Title
Development of novel data applications for improving precipitation-runoff modeling in headwater catchments

Author
Maurer, Tessa

Publication Date
2020

Peer reviewed|Thesis/dissertation

Development of novel data applications for improving precipitation-runoff modeling in headwater catchments

by

Tessa Maurer

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Civil & Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Steven D. Glaser, Chair
Professor Roger C. Bales
Professor Laurel Larsen
Professor Arpad Horvath

Fall 2020

Abstract

Development of novel data applications for
improving precipitation-runoff modeling in headwater catchments

by

Tessa Maurer

Doctor of Philosophy in Civil & Environmental Engineering

University of California, Berkeley

Professor Steven D. Glaser, Chair

Hydrologic modeling is heavily used in water resources for advancing scientific process understanding and supporting operational decision making, but the capabilities of these models are challenged by climate change, land-use changes, and population growth. Novel ground-based and remotely sensed data sources can provide high-quality and reliable information that can enhance hydrologic models to meet current demands. Here, I examine how these data sources can be leveraged to facilitate and improve use of three major model types: data-driven, physically based, and conceptual, which each fall in a different range of complexity (the extent to which a model represents or simplifies physical processes) and scale (the number of simulation or data points). I focus on the state of California, a region that faces significant water management challenges due to its arid Mediterranean climate, highly seasonal precipitation, and high variability in annual precipitation totals. Chapter II investigates snowpack response to rain-on-snow events by leveraging expanded ground-based wireless-sensor networks to drive a machine-learning algorithm. Supported by the high temporal and spatial resolution of these data, I identified a nonlinear and temporally variable relationship between Leaf Area Index and snow-depth change during rain-on-snow events. Chapter III demonstrates how remotely sensed spatial maps can be used to identify the optimal spatial distribution of input data in a physically based hydrologic model through use of a Gaussian-Mixture Model. This novel method addresses issues of overparameterization and improves the efficiency of model design for physically based models while still being grounded in physical principles. Finally, Chapter IV applies new spatially distributed data to a conceptual hydrologic model to quantify the impact of predictable versus unpredictable shifts in the precipitation-runoff relationship during droughts. Collectively, this work demonstrates that 1) data that are gathered strategically by location and water-balance component is most helpful in improving or developing new models; 2) remotely sensed data are most effective when calibrated or coupled with ground-based sensors; 3) scale continues to be a challenge both on the modeling and data sides, whether scaling up physically based models or scaling down conceptual models; and 4) model improvement relies not only on data collection but on making the data usable and easily accessible, particularly for operational applications.

For the trespassers,

and in honor of the life of Richard O. Maurer.

Life is not a stream to be crossed, to be dammed, or to be followed,
because you determine where it flows.
– B. Tinsley

The water is wise.
– Tim O'Brien

Acknowledgements

This work would not have been possible without the support, guidance, and collaboration of my advisors, Steven Glaser, Roger Bales, and Martha Conklin. I am also deeply grateful for my collaborators Francesco Avanzi, Carlos Oroza, Mohammad Safeeq, and the entire Mountain Hydrology Research Group for their ideas and feedback. This work was partially supported by the California Energy Commission under contract EPC-14-067 and the National Science Foundation Graduate Research Fellowship under grant DGE 1106400. Critical support was provided by Kevin Richards, especially for the modeling efforts in Chapter 2. A million thanks go to our field work team that assisted in deploying ground-based sensors in the Feather River, but most especially to Peter Hartsough, Ansel McClelland, and Rob Taylor. I would be remiss to not acknowledge Shelly Okimoto and the CEE staff, whose excellence makes graduate student life better in a million large and small ways every day. My experience in graduate school was deeply enriched by my interactions and friendships with other graduate students, especially the Spring 2020 ACES Graduate Learning cohort. Your thoughtfulness, insights, and support were invaluable in the final push of this dissertation amidst the blazing detritus receptacle of this year. I am forever indebted to my parents, my sibling, and my four wonderful and loving grandparents for their unerring belief in me. Finally, thank you, Baela, for laughing, commiserating, celebrating, discussing, and asking me to let go sometimes. I love you.

Table of Contents

Chapter I

Introduction

The field of water resources relies heavily on hydrologic modeling for simulating and forecasting of hydrologic processes. The overarching goal of these models is to better understand and predict the movement of water in the critical zone, the surface and near-surface environment that supports most of Earth's terrestrial life, but with particular emphasis on predicting streamflow. Processes simulated by hydrologic models may include precipitation phase estimation (rain versus snow); canopy interception of precipitation; snow accumulation and ablation; and infiltration, interflow, and runoff dynamics (Hrachowitz and Clark, 2017; Kirchner, 2006). On the scientific side, researchers make use of models to better understand the mechanisms underlying physical processes, how those mechanisms differ spatially and temporally, and how the hydrologic cycle changes under different scenarios such as greenhouse gas emissions or land-use changes. From an operational standpoint, drinking water agencies, irrigation districts, and hydropower operators use forecasting models to inform and improve decisions on water allocation, use, and storage. They may be seeking to optimize operations under a given set of conditions or to understand the implications of anthropogenic or natural changes for their management plans (Blöschl and Sivapalan, 1995; Kirchner, 2006; Troch et al., 2015).

Increasingly robust computational resources over the past few decades coupled with longstanding, widespread environmental monitoring in the United States and much of the developed world has enabled the expansion of modeling for both research and operational hydrology (Beven, 1989; Kirchner, 2006). Models are currently a critical component of water resources planning and management, particularly in arid Mediterranean climates like the state of California, which face a number of unique water supply challenges. Precipitation in these regions is highly seasonal (Klos et al., 2018), requiring extensive infrastructure and careful operational planning to maintain supplies through the dry season (He et al., 2017). Interannual precipitation totals also vary significantly, leaving these regions prone to drought on the one hand and flood risk during wet years on the other (He et al., 2017; Woodhouse et al., 2010; Ralph and Dettinger, 2011). Furthermore, like other Mediterranean climates, California is primarily dependent on mountain runoff for its water supply (He et al., 2017; García-Ruiz et al., 2011), which creates unique modeling demands given the highly heterogeneous landscape. This topographic variability impacts the water balance through orographic effects, which can create dramatic differences in precipitation rates on either side of mountain ranges (rain-shadow effects; Roe, 2005); elevational gradients of precipitation

amount, precipitation phase, and vegetation distribution (Rungee et al., 2019; Avanzi et al., 2020a); and heterogeneous solar load depending on slope and aspect (Lundquist et al., 2013). Snowpack is critical to the water balance in these regions, providing a natural storage reservoir that helps extend runoff availability beyond the precipitation season. Snowpack also increases the availability of plant-accessible subsurface water storage (Rungee et al., 2019), supporting vegetation growth through the dry season.

The development and use of hydrologic models in regions like California rely on the ongoing availability of readily accessible climate and streamflow data, traditionally comprised of manual or in-situ sensor measurements of variables such as precipitation, wind speed, air temperature, relative humidity, snow water equivalent (SWE), and streamflow volume (Bales et al., 2006; Dozier, 2011). Here again montane topography presents unique challenges, as large portions of the landscape may be inappropriate for installing sensors (such as steep slopes); sensors are subject to harsh weather conditions, particularly in the wintertime; and sensing locations may be inaccessible for maintenance for large portions of the year due to the terrain and snow accumulation (Zhang et al., 2017b). In order to ensure water supply and limit flood risk in these regions, models and data must be leveraged together to understand the hydrologic processes that contribute to the timing and quantity of streamflow.

Hydrologic models approach simulation of these physical processes in a number of ways; two common means of classifying them are complexity and scale (Hrachowitz and Clark, 2017). Complexity describes the extent to which a model simplifies physical systems, while scale refers to the spatial granularity of the model. For example, a detailed precipitation-runoff model that resolves mass and energy balances to explicitly simulate precipitation phase, solar radiation load, canopy interception, snow accumulation and ablation, and infiltration and interflow would be considered a highly complex model. However, such a model may be run at a single, representative point in a basin (low spatial resolution) or at many points, which are then aggregated to give streamflow results at the basin outlet (high spatial resolution). An example of a low-complexity model may be one that simulates the hydrologic cycle as a series of storage "buckets" (e.g., canopy, snowpack, subsurface storage) between which water moves more or less instantaneously. Here, we focus on three major types of models, which we classify based on their overall characterization of the physical system: data-driven, physically based, and conceptual models. Each has unique data needs and exists within a subset of the space defined by the two axes of complexity and simulation scale (Figure I.1).

Data-driven models include traditional statistical regressions and newer machine-learning algorithms. These models do not attempt to identify causality between water-balance components, but instead to form associations that may be informative for forecasting or process understanding (Oyebode and Stretch, 2019; Schmidt et al., 2020; American Society of Civil Engineers, 2000). These models are low complexity and furthest removed from physical principles, but may therefore be easier to leverage, particularly if causal links between observed states or phenomena are not fully understood (Shen, 2018). Statistical modeling is one of the oldest forms of model used in the field of hydrology. For example, the practice of using SWE to predict streamflow volumes in California dates from the early 1900s (California Department of Water Resources, 2012). However, recent developments in the science of machine learning, coupled with an increase in available data, have renewed and expanded applications for this class of model (Shen, 2018; Schmidt et al., 2020). Machine-learning

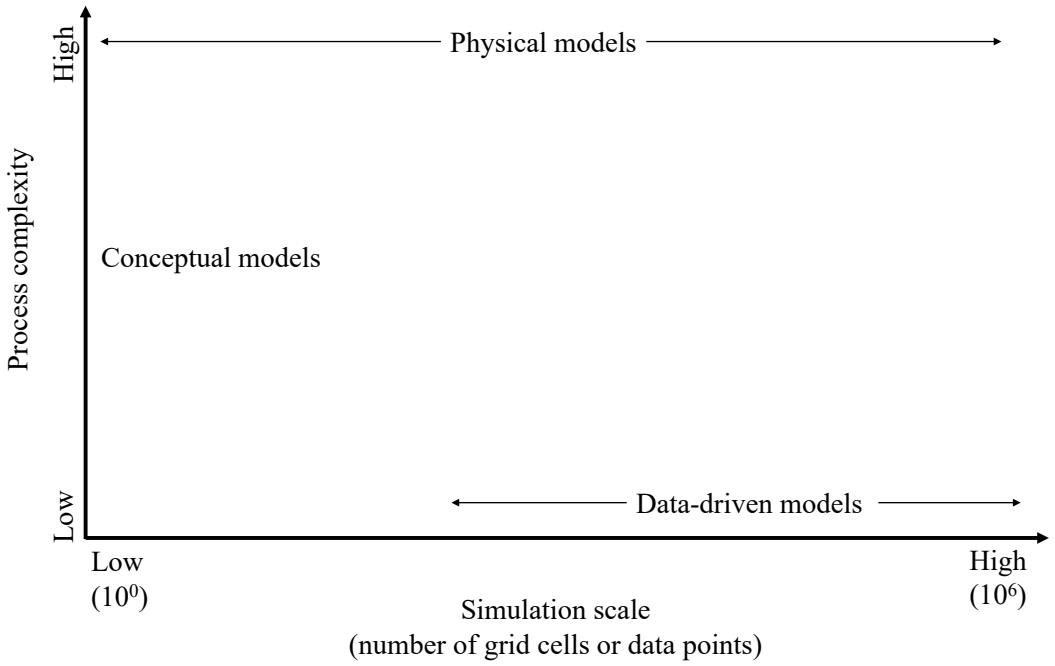


Figure I.1: Hydrologic model classification; adapted from Hrachowitz and Clark (2017).

techniques have the advantage of potentially identifying relationships that are not obvious or that would be too labor intensive to test individually (Shen, 2018). Machine-learning techniques have been applied in hydrology to analyze liquid water flow through snowpack (Avanzi et al., 2019), identify spatial patterns of soil moisture (Oroza et al., 2018), predict flood magnitudes (Schmidt et al., 2020), and forecast precipitation (Kim et al., 2020). Thus far, two of the most popular algorithms in hydrology are Random Forest and Artificial Neural Networks (Schmidt et al., 2020). The former uses an ensemble of decision trees with stochastic elements for regression or classification (Breiman, 2001), while the latter utilizes elements called neurons that exchange and process information via weighted connections to identify correlations between inputs and outputs (Oyebode and Stretch, 2019). Depending on the input data used, machine-learning algorithms can be applied at various spatial scales, ranging from point-scale measurements to spatially distributed data from remote sensing products. Regardless of spatial resolution, however, having a large number of data points is critical (Figure I.1), especially true for machine-learning algorithms, which rely on a high quantity of data to adequately characterize ("learn") the system in question (Hastie et al., 2009). Even with advances in data collection techniques, this can make these models inappropriate for ungauged or under-gauged areas or variables. Moreover, users must be careful that the training data are not biased in a way that may give an incomplete picture of the system, since most machine-learning algorithms are only capable of simulating the conditions present in the training data (Hastie et al., 2009). This also makes data-driven models susceptible to false assumptions of stationarity, since many relationships between hydrologic processes, including that between SWE and streamflow, are predicated on steady-state conditions, an assumption that is rapidly deteriorating due to climate change (Milly et al.,

2008). Understanding the ways that machine learning may be appropriate and useful for the field hydrology is an ongoing area of research (Shen, 2018; Schmidt et al., 2020; Oyebode and Stretch, 2019; Kim et al., 2020).

On the other end of the spectrum, high-complexity physically based models resolve mass and energy balances at various points across the landscape and are the closest to simulating the exact mechanisms controlling hydrologic processes. Usually, these models use point-scale simulations, which are then extrapolated to a broader spatial area ("continuum" model architecture; Hrachowitz and Clark, 2017). If processes are simulated at a single point, taken to be representative of the basin or region as a whole, the model is considered spatially "lumped." Conversely, distributed models simulate processes at multiple points, and outputs such as runoff, evapotranspiration, and other state variables are aggregated to give results at the basin scale (Lobligeois et al., 2014). Since fluxes are explicitly represented, physically based models are typically aimed at replicating catchment response through time based on modeled or observed timeseries inputs. Physically based models may simulate the entire water balance from precipitation through streamflow (see, e.g. the Precipitation-Runoff Modeling System (PRMS; Markstrom et al., 2015), the Soil and Water Assessment Tool (SWAT; see, e.g., Kalcic et al., 2015; Teshager et al., 2016; Qi et al., 2017), and the Regional Hydro-Ecological Simulation System (RHESSys; Tague and Band, 2004)) or focus on specific processes (e.g., the Sacramento Soil Accounting Model (SAC-SMA; National Oceanic and Atmospheric Administration, 2002), MODFLOW (Hughes et al., 2017; Langevin et al., 2017; Provost et al., 2017), SNOW-17 (Anderson, 1973, 1976), and SNOWPACK (Bartelt and Lehning, 2002; Lehning et al., 2002a,b)).

