fires and slash burns (Balfour and Woods, 2008; Ferreira et al., 2008). Qualitative wildfire characterizations (i.e., burn severity) are less precise than temperature or fuel measurements (i.e., burn intensity), hindering quantitative analyses and precise replication of simulated burn severities.

Tradeoff between high intensities and burn coverage is another limitation for prescribed fires and slash burns. As these burns must be managed in a safe, controlled way, larger prescribed burns (40-200 ha) are typically limited to a low burn intensity. Slash burns, which can achieve extreme temperatures, are therefore limited to smaller areas.

Uniformity in spatial heating is considered a limitation for propane torch and heat lamp methods, as well as muffle furnaces. While this characteristic may assist in the attribution of burn effects to specific drivers, it typically does not produce the heterogeneous combustion residues present after a natural wildfire. This is a key source of uncertainty in the representativeness of burned soil structure and composition as compared to natural combustion (i.e., burning fuel) (Hogue and Inglett, 2012). Some studies have

addressed this shortcoming by using controlled heating methods coupled with igniting vegetation or litter on the soil surface (Cancelo-González et al., 2013; Kral et al., 2015; Reardon et al., 2007).

Limitations to low-intensity burns, due to safety and other logistical considerations, is a drawback of litter burns (Busse et al., 2010). The even distribution of fuels is often inconsequential in substantively heating soils, which may not exceed a low intensity burn (Busse et al., 2005).

Heating from all sides is a disadvantage of muffle furnaces. These heating mechanics are categorically different from a natural wildfire, which only heat the side of exposed soil surface (Blank et al., 1994; Hohner et al., 2019b).

2.4.2 Rainfall Simulation

The following paragraphs discuss common *pros* of rainfall simulation experiments. A summary is additionally provided in Table A.2 in *Appendix A*.

Simplicity in design is a key benefit of fixed nozzle-based rainfall simulators, WDPT tests, and leaching (Cancelo-González et al., 2013; Cerdà et al., 1997; Holland, 1969; Wilcox et al., 1986). A stationary nozzle is a relatively simple and inexpensive mechanism to construct or purchase and is often sufficient in terms of coverage area and semblance to natural rainfall. WDPT tests and leaching typically only require simple laboratory equipment.

Transportability and adaptability to steep terrains is a benefit of fixed nozzle-based rainfall simulators. This allows these simulators to be tested on otherwise inaccessible sampling locations and on steep terrains up to 45° (Cerdà et al., 1997; Wilcox et al., 1986; Wilson, 1999).

Rainfall intensities and droplet sizes similar to natural rainfall are key benefits of fixed nozzle-based, dynamic nozzle-based, and drip-style rainfall simulators. The ranges of intensities and droplet sizes for these types of simulators make them representative of typical natural precipitation (Cancelo-González et al., 2013; Cerdà et al., 1997; Dunne et al., 1980; Marcos et al., 2000; Wilcox et al., 1986).

A large area of coverage is a benefit unique to dynamic nozzle-based rainfall simulators. Larger plot-scale analyses allow for larger-scale hydrologic processes to occur which cannot be observed on smaller scales (Fraser et al., 2013; Hamed et al., 2002; Le Bissonnais et al., 1998), and minimizes edge effects (Ferreira et al., 2008).

Spatial and temporal variability in droplet distribution similar to natural rainfall is an important attribute of dynamic nozzle-based rainfall simulators. The horizontal rotation and sweeping motion of the nozzles used in these simulators may create spatial and temporal variability which is more representative of natural rainfall than stationary nozzles (Chevone et al., 1984).

Increased control and precision of droplet size is a benefit of drip-style rainfall simulators. Droplet size can be altered by changing the gage of the tubes and needles used, allowing for control over droplet diameters (Chevone et al., 1984) and subsequently the kinetic energy of raindrops produced.

Direct measurement of water repellency and chemical changes is a benefit unique to WDPT tests and leaching. This can allow for more precise attribution of burn effects, as opposed to the indirect measurements through runoff generation and chemical composition in other rainfall simulation techniques (Debano and Krammes, 1966; Hohner et al., 2019b; Wieting et al., 2017).

The following paragraphs discuss common *cons* of rainfall simulation experiments:

A small area of coverage is a limitation of many fixed nozzle-based rainfall simulators, WDPT tests, and leaching (Cerdà et al., 1997). The smaller area covered limits analyses to the laboratory- or smaller plot-scale for fixed nozzle-based rainfall simulators, and typically only laboratory-scale samples for WDPT tests and leaching. This often means that these simulation experiments do not capture larger-scale hydrologic processes, such as rill erosion (as discussed in *Section 2.5.3*).

Rainfall kinetic energies lower than typical natural rainfall is a limitation for fixed nozzle-based rainfall simulators. Rainfall kinetic energies for these simulators tend to be lower than natural rainfall due to simulated droplets often not reaching terminal velocity before impact, as nozzle heights are constrained by equipment (Cerdà et al., 1997).

Complexity and expense of design are limitations for dynamic nozzle-based rainfall simulators and drip-style rainfall simulators, which may result in logistical and financial challenges for studies.

Difficulty in transportation is another drawback for dynamic nozzle-based rainfall simulators and drip-style rainfall simulators. This can limit plot-scale study sites to ones with accessible roads, as well relatively flat terrain.

Lack of rainfall impact on soils, and therefore lack of droplet kinetic energy, is a limitation of WDPT tests and leaching. These methods do not capture the physical processes of rainfall impact on soil surface and therefore cannot simulate natural constituent transport through runoff (Hohner et al., 2019b).

2.5 Recommendations

Decisions about which experimental methodologies to choose should largely depend on study scope—the hydrologic or chemical responses being analyzed, as well as the

temporal and physical scales of the analysis. Researchers should also carefully consider the geographical setting of their study and incorporate specific regional characteristics into their experimental design, such as soil type, vegetation cover, climate regimes, and terrain slopes. In general, we recommend that experimental design elements should be optimized based on their strengths in analyzing important study elements, while weighing time, logistic, and financial constraints.

For example, a study analyzing the effects of different vegetation types on post-wildfire hydrologic processes may want to focus on plot-scale techniques in order to capture larger-scale erosional processes such as rill erosion. In such a scenario, a prescribed burn method could provide the necessary burning scale, burn the intended fuel type (i.e., the different types of vegetation in question), and also partially represent heterogeneous combustion patterns, intact soil structure, and larger-scale vegetation present in a natural wildfire. In this example, a dynamic nozzle-based rainfall simulator may be the best choice of rainfall simulator to allow for a large area of coverage. Alternatively, if the interaction between burn intensity and vegetation characteristics was of primary interest in the above example, then a more appropriate experimental set-up may use a burn simulation technique which allows for greater incremental control over wildfire intensity. Heat lamps, for example, allow for analysis of targeted vegetation burning at specific intensity levels. However, this type of analysis sacrifices some representation of natural burning, due to low spatial variability in combustion. Additionally, as with any laboratory method, the soil sampling process involved in heat lamp simulation techniques introduces edge effects, potentially disturbed soil structure, and can only represent small-scale hydrologic processes and vegetation. A fixed nozzle-based rainfall simulator may be the best choice in this scenario, as only a small area of coverage would be required.

Researchers should also take the results from previous simulation studies into consideration for their study designs. For example, we showed that plot-scale studies which implemented prescribed burning tended to most frequently produce results that were not statistically significant relative to control samples. This is most likely due to a combination of the heterogeneity of burn intensities and high variability of the natural settings that are subjected to prescribed burning. Conversely, laboratory-scale studies which used muffle furnace heating tended to produce results that were highly statistically significant and were able to assess responses at a high number (up to nine) of burn intensity increments. This level of granularity allowed these studies to infer a more fundamental character of the heating effect, for example a monotonic versus negative parabolic response to heating. However, these types of analyses were typically limited to water quality constituents and water repellency, as larger-scale erosional and hydrologic responses were not captured.

Using precedents set by previous simulation studies as a guide, researchers may choose appropriate methods to fit their research goals. To aid discussion, we categorize strengths and weaknesses of simulation experiments into four important factors: (1) representation of natural processes and settings, (2) analysis of multiple post-wildfire water quality and supply drivers simultaneously and independently, (3) observation of responses on different temporal and spatial scales, and (4) mitigation of uncertainty of results. The following paragraphs discuss the tradeoffs that exist between these four key design factors.

2.5.1 Representation of Natural Processes

Replicating natural processes improves the representativeness of experimental results—furthering understanding of wildfire effects on soil and runoff characteristics. However, this goal must be weighed against logistical challenges of *in situ* collection, as

well as the increased spatial and temporal variability inherent with natural features—which can create difficulties in attribution. Studies typically address these tradeoffs by only choosing natural or unperturbed features most important to the subject and scale of the study. For example, Hogue and Inglett (2012) examined C and N concentrations in naturally combusted residue, using litter burning with spatially variable combustion to replicate natural wildfire mechanisms (Hogue and Inglett, 2012). Similarly, Benavides-Solorio and MacDonald (2005) and Johansen et al. (2001) analyzed wildfire's role in increased rill erosion, a plot-scale erosional process, by employing plot-scale wildfire and rainfall simulation techniques. Therefore, future studies are recommended to first identify the subject and scale(s) of greatest interest, then focus efforts on replicating natural processes for those elements.

2.5.2 Incorporation of Multiple Key Drivers

Incorporation of multiple drivers—i.e., burn severity, rainfall intensity, terrain slope, vegetation type, and soil characteristics—at multiple increments and categories, is often sought to gain a more comprehensive understanding of their relative importance and system interactions. However, studies must evaluate the benefits of including these characteristics, since they can limit the number of replicate samples useful for characterizing uncertainty, as well as require large numbers of samples to be collected. Most studies in this review include burn severity and rainfall intensity in their analyses (Cancelo-González et al., 2013; Ebel et al., 2012b; Moody et al., 2013; Robichaud, 2005a; Smith et al., 2011). Soil structure and composition, terrain slope, climate, vegetation type, and antecedent moisture content are less commonly incorporated, but can have comparable impacts on hydrology and water quality (Certini, 2005). For example, Johansen et al. (2001)

incorporated the percentage of bare soil into their analysis of post-wildfire sedimentation, finding that this driver had a strong correlation with sediment generation in addition to burn severity. Factors involved in wildfire prevention, suppression, and mitigation of effects (i.e. mechanical thinning, mulching, and chemical fire-retardants) are also less commonly incorporated, though could provide insights important to fire management efforts. Understanding how each driver impacts hydrology and water quality independently and jointly can also assist in the creation of catchment-scale predictive models, which typically incorporate multiple drivers as model parameters. Benavides-Solorio and MacDonald (2005), for example, used fire severity, percent bare soil, rainfall erosivity, soil water repellency, and soil texture as model parameters to predict post-wildfire sedimentation.

Responses systematically tested over ranges of drivers allows for an understanding of the shape of the response function (e.g., monotonic, parabolic, etc.). Studies typically use only 2-3 increments of burn severity and rainfall intensity due to logistical and time constraints (Blake et al., 2010; Cancelo-González et al., 2013; Cawley et al., 2017; Robichaud and Hungerford, 2000). However, a higher number of increments proved to be important in Hohner et al. (2019). Here, soil samples were heated in a muffle furnace at five temperature increments ranging from 150-550°C, finding that water extractable organic C and N had a roughly negative parabolic relationship with temperature, peaking around 250-350°C (Hohner et al., 2019). We recommend that for a given number of total samples, future studies carefully consider the tradeoffs between the number of increments and the number of replicates.

2.5.3 Analyses at Multiple Spatial Scales

A limitation of single-scale wildfire and rainfall simulation experiments is the lack of consideration for how properties and processes at one scale may effect water quality and supply responses at larger scales (Ferreira et al., 2008). Incorporating multiple physical scales in simulation experiments can provide insight into upscaling operators which can inform catchment-scale predictions. This is particularly important in sedimentation analyses, as geomorphic and erosional processes vary greatly from the laboratory-scale to the catchment-scale (Kampf et al., 2016b). Post-fire sedimentation mechanisms such as streambed erosion may be entirely missing, even from plot-scale analyses. However, understanding how mechanisms (e.g., rill erosion, streambed erosion, etc.) are introduced and change at increasing scales can allow for indirect estimation and inference about catchment-scale response.

Simulating multiple physical scales is challenging in a laboratory setting due to fixed equipment size (Busse et al., 2005; Cancelo-González et al., 2013; Keesstra et al., 2014c; Klopatek et al., 1988; Wieting et al., 2017). Multi-scale analysis is also uncommon in plot-scale studies. However, Ferreira et al. (2005) analyzed sediment and runoff in post-fire plots on a microplot- (<1 m^2), plot- (16 m^2), and catchment-scale (<1.5 km^2), allowing for comparison of results across varying scales. We recommend that future studies consider analyzing postfire responses at more than one scale, if feasible with their study design and logistical and financial limitations.

2.5.4 Uncertainty Quantification

Uncertainty is most commonly estimated in experimental systems by testing multiple replicate samples, or uniform samples tested under the same conditions

(Benavides-Solorio and MacDonald, 2005; Keesstra et al., 2014c). Quantifying uncertainty can be useful in differentiating the water quality responses of different drivers and can help inform upscaling of results to the catchment scale (Liu and Gupta, 2007; Wagener and Gupta, 2005). High spatial and temporal variability, albeit representative of natural systems, can introduce additional uncertainty due to difficulties in the attribution of responses to specific drivers. Thus, uncertainty analysis and mitigation efforts must consider both replicate uncertainty, as well as the role of natural variability on attribution uncertainty. In Keesstra et al. (2014), for example, soil samples were homogenized—reducing uncertainty from spatially variable soil structure and vegetation, but decreasing the samples' semblance of a natural environment. We recommend opting for greater numbers of replicate samples to quantify uncertainty, while weighing incorporation of multiple drivers at different increments and scales—which may constrain the feasible number of replicates across each study dimension.

2.6 Conclusion

This review provides a synthesis of knowledge on wildfire and rainfall simulation techniques used to understand the impacts of wildfire on water quality and supply. Wildfire and rainfall simulation techniques offer solutions to logistical challenges faced in the collection of *in situ* data, including potentially dangerous post-fire environments, expensive fieldwork expeditions, and lack of control data. However, each technique has unique strengths and weaknesses. Plot-scale analyses are often able to capture a higher spatial variability more representative of natural settings than laboratory-scale analyses, as well as simulate larger-scale hydrologic processes (i.e., erosion). Yet, attribution of responses to specific drivers is often difficult due to high variability of conditions within and across plots.

Laboratory-scale analyses can more precisely control factors in the simulated system, limiting variability and allowing for drivers to be tested at a range of increments. This allows for a more direct attribution of the role of each driver on system responses, independently and jointly across ranges of values. Laboratory-scale experiments also have the benefit of more precise measurement—for example using thermocouples to measure surface temperature, or control over drivers, e.g., a muffle furnace, which assists in the quantitative analysis of results. The downsides of laboratory-scale experiments are that they are less representative of a natural wildfire system, due to limited spatial variability and scale—meaning only small-scale hydrologic processes can be analyzed.

Common design considerations across these studies include representation of natural processes, incorporation of multiple key drivers, analysis at multiple spatial scales, and uncertainty quantification. As studies are limited by time, resources, and logistical constraints, prioritization of these design considerations in future studies must be made based on scale, scope, and subject matter. Representation of natural processes can increase variability, and therefore increase uncertainty in results. Similarly, increased complexities in the study design, such as incorporation of multiple drivers and spatial scales, can decrease the amount of replicate samples at each condition, thereby limiting a robust quantification of uncertainty. Thus, future studies must weigh which design considerations are important for each aspect of their experiment, focusing resources on realistic representation of the key drivers or constituents of interest.

This review seeks to support the advancement of knowledge in the field of wildfire impacts on water quality and supply. These findings may be informative for future practitioners, as well as for water management efforts in mitigation and adaptation strategies for wildfire impacts. As wildfires continue to represent an increasing threat to

water quality and supply, developing advanced techniques to provide further understanding of wildfire effects will become increasingly essential.

Chapter 3

An analysis of wildfire effects on water quality and hydrology through laboratory simulation experiments

The following chapter is currently under review at the International Journal of Wildland Fire:

> Brucker, C.P., Livneh, B., Butler, C.E., Rosario-Ortiz, F., 2023. A laboratory-scale simulation framework for analyzing wildfire hydrologic and water quality effects. International Journal of Wildland Fire.

3.1 Overview

As discussed in Chapter 2, laboratory-scale wildfire simulation techniques present a useful method of increasing understanding of post-wildfire effects on hydrology and water quality. However, many limitations are associated with these experiments, including a narrow focus on only one or two mechanisms important in post-fire environments and experimental limitations which contribute uncertainty to results. Typically, only one to two driving factors are observed in these studies, with one to two increments each, whereas conditions in a natural environment are much more numerous and varied. Additionally,

most laboratory-scale studies test few replicate samples, creating higher uncertainty in results.

In this chapter, we present the design, construction, and evaluation of our unique laboratory-scale wildfire and rainfall simulation experiments which attempt to address some limitations of previous studies. We incorporated a relatively higher number of key drivers (3) and testing increments (2-4) into our experimental framework as compared with previous studies, attempting to capture the wide variability of driving factors in natural settings. Additionally, a greater number of replicates (8) tested at each combination of driver increments was more conducive to uncertainty quantification and characterization. Using this framework, we observed runoff, infiltration, sediment, dissolved organic matter, nitrogen, and turbidity responses from 154 soil samples in a three-dimensional matrix of controls: burn intensity, rainfall intensity, and terrain slope. Results from this experiment were compared to previous *in situ* and simulation wildfire studies and evaluated for similar trends.

3.2 Background

Laboratory-scale wildfire and rainfall simulations facilitate attribution of post-wildfire water quality responses to drivers by controlling and quantifying system factors, but may offer oversimplified representations of natural, highly variable environments (Brucker et al., 2022). Where natural processes involve interactions of numerous drivers with wide ranges, previous simulation studies analyzed only one or two drivers and intensity increments (Badía-Villas et al., 2014; Blank et al., 1994; Busse et al., 2010; Cancelo-González et al., 2013; Robichaud and Hungerford, 2000; Stoof et al., 2010; Wieting et al., 2017). Typically, either burn or rainfall mechanisms were simulated, using propane

torches (Kral et al., 2015), heat lamps (Wieting et al., 2017), litter burning (Busse et al., 2010), and muffle furnaces (Hohner et al., 2019), or nozzle-based (Cancelo-González et al., 2013) and drip-based (Hester et al., 1997) precipitation apparati. Other key processes were either naturally applied, e.g., *in situ* samples collected from wildfire-affected areas (Wang et al., 2015b), or excluded, e.g., samples analyzed using leaching instead of simulating precipitation mechanisms (Hohner et al., 2019). Cancelo-González et al., 2013 and Keesstra et al., 2014 are key exceptions which controlled and quantified both burn and rainfall mechanisms. However, in these studies and others, the few driver increments tested limit driver analyses to binary assessments (e.g., 'burned' versus 'unburned') (Brucker et al., 2022), lacking more continuous information on the shape of response relationships. Finally, analysis of just one or two geophysical drivers is common (Badía-Villas et al., 2014; Blank et al., 1994; Keesstra et al., 2014b; Klopatek et al., 1988), including vegetation types (Blank et al., 1994) and layers of burned detritus (Klopatek et al., 1988), as well as soil moisture (Busse et al., 2010), rock content (Stoof et al., 2011), and aggregate sizes (Keesstra et al., 2014b). These factors are much more numerous and varied in natural environments and can have great influence on post-fire hydrologic and water quality response (Cotrufo et al., 2016; Ebel et al., 2012b; Murphy et al., 2015).

This laboratory-scale wildfire and rainfall simulation experiment attempts to represent complex post-wildfire environments more closely by simulating two key processes—burn and rainfall mechanisms—and one driving topographic feature—terrain slope. These drivers were tested at three to four increments each across intensity ranges which reflected natural settings and captured key mechanisms, observing responses in runoff, infiltration, sediment, DOM, turbidity, and nitrogen. Differing from previous studies with at most two simulated drivers (Busse et al., 2010; Cancelo-González et al., 2013;

Keesstra et al., 2014b; Stoof et al., 2011; Wieting et al., 2017), here the simulator was designed to quantify burn effects both independently and in the context of system interactions—with ours being unique to include varying terrain slopes as a simulated factor (Cancelo-González et al., 2013; Kibet et al., 2014). The influence of drivers was assessed at different levels through relatively numerous (Badía-Villas et al., 2014; Cancelo-González et al., 2013; Keesstra et al., 2014b; Stoof et al., 2011; Wieting et al., 2017) testing increments at ranges common in natural settings. While previous studies have analyzed either sediment (Emmerich and Cox, 1992; Keesstra et al., 2014b; Knight et al., 1983; Marcos et al., 2000; Roundy et al., 1978) or solutes (Cancelo-González et al., 2013) in simulated runoff, observations of both provided insights into sediment's role in solute response, in addition to hydrologic mechanism influences. This chapter describes the design, construction, and evaluation of laboratory-scale wildfire and rainfall simulators, as well as their usefulness in providing insights into post-wildfire water quality assessments.

3.3 Methods

To conduct this simulation experiment, soil samples were first excavated from a site in Colorado (Figure 3.1a), then burn and rainfall treatments applied using custom-designed heat lamp (Figure 3.1b) and nozzle-based (Figure 3.1c) apparati. Collected runoff was then analyzed (Figure 3.1d) for turbidity, total suspended solids (TSS), total dissolved nitrogen (TDN), and dissolved organic carbon (DOC)—measured as a proxy to estimate DOM. Observed responses for the tested soil samples were then assessed over time and across drivers.

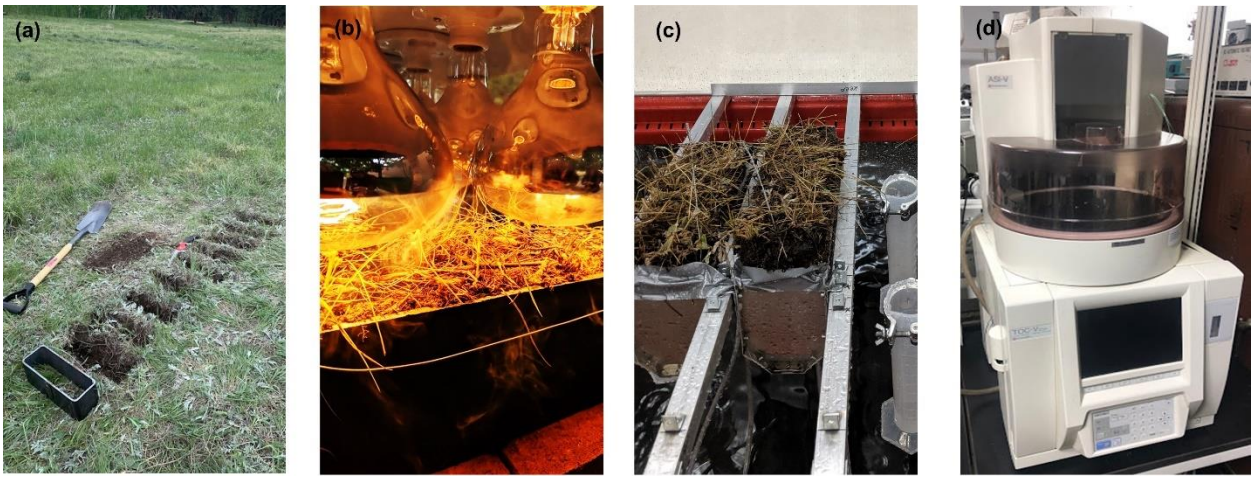

Figure 3.1: The steps involved in the simulation experiment process: (a) excavating soil samples, (b) heating in the wildfire simulator, (c) applying precipitation in the rainfall simulator, and (d) analyzing the chemical properties of runoff in a water quality lab (a Shimadzu TOC-V/TN Analyzer for assessment of dissolved carbon and nitrogen loads pictured here). *Photos courtesy of Carli Brucker.*

3.3.1 Experimental Setup

The experimental framework in this study observed hydrologic and water quality responses within a three-dimensional matrix of burn intensity, rainfall intensity, and terrain slope at four, three, and three different intensity increments, respectively, as shown in Table 3.1. Measuring responses over more than two increments of each driver provided more granular insights into potential response shapes, e.g., linear vs. polynomial, with statistical methods used to assess the strength of drivers' influence, both independently and jointly. A target of eight replicate soil samples were designated for testing at each combination of driver increments. This relatively large number of replicates—compared to the two to five typical across studies reviewed by Brucker et al. (2022)—aided in characterizing and minimizing variability-related uncertainty in response trends. Soil and vegetation characteristics were held as consistent as possible between samples to further isolate effects from the three key drivers. Unburned, or control, samples were also tested to allow for further attribution of burn effects. In addition to the main experimental matrix,

the effects of two sequential rainfall events with a ~24-h drying period in between were tested on 27 additional soil samples, similar to methods in Keesstra et al., 2014.

Table 3.1: Matrix of all combinations of driving factors (burn severities, rainfall intensities, and terrain slopes) at which up to 8 replicate samples were tested. Note, the exact terrain slopes tested were 9.8°, 19.8°, and 29.4°—rounded up to whole numbers for the table.

Burn Intensity (°C)	**Rainfall Intensity (mm/h)**			**Terrain Slope (°)**
	Low (14.3)	**Moderate (26.3)**	**Intense (50.8)**	
	8 replicates	8 replicates	8 replicates	*10*
Unburned (NA)	8 replicates	8 replicates	8 replicates	*20*
	8 replicates	8 replicates	8 replicates	*30*
	8 replicates	8 replicates	8 replicates	*10*
Mild (100-200)	8 replicates	8 replicates	8 replicates	*20*
	8 replicates	8 replicates	8 replicates	*30*
	8 replicates	8 replicates	8 replicates	*10*
Moderate (200-350)	8 replicates	8 replicates	8 replicates	*20*
	8 replicates	8 replicates	8 replicates	*30*
	8 replicates	8 replicates	8 replicates	*10*
Severe (350-600)	8 replicates	8 replicates	8 replicates	*20*
	8 replicates	8 replicates	8 replicates	*30*

At each combination of driver increments and for each replicate, 6 hydrologic and water quality response variables were measured and analyzed: runoff, infiltration, turbidity, total suspended solids (TSS), dissolved organic carbon (DOC), and total dissolved nitrogen (TDN). Runoff and percolation (i.e., liquid drainage through the soil samples) were taken at varying temporal resolutions ranging from 2 minutes to 2 hours (the duration of a rainfall simulation event). This resulted in ~1-17 aliquots of each liquid sample per replicate, allowing for analysis of hydrologic and water quality response over time. The system's water balance equation was then rearranged to calculate change in storage and infiltration terms for each soil sample, assuming no losses occurred, or:

$$\Delta S = P - (R + D + ET)$$

Equation 3.1

and

$$I = D + \Delta S$$

Equation 3.2

where ΔS is change in storage, P is precipitation applied, R is collected runoff, D is drainage or collected percolation, ET is evapotranspiration (assumed to be negligible), and I is infiltration—all in units of mm.

3.3.2 Study area

The Fraser Experimental Forest (FEF) was selected as this study's sampling site due to a nearby burn scar (as seen in Figure 3.2) and previous research providing benchmark burn response data (Lawrence, 2020). The FEF is a 93 km^2 outdoor research laboratory maintained by the U.S. Forest Service, with the St. Louis Creek its main drainage (Alexander and Watkins, 1977). Vegetation regimes are primarily subalpine forests and alpine tundra typical to the central Rocky Mountains, including Engelmann spruce, subalpine fir, and lodgepole pines in the lower elevations (Rhoades et al., 2017). The elevation ranges from 2680 to 3900 m, with about one-third of the forest above the timberline. The climate is generally humid and cool, though it varies greatly with elevation and aspect. Annual precipitation averages 74 cm—nearly two thirds of which is in the form of snow (Essery et al., 2009). Soils are mostly skeletal, sandy loam Dystric and Typic Cryochrepts (Alstatt and Miles, 1983), containing angular gravel and stone with very little silt and clay.

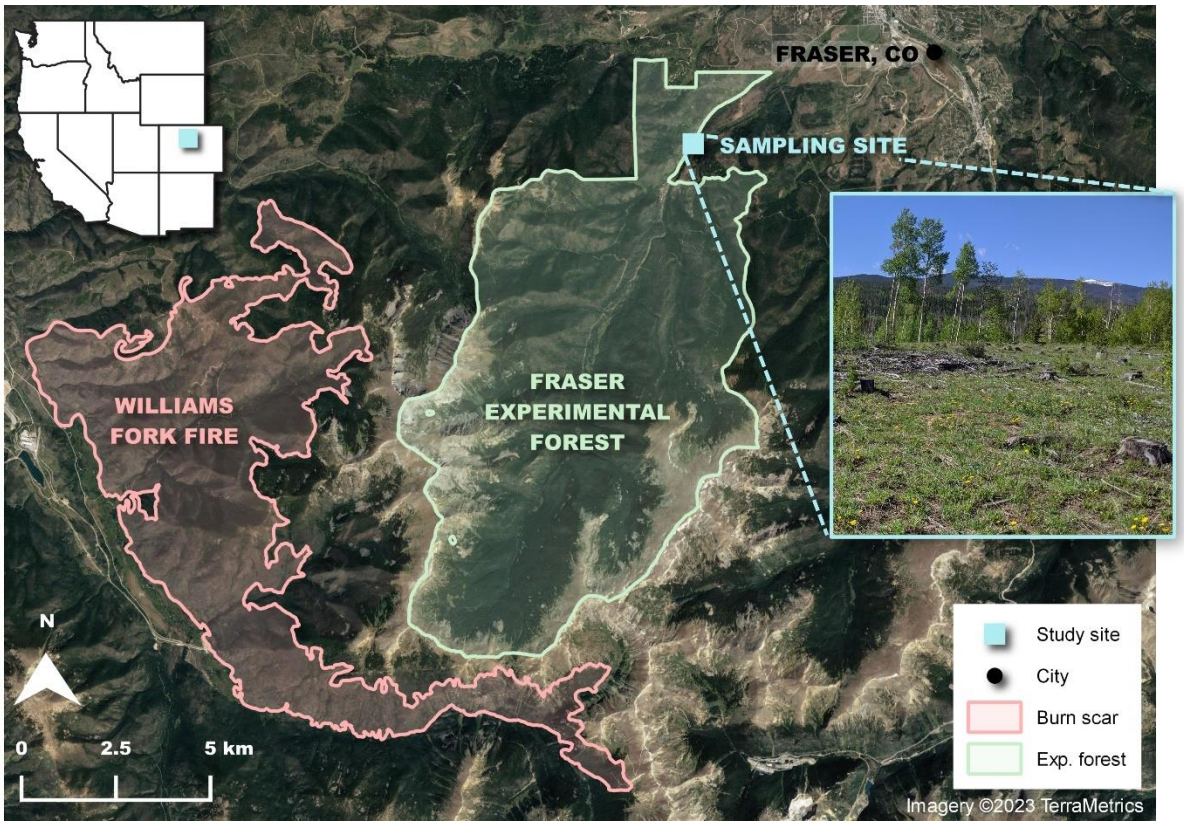

Figure 3.2: Map with the Fraser Experimental Forest overlaid in green and the Williams Fork Fire burn scar overlaid in red. A blue square marker indicates the coordinates of our sampling location, with an inset image of the site. A map of the sampling location in the U.S. West is included in the top-left corner. *Photo courtesy of Carli Brucker.*

The Williams Fork fire occurred several miles southwest of the FEF in 2020, affecting 52 km^2 of the Arapaho National Forest. This fire occurred close to the Williams Fork reservoir and thus the water quality was monitored closely, providing empirical burn data with which to compare experimental data ("Water quality tested in Williams Fork burn area," 2021). A high probability of post-wildfire debris flow was projected for large areas of the burn scar, with downstream water quality closely monitored (Staley and Kean, 2020). Additionally, slash burn simulations conducted in the FEF provided further data about pre- and post-burn soil conditions which can be compared to the simulated results (Lawrence, 2020).

3.3.3 Soil Sample Collection

At a field within this site, soil samples were carefully excavated to preserve natural soil structure and vegetation, then securely fitted into sampling containers designed to minimize structural degradation during transportation and edge effects during testing. Sampling containers 12 x 4 x 4 in (31 x 10 x 10 cm) in size were designed to withstand intense heat from wildfire simulations and allow for collection of sample runoff and percolation. Containers were created from rectangular 30.5 x 10.2 cm steel pipes with 1.3 cm thick walls, which were cut to 10.2 cm segments. Selected for their high melting point, these steel piping-based sampling containers were similar to those in Stoof et al., 2011, also with holes for thermocouple insertion drilled into the sides. This piping-based design allowed for both applied precipitation and percolation collection, with the rectangular shape was more conducive to capturing runoff (Cancelo-González et al., 2013; Keesstra et al., 2014b; Kibet et al., 2014) compared with circular piping used in other sampling containers (Busse et al., 2010; Wieting et al., 2017).

Soil samples were excavated following methods from the USDA Natural Resources Conservation Service South Dakota, 2019. Trimming vegetation to a manageable height (< ~30 cm), sample outlines were cut with a spade then gently lifted from the ground. The sides were shaved down to sampling container dimensions, allowing for gentle sample insertion and minimized compression. The top ~1 in (2.5 cm) of soil was left above the top of the sampling containers, minimizing fall-through and edge gaps from shrinkage after drying, as well as transportation disturbances. All samples were taken from within a ~50 ft (15.2 m) diameter to maximize consistency in soil and vegetation characteristics.

3.3.4 Wildfire Simulator Design

The wildfire simulator apparatus (shown in Figure 3.3) was designed to simulate natural burn mechanisms at a range of intensities through heating and combustion while allowing for repeatable and quantitative burn treatments. Following Cancelo-González et al., 2013 and Klopatek et al., 1988, high-wattage heat lamps were used to heat soil surfaces and subsurfaces (~3 cm below the surface) to temperatures reflective of natural wildfires, or ~100-600°C and ~25-550°C, respectively (Chandler et al., 1983; Jian et al., 2018; Wieting et al., 2017). This mechanism additionally captured combustive and 'top-down' heating mechanisms neglected in other simulation apparati such as muffle furnaces (Brucker et al., 2022). Controlled by the duration of heating applied, burn intensity was precisely measured using soil heating profiles from thermocouple measurements, similar to Busse et al., 2010. While severe wildfires can reach temperatures greater than ~2200°C (Shahlaee et al., 1991), the simulated surface temperatures likely captured most wildfire-driven soil and water chemistry effects (Chandler et al., 1983; Jian et al., 2018; Wieting et al., 2017). Hogue and Inglett, 2012 and Hohner et al., 2019b both showed peak nitrogen and carbon production at temperatures less than ~225-550°C, with marginal loads at higher temperatures due to volatilization. Some wildfire-driven hydrologic effects may have been neglected, however, due to changes in soil physical characteristics which only occur at extreme temperatures (e.g., the destruction of clay at ~800°C) (Neary et al., 2005). Soil subsurface temperatures captured wildfire-driven vegetative root destruction—a key driver of erosion and sedimentation. Temperatures lethal for roots, or >60°C (Busse et al., 2005), were achieved in the subsurface for most mild and all moderate and high severity burns, with medians of 63, 95, and 202°C, respectively. Additionally, ashy combustion residues

produced during simulations had chemical compositions similar to natural fires (Hogue and Inglett, 2012).

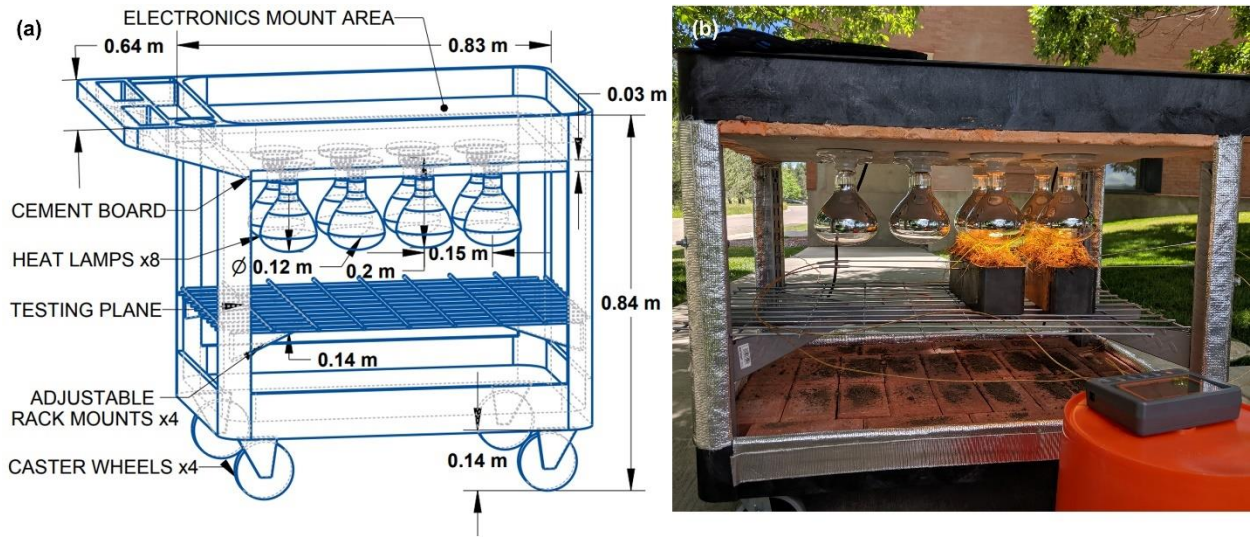

Figure 3.3: (a) Schematic of the wildfire simulator design, including structural components and dimensions. (b) The constructed wildfire simulator applying heat to two soil samples with thermocouples inserted into their sides. *Photo courtesy of Carli Brucker.*

Heat lamps were affixed to a cart apparatus to facilitate transportation and burning of multiple soil samples simultaneously. Eight Philips Infrared 375-Watt Heat Lamps with ~13 cm diameters, similar to those used by Cancelo-González et al., 2013, were affixed to a 0.6 x 1.2 x 0.9 m cart made with a non-conductive material, or high-density polyethylene. Though this material aided in affixing electrical components, its low melting point, or ~125°C (Wei et al., 2010), required fire-resistant cement board to be laid in between the cart and heat lamps, with heat-reflective cloth and fire-resistant spray covering additional exposed areas. Prior to burn simulations, soil surfaces were positioned ~3 cm below the heat lamps using an adjustable metal rack, with aluminum windscreens wrapped around to mitigate wind effects. Two thermocouples were inserted into each sample's surface—underneath a heat lamp and in the center of the sample—with two more at the same lateral positions ~3 cm below. K-type thermocouples, which measured wide temperature ranges

from ~95 to 1260°C (Park, 2010), recorded temperature profiles at 5-s intervals to *Gain Express AZ 4-Channel SD* data loggers.

3.3.5 Burn Severity Characterization

Two burn intensity characterization methods were chosen for analysis based on their conduciveness to quantitative analyses and repeatability of treatments at specific intensities. We used both a temperature-based burn severity scale as well as a metric referred to as *degree-hours*, which incorporates both burn temperature and duration (Blank et al., 1994; Cancelo-González et al., 2012; Keeley, 2009; Lentile et al., 2006; Stoof et al., 2011; Wang et al., 2015c). The temperature-based scale was initially used for completing replicates in the experimental matrix, with samples secondarily characterized by degree hour intensities during post-experimental analyses. These were selected through an evaluation of six potential burn intensity characterization methods completed at the beginning of the study, shown in Figure C.1 in *Appendix C*.