Though physical models aim to faithfully represent physical processes, thereby allowing users to pinpoint and predict the causes of observed phenomena, they also lean heavily on parameters. Parameters are used to represent processes that are not well understood and/or are too data-intensive to be fully characterized (for example, flow of water through snowpack or precipitation-canopy interactions) and to resolve issues of scale in moving from point simulations to basin-wide values (Beven et al., 1988; Beven, 1989; Wood et al., 1988; Blöschl and Sivapalan, 1995; Andréassian et al., 2012; Bai et al., 2009; Pianosi et al., 2015). This reliance on parameters can lead to overparameterization, where input data are not sufficient to adequately constrain the parameter space, and, as a result, the same output can be observed for multiple input parameter sets, an occurrence known as "equifinality" (Beven, 2006). Extensive literature examines how physically based simulations, many of which are based on point-scale process results, can be scaled across a landscape in ways that retain necessary complexity but avoid overparameterization (e.g. Khakbaz et al., 2012; Reed et al., 2004; Lobligeois et al., 2014; Sivapalan and Kalma, 1995; Tran et al., 2018; Hundecha et al., 2016; Samaniego et al., 2010). Overcoming these scaling issues has had mixed success and is an active area of research.

Physically based models use data for direct inputs, to establish parameter values, and for validation. Parameter values may be set manually (for example, percent forest cover derived from vegetation density maps) or determined during calibration, which relies on validation data to assess how a given parameter set impacts model outputs. For spatially distributed physical models, these data must be available at multiple points throughout the landscape in order to define conditions at each simulation point. The length of data record used for calibration depends on the temporal and spatial extent of the model and the

hydrologic process being modeled, but increases with greater climate variability (Gan and Biftu, 1996; Blöschl and Sivapalan, 1995). As a result, previously unobserved extreme climate conditions are challenging the usefulness of these models (Milly et al., 2008), particularly when they have been calibrated to maximize average performance over a long period. More frequent model updates and recalibrations may help in this regard but are often beyond the capacity of operational users with limited time and human resources. Other options may include different model versions, each calibrated to specific conditions (for example, having dry, average, and wet year parameter sets for precipitation-runoff models in California). Adapting models for a climate change regime is an active research area (Troch et al., 2015; Kirchner, 2006).

Finally, conceptual models simulate the water balance based on the major processes governing water allocation, such as evaporative demand and precipitation rate (Hrachowitz and Clark, 2017). They represent a medium complexity between data-driven and physically based alternatives. Instead of scaling up, as is typical with physical models, conceptual models are used to examine average conditions or overarching drivers, meaning that they operate on larger spatial and longer temporal scales at which less dominant controls on the hydrologic process can be ignored. These models can be used to understand broader competing forces that impact the partitioning of water in the critical zone. They are typically "bucket" models (Hrachowitz and Clark, 2017), in which each major reservoir of water in the system is treated as a homogeneous receptacle. While they may employ mass conservation principles, their governing equations are derived to mathematically meet expected behavior under limit conditions and are otherwise not reflective of physical water fluxes. Conceptual models usually rely on no more than a handful of parameters, thus limiting equifinality problems, but may present limitations in identifying causality between observed changes in the water balance and specific physical mechanisms. Because they do not explicitly represent most physical processes, conceptual models are sometimes inappropriate at smaller spatial or temporal scales, such as small catchments (Donohue et al., 2007) or sub-annual timesteps (Du et al., 2016). Conceptual models include the Budyko framework, which conceives of the long-term water balance as a trade-off between supply of water (precipitation) and vegetation demand for water (potential evapotranspiration; Budyko, 1974); the Soil Conservation Service (SCS) curve number, which predicts runoff from specific rainfall events based on a proportionality hypothesis between potential runoff and potential "abstracted" water (water going to infiltration, canopy interception, and ongoing soil wetting; Soil Conservation Service, 1972); and the *abcd* water-balance model, which predicts daily, monthly, or annual ET, soil storage, groundwater, and runoff assuming evaporation opportunity as nonlinear function of available water (Thomas, 1981).

Conceptual models use data similarly to physical models, as direct inputs and to validate calibration of any parameters. Since the number of free parameters is usually much lower in conceptual models, the information content of data is more likely to fully constrain them (see, e.g. Fu, 1981). Data may also be used in conjunction with conceptual models to relate model outputs or parameters with physical characteristics or processes. For example, the free parameter ω in the Fu (1981) and Zhang et al. (2004) Budyko formulation has been related to non-climate basin characteristics including percent forest cover, average slope, and soil characteristics like relative infiltration capacity and relative water storage (Zhang et al., 2001, 2016; Yang et al., 2007; Jaramillo et al., 2018) through the use of additional

data.

Despite their importance both scientifically and operationally, the capacity of all hydrologic models to adequately represent and forecast the natural system is being subjected to greater and greater stressors (Troch et al., 2015; Milly et al., 2008; Kirchner, 2006; Hrachowitz and Clark, 2017). Climate change, population growth, and land-use changes have created previously unobserved conditions and needs. Other system changes are desirable from a long-term resiliency standpoint, but will stress water management systems; for example, increased renewable energy generation is projected to induce more uncertainty in electricity prices, thereby constraining hydropower operational flexibility (Seel et al., 2018). These changes push the boundaries of scientific understanding, especially when they impact hydrologic processes that are already not fully understood (for example, the climate elasticity of evapotranspiration; see Avanzi et al., 2020b). Furthermore, they reduce the acceptable margin of error in modeling results for water resource operators. Fortunately, new and more robust data sources have emerged alongside these challenges, providing an opportunity to increase scientific process understanding, ability to represent those processes in hydrologic models, and operational capacity to meet the water-allocation challenges of the twenty-first century.

Data in hydrologic modeling may be broadly classified as either ground based or remotely sensed (Zhang et al., 2017b), both of which have seen critical innovations in recent decades. Ground-based data are derived from sensors or manual measurements that are co-located with the phenomena being measured: temperature and relative humidity sensors, ultrasonic or laser snow-depth sensors, precipitation gauges, anemometers, lysimeters, snow pillows, and soil moisture sensors. Traditionally, these sensors have been relatively few in number and more heavily concentrated in montane regions at lower elevations and flat, open sites (Bales et al., 2006; Molotch and Bales, 2005). Improvements in chips and data storage have enabled expansion of ground-based data sources, with measurements now often available at high temporal resolutions (as frequently as every fifteen minutes or less). Furthermore, wireless connectivity has increased the flexibility and availability of ground-based measurements, with wireless-sensor networks allowing for data collection across a variety of vegetation and topographic conditions and making the data available in real time via cell or satellite signal (Zhang et al., 2017a). Ground-based sensors are susceptible to a number of challenges that may impact data quality, including environmental stressors, such as severe climate effects or animal interference; accessibility of the site, which may limit ability to collect information under certain conditions or perform maintenance; and data reliability, since connectivity issues may prevent the data from being broadcast in real time (Zhang et al., 2017b).

Remotely sensed data are measured from airborne or satellite sensors and are among the new data innovations helping to advance hydrologic modeling. These datasets include the Jet Propulsion Lab's Airborne Snow Observatory data (Painter et al., 2016), NASA's LandSat (U.S. Geological Survey, 2016) and MODIS (Friedl and Sulla-Menashe, 2019) missions, and the European Space Agency's Sentinel missions (Copernicus, 2020). Products derived from these satellite missions include a range of variables including climatic data (precipitation and temperature), water balance data (such as SWE), and land cover and vegetation data (such as Normalized Difference Vegetation Index and Leaf Area Index). These data frequently cover large spatial extents, anywhere from river basins to global extent. Spatial resolution of these products varies from the order of tens to hundreds of meters, while temporal resolution

can vary from a few days to weeks. Remote sensing instruments face different data quality issues than ground-based sensors: satellite readings may be incomplete or less reliable due to cloud cover and products may not be available on as fine a temporal resolution as ground-based data. While connectivity is less of an issue with remote sensing products, some data may require post-processing that also limits their real-time use.

Expansion of both ground-based and remote-sensing products is a valuable development and critical to fully leveraging all types of hydrologic models. Data can enable greater understanding of poorly characterized processes; for example, wireless-sensor networks placed in montane regions of the Sierra Nevada across a variety of canopy and topographic conditions have helped identify drivers of soil storage at different times of year (Oroza et al., 2018). Since traditional ground-based sensors are biased towards flat, open areas, these sensor networks also provide a more comprehensive and less biased understanding of landscape conditions (Avanzi et al., 2018). Other datasets provide more information at finer spatial and temporal time scales, enabling and/or expanding use of models in ways that were not previously possible. For example, satellite products have enabled the creation of spatial SWE maps across large portions of California and the Western U.S. (Margulis et al., 2016; Molotch and Margulis, 2008), which can help validate the snow simulations in physically based models. Parameter-elevation Relationships on Independent Slopes Model (PRISM) climate data provide, among other products, improved precipitation and temperature data, particularly in areas with sparse ground measurements (PRISM Climate Group, 2004). While even new data sources are not sufficient in quantity or scope to make highly parameterized physically based models fully identifiable, they can lessen the problem by constraining key processing (for example, through higher information content or multi-objective calibration; Avanzi et al., 2020a; Hay and Umemoto, 2006; Hay et al., 2006) or by eliminating the need for some parameters if processes can be more fully characterized.

Understanding how to best leverage these new data is critical to confronting the challenges facing hydrologists and water resources managers. This dissertation explores how novel data sources, both ground-based and remotely sensed, can be coupled with models to improve hydrologic forecasting in snow-dominated montane headwater catchments in California. California is uniquely suited to exploring improvements in hydrologic modeling, as the region's water supply challenges reflect critical needs world-wide: Mediterranean climates like California's are found on all continents except Antarctica and cover major population centers such as Rome, Barcelona, Jerusalem, Los Angeles, Santiago, Cape Town, and Perth, while over half the world's population relies on montane regions for water supply (Mountain Partnership, 2014; Immerzeel et al., 2020; Viviroli et al., 2007a). Each chapter of this dissertation addresses a different type of hydrologic model and leverages one or more novel data sources. The unprecedented quantity of ground-based data available through wireless-sensor networks is used in Chapter II to drive a machine-learning algorithm exploring the response of the snowpack to rain-on-snow events. Chapters III and IV both use a variety of spatial maps derived from remote sensing products and calibrated to ground-based data. In Chapter III, these maps characterize basin features that influence runoff production, allowing for more efficient identification of the optimal modeling locations for a spatially distributed, physically based precipitation-runoff model. The methods in Chapter III also incorporate a machine-learning algorithm, the Gaussian Mixture Model, thereby exploring ways model types can be coupled to overcome the limitations of each. Finally, Chapter IV uses newly

available spatial maps in conjunction with a conceptual water balance model to explore shifts in the precipitation-runoff relationship during droughts and quantify the impact of predictable versus unpredictable shifts on runoff production. Broadly, I aim to explore the following questions: 1) Do novel data sources improve accuracy of existing hydrologic models and/or enable use of new models?; 2) What are the opportunities and limitations of novel data sources for data-driven, physically based, and conceptual water balance models?; 3) What are the implications of the relative strengths and weaknesses of each of these model types for scientific understanding and operational water management?

Chapter II

Drivers of snowpack response to rain-on-snow events in forested terrain

Based on the paper of the same name submitted for publication by:
Tessa Maurer, Francesco Avanzi, Safeeq Khan, Steven D. Glaser, Roger C. Bales, Martha Conklin

Abstract

Interactions between canopy and snowpack during rain-on-snow (ROS) events are poorly understood, particularly in complex, forested terrain. Seven wireless-sensor networks with 74 measurement sites in the southern Sierra Nevada of California were used to investigate the response of snowpack to rain events. We identified a total of 496 ROS events that occurred across a range of topographic and canopy conditions over a ten-year period (2008–2017). Leveraging climate, vegetation, topographic, and timing data of the events, we trained a Random Forest algorithm to predict change in snow depth during ROS events. The model performed well across all validation events (R^2 = 0.87, RMSE = 3.91 cm, bias = 2.97 cm, Kling-Gupta Efficiency = 0.72), allowing us to identify four drivers of snow-depth change. First, the order of precipitation phase (rain before snow or snow before rain) in mixed rain-and-snow ROS events causes different responses in terms of the sign of snow-depth change. Second, variability in year-to-year timing of snow accumulation and ablation makes prediction more uncertain in the middle part of the water year. Third, there is a nonlinear relationship between Leaf Area Index and snow-depth change during ROS, supporting more-quantitative characterization of forest structure than simply "open" and "closed" in ROS studies. Finally, the impact of vegetation on snow-depth response shifts from being both interception- and energy-exposure-related early in the season to being dominated by energy-exposure patterns at the end of the accumulation season. This work more fully characterizes the impact of precipitation phase and vegetation patterns on ROS response of the snowpack, providing a foundation for improving predictive modeling of these events in forested regions.

II.1 Introduction

Rainfall on a pre-existing snowpack, known as rain-on-snow (ROS; Harr and Berris, 1983), affects montane regions in the rain/snow transition zone across continents (Musselman et al., 2018; Würzer et al., 2017). ROS has been observed in the European Alps (Würzer and Jonas, 2018; Garvelmann et al., 2015; Wever et al., 2014; Rössler et al., 2014), the Canadian Rockies (Pomeroy et al., 2016a), the Pacific Northwest of the U.S. (Safeeq et al., 2015; McCabe et al., 2007; Marks et al., 1998; Berris and Harr, 1987), and California's Sierra Nevada (Musselman et al., 2018; McCabe et al., 2007; Kattelmann, 1997; Berg et al., 1991). These storms may result in extreme and dangerous flood events (e.g., Pomeroy et al., 2016a; Garvelmann et al., 2015; Wever et al., 2014; Rössler et al., 2014; McCabe et al., 2007; Marks et al., 1998; Leathers et al., 1998; Kattelmann, 1997), including the February 2017 storm that prompted the evacuation of 190,000 people from below Oroville Dam in northern California (Musselman et al., 2018) and the June 2013 storm that became the costliest natural disaster in Canadian history (damages exceeding USD$6 billion; Pomeroy et al., 2016b). However, modeling and forecasting precipitation and runoff during these events continues to prove challenging (Fehlmann et al., 2019; Würzer and Jonas, 2018; Pomeroy et al., 2016a; Rössler et al., 2014; Marks et al., 1998).