The temperature-based characterization was based on peak temperatures achieved at the 'hottest' lateral location on each soil sample's surface during burn simulations. This was derived from temperature measurements from thermocouples placed at multiple lateral positions across samples' surfaces, as described in *Section 3.3.4*. As of yet, no universal surface temperature-based wildfire intensity scale exists (Keeley, 2009; Moody et al., 2013), as qualitative characterizations of burn severity such as the type of vegetation consumed (Emmerich and Cox, 1992; Nyman et al., 2014) or ash color (Moreno and Oechel, 1989) are more typical in wildfire studies (Brucker et al., 2022; Keeley, 2009). Thus, we created a unique scale based on temperature-severity characterizations referenced across

previous studies (Chandler et al., 1983; Hohner et al., 2019; Jian et al., 2018; Robichaud and Hungerford, 2000; Wang et al., 2015c; Wieting et al., 2017), as seen in Table 3.2.

Table 3.2: Burn intensity levels used in this study, characterized using a temperature-based scale and degree hours, and the median heating duration to achieve each intensity during wildfire simulation.

Burn Severity	Surface Temperature Characterization (°C)	Degree Hour Characterization (°C-h)	Median Heating Duration (min)
Low	100-200	8-39	2.5
Moderate	200-350	39-110	5.8
Severe	350-600	110-993	9.3

Degree-hours were calculated for each sample by interpolating under time-temperature curves associated with the 'hottest' area of samples' surfaces, i.e., using measurements from the same thermocouple as the temperature-based characterization. As seen in Figure 3.5 in *Section 3.4.1*, the area under the time-temperature curve was summed at 5 s intervals from the beginning of the simulation until temperatures had cooled to less than ~100°C, with initial ambient temperature subtracted. The equation used to calculate this metric was a slightly altered version of the equation from Cancelo-González et al., 2012:

$$DH = \sum (T_S - T_{Amb.}) * (5\ \text{sec}) * (\frac{1\ hr}{3600\ \text{sec}})$$

Equation 3.3

where DH = degree-hour in °C-h, T_S = the surface temperature at each 5 s interval in °C, and $T_{Amb.}$ = the ambient temperature recorded at the start of the wildfire simulation in °C. As no standardized intensity categorization of degree-hours exists, burned samples were

characterized as ‘mild’, ‘moderate’, or ‘severe’ using terciles of all degree-hour values calculated from the experiment. Note, as burn intensities were characterized using the thermocouple location with the highest achieved surface temperatures for both methods, soil samples in aggregate likely achieved lower, more spatially variable burn intensities.

3.3.6 Wildfire Simulation Setup and Procedure

The wildfire simulation procedure applied heating to soil samples for durations necessary to achieve target burn intensities, recording the soil heating profiles produced. Burn intensity was characterized using both a temperature-based and temperature-time (degree hours) scale (shown in Table 3.2: Burn intensity levels used in this study, characterized using a temperature-based scale and degree hours, and the median heating duration to achieve each intensity during wildfire simulation.), as discussed in the above section. Heat lamps and data loggers were first switched on, heating samples until temperatures associated with desired temperature-based burn intensities were achieved at the hottest areas of their surfaces. The lamps were then switched off, while the data loggers continued recording until surfaces returned to near-ambient temperatures, typically ~1-1.5 hours from the start of the simulation. Experiments were only scheduled for days with no freezing weather, precipitation, or high winds to maximize consistency in ambient climate.

3.3.7 Rainfall Simulator Design

The rainfall simulator was designed to simulate natural rainfall mechanisms at a wide range of intensities by producing similar droplet sizes, kinetic energies, and distributions. Based on apparati used in Cancelo-González et al., 2013 and Kibet et al., 2014, rainfall was generated by differently-sized nozzles which produced precipitation intensities common in the FEF, as seen in Figure 3.4. These nozzles were affixed atop of a

tall (2.3 m) steel frame, providing sufficient height for even droplet distribution and adequate fall time to achieve natural droplet kinetic energies. *FullJet*® nozzles (Spraying Systems Co., Bloomingdale, IL, United States) with conical, downward spray were used—rated to produce natural droplet sizes (~0.5-4 mm (Ulbrich, 1983)) and kinetic energies (~0.1-28 $J/m^2 \cdot mm$ (Yonter and Houndonougbo, 2022)) when operated at appropriate pressures and heights. As seen in Table 3.3, nozzle sizes HH-4.3W (small), HH-8W (medium), and HH-20W (large) were selected, achieving average 14.4 mm/h, 26.4 mm/h, and 51.3 mm/h intensities, respectively, which corresponded roughly to historical average 10-year, 200-year, and 1000-year 2-h rainfall events within the FEF ("Precipitation Frequency Data Server," 2017). The frame height used here was similar to previous studies testing the same nozzles which reported kinetic energies comparable to ~90% of natural values (Yonter and Houndonougbo, 2022) and was sufficient for producing even spatial distributions of droplets, despite the conical spray produced by nozzles. Assessment of rainfall distribution, selection of nozzles' optimal operating pressures (shown in Table 3.3), and interpolated precipitation estimates for each soil sample is discussed further in *Section 3.4.2*.

Table 3.3: Rainfall intensity levels and their associated *FullJet*® nozzles, as well as return intervals for the FEF, median rainfall intensities produced, optimal operating pressures (or

those which produced the lowest rainfall spatial variability (Kibet et al., 2014; Tossell et al., 1987; Yonter and Houndonougbo, 2022)), and average system flow rates produced.

Rainfall Level	Nozzle Size	Return Int. (yr)	Median Rainfall Intensity (mm/h)	Optimal Operating Pressure (kPa)	Average Flow Rates (L/m)
Low	HH-4.3W	10	14.4	69	1.5
Medium	HH-8W	200	26.4	62	1.9
High	HH-20W	1000	51.3	62	6.8

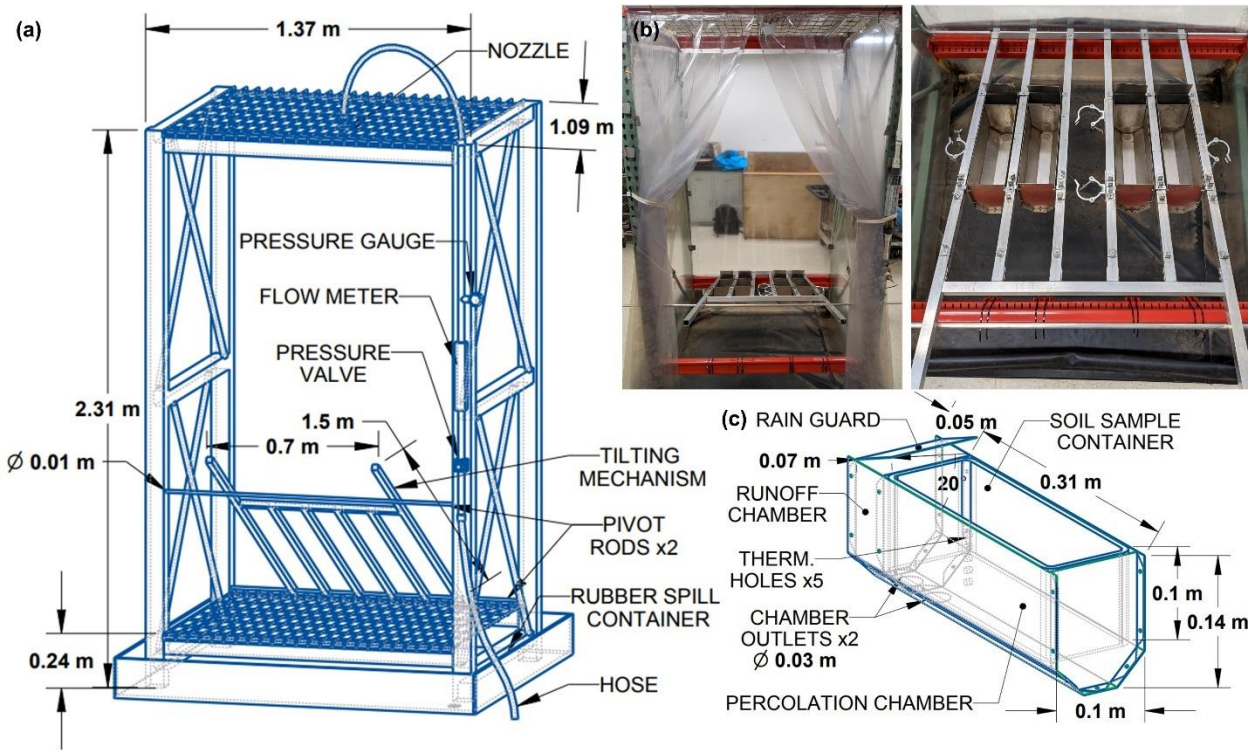

Figure 3.4: (a) Schematic of the rainfall simulator design including structural component dimensions and plumbing features. (b) Front view of the rainfall simulator and a top-down view of the tilting mechanism inside, with custom funnels put in place. (c) A similar schematic for the custom funnels, shown with an inserted soil sampling container. *Photos courtesy of Carli Brucker.*

The simulator plumbing and structural features were designed to facilitate this nozzle system. As seen in Figure 3.4a, a flexible hose attached to the lab sink supplied nozzles with a consistent flow of tap water, with a pressure gauge, pressure valve, flow meter, and ball valve installed inline similar to Kibet et al., 2014. While previous studies have used pump systems to regulate flow (Keesstra et al., 2014b; Kibet et al., 2014), the lab sink supplied sufficient and consistent pressures and flow rates to the system (Kibet et al., 2014). The tap water quality was generally similar to typical rainfall in the FEF, with a pH of ~7 and DOC and TDN concentrations of 1.1 and 0.1 mg/L, respectively. However, the higher pH compared to typical rainwater (~5.6) may have contributed so slightly lower rates of DOM transport (Meyer-Jacob et al., 2019). The simulator frame had a 1.2 x 1.1 m lateral testing plane, with clear plastic covering the sides and an industrial-grade rubber spill container underneath to drain excess rainfall.

3.3.8 Tilting Mechanism and Custom Funnel Design

Affixed in the rainfall simulator, an aluminum tilting mechanism (Figure 3.4b) held four soil samples in at a range of terrain slopes typical in the FEF, or ~9.3°, 19.8°, and 29.4°—interfaced using custom soil sample funnels designed to capture and separate hydrologic responses (Figure 3.4c). These angles were identified as common low, medium, and high terrain slopes in the FEF using a frequency analysis performed on the area's digital elevation model. A 0.7 x 1.0 m aluminum frame was constructed to pivot at the back of the rainfall simulator while resting on an adjustable rod at the front, placed at different vertical locations to set terrain slopes. Graduated cylinders at all four sides of the tilting mechanism measured simulated precipitation, affixed using bracket sets which positioned them at constant lateral locations for all three terrain slopes. Following a similar design

concept as Kibet et al., 2014, custom aluminum funnels were created to hold soil samples within their containers, separating and collecting runoff and percolation generated by precipitation. Bracket sets for the funnels were placed on the tilting frame to hold their lateral positions constant, similar to the graduated cylinders. The runoff collection chamber was placed at the 'downhill' end of the funnel, covered by an angled piece of aluminum to deflect precipitation, with the percolation chamber underneath the sample. Small plastic funnels were affixed to the bottom of each chamber, channeling liquids through attached plastic tubes which led to sample bottles placed outside the simulator.

3.3.9 Rainfall Simulator and Tilting Mechanism Setup and Procedure

To prepare the apparatus for testing, first a nozzle was screwed into its fixture, the tilting mechanism slope set, and plastic tubing, sample bottles, and graduated cylinders for measuring precipitation put in place. Four soil samples in their sampling containers were then placed in the custom funnels in the tilting mechanism. Following methods similar to Wieting et al., 2017, petroleum jelly and duct tape were used to seal any preferential flow paths around the edges of the soil samples, or holes allowing excessive amounts of infiltration to run through. Next, simulated precipitation was applied for two hours by opening the valve in the nozzle plumbing. 1000 mL plastic jars collected the infiltration and 300 mL plastic bottles collected the runoff, with full bottles switched for empty ones as needed. Runoff from one of four samples in each rainfall simulation was collected in smaller aliquots (60 mL), at 2-minute intervals for the first 10 minutes post-runoff initiation, then 5-minute intervals for the next 10 minutes, then every subsequent 10 minutes. Runoff and infiltration samples were weighed, then the runoff samples frozen, as well as tap water samples taken from each run.

3.3.10 Water Chemistry Analysis

Runoff samples were defrosted, then analyzed in a lab to assess total suspended solids (TSS), dissolved organic carbon (DOC), and total dissolved nitrogen (TDN) levels, as well as turbidity. Turbidity was first measured by gently inverting each sample several times to homogenize suspended solids, then testing a 30 mL subset in a Hach 2100N Turbidimeter following the U.S. Environmental Protection Agency (USEPA) Method 180.1 (Hatch Corporation, 2014; O'Dell, 1993). Next, TSS was measured following the Total Suspended Solids Method 2540 D from *Standard Methods for the Examination of Water and Wastewater, 22nd ed.* (American Public Health Association and American Water Works Association, 2012). Using a vacuum pump, samples were filtered through 0.7 µm glass filters, dried in an ~104°C oven for 1 h, and weighed before and after filtering to determine collected sediment mass. The filtered liquid was then tested for DOC and TDN using a Shimadzu TOC-V/TN Analyzer. Samples were prepared by pouring them into 24 mL glass vials and adding 1M of hydrochloric acid, acidifying them to a pH of 2-3 (Shimadzu Corporation, 2001). The Shimadzu instrument then measured DOC by sparging samples with high-purity air, removing inorganic carbon (i.e. carbonates and bicarbonates), then determining the non-purgeable organic carbon (Shimadzu Corporation, 2001). TDN was measured simultaneously through a similar oxidation process. The machine's calibration curves were created using a standard, with peak DOC and TDN concentrations of 25 and 5 mg/L, respectively. Samples with higher concentrations were re-tested with a 1:1 dilution.

Though the USEPA method 415.3 defines dissolved constituents as < 0.45 µm (Potter and Wimsatt, 2012), the 0.7 µm filters used for TSS analysis were also deemed sufficient for DOC and TDN testing. Samples filtered through this size had similar solute concentrations to those filtered at 0.45 µm and fulfilled requirements for testing on the

Shimadzu TOC-V/TN Analyzer (Shimadzu Corporation, 2001). Pairwise t-tests performed on six samples filtered separately through 0.7 μm and 0.45 μm filters showed no significant differences in dissolved organic carbon (DOC) and total dissolved nitrogen (TDN) concentrations (p = 0.74 and p = 0.50, respectively), with median absolute differences of 5.8% and 4.0% (ranges of -7.2 to 10.9% and -5.7 to 15.3% differences), respectively, between the two sizes. Comparatively, median absolute differences in duplicates (samples tested multiple times during the same run) across all samples were slightly lower, or 1.4% and 1.8% for DOC and TDN concentrations, respectively, but with wider ranges of percent differences, or -48.8 to 14.2% and -63.6 to 15.6%, respectively. Thus, the differences in concentrations produced by the two filter sizes were deemed marginal—no higher than the existing variability between duplicate samples generated by experimental error. Additionally, 0.7 μm filter sizes were within specifications for testing on the Shimadzu TOC-V/TN Analyzer, which required use of 1.5 μm filter sizes or finer (Shimadzu Corporation, 2001).

3.3.11 Statistical Analysis

Observed hydrologic and water quality responses were assessed for significant changes and trends across burn intensity, rainfall intensity, and terrain slope increments using statistical analyses. First, ratios of responses over runoff and precipitation values were calculated to assess changes independent of varying rainfall intensity or hydrologic mechanisms. Runoff ratios, a unitless metric, were calculated by dividing generated runoff by the precipitation applied to each sample and suspended sediment concentrations (SSC) were calculated by dividing TSS by total runoff. A one-way analysis of variance (ANOVA) test was performed for each driver-response comparison to assess for significant changes

across driver increments. Pairwise t-tests further assessed significant changes between groups at specific increments. Linear regressions and R-squared values evaluated the linearity of responses. A significance level of $\alpha = 0.05$ was used for all tests, where applicable. A preliminary, small-scale physical modeling analysis of water movement through the tested soil samples is discussed in *Appendix E*.

3.4 Results

First, a validation of the wildfire simulator's ability to emulate natural burn mechanisms at a range of intensities with repeatable, quantifiable burn treatments is presented. Next, selection of the rainfall simulator's operation settings using rainfall distribution tests is discussed, as well as a validation of the simulator's height and more accurate precipitation intensity estimates. Finally, key runoff, sediment, and solute responses from soil samples tested in the experimental matrix are presented.

3.4.1 Wildfire Simulator Calibration

The duration of heat applied from the heat lamps, i.e., the forcing mechanism for driving varying burn intensities, was generally linearly related to peak temperatures achieved ($R^2 = 0.32$). The median time to achieve low, moderate, and severe burn intensities was 2.5, 5.8, and 9.3 min, respectively. However, antecedent soil moisture and ambient temperature during burning may also have influenced heating profiles (Busse et al., 2010; Klopatek et al., 1988; Reardon et al., 2007; Stoof et al., 2011), with high moistures and low temperatures driving outliers when heating durations were plotted against peak temperatures, as shown in Figure 3.5a and Figure 3.5b. The R^2 values for these plotted variables rose to 0.74 and 0.5, respectively, when only moistures in the 1st tercile and ambient temperatures in the 2nd tercile were included.

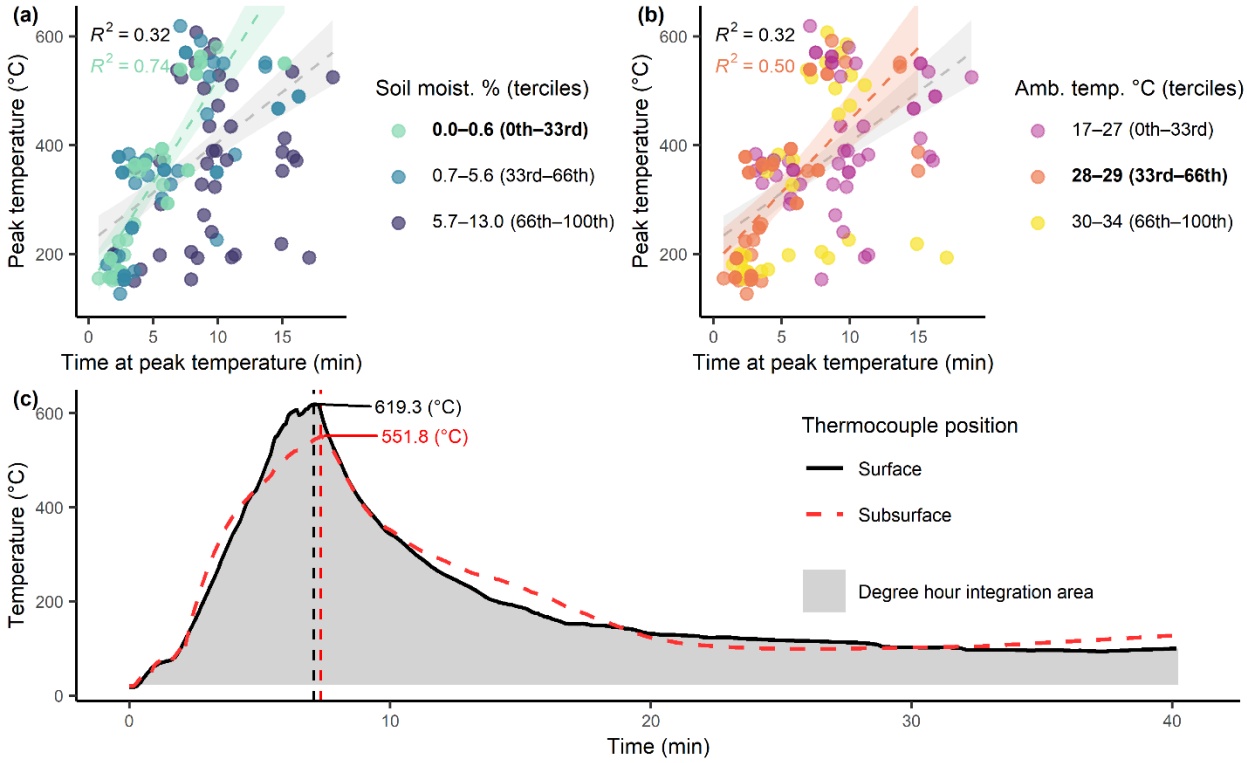

Figure 3.5: (a) Comparison of peak temperatures and heating durations, with colors showing different antecedent moisture contents split into terciles. The dashed lines show the best linear fit of the data, with the gray line representing all data (R^2 = 0.32) and the teal line representing data just in the 1st antecedent moisture tercile (R^2 = 0.74). This tercile (i.e., the one with the highest correlation of temperatures and durations) is bolded and greyed areas represent confidence intervals (level = 0.95). (b) Peak temperature and heating similarly plotted, but with colors showing different ambient temperatures during the simulation. The orange dashed line represents data in just the middle tercile (bolded) with R^2 = 0.50. (c) Time-temperature curves from two thermocouples placed at the soil surface (solid line) and 3 cm below the soil surface (dashed, red line) during a severe burn simulation. Dashed lines show the time when peak temperatures were achieved, or 619.3°C and 551.8°C for the surface and subsurface, respectively, and the gray area represents the area under the surface temperature curve integrated to calculate degree hours.

Time-temperature curves from the thermocouple measurements showed surface temperatures as high as ~600°C achieved, capturing the full range of wildfire effects as discussed in *Section 3.3.4*. Higher values were likely limited by heat loss in the space between heat lamps and soil surfaces, though temperatures were still notably higher than those in previous studies using similar burning techniques (Cancelo-González et al., 2013; Klopatek et al., 1988). As expected, subsurface heating profiles showed a delayed and

muted response to heating, similar to previous studies (Cancelo-González et al., 2013; Stoof et al., 2011), with temperatures necessary for root destruction achieved in most cases (median values of ~63-202°C).

3.4.2 Rainfall Simulator Calibration

In addition to assessing even spatial distributions of droplets produced at the rainfall simulator's height, rainfall distribution testing was used to select nozzles' optimal operating pressures and create more accurate estimates of precipitation applied to each soil sample. To evaluate distribution, 20 measuring devices were placed across the bottom of the rainfall simulator. These collected volume measurements from three 1-h rainfall events for each nozzle and at three different pressures (or 27 trials total). Optimal operating pressures were then selected based on the lowest-variability distribution tests (Kibet et al., 2014; Tossell et al., 1987; Yonter and Houndonougbo, 2022), furthering uniformity of simulated rainfall distribution. 62 kPa produced the lowest variability for both the HH-8W and HH-20W nozzles, with coefficients of variation (standard deviation / mean) of 0.0071 and 0.0137, respectively. 69 kPa produced the lowest variability for the HH-4.3W nozzle, with a coefficient of variation of 0.0069.

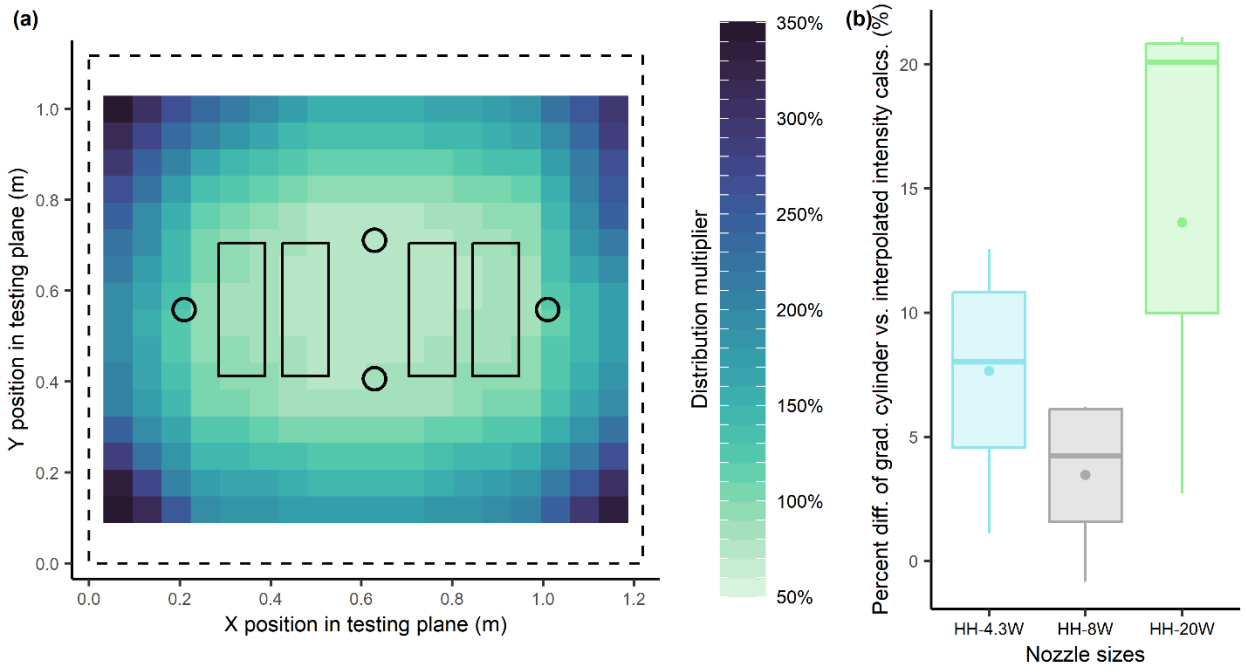

Figure 3.6: (a) Interpolated schematic of rainfall distribution across the footprint of the rainfall simulator for the HH-20W nozzle as an example, normalized by average graduated cylinder measurements to allow for localized estimates of precipitation applied to each sample. The dashed line represents the full extent of the testing plane. (b) Percent differences of graduated cylinder rainfall intensity estimates from values interpolated from distribution maps for each nozzle size.

As rainfall was spatially variable across the testing plane, total rainfall interception for each soil sample was interpolated using both the original graduated cylinder precipitation estimates, as well as distribution test measurements. First, an estimate of rainfall distribution for each simulated event was calculated by projecting the precipitation measured by the graduated cylinders onto an interpolated rainfall distribution map for the appropriate nozzle size. Intensities were then averaged over each soil samples' area in the x-y plane of the simulator—altered based on simulated terrain slope. Though all nozzles' coefficients of variation were satisfactorily low, these results indicated that rainfall intensity was more evenly distributed for the HH-4.3W and HH-8W nozzles than the HH-20W. Due to this higher variability, interpolated rainfall estimates were the most different from the original graduated cylinder measurements for high rainfall intensities, or up to

20% as shown in Figure 3.6. Interpolated estimates for medium and low rainfall intensities were different from original estimates by as much as 7 and 11%, respectively.

3.4.3 Simulation Experiment Responses

Results from tested soil samples were highly variable, but generally consistent with previous studies (Bladon et al., 2014; Hohner et al., 2019; Johansen et al., 2001; Moody et al., 2013). Runoff ratio, SSC, and turbidity generally increased monotonically with increasing burn severity, while both DOC and TDN tended to show an inverse "U" shaped trend with increasing burn severity similar to past research (Becker et al., 2018a; Hohner et al., 2019; Johansen et al., 2001). A total of 154 soil samples were tested in the experimental matrix, though a low number of replicates were achieved for the lowest rainfall intensity and terrain slope settings and thus were excluded from analysis. This was due to limitations in time and resources, as these settings resulted in acutely delayed responses during rainfall simulation, with runoff generation only initiating 4+ h into the event. Grain size distribution and hydrometer analyses of soil samples, following methods by Das and Sobhan, 2010 and USDA Particle-Size Classifications (García-Gaines and Frankenstein, 2015), showed a composition of ~5% gravel, ~82% sand, ~10% silt, and ~3% clay—similar to soils characterized as loamy sand.

Hydrologic, Sediment, and Turbidity

General increases in runoff, percolation, and sedimentation rates, as well as turbidity, were observed at increasing burn intensities, though with high variability. A boxplot with each of these constituents plotted at each driver increment is included in Figure C.3 in *Appendix C*. Runoff ratios generally increased monotonically with increasing burn intensity at 30° terrain slopes, reflecting results in previous literature (Bladon et al.,

2014), with a similar, but less clear trend at 20° terrain slopes. Unburned samples had anomalously high responses, however, likely due to differences in soil sample handling as discussed in *Section 3.5.2*, lower loss rates as discussed in *Appendix D*. When unburned samples were removed, an ANOVA test for responses at 30° terrain slopes had a *p*-value of 0.047, meaning runoff ratios significantly increased with each burn intensity increment. Additionally, values at severe burn intensities were ~50% higher than at mild burn intensities. Runoff ratios were also observed over time, as seen in Figure 3.7a. Runoff ratio increased more rapidly over the duration of simulated precipitation for the high burn intensity samples as compared to the other burn categories, then leveled off. Figure 3.7 shows that runoff ratios during the first 60 min increased monotonically with burn intensity, with severely burned samples significantly higher ($\alpha = 0.05$) and almost twice the median value compared to mild burn samples. Similar to reports of 'flashy' runoff in fire-affected areas in previous studies (Bladon et al., 2014), this may be due to initial hydrophobic properties in burned soils, which may have resolved later on in the simulation.

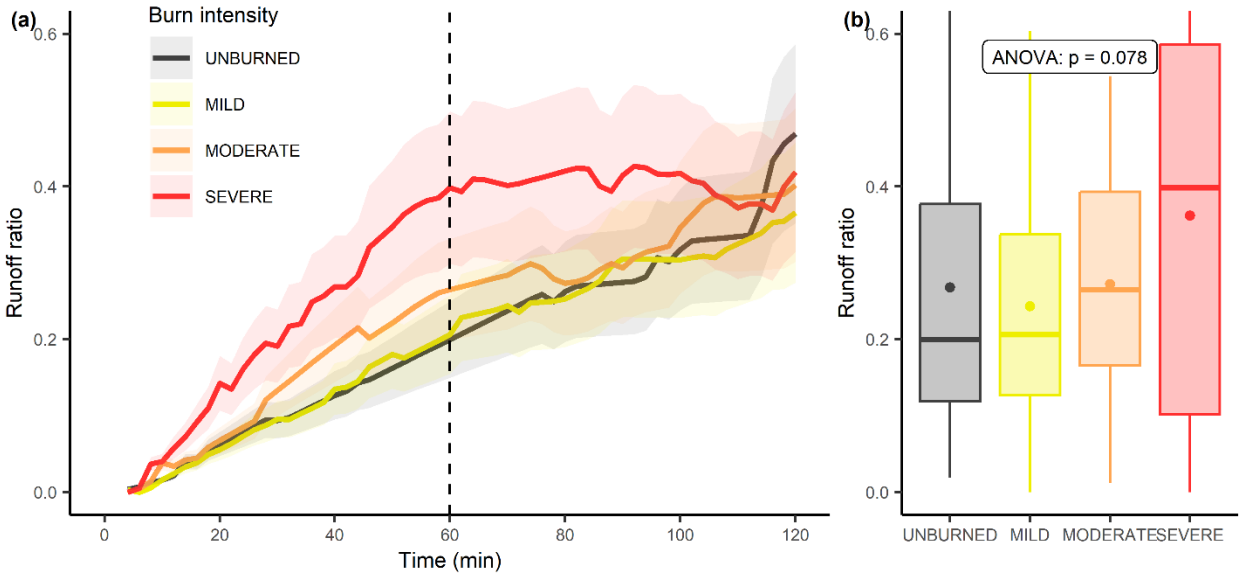

Figure 3.7: (a) Time-series plot of median runoff ratio, beginning when runoff was first produced and ending at the completion of the 2-h simulated rainfall event. Colored lines represent burn severities, with

shaded regions indicating inner quartile ranges. A dashed line shows the 60-min mark. (b) Box plots of runoff ratios calculated for the first 60 min of rainfall simulations. An ANOVA p-value of 0.078 indicated that each burn intensity group was not significantly different from all other groups ($\alpha = 0.05$). However, a t-test between the severe and mild burn groups had a p-value of 0.01, indicating that the severe burn group was significantly higher.

Percolation also showed a significant, monotonic increase with increasing burn intensity when characterized using degree hours, with infiltration calculations showing no relationship. Median percolation increased by ~170% from unburned to severe burn intensities and an ANOVA test over all intensities had a *p*-value of 0.0089, indicating values increased significantly at each increment. This trend may have been partially due to greater structural disturbance in burned samples from burn-induced soil destruction, as well as transportation to and from the burn simulation site, as discussed *Section 3.5.2*. Additionally, runoff responses may have been muted by these effects, with increased runoff and decreased percolation responses in the less-disturbed unburned samples. Infiltration was calculated by summing percolation and estimated storage, then infiltration ratio derived by dividing infiltration by total precipitation for each sample. These values had no significant trends with increasing burn intensity, though a slight decrease (~10%) was observed from 20° to 30° terrain slopes with an ANOVA *p*-value of 0.055.

SSC generally increased with increasing burn intensity, with the clearest trends observed when compared to degree hour characterizations, though anomalies existed for unburned samples and at moderate rainfall intensities and 30° slopes. Unburned samples often produced similar or greater SSC values as compared with the burned samples—contrasting from previous studies which showed SSC increases with burning (Blake et al., 2010; Knight et al., 1983; Shahlaee et al., 1991). This may have been due to experimental error from different testing methods applied to burned and unburned samples, as discussed in *Section 3.5.2*, or higher loss rates for burned samples as discussed in *Appendix D*.

However, when unburned samples were removed from analyses, a clear monotonic trend existed between SSC and burn intensity characterized by degree hours, as seen in Figure 3.8a, similar to previous literature (Robichaud, 2005b; Smith et al., 2011). An ANOVA test for SSC over mild to severe burn intensities had a *p*-value of 0.013, indicating a significant increase with each increasing burn intensity. Sediment trapping in the system was evident, however, from a 44% decrease in median TSS from the 20° to 30° slope angle, despite median runoff, sediment's main transport mechanism, experiencing a 41% increase. The higher angle may have increased sediment settling in the corners of the custom funnels or allowed for the transport of larger sediment particles which clogged the tubing system.

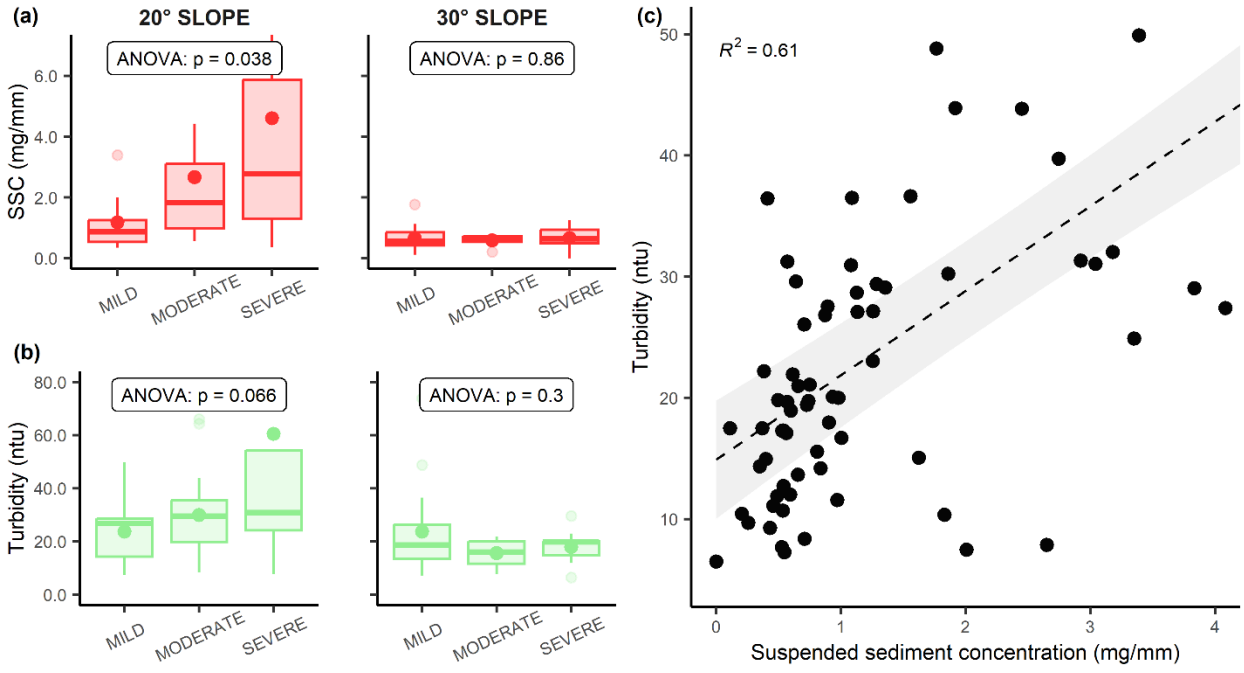

Figure 3.8: (a) Boxplots of SSC with increasing burn intensity characterized by degree hours, with unburned samples removed, for 20° and 30° terrain slopes. (b) Similar boxplots, but for turbidity. ANOVA *p*-values displayed indicate that intensity groups are not significantly different from all other groups for each case. (c) Turbidity plotted against SSC with unburned samples removed. The dashed line represents the best linear fit of data (R^2 = 0.61), with the greyed-out area representing the confidence interval (level = 0.95).

Turbidity also generally increased monotonically with burn intensity characterized by degree hours and had a strong relationship with SSC. Similar to SSC response, turbidity was unexpectedly high for unburned samples, but showed a clear linear trend with burn intensity when unburned samples were removed as seen in Figure 3.8b. For the 20° slope case, turbidity increased significantly with each increasing burn intensity similar to previous literature (Becker et al., 2018a), evident from an ANOVA p-value of 0.033. Here, mean turbidity was 55% higher for severely burned samples than mild burn samples. Additionally, turbidities were strongly correlated with SSC with an $R^2 = 0.61$, as seen in Figure 3.8c.

Dissolved Constituent Response

As seen in Figure 3.9a and Figure 3.9b, both DOC and TDN increased incrementally from unburned to moderate burn intensity groups, peaked at 200-300°C, then decreased from moderate to severe burn samples. Median concentrations at moderate intensities were 44 and 112% higher than unburned samples, respectively, and 65 and 15% higher than severe burn samples. ANOVA p-values of 0.044 and 0.003 for DOC and TDN, respectively, meaning the observed inverse 'U' shaped trends were significant ($\alpha = 0.05$). Similar to findings in Hohner et al., 2019b, this was likely due to low to moderate temperatures (i.e., below ~350°C) releasing carbon and nitrogen, then more extreme temperatures (~350-600°C) volatilizing constituents. Additional discussions of simulated responses, including results from sequential rainfall applications, are included in *Appendix C*.

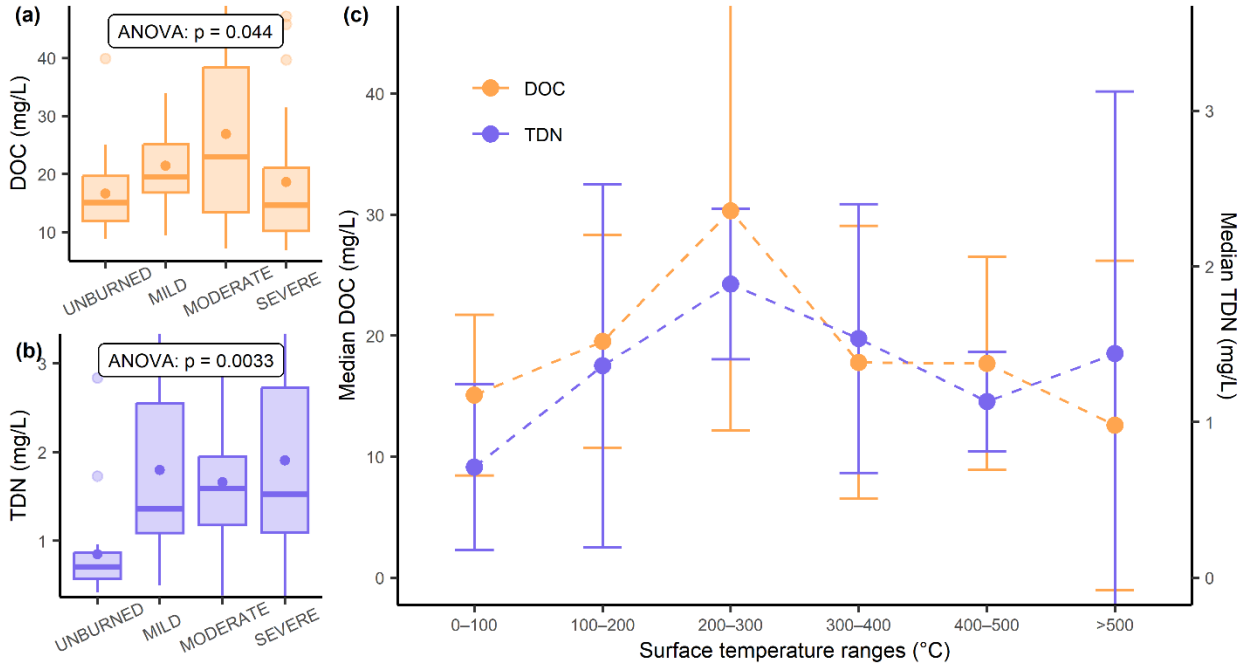

Figure 3.9: (a) Boxplot of DOC concentrations with increasing burn intensity levels, showing an ANOVA test *p*-value of 0.044. (b) A similar boxplot for TDN concentrations, with an ANOVA *p*-value of 0.003. (c) Median DOC and TDN values binned into increasing surface temperature ranges. Error bars show medians plus and minus one standard deviation.

3.5 Discussion

Calibration and validation testing results from the simulators confirmed their ability to generate controlled, replicate intensities at ranges capturing key mechanisms and wide variability in post-wildfire environments. The controlled setting of the experiment additionally allowed for isolation of factors and provided key insights into independent and joint interactions of drivers with responses. However, anomalous responses also revealed experimental limitations, contributing uncertainty to results.

3.5.1 Experiment Validation and Key Insights

The wildfire simulator achieved temperature ranges, validated by time-temperature curves, and combustion processes important in natural burn mechanisms. The driving mechanism—duration of heating applied—was strongly correlated with peak temperature,

allowing for control over burn intensity achieved for each sample. The rainfall simulator achieved rainfall intensities typical for the FEF with droplet size and kinetic energy targeting natural ranges. Detailed quantification of precipitation distribution, as well as control over its timing, allowed for application of targeted precipitation amounts.

Observed hydrologic and water quality responses generally reflected trends consistent with *in situ* and simulated data in previous literature (Badía-Villas et al., 2014; Hohner et al., 2019; Moody and Martin, 2009a), with the strongest correspondence of trends observed in solute responses. Runoff ratio showed a relatively clear, monotonic relationship with increasing burn intensity, suggesting sensitivity to burn effects and possible increased soil water repellency (Badía-Villas et al., 2014; Robichaud and Hungerford, 2000). Sedimentation response was similarly related to burn intensity, reflecting *in situ* studies, though high variability obscured trends in some cases (Lane et al., 2006; Larsen and MacDonald, 2007; Moody and Martin, 2009b; Shahlaee et al., 1991). DOC and TDN responses had shapes which closely reflected results from previous laboratory-scale wildfire studies, such as Hohner et al., 2019b, with lower variability and higher statistical significance than hydrologic and sediment responses. The similar response trends in these two constituents is additionally consistent with previous studies showing similar release mechanisms in both DOM and nutrients, where combustion of soils and vegetation leads to increased constituent deposition and sorption rates (Certini, 2005; J.-J. Wang et al., 2015b). Trends in turbidity were less clear, though the highest turbidities were observed at high burn intensities, similar to previous literature (Becker et al., 2018a; Hohner et al., 2016).

3.5.2 Experimental Limitations

Inconsistencies in sample testing methods, as well as pooling and unintended flow paths in the rainfall simulator setup, contributed uncertainty to results, evident from anomalous results. Where lower sediment response from unburned samples was expected (Moody and Martin, 2009a), median SSC was 26% greater for unburned than burned samples—potentially due to greater disturbance via additional handling of burned samples. While unburned samples were never removed from the lab, burned samples were disturbed during transportation to an outdoor testing area, exposed to light winds, and inserted with thermocouples, disturbing soil structure. Significant mass loss as high as 15% occurred during this step, some of which may be attributed to volatilization of vegetation and soils, but also due to soil loss from disturbances during the testing procedure.

Differences between estimated water storage and change in soil moisture, as shown in Figure 3.10, may be suggestive of trapped water and unaccounted flow paths in the system. Storage estimates were made assuming a closed water balance with no losses for each soil sample, however a median ~15 mm difference in storage from change in soil moisture reflects experimental uncertainties. Water may have been trapped in the custom funnels, blocked by sediment lodged in the tubing system, or flowed laterally over the sides of samples. A thin (inner diameter ~0.6 cm) plastic tubing was used to transport liquids from funnels' runoff and infiltration chambers to drive a higher flowrate capable of transporting sediment, however larger particles frequently became lodged in the tubing. Additionally, the system relied on gravity to transport liquids and sediment to the sample bottles placed outside the simulator, however sections of tubing often sagged, temporarily trapping liquids. Sediment trapping was also evident from a 44% decrease in median TSS

from the 20° to 30° slope angle, potentially due to increased sediment settling in the corners of the custom funnels at higher angles.

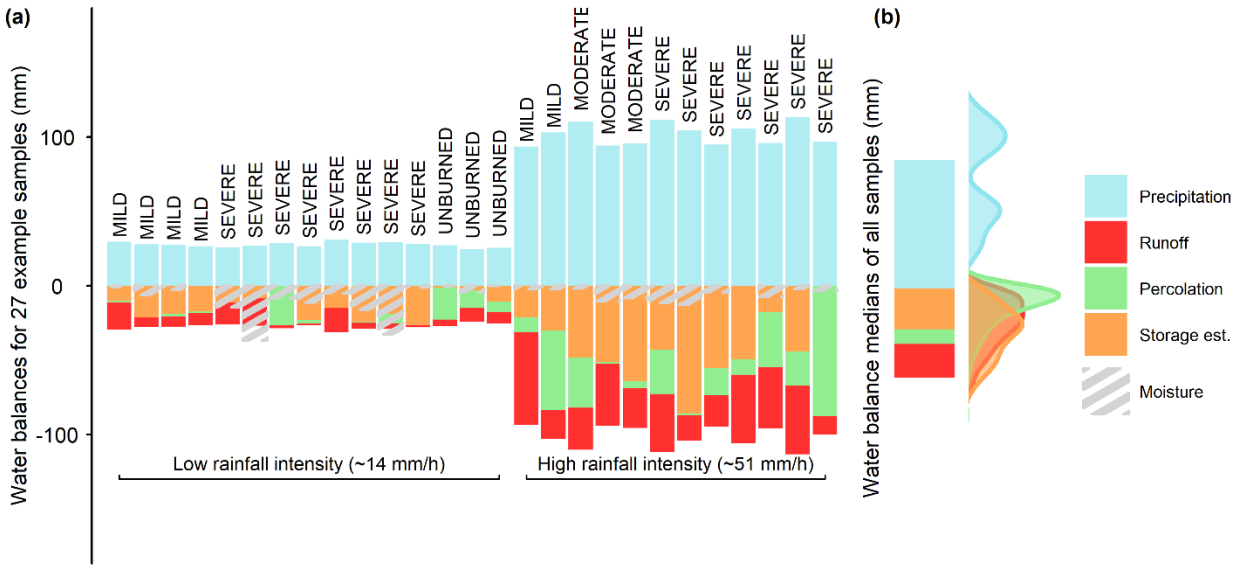

Figure 3.10: (a) A composite bar graph of water budget components for several soil samples, showing water applied to samples through precipitation, as well as subsequent runoff, percolation, and estimated storage in units of mm (colored bars). The change in soil moisture, which should approximate estimated storage, is overlayed (gray hatches). (b) Median values for water balance components across all samples, as well as their distributions shown by density plots on the right of the bar chart.

The small scale of the soil samples tested in our experiment also limited observable hydrologic and sedimentation processes which occur in wildfire-affected basins. On hillslope and basin-scales, erosional effects, such as rill erosion which expands channel networks, are enhanced by loss of both vegetation (including ground and canopy cover) and root structure (Kampf et al., 2016b; Larson-Nash et al., 2018; Robichaud et al., 2016) from burning. Greatly contributing to post-wildfire sedimentation rates, these larger-scale erosional processes are not represented on the small scale due to lack of streamlet connectivity. While hydrologic effects of hydrophobicity, as well as sedimentation from ash and burned soils, were reflected at the small-scale, additional hydrologic and erosional factors would need to be considered in upscaling efforts.

3.6 Conclusion

The experimental framework and design in this study sought to create observations of post-wildfire effects on water quality and supply, towards the goal of capturing key mechanisms associated with wildfire, rainfall, and terrain slope drivers. The simulation apparati controlled driver intensities, allowing for repeatable and quantifiable treatments. Burn and rainfall intensity ranges were prescribed to values similar to natural settings. The experiment was further validated by the simulated responses, which were generally similar to previous studies (Bladon et al., 2014; Hohner et al., 2019; Moody et al., 2013). Runoff and SSC increased monotonically with increasing burn intensity, while DOM and TDN peaked at moderate burn severities with an inverted 'U' shape, likely due to volatilization at higher temperatures. However, experimental limitations were also apparent through anomalous responses, introducing uncertainty into results.

This study additionally sought to contribute knowledge to wildfire and rainfall simulation experiment design considerations. In addition to providing detailed descriptions of simulation apparati, strengths and weaknesses of the scales, dimensions, and techniques used were also discussed. Future researchers may be able to leverage these findings to optimize and alter design components for their own simulation studies. These contributions to the field of wildfire effects on water quality and supply can ultimately help inform water managers in the preparation and mitigation of wildfire effects, which will become increasingly important in the coming years.

Chapter 4

Data-driven Modeling of Post-Wildfire Water Quality Impacts

The following chapter is currently in preparation and will be submitted to the International Journal of Wildland Fire in Summer 2023:

> Brucker, C.P., Livneh, B., Rosario-Ortiz, F., 2023. Increased post-wildfire water quality response across U.S. West watersheds. Environmental Research Letters *(in prep).*

4.1 Overview

Post-wildfire water quality effects can result in severe implications for human and freshwater systems, creating hazards for water treatment plants (WTPs) (Bladon et al., 2014). Basin-scale predictive models can be a powerful tool in informing mitigation and prevention efforts. However, water quality data with high spatial and temporal resolution, as well as a continuous sampling period over a wildfire event, is sparce—limiting modeling efforts and significance of analyses (Yu et al., 2019). Though several modeling techniques have been developed for post-wildfire sediment response and erosional effects (Blake et al., 2020; Kampf et al., 2020; Langhans et al., 2016; Rengers et al., 2016; Surfleet et al., 2014; Zema et al., 2020), models of nutrient and dissolved constituent concentrations have been

especially limited by data scarcity (Basso et al., 2022). Most existing models are either physical or process-based (Kampf et al., 2020), requiring numerous geophysical and climate parameters with high spatial and temporal resolution, as well as large computational resources. Though lacking the complex relationships simulated in physically-based models, data-driven modeling approaches typically have a greater capacity for handling low quality data (Yu et al., 2019). Use of geostatistical models and machine learning approaches in post-wildfire water quality modeling has been largely unexplored, however, with only a few previous studies utilizing this analysis type (Beyene et al., 2023; Gannon et al., 2022; Jain et al., 2020; Pennino et al., 2022; Rhea et al., 2022; Rust et al., 2019).

In this chapter, I present a statistical and machine learning analytical framework which optimizes available empirical water quality data in characterizing post-wildfire dissolved organic matter (DOM), nutrient, and sediment responses in 241 forested catchments across the U.S. West. Wildfire-driven responses were assessed using data from all basins in aggregate, as changes in water quality are often insignificant in individual basins due to data scarcity (Rust et al., 2018). Broad changes in water quality in each post-fire year were first evaluated using a regression-based approach, then inter-site variability in responses characterized by watershed physiography and wildfire characteristics using correlation and Random Forest modeling techniques. To optimize data sample sizes for analyses, an initial data-mining process identified forested basins with data which met availability and quality criteria spanning the entire U.S. West. Contributing areas to each water quality monitoring station were defined using a watershed delineation process, then basins were filtered for burn impacts, land cover type, and data availability. Post-fire response was then assessed in each of the selected basins by building models using pre-fire water quality data and hydroclimatic predictor variables, then creating predictions of post-

fire years and calculating residuals. Post-fire responses, or residuals, were aggregated across basins and compared to similar residuals calculated in 258 paired, unburned basins across the same time periods to assess significance. Finally, variability in the magnitude of post-fire responses was attributed to physical basin and wildfire characteristics through a correlation analysis, as well as by determining variable importance in random forest models trained on post-fire residuals. By providing information on both the magnitude of water quality responses over time, as well as the influence of basin and wildfire features, this analysis may help inform water managers in assessing watersheds' vulnerability to post-wildfire water quality effects.

4.2 Background

Water treatment plants (WTPs) rely on high quality water from forested catchments in the U.S. West, making them vulnerable to disturbances in their source water collection areas—including wildfires (Becker et al., 2018b; Writer et al., 2014). After nearby wildfire events, WTPs often experience increased turbidity, sediment and suspended solids, as well as increases in and changes to the character of natural organic matter. As these water quality parameters are key factors in WTPs' capacity requirements and process designs, significant changes can lead to operational issues or changes in treatment methods (Becker et al., 2018b). In the short term, process changes are often implemented to combat poor post-wildfire water quality. Costly alternate disinfectants or precursor removal strategies are often implemented instead of chlorination, which can form carcinogenic disinfection byproducts with high levels of DOM (Hua and Reckhow, 2007; Wang et al., 2015b). In extreme cases, low quality water may be diverted all-together (Writer et al., 2014)—adding further stress to regional water resources. Additionally, though wildfire effects are typically

most severe in the weeks and months after a wildfire event, an increasing number of studies have shown longer-term implications for WTPs spanning months and years (Smith et al., 2011; Yu et al., 2019). These longer-term effects may require costly changes to WTP infrastructure and processes, as well as increased planning and adaptability (Becker et al., 2018b). The recent increased occurrence and severity of wildfires in the U.S. West (Abatzoglou and Williams, 2016; Marlon et al., 2009), coupled with the threat of decades-long effects, have highlighted the need for advancements in post-wildfire water quality impact assessments.

Basin-scale models and analyses of post-wildfire water quality response are important tools in mitigation efforts (Basso et al., 2022; Nunes et al., 2022), however data scarcity issues in post-wildfire water quality limit these analyses. While numerous previous studies have analyzed wildfire effects on water quality and supply (Abraham et al., 2017; Smith et al., 2011), most are case studies which analyze pre- and post-wildfire *in situ* data on a small regional scale (Murphy et al., 2015; Rhoades et al., 2019b; Smith et al., 2011; Uzun et al., 2020; Wang et al., 2015b; Writer et al., 2014). Modeling efforts of wildfire effects are lacking in comparison, due to challenges surrounding lack of pre-wildfire data with adequately long periods of record, as well as high spatial and temporal resolution (Yu et al., 2019). Several previous modeling efforts have characterized post-wildfire responses in runoff, debris flow, and sediment (Blake et al., 2020; Cannon et al., 2010; Culler et al., 2023; Gannon et al., 2022; Kampf et al., 2020; Langhans et al., 2016; Rengers et al., 2016; Surfleet et al., 2014; Williams et al., 2022; Zema et al., 2020). However, the majority of these efforts have been focused on physical- or process-based modeling requiring numerous geophysical variables with high spatial and temporal resolution as parameters, as well as large computational resources (Yu et al., 2019). As availability of post-fire chemical

constituent data such as DOM and nutrients are especially limited, few modeling efforts have examined these responses (Basso et al., 2022; Beyene et al., 2023; Bladon et al., 2008; Pennino et al., 2022; Rhea et al., 2022).

Machine learning and geostatistical analyses are alternative modeling techniques which typically have lower data and computational requirements than physical models (Yu et al., 2019). Only several studies have explored their use in post-wildfire water quality applications, however (Beyene et al., 2023; Gannon et al., 2022; Jain et al., 2020; Pennino et al., 2022; Rhea et al., 2022; Yu et al., 2019)—an identified gap in wildfire water quality predictive efforts (Mishra et al., 2021). Beyene et al., 2023 and Yu et al., 2019 both used regression-based analyses to assess significant changes in trace elements and nutrients, respectively, across wildfire events. Pennino et al., 2022 used bootstrapping techniques to assess significant wildfire impacts on nitrate, volatile organic carbon, arsenic, and disinfection byproducts. Similarly, Rust et al., 2018 used non-parametric, paired significance tests and change-point analyses to detect significant post-fire changes in 89 physical and chemical water quality parameters. Studies also used statistical and machine learning techniques to characterize inter-basin variability in responses, leading to conclusions about the influence of basin and wildfire characteristics in post-wildfire water quality response. Rust et al., 2019 assessed relationships between basin and fire characteristics and post-wildfire water quality response by calculating the correlations between every driver-response combination. Rhea et al., 2022, on the other hand, used linear mixed model selection to identify the influence of topographic and vegetation basin characteristics on post-wildfire nitrate and sodium concentrations, while Beyene et al., 2023 used Random Forest models to assess driver strengths.

This project will advance efforts in empirical assessments of post-wildfire water quality response by focusing on key contaminants across a broad regional scope—using a novel machine learning and statistical analysis framework. Here, post-fire responses in DOM, nutrients, turbidity, and sediment—key contaminants in WTP process planning—were assessed across 241 forested catchments in the U.S. West. Similar to study areas in Beyene et al., 2023, Pennino et al., 2022, and Rust et al., 2018, this scope was large enough to maximize data availability for modeling efforts, while maintaining consistency in vegetation, climate, and topographic regimes. However, while previous studies tended to borrow basin definitions and characteristics from existing datasets, e.g., the Geospatial Attributes of Gages for Evaluating Streamflow (GAGES II) dataset (Beyene et al., 2023; Rust et al., 2019), the custom-defined basin delineations here enhanced data availability and information at each model site. The use of paired, unburned basins additionally allowed for characterization of constituents' natural variability in comparison to responses in wildfire-affected basins, validating responses' significance using methods similar to Salavati et al., 2016 and Williams et al., 2022. Finally, with the exception of Rust et al., 2018 and 2019, this project was the first to analyze post-fire DOM and nutrient response across the entire U.S. West, and employed more advanced analysis techniques which built upon the original framework created by Rust et al. By levering analyses of post-fire data in aggregate across numerous basins, this study seeks to provide significant insights into the responses of sparce post-wildfire water quality data.

4.3 Methods

For this analysis, basins delineated from water quality monitoring stations across 11 states in the U.S. West were screened for analyses based on temporal resolution and data

quality, creating a subset of both wildfire-affected and paired unburned basins (*Section 4.3.1*). With these basin subsets, significance in constituent response was assessed by controlling for local factors through linear models built with climate and streamflow data for each basin, trained on pre-fire data, then analyzing the residuals calculated from post-fire predictions (*Section 4.3.2*). These results were compared to both pre-fire residuals generated from the same basin, as well as nearby paired unburned basins which had models trained on equivalent pre- and post-fire years. Finally, inter-basin variability in water quality responses were attributed to basin and fire characteristics by assessing their correlation to burned-basin model residuals, as well as using a Random Forest model to assess variable importance (*Section 4.3.3*). As discussed in *Section 4.3.4*, the watershed delineation process used pour points at the coordinates of monitoring stations with relevant water quality measurements to maximize data availability and usability. Finally, the preparation of initial datasets for analysis through data mining of *in situ* pre- and post-wildfire water quality characteristics is discussed in *Section 4.3.5*.

4.3.1 Site selection

Burned and unburned basin subsets were created and filtered for analysis by first delineating watersheds from a set of 57,979 water quality stations with appropriate data, then screening the resulting 48,577 successfully-delineated basins (shown in Figure 4.1a) by burn extent and data availability, as well as land cover data requirements. The water quality monitoring station selection process is described further in Section 4.3.5 and the basin delineation process is discussed in Section 4.3.4. The set of custom-delineated basins were first categorized for burn effects by assessing intersections between basins and wildfire polygons. The overlap of burn scar area with each basin was calculated as a percent

of the total basin extent, with wildfires occurring within the same water year merged together as an individual fire. “Burned” basins were designated as watersheds with greater than 5% burn area from an individual wildfire and “unburned” basins were required to have either no intersecting wildfires or less than 0.5% burn area from any wildfire—thresholds where burn impacts were, respectively, noticeable or marginal (Beyene et al., 2023; Rust et al., 2018; Williams et al., 2022). In basins where multiple wildfire events had occurred, events which occurred less than six years after a different event were discarded.

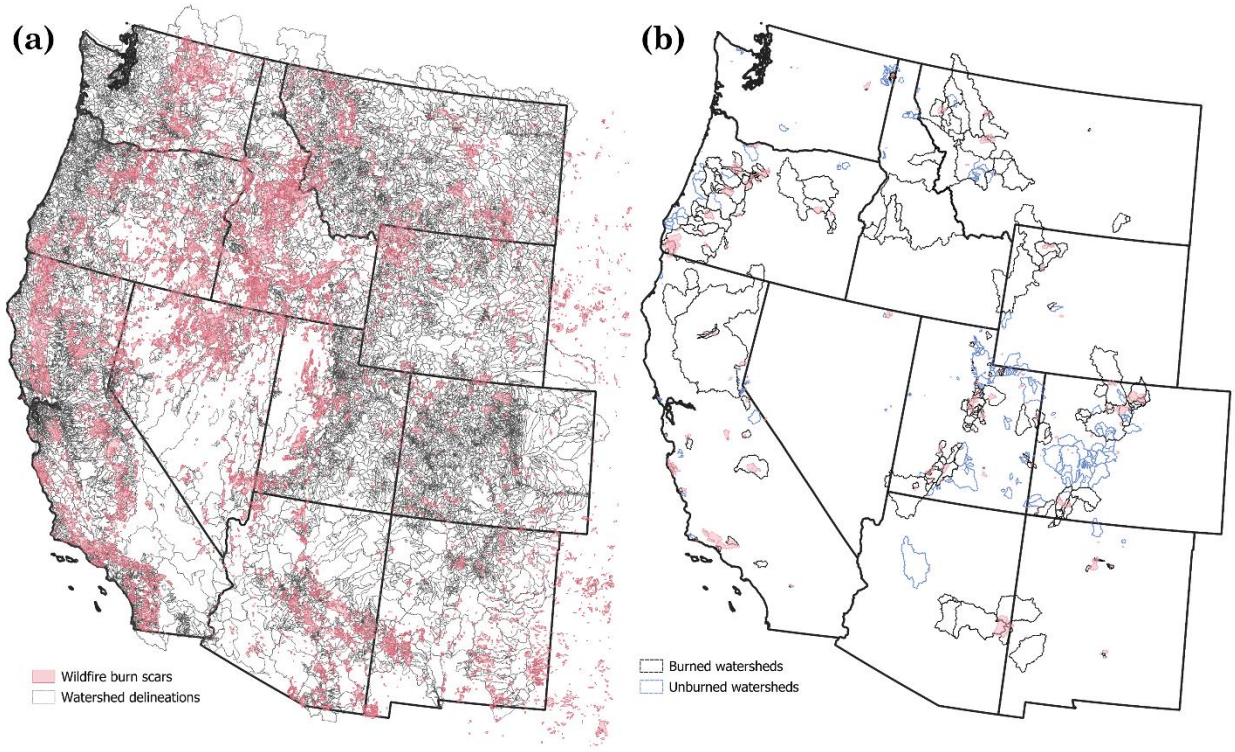

Figure 4.1: (a) Initial 48,577 watershed delineations (outlined in gray) created from coordinates of water quality monitoring stations. (b) 241 burned basins (outlined in black) and 258 paired unburned basins (outlined in blue) selected as modeling sites. Wildfire burn scar outlines are overlaid in pink.

Burned basins were then filtered for data availability, requiring at least 20 data points pre-fire and at least 10 data points post-fire, as well as a period of record spanning more than 3 years before and after the date of wildfire ignition. Although these sample

sizes might be considered low for modeling individual basins, similar data criteria produced significant results when analyzing water quality data in aggregate in basins across the U.S. West in previous studies (Beyene et al., 2023; Rust et al., 2018). Paired unburned basins were filtered to those with outlets < 50 km away from at least one burned basin outlet, as well as similar periods of record and data frequency as their paired burned basins. Paired basins' proximity increased consistency in vegetation and climate regimes to their paired burned basins (Salavati et al., 2016). As shown in Table 4.1, basin attributes were similar between burned and unburned subsets. Both burned and unburned basins were additionally filtered for land cover criteria, characterized using the NLCD data. Basins with greater than 25% forest coverage and less than 5% developed area (i.e., cities and residential areas) were selected to further maximize consistency in basins' geophysical characteristics—similar to thresholds used in Williams et al., 2022 and Beyene et al., 2023, respectively. Though the 25% forested area threshold is relatively low for an analysis of "forested" basins, increasing this threshold to, for example, 50% decreased the number of candidate basins by about a third. Thus, the lower threshold was kept for increased data availability. In total, 241 and 258 burned and paired unburned basins met the above filtering criteria and were selected as modeling sites, as shown in Figure 1b.

Table 4.1: Summary of basin characteristics for burned and unburned basin subsets. Number of basins, basin sizes, percent burn extent, wildfire distance from basins' outlets, percent forested extent, and percent developed extent are displayed. Median values are shown, with min-max ranges displayed in parentheses.

Basin subset	Number of basins	Size (km^2)	Burn extent (%)	Fire-outlet dist. (km)	Forested extent (%)	Developed extent (%)
Burned	241	490.7 (5.1-63,407.8)	9.8 (5.0-100.0)	10.4 (0.0-112.1)	53.0 (25.0-89.5)	1.1 (0.0-4.9)
Unburned	258	139.4 (5.0-20,496.3)	N/A	N/A	64.5 (25.1-97.6)	1.3 (0.0-5.0)

4.3.2 Wildfire impact on water quality constituents

Broad changes in water quality constituents were assessed for each post-fire year using a regression-based modeling approach between hydroclimatic and fire variables, following similar methods as Beyene et al., 2023 and Williams et al., 2022. This analysis sought to control the influence of climate variables on water quality concentrations, allowing for subsequent isolation of changes driven by wildfire activity. Precipitation, aridity characterizations, and streamflow have been shown to be highly influential on water quality variables (Murphy et al., 2015; Rhea et al., 2022; Rust et al., 2019), thus total daily precipitation, potential evapotranspiration, maximum temperature, estimated runoff, and remote *in situ* runoff in unburned basins were chosen as potential predictors, or covariates, for model building. Natural variability in water quality was characterized by responses in pre-fire data, as well as the paired unburned basins. Use of paired basins is common in assessments of watershed disturbances on water quality and runoff response (Bladon et al., 2008; Salavati et al., 2016; Williams et al., 2022; Yu et al., 2019) , allowing for characterizations of response significance and magnitude outside of what would be expected solely from natural variability in undisturbed basins (Salavati et al., 2016; Williams et al., 2022; Yu et al., 2019). Models were built for each burned and unburned site using climate variables and trained on pre-fire data, then applied to post-fire years—analyzing residual distributions between pre-fire and post-fire data. “Pre-fire” and “post-fire” data for each parallel unburned basin were designated in relation to the date of the fire occurrence in its paired, burned basin.

Linear modeling

Linear models for each watershed site were built using a step-wise approach and trained on pre-fire data, as illustrated in Figure 2. Step-wise model building prevents overfitting and is commonly used in applications with high numbers of potential predictors and low sample sizes (Beyene et al., 2023; Williams et al., 2022). First, the covariate with the strongest (maximum absolute) Pearson's correlation with the response variable (pre-fire water quality constituent) was chosen and used to condition a single-variable linear model. Model skill and complexity were assessed with the Akaike information criterion with a bias correction for small sample sizes (AICc) (HURVICH and TSAI, 1989). Akaike information criterion scores evaluate model fit while adding penalties for complexity. This bias corrected information criterion is recommended for model building applications where the ratio of sample datapoints to covariates is less than 40 (Burnham and Anderson, 1998), which was rarely exceeded here due to small numbers of available water quality data and high numbers of candidate covariates tested (49). The model was then used to create predictions of the pre-fire training dataset, with residuals calculated from the difference between observed and estimated values. Residuals were compared to each of the remaining predictor variables and the predictor with the strongest correlation added to the model. AICc was then calculated for the new model and the additional covariate retained if it lowered the AICc value by more than 2. This process was repeated until the addition of a new covariate did not satisfy the delta AICc requirement.

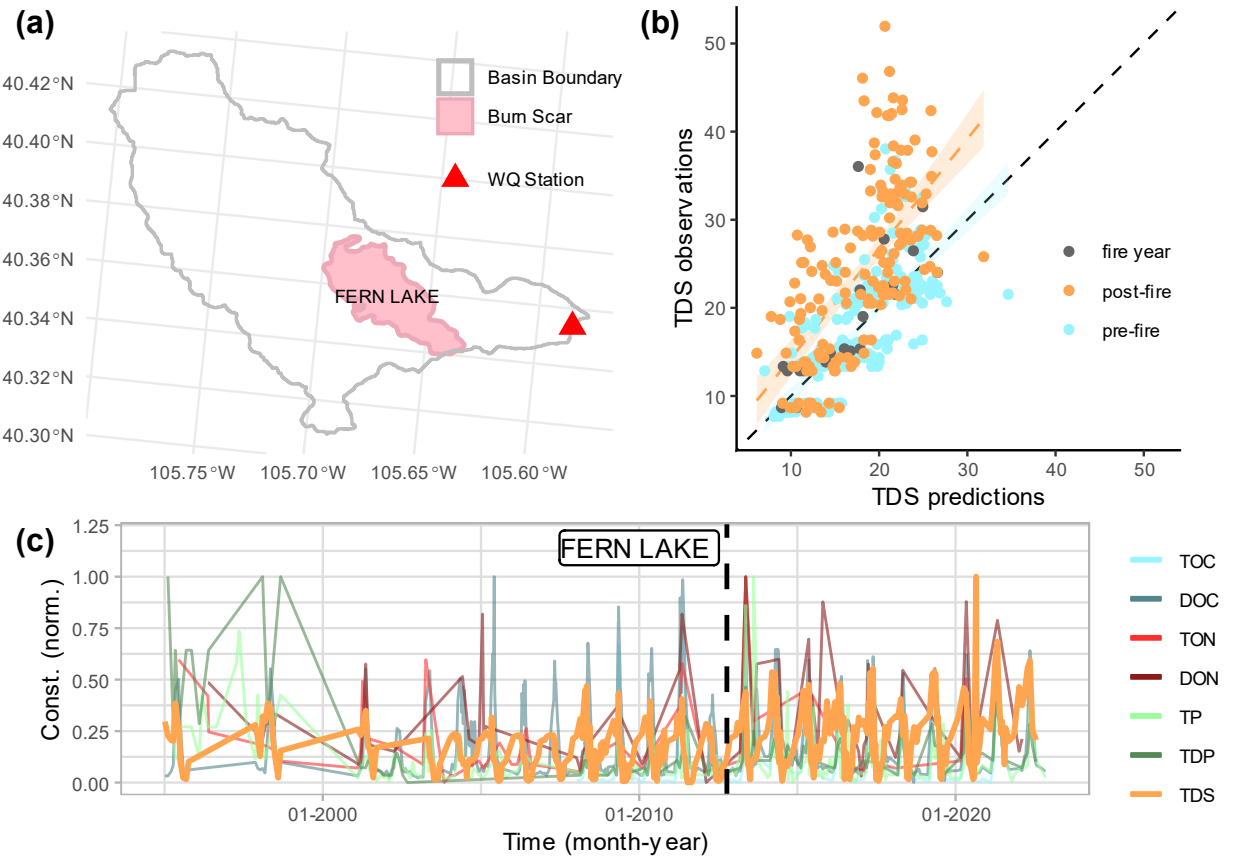

Figure 4.2: (a) Water quality monitoring station and associated basin definition, as well as a burn scar from a wildfire (the 2012 Fern Lake fire in Colorado as an example) affecting the basin. (b) Pre- and post-fire observations and predictions created from a linear model trained on pre-fire data. Inner-quartile ranges or best fit linear models for pre-fire and post-fire data are shown as transparent blue and orange ribbons, respectively. (c) Water quality data which spanned the wildfire occurrence (vertical dashed line) in the basin.

As shown in Table 4.2, 45 total covariates were prepared for building each model. Before the model building process, however, a correlation analysis was completed to remove highly correlated values and reduce the effects of collinearity. Daily total precipitation, potential evapotranspiration, max temperature, and surface runoff were averaged across each basin for the entire period of record, 1974-2022. Additionally, where available, median daily streamflow values were calculated from nearby unburned basins with stream gages < 50 km away from each basin's outlet. This variable represented median regional and seasonal streamflow's role in driving constituent responses. For each of these variables,

average moving windows were calculated to represent the effects of longer-term shifts in climate variables on constituent response (Pennino et al., 2022). 7-, 30-, and 90-day average moving windows were chosen to represent the effects of rapid changes in climate variables, as well as seasonal changes. Additional transformed covariates were calculated for each variable-moving window size combination, as variable distributions were mostly non-normal (assessed using density distributions and Q-Q plots). The square was taken for all climate variables and the log taken of precipitation and surface runoff variables. Before being input into the model building process, covariates with a > 0.9 absolute maximum Pearson's correlation with another variable were removed, keeping the variable with slightly higher correlation to the predictand. Sources of covariate data and preprocessing steps are discussed further in *Section 4.3.5*.

Table 4.2: Table listing all candidate covariates used in the model building process for each basin and constituent. Additional moving average window, log, and square transformations tested are shown.

Type	Variable	Transformations
Climate	Total precipitation	7-, 30-, and 90-day average moving windows; log and square
	Total potential evaporation	7-, 30-, and 90-day average moving windows
	Peak temperature	7-, 30-, and 90-day average moving windows
Hydrologic	Surface runoff	7-, 30-, and 90-day average moving windows; log and square
	Median remote streamflow	7-, 30-, and 90-day average moving windows; log and square
Seasonal	Day in water year	NA

The final models built using the step-wise AICc method were additionally evaluated for performance using a leave-one-out cross-validation (LOOCV) method applied to pre-fire data for each model—similar to methods used by Beyene et al., 2023 and McManus et al., 2020. LOOCV involves first assigning one day of covariate and predictand variables as

testing data, calibrating a model with data from the remaining days, then using that model to predict the response on the testing day. This process is repeated for each available day, then model performance metrics calculated from the observed and predicted data. Metrics commonly used in water quality modeling were selected to evaluate each model (Moriasi et al., 2007): the Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970), percent bias (PBIAS), and the ratio of the root mean squared error to standard deviation (RSR). These metrics describe, respectively, the model error relative to the total variation, the tendency of the model to overpredict or underpredict, and the goodness-of-fit of the model (Beyene et al., 2023; Cho and Lee, 2018). Models were considered to have satisfactory performance with reliable predictions if NSE exceeded 0.5, PBIAS was less than 30%, and RSR was less than 0.7, based off of guidelines by Moriasi et al., 2007.

Models built for each basin-constituent combination were then used to create predictions over the entire available period—both pre-fire and post-fire. Residuals, or offsets, were calculated for each pre- and post-fire water quality data point as the difference between predicted and observed values. Offsets from post-fire years were interpreted to represent constituent change after fire events, whereas offsets from pre-fire years were used to quantify natural variability in constituent responses. Median offsets were calculated for each pre- and post-fire year and fire-related change was considered significant ($p < 0.05$) if greater than two standard errors (RMSE).

4.3.3 Inter-Basin Variability

Two methods were used to attribute different post-fire constituent responses across basins to geophysical watershed and wildfire characteristics: A correlation analysis and covariate importance assessment using Random Forest models. Analyzed watershed and

wildfire characteristics included percent forested and developed areas in each basin, as well as percent burn extent and the distance of wildfire burn scars from basins' outlets—factors shown to be highly influential on water quality response in previous studies (Beyene et al., 2023; Rhea et al., 2022; Rust et al., 2019; Williams et al., 2022). Sources and methods to calculate these factors for each fire-basin combination are discussed further in Section 4.3.5.

Correlation Analysis

Similar to methods used in Williams et al., 2022, post-fire model residuals were first plotted against each watershed and fire characteristic to visually assess their linear relationships. A best fit linear model was applied to each combination, with an R^2 calculated to assess the strength of relationships. Correlations were then assessed by calculating the Pearson's correlation between every response and factor combination.

Random Forest modeling

Constituent residuals and watershed and fire characteristics were then input into Random Forest models, assessing the importance of covariates. Random Forest is a supervised machine learning algorithm commonly used in regression analyses which uses the output of multiple decision trees to determine a single result (Breiman, 2001). This algorithm is capable of handling numerous, correlated covariates with limited predictand sample sizes, thus all potential physical covariates were input into the model. Using the resulting model, each covariate's importance, or the loss of model skill with its removal, was calculated to determine the most influential characteristics in each constituent's post-fire response. The *randomForest* package in R (Cutler and Wiener, 2022) was used to build and evaluate this model.

4.3.4 Hydrologic conditioning and basin delineation

Watersheds were delineated using the coordinates of each of the 57,979 water quality monitoring stations as a pourpoint, or outlet of a contributing drainage area—resulting in a total subset of 48,577 successfully delineated watersheds. As shown in Figure 3, DEM segments were first prepared through hydrologic conditioning, then flow information extracted, before finally moving pourpoints to be coincident with streams and executing watershed delineation algorithms. These processing steps were completed using the *whitebox* package in R (Wu and Brown, 2023). Hydrologic conditioning algorithms prepare DEMs for hydrologic analyses by smoothing out depressions and correcting disrupted flow paths (Lindsay et al., 2008). Here, two subsequent processes were applied to the HUC 4 DEM segments: stream burning and filling. Stream burning is a flow enforcement technique which corrects surface drainage patterns in DEMs by lowering grid cells underneath overlaid stream vector shapefiles (Lindsay, 2016). 1:100,000 resolution flowline vectors from the NHD were trimmed to the extent of each of the 77 HUC 4 segments, overlaid on the associated DEM subdivision, then the flow paths burned in. This was executed using the Whitebox Tools *FillBurn* function, interfaced in R using the *whitebox* package. Next, depressions in the DEMs' topography which were disruptive to flow paths were identified and "filled", i.e., grid cell elevations raised, using a method described in Wang and Liu, 2006. Implemented with the *FillDepressionsWangAndLiu* Whitebox tool, this algorithm was chosen due to its computational efficiency, as it simultaneously determines flow paths and spatial partitions of watersheds with one pass of processing (Wang and Liu, 2006).