In particular, the impact of vegetation on snowpack response to ROS is an important knowledge gap (Garvelmann et al., 2015; Wayand et al., 2015). Vegetation influences snowpack conditions generally through interception, sheltering from wind and shortwave radiation, and longwave-radiation enhancement (Broxton et al., 2015; Revuelto et al., 2015; Varhola et al., 2010; Veatch et al., 2009). However, these effects are not consistent across climates (Lundquist et al., 2013). During ROS, turbulent fluxes leading to condensation of moisture on the snow surface are often identified as the most significant energy input to the snowpack (Harr, 1981; Marks et al., 1998), but radiation may become the primary driver of snowmelt if wind speeds are low (Mazurkiewicz et al., 2008). These storm-specific energy dynamics affect the role of vegetation, which will tend to mitigate the former and increase the latter (Garvelmann et al., 2015; Marks et al., 1998; Mazurkiewicz et al., 2008). While energy inputs from the rain itself are not considered significant (Harr, 1981), interception by vegetation may also change the mass and energy balance of the snowpack during ROS.

Previous ROS studies have reported mixed findings with regard to the influence of canopy cover. Using an energy-balance model and data from paired clear-cut and forested sites in western Oregon, (Harr, 1981) concluded that clear-cut logging would increase snowmelt (i.e., loss of Snow Water Equivalent, SWE) between 8% and 25% compared to a forested site. Berris and Harr (1987) also found greater melt in the clear-cut plot, attributing this to the greater exposure of the open site to wind and thus turbulent heat transfer. In another more-recent paired-site study, Marks et al. (1998) used an energy balance model to simulate snowmelt during an ROS-induced flood event at open and forested locations. The model showed significantly less melt under canopy than in the nearby open site, which was due to lower turbulent fluxes (rather than lower snow accumulation) at vegetated sites. On the other hand, Garvelmann et al. (2015), Berg et al. (1991), and Kattelmann (1987) found little difference between SWE reduction at open and forested plots as the result of ROS. This discrepancy in the literature is discussed by Garvelmann et al. (2015), who noted that small clearings and large meadows may both be considered "open" by researchers, but will

be subject to different snow-accumulation and energy-balance patterns. They further note that the difference in snow-cover patterns between open and canopied sites is exacerbated later in the season, so timing of the event(s) being analyzed may lead to different conclusions. More recently, Wayand et al. (2015) examined the impact of basin hypsometry and vegetative cover during ROS events using a physically based hydrologic model and concluded that forest cover reduced basin-wide snowmelt compared to the same basin covered in grass. However, the magnitude of that reduction depended on storm and initial snowpack characteristics.

Albeit promising, existing work on canopy-snow interactions during ROS is mostly site- or storm-specific. Analyses covering a broad spectrum of forest structures and storm properties across an extended time period are needed to better understand the response of snowpack to ROS across the landscape. In regions where forest patterns are changing, whether because of rising temperatures (Cannone et al., 2007), wildfires (Boisramé et al., 2017), or forest thinning (Saksa et al., 2017), improved understanding of this interaction can better quantify long-term effects on water and energy supply. More immediately, this work has implications for reservoir managers and flood forecasters (Marks et al., 1998; Wayand et al., 2015).

The California Sierra Nevada typifies conditions that lead to predictive uncertainty of ROS events. The region is usually snow covered during the winter and spring (December through May), but year-to-year variations in precipitation levels and snow-line elevation mean that peak SWE can change significantly (Margulis et al., 2016). ROS and related flooding are a regular occurrence in the Sierra Nevada, which is prone to warm storms even during the winter months (Guan et al., 2016; Kattelmann, 1997; Musselman et al., 2018). Finally, the Sierra Nevada is densely forested (Knapp et al., 2013; Miller et al., 2012), which makes it appropriate for examining the impact of canopy during ROS. However, the time and spatial scales at which ROS impacts the snowpack in forested regions cannot be sufficiently captured with operational sites currently used for forecasting, which are usually located in open areas (Garvelmann et al., 2015; Wayand et al., 2015).

We can revisit this question using recent advances in hydrologic monitoring, namely distributed wireless-sensor networks (WSNs) and LiDAR data, which provide finer-resolution climate data across a range of vegetation characteristics. In this paper, we combine the WSN and LiDAR data with a Random Forest regression algorithm to predict changes in snow depth across ten years of ROS storms in the southern Sierra Nevada. Data-driven models like Random Forest can overcome many of the limitations of traditional statistical techniques (Hastie et al., 2009), such as assumptions of linearity and normality of residuals in multilinear regression, and have been increasingly applied in hydrologic contexts (see, e.g., Wang et al., 2015; Tinkham et al., 2014; Snelder et al., 2013). We focus on snow depth as a response variable (following McCabe et al., 2007) since it is widely measured by WSNs and can easily be assessed in a variety of canopy conditions and at fine temporal resolution, unlike SWE. The primary aim of this paper is to characterize drivers of snowpack response to ROS in montane regions with complex physiographic features. Using the southern Sierra Nevada of California as our study area, we address three questions. First, what is the relative importance of weather and snow conditions, including the mix of rainfall and snowfall during a storm, to snowpack response to ROS events? Second, how does the response of the snowpack to these weather conditions change based on local physiography? Third, to what extent does canopy modulate or exacerbate the impact of weather on the snowpack during ROS events, and is this effect consistent across and throughout seasons?

II.2 Methods and data

This section presents an overview of the study sites and data (Sections II.2.1 and II.2.2), followed by methods for identifying ROS events and designing the Random Forest model (Sections II.2.3 through II.2.6).

II.2.1 Study sites and LiDAR data

The Sierra Nevada (Figure II.1) has a Mediterranean climate, with strong seasonal precipitation (October–May is the normal wet season, with June–September dry). Snowpack falls between alpine and maritime according to Sturm and Holmgren's classification system (1995). At low elevations, small melt events occur throughout the winter season, but higher in the range the snowpack is typically spatially continuous from the early winter through April, with the exception of very low-snow years. Historically, April 1st has been taken as the day of peak SWE (Montoya et al., 2014).

All seven WSNs used in this study are operated through the Southern Sierra Critical Zone Observatory (SSCZO). Upper Met (UPM), Lower Met (LOM), and P301 were installed within the Providence Creek catchment in the Kings River (Bales et al., 2013) while SEKI 1, SEKI 2, SEKI 3, and SEKI 4 are located in the Wolverton catchment of the Kaweah River (Bales et al., 2018b, 2017). These networks provide spatially dense climate and snowpack data across each site, including measurements taken under canopy, at the canopy drip edge, and in clearings. Henceforth, each measurement location will be referred to as a "node." These sites were selected for the availability of data and the fact that they span a range of elevations from the rain/snow transition up to the snow-dominated Wolverton catchments (Table II.1; Bales et al., 2018b).

High-resolution LiDAR point cloud data, collected during snow-off conditions between August 5 and 15, 2010, were converted to rasters and used to characterize forest structure (Leaf Area Index or LAI, canopy height, and open fraction) and topographic characteristics (elevation, slope, and aspect) at each node. LAI is defined as half of the leaf surface area projected on a unit surface area and open fraction is a measure of the proportion of each pixel that is vegetation-free. Each feature was calculated at 1-m and 10-m resolution to capture processes that may be important at either or both scales with the exception of LAI, for which the density of the LiDAR point cloud was deemed insufficient for calculating 1-m values (Korhonen et al., 2011; Ma et al., 2017). Details on processing can be found in Appendix A.1.

II.2.2 Climate data

Climate data included temperature, relative humidity, precipitation, and wind measurements. Dewpoint temperature was calculated using temperature and relative humidity data following (Zhang et al., 2017a). In addition, storms were classified as related to atmospheric rivers using data from the Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA 2; Rutz et al., 2014). Details on climate data processing can be found in Appendix A.1.

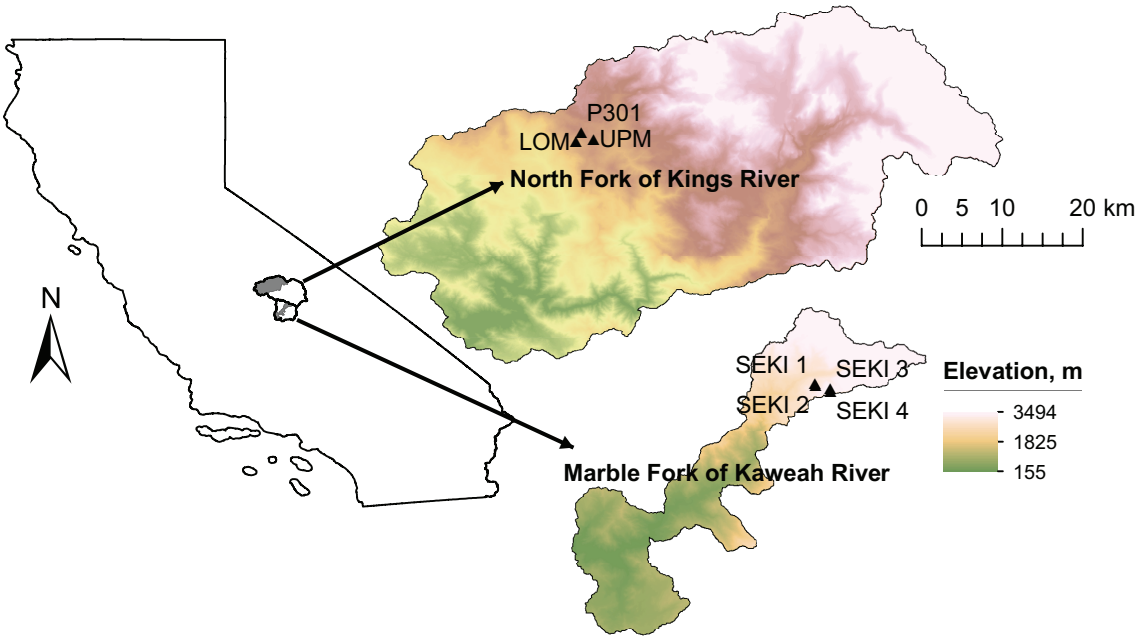


Figure II.1: Map of California indicating the Kings and Kaweah River basin extents with relevant subbasins shaded. Inset shows site locations along the North Fork of the Kings River and the Marble Fork of the Kaweah River.

Where necessary, WSN data were filled or supplemented with climatic data from other sources (Table A.2). Supplemental sites were chosen for their proximity and/or similar elevation to the WSNs. All data were hourly unless otherwise noted in Appendix A.1. Data were processed to the Level 2 hourly timestep. This included removing nonphysical outliers and masking data to physical ranges. Snow-depth data were smoothed using a 24-hour moving-average window to remove fluctuations due to sensor temperature changes (commonly known as "flutter"; Avanzi et al., 2014). An overview of each WSN cluster is presented in Table II.1.

II.2.3 Identifying ROS events

ROS events have been defined differently in the literature, depending on the focus of study (Würzer et al., 2016). We identified events on a node-by-node basis, since weather and snow conditions may vary within a WSN. For example, the same storm may create ROS conditions at an open location with significant snowpack but not at a location under canopy with little to no snowpack. Since one of the aims of this paper was to assess a broad range of conditions, we also aimed to identify majority-rain events without excluding mixed rain-and-snow periods (Berg et al., 1991).

We used criteria related to precipitation amount, snowpack depth, and minimum time period of non-ROS conditions triggering the end of an event. We also imposed restrictions based on dewpoint temperature since precipitation phase was not known *a priori*. First, each

hourly timestep was analyzed separately and designated as an ROS hour if precipitation and dewpoint temperature were greater than zero and snow depth was greater than 10 cm. Any period between ROS hours shorter than four hours was also designated as ROS-related to ensure that ROS conditions arising from the same storm were grouped together. Each continuous period of ROS conditions was then treated as an ROS event. Events were subject to a minimum length (8 hours) and precipitation amount (10 mm). These thresholds are in the range of those used by Berg et al. (1991); Mazurkiewicz et al. (2008); Musselman et al. (2018); Würzer and Jonas (2018), though we note that much existing literature has focused on events lasting at least 24 hours. In some cases, the studies focused on daily time steps (e.g. Musselman et al., 2018), while in other cases the authors used longer time periods to capture subsequent snowpack runoff after precipitation ended (e.g. Würzer et al., 2017). We chose to not exclude shorter events, using a minimum precipitation threshold to ensure that total event precipitation was not negligible. We also chose not to automatically eliminate periods of increasing snow depth in order to include periods of mixed rain-and-snow precipitation. However, we discarded all-snow events, defined as those with monotonically increasing snow depth in which the measured snowfall, converted to SWE (using a standard assumption of 100 kg/m^3 for the density of newly fallen snow; see, e.g., Rössler et al., 2014), was greater than or equal to 90% of the measured precipitation. Finally, we performed a visual scan to identify events affected by noise or other obvious data-quality issues (e.g., a long event with only a couple hours of decreasing snow depth at the beginning or end). Twenty-two events (less than 5%) were removed, giving a total of 496 ROS events across the seven WSNs (Table II.2).

To verify that we identified primarily rain events but did not exclude mixed events, the percent of precipitation that fell as rain during each event was estimated following Zhang et al. (2017a). To analyze the impact of phase timing during ROS events, we categorized mixed events by whether rain fell before snow ("rain-first"), snow fell before rain ("snow-first"), or there was no appreciable difference in the timing of the phases ("coincident-timing"). After categorizing precipitation phase on an hourly basis during ROS events, we calculated the center of mass of snow and rain in hours from the start of the storm. Storms in which the center of mass of snow and rain fell within one hour of each other were considered coincident. Otherwise, storms were considered rain-first or snow-first, collectively referred to as "transitional" storms henceforth.

II.2.4 Random Forest models

The Random Forest (Breiman, 2001) algorithm uses an ensemble of decision trees to predict a target output variable, which may be a category (classification) or, as in this study, a numeric value (regression). Random Forest has several advantages for hydrologic applications. Unlike multivariate regression, it allows for nonlinear regression with arbitrarily complicated decision boundaries. It is relatively interpretable among machine-learning methods and can thus be used for process understanding as well as prediction. Finally, it is also relatively robust to irrelevant features, meaning that such features will not significantly increase model variance, and can be used with both categorical and quantitative features. Each decision tree in a Random Forest uses a series of binary decisions to arrive at a prediction. The trees are trained using a subset of observations, and features are randomly subsampled when training

Table II.1. Summary of wireless-sensor networks

Name	Code	Average elevation, m	Longitude & latitude, degrees	Water Year record	Number of nodes†	Number of ROS events
Lower Met	LOM	1735	-119.20, 37.06	2008–2017	10	166
Upper Met	UPM	1979	-119.18, 37.06	2008–2017	17	163
P301 network	P301	1997	-119.20, 37.07	2010–2016	23	41
SEKI 1	SK1	2226	-118.73, 36.59	2008–2016	6	27
SEKI 2	SK2	2261	-118.73, 36.59	2008–2016	6	52
SEKI 3	SK3	2596	-118.72, 36.59	2008–2016	6	22
SEKI 4	SK4	2639	-118.72, 36.59	2008–2016	6	25

† Number of nodes is total number of sensing locations considered for this study, including those that did not experience any rain-on-snow (ROS) events.