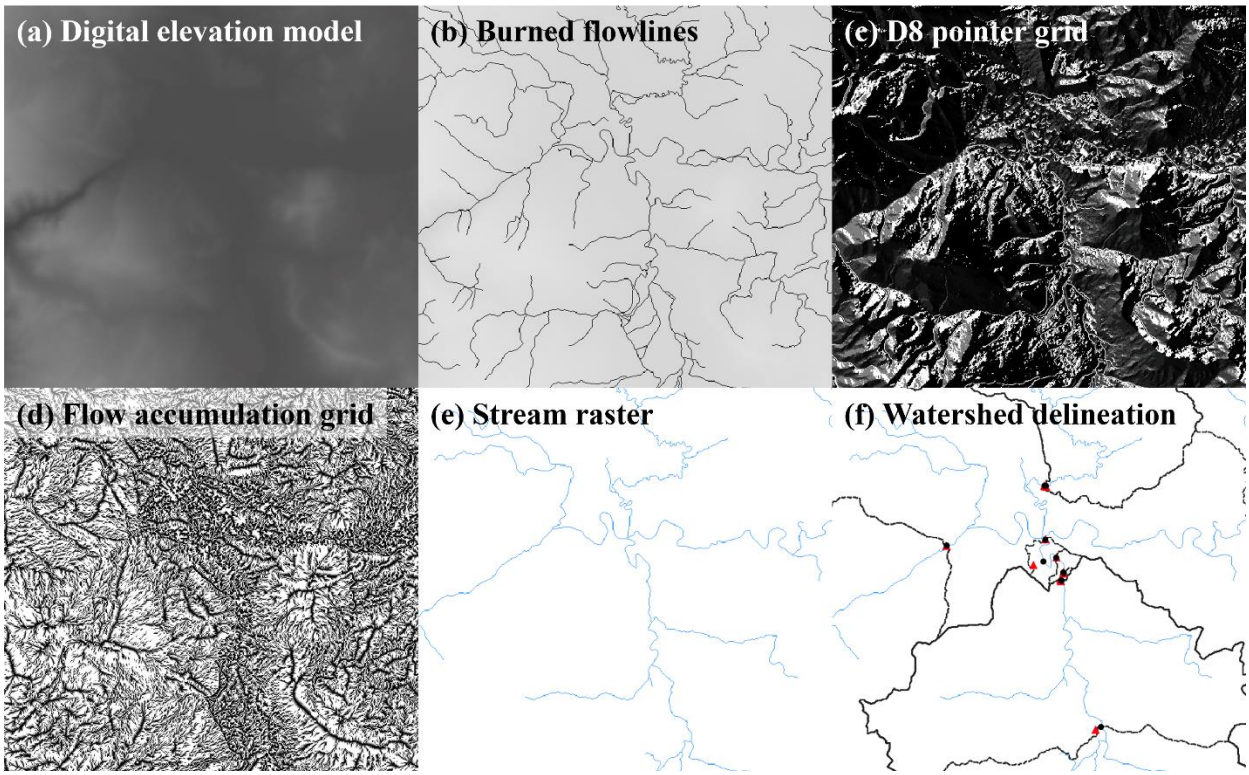

Figure 4.3: Example steps involved in delineating watersheds from pourpoints. (a) A digital elevation model was tiled for the U.S. West and segmented to HUC 4 subdivisions. (b) NHD flowline streams were burned into the DEM segments, or the elevations of raster grid cells coincident with streamlines lowered. (c) A D8 pointer, or flow direction, grid was extracted, where one of eight possible flow directions was assigned to each grid cell based on the slope and aspect of surrounding cells. (d) Flow accumulation was calculated for each grid cell as the total upstream or contributing area. (e) Stream grid cells were designated where flow accumulation exceeded a certain threshold (5 km^2). (f) Water quality station pourpoints (red triangles) were snapped (black points), or moved to be coincident, with stream grid cells, then a watershed delineation algorithm applied which evaluated contributing areas to each pourpoint (black outlines).

Once DEM segments were conditioned, flow direction, flow accumulation, and stream raster grids were extracted to execute the basin delineation algorithm. Flow direction, or pointer, grids contain information about the direction of flow in each grid cell, calculated based on the slope and aspect of surrounding grid cells. A D8 algorithm, which assumes 8 possible flow directions from each grid cell (O'Callaghan and Mark, 1984), was implemented for each conditioned HUC 4 DEM segment using the *D8Pointer* tool. These pointer grids were then input into a flow accumulation algorithm—the

D8FlowAccumulation tool—which determines the contributing number of cells, or area of drainage, to each grid cell. Finally, stream network rasters were then extracted from the flow accumulation grids using the *ExtractStreams* tool, which designated stream grid cells at high flow accumulation locations. A threshold contributing area of 5 km^2 was determined by testing several values and assessing the resulting stream raster's similarity to the 1:100,000 resolution NHD flowlines.

Once stream rasters were created for each HUC 4 segment, pourpoints which fell within the extent of each segment were snapped or moved to be coincident with a stream raster grid cell. This step ensured limited truncation in delineations due to slight discrepancies in the pourpoint coordinates and the location of flow paths defined through the DEM conditioning process. The Jenson algorithm was used, which snaps pourpoints to the nearest stream grid cell (Jenson, 1991)—commonly preferred over less sophisticated algorithms which snap pourpoints to the largest flow accumulation grid cells within a certain radius (Lindsay et al., 2008). This was implemented using the *JensonSnapPourPoints* tool, with a 5 km snap radius threshold.

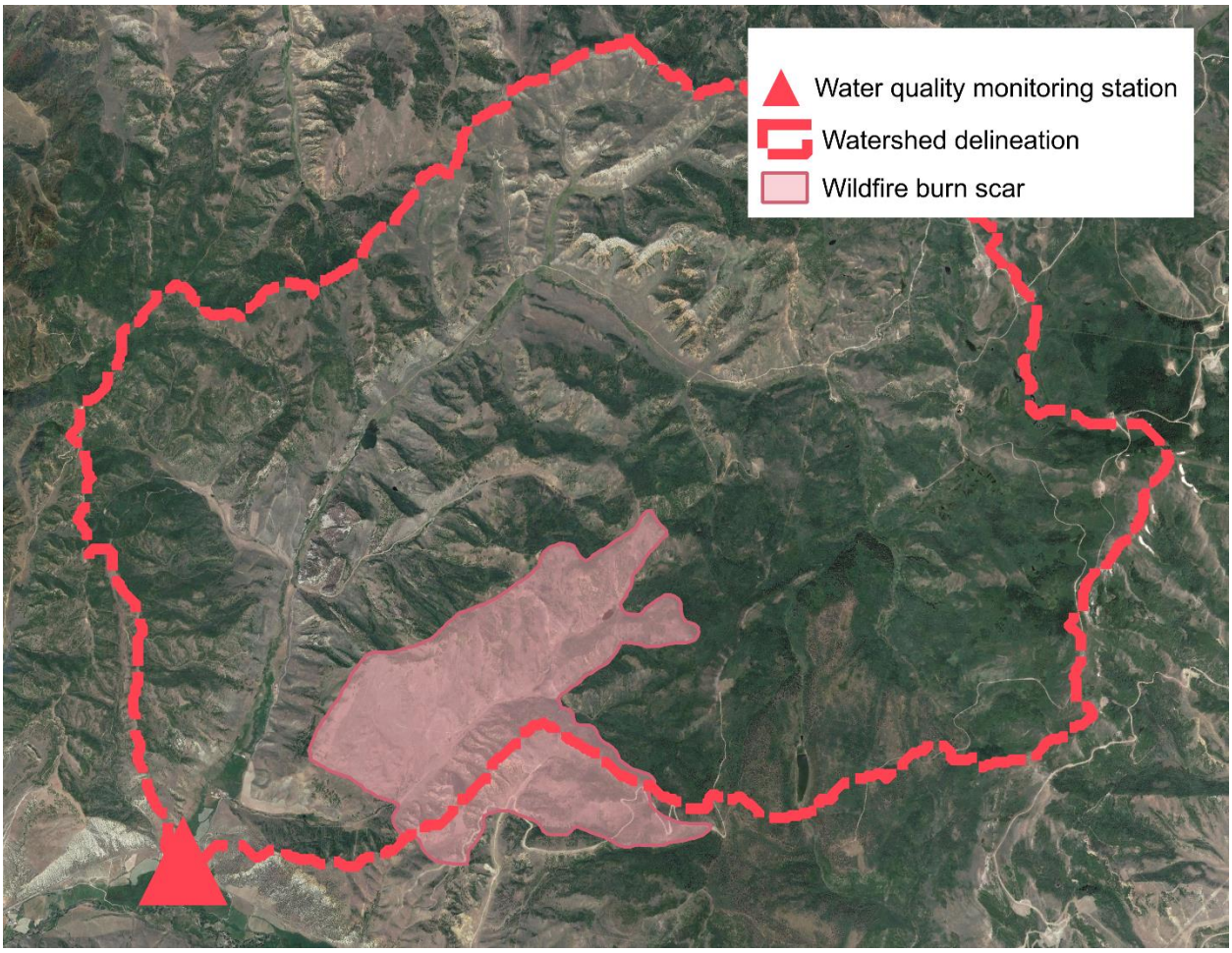

Figure 4.4: Water quality monitoring station pourpoint and resulting watershed delineation calculated by the contributing drainage area. A wildfire burn scar is shown in pink.

The final step in the basin delineation process was to execute a search algorithm from each snapped pourpoint which looked upslope and determined the contributing area, as seen in Figure 4. Water quality monitoring station coordinates which snapped to the same 30 m grid cell, thus resulting in the same delineated watershed, were assigned a unique basin ID, and their available data combined. In total, 48,577 basins were successfully delineated from 57,979 water quality monitoring station locations. Stations within the same 30 x 30 m grid cell were merged together to form a single pourpoint and their data combined. 285 water quality stations were screened from the analysis due to lack

of proximity to stream raster grid cells. This was likely due to slight inaccuracies or low resolution of recorded water quality monitoring station coordinates, or inaccuracies in the extracted stream raster to natural flowlines. USGS water quality monitoring stations which had equivalent basin delineations in the GAGES II dataset were overlaid and visually compared to available GAGES II delineations to validate the implemented basin delineation process.

4.3.5 Data sources and pre-processing

Empirical water quality and stream gage data, as well as remotely sensed geospatial data, climate variables, and wildfire information were compiled to assess their availability in 11 states across the U.S. West. The following databases were used:

- **Water quality observations:** The Water Quality Portal (WQP) created by the National Water Quality Monitoring Council (Read et al., 2017) provided empirical water quality data for sites across the U.S. West. The database includes publicly available data from the United States Geological Survey (USGS), the Environmental Protection Agency (EPA), and over 400 state, federal, tribal, and local agencies. Data available from this portal include descriptions of both biological and chemical water quality constituents and contaminants, as well as physical characteristics, for streams, lakes, reservoirs, and other water sources as varying temporal resolutions.
- **Wildfire data**: The U.S. Forest Service's Monitoring Trends in Burn Severity (MTBS) database (Finco et al., 2012) supplied data for wildfires in the U.S. West occurring from 1984-2022 which were greater in size than 1000 ac (404.7 ha). This satellite-derived data has a 30-m spatial resolution of both

burn intensity and burn perimeter delineations. Both categorical (low, moderate, and severe) and continuous (delta normalized burn ratio (dNBR)) burn intensity characterizations are available for individual wildfires as rasterized images, or "Thematic Burn Severity" and "NBR/dNBR/RdNBR Images" on the MTBS website, respectively. Polygon and point shapefile datasets are available for all wildfire perimeters and centroids, or the "National – Burned Area Boundaries Dataset" and the "National – Fire Occurrence Dataset", respectively.

- **Physiographic data**: The 1-Arc second Global Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) (EROS, 2017) is a near-global land elevation dataset acquired from radar data by the National Aeronautics and Space Administration (NASA) and the National Geospatial-Intelligence Agency (NGA). This dataset covers over 80% of the Earth's land surface between 60° north and 56° south latitude, with datapoints from every 1 arc-second (~30 m). The SRTM Void Filled elevation data were used in this analysis, which has additional processing to address areas of missing data, or voids where initial processing did not meet quality specifications. The National Land Cover Database (NLCD) (Dewitz, 2019) is a land cover database for the U.S. with 28 different land cover products characterizing land cover and land cover changes across 8 epochs from 2001-2019. The Watershed Boundary Dataset (WBD) (Laitta et al., 2004) is a comprehensive aggregated collection of hydrologic unit data, created by the USGS, U.S. Department of Agriculture – Natural Resource Conservation Service (USDA NRCS), and the EPA. It defines the extent of surface water drainage to a

certain point, delineating watershed boundaries of varying sizes. The National Hydrography Dataset (NHD) (USGS, 2019) represents the water drainage network of the U.S., defining rivers, streams, canals, lakes, ponds, dams, and stream gages. Mapped at a 1:24,000 or larger scale, these data are updated and maintained through Stewardship partnerships and other collaborative bodies.

- **Hydroclimate data**: The European Environment Agency ERA5-Land (Sabater, 2019) climate data is a reanalysis dataset which provides information on changing land variables over several decades. At a 0.1° x 0.1° spatial resolution and hourly timestep, this dataset combines model data with observations across the world to form a globally complete and consistent dataset using the laws of physics. The temperature of air at 2 m above the surface of the Earth, potential evaporation calculated through the surface energy balance, surface runoff calculated as the total amount of water accumulated during the forecast duration, and total precipitation, or the sum of accumulated liquid and frozen water, were used in this analysis. The Geospatial Attribute of Gages for Evaluating Streamflow (GAGES II) database (Falcone, 2011) provide geospatial data and classifications for 9,322 stream gages maintained by the U.S. Geological Survey (USGS). Information exists for gages which have either 20+ complete years of discharge record since 1950 or are currently active. Geospatial data includes watershed definitions for the contributing area to each gage, as well as several hundred watershed characteristics compiled from national data sources.

Accompanying historical and current streamgage records are available for each gage through the USGS website.

Data from the above databases were downloaded, filtered, and pre-processed to prepare them for overlaying and site identification processes, as well as for use as covariates in model building. Data were filtered to the temporal and spatial scope of this study, or from 1974-2022 in the U.S. West. Pre-processing steps were completed using R Statistical Software (R Core Team, 2022) and Python programming languages, as well as QGIS and the Climate Data Operators (CDO) command-line tool.

Water quality observations

Sediment, DOM, nutrient, and turbidity data from the WQP were filtered for the temporal and spatial scope of this study's analyses, minimized anthropogenic contaminants, and maximized consistency in site and data types. Data was first filtered for sites within 11 states in the U.S. West: Arizona, California, Colorado, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington and Wyoming. Date ranges from 01-01-1974 to 01-01-2023 were chosen to span the ignition dates of wildfires available through the MTBS database, with 10 years of buffer on the front end to construct and fit pre-wildfire models. Sample media was limited to "Water" and "water" (different terms were used for various agencies) and monitoring station types were limited to "streams" and "rivers" for increased consistency in site types.

As measurement types and descriptions of sediment, DOM, nutrients, and turbidity varied widely across agencies, a frequency analysis was completed to identify sample search queries for each constituent—selecting the ones most common across agencies. For example, when searching for water quality stations in Colorado, filtering for "Sediment"

resulted in only 94 available sites monitored by the USGS (the NWIS database). However, filtering for "Total Suspended Solids" resulted in 4,474 sites monitored by the USGS (the NWIS database) and the EPA (the STORET database). Search term frequencies for water quality data in the U.S. West are shown in Figure 5. Thus, "Total Organic Carbon" and "Dissolved Organic Carbon" were the chosen terms for carbon (abbreviated as TOC and DOC, respectively); "Total Nitrogen", "Total Organic Nitrogen", and "Dissolved Organic Nitrogen" were the chosen terms for nitrogen (abbreviated as TN, TON, and DON, respectively); "Total Phosphorus" and "Total Dissolved Phosphorous" were the chosen terms for phosphorus (abbreviated as TP and TDP, respectively); and "Suspended Sediment Concentration", "Suspended Sediment Discharge", "Total Suspended Solids", and "Total Dissolved Solids" were the chosen terms for sediment and dissolved organic matter (abbreviated as SSC, SSD, TSS, and TDS, respectively). Note, carbon characterizations of DOC and TOC were also selected due to being common analytical measurements associated with DOM. A "Turbidity" search term was additionally used for turbidity measurements, abbreviated as "TURB" in this manuscript. The identified water quality stations were then used as the pourpoints for basin delineations.

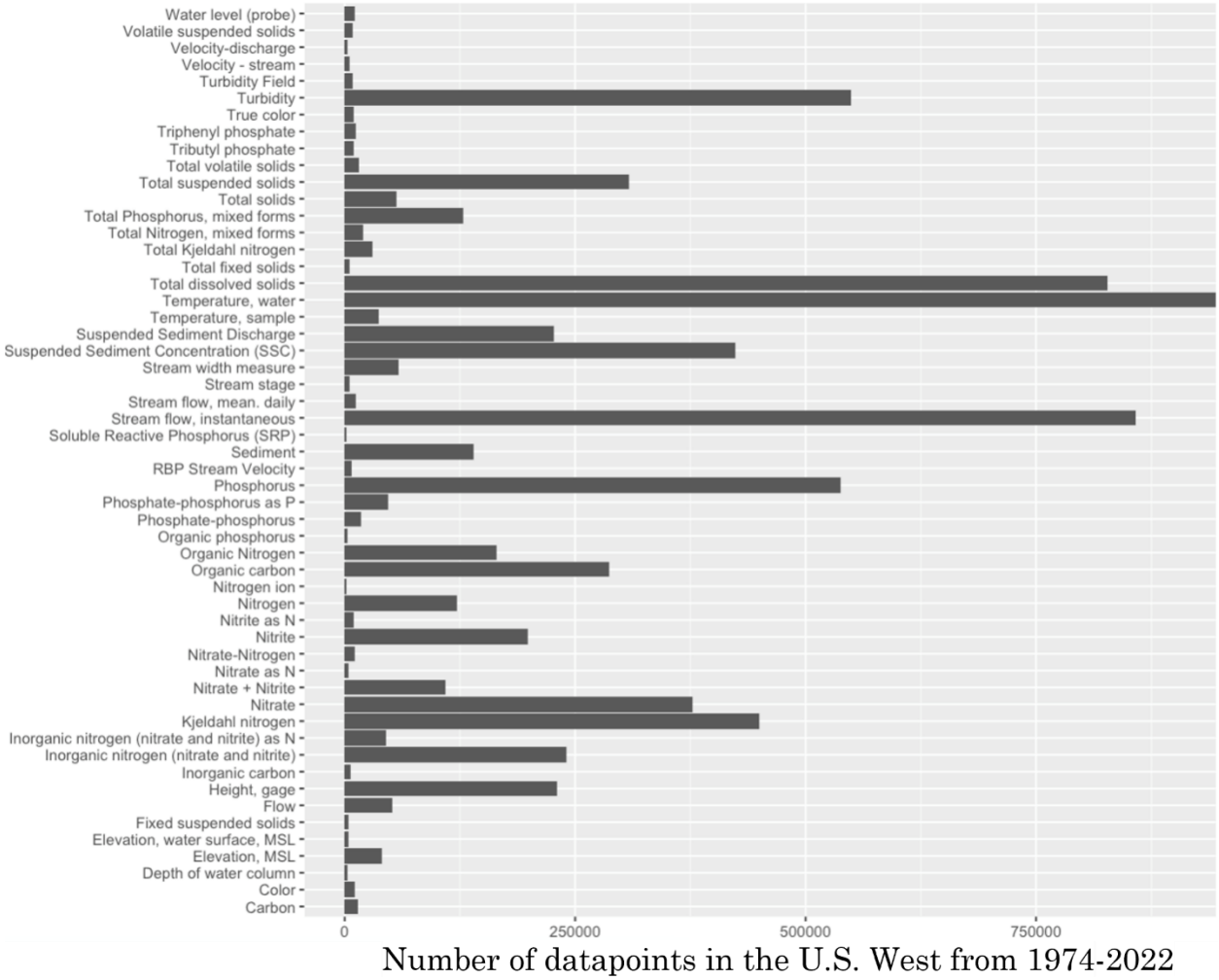

Figure 4.5: A bar chart showing the frequency of various search terms for the constituents in this analysis.

The collection time and date of each measurement was recorded on the granularity of one second, but the actual time interval between successive measurements was irregular, ranging from several minutes to years. Values were aggregated to an average daily timescale (total daily values for flux measurements, e.g., units of "kg") where applicable for increased consistency, but high enough temporal resolution to capture rapid changes in concentrations. In total, 57,979 water quality stations were available for analysis and ~2.5 million water quality datapoints.

Fire data

The national wildfire perimeter and centroid datasets were used for initial modeling site identification and filtering, then individual dNBR grid rasters compiled for fires associated with final modeling sites. The dNBR burn intensities averaged over each basin were evaluated as potential drivers during the inter-site variability analysis. These national datasets were cropped to those within 11 states in the U.S. West (Arizona, California, Colorado, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington and Wyoming), with 11,400 wildfires were available for analysis.

Physiographic data

1-degree x 1-degree SRTM DEM rasters were first tiled together for the extent of the U.S. West, or latitudes from 31°N to 49°N and longitudes from 125°W to 102°W. To create more manageable file sizes, the DEM was then cropped to 10 HUC 2 watershed extents, then those cropped to 77 HUC 4 watershed extents, using HUC delineation polygons from the WBD. These pre-existing watershed boundaries were used to ensure that DEM subdivisions would not cut short natural flow paths during the basin delineation process.

Hydroclimatic observations

Within the extent of the U.S. West, hourly ERA5-Land data was downloaded from 1974-2022 for total precipitation, potential evapotranspiration (PET), maximum 2 m temperature, and estimated surface runoff. This data was first aggregated to a daily timescale to match water quality and streamflow data temporal resolutions. Total daily values were calculated for precipitation, PET, and runoff, and max daily values calculated for temperatures. Then, moving window averages were calculated for each variable using varying window sizes (7, 30, and 90 days) to test as potential covariates in model building.

Finally, the squares of daily values and moving averages for calculated for each variable, as well as the logarithm of precipitation and runoff values.

Streamflow data from USGS GAGES II stream gages within the U.S. West were filtered for their proximity to chosen modeling sites (< 50 km from basin outlets), as well as data availability for the period of record of the associated water quality monitoring basins. Associated streamgage basins were further filtered by the same land cover criteria as the water quality basins, as well as categorized as burned or unburned. Data was aggregated to a daily timescale, calculating mean daily flowrates. Where data gaps of less than 10 consecutive days existed, an infilling processing using linear interpolation was applied to estimate missing values. In total, 450 gages and 3.6 million daily streamflow datapoints were available for analysis.

4.4 Results

Results are presented in three sections. First, data mining and basin delineation results, as well as an overview of water quality data are presented in Section 4.4.1, next broad constituent response in residuals across all basins from the regression analysis (Section 4.4.2), and Section 4.4.3 provides an attribution of responses to watershed physiographic and wildfire characteristic variables.

4.4.1 Data mining and assessment

To validate the custom delineation processes used in this study, the 48,577 watersheds were compared to geometries of GAGES II basins with similar outlet coordinates, where available. The two basin types were first overlaid, then their overlaps with each other calculated. Out of 1843 compared delineations, 1674 (~91%) had greater than 75% area overlapping for both basin types. Through a visual assessment in QGIS, the

9% erroneous delineations were mostly attributed to pourpoints snapped to incorrect streamlines. Of these, 33% had greater than 5% "developed" land cover and 54% had less than 25% "forested" land cover, indicating that a high number of erroneous delineations were in human-affected areas or relatively flat terrain. This is consistent with previous studies, which have shown high error rates in watershed delineation processes in urban areas and flat terrain (Datta et al., 2022). However, inaccuracies in the hydrologic conditioning processes may have also contributed to erroneous delineations. Generated flow direction and flow accumulation rasters for several HUC 4 basins were visually compared to similar grids available through the NHDplus geospatial database to validate conditioning methods used. While the flowlines used in this study were generally similar to NHDplus, especially for larger mainstems, anomalous areas with acute differences were apparent. As anomalies were most present in areas with smaller, more sparse stream networks, smaller basin delineations were likely most affected.

Of the 48,577 evaluated basins, just 241 (0.5%) met the filtering criteria required for adequate burn impacts and data availability for modeling. Thus, though ~2.5 million water quality datapoints were initially compiled from water quality monitoring stations, just 101,307 datapoints for burned basins were available for analyses. An additional 258 basins and 117,171 datapoints were designated as a paired, unburned dataset. As shown in Figure 6, turbidity, total dissolved solids, and total suspended solids had the greatest sample sizes, making up 21, 18, and 16% of the total available water quality data, respectively. Total nitrogen, total organic nitrogen, and dissolved organic nitrogen had the least availability, with the sum of these datapoints comprising less than 3% of the total available water quality data.

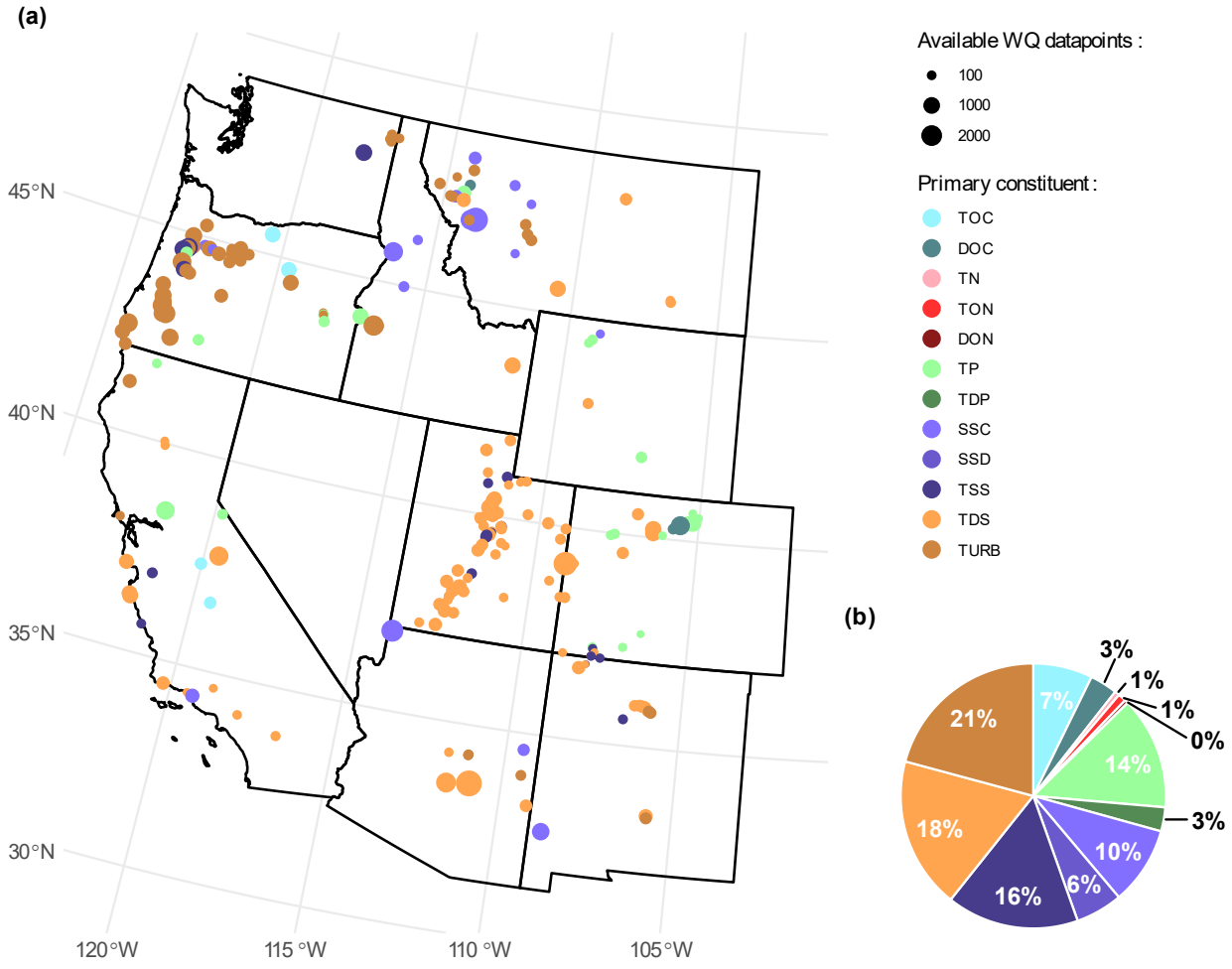

Figure 4.6: (a) Map of burned basins across the U.S. West and their availability of nutrient, DOM, and sediment water quality measurements from 1974-2022. (b) The proportion of each water quality variable to the total data available.

Raw water quality data from burned basins were visualized and assessed for changes across wildfire events. As shown in Figure 4.7, density plots were created for each constituent to assess their distributions. Constituent response distributions were positively skewed and non-normal, with all constituents failing Shapiro-Wilk normality tests ($p < 0.05$). SSC, SSD, and TSS constituents were especially greatly skewed in a positive direction, thus were log-transformed to visually assess their distributions. This is

consistent with previous studies which typically report non-normal distributions in water quality data (Bladon et al., 2008; Rhea et al., 2022).

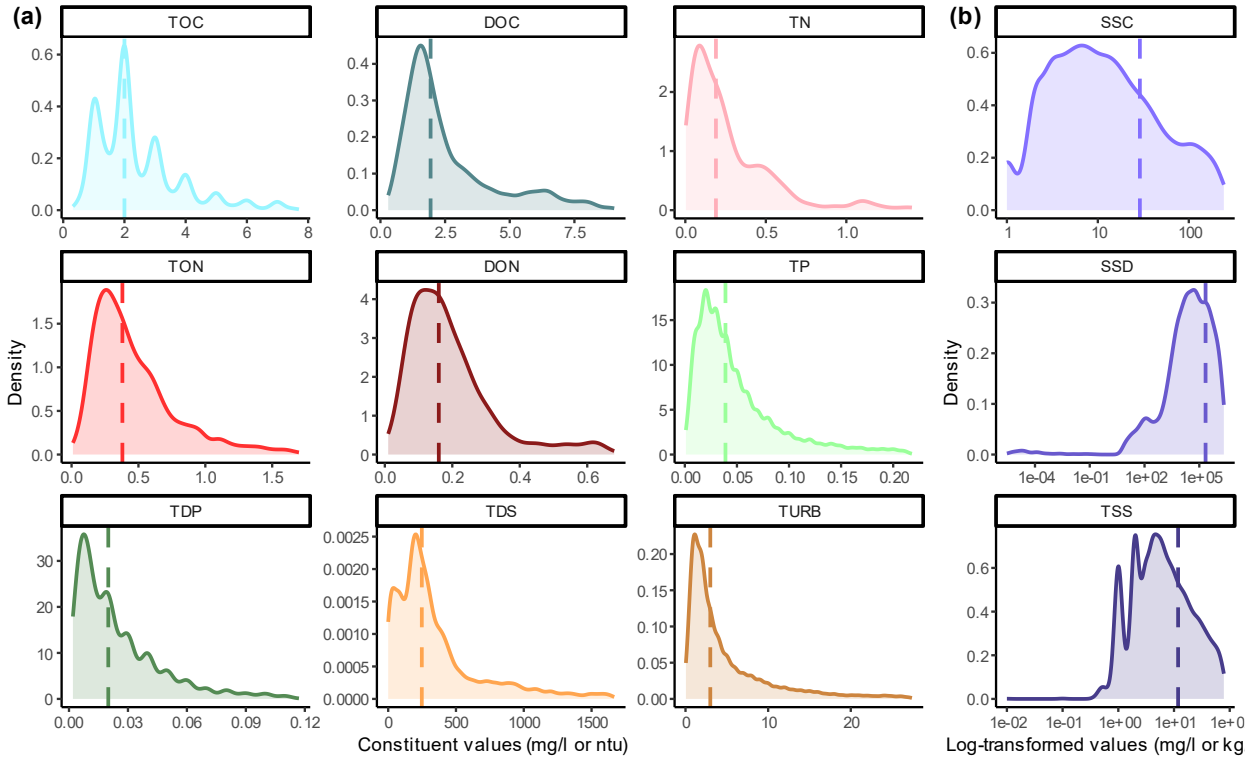

Figure 4.7: (a) Distribution plots all data from burned basins. (b) Similar, log-transformed distributions of SSC, SSD, and TSS. Population medians are designated by the dashed, vertical lines.

For each basin, mean and median change in constituent responses between pre-fire years and the first two years post-fire were calculated, with the significance of differences assessed using Mann-Whitney U tests. These values were then averaged across basins to assess overall changes in raw data. The percent of all basins which exhibited significant changes for each constituent was also calculated. As seen in Table 4.3, average differences in mean concentrations ranged from 21-121% for carbon constituents, -22-1081% for nitrogen constituents, 74-94% for phosphorus constituents, 14-155% for sediment constituents, and 14-1115% for TDS and TURB. The percent of basins exhibiting significant

responses ranged from 19% for TSS to 80% for TN, with ~20-30% significance rates typical for most of the other constituents.

Table 4.3: Median and mean constituent percent differences between pre-fire data and data within two years of a wildfire event—averaged across all burned basins. The percent of all basins exhibiting a significant change, tested using Mann-Whitney U-tests, is also displayed. Total number of datapoints is shown for all available data across burned basins for each constituent.

Constituent	Total number of datapoints	Average median change (%)	Average mean change (%)	Percent significant (%)
TOC	7345	42	121	39
DOC	3406	17	21	22
TN	588	110	90	80
TON	995	445	1081	25
DON	438	-21	-22	25
TP	13,948	131	94	32
TDP	2968	76	74	27
SSC	9655	259	155	27
SSD	5801	8797	14	22
TSS	16,355	47	83	19
TDS	18,732	2164	14	28
TURB	21,076	288	1115	26

4.4.2 Wildfire impact on water quality constituents

Linear models built for each basin-constituent combination ranged in number of and type of covariates used, number of training and testing datapoints, and model performance metrics. The number of covariates incorporated in the final models for each basin ranged from 1 to 9, though the median number of covariates used in models for each type of constituent was 1 for most constituents and only as high as 3 for SSD. Median numbers of training pre-fire data for each of the models ranged from 27 for TN to 99 for TDP, and median post-fire observations for each model ranged from 16 for DOC to 47 for SSC. Model performance metrics also had broad ranges, with median RSR scores ranging from 0.87 for TDS to 0.97 for TOC and TURB, as shown in Figure 4.8.

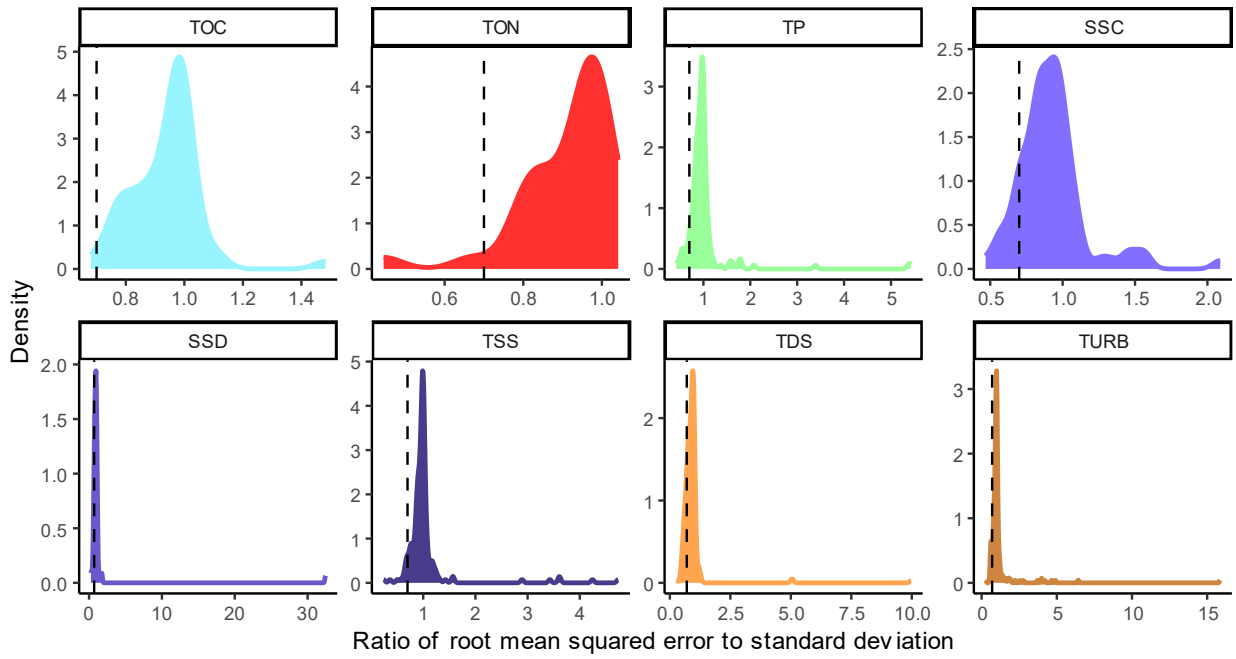

Figure 4.8: Ratio of root mean squared error to standard deviation performance metrics from LOOCV tests for each constituent. The dashed line shows the criteria for this performance metric (< 0.7). Here, 8 of the total 12 constituents with the highest data availability are shown.

As shown in Figure 4.9, the covariates selected in model building varied across different regions in the U.S. West. For TSS, TDS, and TURB constituents, temperature-related covariates (i.e., temperature and potential evaporation) tended to have a greater influence in mountainous and plains regions, whereas water-related covariates (i.e., precipitation and streamflow) were more prevalent in coastal regions. Water-related covariates were consistently dominant over all regions for TOC, TON, TP, SSC, and SSD.

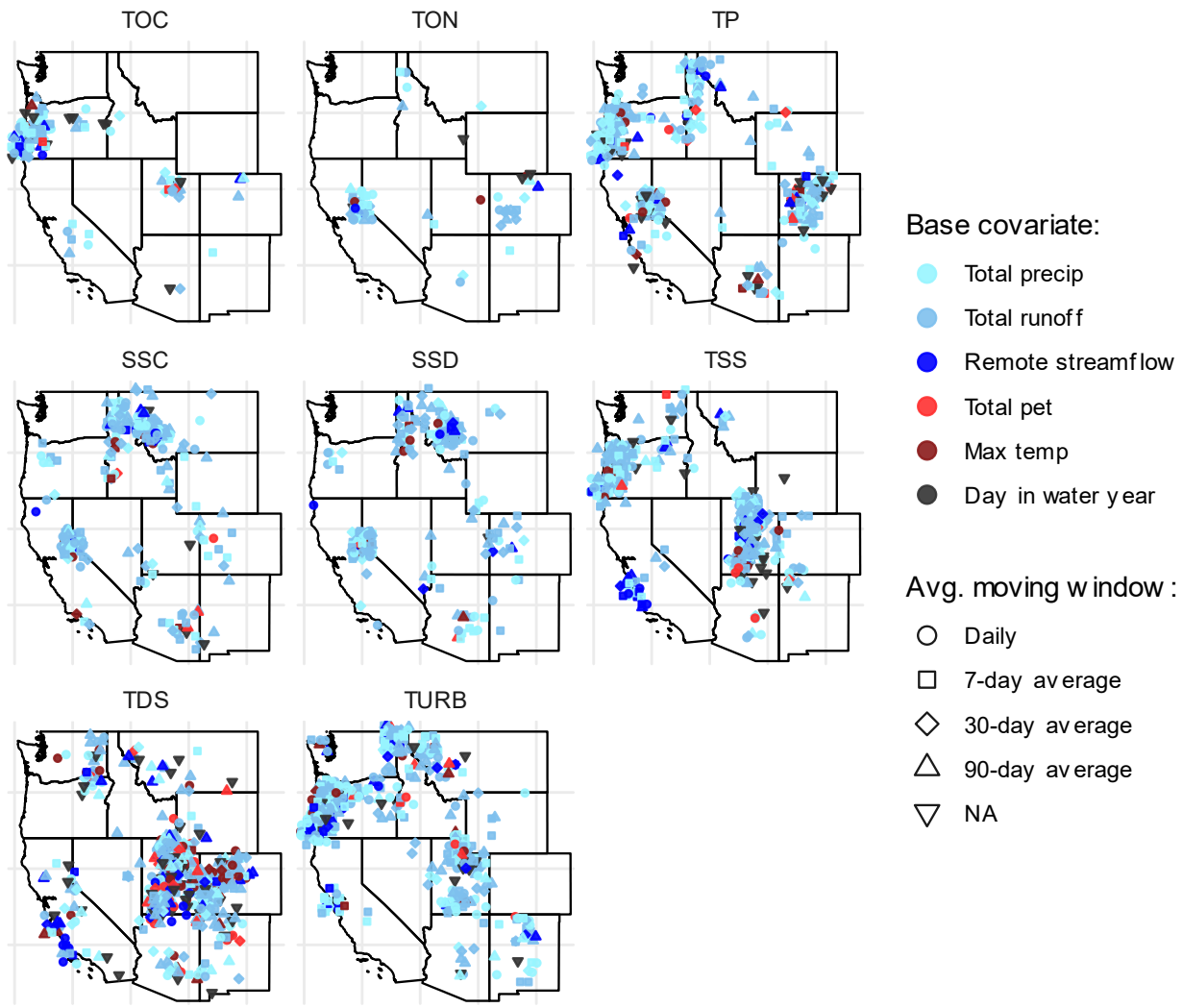

Figure 4.9: Regional distribution of covariates used in linear model building for each constituent. Covariate types are represented by different colors and averaging moving window transformations are represented by different shapes. Here, 8 of the total 12 constituents with the highest data availability are shown.

Pre- and post-fire residuals from the regression analysis were assessed for each year leading up to and following wildfire events. For this analysis, eight constituents with higher data availability were the primary focus, as seen in Figure 4.10. Mean residuals across basins were calculated for each pre- and post-fire year for each constituent. Significant responses for each post-fire year were assessed by comparing means to the variability of paired basin responses, as well as overall variability in pre-fire years. Mean responses for a

given post-fire year were considered significant if they were outside of the 90% confidence interval of paired basin residuals from that year, as well the 90% confidence bounds calculated for all aggregated pre-fire years. With this criterion, significant responses were observed in the fire year for all constituents except TDS, as well as in the following year for all constituents except TSS. Responses of TON, SSC, SSD, and TURB were significantly elevated beyond these first two years post-fire, with significant elevated responses observed 4-5 years after wildfire occurrence. While most constituents showed the highest response during the first two years post-fire, SSD, TDS, and TURB all saw the highest constituent responses 3-5 years after wildfire occurrence.

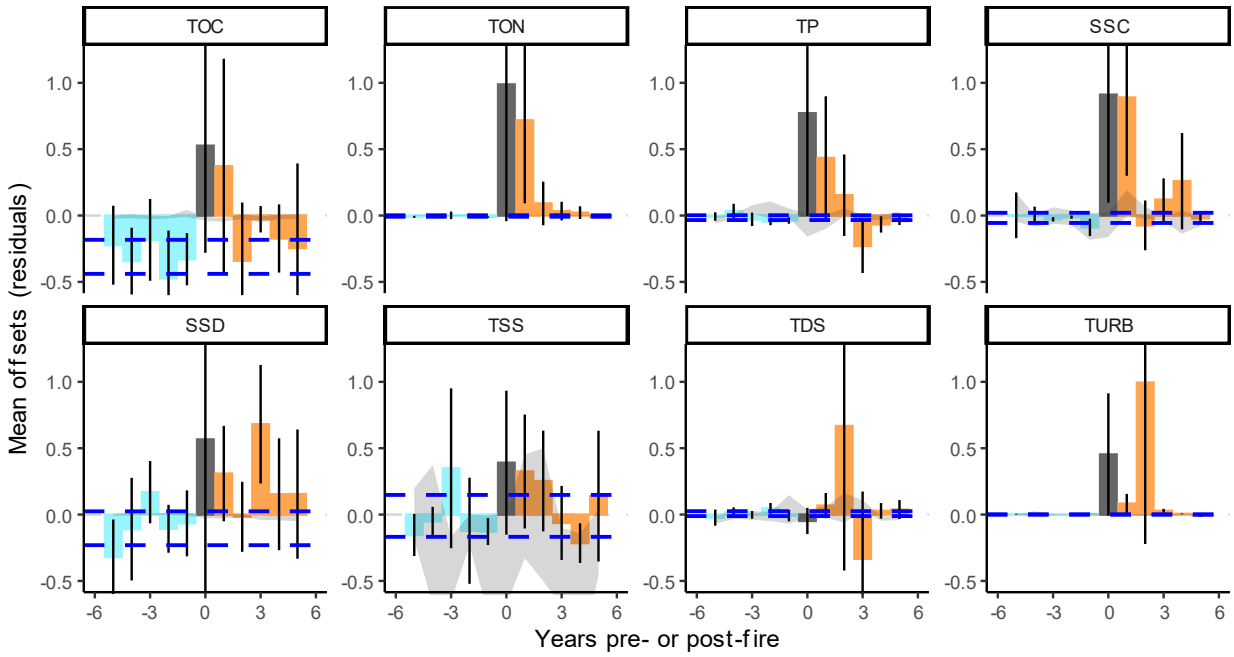

Figure 4.10: Model residuals for each constituent for all basins. Mean residuals for each pre-fire year are shown in light blue with post-fire years shown in orange. Fire year residuals are shown in dark gray. The black vertical lines on each bar represent that year's 90% confidence interval and the gray ribbon represents the 90% confidence interval of paired basins' mean residuals for each pre- and post-fire year. The horizontal, dashed blue lines represent the overall confidence interval bounds for all pre-fire years in aggregate. All data shown here was divided by a normalization factor for each constituent based on the range of mean residuals for plotting and comparison purposes.

This analysis was repeated for models which met satisfactory performance metrics, as seen in Figure 4.11. Few TOC models met performance criteria, thus this constituent was excluded. High performance model residuals generally reflected trends observed in all model residuals, though in some cases with less significance and clarity. SSC, SSD, and TSS showed increasing, significant responses in the first 4-5 years post-fire. TON, TP, and TURB showed increasing responses in the first 1-2 years post-fire, though not all years exhibited a significant change. TDS, however, showed a significant decrease in the 1st through 3rd and 6th years post-fire, largely different from the responses from all model residuals for this constituent. Though these models may show more reliable results, only 10% of models fulfilled the performance criteria and biases from smaller sample sizes may have also affected residual responses.

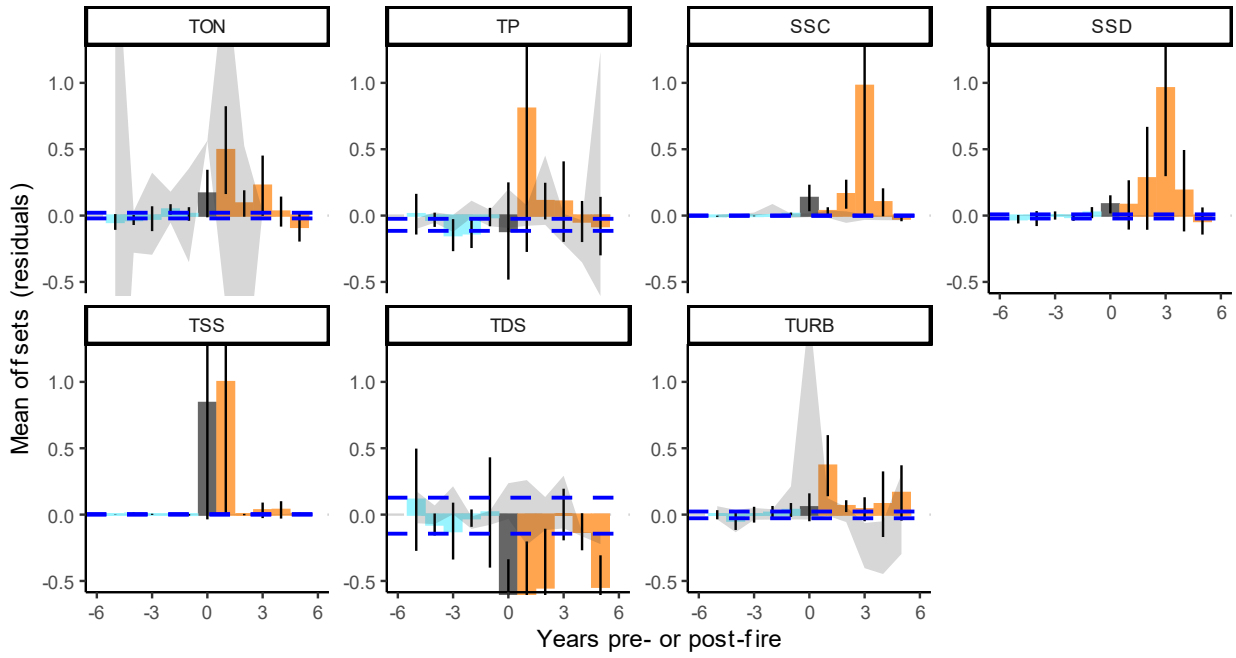

Figure 4.11: Model residuals for each constituent from only high performance models. Mean residuals for each pre-fire year are shown in light blue with post-fire years shown in orange. Fire year residuals are shown in dark gray. The black vertical lines on each bar represent that year's 90% confidence interval and the gray ribbon represents the 90% confidence interval of paired basins' mean residuals for each pre- and post-fire year. The

horizontal, dashed blue lines represent the overall confidence interval bounds for all pre-fire years in aggregate. All data shown here was divided by a normalization factor for each constituent based on the range of mean residuals for plotting and comparison purposes.

4.4.3 Attribution of inter-site variability

Variability in post-fire residuals were compared to four physiographic watershed variables and watershed characteristics—percent forested area, percent developed area, percent burn area, and distance of wildfires to basin outlets—to assess their strength as predictors. As seen in Figure 4.12, Pearson's correlations between physical variables and constituent residuals within two years post-wildfire were calculated to determine whether significant linear relationships existed. P-values of each correlation were additionally calculated to assess if the correlations were significant. Figure 4.12 shows the correlations for each physical variable-constituent combination, with significant values ($p < 0.05$) bolded. Percent forested area had a significant, moderately-strong (>= 0.12) positive relationship with TOC, TON, TP, SSC, and TSS. Distance of the wildfire from the basin outlet had a somewhat negative (<= -0.11), significant correlation with TOC, TP, and TSS.

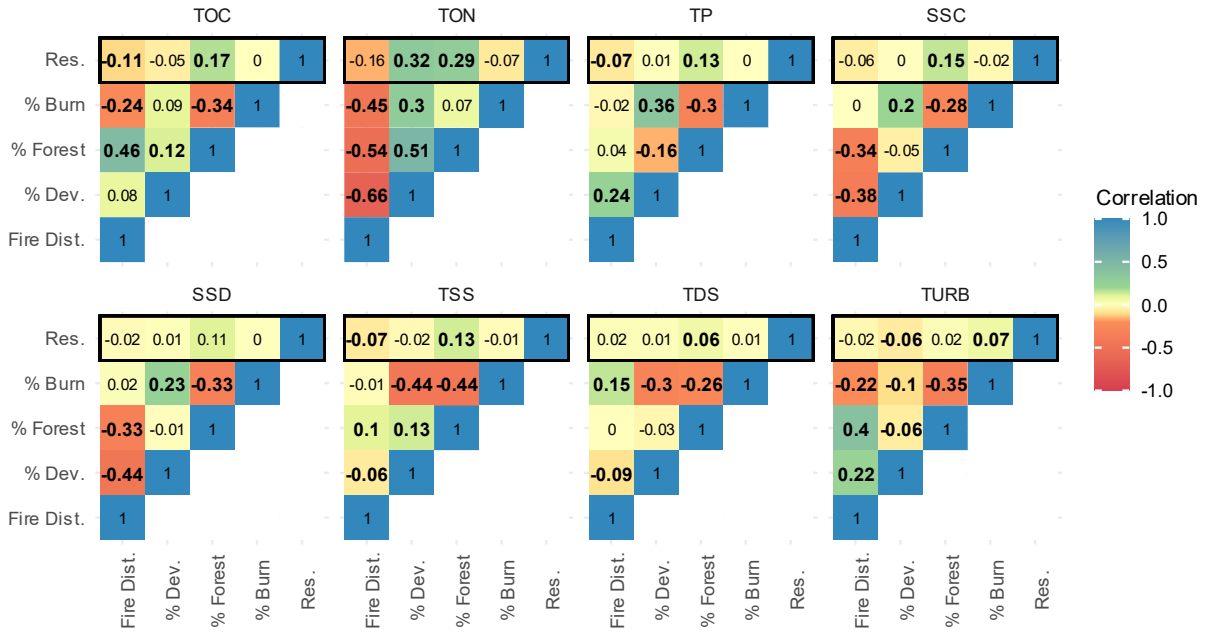

Figure 4.12: A correlation analysis between model residuals for each constituent and watershed and fire characteristics. Residual correlations are shown in the outlined black box, and significant ($p < 0.05$) correlations are bolded and slightly expanded.

To assess predictors' role in constituents' variability, post-fire residuals were also plotted against each analyzed physical predictor, as seen in Figure F.1 in Appendix F. The mean of residuals from within two years of a wildfire event were calculated for each basin-fire combination, then compared to the associated physical predictor values. Best fit linear models were applied to each predictor-residual combination, with the R^2 calculated to assess predictors' roles in explaining response variability. R^2 values were generally low (< 0.05), however percent burn extent explained 10% of the variability in TURB responses, percent forested area explained 9% of the variability in TON responses, and percent developed area explained 29% of the variability in TON responses.

4.5 Discussion

Overall, sharp and significant constituent responses were observed after wildfire events across analyzed basins—consistent with information from previous wildfire research (Bladon et al., 2008; Smith et al., 2011). The significant results from the regression-based analysis validated this study's analytical framework and focused on constituents with sufficient data availability for significant analyses. The inter-site variability analysis showed largely low predictive power of the four tested physiographic watershed variables and wildfire characteristics, however percent forested area emerged as an influential factor for some water quality responses. The following sections will discuss the implications of results from this study, as well as limitations and future improvements.

4.5.1 Implications of data availability and study findings

The study's framework of using custom watershed delineations for increased data availability in analyses resulted in high numbers of available water quality sample sizes compared with previous post-wildfire water quality studies, allowing for significant empirical assessments. Due to the data-scarce nature of post-wildfire water quality analyses, custom delineations were necessary for creating a data-rich analytic setup for analysis. This is exemplified in this study by the filtering process applied to the initial ~50,000 delineated basins, where fewer than 1% had the appropriate spatial and temporal overlap in data for post-wildfire water quality analyses. The use of paired unburned basins in this study additionally increased data availability for characterization of natural variability in constituent responses, supplementing available pre-fire data in burned basins to serve as a comparison to post-fire responses. The analysis of 241 burned basins, along with 258 paired unburned basins and over 200,000 combined datapoints was considerably more numerous than burned basins evaluated in previous studies which relied on GAGES

II basin delineations, or approximately 65 to 153 (Beyene et al., 2023; Rust et al., 2018; Williams et al., 2022).

Though inaccuracies in the delineation process resulted in an estimated 10% of erroneously delineated basins, the high number of available basins used in the analysis likely offset any effects of incorrect basin definitions. This was validated by post-fire responses in the raw water quality data which generally reflected findings from previous studies. For example, the Rust et al., 2018 study calculated an average 2350, 245, and 23% in percent differences in turbidity, TSS, and DOC values, respectively, before and after wildfires across basins in the U.S. West, where 1115, 83, and 21% percent changes, respectively, were calculated here. DON overall negative change when percent differences in means were averaged across basins, which is also consistent with findings in Rust et al., 2018. Thus, the number of basins incorrectly categorized as "burned" or "unburned" due to incorrect delineations seemed to have a low influence on available burned and unburned water quality values.

However, significance of responses in individual basins was low, ranging from 19-39% for each constituent, with the exception of an 80% significance rate for TN—likely due to high variability in compounding geophysical variables driving constituent response. The regression-based analysis across all basins, on the other hand, allowed for control and characterization of hydroclimatic variables' influence, and thus exhibited high significance in post-wildfire response in each constituent. By analyzing data in aggregate across basins, significance could be assessed with the relatively higher data sample sizes.

Most constituents exhibited significantly elevated responses in the first one to two years after a wildfire event, with sharp declines in subsequent years. This reflects previous studies which have observed highest post-wildfire water quality responses within the first

two to three years after a wildfire event (Smith et al., 2011). TOC and TP levels were significantly elevated for the first 2-3 years post-fire, then returned to pre-fire levels in subsequent years. TON, however, had a sustained, significant response for the first 5 years after wildfire occurrence before returning to pre-fire levels. Elevated stream nitrogen has been observed in previous studies after wildfires, likely due to lack of vegetation recovery which can be a dominant driver of nitrogen (Rhoades et al., 2019a). Sediment characteristics SSC and SSD also showed elevated responses 5-6 years after wildfire occurrence. This is consistent with previous studies which have shown extended sediment responses post-wildfire due to exacerbated erosional effects from loss of vegetative roots (Larson-Nash et al., 2018; Robichaud et al., 2016; Smith et al., 2011). TSS showed a similarly extended response, though not significantly different from unburned basin responses in most years. TDS, on the other hand, exhibited high variability between positive and negative responses in post-fire years. This may have been caused by a higher influence of background sources on this constituent than others. Finally, turbidity exhibited a similar, prolonged response to sediment characteristics, consistent with previous literature showing a strong correlation between these two constituents (Brucker et al., 2023), though a sharp decline occurred after just 4 years. Residuals from high performance models largely reflected these results, providing further validation of the significance of these results.

4.5.2 Limitations and future work

Potentially erroneous watershed delineations and low performance metrics in linear models contributed uncertainty to results in this study. The 9% error in the basin delineation process was calculated from a subset of basins, meaning the actual rate of

erroneous delineations is unknown and may be larger. The measure of inaccurately-delineated basins is also based the assumption that GAGES II delineations are the "true" watershed definitions, though error may exist in this dataset as well. As basin delineations were created from almost 50,000 pourpoints, it was unrealistic to visually inspect each pourpoint and its associated basin. However, previous studies have developed and tested more advanced, automatic pourpoint snapping algorithms for similar mass delineation cases, reporting ~50 to 75% reduction in error from more traditional methods (Lindsay et al., 2008; Xie et al., 2022). Future developments in this work will consider implementing a similar, more advanced snapping algorithm. Additionally, though the flow direction and flow accumulation grids in this study were generally similar to those available through the NHDplus database, re-doing the analysis with NHDplus data may further decrease erroneous basin delineations.

Model performance metrics from many of the linear models developed were often too poor to have confidence in their predictive abilities, with only 10% of models meeting satisfactory performance criteria. RSR scores were generally poorest for sediment and turbidity characteristics, with medians ranging from 0.90 to 0.98. This is likely due to high variability and frequent outliers typical in sediment responses (Smith et al., 2011). Median RSR scores for dissolved constituents were only slightly lower, however, ranging from 0.87 to 0.97. Future analysis in this work will consider potential data transformations prior to modeling and analysis to improve model performance, as well as applying models to aggregated data from within specific regions in the U.S. West for increased data availability. As confidence intervals are currently used as a method of significance assessment, future work will also consider bootstrapping as an alternative, more robust method of significance testing.

Additional analyses of inter-site variability in constituent responses will also be completed in future work. Previous studies have shown NDVI (Rhea et al., 2022; Rust et al., 2019), burn severity levels (Beyene et al., 2023; Brucker et al., 2023), and soil characteristics (Rust et al., 2019) to be strong determining factors in post-wildfire water quality response, thus these and other geophysical watershed variables and wildfire characteristics should be assessed as predictors of constituent residuals in this analysis. Additionally, though an initial random forest modeling framework has been developed tested as a method of inter-site variability characterization, this analysis will be fully implemented for all constituents and potential physical predictors. This will serve as an additional method of characterizing physical variables' influence on residual variability though calculations of variable importance.

4.6 Conclusion

Through optimized data availability using custom basin delineations, as well as statistical analyses of water quality data aggregated across 499 burned and unburned basins, this study provided significant insights into broad, post-fire changes in DOM, nutrients, and sediment across the U.S. West. Significant responses were observed in the year post-wildfire for all constituents except TDS. Though most constituents saw a sharp decrease in concentrations after the second or third year following a wildfire event, continued elevated responses up to six years post-wildfire were observed for nitrogen and sediment characteristics. These findings generally reflected results from previous *in situ* and modeling analyses, in terms of longevity of elevated post-fire concentrations.

An analysis of inter-site variability between constituent responses provided additional information about potential compounding factors driving responses in water

quality. Though predictive power of the evaluated physical watershed and wildfire characteristics was low overall, percent forested area emerged as a dominant driver of response, especially for the dissolved constituents. Overall, the information provided here on the longevity of elevated water quality response, as well as influential basin and wildfire characteristics may help inform water managers in planning and mitigative efforts of wildfire effects.

Chapter 5

Conclusion

5.1 Overview

This dissertation presented a multi-scale analysis of wildfire effects on soil and water physical and chemical properties, and resulting responses in runoff, sediment, DOM, and nutrients. A two-pronged analysis was used to investigate both small-scale effects and broad, regional responses in watersheds across the U.S. West. This framework aimed to assess drivers and magnitudes of post-wildfire response using both ground-up and top-down approaches: The small-scale analysis provided insights into underlying mechanisms driving post-fire responses, while the regional analysis used empirical water quality data to derive predictors' influence on responses. Key findings from the initial small-scale analysis additionally informed the design of and motivation for the subsequent, large-scale analysis—guiding the selection of predictor variables and factors included.

Though vastly different in scale and methods, both analysis types were designed to address a key issue in previous analysis efforts of post-wildfire water quality response: variability and data scarcity. Previous studies have cited variability in post-wildfire responses due to highly variable natural environments, as well as data scarcity, and

significant hinderances in creating significant assessments. The small-scale analysis used laboratory-based simulation experiments to observe post-fire responses in a controlled environment, limiting variability and isolating driver effects. The large-scale analysis utilized high numbers of datapoints aggregated across numerous basins, as well as geostatistical and machine learning techniques to characterize hydroclimatic variables' role in constituent variability, allowing for subsequent isolation of wildfire effects.

Key factors, trends, and limitations in the laboratory-scale simulation experiment informed the hypotheses explored in the regional, statistical analysis. Rainfall characteristics in the laboratory experiment were strongly influential on constituent responses and in some cases overpowered burn effects. This informed the selection of climate variables as predictors in the regional regression-based analysis. The regional analysis additionally attempted to highlight the influence of extraneous factors on the basin-scale which were not captured in the laboratory simulations. For example, sedimentation mechanisms which exist in natural settings, i.e. erosional and geomorphological forces (Kampf et al., 2016b; Larson-Nash et al., 2018; Robichaud et al., 2016) were not observable on the small scale—resulting in lower responses compared with *in situ* studies (Blake et al., 2010; Knight et al., 1983; Shahlaee et al., 1991). As all driving mechanisms of sedimentation were captured in the empirical analysis, response differences provided insight into the individual contributions of larger-scale erosional forces and smaller-scale ash and combusted vegetation transport in total post-fire sedimentation. Additionally, DOM and nutrient responses closely reflected each other in shape and magnitude on the small scale, and were both strongly influenced by burn effects. However, at the basin scale and over longer periods of time, these responses are typically not as closely correlated due to the influence of vegetation regrowth on nutrient cycling (Ice et al.,

2004). Differences in responses in the regional analysis provided insight into the influence of vegetation and other basin-scale factors on DOM and nutrient responses as compared with burn effects.

Though each analysis type captured different mechanisms driving post-wildfire water quality response, similar themes were apparent throughout the small- and regional-scales. High variability existed in observations of sediment responses across scales, as compared with dissolved constituents which had more consistent trends with burn effects. This observation has been made by previous studies as well and is typically attributed to the complex interactions of ash, loss of vegetation, and increased runoff rates with erosional and geomorphic processes (Cotrufo et al., 2016; Ebel et al., 2012b; Shakesby and Doerr, 2006). Non-normality and non-linearity of constituent responses was also apparent across scales. This was exemplified by the inverse "U" trends in organic carbon and nitrogen responses with increasing burn severity in the laboratory-scale simulation, and highly right-skewed distributions of sediment response data in the regional analysis.

The combined results from this dissertation provided comprehensive insights into short-term (hours), small-scale (< 1 m^2) responses to burn effects, as well as long-term (years), basin-scale effects (~5-20,000 km^2). These paired analysis inform nuances of complex driver interactions in post-wildfire environments, but also provided information on broad trends and influential physical variables. Together, these key insights may help inform basins' vulnerability to wildfire effects, assisting water managers in planning and mitigative efforts.

5.2 Contributions

The novel contributions of this research are as follows:

Chapter 2

- The first review on wildfire effects on water quality and supply which focuses exclusively on methodological techniques (i.e., wildfire and rainfall simulation) used to observe laboratory- and plot-scale burn effects.
- Recommendations for future researchers in the design and implementation of wildfire and rainfall simulation experiments.

Chapter 3

- Development of laboratory-scale wildfire and rainfall simulation experiments which analyze burn effects on water quality and supply through incorporation of three key drivers at multiple increments.
- Testing the implementation of this unique simulation framework in its ability to create reliable observations of burn effects on runoff, sediment, dissolved organic matter, and nitrogen generation, as well as turbidity.

Chapter 4

- Delineation and compilation of a unique dataset of 646 basins with ~250,00 empirical sediment, nutrient, and DOM observations for assessing post-fire water quality across the western U.S.
- Development of a data-driven statistical and machine learning analytical framework which used in situ and satellite-derived data to identify elevated post-wildfire sediment, nutrient, and DOM responses in basins across the U.S. West.

- Identification of regionally-specific covariates which were used in the highest skill predictive models—allowing for a characterization of basins' vulnerability to water quality effects.

5.3 Dissemination

All three chapters of this dissertation will be disseminated as journal publications. Chapter 2 has already been published in a peer-reviewed journal, and Chapter 3 has been submitted and is in review for publication in a peer-reviewed journal. The contents of Chapter 4 will be refined slightly before submission to a journal. Specifically, additional analyses of inter-site variability will be included. This process will be completed during Summer 2023, with the manuscript submitted before the end of the summer.

Bibliography

Abatzoglou, J.T., Williams, A.P., 2016. Impact of anthropogenic climate change on wildfire across western US forests. Proc Natl Acad Sci USA 113, 11770–11775. https://doi.org/10.1073/pnas.1607171113

Abraham, J., Dowling, K., Florentine, S., 2017. Risk of post-fire metal mobilization into surface water resources: A review. Science of The Total Environment 599–600, 1740–1755. https://doi.org/10.1016/j.scitotenv.2017.05.096

Alexander, R.R., Watkins, R.K., 1977. The Fraser Experimental Forest, Colorado. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station.

Alstatt, D., Miles, R.L., 1983. Soil survey of Grand County area, Colorado. The Service.

American Public Health Association, American Water Works Association, 2012. Standard methods for the examination of water and wastewater, in: Standard Methods for the Examination of Water and Wastewater.

Badía-Villas, D., González-Pérez, J.A., Aznar, J.M., Arjona-Gracia, B., Martí-Dalmau, C., 2014. Changes in water repellency, aggregation and organic matter of a mollic horizon burned in laboratory: Soil depth affected by fire. Geoderma 213, 400–407. https://doi.org/10.1016/j.geoderma.2013.08.038

Balfour, V., Woods, S., 2008. Causes of Variability in the Effects of Vegetative Ash on Post-Fire Runoff and Erosion. AGU Fall Meeting Abstracts 11, H11C-0778.

Basso, M., Serpa, D., Mateus, M., Keizer, J.J., Vieira, D.C.S., 2022. Advances on water quality modeling in burned areas: A review. PLOS Water 1, e0000025. https://doi.org/10.1371/journal.pwat.0000025

Becker, W.C., Hohner, A., Rosario-Ortiz, F., DeWolfe, J., 2018a. Preparing for Wildfires and Extreme Weather: Plant Design and Operation Recommendations. Journal: American Water Works Association 110, 32–40. https://doi.org/10.1002/awwa.1113

Becker, W.C., Hohner, A., Rosario-Ortiz, F., DeWolfe, J., 2018b. Preparing for Wildfires and Extreme Weather: Plant Design and Operation Recommendations. Journal: American Water Works Association 110, 32–40. https://doi.org/10.1002/awwa.1113

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

Benavides-Solorio, J., MacDonald, L.H., 2001. Post-fire runoff and erosion from simulated rainfall on small plots, Colorado Front Range. Hydrological Processes 15, 2931–2952. https://doi.org/10.1002/hyp.383

Beyene, M.T., Leibowitz, S.G., Dunn, C.J., Bladon, K.D., 2023. To burn or not to burn: An empirical assessment of the impacts of wildfires and prescribed fires on trace element concentrations in Western US streams. Science of The Total Environment 863, 160731. https://doi.org/10.1016/j.scitotenv.2022.160731

Bladon, K.D., Emelko, M.B., Silins, U., Stone, M., 2014. Wildfire and the Future of Water Supply. Environ. Sci. Technol. 48, 8936–8943. https://doi.org/10.1021/es500130g

Bladon, K.D., Silins, U., Wagner, M.J., Stone, M., Emelko, M.B., Mendoza, C.A., Devito, K.J., Boon, S., 2008. Wildfire impacts on nitrogen concentration and production from headwater streams in southern Alberta's Rocky Mountains. Can. J. For. Res. 38, 2359–2371. https://doi.org/10.1139/X08-071

Blake, D., Nyman, P., Nice, H., D'Souza, F.M.L., Kavazos, C.R.J., Horwitz, P., Blake, D., Nyman, P., Nice, H., D'Souza, F.M.L., Kavazos, C.R.J., Horwitz, P., 2020. Assessment of post-wildfire erosion risk and effects on water quality in south-western Australia. Int. J. Wildland Fire 29, 240–257. https://doi.org/10.1071/WF18123

Blake, W.H., Theocharopoulos, S.P., Skoulikidis, N., Clark, P., Tountas, P., Hartley, R., Amaxidis, Y., 2010. Wildfire impacts on hillslope sediment and phosphorus yields. J Soils Sediments 10, 671–682. https://doi.org/10.1007/s11368-010-0201-y

Blank, R.R., Allen, F., Young, J.A., 1994. Extractable Anions in Soils following Wildfire in a Sagebrush-Grass Community. Soil Science Society of America Journal 58, 564–570. https://doi.org/10.2136/sssaj1994.03615995005800020045x

Breiman, L., 2001. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). Statistical Science 16, 199–231. https://doi.org/10.1214/ss/1009213726

Brucker, C., Livneh, B., Butler, C., Rosario-Ortiz, F., 2023. A laboratory-scale simulation framework for analyzing wildfire hydrologic and water quality effects. https://doi.org/10.5281/zenodo.7799976

Brucker, C.P., Livneh, B., Minear, J.T., Rosario, F., 2022. A review of simulation experiment techniques used to analyze wildfire effects on water quality and supply. Environmental Science: Processes & Impacts.

Burnham, K.P., Anderson, D.R., 1998. Practical Use of the Information-Theoretic Approach, in: Burnham, K.P., Anderson, D.R. (Eds.), Model Selection and Inference: A Practical Information-Theoretic Approach. Springer, New York, NY, pp. 75–117. https://doi.org/10.1007/978-1-4757-2917-7_3

Busse, M.D., Hubbert, K.R., Fiddler, G.O., Shestak, C.J., Powers, R.F., Busse, M.D., Hubbert, K.R., Fiddler, G.O., Shestak, C.J., Powers, R.F., 2005. Lethal soil temperatures during burning of masticated forest residues. Int. J. Wildland Fire 14, 267–276. https://doi.org/10.1071/WF04062

Busse, M.D., Shestak, C.J., Hubbert, K.R., Knapp, E.E., 2010. Soil Physical Properties Regulate Lethal Heating during Burning of Woody Residues. Soil Science Society of America Journal 74, 947–955. https://doi.org/10.2136/sssaj2009.0322

Cancelo-González, J., Barros, N., Rial-Rivas, M.E., Díaz-Fierros, F., 2012. Assessment of the impact of soil heating on soil cations using the degree-hours method . Spanish Journal of Soil Science 2. https://doi.org/10.3232/SJSS.2012.V2.N3.04

Cancelo-González, J., Rial-Rivas, M.E., Díaz-Fierros, F., 2013. Effects of fire on cation content in water: a laboratory simulation study. Int. J. Wildland Fire 22, 667–680. https://doi.org/10.1071/WF12178

Cannon, S.H., Gartner, J.E., Rupert, M.G., Michael, J.A., Rea, A.H., Parrett, C., 2010. Predicting the probability and volume of postwildfire debris flows in the intermountain western United States. GSA Bulletin 122, 127–144. https://doi.org/10.1130/B26459.1

Cawley, K.M., Hohner, A.K., Podgorski, D.C., Cooper, W.T., Korak, J.A., Rosario-Ortiz, F.L., 2017. Molecular and Spectroscopic Characterization of Water Extractable Organic Matter from Thermally Altered Soils Reveal Insight into Disinfection Byproduct Precursors. Environ. Sci. Technol. 51, 771–779. https://doi.org/10.1021/acs.est.6b05126

Certini, G., 2005. Effects of fire on properties of forest soils: a review. Oecologia 143, 1–10. https://doi.org/10.1007/s00442-004-1788-8

Chandler, C., Cheney, P., Thomas, P., Trabaud, L., Williams, D., 1983. Fire in forestry. Volume 1. Forest fire behavior and effects. Volume 2. Forest fire management and organization. John Wiley & Sons, Inc., New York.

Cho, J.H., Lee, J.H., 2018. Multiple Linear Regression Models for Predicting Nonpoint-Source Pollutant Discharge from a Highland Agricultural Region. Water 10, 1156. https://doi.org/10.3390/w10091156

Cotrufo, M.F., Boot, C.M., Kampf, S., Nelson, P.A., Brogan, D.J., Covino, T., Haddix, M.L., MacDonald, L.H., Rathburn, S., Ryan-Bukett, S., Schmeer, S., Hall, E., 2016. Redistribution of pyrogenic carbon from hillslopes to stream corridors following a large montane wildfire. Global Biogeochemical Cycles 30, 1348–1355. https://doi.org/10.1002/2016GB005467

Culler, E.S., Livneh, B., Rajagopalan, B., Tiampo, K.F., 2023. A data-driven evaluation of post-fire landslide susceptibility. Natural Hazards and Earth System Sciences 23, 1631–1652. https://doi.org/10.5194/nhess-23-1631-2023

Cutler, F. original by L.B. and A., Wiener, R. port by A.L. and M., 2022. randomForest: Breiman and Cutler's Random Forests for Classification and Regression.

Das, B.M., Sobhan, K., 2010. Principles of geotechnical engineering, CENGAGE Learning. Stamford, Connecticut.

Datta, S., Karmakar, S., Mezbahuddin, S., Hossain, M.M., Chaudhary, B.S., Hoque, Md.E., Abdullah Al Mamun, M.M., Baul, T.K., 2022. The limits of watershed delineation: implications of different DEMs, DEM resolutions, and area threshold values. Hydrology Research 53, 1047–1062. https://doi.org/10.2166/nh.2022.126

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

Debano, L.F., Krammes, J.S., 1966. WATER REPELLENT SOILS AND THEIR RELATION TO WILDFIRE TEMPERATURES. International Association of Scientific Hydrology. Bulletin 11, 14–19. https://doi.org/10.1080/02626666609493457

Doerr, S.H., Shakesby, R.A., Blake, W.H., Chafer, C.J., Humphreys, G.S., Wallbrink, P.J., 2006. Effects of differing wildfire severities on soil wettability and implications for hydrological response. Journal of Hydrology 319, 295–311. https://doi.org/10.1016/j.jhydrol.2005.06.038

Ebel, B.A., Moody, J.A., 2017. Synthesis of soil-hydraulic properties and infiltration timescales in wildfire-affected soils. Hydrological Processes 31, 324–340. https://doi.org/10.1002/hyp.10998

Ebel, B.A., Moody, J.A., Martin, D.A., 2012a. Hydrologic conditions controlling runoff generation immediately after wildfire. Water Resources Research 48. https://doi.org/10.1029/2011WR011470

Ebel, B.A., Moody, J.A., Martin, D.A., 2012b. Hydrologic conditions controlling runoff generation immediately after wildfire. Water Resources Research 48. https://doi.org/10.1029/2011WR011470

Emmerich, W.E., Cox, J.R., 1992. Hydrologic characteristics immediately after seasonal burning on introduced and native grasslands. Rangeland Ecology & Management / Journal of Range Management Archives 45, 476–479.

Essery, R., Rutter, N., Pomeroy, J., Baxter, R., Stähli, M., Gustafsson, D., Barr, A., Bartlett, P., Elder, K., 2009. SNOWMIP2: An Evaluation of Forest Snow Process Simulations. Bulletin of the American Meteorological Society 90, 1120–1136. https://doi.org/10.1175/2009BAMS2629.1

Ferreira, A.J.D., Coelho, C.O.A., Ritsema, C.J., Boulet, A.K., Keizer, J.J., 2008. Soil and water degradation processes in burned areas: Lessons learned from a nested approach. CATENA, Fire Effects on Soil Properties 74, 273–285. https://doi.org/10.1016/j.catena.2008.05.007

Gannon, B.M., Wei, Y., Thompson, M.P., Scott, J.H., Short, K.C., 2022. System Analysis of Wildfire-Water Supply Risk in Colorado, USA with Monte Carlo Wildfire and Rainfall Simulation. Risk Analysis 42, 406–424. https://doi.org/10.1111/risa.13762

García-Gaines, R.A., Frankenstein, S., 2015. USCS and the USDA Soil Classification System : Development of a mapping scheme (Report), This Digital Resource was created in Microsoft Word and Adobe Acrobat. Cold Regions Research and Engineering Laboratory (U.S.).

Hatch Corporation, 2014. Hach 2100N User Manual, 5th Edition.

Hester, J.W., Thurow, T.L., Taylor, C.A., 1997. Hydrologic characteristics of vegetation types as affected by prescribed burning. Rangeland Ecology & Management / Journal of Range Management Archives 50, 199–204.