Table II.2. Summary of ROS events by site

Site	Number of events per node per WY	Event length, hr average (max, min)	Event precipitation, mm average (max, min)	Events per season, % Nov–Dec	Jan–Mar	Apr–May
LOM	1.7	41.6 (97, 12)	20 (89, 10)	13.9	83.1	3.0
UPM	0.95	38.3 (73, 9)	27 (89, 10)	24.5	68.7	6.7
P301	0.25	28.3 (64, 8)	33 (89, 10)	70.7	24.4	4.9
SEKI 1	0.5	29.9 (54, 15)	20 (39, 12)	0.0	66.7	33.3
SEKI 2	0.96	20.7 (43, 8)	19 (64, 10)	0.0	69.2	30.8
SEKI 3	0.41	29.9 (74, 8)	39 (76, 10)	18.2	54.5	27.3
SEKI 4	0.46	30.2 (54, 10)	38 (78, 12)	20.0	56.0	24.0

each internal decision. These bootstrapping (or "bagging") mechanisms help reduce bias in the model. The results of all decision trees are averaged to produce the final results. The Random Forest regression algorithm was implemented in Python using Scikit-learn version 0.20.3 (Pedregosa et al., 2011). We implemented a five-fold cross validation on the RandomForestRegressor class during training to tune the following metaparameters: number of trees (between 200 and 500 in increments of 50); maximum tree depth (9 to 30 in increments of 2); maximum number of features (n) considered at each split (we tested n, n divided by three, the square root of n, and the binary logarithm of the n). After training and cross-validating, we verified that the optimal metaparameter values fell within the search range.

For the primary regression analysis, the node-events were divided into testing and training sets using a randomized 80%/20% split. The predictions of the best-performing model after cross validation were assessed using root mean square error (RMSE), absolute bias, and the modified Kling-Gupta Efficiency (KGE; Gupta et al., 2009; Kling et al., 2012). These metrics were calculated for both calibration and validation data to check for overfitting (indicated by vastly better calibration than validation results). In addition, three other Random Forest models were trained following the primary analysis, but with a subset of all features (see Section II.2.5). These models were used to assess the interactions and marginal improvement attributable to different groupings of features.

In addition, results from ten Random Forest models trained on all events except those from a particular water year (2008–2017) are presented in Appendix A.1.4.

II.2.5 Feature selection

Features considered for the Random Forest model (Table II.3) can be broadly classified as pertaining to storm conditions; timing and seasonality of the storm event; and physiographic site characteristics (topographic and vegetation). Features were developed based on established factors influencing runoff and snowpack during ROS events (see e.g., Berg et al., 1991) as well as visual observations of dewpoint temperature, snow depth, and precipitation.

Storm conditions, represented by snow and climate features, were selected to capture mass (e.g., total precipitation) and energy (e.g., temperature and wind features) inputs to the snowpack. Average wind speed was used as a proxy for energy from turbulent fluxes. Average dewpoint temperature provides a benchmark of precipitation phase, while standard deviation, coefficient of variation, and autocorrelation were calculated to capture how phase might shift during an event. In addition, we observed through visual screening that periods of increasing snow depth frequently exhibited relatively stable temperature conditions. In contrast, during periods where snow depth was constant or decreasing (suggestive of rain rather than snow), dewpoint temperature was more variable. We therefore included as a feature the length of the longest period during which dewpoint temperature changed by less than half a degree (i.e., the longest period of stable dewpoint temperature). Finally, we included a binary feature classifying each event as an atmospheric river or not based on Rutz et al. (2014)'s atmospheric river catalogue (`http://www.inscc.utah.edu/~rutz/ar_catalogs/merra_0.5/`). Atmospheric rivers are characterized as long, narrow plumes carrying high concentrations of water vapor, often warmer than arctic storms and with the potential to generate extreme conditions (Guan et al., 2016; Lamjiri et al., 2018). Details on data processing can be found in Appendix A.1.

Table II.3. Random Forest features by category[†]

Climate & snow
Initial snow depth
Storm length
Total precipitation
Average temperature*
Change in air temperature
Average wind speed
Average dewpoint temperature
Standard deviation of dewpoint temperature
Coefficient of variation of dewpoint temperature
Autocorrelation of dewpoint temperature
Maximum duration of stable dewpoint temperature
Atmospheric river or not
Topography & vegetation
Aspect (10 m)
Aspect (1 m)
Canopy height (10 m)
Canopy height (1 m)
Elevation (10 m)*
Elevation (1 m)
Leaf Area Index (LAI) (10 m)
Open fraction (10 m)
Open fraction (1 m)
Sheltering coefficient
Slope (10 m)*
Slope (1 m)
Event timing
Day of water year (DOWY)
Number of previous events
Start time

[†] Features with an asterisk were removed prior to training due to high correlation with other features.

Timing features included storm start date as day of water year (DOWY; days since the previous October 1), number of previous ROS events that season, and start time by hours since midnight. The DOWY on which the event began was included due to the strong seasonality of snow accumulation and ablation patterns in the Sierra Nevada (Guan et al., 2013; Margulis et al., 2016), which has implications for the condition of the snowpack, including depth, cold content, and liquid-water content. Cold content and liquid-water content relate to the energy deficit and retention-storage capacity of the snowpack, respectively; in other words, they relate to the ability of the snowpack to buffer energy and mass inputs (Avanzi et al., 2015; Garvelmann et al., 2015; Jennings et al., 2018; Wever et al., 2014). The number of previous ROS events that season was included to capture the possible effects of repeated rainfall on the snowpack, such as the formation of preferential flow paths, creation of deep ice layers, and increased densification of the snowpack that may exist in alpine/maritime snow regions like the Sierra Nevada (Avanzi et al., 2017; Sturm and Holmgren, 1995). The hour of the day at which the storm started aimed to capture any diurnal effects on storm conditions. While incident shortwave radiation is generally negligible during ROS events due to cloud cover (Mazurkiewicz et al., 2008), net shortwave radiation during daytime storms may compensate for negative net longwave radiation (Garvelmann et al., 2015; Würzer et al., 2016).

Topographic features include elevation, slope, and aspect, all of which are known to influence the distribution and condition of the snowpack (Oroza et al., 2016). They may also influence storm conditions by changing exposure to turbulent energy fluxes and precipitation (e.g. through rain shadowing). Vegetation features were LAI, canopy height, open fraction, and a "sheltering coefficient." LAI indicates the extent to which a site is sheltered from direct incoming precipitation. Both canopy height and open fraction are related to how exposed a site is to incoming precipitation and energy fluxes. The sheltering coefficient, defined as the ratio between open fraction at 1-m and 10-m resolutions, was intended to capture the interactions between vegetation coverage and turbulent fluxes, a key snowmelt factor during ROS (Harr, 1981; Marks et al., 1998). This interaction depends not only on a one-dimensional feature like open fraction but also on how forest patterns vary across the landscape. For example, small clearings may see greater SWE than both nearby large open meadows and densely vegetated sites since they are more sheltered from turbulent fluxes than meadows but subject to lower interception rates than vegetated areas (Garvelmann et al., 2015). In a spatially homogeneous forest, the sheltering coefficient tends to 1; in areas where the forest density increases (decreases) with the distance from the node, it tends to infinity (zero). The 0 / 0 limit case (open on both the 1-m and 10-m scale) was flagged with -999 because it is conceptually different from the case of a homogeneous forest.

Of the three Random Forest models trained on feature subsets, the first used climate features only, expected to be the primary drivers of snowpack response (Marks et al., 1998), to provide a point of comparison for determining if and to what extent vegetation and timing features added new information to the model. The other two models coupled climate features with each of other feature groups (physiographic features and timing features; Table II.3).

II.2.6 Feature importance and correlation

Feature importance, which can aid in process understanding, was calculated based on the mean decrease in impurity across nodes and regression trees (Breiman et al., 1984). Feature importance calculations can be impacted by correlated features (Gregorutti et al., 2013). Once the model selects a given predictor, the predictive power of all correlated features is reduced and a strong feature may score poorly on a feature-importance test. Therefore, to avoid spurious feature-importance results, variable-inflation factors were computed for each feature and correlation coefficients computed for each pair of features. Variable-inflation factors indicate how much the variance of the regression coefficients is inflated due to multicollinearity between features. Features that are starred in Table II.3 were removed before running the Random Forest algorithm to ensure that all variable inflation factors were below four, a slightly conservative threshold (five is a commonly used cutoff; see Čeh et al., 2018). The decision of which of two correlated variables to keep is necessarily subjective; in this analysis we kept the features we believed to be most informative. An alternative Random Forest model was trained keeping the starred features in Table II.3 and dropping the other correlated features (average dewpoint temperature and slope and elevation at 10-m resolution). No major differences in performance were detected (Appendix A.1.1).

II.3 Results

II.3.1 Climatology of ROS events

An examination of the climate, timing, and physiography data suggests that climate factors most strongly influence snowpack response to ROS events. Climate features are more highly correlated to change in snow depth than features in other categories (Figure II.2). Average air and dewpoint temperature are negatively correlated with the change in snow depth, as colder temperatures are more likely to produce snowfall. Snow-depth change increases with increasing wind speed and total precipitation, suggesting that mixed-phase ROS events, which may result in positive snow-depth change, also tend to see higher wind speeds and total precipitation.

Storm length is moderately negatively correlated with change in snow depth. The longer storms may result in greater precipitation, but since greater precipitation is positively correlated with snow-depth change, it is possible that long storms also tend to be rainier and thus more likely to decrease snow depth. The negative correlation between snow-depth change and DOWY may be due to more precipitation falling as rain in the later part of the season as well as to snowpack that is isothermal and more prone to melting. Both conditions would make snow-depth increases less common.

Unlike storm and timing characteristics, there were no strong correlations between physiographic features and change in snow depth. However, the distribution of events with respect to each feature (Figure A.2 in Appendix A.1.2) shows patterns of decreasing frequency of ROS events with increasing canopy height and open fraction. Lower canopy heights may more frequently create ROS conditions due to greater snow accumulation. Open fraction, while negatively correlated to LAI and canopy height, also showed a greater frequency of ROS events at low values. While more events at lower open fraction may be partly due to the

prevalence of vegetation in the Sierra Nevada (areas that are mostly clear of any vegetation are less common), this is also evidence that the presence of vegetation alone may not limit ROS conditions. Though only one-dimensional relationships are presented here, these findings tally with previous, moderately successful efforts using multilinear regression to relate climate and snow characteristics with snowpack runoff during ROS events (e.g. Berg et al., 1991).

II.3.2 Performance of Random Forest model

The cross-validated Random Forest model performed well across all validation events (RMSE = 3.91 cm; absolute bias = 2.98 cm; KGE = 0.72, R^2 = 0.87; Table II.4). The model slightly underestimates the magnitude of snow-depth change for events characterized by high absolute change in snow depth (Figure II.3). Underestimation of extremes is relatively common for machine-learning models, since these events tend to be underrepresented in training data and are thus harder for the model to capture. However, as mentioned, this tendency is minimal in this model. Model performance is also high with respect to the KGE (Table II.4). This metric was developed to capture not only average model performance in the form of mean squared error but also the ability of the model to match correlation and variations in the data (Gupta et al., 2009; Kling et al., 2012).

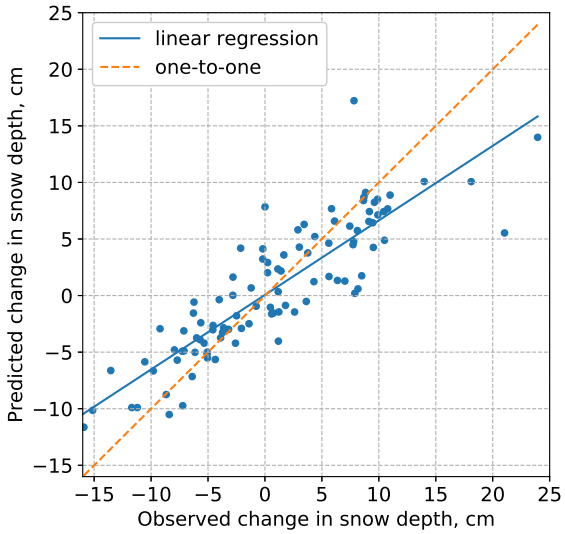


Figure II.3: Observed vs. predicted change in snow depth across all validation events ($R^2 = 0.876$ and $p < 0.001$).

All three models trained on feature subsets performed comparably to the baseline model in terms of RMSE and bias (Table II.4). However, the KGE was significantly lower in models without vegetation data than for the all-feature model, suggesting that despite low correlations with snow-depth change, vegetation and topography features do play a role in accurately modeling snowpack response. The differences in KGE values between versions of the model were driven by differences in the ratio of the means and the ratio of the variations,

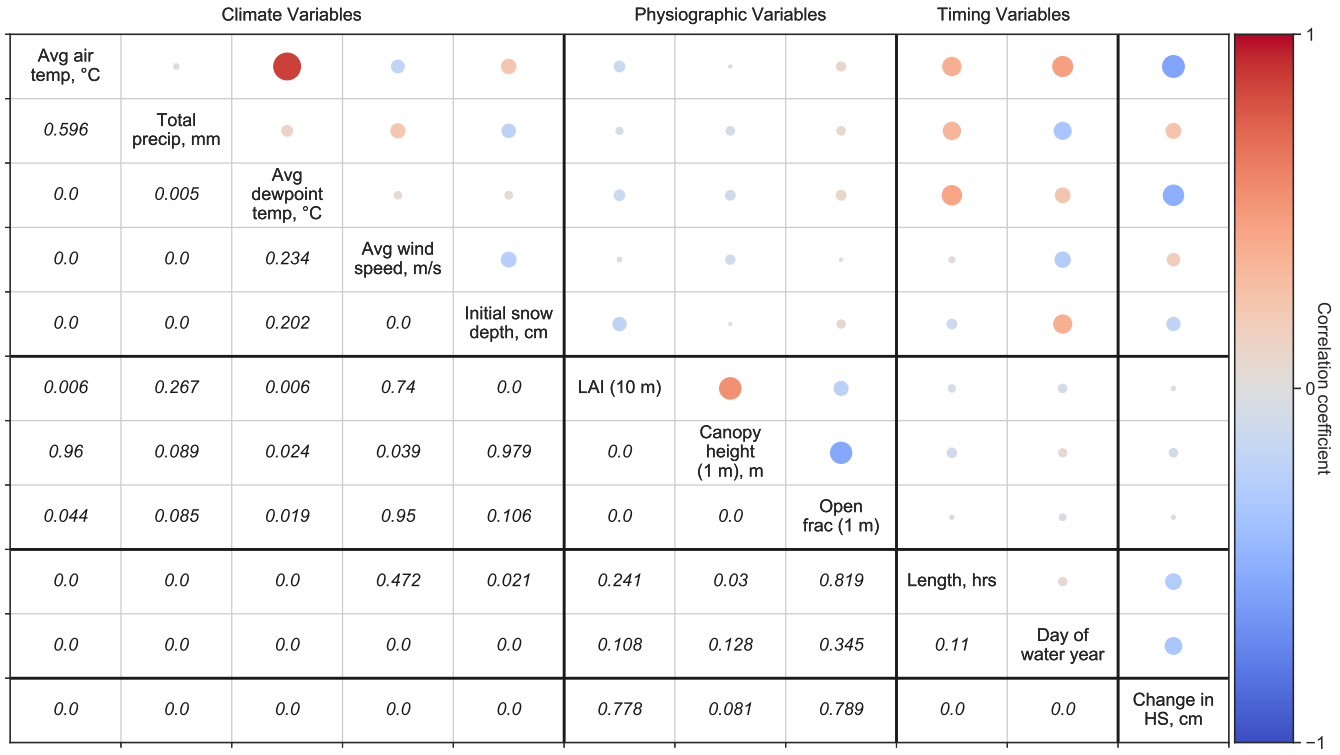


Figure II.2: Correlation plots of storm features and change in snow depth (HS). All physiographic features are at 1-m resolution except LAI (10-m resolution). The upper triangle shows the strength of the correlation and the bottom triangle displays the p-value. A subset of all features are reported for brevity and full correlation plots may be found in Appendix A.1.2.

Table II.4. Random Forest validation performance

Model	RMSE, cm	Absolute bias, cm	Kling-Gupta efficiency
Baseline	3.91	2.98	0.72
Climate features only	3.94	2.98	0.61
Climate + topography & vegetation	3.97	3.07	0.69
Climate + timing	3.91	2.88	0.56

two of three metrics included in the KGE (Table A.4). Thus, including vegetation features allowed the model to more accurately capture average snow-depth response and the range of responses compared to the climate-only model.

II.3.3 Feature importance

As expected, climate features were the most important factors in the Random Forest model (Berg et al., 1991; Würzer et al., 2016; Marks et al., 1998). Average dewpoint temperature and total precipitation had comparatively high scores of 0.17 and 0.13, respectively, with DOWY, average wind speed, the standard deviation of dewpoint temperature, and length of storm showing comparatively moderate importance (between 0.072 and 0.088; Figure II.4). We note that two of these six most important features were related to dewpoint temperature, the primary predictor for precipitation phase (Marks et al., 2013). Timing and physiographic features other than DOWY were relatively low-scoring despite the fact that including these features improve model performance (Section II.3.2), suggesting that a low feature importance score does not mean a feature is irrelevant.

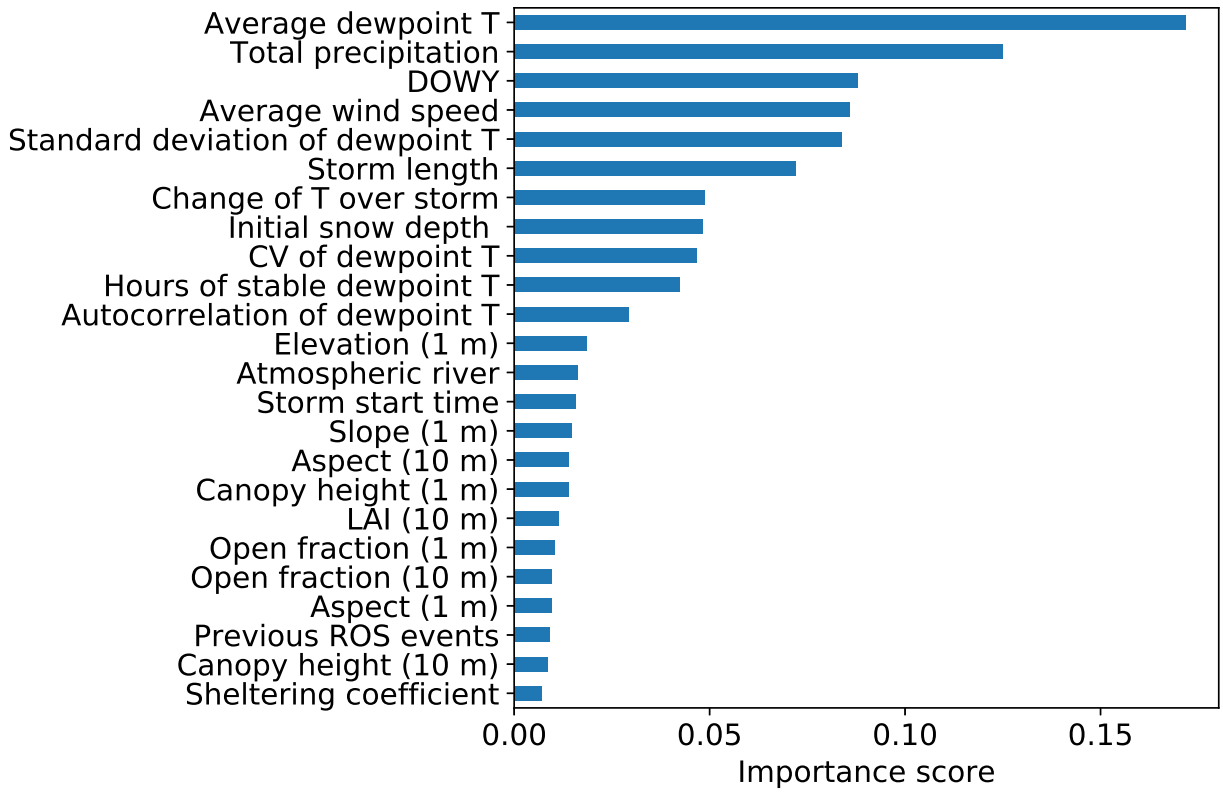


Figure II.4: Feature importance scores of the baseline Random Forest model ranked from highest to lowest. Resolutions at which physiographic variables were calculated are given in parentheses. ROS is rain-on-snow, CV is coefficient of variation, T is temperature, DOWY is day of water year, and LAI is Leaf Area Index.

II.3.4 Mix of precipitation phases during storms

The ROS identification method successfully identified mixed, but mostly rain, events (Figure II.5). Snowpack response to mixed-phase storms was different than to all-rain events. Mixed events resulted in a snow-depth increase (average change of +0.22 cm), whereas all-rain events resulted in a decrease (average change of -3.14 cm). The range of errors associated with mixed events was larger than with all-rain events (-15.50 to 9.40 cm and -9.01 to 3.37 cm, respectively).

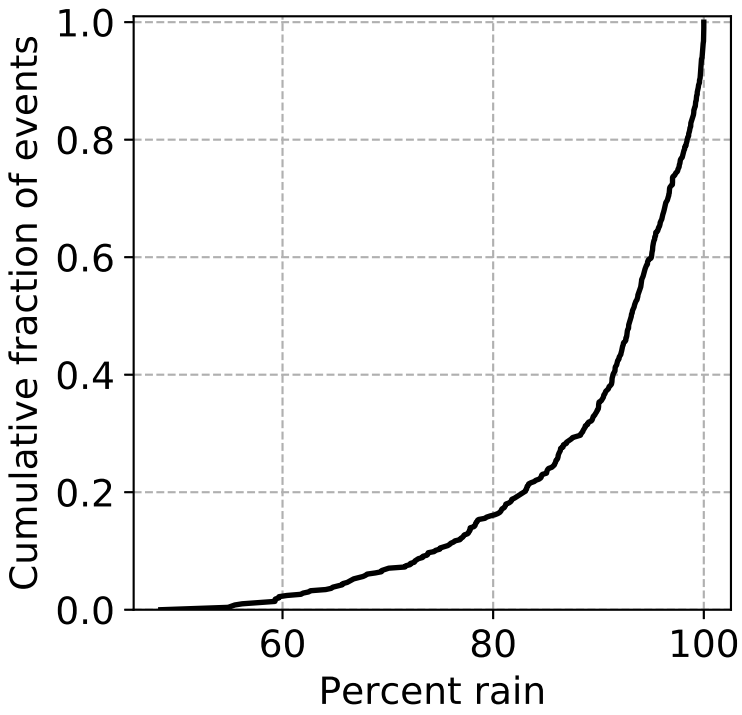


Figure II.5: Cumulative density function of storm precipitation phase.

Phase timing within a storm impacted the response of the snowpack, with rain-first events resulting in close to zero change in depth on average (+0.05 cm) and snow-first events resulting in a negative change (-0.62 cm). Both were also different from the average change of all-rain events. The distributions of snow-depth change between these three event types were all statistically different to the $\alpha = 0.05$ level based on a two-sample Kolmogorov-Smirnov test (Figure II.6a). We also see a greater range of errors in predicting snowpack response to transitional events than all-rain or coincident events, suggesting greater uncertainty in modeling these events.

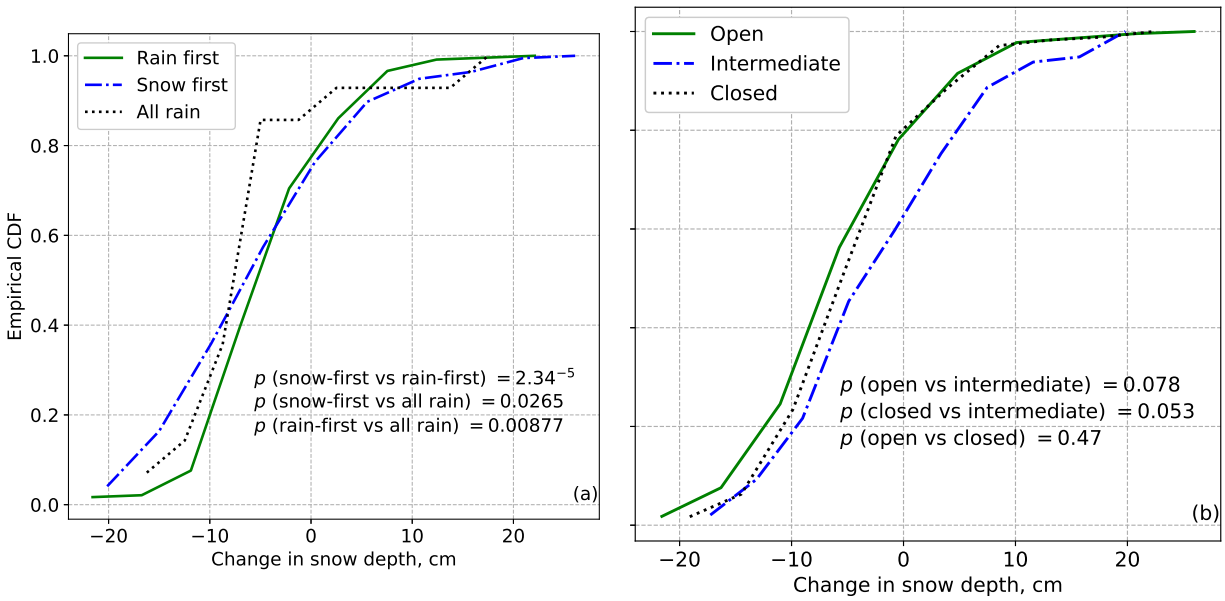


Figure II.6: Cumulative distribution of snow-depth change during rain-on-snow events by (a) phase timing and (b) LAI. "Open" refers to LAI ≤ 0.86; "Intermediate" to $0.86 <$ LAI ≤ 2.05; and "Closed" to LAI > 2.05. The Intermediate events are statistically different from the other two to the $\alpha = 0.1$ level based on a Kolmogorov-Smirnov test.

II.3.5 Physiography and seasonal timing of storms

To better understand the role of physiographic and timing features, we assessed model performance by quintile of each of feature and found noteworthy variations with respect to LAI and DOWY. Moderate LAI is associated with the most uncertainty in the model based on RMSE and bias (Figure II.7a). (KGE was not calculated, since this metric assumes that data are normally distributed, which is not true when the data are binned into subgroups; Pool et al., 2018). In addition, performance with respect to DOWY shows overall improvement as the season progresses, with later-season events associated with least predictive uncertainty. We focus on these two features in the discussion; corresponding plots for other features are shown in the appendix (Figures A.7 to A.10).

Events in the intermediate LAI range ($0.86 \leq$ LAI < 2.05; Figure II.7a) are associated with high uncertainty and show statistically different snowpack response from events with lower or higher LAI to the $\alpha = 0.1$ level based on a two-sample Kolmogorov-Smirnov test (Figure II.6b). Average snow-depth change at intermediate LAIs is +1.53 cm (range = -17.25 to 23.94 cm). For events with lower and higher LAIs, average change in depth is negative (-0.68 cm and -0.50 cm, respectively) and the ranges are slightly larger (-21.58 to 31.24 cm and -19.09 to 26.96 cm, respectively).

Since vegetation and timing features appear to interact to improve model results (Section II.3.2), we extended this analysis to explore the seasonality of snow-canopy interactions. We used DOWY quintile to categorize all ROS events as early (quintiles 1 & 2), middle- (quintiles

3 & 4), or late-season (quintile 5; as shown in Figure II.7b). For these early, middle, and late periods, we then binned events into groups by LAI and compared storm behavior across those groups.

We employed two strategies for binning LAI: first, based on quintile as in Figure II.7a and second, applying a threshold to divide the data into two categories. When binning by quintile, we again found that the intermediate LAI category responds statistically differently, specifically in the early season ($p < 0.01$). Though all sites see an average positive change in snow depth during ROS at this point in the season, intermediate LAIs have a significantly greater average increase (6.77 cm vs. 2.03 cm at open sites and 2.26 cm at closed sites) and also a narrower range of snow-depth changes overall.

We tested the threshold-binning method in 0.5 increments from LAI = 0.5 to 3.5 and found statistically significant differences when dividing events at an LAI threshold of 0.5 for the middle-season period ($p = 0.064$). All together, these results indicate that early season events show a distinct response at intermediate LAIs, but by the middle part of the season that difference shifts to open locations with LAI < 0.5. At this point in the season, open sites (LAI ≤ 0.5) see less decrease (-0.931 cm) than canopied sites (LAI > 0.5; -3.20 cm). These results suggest that the influence of vegetation on snowpack response to ROS is not consistent throughout the season.