Hogue, B.A., Inglett, P.W., 2012. Nutrient release from combustion residues of two contrasting herbaceous vegetation types. Science of The Total Environment 431, 9–19. https://doi.org/10.1016/j.scitotenv.2012.04.074

Hohner, A.K., Cawley, K., Oropeza, J., Summers, R.S., Rosario-Ortiz, F.L., 2016. Drinking water treatment response following a Colorado wildfire. Water Research 105, 187–198. https://doi.org/10.1016/j.watres.2016.08.034

Hohner, A.K., Summers, R.S., Rosario-Ortiz, F.L., 2019. Laboratory simulation of postfire effects on conventional drinking water treatment and disinfection byproduct formation. AWWA Water Science 1, e1155. https://doi.org/10.1002/aws2.1155

Hua, G., Reckhow, D.A., 2007. Comparison of disinfection byproduct formation from chlorine and alternative disinfectants. Water Research 41, 1667–1678. https://doi.org/10.1016/j.watres.2007.01.032

HURVICH, C.M., TSAI, C.-L., 1989. Regression and time series model selection in small samples. Biometrika 76, 297–307. https://doi.org/10.1093/biomet/76.2.297

Ice, G.G., Neary, D.G., Adams, P.W., 2004. Effects of Wildfire on Soils and Watershed Processes. j for 102, 16–20. https://doi.org/10.1093/jof/102.6.16

Jain, P., Coogan, S.C.P., Subramanian, S.G., Crowley, M., Taylor, S., Flannigan, M.D., 2020. A review of machine learning applications in wildfire science and management. Environ. Rev. 28, 478–505. https://doi.org/10.1139/er-2020-0019

Jenson, S.K., 1991. Applications of hydrologic information automatically extracted from digital elevation models. Hydrological Processes 5, 31–44. https://doi.org/10.1002/hyp.3360050104

Jian, M., Berli, M., Ghezzehei, T.A., 2018. Soil Structural Degradation During Low-Severity Burns. Geophysical Research Letters 45, 5553–5561. https://doi.org/10.1029/2018GL078053

Johansen, M.P., Hakonson, T.E., Breshears, D.D., 2001. Post-fire runoff and erosion from rainfall simulation: contrasting forests with shrublands and grasslands. Hydrological Processes 15, 2953–2965. https://doi.org/10.1002/hyp.384

Kampf, S.K., Brogan, D.J., Schmeer, S., MacDonald, L.H., Nelson, P.A., 2016a. How do geomorphic effects of rainfall vary with storm type and spatial scale in a post-fire landscape? Geomorphology 273, 39–51. https://doi.org/10.1016/j.geomorph.2016.08.001

Kampf, S.K., Brogan, D.J., Schmeer, S., MacDonald, L.H., Nelson, P.A., 2016b. How do geomorphic effects of rainfall vary with storm type and spatial scale in a post-fire landscape? Geomorphology 273, 39–51. https://doi.org/10.1016/j.geomorph.2016.08.001

Kampf, S.K., Gannon, B.M., Wilson, C., Saavedra, F., Miller, M.E., Heldmyer, A., Livneh, B., Nelson, P., MacDonald, L., 2020. PEMIP: Post-fire erosion model inter-comparison project. Journal of Environmental Management 268, 110704. https://doi.org/10.1016/j.jenvman.2020.110704

Keeley, J.E., 2009. Fire intensity, fire severity and burn severity: a brief review and suggested usage. Int. J. Wildland Fire 18, 116–126. https://doi.org/10.1071/WF07049

Keesstra, S.D., Maroulis, J., Argaman, E., Voogt, A., Wittenberg, L., 2014a. Effects of controlled fire on hydrology and erosion under simulated rainfall. Cuadernos de Investigación Geográfica 40, 269–294. https://doi.org/10.18172/cig.2532

Keesstra, S.D., Maroulis, J., Argaman, E., Voogt, A., Wittenberg, L., 2014b. Effects of controlled fire on hydrology and erosion under simulated rainfall. Cuadernos de Investigación Geográfica 40, 269–294. https://doi.org/10.18172/cig.2532

Keesstra, S.D., Maroulis, J., Argaman, E., Voogt, A., Wittenberg, L., 2014c. Effects of controlled fire on hydrology and erosion under simulated rainfall. Cuadernos de Investigación Geográfica 40, 269–294. https://doi.org/10.18172/cig.2532

Kibet, L.C., Saporito, L.S., Allen, A.L., May, E.B., Kleinman, P.J.A., Hashem, F.M., Bryant, R.B., 2014. A Protocol for Conducting Rainfall Simulation to Study Soil Runoff. J Vis Exp. https://doi.org/10.3791/51664

Klopatek, C.C., Debano, L.F., Klopatek, J.M., 1988. Effects of simulated fire on vesicular-arbuscular mycorrhizae in pinyon-juniper woodland soil. Plant Soil 109, 245–249. https://doi.org/10.1007/BF02202090

Knight, R.W., Blackburn, W.H., Scifres, C.J., 1983. Infiltration Rates and Sediment Production following Herbicide/Fire Brush Treatments. Journal of Range Management 36, 154–157. https://doi.org/10.2307/3898151

Kral, K.C., Limb, R.F., Hovick, T.J., McGranahan, D.A., Field, A.L., O'Brien, P.L., 2015. Simulating Grassland Prescribed Fires Using Experimental Approaches. fire ecol 11, 34–44. https://doi.org/10.4996/fireecology.1103034

Lane, P.N.J., Sheridan, G.J., Noske, P.J., 2006. Changes in sediment loads and discharge from small mountain catchments following wildfire in south eastern Australia. Journal of Hydrology 331, 495–510. https://doi.org/10.1016/j.jhydrol.2006.05.035

Langhans, C., Smith, H.G., Chong, D.M.O., Nyman, P., Lane, P.N.J., Sheridan, G.J., 2016. A model for assessing water quality risk in catchments prone to wildfire. Journal of Hydrology 534, 407–426. https://doi.org/10.1016/j.jhydrol.2015.12.048

Larsen, I.J., MacDonald, L.H., 2007. Predicting postfire sediment yields at the hillslope scale: Testing RUSLE and Disturbed WEPP. Water Resources Research 43. https://doi.org/10.1029/2006WR005560

Larson-Nash, S.S., Robichaud, P.R., Pierson, F.B., Moffet, C.A., Williams, C.J., Spaeth, K.E., Brown, R.E., Lewis, S.A., 2018. Recovery of small-scale infiltration and erosion after wildfires. Journal of Hydrology and Hydromechanics 66, 261–270. https://doi.org/10.1515/johh-2017-0056

Lawrence, R.L., 2020. Addressing constraints to restoration of highly disturbed ecosystems affected by cheatgrass invasion and slash pile burning (Text). Colorado State University.

Lentile, L.B., Holden*, Z.A., Smith*, A.M.S., Falkowski, M.J., Hudak, A.T., Morgan, P., Lewis, S.A., Gessler, P.E., Benson, N.C., 2006. Remote sensing techniques to assess active fire characteristics and post-fire effects. Int. J. Wildland Fire 15, 319–345. https://doi.org/10.1071/WF05097

Lindsay, J.B., 2016. The practice of DEM stream burning revisited. Earth Surface Processes and Landforms 41, 658–668. https://doi.org/10.1002/esp.3888

Lindsay, J.B., Rothwell, J.J., Davies, H., 2008. Mapping outlet points used for watershed delineation onto DEM-derived stream networks. Water Resources Research 44.

Liu, Y., Gupta, H.V., 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. Water Resources Research 43. https://doi.org/10.1029/2006WR005756

Marcos, E., Tarrega, R., Luis-Calabuig, E., 2000. Comparative Analysis of Runo and Sediment Yield with a Rainfall Simulator After Experimental Fire. Arid Soil Research and Rehabilitation 14, 293–307. https://doi.org/10.1080/089030600406699

Marlon, J.R., Bartlein, P.J., Walsh, M.K., Harrison, S.P., Brown, K.J., Edwards, M.E., Higuera, P.E., Power, M.J., Anderson, R.S., Briles, C., Brunelle, A., Carcaillet, C., Daniels, M., Hu, F.S., Lavoie, M., Long, C., Minckley, T., Richard, P.J.H., Scott, A.C., Shafer, D.S., Tinner, W., Umbanhowar, C.E., Whitlock, C., 2009. Wildfire responses to abrupt climate change in North America. PNAS 106, 2519–2524. https://doi.org/10.1073/pnas.0808212106

McManus, M.G., D'Amico, E., Smith, E.M., Polinsky, R., Ackerman, J., Tyler, K., 2020. Variation in stream network relationships and geospatial predictions of watershed conductivity. Freshwater Science 39, 704–721. https://doi.org/10.1086/710340

Meyer-Jacob, C., Michelutti, N., Paterson, A.M., Cumming, B.F., Keller, W. (Bill), Smol, J.P., 2019. The browning and re-browning of lakes: Divergent lake-water organic carbon trends linked to acid deposition and climate change. Sci Rep 9, 16676. https://doi.org/10.1038/s41598-019-52912-0

Mishra, A., Alnahit, A., Campbell, B., 2021. Impact of land uses, drought, flood, wildfire, and cascading events on water quality and microbial communities: A review and analysis. Journal of Hydrology 596, 125707. https://doi.org/10.1016/j.jhydrol.2020.125707

Moody, J.A., Martin, D.A., 2009a. Synthesis of sediment yields after wildland fire in different rainfall regimes in the western United States. Int. J. Wildland Fire 18, 96–115. https://doi.org/10.1071/WF07162

Moody, J.A., Martin, D.A., 2009b. Synthesis of sediment yields after wildland fire in different rainfall regimes in the western United States. Int. J. Wildland Fire 18, 96–115. https://doi.org/10.1071/WF07162

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

Moreno, J.M., Oechel, W.C., 1989. A simple method for estimating flre intensity after a burn in California chaparral. Acta Oecol. 13.

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

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

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

Neary, D.G., Ryan, K.C., DeBano, L.F., 2005. Wildland fire in ecosystems: effects of fire on soils and water (No. RMRS-GTR-42-V4). U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Ft. Collins, CO. https://doi.org/10.2737/RMRS-GTR-42-V4

Nunes, J.P., Aparício, B., Carvalho, M., Brito, C., Jahanianfard, D., Benali, A., Parente, J., Nitzsche, N., Faria, B., Baartman, J., Dias, L.F., 2022. Assessing and predicting the impacts of wildfires on water quality in Portugal: Project FRISCO (No. EGU22-10710). Presented at the EGU22, Copernicus Meetings. https://doi.org/10.5194/egusphere-egu22-10710

Nyman, P., Sheridan, G.J., Smith, H.G., Lane, P.N.J., 2014. Modeling the effects of surface storage, macropore flow and water repellency on infiltration after wildfire. Journal of Hydrology 513, 301–313. https://doi.org/10.1016/j.jhydrol.2014.02.044

O'Callaghan, J.F., Mark, D.M., 1984. The extraction of drainage networks from digital elevation data. Computer Vision, Graphics, and Image Processing 28, 323–344. https://doi.org/10.1016/S0734-189X(84)80011-0

O'Dell, J.W., 1993. Method 180.1: Determination of turbidity by nephelometry. Environmental Monitoring Systems Laboratory Office of Research and Development, US Environmental Protection Agency, Cincinnati, OH 8.

Park, R.M., 2010. Thermocouple fundamentals. Course Tech., Temp 2–1.

Parson, A., Robichaud, P.R., Lewis, S.A., Napper, C., Clark, J.T., 2010. Field guide for mapping post-fire soil burn severity. Gen. Tech. Rep. RMRS-GTR-243. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 49 p. 243. https://doi.org/10.2737/RMRS-GTR-243

Pennino, M.J., Leibowitz, S.G., Compton, J.E., Beyene, M.T., LeDuc, S.D., 2022. Wildfires can increase regulated nitrate, arsenic, and disinfection byproduct violations and concentrations in public drinking water supplies. Science of The Total Environment 804, 149890. https://doi.org/10.1016/j.scitotenv.2021.149890

Potter, B.B., Wimsatt, J.C., 2012. USEPA method 415.3: Quantifying TOC, DOC, and SUVA. Journal-American Water Works Association 104, E358–E369.

Precipitation Frequency Data Server [WWW Document], 2017. . NOAA's National Weather Service Hydrometeorological Design Studies Center. URL https://hdsc.nws.noaa.gov/hdsc/pfds/ (accessed 12.14.20).

R Core Team, 2022. R: A language and environment for statistical computing.

Rainfall Simulator: How to Properly Collect and Store Large Soil Samples, 2019.

Reardon, J., Hungerford, R., Ryan, K., 2007. Factors affecting sustained smouldering in organic soils from pocosin and pond pine woodland wetlands. Int. J. Wildland Fire 16, 107. https://doi.org/10.1071/WF06005

Rengers, F.K., McGuire, L.A., Kean, J.W., Staley, D.M., Hobley, D.E.J., 2016. Model simulations of flood and debris flow timing in steep catchments after wildfire. Water Resources Research 52, 6041–6061. https://doi.org/10.1002/2015WR018176

Rhea, A.E., Covino, T.P., Rhoades, C.C., Brooks, A.C., 2022. Use of geostatistical models to evaluate landscape and stream network controls on post-fire stream nitrate concentrations. Hydrological Processes 36, e14689. https://doi.org/10.1002/hyp.14689

Rhoades, C.C., Chow, A.T., Covino, T.P., Fegel, T.S., Pierson, D.N., Rhea, A.E., 2019a. The Legacy of a Severe Wildfire on Stream Nitrogen and Carbon in Headwater Catchments. Ecosystems 22, 643–657. https://doi.org/10.1007/s10021-018-0293-6

Rhoades, C.C., Hubbard, R.M., Elder, K., 2017. A Decade of Streamwater Nitrogen and Forest Dynamics after a Mountain Pine Beetle Outbreak at the Fraser Experimental Forest, Colorado. Ecosystems 20, 380–392. https://doi.org/10.1007/s10021-016-0027-6

Rhoades, C.C., Nunes, J.P., Silins, U., Doerr, S.H., Rhoades, C.C., Nunes, J.P., Silins, U., Doerr, S.H., 2019b. The influence of wildfire on water quality and watershed processes: new insights and remaining challenges. Int. J. Wildland Fire 28, 721–725. https://doi.org/10.1071/WFv28n10_FO

Robichaud, P.R., 2005a. Measurement of post-fire hillslope erosion to evaluate and model rehabilitation treatment effectiveness and recovery. Int. J. Wildland Fire 14, 475–485. https://doi.org/10.1071/WF05031

Robichaud, P.R., 2005b. Measurement of post-fire hillslope erosion to evaluate and model rehabilitation treatment effectiveness and recovery. Int. J. Wildland Fire 14, 475–485. https://doi.org/10.1071/WF05031

Robichaud, P.R., Hungerford, R.D., 2000. Water repellency by laboratory burning of four northern Rocky Mountain forest soils. Journal of Hydrology 231–232, 207–219. https://doi.org/10.1016/S0022-1694(00)00195-5

Robichaud, P.R., Wagenbrenner, J.W., Pierson, F.B., Spaeth, K.E., Ashmun, L.E., Moffet, C.A., 2016. Infiltration and interrill erosion rates after a wildfire in western Montana, USA. CATENA 142, 77–88. https://doi.org/10.1016/j.catena.2016.01.027

Rosso, R., Rulli, M.C., Bocchiola, D., 2007. Transient catchment hydrology after wildfires in a Mediterranean basin: runoff, sediment and woody debris. Hydrology and Earth System Sciences Discussions 11, 125–140.

Roundy, B.A., Blackburn, W.H., Eckert, R.E., 1978. Influence of Prescribed Burning on Infiltration and Sediment Production in the Pinyon-Juniper Woodland, Nevada. Journal of Range Management 31, 250–253.

Rust, A.J., Hogue, T.S., Saxe, S., McCray, J., Rust, A.J., Hogue, T.S., Saxe, S., McCray, J., 2018. Post-fire water-quality response in the western United States. Int. J. Wildland Fire 27, 203–216. https://doi.org/10.1071/WF17115

Rust, A.J., Saxe, S., McCray, J., Rhoades, C.C., Hogue, T.S., 2019. Evaluating the factors responsible for post-fire water quality response in forests of the western USA. Int. J. Wildland Fire 28, 769. https://doi.org/10.1071/WF18191

Salavati, B., Oudin, L., Furusho-Percot, C., Ribstein, P., 2016. Modeling approaches to detect land-use changes: Urbanization analyzed on a set of 43 US catchments. Journal of Hydrology 538, 138–151. https://doi.org/10.1016/j.jhydrol.2016.04.010

Shahlaee, A.K., Nutter, W.L., Burroughs, E.R., Morris, L.A., 1991. Runoff and Sediment Production from Burned Forest Sites in the Georgia Piedmont1. JAWRA Journal of the American Water Resources Association 27, 485–493. https://doi.org/10.1111/j.1752-1688.1991.tb01449.x

Shakesby, R.A., Doerr, S.H., 2006. Wildfire as a hydrological and geomorphological agent. Earth-Science Reviews 74, 269–307. https://doi.org/10.1016/j.earscirev.2005.10.006

Shimadzu Corporation, 2001. TOC-Vcsh/csn Total Organic Carbon User Manual.

Smith, H.G., Sheridan, G.J., Lane, P.N.J., Nyman, P., Haydon, S., 2011. Wildfire effects on water quality in forest catchments: A review with implications for water supply. Journal of Hydrology 396, 170–192. https://doi.org/10.1016/j.jhydrol.2010.10.043

Staley, D.M., Kean, J.W., 2020. Emergency Assessment of Post-Fire Debris-Flow Hazards [WWW Document]. USGS Landslide Hazards Program Williams Fork (Arapaho and Roosevelt National Forests, CO). URL https://landslides.usgs.gov/hazards/postfire_debrisflow/detail.php?objectid=298 (accessed 3.12.23).

Stoof, C.R., De Kort, A., Bishop, T.F.A., Moore, D., Wesseling, J.G., Ritsema, C.J., 2011. How Rock Fragments and Moisture Affect Soil Temperatures during Fire. Soil Science Society of America Journal 75, 1133–1143. https://doi.org/10.2136/sssaj2010.0322

Stoof, C.R., Wesseling, J.G., Ritsema, C.J., 2010. Effects of fire and ash on soil water retention. Geoderma 159, 276–285. https://doi.org/10.1016/j.geoderma.2010.08.002

Surfleet, C.G., Dietterick, B., Skaugset, A., 2014. Change detection of storm runoff and sediment yield using hydrologic models following wildfire in a coastal redwood forest, California. Can. J. For. Res. 44, 572–581. https://doi.org/10.1139/cjfr-2013-0328

Tossell, R., Dickinson, W., Rudra, R., Wall, G., 1987. A portable rainfall simulator. Can. Agric. Eng 29, 155–162.

Ulbrich, C.W., 1983. Natural Variations in the Analytical Form of the Raindrop Size Distribution. Journal of Applied Meteorology and Climatology 22, 1764–1775. https://doi.org/10.1175/1520-0450(1983)022<1764:NVITAF>2.0.CO;2

Uzun, H., Dahlgren, R.A., Olivares, C., Erdem, C.U., Karanfil, T., Chow, A.T., 2020. Two years of post-wildfire impacts on dissolved organic matter, nitrogen, and precursors of disinfection by-products in California stream waters. Water Research 181, 115891. https://doi.org/10.1016/j.watres.2020.115891

Wagener, T., Gupta, H.V., 2005. Model identification for hydrological forecasting under uncertainty. Stoch Environ Res Ris Assess 19, 378–387. https://doi.org/10.1007/s00477-005-0006-5

Wang, J.-J., Dahlgren, R.A., Chow, A.T., 2015a. Controlled Burning of Forest Detritus Altering Spectroscopic Characteristics and Chlorine Reactivity of Dissolved Organic Matter: Effects of Temperature and Oxygen Availability. Environ. Sci. Technol. 49, 14019–14027. https://doi.org/10.1021/acs.est.5b03961

Wang, J.-J., Dahlgren, R.A., Erşan, M.S., Karanfil, T., Chow, A.T., 2015b. Wildfire Altering Terrestrial Precursors of Disinfection Byproducts in Forest Detritus. Environ. Sci. Technol. 49, 5921–5929. https://doi.org/10.1021/es505836m

Wang, J.-J., Dahlgren, R.A., Erşan, M.S., Karanfil, T., Chow, A.T., 2015c. Wildfire Altering Terrestrial Precursors of Disinfection Byproducts in Forest Detritus. Environ. Sci. Technol. 49, 5921–5929. https://doi.org/10.1021/es505836m

Wang, L., Liu, H., 2006. An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. International Journal of Geographical Information Science 20, 193–213. https://doi.org/10.1080/13658810500433453

Water quality tested in Williams Fork burn area [WWW Document], 2021. . KUSA.com. URL https://www.9news.com/article/weather/weather-colorado/water-quality-williams-fork-burn-area/73-4c344625-ef06-49c1-a23a-9274c0828ea0 (accessed 11.14.21).

Wei, H., Thompson, R., Park, C., Chen, P., 2010. Surface tension of high density polyethylene (HDPE) in supercritical nitrogen: Effect of polymer crystallization. Colloids and Surfaces A: Physicochemical and Engineering Aspects 354, 347–352.

Wieting, C., Ebel, B.A., Singha, K., 2017. Quantifying the effects of wildfire on changes in soil properties by surface burning of soils from the Boulder Creek Critical Zone Observatory. Journal of Hydrology: Regional Studies 13, 43–57. https://doi.org/10.1016/j.ejrh.2017.07.006

Williams, A.P., Livneh, B., McKinnon, K.A., Hansen, W.D., Mankin, J.S., Cook, B.I., Smerdon, J.E., Varuolo-Clarke, A.M., Bjarke, N.R., Juang, C.S., Lettenmaier, D.P., 2022. Growing impact of wildfire on western US water supply. Proceedings of the National Academy of Sciences 119, e2114069119. https://doi.org/10.1073/pnas.2114069119

Writer, J.H., Hohner, A., Oropeza, J., Schmidt, A., Cawley, K.M., Rosario-Ortiz, F.L., 2014. Water treatment implications after the High Park Wildfire, Colorado. Journal - AWWA 106, E189–E199. https://doi.org/10.5942/jawwa.2014.106.0055

Wu, Q., Brown, A., 2023. whitebox: "WhiteboxTools" R Frontend.

Xie, J., Liu, X., Bai, P., Liu, C., 2022. Rapid Watershed Delineation Using an Automatic Outlet Relocation Algorithm. Water Resources Research 58, e2021WR031129. https://doi.org/10.1029/2021WR031129

Yonter, G., Houndonougbo, H.M., 2022. Comparison of different fulljet nozzles used in laboratory type rain simulator in terms of some rainfall characteristics. Ege Üniversitesi Ziraat Fakültesi Dergisi 59, 33–41. https://doi.org/10.20289/zfdergi.865324

Yu, M., Bishop, T.F.A., Van Ogtrop, F.F., 2019. Assessment of the Decadal Impact of Wildfire on Water Quality in Forested Catchments. Water 11, 533. https://doi.org/10.3390/w11030533

Zema, D.A., Lucas-Borja, M.E., Fotia, L., Rosaci, D., Sarnè, G.M.L., Zimbone, S.M., 2020. Predicting the hydrological response of a forest after wildfire and soil treatments using an Artificial Neural Network. Computers and Electronics in Agriculture 170, 105280. https://doi.org/10.1016/j.compag.2020.105280

Appendix A

Summary of Simulation Technique Comparisons

A.1 Wildfire Simulation Techniques

Table A.1: Pros and cons of wildfire simulation techniques covered in the review, as well as the studies referenced and their scales. "WP1", "WP2", etc. represents "Wildfire Simulation Pro 1", "Wildfire Simulation Pro 2", etc. Similarly, "WC1", "WC2", etc. represents "Wildfire Simulation Con 1", "Wildfire Simulation Con 2", etc.

Simulation Technique	Pros	Cons	Scale	References
Prescribed Fire or Slash Burn	WP1-Heterogeneous combustion similar to natural wildfires WP2-Similar intensity and duration as a natural wildfire	WC1-Qualitative wildfire characterizations WC2-Tradeoff between intensity and size of burn	Plot	Emmerich and Cox, 1992; Hester et al., 1997; Marcos et al. 2000; Roundy et al., 1978; Santin et al., 2013
Propane Torch or Heat Lamp	WP3-Control over burn intensity and spatial distribution WP4-Low variability in heating WP5-Allowance for measurement of heating profile WP6-Control over duration of heating	WC3-Uniformity in spatial distribution of heating	Laboratory	Badía-Villas et al., 2014; Cancelo-Gonzalez, et al., 2012; Cancelo-Gonzalez, et al., 2013; Cancelo-Gonzalez, et al., 2015; Klopatek et al., 1988; Robichaud and Hungerford, 2000; Stoof et al., 2011; Wieting et al., 2017
Litter Burns	WP1-Heterogeneous combustion similar to natural wildfires WP3-Control over burn intensity and spatial distribution WP5-Allowance for measurement of heating profile	WC4-Limited to low-intensity burns	Laboratory Laboratory/Plot	Busse, et al., 2010; Keesstra, et al., 2014 Kral et al.
Muffle Furnaces	WP4-Low variability in heating	WC3-Uniformity in spatial distribution of heating	Laboratory	Blank et al., 1994; Cawley et al., 2017; Debano and

	WP5-Allowance for measurement of heating profile WP6-Control over duration of heating WP7-Incremental control of burn intensity	WC5-Heating occurs from all sides	Krammes, 1966; Hohner et al., 2019

A.2 Rainfall Simulation Techniques

Table A.2: Pros and cons of rainfall simulation techniques covered in the review, as well as the studies referenced and their scales. "RP1", "RP2", etc. represents "Rainfall Simulation Pro 1", "Rainfall Simulation Pro 2", etc. Similarly, "RC1", "RC2", etc. represents "Rainfall Simulation Con 1", "Rainfall Simulation Con 2", etc.

Simulation Technique	Pros	Cons	Scale	References
Fixed Nozzle-Based Rainfall Simulators	RP1-Simplicity in design RP2-Transportability and adaptability to steep terrains RP3-Intensities and droplet sizes similar to natural rainfall	RC1-Small area of coverage RC2-Rain kinetic energies lower than natural rain	Laboratory	Cancelo-Gonzalez, et al., 2012; Cancelo-Gonzalez, et al., 2013; Cancelo-Gonzalez, et al., 2015; Kibet, et al., 2014
			Plot	Cerda et al., 1997; Ferreira et al., 2005; Holland, 1969; Marcos et al., 2000; Rosso et al., 2007; Wilcox et al., 1986; Wilson, 1999
Dynamic Nozzle-Based Rainfall Simulators	RP3-Intensities and droplet sizes similar to natural rainfall RP4-Large area of coverage RP5-Variability in droplet dist. similar to natural rainfall	RC3-Complexity and expense of design RC4-Difficulty in transportation	Laboratory	Keesstra et al., 2014;
			Plot	Benavides-Solorio and MacDonald, 2001; Emmerich and Cox, 1992; Johansen, et al., 2001; Robichaud et al., 2016; Simanton et al., 1986; Swanson, 1965; Woods and Balfour, 2008; Betrand, 1961
Drip-Style Rainfall Simulators	RP3-Intensities and droplet sizes similar to natural rainfall RP6-Increased precision in droplet size	RC3-Complexity and expense of design RC4-Difficulty in transportation	Laboratory	Chevone et al., 1984
			Plot	Blackburn et al., 1974; Blake et al., 2010; Chevone et al., 1984; Hester et al., 1997; Knight et al., 1983; Roundy et al., 1978
WDPT Tests or Leaching	RP1-Simplicity in design RP7-Direct measurement of water repellency and quality	RC1-Small area of coverage RC5-No rainfall impact on soil surface	Laboratory	Badía-Villas et al., 2014; Blank et al., 1994; Cawley et al., 2017; Debano and Krammes, 1966; Hogue and Inglett, 2012; Hohner et al., 2019; Robichaud and Hungerford, 2000; Wang et al., 2015; Wieting et al., 2017

Appendix B

Compilation of Simulation Experiment Information

The studies included in this review have a complex set of study design factors and wide range of foci on different burn responses. However, for the purpose of cross-study analyses, we created simplified tables reporting five key hydrologic and water quality responses. These topically focused tables allow for useful cross-comparisons between simulation studies, in terms of the variability in measurements, the magnitude of responses for increasing burn intensity, and the level of statistical significance in results. Key responses synthesized include runoff, infiltration, sediment production, water repellency, and solutes/soil chemical properties—i.e., metals, nutrients, and organic matter in leachates and runoff, as well as in burned material. Results are divided into approximate burn intensities (unburned, mild, moderate, and severe), with peak temperature approximations selected by compiling temperature-based burn intensity scales in previous studies(Chandler et al., 1983; Hohner et al., 2019; Jian et al., 2018; Robichaud and Hungerford, 2000; Wang et al., 2015b; Wieting et al., 2017). A summary column provides an approximate synthesis of each study's results, with estimated or reported error and variability displayed in the adjacent column. These summaries are reported with respect to

unburned conditions unless otherwise stated, e.g., a *percent decrease* indicates the percent decrease of a response after burning from the control or unburned case in the study. Results were limited to those showing response exclusively to burning or fire exposure in various environments, while results from other joint treatments such as logging were excluded.

Table B.1: Runoff data reported from simulation studies. Time elapsed since the burn event (simulated or natural wildfire) and precipitation information are shown for each study. ± indicates the upper and lower bounds of the 95% confidence interval and parentheses indicate standard deviations. An *N/A* entry denotes an unreported study characteristic.

Time elapsed post-burn	Precipitation intensity and duration	Runoff				Summary	Error and Variability	Source
		Unburned	Mild Burn (~100-200 °C)	Moderate Burn (~200-350 °C)	Severe Burn (~<350 °C)			
0 years	66-94 mm/hr (mean: 79 mm/hr) for 60 mins	Runoff ratio: 55%		Runoff ratio: 58%	Runoff ratio: 66%	Runoff ratios increased 5% for moderate burns and 20% for severe burns	N/A	Benavides-Solorio and MacDonald, 2001
1 year		Runoff ratio: 61%		Runoff ratio: 48%	Runoff ratio: 61%	Runoff ratios decreased 20% for moderate burns and 0% for severe burns		
6 years		Runoff ratio: 59%		N/A	Runoff ratio: 36%	Runoff ratios decreased 39% for severe burns		
Same day	55 mm/hr for 45 mins	Runoff ratio: 5 (8)%	Runoff ratios: 6 (10)%	N/A	N/A	Runoff ratio increased 19% for mild burns	Standard deviations ranged from 114% to 179% of the means	Emmerich and Cox, 1992
	110 mm/hr for 15 mins (subsequent)	Runoff ratio: 14 (16)%	Runoff ratio: 17 (22)%			Runoff ratio increased 23% for mild burns		
	Total: average 151.5 mm/hr for 60 mins	Runoff ratio: 8 (11)%	Runoff ratio: 10 (14)%			Runoff ratio increased 21% for mild burns		
N/A	Microplots: 50.5 mm/hr for 45-60 mins Plots and catchments: variable natural precipitation	Runoff ratios: 5.5% and <1% for microplots, 0.09% and 0.05% for plots (two replicates)	Runoff ratios: 14.4% for microplots, 0.08% for plots (prescribed fire)	Runoff ratios: 23.2% for microplots (experimental fire)	Runoff ratios: 65.1% and 54.9% for microplots (two replicates), 11.6% for plots (natural wildfire)	Runoff ratios increased 343% and 14% for mild burns (microplots and plots, respectively), 614% for moderate burns (microplots), and 1746% and 16,471% for severe burns (microplots and plots, respectively)	N/A	Ferreira et al., 2005
80 days	60 mm/hr for 60 mins	Runoff ratios: 19% and 21% (two replicates)	N/A	N/A	Runoff ratios: 25% and 58% (two replicates)	Runoff ratios on average increased 104% for severe burns	Standard deviations ranged from 0% to 56% of the means	Johansen, et al., 2001
	60 mm/hr for 30 mins (subsequent)	Runoff ratios: 20% and 18% (two replicates)			Runoff ratios: 35% and 58% (two replicates)	Runoff ratios on average increased 149% for severe burns		
	60 mm/hr for 30 mins (subsequent)	Runoff ratios: 31% and 31% (two replicates)			Runoff ratios: 32% and 64% (two replicates)	Runoff ratios on average increased 55% for severe burns		
Same day	33 mm/hr for 4 hrs	Runoff rates: 25-28 and 0.2-0.8 mm/hr peak (bare and litter-covered soils, respectively)	N/A	Runoff rates: 20-23 and 12-14 mm/hr peak (soils without and with ash, respectively)	N/A	Peak runoff rates decreased 2% and increased 2500% for moderate burns (samples without and with litter/ash, respectively)	N/A	Keesstra, et al., 2014
	33 mm/hr for 2 hrs (subsequent)	Runoff rates: 23 and 0.2-0.6 mm/hr peak (bare and litter-covered soils, respectively)		Runoff rates: 20-25 mm/hr steady state for all samples		Peak runoff rates decreased 19% and increased 5525% for moderate burns (samples without and with litter/ash, respectively)		
0 months	180 mm/hr for 5 mins	Runoff ratios: 9.1% total, and 8.7%, 14.7%, 1.4%, 18.2%, and 10.5%*	N/A	N/A	Runoff ratios: 34.0% total, and 8.7%, 17.9%, 1.7%, 15.2%, and 9.4%*	Runoff ratios increased 275% for the entire plot and on average increased 3% (range: -16% to 22%) for individual subplots for severe burns	N/A	Marcos et al., 2000

18 months		N/A			Runoff ratios: 1.6%, 12.8%, 31.2%, 24.5%, and 1.1%*	Runoff ratios on average increased 396% (range: -89% to 2129%) for individual subplots for severe burns		
0 years	100 mm/hr for 60 mins	Runoff ratio: 58 (3.5)%	N/A	N/A	Runoff ratio: 70 (1.1)%	Runoff ratio increased 21% for severe burns	Standard deviations ranged from 2% to 11% of the means	Robichaud et al., 2016
1 year		Runoff ratio: 56 (3.9)%			Runoff ratio: 61 (1.4)%	Runoff ratio increased 9% for severe burns		
2 years		Runoff ratio: 53 (3.6)%			Runoff ratio: 62 (1.3)%	Runoff ratio increased 17% for severe burns		
5 years		Runoff ratio: 53 (3.4)%			Runoff ratio: 14 (1.5)%	Runoff ratio decreased 74% for severe burns		
6 weeks	Unburned: 58 mm/hr for 42 mins Burned: 74 mm/hr for 60 mins	Runoff ratio: 0% Peak runoff: 0 mm/hr	N/A	N/A	Runoff ratio: 21% Peak runoff: 33.5 mm/hr	Runoff ratio and peak runoff increased from no response, to 21% and 33.5 mm/hr, respectively, for severe burns	N/A	Rosso et al., 2007
	Unburned: 74 mm/hr for 61 mins Burned: 68 mm/hr for 55 mins (several days later)	Runoff ratio: 0.4% Peak runoff: 0.7 mm/hr			Runoff ratio: 24% Peak runoff: 25.3 mm/hr	Runoff ratio and peak runoff increased 5900% and 3514%, respectively, for severe burns		
	Unburned: 74 mm/hr for 71 mins Burned: 76 mm/hr for 60 mins (2 hrs later)	Runoff ratio: 2% Peak runoff: 2.2 mm/hr			Runoff ratio: 41% Peak runoff: 41.3 mm/hr	Runoff ratio and peak runoff increased 1950% and 1777%, respectively, for severe burns		
0 years	Unburned: 56.1, 57.3, 43.7, 44.6 mm/hr for 60 mins Severe burn: 33.5, 33.9 mm/hr for 60 mins	Runoff ratios: 0.0%, 0.3%, 0.0%, and 0.4% (four replicates)	N/A	N/A	Runoff ratios: 0.3% and 2.9% (two replicates)	Runoff ratio increased 715% for severe burns	Standard deviations ranged from 1.6% to 117% of the means	Simanton et al., 1986
	Unburned: 48.0, 46.8, 46.4, 49.2 mm/hr for 30 mins Severe burn: 56.0, 51.4 mm/hr for 30 mins (24 hrs later)	Runoff ratios: 1.7%, 0.9%, 0.4%, and 3.3% (four replicates)			Runoff ratios: 3.6% and 7.8% (two replicates)	Runoff ratio increased 266% for severe burns		
	Unburned: 69.6, 52.2, 91.4, 54.8 mm/hr for 30 mins Severe burn: 66.4, 71.6 mm/hr for 30 mins (subsequent)	Runoff ratios: 5.3%, 2.7%, 0.2%, and 4.7% (four replicates)			Runoff ratios: 11.7% and 14.8% (two replicates)	Runoff ratio increased 314% for severe burns		
1/2 year	41.4, 45.8 mm/hr for 60 mins	N/A	N/A	N/A	Runoff ratios: 0.2% and 1.3% (two replicates)	Runoff ratio increased 289% for severe burns		
	53.8, 55.0 mm/hr for 30 mins (24 hrs later)				Runoff ratios: 7.4% and 7.3% (two replicates)	Runoff ratio increased 374% for severe burns		
	59.2, 59.4 mm/hr for 30 mins (subsequent)				Runoff ratios: 21.3% and 17.8% (two replicates)	Runoff ratio increased 511% for severe burns		

1 year	42.9, 44.2 mm/hr for 60 mins	N/A	Runoff ratios: 1.2% and 0.2% (two replicates)	N/A	N/A	Runoff ratio increased 249% for mild burns		
	48.4, 51.4 mm/hr for 30 mins (24 hrs later)		Runoff ratios: 1.7% and 0.8% (two replicates)			Runoff ratio decreased 22% for mild burns		
	51.8, 51,4 mm/hr for 30 mins (subsequent)		Runoff ratios: 6.6% and 2.3% (two replicates)			Runoff ratio increased 39% for mild burns		
6 months	Unburned: 47 mm/hr for 44 mins Burned: 35, 70 mm/hr for 35, 21 mins, respectively	Peak runoff: 2 mm/hr	N/A	N/A	Peak runoff: 1 and 4 mm/hr (two replicates)	Peak runoff rates increased 25% for severe burns	Standard deviations ranged from 13% to 85% of the means	Wilson, 1999
	Unburned: 75 mm/hr for 21 mins Burned: 38, 68 mm/hr for 49, 21 mins, respectively	Peak runoff: 8 mm/hr			Peak runoff: 1 and 4 mm/hr (two replicates)	Peak runoff rates decreased 69% for severe burns		
	Unburned: 162 mm/hr for 10 mins Burned: 143, 110 mm/hr for 10, 14 mins, respectively	Peak runoff: 45 mm/hr			Peak runoff: 43 and 36 mm/hr (two replicates)	Peak runoff rates decreased 12% for severe burns		
1 month	75 ± 0.4 mm/hr for 1 hr	N/A	N/A	N/A	Final runoff: 23 and 42 mm/hr Runoff ratio: 15.8% and 44.2% (ash and no-ash plots, respectively)	Final runoff rates and runoff ratios increased 83% and 180%, respectively, when ash was removed for severe burns	N/A	Woods and Balfour, 2008
9 months	76 ± 0.4 mm/hr for 1 hr				Final runoff: 25 and 23 mm/hr Runoff ratio: 27.5% and 21.4% (ash and no-ash plots, respectively)	Final runoff rates and runoff ratios decreased 8% and 22%, respectively, when ash was removed for severe burns		
12 months	75 ± 0.4 mm/hr for 1 hr				Final runoff: 46 and 45 mm/hr Runoff ratio: 36.5% and 36.5% (ash and no-ash plots, respectively)	Final runoff rates and runoff ratios decreased 2% and 0%, respectively, when ash was removed for severe burns		

*for plots sown with herb species; herb and shrub species; herb and oak species; herb, shrub, and oak species; and unsown, respectively

Table B.2: Infiltration data reported from simulation studies. Time elapsed since the burn event (simulated or natural wildfire) and precipitation information are shown for each study. ± indicates the upper and lower bounds of the 95% confidence interval and parentheses indicate standard deviations. An *N/A* entry denotes an unreported study characteristic.