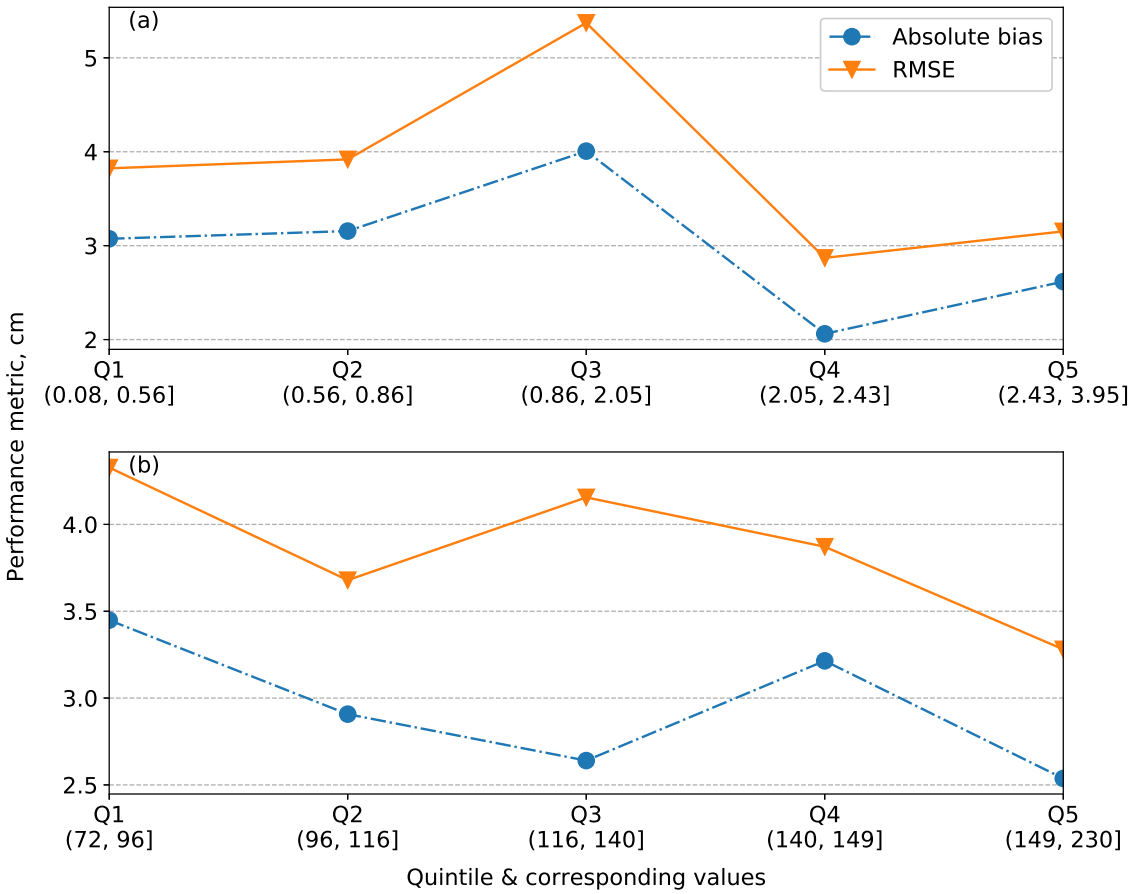


Figure II.7: Model performance calculated by quintile of (a) Leaf Area Index and (b) day of water year.

II.4 Discussion

II.4.1 Mechanisms driving ROS response

These results allow us to identify four primary drivers of snowpack response to ROS events in forested regions: 1) phase fluctuations during storms; 2) temporal variations in snowpack characteristics; 3) forest structure; and 4) seasonal shifts in canopy-snow interactions.

Mix of precipitation phases

Mixed-phase events present the most uncertainty for snow-depth change (Section II.3.4), as they are the only ROS events that may result in an increase in depth. However, settling, densification, and melt may negate any increases, especially in snow-first storms, where rain falls on fresh snow rather than the older, denser snow of rain-first storms. This results in the more-negative snow-depth changes for snow-first storms and may also account for the fact that snow-first events have the largest error range among all storm types. Though snowfall preceding rainfall may decrease snowpack temperature and increase cold content,

thereby increasing the buffer capacity, these storms may also lead to greater subsequent SWE available for melt and runoff. Several case studies of ROS-induced floods, including Pomeroy et al. (2016a); Rössler et al. (2014); Wever et al. (2014), and Marks et al. (1998), cite significant snowfall before rainfall as a driver of subsequent excessive runoff.

Many studies have focused specifically on events in which the precipitation was entirely rain, limiting discussion of phase fluctuations in the literature (e.g., Rössler et al., 2014; Garvelmann et al., 2015; Wever et al., 2014). A notable exception is Pomeroy et al. (2016a), who included temperature-based phase as part of an energy balance model and discussed the implications of late-event snowfall on total event runoff. The high percentage of mixed ROS events, the high feature importance of dewpoint temperature features, and the enhanced modeling challenges of such events support a renewed focus on the precipitation patterns during ROS storms. We suggest that rain-first and snow-first events are fundamentally different, and attention should be given to phase shifts during mixed ROS storms.

Seasonal and interannual variations in snowpack characteristics

Changes in model performance by DOWY (Figure II.7b) and the high feature importance score of storm start date (Figure II.4) suggest a change in snowpack response as the season progresses. We attribute these findings to seasonal and year-to-year shifts in snow accumulation and ablation patterns.

The date of peak snow depth at our sites is highly variable, particularly since the study includes some of the driest (e.g. WY 2015) and wettest (e.g. WY 2017) years on record (Figure II.8). We use date of peak snow depth to estimate date of peak SWE, when the snowpack shifts from accumulation to ablation. Though data are not available to estimate snowpack temperature or cold content, we can expect that the maritime snowpacks in the Sierra Nevada will be at isothermal at the beginning of the season, when snow depth and SWE are low, before becoming colder as the season progresses and snow cover becomes deeper and spatially continuous (Sturm and Holmgren, 1995). By the early ablation season, the snowpack returns to isothermal as it melts.

Thus, during the first DOWY quintile (DOWY 72 to 96; December 12$^{\text{th}}$ to January 5$^{\text{th}}$), when the snowpack is still being established, snow is spatially patchy and interannual variability may be high. This limits the transferability of information from other locations and seasons, making predication more uncertain. Between DOWY 96 (January 5$^{\text{th}}$) and 116 (January 25$^{\text{th}}$), snowpack in most years is well established and still accumulating, leading to more-predictable conditions and lower model error. Across all site-years, the period covering quintiles 3 & 4 (DOWY 116–149; January 25$^{\text{th}}$–February 27$^{\text{th}}$) may or may not correspond to peak SWE (Figure II.8), so ROS storms will encounter a snowpack at peak SWE in some years and a snowpack still in the accumulation period in other years. This variability makes prediction more uncertain. Finally, the improved performance later in the season is due to more-predictable snowpack conditions. During this period snow is more consistently at peak SWE or in the ablation phase, likely corresponding to isothermal temperature and maximum density, meaning that decreases are due strictly to snowmelt rather than settling (De Michele et al., 2013). These findings indicate that while DOWY is important for the model, variability in snowpack characteristics, especially the timing of peak SWE, can make prediction more challenging during certain periods.

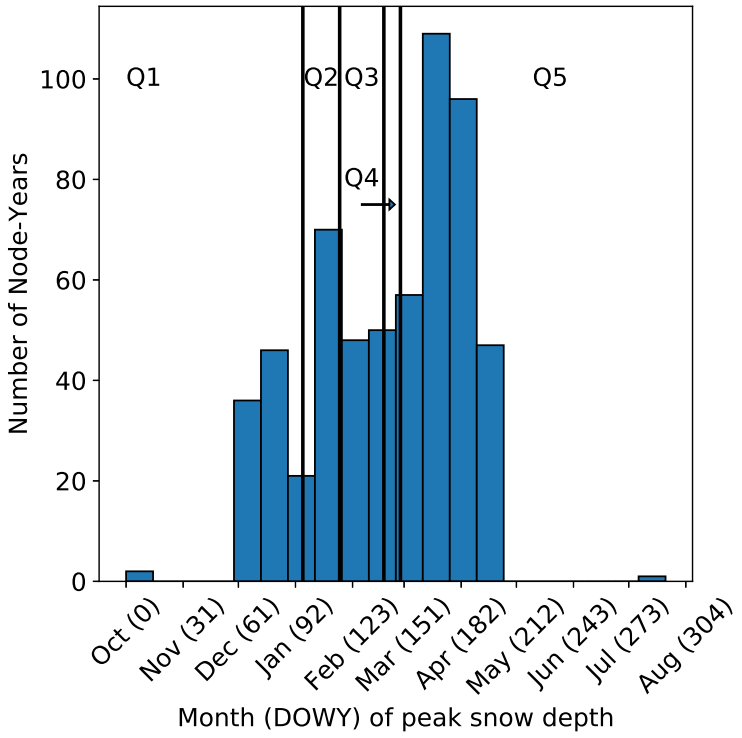


Figure II.8: Histogram of day of water year (DOWY) on which peak snow depth was observed across all nodes and water years. Quintiles are delineated as defined in Figure II.7b.

Vegetation density

At intermediate LAIs (0.86 to 2.05), the snowpack shows a distinct response during ROS and also greater uncertainty in modeling that response (Section II.3.5). Sites at lower LAI will be more consistently exposed to shortwave radiation and turbulent fluxes; sites at higher LAI will be more consistently sheltered from shortwave radiation and turbulent fluxes and exposed to longwave radiation (Lundquist et al., 2013; Musselman et al., 2008). However, at intermediate LAIs, factors including the exact forest structure around the site, time of day, slope, and aspect may influence radiation exposure, snow accumulation patterns, snow conditions, and interception rates. Coupled with the possibility for significant snowpack accumulation at intermediate-LAI sites, both during mixed-ROS events and snowfall events (Garvelmann et al., 2015; Veatch et al., 2009; Broxton et al., 2015), these variations in mass and energy input can make snow-depth change harder to predict.

Interactions between energy and mass inputs may account for intermediate-LAI sites having the opposite sign of snow-depth change than other locations (Garvelmann et al., 2015; Veatch et al., 2009; Dunford and Niederhof, 1944). Intermediate-LAI sites see greater sheltering from turbulent fluxes than low-LAI sites, but lower interception rates than high-LAI sites. The lower range of snow-depth changes at intermediate-LAI sites may be a sign that these drivers of mass and energy inputs are at least partially negating each other, resulting in the vegetation at these densities modulating snowpack response compared to more open or closed sites.

Quantified descriptions of vegetation coverage and forest structure during ROS has not been widely attempted in the literature (Garvelmann et al., 2015). Paired open and vegetated sites (see, e.g., Beaudry and Golding, 1983; Harr and Berris, 1983; Berris and Harr, 1987; Kattelmann, 1987; Berg et al., 1991; Marks et al., 1998) examine only one of a wide range of possible forest structures. Garvelmann et al. (2015) pointed out that studies that found large differences in ROS snowpack response between open and closed sites generally compared sites in small clearings with forested areas. On the other hand, studies such as Garvelmann et al. (2015) and Kattelmann (1987) that compared large open meadows with forested areas found less difference. Our findings support this hypothesis: we found no statistical difference between the distributions of snow-depth change at LAI < 0.86 and LAI > 2.05 (Figure II.6b). Our results suggest that more-precise characterizations of forest structure are necessary for better understanding and predicting the interactions between vegetation and snow during ROS events.

Seasonality of canopy effects

As shown in Section II.3.2, including timing features in the absence of vegetation features in a Random Forest model resulted in worse performance with respect to the KGE than not including them at all (Table II.4). This suggests that interactions between vegetation and snowpack evolve through the year. In the early season, the distinctions discussed in the previous section that distinguish intermediate-LAI sites (greater sheltering from turbulent fluxes and exposure to longwave radiation than low-LAI sites; less interception capacity than high-LAI sites) are all in effect. As a result, sites at intermediate LAI respond differently during this part of the season. The combination of lower turbulent-flux input and less inter-

ception may explain the overall greater increase in snow depth at these sites. As the season progresses, the distinction between intermediate- and high-LAI sites (i.e., interception) becomes less pronounced as canopy capacity fills, and intermediate sites begin to behave more like high-LAI sites. However, the differences between intermediate- and high-LAI sites and low-LAI sites (sheltering from turbulent fluxes and exposure to longwave radiation) remain, leading to a situation where open sites behave differently from the rest. By the end of the season, when precipitation is less likely to be rain than snow, the effect of interception is further reduced.

These findings suggest that for ROS events, the effect of canopy on snowpack shifts from being related to both energy exposure and interception in the early to middle parts of the season to being more related to energy-exposure patterns later in the season, in keeping with previous findings (Lundquist et al., 2013). These early, middle, and late season periods may also be correlated with spatial or year-to-year variability in snowpack conditions. In the Sierra Nevada, the early and middle-season periods (approximately December through February), is also associated with many historical ROS-induced floods (Kattelmann, 1997). We therefore suggest a focus on vegetation controls on snowpack during midwinter period as a topic for future work.

II.4.2 Data-driven models for ROS simulation

Our results suggest that simultaneous climate and ground-cover data provide a promising basis for a data-driven model of snowpack response to ROS that could support forecasting and decision-making related to ROS events. The model captured the correct sign change in 86% of validation cases and performed better for events with a negative change in snow depth (average absolute error of 2.46 cm) than positive change (3.43 cm). This is important for forecasting because events associated with a decrease in depth may be more likely to be associated with increases in runoff, though this could only be verified by extending this work to include SWE changes and/or runoff measurements.

The Random Forest model used here performed well across a variety of vegetation and topographic features, but like all machine-learning algorithms, relies on sufficient data under a representative set of conditions. The hydrology of California may require multiyear datasets for accurate results, since water year precipitation totals often fall well below or above the average (Dettinger et al., 2011). These varying conditions can cause snowpack distribution and characteristics, including overall snow depth, liquid-water content, and cold content of the snowpack, to vary widely from year to year. A set of Random Forest models trained by leaving out specific water years for validation supports this idea (Figure A.11 in Appendix A.1.4). The baseline ROS model with information across years showed better performance than nearly all the leave-one-out models, the exception being especially dry years, likely due to the low overall snowpack.

California's already-variable climate may be further exacerbated by climate change. The southern Sierra Nevada and regions at similar elevation are projected to see increases in ROS events as temperatures rise, since their snowpack will persist while still seeing an increase in warm storms (Musselman et al., 2018). Climate change may also impact the timing of these events, with flood regimes shifting from snowmelt-driven spring events to rain-driven winter events (Arnell and Gosling, 2016). Finally, there is also evidence that in warm snowpacks

such as found in California, the variability of peak SWE will increase (Marshall et al., 2019) and precipitation patterns will become more extreme (more very wet and very dry years; see Swain et al., 2018). These changes could negatively impact model performance for unobserved conditions and exacerbate issues with interannual information transfer.