Time elapsed post-burn	Precipitation intensity and duration	Infiltration				Summary	Error and Variability	Source
		Unburned	Mild Burn (~100-200 °C)	Moderate Burn (~200-350 °C)	Severe Burn (~<350 °C)			
3 months	20 for 15 mins	Saturated hydraulic conductivity: 505 ± 24, 812 ± 29, and 1114 ± 97 mm/hr (three replicates)	N/A	Saturated hydraulic conductivity: 321 ± 104, 510 ± 96, and 263 ± 70 mm/hr (three replicates)	Saturated hydraulic conductivity: 62 ± 30, 69 ± 11, and 289 ± 35 mm/hr (three replicates)	Saturated hydraulic conductivity decreased on average 50% for moderate burn and 84% for severe burn areas	95% confidence intervals ranged 4%-48% of the means	Blake et al., 2001
	40 mm/hr for 15 mins							
	60 mm/hr for 15 mins							
5 months	203 mm/hr for 50 mins	Infiltration rates*: 202, 183, 146, and 105 mm/hr	Infiltration rates*: 129, 162, 110, and 76 mm/hr	N/A	N/A	Infiltration rates on average decreased 25% (range: 11%-36%) for mild burn areas	Standard deviations ranged 27% to 30% of the means	Hester et al., 1997
Same day	33 mm/hr for 4 hrs	Infiltration rates: 1.5 and 33 mm/hr (bare and litter-covered soils, respectively)	N/A	Infiltration rates: ~33 mm/hr peak, then decrease for all samples (ash was slightly higher than without ash)	N/A	Peak infiltration rates increased 2100% and decreased 0% for moderate burns (samples without and with litter/ash, respectively)	N/A	Keesstra, et al., 2014
	33 mm/hr for 2 hrs (subsequent)	Infiltration rates: 1.6 and 33 mm/hr (bare and litter-covered soils, respectively)		Infiltration rates: 2-7 mm/hr steady state for all samples		Peak infiltration rates increased 181% and decreased 86% for moderate burns (samples without and with litter/ash, respectively)		
10 months	203 mm/hr for 30 mins	Infiltration rates: 124 and 128 mm/hr (two replicates)	Infiltration rate: 128 mm/hr	N/A	N/A	Infiltration rates increased 1.6% for mild burns	Standard deviation was 2% of the mean	Knight et al., 1983
0 years	100 mm/hr for 60 mins	Infiltration rate: 44 (3.5) mm/hr	N/A	N/A	Infiltration rate: 31 (1.0) mm/hr	Infiltration rate decreased 30% for severe burns	Standard deviations ranged from 2% to 9% of the means	Robichaud et al., 2016
1 year		Infiltration rate: 45 (4.0) mm/hr			Infiltration rate: 38 (1.4) mm/hr	Infiltration rate decreased 16% for severe burns		
2 years		Infiltration rate: 48 (3.6) mm/hr			Infiltration rate: 37 (1.3) mm/hr	Infiltration rate decreased 23% for severe burns		
5 years		Infiltration rate: 47 (3.4) mm/hr			Infiltration rate: 84 (1.5) mm/hr	Infiltration rate increased 79% for severe burns		
1-2 months	83.8 mm/hr for 60 mins	Infiltration rates**: 83.2 and 82.8, 78.2 and 80.2, 80.7 and 81.5, and 57.6 and 41.2 mm/hr for dry conditions; 82.6 and 82.2, 72.5 and 74,2, 77.6 and 78.2, and 37.5 and 19.5 mm/hr for field capacity conditions (two replicates)	Infiltration rates**: 83.2 and 82.8, 78.2 and 80.2, 80.7 and 81.5, and 57.6 and 41.2 mm/hr for dry conditions; 82.6 and 82.2, 72.5 and 74,2, 77.6 and 78.2, and 37.5 and 19.5 for field capacity conditions (two replicates)	N/A	N/A	Infiltration rates on average decreased 2.3% (range: 0.6% to 3.8%) and 10.6% (range: -1.4% to 14.9%) (dry and field capacity moisture conditions, respectively) for mild burns	Standard deviations ranged from 0% to 30% of the means	Roundy et al., 1978

1 year		N/A	Infiltration rates**: 82.0 and 81.3, 72.3 and 75.5, 77.1 and 78.4, and 45.3 and 59.7 mm/hr for dry conditions; 78 and 75.8, 57.2 and 58.3, 67.6 and 67.1, and 22.5 and 38.5 mm/hr for field capacity conditions (two replicates)			Infiltration rates on average decreased 1.6% (range: -6.3% to 6.7%) and 8.2% (range: -7.0% to 21.3%) (dry and field capacity antecedent moisture contents, respectively) for mild burn areas		
Same day	23 mm/hr for 80 mins	Field saturated hydraulic conductivity: 3.7 x 10^{-4} mm/s	N/A	Field saturated hydraulic conductivity: 2.0 x 10^{-4} mm/s	Field saturated hydraulic conductivity: 14 x 10^{-4} mm/s	Field saturated hydraulic conductivity decreased 46% for moderate burns and increased 278% for severe burns	N/A	Wieting et al., 2017
6 months	Unburned: 47 mm/hr for 44 mins Burned: 35, 70 mm/hr for 35, 21 mins, respectively	Infiltration rate: 45 mm/hr	N/A	N/A	Infiltration rates: 37 and 66 mm/hr (two replicates)	Infiltration rate increased 14% for severe burns	Standard deviations ranged from 30% to 40% of the means	Wilson, 1999
	Unburned: 75 mm/hr for 21 mins Burned: 38, 68 mm/hr for 49, 21 mins, respectively	Infiltration rate: 67 mm/hr			Infiltration rates: 37 and 64 mm/hr (two replicates)	Infiltration rate decreased 25% for severe burns		
	Unburned: 162 mm/hr for 10 mins Burned: 143, 110 mm/hr for 10, 14 mins, respectively	Infiltration rate: 117 mm/hr			Infiltration rates: 114 and 74 mm/hr (two replicates)	Infiltration rate decreased 20% for severe burns		
1 month	75 ± 0.4 mm/hr for 1 hr	N/A	N/A	N/A	Infiltration rates: 62 and 42 mm/hr (ash and no-ash plots, respectively)	Infiltration rates decreased 32% when ash was removed for severe burns	N/A	Woods and Balfour, 2008
9 months	76 ± 0.4 mm/hr for 1 hr				Infiltration rates: 56 and 58 mm/hr (ash and no-ash plots, respectively)	Infiltration rates increased 4% when ash was removed for severe burns		
12 months	75 ± 0.4 mm/hr for 1 hr				Infiltration rates: 48 and 48 mm/hr (ash and no-ash plots, respectively)	Infiltration rates did not change when ash was removed for severe burns		

*oak, juniper, bunchgrass, and shortgrass sites, respectively

**for tree coppice, shrub coppice, all coppice, and interspace, respectively

Table B.3: Sedimentation and erosion data reported from simulation studies. Time elapsed since the burn event (simulated or natural wildfire) and precipitation information are shown for each study. ± indicates the upper and lower bounds of the 95% confidence interval and parentheses indicate standard deviations. An *N/A* entry denotes an unreported study characteristic.

Time elapsed post-burn	Precipitation intensity and duration	Sedimentation and Erosion				Summary	Error and Variability	Source
		Unburned	Mild Burn (~100-200 °C)	Moderate Burn (~200-350 °C)	Severe Burn (~<350 °C)			
0 years	66-94 mm/hr (mean: 79 mm/hr) for 60 mins	Total sediment: 43 g Sediment concentration: 1.9 g/L		Total sediment: 89 g Sediment concentration: 4.0 g/L	Total sediment: 428 g Sediment concentration: 23.5 g/L	Total and suspended sediment increased 107% and decreased 79%, respectively, for moderate burns and increased 895% and 1137%, respectively, for severe burns	N/A	Benavides-Solorio and MacDonald, 2001
1 year		Total sediment: 13 g Sediment concentration: 0.6 g/L		Total sediment: 76 g Sediment concentration: 3.5 g/L	Total sediment: 342 g Sediment concentration: 17.1 g/L	Total and suspended sediment increased 485% and 483%, respectively, for moderate burns and increased 2392% and 2750%, respectively, for severe burns		
6 years		Total sediment: 13 g Sediment concentration: 0.5 g/L		N/A	Total sediment: 27 g Sediment concentration: 1.7 g/L	Total and suspended sediment increased 108% and 240%, respectively, for severe burns		
3 months	20 for 15 mins	Sediment concentration: ~2.5 g/L	N/A	Sediment concentration: ~3 g/L	Sediment concentration: ~5 g/L	Suspended sediment concentration increased 20% for moderate burns and 100% for severe burns	N/A	Blake et al., 2001
	40 mm/hr for 15 mins	Sediment concentration: ~1 g/L		Sediment concentration: ~1.5 g/L	Sediment concentrations: ~2.5 g/L	Suspended sediment concentration increased 50% for moderate burns and 150% for severe burns		
	60 mm/hr for 15 mins	Sediment concentration: < 0.1 g/L		Sediment concentration: ~1 g/L	Sediment concentrations: ~2 g/L	Suspended sediment concentration increased 900% for moderate burns and 1900% for severe burns		
Same day	55 mm/hr for 45 mins	Sediment yield: 2.5 (4.7) g/m²	Sediment yield: 3.4 (6.7) g/m²	N/A	N/A	Sediment yield increased 36% for mild burns	Standard deviations ranged from 112% to 197% of the means	Emmerich and Cox, 1992
	110 mm/hr for 15 mins (subsequent)	Sediment yield: 5.0 (5.6) g/m²	Sediment yield: 7.2 (8.7) g/m²			Sediment yield increased 44% for mild burns		
	Total: average 151.5 mm/hr for 60 mins	Sediment yield: 7.6 (9.8) g/m²	Sediment yield: 10.6 (14.5) g/m²			Sediment yield increased 39% for mild burns		
5 months	203 mm/hr for 50 mins	Sediment yield: 0.2, 3.4, 30, and 130 g/m²*	Sediment yield: 450, 193, 446, and 577 g/m²*	N/A	N/A	Sediment yield on average increased 58,049% (range: 344%-224,900%) for mild burn areas	Standard deviations ranged from 39% to 149% of the means	Hester et al., 1997

80 days	60 mm/hr for 60 mins	Sediment yield per mm precip: 0.3 and 0.4 g/m² ·mm (two replicates)	N/A	N/A	Sediment yield per mm precip: 2.8 and 11 g/m² ·mm (two replicates)	Sediment yield per mm precip increased 1742% for severe burns	Standard deviations ranged from 0% to 84% of the means	Johansen, et al., 2001
	60 mm/hr for 30 mins (subsequent)	Sediment yield per mm precip: 0.2 and 0.3 g/m² ·mm (two replicates)			Sediment yield per mm precip: 4.8 and 11.3 g/m² ·mm (two replicates)	Sediment yield per mm precip increased 2983% for severe burns		
	60 mm/hr for 30 mins (subsequent)	Sediment yield per mm precip: 0.3 and 0.4 g/m² ·mm (two replicates)			Sediment yield per mm precip: 4.1 and 11.3 g/m² ·mm (two replicates)	Sediment yield per mm precip increased 1996% for severe burns		
Same day	33 mm/hr for 4 hrs	Sedimentation rates: 168 and 0 g/m² ·hr (bare and litter-covered soils, respectively)	N/A	Sedimentation rates: ~60-90 and ~30 g/m² ·hr (soils without and with ash, respectively)	N/A	Peak sedimentation rates decreased 55% for samples without litter/ash and samples with littler/ash increased from no response to 40% of the bare samples for moderates	N/A	Keesstra, et al., 2014
	33 mm/hr for 2 hrs (subsequent)	Sedimentation rates: 222-252 and 0 g/m² ·hr (bare and litter-covered soils, respectively)		Sedimentation rates: ~105-120 and ~105 g/m² ·hr (soils without and with ash, respectively)		Peak sedimentation rates decreased 53% for samples without litter/ash and samples with littler/ash increased from no response to 93% of the bare samples for moderate burns		
10 months	203 mm/hr for 30 mins	Sediment yield: 73 and 196 g/m² (two replicates)	Sediment yield: 130 g/m²	N/A	N/A	Sediment yield decreased 3% for mild burns	Standard deviation was 65% of the mean	Knight et al., 1983
0 months	180 mm/hr for 5 mins	Sediment yield: 1.18 g/m² total, and 1.66, 20.11, 5.16, 8.80, and 4.94 g/m²	N/A	N/A	Sediment yield: 10.32 g/m² total, and 17.06, 10.38, 5.11, (lost data), and 11.23 g/m²	Sediment yields increased 775% for the entire plot and on average increased 241% (range: -48% to 928%) for individual subplots for severe burns	N/A	Marcos et al., 2000
18 months		N/A			Sediment yield: 0.32, 17.99, 12.07, 14.23, and 4.49 g/m²	Sediment yields on average increased 19% (range: -81% to 134%) for individual subplots for severe burns		
0 years	100 mm/hr for 60 mins	Sediment yield: 50 (7.1) g/m²	N/A	N/A	Sediment yield: 1757 (114) g/m²	Sediment yield increased 3414% for severe burns	Standard deviations ranged from 5% to 41% of the means	Robichaud et al., 2016
1 year		Sediment yield: 15 (3.0) g/m²			Sediment yield: 1099 (60) g/m²	Sediment yield increased 7227% for severe burns		
2 years		Sediment yield: 7 (1.4) g/m²			Sediment yield: 391 (47) g/m²	Sediment yield increased 5486% for severe burns		
5 years		Sediment yield: 54 (22) g/m²			Sediment yield: 15 (3.6) g/m²	Sediment yield decreased 72% for severe burns		

6 weeks	Unburned: 58 mm/hr for 42 mins Burned: 74 mm/hr for 60 mins	Sedimentation rate: 0 $g/m^2 \cdot hr$	N/A	N/A	Sedimentation rate: 16.8 $g/m^2 \cdot hr$	Sedimentation rate increased from no response to 16.8 g/m^2h for severe burns	N/A	Rosso et al., 2007
	Unburned: 74 mm/hr for 61 mins Burned: 68 mm/hr for 55 mins (several days later)	Sedimentation rate: 0.1 $g/m^2 \cdot hr$			Sedimentation rate: 17.8 $g/m^2 \cdot hr$	Sedimentation rate increased 17700% for severe burns		
	Unburned: 74 mm/hr for 71 mins Burned: 76 mm/hr for 60 mins (2 hrs later)	Sedimentation rate: 0.2 $g/m^2 \cdot hr$			Sedimentation rate: 32.9 $g/m^2 \cdot hr$	Sedimentation rate increased 16350% for severe burns		
1-2 months	83.8 mm/hr for 60 mins	Sediment yields**: 49 and 105, 397 and 537, 140 and 240, and 932 and 1445 kg/ha for dry conditions; 50 and 98, 572 and 813, 157 and 282, and 850 and 1820 kg/ha for field capacity conditions (two replicates)	Sediment yields**: 181 and 389, 630 and 794, 338 and 550, and 666 and 1047 kg/ha for dry conditions; 194 and 603, 1046 and 1096, 456 and 813, and 870 and 2188 kg/ha for field capacity conditions (two replicates)	N/A	N/A	Sediment yields on average increased 107% (range: -28% to 270%) and 174% (range: 15% to 439%) (dry and field capacity antecedent moisture contents, respectively) for mild burns	N/A	Roundy et al., 1978
1 year		N/A	Sediment yields**: 55 and 457, 1229 and 1148, 262 and 724, and 1697 and 997 kg/ha for dry conditions; 190 and 741, 1861 and 1585, 602 and 1072, and 2753 and 1145 kg/ha for field capacity conditions (two replicates)			Sediment yields on average increased 140% (range: 13% to 232%) and 251% (range: 46% to 529%) (dry and field capacity antecedent moisture contents, respectively) for mild burns		

0 years	Unburned: 56.1, 57.3, 43.7, 44.6 mm/hr for 60 mins Severe burn: 33.5, 33.9 mm/hr for 60 mins	Sediment yield: 0.0, 1.2, 0.0, and 0.5 g/m² (four replicates)	N/A	N/A	Sediment yield: 0.0 and 7.3 g/m² (two replicates)	Sediment yield increased 798% for severe burns	Standard deviations ranged from 15% to 141%	Simanton et al., 1986
	Unburned: 48.0, 46.8, 46.4, 49.2 mm/hr for 30 mins Severe burn: 56.0, 51.4 mm/hr for 30 mins (24 hrs later)	Sediment yield: 1.1, 1.7, 0.1, and 2.0 g/m² (four replicates)			Sediment yield: 11 and 17.1 g/m² (two replicates)	Sediment yield increased 1044% for severe burns		
	Unburned: 69.6, 52.2, 91.4, 54.8 mm/hr for 30 mins Severe burn: 66.4, 71.6 mm/hr for 30 mins (subsequent)	Sediment yield: 7.9, 9.8, 0.5, and 4.8 g/m² (four replicates)			Sediment yield: 40.8 and 50.3 g/m² (two replicates)	Sediment yield increased 692% for severe burns		
1/2 year	41.4, 45.8 mm/hr for 60 mins	N/A	N/A		Sediment yield: 0.06 and 1.9 g/m² (two replicates)	Sediment yield increased 144% for severe burns		
	53.8, 55.0 mm/hr for 30 mins (24 hrs later)				Sediment yield: 18.8 and 13.6 g/m² (two replicates)	Sediment yield increased 1216% for severe burns		
	59.2, 59.4 mm/hr for 30 mins (subsequent)				Sediment yield: 65.1 and 28.1 g/m² (two replicates)	Sediment yield increased 711% for severe burns		
1 year	42.9, 44.2 mm/hr for 60 mins	N/A	Sediment yield: 0.5 and 0.05 g/m² (two replicates)		N/A	Sediment yield increased 534% for mild burns		
	48.4, 51.4 mm/hr for 30 mins (24 hrs later)		Sediment yield: 3.2 and 1.3 g/m² (two replicates)			Sediment yield increased 84% for mild burns		
	51.8, 51,4 mm/hr for 30 mins (subsequent)		Sediment yield: 18.9 and 8.3 g/m² (two replicates)			Sediment yield increased 136% for mild burns		

6 months	Unburned: 47 mm/hr for 44 mins Burned: 35, 70 mm/hr for 35, 21 mins, respectively	Sediment yield: 0.09 g/m²	N/A	N/A	Sediment yield: 0.03 and 1.5 g/m² (two replicates)	Sediment yield increased 730% for severe burns	Standard deviations ranged from 15% to 136% of the means	Wilson, 1999
	Unburned: 75 mm/hr for 21 mins Burned: 38, 68 mm/hr for 49, 21 mins, respectively	Sediment yield: 1.3 g/m²			Sediment yield: 0.03 and 1.2 g/m² (two replicates)	Sediment yield decreased 53% for severe burns		
	Unburned: 162 mm/hr for 10 mins Burned: 143, 110 mm/hr for 10, 14 mins, respectively	Sediment yield: 20 g/m²			Sediment yield: 81 and 100 g/m² (two replicates)	Sediment yield increased 353% for severe burns		
1 month	75 ± 0.4 mm/hr for 1 hr	N/A	N/A	N/A	Sediment yield: 137 ± 98 and 520 ± 225 g Sediment concentration: 22 ± 9 g/L and 35 ± 23 g/L (ash and no-ash plots, respectively)	Sediment yield and sediment concentration increased 280% and 59%, respectively, when ash was removed for severe burns	Standard deviations ranged from 41%-140% and 43%-100% of the means for sediment yield and sediment concentration, respectively	Woods and Balfour, 2008
9 months	76 ± 0.4 mm/hr for 1 hr				Sediment yield: 207 ± 247 and 58 ± 40 g Sediment concentration: 17 ± 12 g/L and 8.0 ± 6.0 g/L (ash and no-ash plots, respectively)	Sediment yield and sediment concentration decreased 72% and 53%, respectively, when ash was removed for severe burns		
12 months	75 ± 0.4 mm/hr for 1 hr				Sediment yield: 33 ± 33 and 16 ± 16 g Sediment concentration: 3.0 ± 4.2 g/L and 1.4 ± 1.2 g/L (ash and no-ash plots, respectively)	Sediment yield and sediment concentration decreased 52% and 53%, respectively, when ash was removed for severe burns		

*oak, juniper, bunchgrass, and shortgrass sites, respectively
**for tree coppice, shrub coppice, all coppice, and interspace, respectively

Table B.4: Water repellency data reported from simulation studies. ± indicates the upper and lower bounds of the 95% confidence interval and parentheses indicate standard deviations. An *N/A* entry denotes an unreported study characteristic, *O horizon* is *organic soil horizon*, *Ah horizon* is *soil mineral horizon*, and *WDPT* is *water drop penetration time*.

Water repellency				Summary	Error and Variability	Source
Unburned	**Mild Burn (~100-200 °C)**	**Moderate Burn (~200-350 °C)**	**Severe Burn (~<350 °C)**			
Ethanol %: O horizon: 36.0 ± 0; Ah horizon: 24.0 ± 0, 22.0 ± 3.5, and 20.0 ± 3.5*	N/A	Ethanol %: O horizon: 0 ± 0 and 1.7 ± 2.9 (ashes and O charred, respectively); Ah horizon: 6.2 ± 2.0, 18.3 ± 5.5, and 18.3 ± 5.5*	N/A	Ethanol % decreased 98% for the O horizon and 74%, 17%, and 9% for 3 depths in the Ah horizon*, respectively, for moderate burns	N/A	Badía-Villas et al., 2014
Max WDPT tests: ~110, 120, and 90 s (three replicates)		Max WDPT tests: ~100 and 100 s (three replicates)	Max WDPT tests: ~100, 120, and 0 s (three replicates)	Max WDPT tests decreased ~6% for moderate burns and ~31% for severe burns	Standard deviations ranged from 0% to 88% of the means	Benavides-Solorio and MacDonald, 2001
N/A	149 °C (300 °F) for 5, 10, 15, and 20 mins: no increase in non-wettability	204 °C (400 °F) for 5 and 10 mins: no increase in non-wettability, increase after 10 and 20 minutes; 260 °C (500 °F): impenetrability reached after 10 mins; > 316 °C (600 °F): impenetrability reached after 5 minutes	427 °C (800 °F) and 482 °C (900 °F): non-wettability started decreasing after 10 mins	Non-wettability: no increase for low burns, increase after 10-20 mins for mild-moderate burns; impenetrability reached in 5-10 mins for moderate burns; started decreasing after 10 mins for severe burns	N/A	Debano and Krammes, 1966
Ethanol %: ~0-6	Ethanol %: 4.3 (prescribed fire)	Ethanol %: 9.46 (experimental fire)	N/A	Ethanol % increased 43% for mild burns and 215% for moderate burns	N/A	Ferreira et al., 2005
N/A	N/A	WR test** before precip: 1.07 and 1.48 (4- and 15-mm aggregate, respectively), ash removed; 1.49 (15-mm aggregate) with ash After precip: 1.28 and 1.23 (4- and 15-mm aggregate, respectively), ash removed; 1.13 and 1.02 (4- and 15-mm aggregate, respectively), with ash	N/A	WDPT tests decreased by 1% for samples with removed ash before the precip and on average increased 17% (range: 13%-21%) for samples with removed ash after the precip	N/A	Keesstra, et al., 2014
WDPT test ranges***: 5 to 60, 5 to 60, > 60, and 5 to 60 s (dry conditions); 5 to 60, > 60, < 5, and <5 s (wet conditions)	WDPT test ranges***: (missing data), > 60, 5 to 60, and 5 to 60 s (dry conditions); 5 to 60, 5 to 60, 5 to 60, and 5 to 60 s (wet conditions)	WDPT test ranges***: < 5, 5 to 60, 5 to 60, and 5 to 60 s (dry conditions); < 5, (missing data), (missing data), and 5 to 60 s (wet conditions)	WDPT test ranges***: < 5, < 5, < 5, and 5 to 60 s (dry conditions); 5 to 60, < 5, 5 to 60, and 5 to 60 s (wet conditions)	WDPT tests increased ~70% and ~600% for mild burn, decreased ~40% and increased ~1000% for moderate burns, and decreased ~70% and increased ~600% for severe burns (dry and wet conditions, respectively)	N/A	Robichaud and Hungerford, 2000
0 yrs post-burn: Water repellency occurrence: 93 (3.9) %	N/A	N/A	0 yrs post-burn: Water repellency occurrence: 88 (2.3) %	0 yrs post-burn: Water repellency occurrence decreased 5% for severe burns	Standard deviations ranged from 3% to 22% of the means	Robichaud et al., 2016
1 yr post-burn: Water repellency occurrence: 39 (8.6) %			1 yr post-burn: Water repellency occurrence: 79 (3.5) %	1 yr post-burn: Water repellency occurrence increased 103% for severe burns		
2 yrs post-burn: Water repellency occurrence: 51 (8.8) %			2 yrs post-burn: Water repellency occurrence: 48 (4.4) %	2 yrs post-burn: Water repellency occurrence decreased 6% for severe burns		
5 yrs post-burn: Water repellency occurrence: 94 (4.1) %			5 yrs post-burn: Water repellency occurrence: 45 (5.2) %	5 yrs post-burn: Water repellency occurrence decreased 52% for severe burns		
WDPT test**** ranges: < 5 to 225, < 5 to 164, and < 5 to 64 s (three replicates)	N/A	WDPT test**** ranges: 440 to > 3600, 570 to > 3600, and 680 to > 1990 s (three replicates)	WDPT test**** ranges: < 5 to 75, < 5, and < 5 to 20 s (three replicates)	WDPT tests increased ~2500% for moderate burn and decreased ~75% for severe burns	N/A	Wieting et al., 2017

*0-1 cm, 1-2 cm, and 2-3 cm depths, respectively; **1 = hydrophilic (< 5 s), 2 = slight (6-10 s), 3-4 = moderate (11-60 s), 5-6 = strong (61-300 s), and 7-8 = severe (> 300 s); ***for ash-cap, mixed ash-cap, no ash-cap, and granitic soils, respectively; ****Repellency rating: ≤ 5 s = wettable, 5-80 = slight, 80-600 = strong, 600-3600 = severe, > 3600 = extreme

Table B.5: Water quality and soil chemical composition data reported from simulation studies. ± indicates the upper and lower bounds of the 95% confidence interval and parentheses indicate standard deviations. An *N/A* entry denotes an unreported study characteristic, *O horizon* is *organic soil horizon*, *Ah horizon* is *soil mineral horizon*, *TOC* is *total organic carbon*, *PyC* is *pyrolyzed carbon*, *DOC* is *dissolved organic carbon*, and *DON* is *dissolved organic nitrogen*.

Runoff Solutes and Soil Chemical Composition				Summary	Error and Variability	Source
Unburned	**Mild Burn (~100-200 °C)**	**Moderate Burn (~200-350 °C)**	**Severe Burn (~<350 °C)**			
TOC: O horizon: 20.4 ± 2.5%; Ah horizon: 16.1 ± 3.2%, 13.2 ± 2.1%, and 10.3 ± 1.6%* PyC: O horizon: 45.3 ± 6.9%; Ah horizon: 29.8 ± 6.6%, 28.5 ± 3.0%, and 26.7 ± 5.1%*	N/A	TOC: O horizon: 11.2 ± 3.7% and 10.4 ± 0.6% (ashes and O charred, respectively), Ah horizon: 11.4 ± 1.0%, 12.2 ± 3.1%, and 11.5 ± 1.1%* PyC: O horizon: 10.4 ± 3.4% and 13.6 ± 1.8% (ashes and O charred, respectively), Ah horizon: 18.5 ± 1.2%, 23.5 ± 0.9%, and 27.4 ± 6.2%*	N/A	TOC and PyC decreased 47% and 74%, respectively, for the O horizon and decreased 29% and 38%, decreased 8% and 18%, and increased 12% and 3% for 3 depths in the Ah horizon*, respectively, for moderate burns	N/A	Badía-Villas et al., 2014
Total P: 1112 ± 410, 709 ± 472, and 1111 ± 366 mg/kg (three different plots)	N/A	Total P: 1023 ± 322, 944 ± 468, and 1488 ± 411 mg/kg (three replicates)	Total P: 1676 ± 181, 1243 ± 462, and 1357 ± 447 mg/kg (three replicates)	Total P on average increased 20% for moderate burn and 49% for severe burn areas	Standard deviations ranged from 11% to 67% of the means	Blake et al., 2001
Leached cations: 2.2, 1.8, 21.7, and 14.2 kg/ha for Ca, Mg, K, and Na, respectively	N/A	Leached cations: 3.6 and 61.6, 6.7 and 12.1, 27.4 and 18.5, and 14.2 and 26.45 kg/ha for Ca, Mg, K, and Na, respectively (two replicates)	Leached cations: 36.5 and 61.4, 8.3 and 12.7, 28.4 and 39.9, and 13.4 and 25.4 kg/ha for Ca, Mg, K, and Na, respectively (two replicates)	Leached Ca, Mg, K, and Na on average increased 1382%, 422%, 6%, and 43%, respectively, for moderate burns and 2125%, 483%, 57%, and 37%, respectively, for severe burns	Standard deviations ranged from 24% to 126% of the means	Cancelo-Gonzalez et al., 2013
Leached metals: 2.52 and 3.84 mg (Al and Fe, respectively)		Leached metals: 3.01 and 3.55 mg (Al and Fe, respectively)	Leached metals: 3.62 and 6.37 mg (Al and Fe, respectively)	Leached Al and Fe increased 19% and decreased 8%, respectively, for moderate burns and increased 44% and 66%, respectively, for severe burns	N/A	Cancelo-Gonzalez et al., 2015
DOC: 9.35 and 18.4 mg_C/L Fe: 0.050 and 0.026 ppm (two replicates)	N/A	DOC: 24.4 and 26.7 (burned at 225 °C), 15.5 and 8.54 mg_C/L (burned at 350 °C) Fe: 0.058 and 0.100 (burned at 225 °C), 0.034 and 0.009 ppm (burned at 350 °C) (two replicates)	DOC: < 0.2 and < 0.2 mg_C/L Fe: 0.007 and 0.012 ppm (two replicates)	DOC and Fe increased 84% and 108%, respectively, for mild-moderate burns, decreased 13% and 43%, respectively, for moderate burns, and decreased 99% and 75%, respectively, for severe burns	N/A	Cawley et al., 2017
Total C: 410 ± 18 and 444 ± 8 g/kg Total P: 165 ± 10 and 95 ± 8 mg/kg Total N: 8.1 ± 2.1 and 5.3 ± 0.2 g/kg (low nutrient and high nutrient plant biomass, respectively)	Total C: 595 and 554 g/kg Total P: 88 and 107 mg/kg Total N: ~6 and 11 g/kg (low nutrient and high nutrient plant biomass, respectively) (muffled at 200 °C)	Total N: ~35 and 30 g/kg (muffled at 350 °C) (low nutrient and high nutrient plant biomass, respectively)	Total C: 29 and 66 (muffled at 550 °C), 199 and 181 g/kg (10 g lab burn at 460 °C) Total P: 3244 and 4454 (muffled at 550 °C), 1761 and 2090 mg/kg (10 g lab burn at 460 °C) Total N: 5.1 and 4.0 g/kg (10 g lab burn at 460 °C) (low nutrient and high nutrient plant biomass, respectively)	Total C, P, and N on average increased 35%, decreased 25%, and increased 27%, respectively, for mild burns; decreased 89%, increased 2851%, and decreased to no response, respectively, for (muffle furnace) severe burn; and decreased 56%, increased 1381%, and decreased 32%, respectively, for (litter burn) severe burns (total N peaked at moderate burn, increasing 385%)	N/A	Hogue and Inglett, 2012
DOC: 3.3 (0.7) mg_C/g; DOC:DON: 32.7 (10.5) mg_C/mg_N	N/A	DOC: 1.6 (0.7) mg_C/g; DOC:DON: 23.9 (7.6) mg_C/mg_N	N/A	DOC and DOC:DON decreased 52% and 27%, respectively, for moderate burn areas	Standard deviations ranged from 21% to 44% of the means	Hohner et al., 2019

N/A	N/A	N/A	PyC median mass loss: 6.6 (0-70.4)% and 15.1 (0-100)% (experimental fire); 15.1 (13.3-16.0)% and 1.0 (0.7-25.4)% (lab heating at 370 °C); 27.6 (25.4-34.2)% and 63.3 (20.9-76.7)% (lab heating at 470 °C); 25.7 (21.8-29.0)% and 36.7 (16.7-81.4)% (lab heating at 570 °C) (granular PyC and PyC pieces, respectively)	PyC median mass decreased 6.6% and 15.1% for granular PyC and PyC pieces, respectively, for experimental fire and on average 22.8% and 34%, respectively, for lab muffle burns	Wide range of variability (up to 533% of the medians)	Santin et al., 2013
TOC: 7.5%, 13%, and 9.6% (three replicates)	N/A	TOC: 17%, 23%, and 7.9% (three replicates)	TOC: 9.7%, 9.2%, and 7.8% (three replicates)	TOC increased 60% for moderate burns and decreased 11% for severe burns	N/A	Wieting et al., 2017

*0-1 cm, 1-2 cm, and 2-3 cm depths, respectively

Appendix C

Simulation Experiment Results

C.1 Burn Intensity Characterization Analysis Results

Six burn intensity characterization methods were initially evaluated for the wildfire simulation experiment, as shown in Figure C.1. 'Peak temperature' and 'Peak temperature (data logger)' were measurements of peak soil sample surface temperatures achieved, derived from visual inspections of the data loggers during burning and a post-burn assessment of time-temperature curves, respectively. These temperatures were then binned into a temperature-based burn intensity scale derived from previous literature. 'Degree hours' were characterizations based on both temperature and time, calculated by integrating under samples' entire time-temperature curves. 'Modified degree hours' were similarly based on temperature and time, though this metric was calculated by integrating under time temperature curves just until peak surface temperatures were achieved. Both of these metrics were binned into discrete burn intensity characterizations with cutoffs based on their terciles. The 'Luminance' characterization metric was based on the reflectance of samples derived from image processing, with intensity bins similarly based on the metric's terciles. Finally, 'Visual' characterizations of burn intensity were completed by two separate

researchers, using U.S. Forest Service burn severity methods as a guide (Parson et al., 2010). Though this guide additionally recommends burn severity characterization through assessment of vegetation and root destruction, this aspect was neglected as visual characterization was completed after the experiment using images of soil samples. Analysis of variance (ANOVA) tests for each characterization method show whether all burn intensity groups are significantly difference from all other groups ($\alpha = 0.05$).

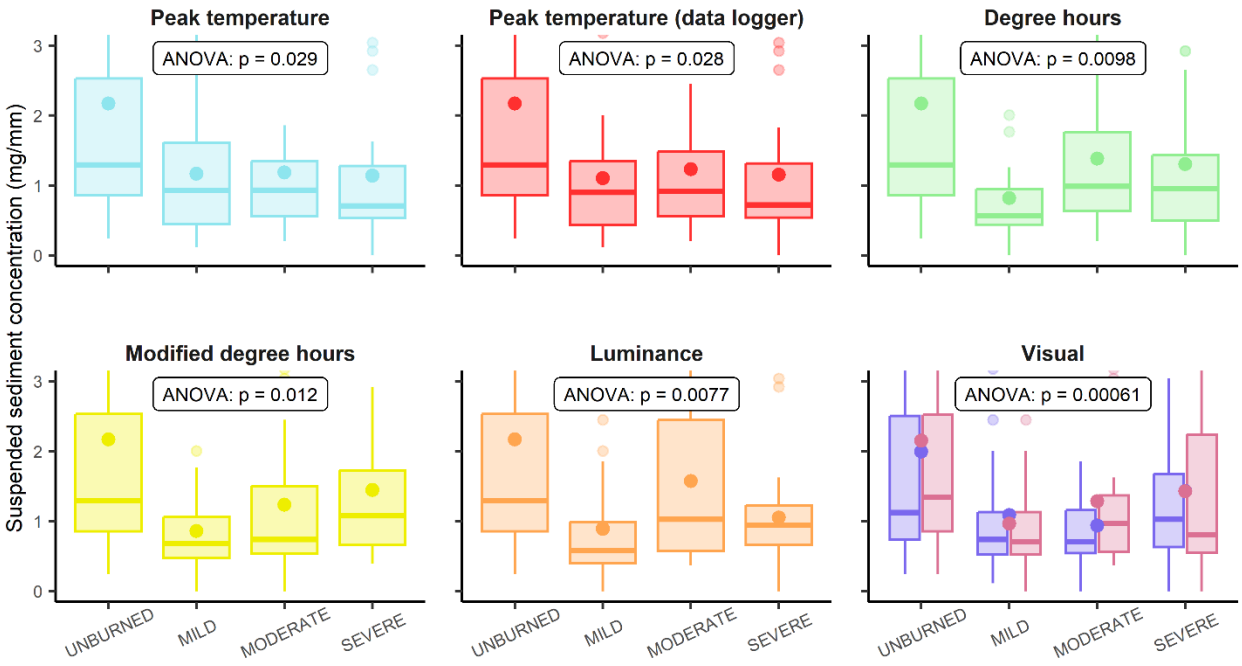

Figure C.1: Boxplots of suspended sediment concentration, as an example, with increasing burn intensity increments characterized using six different methods.

While temperature-based and degree hour characterizations were similar for each burned soil sample, important differences resulted in distinct relationships with hydrologic and water quality responses. Degree hours generally trended linearly with surface temperatures as shown in Figure C.2, with an $R^2 = 0.77$ correlation when degree hour outliers (values greater than one standard deviation above the median) were removed. However, greater differences in degree hours from peak temperatures existed for samples

raised to temperatures above ~500°C, due to high variability of heating ramp-up and cool-down times during severe burn simulations. Due to these differences, results showed that peak temperature (highlighted by the temperature-based characterization) may have been a stronger driving mechanism for runoff and solute responses, whereas heating durations (highlighted by degree hours) affected sedimentation and turbidity responses more strongly. When compared to the temperature-based scale, runoff ratios showed significant monotonic increases at 30° terrain slopes with unburned samples excluded, with an ANOVA test p-value of 0.036, while trends were non-existent when compared to degree hours. DOC and TDN showed significant inverse 'U' shapes (ANOVA p-values of 0.044 and 0.0033, respectively) with increasing temperature-based burn intensities, similarly lacking trends when compared using degree hour intensities. These significant ($\alpha = 0.05$) ANOVA tests indicated that each burn intensity group was significantly different from all other groups. Suspended sediment concentration (SSC) and turbidity were more strongly correlated with degree hours than surface temperatures, however, showing significant monotonic increases with increasing degree hours when unburned samples were removed and at a 20° terrain slope for turbidity (ANOVA p-values of 0.013 and 0.033, respectively).