Changes in vegetation may also impact model robustness, especially in the Sierra Nevada, where forest thinning practices are regularly employed (Saksa et al., 2017). Our findings suggest that forecasting uncertainties may be greater in areas with moderate vegetation cover and small clearings, rather than dense vegetation or large, open clearings. Thus, different thinning patterns may lead to different impacts on snowpack response. For example, uniform thinning that leads to more areas at intermediate LAI may make large-scale prediction more uncertain. However, this may not be observed using variable thinning, in which some areas are clear cut leaving small clusters of dense vegetation (Pickard, 2015). Instead, variable thinning may lead to greater spatial variability of ROS response during the middle part of the water year when nearby open areas will behave differently from dense tree stands.

Vegetation changes may also occur due to fire or tree mortality, which was particularly intense in the southern Sierra Nevada in the aftermath of several extremely dry years between 2012 and 2015. Meanwhile, wildfires are a regular and growing occurrence in California and can dramatically change vegetation coverage (Hu et al., 2019; Miller et al., 2012) and snowpack energy balance (Gleason and Nolin, 2016). From a forecasting standpoint, major changes in vegetation should be corrected for when making predictions. Observational networks and remote sensing data can support these updates and are thus complementary to machine-learning models.

II.5 Conclusion

Using a data-driven Random Forest algorithm, we identified four drivers of snowpack response to ROS events in a montane region under a variety of canopy conditions. Weather conditions, in particular when rain versus snow occurs during an ROS storm, is important to snowpack response. Snow-first storms are more likely than rain-first storms to lead to negative changes in snow depth. The condition of the snowpack itself is also important, with variability in the timing of peak SWE limiting the transferability of information across water years during certain parts of the season. Thus, climate-change impacts on the timing of peak SWE in warm snowpacks may increase model predictive uncertainty. Vegetation coverage changes snowpack response to ROS, but this pattern is not constant throughout the season. Early in the water year, the range of snow-depth change is mitigated at intermediate vegetation densities compared to more open or closed sites due to a combination of low interception rates and sheltering from turbulent wind fluxes. As the season progresses, differences in interception rates across vegetation densities decreases, but differences in energy exposure remain. In this period, all canopied areas experience greater negative change in snow depth on average compared to open areas. These results support a renewed focus on ROS events in montane regions, particularly where changes in vegetation patterns due to forest thinning, wildfires, or tree mortality may influence snowpack response to ROS events.

Chapter III

Optimizing spatial distribution of watershed-scale hydrologic models using Gaussian Mixture Models

Based on the paper of the same name submitted for publication by:

Tessa Maurer, Francesco Avanzi, Carlos A. Oroza, Steven D. Glaser, Martha Conklin, Roger C. Bales

Abstract

We propose the use of Gaussian Mixture Models (GMMs) for identifying the optimal modeling locations for spatially distributed hydrologic models. Common methods for spatial distribution of model input data and parameters, such as defining inputs for individual hydrologic response units, are subjective, time-consuming, and fail to capture the full range of basin attributes. Recent advances in machine-learning techniques and an increase in spatial data availability from remote sensing products allow for new approaches to this problem. GMMs select the set of points that best represent the multivariate distribution of a set of data features. By defining these features to be the watershed characteristics most relevant to runoff production, GMM identifies the optimal modeling locations and provides an efficient, objective way to distribute model input data across the landscape. We demonstrate this method in two hydrologically distinct headwater catchments of the Sierra Nevada and show that it meets or exceeds the performance of traditionally distributed models for multiple metrics across the water balance at a fraction of the time cost. Finally, we use univariate GMMs to identify the most important drivers of hydrologic processes in a basin. The advantages of the GMM method allow for more robust models, which are critical for research and operational decision-making.

III.1 Introduction

Spatial heterogeneity of hydrologic processes within a watershed is fundamentally impacted by basin topography. Topographic variations affect vegetation characteristics directly via climatic controls and indirectly via impacts on soil profile and water and nutrient availability (Fan et al., 2020; Tian et al., 2020a; Zhang et al., 2011; Qiu et al., 2001). All three – topography, soil, and vegetation – combine to impact hydrologic processes including evapotranspiration, infiltration, runoff, and interflow (see, e.g., Ghestem et al., 2011; Wilcke et al., 2011; Obojes et al., 2015; Young et al., 1997). Both vegetation and topographic variations such as slope and aspect impact snow accumulation and ablation patterns through controls on short- and longwave radiation, wind, and interception (Lundquist et al., 2013; Maxwell et al., 2019; Varhola et al., 2010). In montane regions, heterogeneity of the landscape can have profound implications for all portions of the water balance: orographic effects can create dramatic differences in precipitation rates on either side of mountain ranges as well as influence the phase (rain versus snow) of that precipitation (Roe, 2005). Landscape variability in these regions is of particular interest due to the role these river basins play in the "waterscape" connecting natural headwaters with human needs (Karpouzoglou and Vij, 2017). These "water towers of the world", supply water to over half of the human population (Mountain Partnership, 2014; Immerzeel et al., 2020; Viviroli et al., 2007a). Understanding the variations of hydrologic processes that contribute to the timing and quantity of streamflow from these basins is a fundamental goal for both scientific researchers and, increasingly, operational forecasters in the water management sector. These questions have become all the more pressing in regions where climate change is inducing shifts in the water balance not previously seen.

In order to meet these needs and spurred by increases in computational resources, the use of physically based, spatially distributed hydrologic models is becoming more common. Physical models use an explicit approach in which mass and energy balances are resolved; with a spatially distributed set-up, these simulations are performed at multiple points across a river basin and then aggregated. Models may be partially (or semi-) distributed, where some model components (e.g. input data) are varied across the landscape but others (e.g. parameters) are held constant, or fully distributed (all components are spatially variable). In spatially lumped models, on the other hand, all hydrologic processes occurring within a basin are simulated at a single point and output is given as a single time series for the basin (usually streamflow at the basin outlet). Much research has focused on calibration approaches for distributed models in an effort to address concerns of overparameterization, non-identifiability of parameters (equifinality) and scale consistency (Beven et al., 1988; Beven, 1989; Wood et al., 1988; Blöschl and Sivapalan, 1995; Andréassian et al., 2012; Bai et al., 2009; Pianosi et al., 2015). Comparisons between lumped, semi-distributed, and fully distributed calibration techniques (e.g. Khakbaz et al., 2012; Reed et al., 2004; Lobligeois et al., 2014; Carpenter and Georgakakos, 2006) have attempted to characterize the relationship between spatial distribution and model performance. Research on scale consistency focuses on reconciling parameter values across scales in an effort to avoid sudden changes in results when spatial resolution is changed (Sivapalan and Kalma, 1995). These approaches may be top-down (calibrating a lumped model and disaggregating parameter values; e.g. Tran et al., 2018) or bottom-up ("regionalization"; e.g. Blöschl and Sivapalan, 1995; Arse-

nault and Brissette, 2014; Hundecha et al., 2016; Samaniego et al., 2010).

These approaches focus on the calibration step of model set-up, but less attention has been given to the prior step of selecting which and how many specific simulation locations within the basin to include in the model. This critical first step of spatially distributing a model impacts all subsequent set-up, including input data distribution and parameter definition and calibration. Methods proposed in the literature for selecting modeling locations or otherwise partitioning the basin include Representative Elementary Areas (REAs), an intermediate scale at which neither small- nor large-scale processes dominate (Wood et al., 1988; Blöschl et al., 1995); Representative Elementary Watersheds (REWs), units derived based on the streamflow network and over which equations of mass and energy fluxes are integrated (Reggiani et al., 2000; Reggiani and Rientjes, 2005); and landform classes based on the UPNESS index (Summerell et al., 2005; Roberts et al., 1997). These methods showed promise in capturing spatial variability, but were limited by detailed data or catchment-monitoring requirements, inability to simulate multiple processes, and/or assumptions in the derivation process. More recently, pixel-based distribution approaches have risen in popularity to be compatible with gridded remote-sensing products. Though convenient, this approach is disconnected from the physical characteristics of a basin: pixels may straddle discontinuities in topography or land use, introducing uncertainty into simulations and/or runoff routing. In addition, pixel-based approaches typically results in hundreds or even thousands of simulation points for a moderately sized basin (see, e.g., Tran et al., 2018), since model resolution is frequently dictated by input data resolution. Not only does the high number of modeling locations raise equifinality concerns, these models often have higher simulation times and increased computational requirements. This can be particularly problematic for time- or resource-constrained applications such as real-time flood forecasting. In montane regions, elevation bands are sometimes used as a simple alternative to capture spatial variability (e.g., Bongio et al., 2016; Valéry et al., 2014), but are also often arbitrarily defined and may not align with topographic features.

Alongside pixel-based methods, the most widely used approach for spatially representing a basin is Hydrologic Response Units (HRUs; Leavesley et al., 1983; Flügel, 1995, 1997), defined as areas of a basin that can be considered homogeneous in all respects influencing the water balance (e.g. topography, land cover and vegetation density, and soil type). Conceptually simple, HRUs are favored by some modelers as having a stronger connection to physical basin characteristics than pixel-based models. HRUs are the default distribution method in several major hydrologic models, including the Precipitation-Runoff Modeling System (PRMS; Markstrom et al., 2015), the Soil and Water Assessment Tool (SWAT; see, e.g., Kalcic et al., 2015; Teshager et al., 2016; Qi et al., 2017), Precipitation-Runoff Evapotranspiration Hydrotope Model (PREVAH; Viviroli et al., 2007b), the Sacramento Soil Accounting Model (SAC-SMA; National Oceanic and Atmospheric Administration, 2002) and the Regional Hydro-Ecological Simulation System (RHESSys; Tague and Band, 2004). In addition, HRUs are used by many large water-management agencies that rely on physical hydrologic models, including California's Department of Water Resources (DWR) and Pacific Gas & Electric (PG&E) utility company.

Despite their popularity, HRU-based distribution presents both theoretical and practical problems. There is inherent tension between having more, smaller HRUs that are more likely to conform to the assumption of homogeneity and the need to reduce unnecessary

model complexity. In addition, though HRUs are meant to represent a distributed sub-area of a basin, hydrologic processes are simulated at a particular point, usually the geometric centroid of the HRU. This necessarily limits the points of the basin that can simulated with HRUs; for example, the geometric centroid will always be lower than a peak or ridge, meaning that the model is likely to miss the highest elevations. HRUs are frequently delineated using a GIS-based approach, starting with a digital elevation model and using topography, including drainage divides, slope, and aspect, to partition the study area (see, e.g., Flügel, 1995, 1997; Koczot et al., 2005). Though tools such as the ArcMap-based tool GIS Weasel (Viger and Leavesley, 2007) have been built to assist with this process, this method of HRU delineation involves significant subjective decision-making, such as selecting minimum HRU size and stream-segment resolution. All of this can translate to multiple days of hands-on work. Other methods have been proposed for HRU delineation, including Khan et al. (2013) and Khan et al. (2016), who overlaid soil and stream-network data on a set of identified landform classes, and Fiddes and Gruber (2012), who used a sub-grid sampling method to include the effects of topography in a lumped model. While promising, these approaches both rely on assumptions that are not generalizable across catchments and/or all aspects of the water balance. Ultimately, HRU delineation (and, by extension, selection of modeling locations) involves subjective decisions and significant time investment.

Given these issues, there is need for a simple, rapid, and objective approach to selecting modeling locations for spatially distributed hydrologic models. Recent advances in machine learning have made possible alternative approaches to this problem. In particular, mixture models have emerged as a way of optimally identifying a set of underlying components that best describes a population. We propose the application of Gaussian Mixture Models (GMMs) as an objective, efficient, and physically based spatial-distribution technique that addresses both the theoretical and practical shortcomings of existing methods. Using basin characteristics that influence the water balance, mixture models permit us to identify a set of modeling locations that optimally characterize the rainfall-runoff relationship throughout the basin as a whole. Gaussian Mixture Models have been successfully used to capture spatial patterns in other hydrologic contexts such as snow-water equivalent (SWE) distribution across the landscape (Oroza et al., 2016).

We demonstrate the GMM-based distribution method in two contrasting headwater catchments of California's Sierra Nevada using the Precipitation-Runoff Modeling System (PRMS), a physically based rainfall-runoff model commonly used in water management. Owing to their widespread use by researchers and forecasters, we use a GIS HRU-based PRMS model as a baseline with to compare the performance of the GMM-based models. In the research reported here, we address the following:

1. What is the measurable impact of a GMM-based spatial-distribution method versus an HRU-based method on predictive accuracy?

2. Are these spatial-distribution methods robust to unobserved, extreme hydrologic events? Which hydrologic process(es) drive improvements or declines in modeled performance?

3. What attributes are the most important drivers of predictive accuracy in montane catchments?

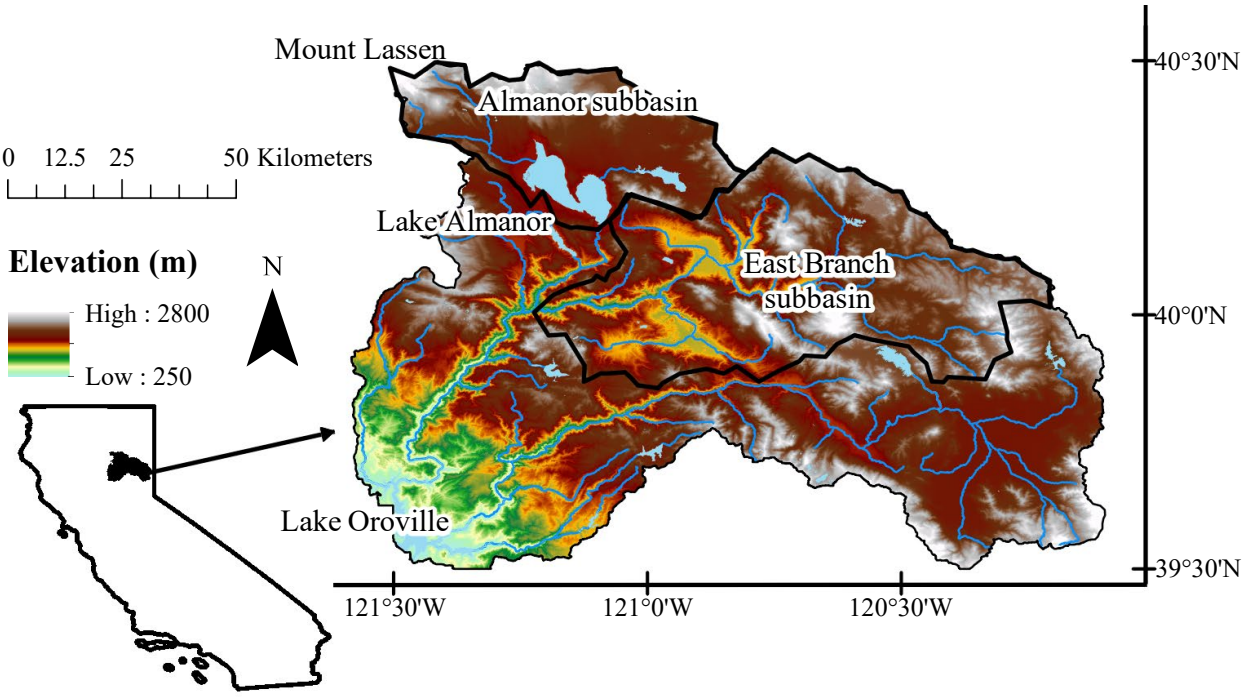


Figure III.1: Map of California, indicating extent of Feather River and the headwater catchments, Almanor and the East Branch, that were the subjects of this study.