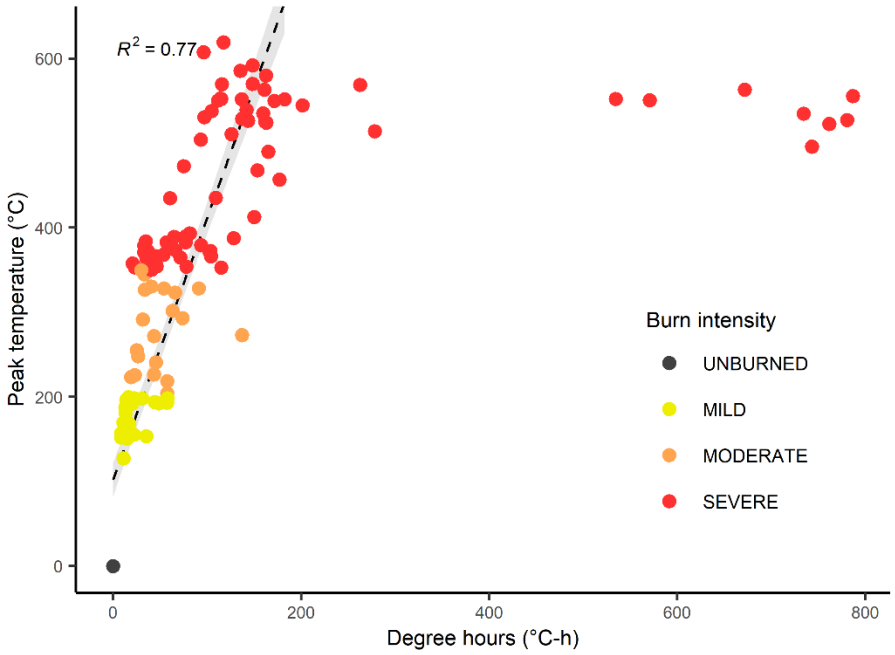

Figure C.2: Peak soil surface temperatures during burn simulations plotted against calculated degree hours. Different colors represent samples' burn intensities characterized using the temperature-based scale. The dashed lines show the best linear fit of data with degree hour outliers (or values greater than one standard deviation above the median) removed (R^2 = 0.77). The shaded area represents the confidence intervals (level = 0.95).

C.2 Hydrologic and Sedimentation Responses

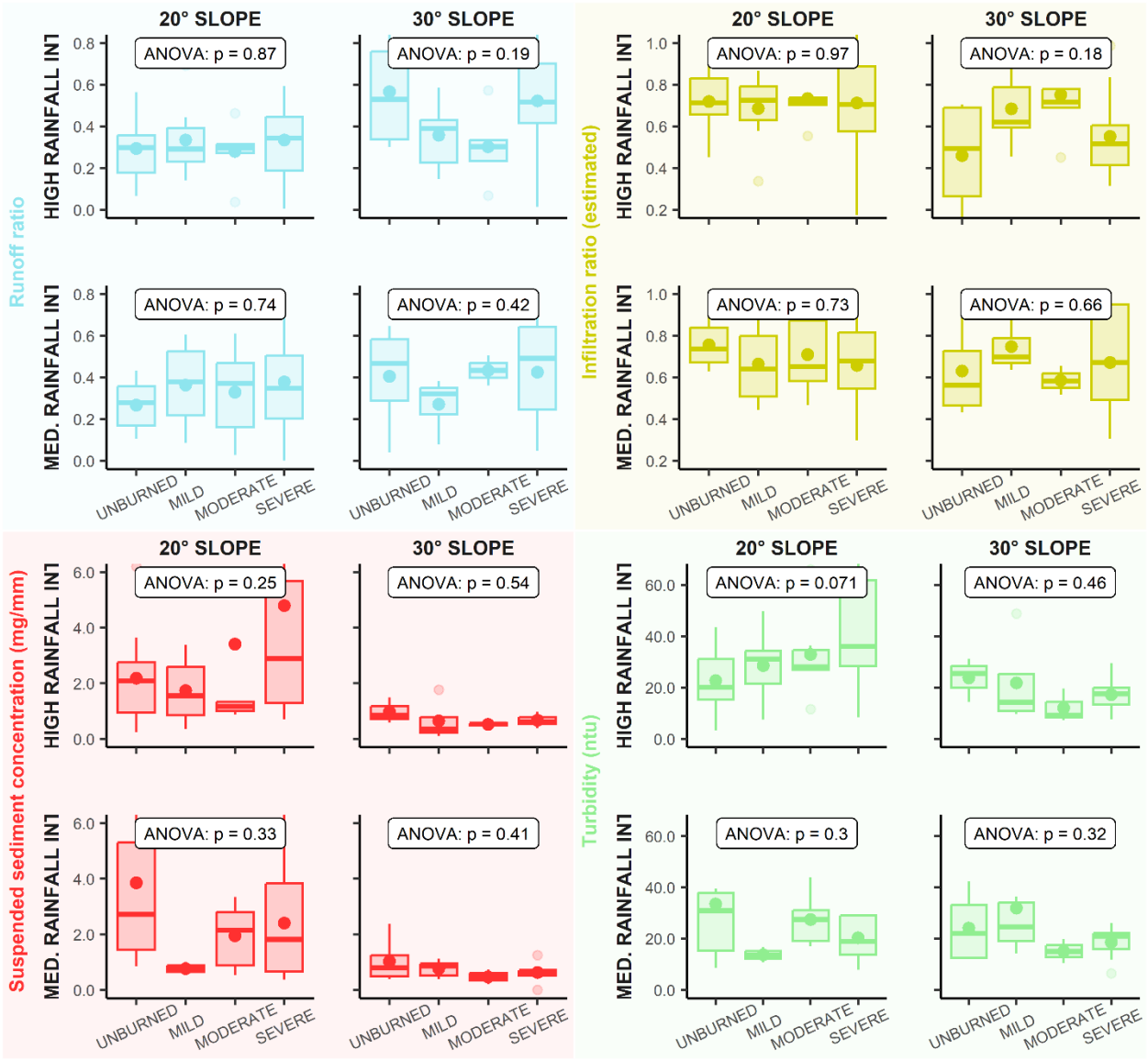

Figure C.3: Boxplots of runoff ratio, infiltration ratio (defined as percolation summed with estimated storage, then divided by total precipitation), SSC, and turbidity responses from soil samples at moderate and high rainfall intensities and 20° and 30° terrain slopes for increasing burn intensities. ANOVA p-values were not significant at any combination, indicating that burn intensity groups were not statistically different from all other groups. Turbidity is reported in nephelometric turbidity units (ntu).

C.3 Sequential Rainfall Treatments

Outside of samples tested in the experimental matrix, 27 additional samples from varying burn intensities were subjected to two sequential rainfall treatments. These samples either received first low intensity (~14.4 mm/h) precipitation for 2 h, then high intensity (~51.3 mm/h) for 2 h, or vice versa, with a ~24 h drying period in between. As seen in Fig. S5, the second precipitation events typically generated more runoff than first events with equivalent rainfall intensities. Median runoff ratios for the second treatments were almost 50% higher than median values for the first treatments. This relationship was slightly heightened with increasing burn intensity, with median runoff ratios for second treatments ~73% higher than first treatments for severely burned samples. As seen in Fig. S6, water quality responses were slightly lower for the second treatments than the first, with median values ~20, 5, 42, and 28% lower than the first treatments for SSC, DOC, TN, and turbidity, respectively.

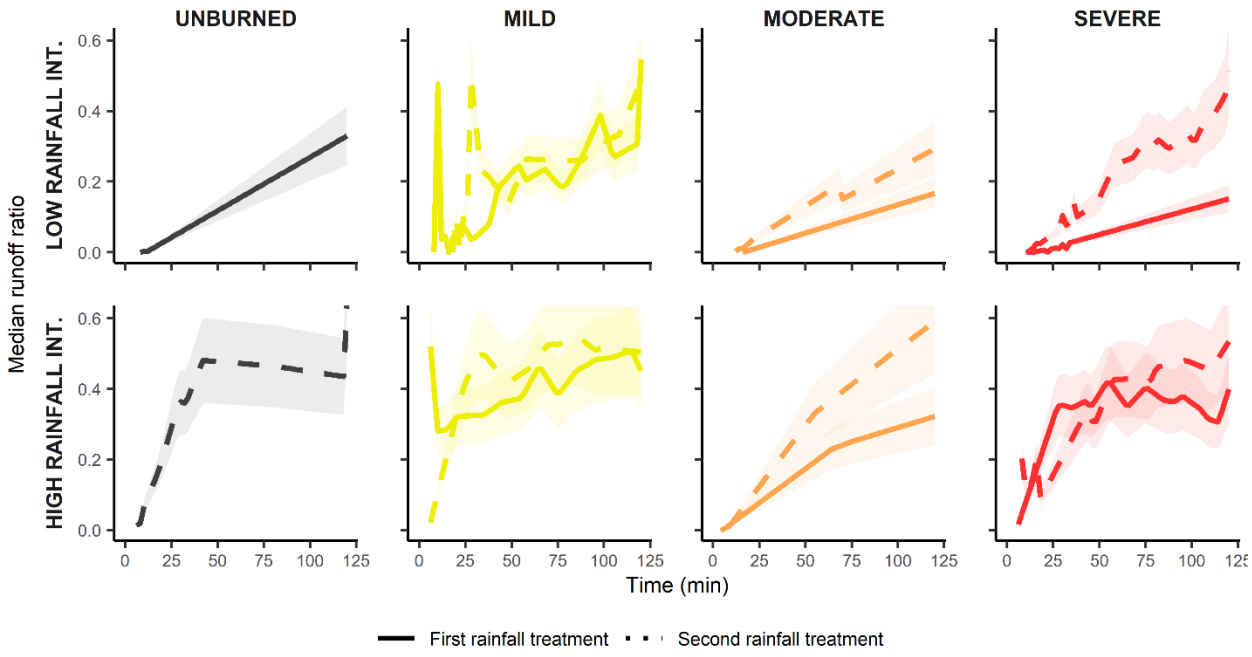

Figure C.4: Time series plots of median runoff ratios from soil samples which received two sequential rainfall treatments, with a ~24 h drying period in between. Responses for low (~14.4 mm/h) and high (~51.3 mm/h) rainfall intensities are shown in the top and bottom rows, respectively. Colors represent different burn intensities, with the solid lines showing responses from the first rainfall treatments and the dashed lines representing the second treatments. Shaded areas represent the interquartile ranges.

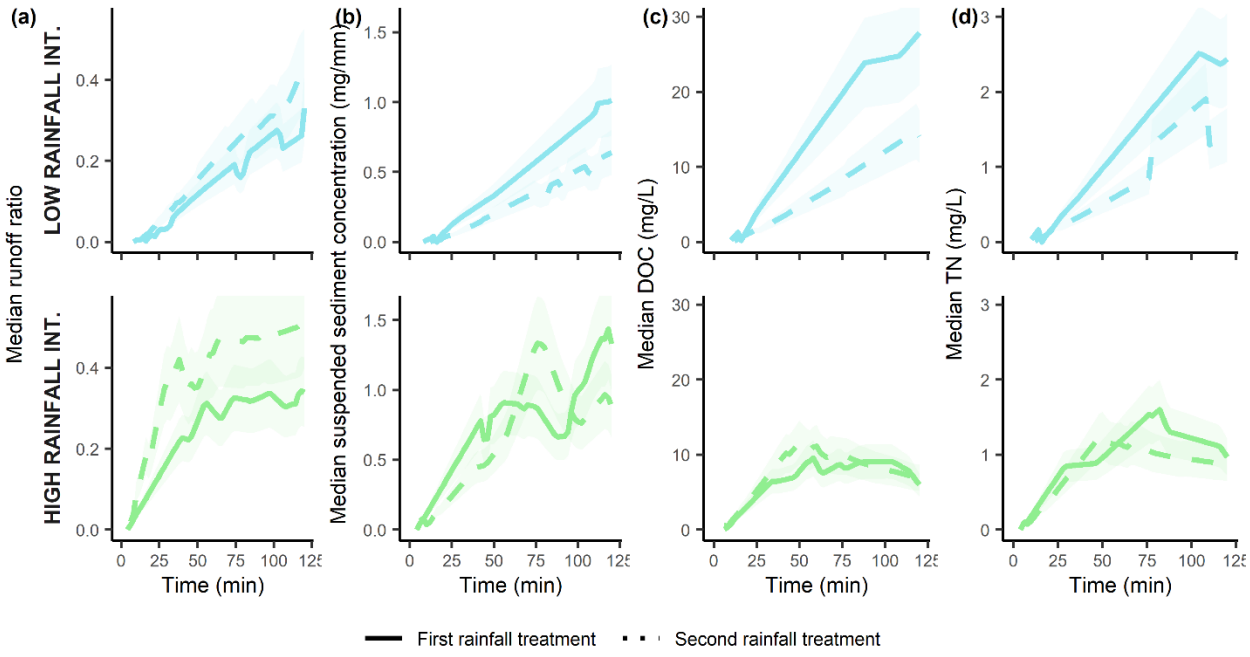

Figure C.5: Time series plots of median runoff ratio, suspended sediment concentration, DOC, and TDN responses from samples subjected to two sequential rainfall events, with a ~24 h drying period in between. Responses for low (~14.4 mm/h) and high (~51.3 mm/h) rainfall intensities are shown in the top and bottom rows, respectively. Colors represent these two treatment options, with the solid lines showing responses from the first rainfall treatments and the dashed lines representing the second treatments. Shaded areas represent the interquartile ranges.

Appendix D

Simulation Experiment Limitations

D.1 Mass Loss During Burn Simulation

As discussed in Section *3.5.2* of the manuscript, extra handling of burned samples may have contributed to higher sediment response in unburned samples. Where unburned samples were never removed from the lab, burned samples were disturbed during transportation to an outdoor testing area, exposed to light winds, and inserted with thermocouples, degrading soil structure. High mass loss occurred during this step up to ~15% of samples' weights, as seen in Fig. S7. Though some of this may have been due to volatilization during burning, soil loss from wind and shaking during transportation was likely.

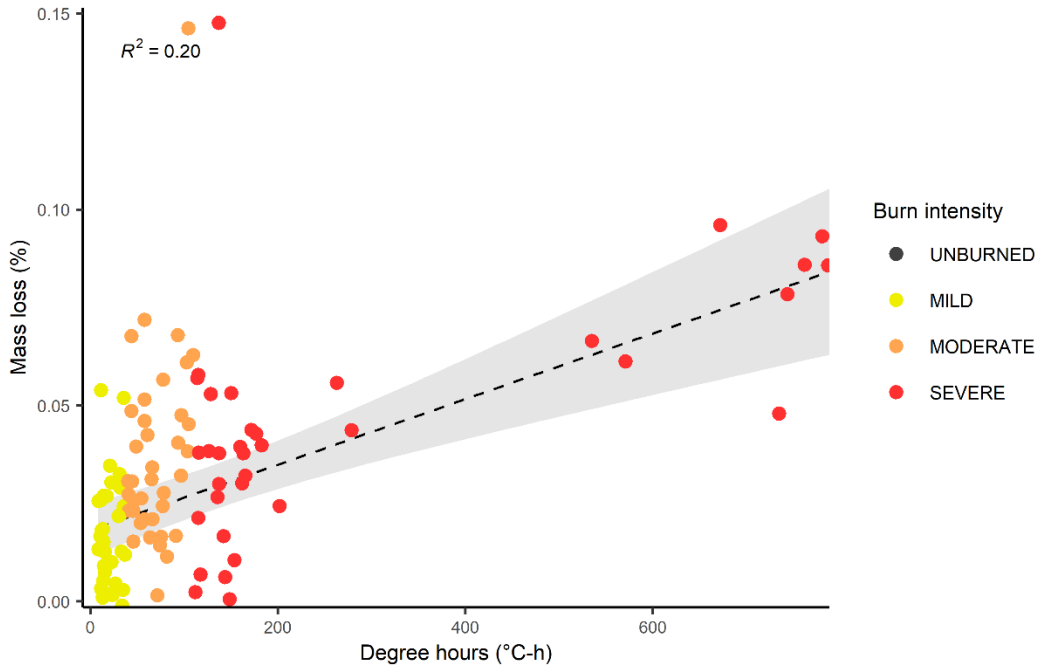

Figure D.1: Soil sample mass loss, or the percent change in weight after burning, plotted against degree hours achieved. Point colors represent different burn intensities. The dashed line represents the best linear fit of the data, with an R^2 = 0.20, and the shaded area shows the confidence interval (level = 0.95).

D.2 Water and Sediment Loss During Rainfall Simulation

Comparisons of storage estimates and changes in soil moisture for each sample revealed system losses during rainfall simulation. Storage was estimated by closing samples' water balance equations, as discussed in Section *3.5.2* in the manuscript, made under the assumption that no losses occurred during simulation. However, this was unlikely due to abstractions and unaccounted flow paths throughout the system. Thus, these estimates were compared to the change in volumetric soil moisture in several soil samples before and after simulation, with values converted to depths by using the samples' weights and common physical characteristics for loamy sand. As shown in Fig. S8a, little to no trend was apparent in storage estimates with increasing burn intensity, whereas change in soil moisture generally increased monotonically with increasing intensities—indicating

potential inaccuracies in the estimates. Storage values were somewhat linearly related to change in soil moisture, as shown in Fig. S8b, with R^2 values up to 0.47 for specific burn intensities when values were expressed as percentages of total precipitation. However, these estimates were generally greater than soil moisture changes, or up to a difference of 91% of precipitation. Storage estimates were closest to changes in soil moisture for unburned samples, with a median difference of storage from moisture change of 2.9% of precipitation. Mild, moderate, and severe intensities had median differences of 43, 43, and 28%, respectively.

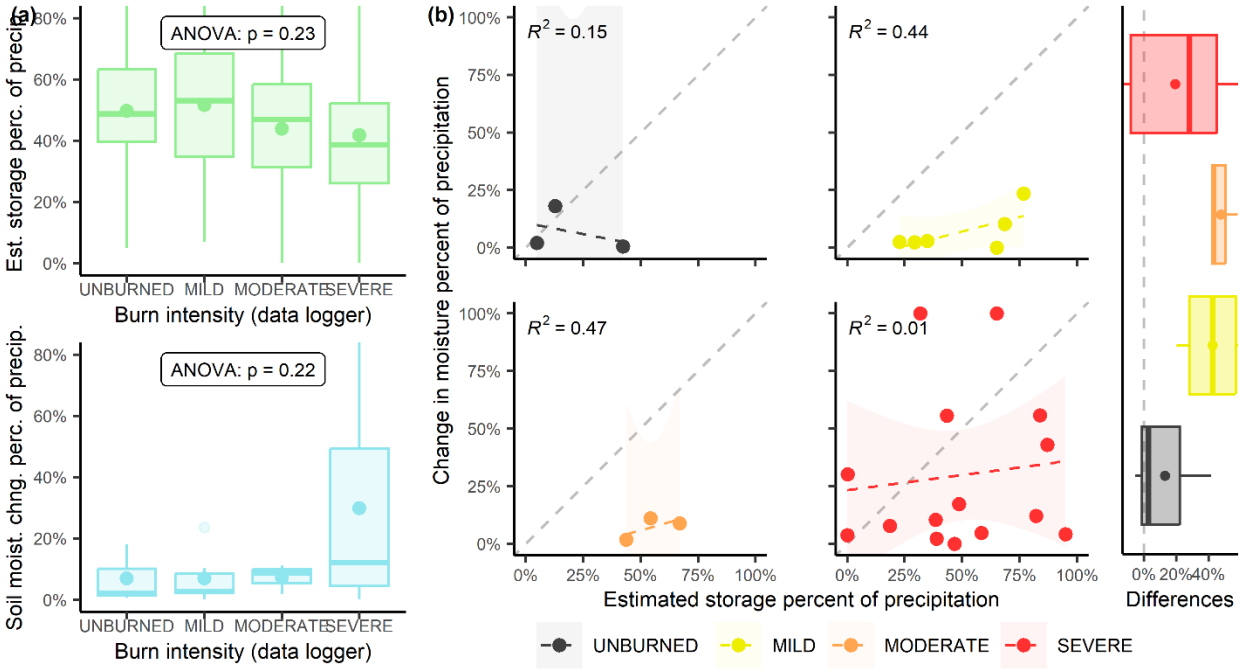

Figure D.2: (a) Boxplots of estimated storage and soil moisture change values expressed as percentages of total precipitation for 27 individual soil samples, plotted with increasing burn intensity. (b) Change in soil moisture during rainfall simulation for each sample plotted against estimated storage during the event, expressed as percentages of total precipitation. Plots are divided into unburned, mild, moderate, and severe burn intensities, with best linear fits represented by dashed lines and R^2 values of 0.15, 0.44, 0.47, and 0.01, respectively. Shaded areas represent confidence intervals (level = 0.95), with a gray, dashed 1:1 line in the background. A boxplot displays differences in storage from change in soil moisture for each burn intensity, expressed as percentages of total precipitation. Colors represent different burn intensities.

Losses during rainfall simulation provide potential explanations for anomalously high experimental results. As previously discussed, runoff and sediment response for unburned samples was anomalously high. Greater differences in storage estimates from changes in soil moisture in burned samples as compared with unburned samples indicated greater liquid and sediment loss during rainfall simulations for these samples. Liquids and sediment may have gotten trapped in the custom funnels, been blocked by sediment in the tubing system, or run over the sides of samples at greater rates for burned samples, potentially due to higher sediment generation. Runoff response in burned samples was likely further muted by preferential flow paths in soil samples. Though minimized by petroleum jelly and duct tape, these holes allowing excessive precipitation to flow through were observed at samples' interfaces with their steel containers. Due to greater handling and disturbances, these holes were especially common in burned samples, reflected by large increases in percolation with increasing burn intensity. Sediment trapping in the rainfall simulator system was also evident from a 44% decrease in median TSS from the 20° to 30° slope angles, despite median runoff, sediment's main transport mechanism, experiencing a 41% increase. The higher angle may have increased sediment settling in the corners of the custom funnels or allowed for the transport of larger sediment particles which clogged the tubing system.

Appendix E

Small-Scale Physical Modeling

E.1 Background

Small-scale observations of wildfire effects can provide insight into specific driving mechanisms of soil and water physical and chemical changes, however difficulties in upscaling can inhibit usefulness in catchment-scale predictions (Beeson et al., 2001; Ferreira et al., 2008). Hydrologic mechanisms such as streamlet connectivity are important in solute transport and geomorphological effects (Kampf et al., 2016; Wilson et al., 2021), but are typically not captured in small-scale observations. Similarly, erosional, sediment-producing processes such as streambed erosion and the formation of rills and gullies also vary greatly from small- to catchment-scales (Robichaud et al., 2016).

Here, we explored a novel approach to catchment-scale post-wildfire hydrologic and water quality modeling by using the physical models HYDRUS 1D and 2D with parameters constrained by laboratory-scale wildfire simulation experiment data. HYDRUS 1D and 2D are hydrology- and solute-modeling one- and two-dimensional environments, respectively (Li et al., 2021). These models solve the Richards equation to simulate water flow and the advection-dispersion equation to simulate heat and solute transport vertically and laterally

in variably saturated subsurface media (Yu and Zheng, 2010). Though able to model fine-scale hydrologic and transport processes, HYDRUS 1D and 2D have also been used in a variety of basin-scale applications (Anlauf et al., 2018; Li et al., 2021; Varvaris et al., 2021). Coupled with HYDRUS 1D and 2D, the Modified Universal Soil Loss Equation (MUSLE) is a soil erosion model (Benavidez et al., 2018) which will be used to simulate sedimentation rates at the catchment scale. This experimental framework will allow for the assessment of small-scale burn effects on catchment-scale processes, providing insight into the formation of upscaling factors. As simplistic cause-and-effect diagram of small-scale burn effects on catchment-scale hydrologic and water quality processes is shown in Figure E.1.

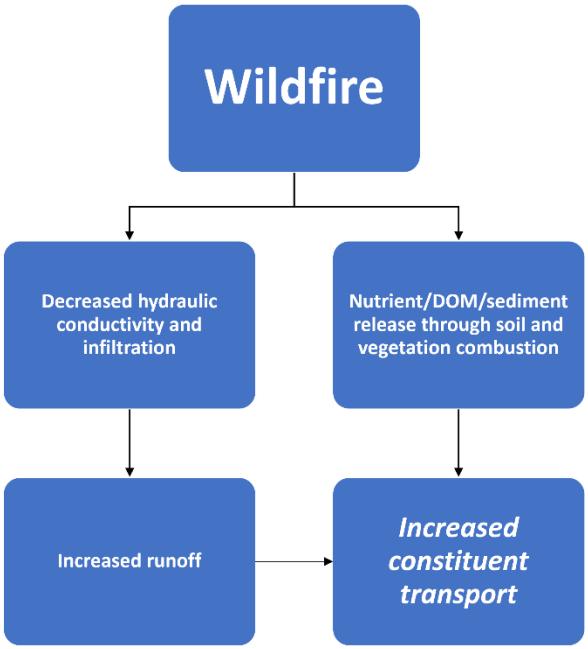

Figure E.1: Diagram of key wildfire effects on soil physical and chemical properties and subsequent hydrologic and water quality responses.

E.2 Methods

First, HYDRUS 1D parameters will be constrained with experimental data, reflecting burn induced water repellency and ground cover changes. Next, a HYDRUS 2D model will be applied to burn-affected basins to predict catchment-scale hydrologic and

water quality response, and MUSLE will use hydrologic model outputs to simulate post-fire sedimentation rates.

Preliminary HYDRUS 1D Model Set-up

HYDRUS 1D was first set up to reflect experiment conditions from Chapter 3, using hydraulic parameters for 10.16 cm thick soil samples comprised of sandy loam (common in the Fraser Experimental Forest ("Web Soil Survey," 2019)) at tested terrain slopes and rainfall intensities. The Durner soil hydraulic model (Durner, 1994) was used in HYDURS 1D to allow eventual representation of two types of porous media (i.e., burned and unburned), with the following equation:

$$S_e = w_1[1 + (\alpha_1 h)^{n_1}]^{-m_2} + w_2[1 + (\alpha_2 h)^{n_2}]^{-m_2}$$

Equation E.1

$$K(S_e) = K_S \frac{(w_1 S_{e_1} + w_2 S_{e_2})^l \left(w_1 \alpha_1 \left[1 - \left(1 - S_{e_1}^{\frac{1}{m_1}} \right)^{m_1} \right] + w_1 \alpha_1 \left[1 - \left(1 - S_{e_2}^{\frac{1}{m_2}} \right)^{m_2} \right] \right)^2}{(w_1 \alpha_1 + w_2 \alpha_2)^2}$$

Equation E.2

where w_i are weighting factors for the two overlapping regions, and α_i, n_i, m_i (=1-1/n_i), and I are empirical parameters of the separate hydraulic functions (i=1,2) (Simunek et al., 1998). Values of the saturated soil water content, parameters n and α (for the soil water retention function), saturated hydraulic conductivity, tortuosity, and other parameters necessary for the Durner's model were then automatically assigned by a built in Soil Catalogue. The residual soil moisture content, however, was set to the average moisture content of soil samples before rainfall simulation (~4.5%). This model was then run at a range of flow path angles (i.e., terrain slopes) (5° to 45°), using the highest simulated rainfall intensity (6.35 cm/hr) and assuming negligible evapotranspiration. Runoff and

infiltration measurements were recorded at a 0.1 hr intervals over a 1-hr simulation period, with timesteps ranging from 0.001-0.1 hrs (HYDRUS automatically selects the optimal timestep). This model framework will be expanded to include simulations at all 16 driver increment combinations (4 burn severities, 2 rainfall intensities, and 2 terrain slopes) as described in Chapter 3. Additionally, simulation length will be extended to 2 hours to reflect experimental procedures, and solute transport will also be simulated.

To simulate burn effects, different values of hydraulic conductivities will be tested iteratively at 1-3 cm below the soil surface, i.e. where burn induced water repellency is typically observed (Larson-Nash et al., 2018). Yochum et al. (2015) suggests multiplying soil hydraulic conductivity by factors of 1.5, 1.75, and 2.0 to reflect effects of mild, moderate, and severe fires, thus these multipliers will be used as initial guesses. Simulated results will be compared to local-scale runoff and dissolved carbon (i.e., solute) transport and assessed for skill—identifying most accurate parameter values across varying terrain slopes and rainfall intensities.

HYDRUS 2D Application to the Catchment Scale

HYDRUS 2D will then be applied to wildfire-affected basins by simulating the connectivity of individual grid cells. Boundary conditions—such as antecedent moisture, vegetation effects, etc.—of the HYDRUS 2D model may be further informed by Variable Infiltration Capacity (VIC) models (Hamman et al., 2018). Incorporation of VIC would allow for investigation of ground cover and vegetation changes in post-fire environments, which can be strong hydrologic and water quality drivers (Miller et al., 2003). The SoilGrids database (Hengl et al., 2017) will be used to supply detailed soil composition information for modeled basins, limiting grid-cell sizes to 250 m resolution.

Model simulations will be run at an hourly time-step to simulate individual storm events, and a daily time-step to evaluate catchment response over entire seasons. Runoff and solute (i.e., dissolved nitrogen, heavy metals, etc.) transport will first be modeled based on pre-fire conditions, assessing for skill by comparing to historical pre-burn data. Next, grid cells will be constrained to reflect burn effects, rerunning the model and comparing to post-wildfire water quality response. From there, basin soil and vegetation characteristics can be calibrated to increase skill—providing insight into key factors which increase basin vulnerability to post-wildfire water quality effects.

MUSLE Sedimentation Model

Hydrologic outputs from the HYDRUS model will then be input into MUSLE to create catchment-scale sedimentation predictions. Gridded datasets from the HYDRUS catchment modeling efforts were used for additional MUSLE inputs: soil and land cover data, as well as digital elevation models. MUSLE incorporates 7 soil, topographic, hydrologic, and vegetation characteristics into its erosion estimate: the total volume of storm runoff (Q) in m^3, the peak flow rate (q_p) in m^3/s, the soil erodibility factor (K), the slope length factor (L), the slope steepness factor (S), the cover-management factor (C), and the support practice factor (P). These factors calculate sediment yield (Y) in metric tons through the following equation (Renard et al., 1997; Yochum and Norman, 2015; Zhang et al., 2009):

$$Y = 11.8(Q * q_p)^{0.56} * K * L * S * C * P$$

Equation E.1

Q and q_p factors will be taken from HYDRUS outputs, K from a soil characteristic database ("Web Soil Survey," 2019), and C from a land cover database (Homer et al., 2012).

P will be set to 1, as little crop support practices are used the modeled basins. Digital elevation models (Hastings and Dunbar, 1993) will be used to derive terrain slope, represented in the L and S variables with the following equations:

$$L = (\frac{\lambda}{22.13})^m$$

Equation E.2

$$m = (\frac{\beta}{1 + \beta})$$

Equation E.3

$$\beta = \frac{\frac{\sin\theta}{0.0896}}{3(\sin\theta)^{0.8} + 0.56}$$

Equation E.4

$$S = 3(sin\theta)^{0.8} + 0.56$$

Equation E.5

Where λ is the horizontal slope length (m), m is the variable slope-length exponent, β is the mean slope angle, and θ is the slope angle (degrees). A test of this process was completed for small-scale models, as shown in Figure E.1.

For each basin, MUSLE will first be run with unburned parameters and inputs and compared to pre-fire historical data. Then, soil and crop parameters, specifically *K* and *C*, will be adjusted to reflect burn effects, based on the HYDRUS 2D model parameters, and model outputs compared to post-fire response. The residuals will then be quantified and characterized to define sedimentation upscaling operators. In addition to using MUSLE, I will also explore developing a simple statistical model constrained with local-scale data which relates HYDRUS 2D outputs to basin-scale sedimentation response.

E.3 Initial Results

Initial HYDRUS 1D and MUSLE simulations show cumulative runoff, drainage through samples, and sedimentation rates for different terrain slopes, as shown in Figure E.1. As expected, increasing the terrain slope produces increased runoff and decreased drainage in the HYDRUS model, indicating the model is running correctly.

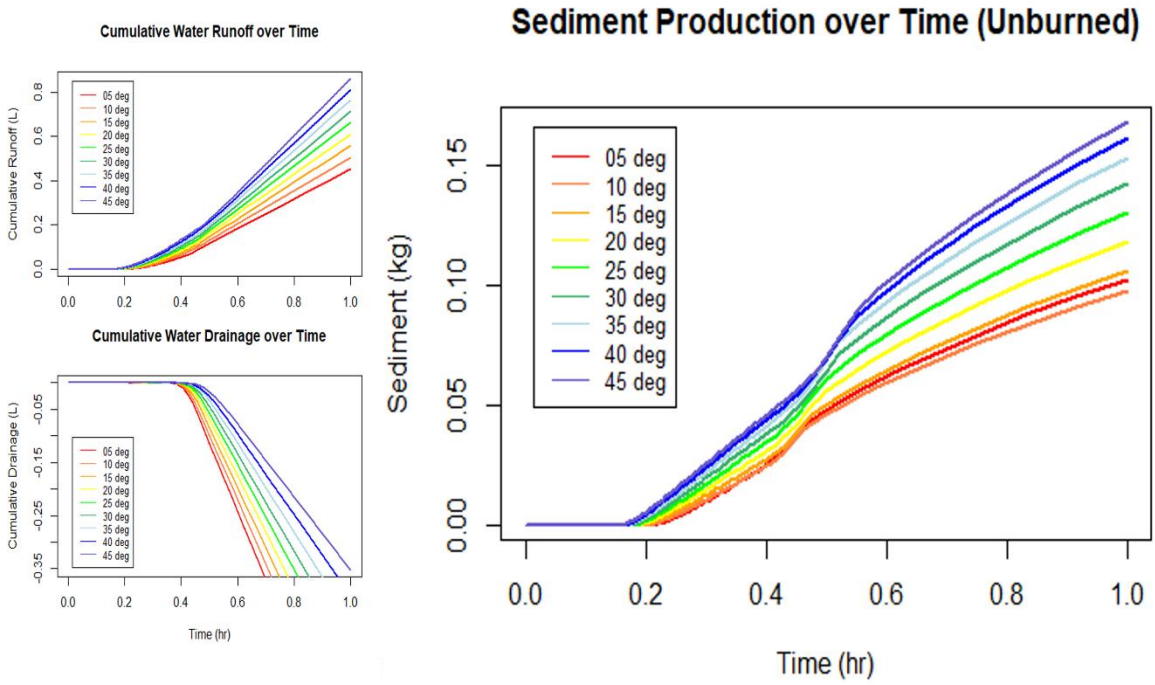

Figure E.1: Simulated HYDURS 1D cumulative runoff (upper left) and drainage (lower left) from a single soil sample, as well as simulated MUSLE sediment generation (right). Simulations were run with a rainfall intensity of 6.35 cm/hr at a range of terrain slopes.

Sediment generation similarly increased with increasing terrain slope, indicating that the MUSLE simulation was set up correctly. However, simulated sedimentation rates are very high compared to experimental results, indicating an inaccuracy somewhere in the model creation methods.

Appendix F

Post-Wildfire Water Quality Inter-Site Variability Analysis

Residuals from the models built for each constituent-basin combination for the first two years post-fire were plotted against several watershed and wildfire physical characteristics. This provided insight into drivers of variability in the magnitude of response for each constituent, informing characteristics more common in high post-fire response events. To visualize these relationships, the mean of residuals from the first two years post-fire were plotted against characterization metrics of each watershed and wildfire variable.

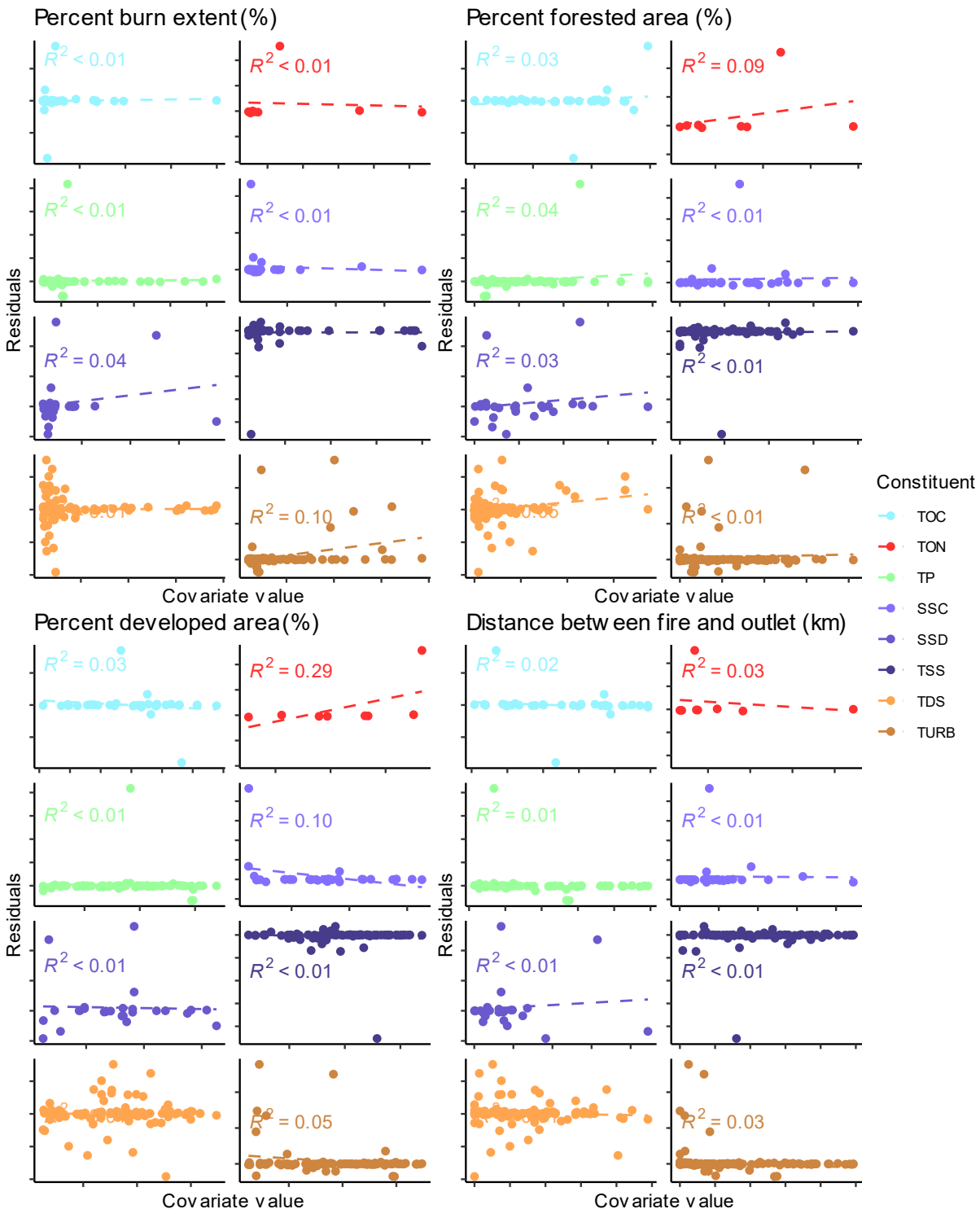

Figure F.1: Four physical watershed and wildfire characteristics plotted against mean model residuals for each basin-wildfire combination from the first two years after wildfire events for each water quality constituent. The best linear fit was calculated for each combination, and the associated R^2 value displayed.