III.2 Methods and data

III.2.1 Study Area

We focus on Almanor and the East Branch, two headwater catchments of the North Fork of the Feather River, the northernmost basin of the California Sierra Nevada (Figure III.1). The Feather River is important for both water resources and energy production. PG&E, California's largest utility company, operates a series of hydropower plants on the North Fork totaling 740 MW of installed capacity, about 19% of company's overall hydropower portfolio. The basin also drains to Lake Oroville, the primary storage reservoir for the State Water Project operated by the California DWR and serving drinking water and agricultural water needs in the central and southern parts of the state.

As a lower-elevation Sierra Nevada basin (peak elevation 2950 m), the Feather is susceptible to climate change effects as more precipitation falls as rain rather than snow. The Feather River can therefore be thought of as an early example of how other basins in the Sierra Nevada may change with rising temperatures (Freeman, 2011). The main stem of the North Fork of the Feather originates in the Almanor catchment to the northwest and is regulated at the outlet of Lake Almanor. Almanor drains an area of approximately 1150 km^2 and contains Mount Lassen, the highest and wettest point in the Feather River at about 2900 m elevation and 3000 mm of annual precipitation (Koczot et al., 2005). Geologically, Almanor is part of the Cascade Mountain range rather than the Sierra Nevada, making it

distinct from the rest of the basin. The subsurface is largely composed of more-permeable volcanic rocks, and baseflow makes up a higher percentage of flow than in other subbasins (Freeman, 2008).

The East Branch is a tributary of the North Fork and drains an area of approximately 2650 km^2. It meets the North Fork south of Lake Almanor. The East Branch is rain-shadowed due to the eastern ridge of the Upper North Fork Canyon on its western edge and is thus considerably drier than Almanor, with an average annual precipitation of about 300 mm. The subbasin has a largely granitic subsurface and low baseflow (Freeman, 2008). It is also mostly unregulated.

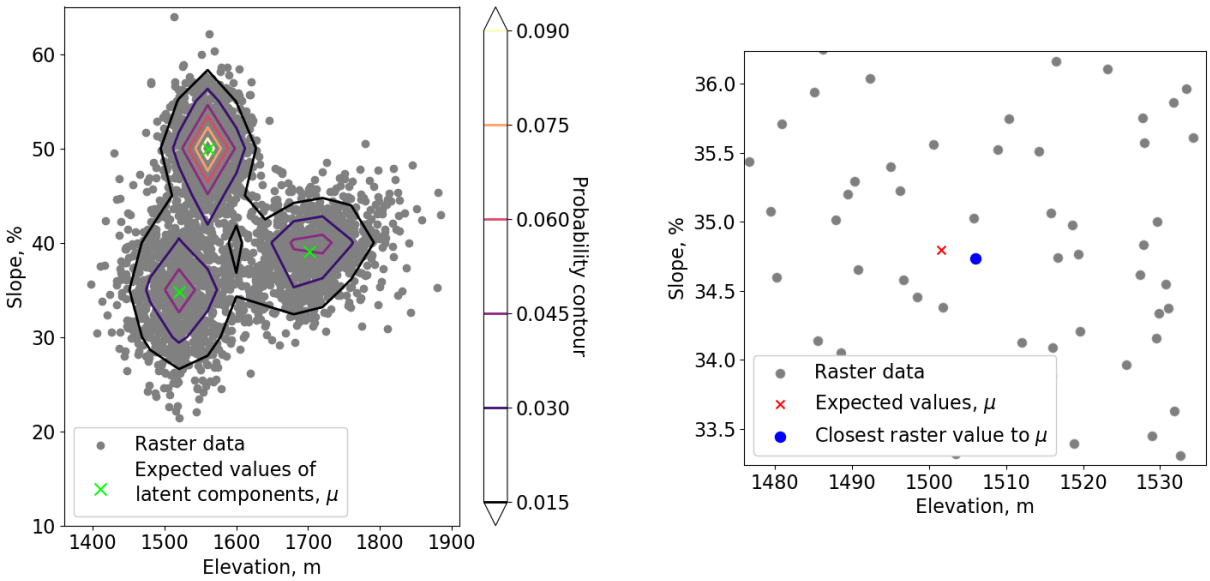
III.2.2 Gaussian Mixture Models for spatial distribution

Gaussian Mixture Models are a machine-learning algorithm that can be used to identify a subset of discrete points that best represent a feature space. Here, "feature" is a measurable characteristic that describes a phenomenon being observed (Bishop, 2006). For example, in describing runoff from a basin, a feature may be the basin's elevational distribution. A "feature space" is the single (if there is only one feature) or multidimensional (if there is more than one) range collectively defined by the feature data. For example, in Figure III.2, there are two features, elevation and slope, creating a two-dimensional feature space. All data points (shown as dots) fall somewhere in the feature space. Features must be continuous numeric variables for use in a standard GMM, but otherwise may be defined at the discretion of the modeler.

The GMM algorithm assumes that the feature space can be represented by superimposing a finite number (M) of "latent components", which are normally distributed. Figure III.2a shows three latent components, each of which can be uniquely described by a mean (expected value; μ) and covariance (Σ). Each is also assigned a mixing parameter (π) based on the prior probability of observing that component (essentially, a weighting factor). Thus, the parameters that are defined in fitting a GMM to a particular dataset are the means, covariances, and mixing parameters for each latent component.

We take the expected values of the latent components as the set of points that best describes the feature space. In concrete terms, and relating the example schematic in Figure III.2a, the three μ values (shown as red X's) are the points that best represent the distribution of elevation and slope in this hypothetical basin. However, a point that exists in feature space (say, for example, an elevation of 2300 m and a slope of 85) may not exist physically in the basin. Thus, once the means have been identified, we use a Nearest Neighbors approach to find the physical location that is closest to the means of feature space (Figure III.2b). These locations define the spatial distribution of the GMM-based PRMS models (henceforth, "modeling locations") and are analogous to the HRU centroids that define the spatial distribution of traditional models (Figure III.2c). Maps of the actual selected modeling locations for each subbasin are available in Appendix B.3 (Figures B.1 and B.2).

Formally, the ability of a GMM latent component to represent the feature space is modeled as a multivariate normal distribution, $\mathcal{N}$, with expected value μ and covariance Σ



(a) (b)

(c)

Figure III.2: Conceptual schematic illustrating the process of selecting modeling locations using GMM. (a) Example of two-dimensional GMM with three latent components. Lines show equal probability contours. Green X's indicate the expected values of each component. (b) Close-up of Figure III.2a showing the Nearest-Neighbor selection of the datapoint closest to the expected value of a latent component. The blue dot represents the raster pixel that is selected as the modeling location. (c) Hypothetical river basin indicating pixels selected as modeling locations

applied to a D-dimensional vector of empirical data $\mathbf{x}$ (Equation III.1).

$$\mathcal{N}(\mathbf{x}|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\Sigma|^{1/2}}\exp\Big\{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\Big\} \tag{III.1}$$

The collective ability of the M components to reproduce the feature space is calculated by superimposing each $\mathcal{N}_m$, weighted with its mixing parameter, π_M. The number of features (i.e., the length of $\mathbf{x}$) determines the dimension of each $\mathcal{N}_m$ distribution. The mixing parameters sum to 1. The expected values, covariance, and mixing parameters that best represent the data were identified by maximizing the likelihood function, given by Equation III.2.

$$\ln\ \mathrm{p}(\mathbf{x}_n|\pi,\mu,\Sigma) = \sum_{n=1}^{N}\ln\Big\{\sum_{m=1}^{M}\pi_m\mathcal{N}(\mathbf{x}_n|\mu_m,\Sigma_m)\Big\} \tag{III.2}$$

In this study, the number of dimensions was five, and the features were basin elevation, slope, aspect, vegetation coverage, and soil hydraulic conductivity (k_{sat}). Together these features capture the major drivers of the water balance endogenous to the basin (i.e., not driven by climate or weather inputs), including spatial distribution of the snowpack, an important if not dominant component of the hydrologic cycle in the Feather River; evapotranspiration; and infiltration characteristics.

Elevation, slope, and aspect were defined using the USGS National Elevation Dataset (NED; EROS Data Center, 1999), and vegetation data were obtained from the 2013 U.S. Forest Service LANDFIRE dataset (LANDFIRE, 2013a,b). (See Appendix B.1.2 for full details on the LANDFIRE dataset). Topographic and vegetation data rasters were both at 30-m resolution. These rasters were masked to the extent of the subbasins in the study, filtered to remove pixels with undefined values (for example, flat areas with undefined aspect), and passed through a 1-in-2 resampling algorithm (see https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.sample.html) to make processing computationally feasible. Subsampling was performed with an initial seed for reproducibility.

Soil hydraulic conductivity, k_{sat}, was derived from STAGO2 data (Soil Survey Staff Natural Resources Conservation Service, 2019), which are available as shapefiles indicating the extent of different "geologic groups"; i.e., clusters of one or more soil types, each of which is associated with a set of unique soil properties. Properties were first depth integrated, then spatially averaged using the percent of each soil type in a geologic group. This averaged property (i.e., k_{sat}) was assumed to be spatially homogeneous across the geologic group. The distribution of the averaged k_{sat} was used as the GMM input feature.

Using the five rasters as inputs, the GMM algorithm was run using Scikit-learn's mixture.GaussianMixture class (Pedregosa et al., 2011), which uses the Expectation Maximization algorithm for optimization (McLachlan and Peel, 2004; Pedregosa et al., 2011). This is an iterative gradient descent method that aims to identify the most likely mixing parameters, covariance, and means to explain the data. The optimization terminates when a maximization step no longer increases the log-likelihood. As noted, the optimal expected values (means) of the latent components in feature space were translated to physical modeling locations using a nearest neighbors algorithm. Features were scaled with equal weight to prevent features with higher magnitude values from dominating the nearest-neighbor search.

In addition to these five-dimensional multivariate GMMs, we ran five additional GMMs in each basin, each driven by only one of the features (univariate GMMs). This was done to assess the usefulness of each individual GMM input feature, and we show that this can inform feature selection and their relevance to different hydrologic processes.

III.2.3 Precipitation-Runoff Modeling System

The Precipitation-Runoff Modeling System (PRMS) is a distributed-parameter hydrologic model developed by the U.S. Geological Survey (Markstrom et al., 2015). The model runs on the daily time step, taking as inputs daily precipitation and minimum and maximum temperature, which are distributed to each HRU either ahead of time by the modeler or through an interpolation scheme within the model. PRMS simulates mass and energy balances beginning with calculation of solar radiation and precipitation phase partitioning and ending with computation of total streamflow. Intermediate processes include snow accumulation and ablation, canopy interception and evapotranspiration, infiltration, surface runoff, interflow, and groundwater recharge. PRMS is executed in a linear fashion at each time step in the simulation, with each hydrologic process represented by a module of code. For some processes, users may specify a desired calculation method by selecting from multiple possible modules.

The spatial distribution in PRMS is achieved by partitioning the modeling area into HRUs, represented by a specific geographical point in the basin (by default, the geometric centroid; Koczot et al., 2005; Markstrom et al., 2015) to which input data are distributed and at which the water balance is simulated. The water balance is simulated separately at each HRU and scaled according to the surface area of the HRU. Outflow is aggregated across the basin based on the selected streamflow routing method.

PRMS is currently used throughout the California Sierra Nevada for streamflow modeling by PG&E (Richards, 2018). It is also being actively developed for new river basins by DWR (see, e.g., Burley and Fabbiani-Leon, 2018). Its widespread use for water-resources planning as well as its commonalities with other distributed-parameter models makes it ideal for this study. The model is publicly available at `https://www.usgs.gov/software/precipitation-runoff-modeling-system-prms`. The latest release is version 5, but at the time of this research, version 4.0.3 was the most updated available. Major changes between the two versions do not affect the modules used in this study.

PRMS models used in this study

Required input data for PRMS are daily temperature range and precipitation amount, which are spatially distributed based on a user-selected method. Here, both temperature and precipitation were pre-distributed to each HRU or modeling location before executing PRMS. Precipitation was distributed using an algorithm called DRAPER, in which spatially distributed long-term average monthly precipitation surfaces from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) dataset are adjusted using daily ground-based values (Daly et al., 2008). Each day, the PRISM surface is multiplied by the ground-based measurements as a percent of long-term daily average. This effectively "tilts" the PRISM surface to reflect the daily observations. For more details on the DRAPER al-

gorithm, see Donovan and Koczot (2019). For a full discussion of the implementation of DRAPER in this study, see Appendix B.1.4.

Daily minimum and maximum temperature were also distributed externally to PRMS. This procedure was selected based on an analysis of temperature variability across the Feather River basin that showed both temporal (sub-monthly) and spatial (intra-subbasin) variability in lapse rates (Avanzi et al., 2020a). Internal PRMS distribution methods did not permit this variability to be accounted for. Instead, temperature was distributed as a two-step process in which we regressed elevation against minimum and maximum temperature for several training stations in each subbasin to establish daily basin-wide lapse rates. Residuals between temperature predicted using these lapse rates and observed temperature at a set of evaluation stations were distributed using multilinear regression with elevation and the temperature at a designated seed station. The final values of maximum and minimum temperature were obtained by subtracting the residual from the first-guess temperature obtained using the lapse rates computed in the first step. For more details on this process, see Appendix B.1.4.

In this study, solar radiation was calculated in PRMS with a degree-day approach, which uses daily maximum air temperature to obtain actual daily solar radiation (*ddsolrad_hru* module). Evapotranspiration (ET) was calculated using the Jensen-Haise formulation (*potet_jh*). PRMS also requires values for several dozen parameters that may be spatially and temporally global or may be set on a per-month or per-HRU basis. Values based on topography and canopy cover were computed based on the USGS NED (EROS Data Center, 1999) and the US Forest Service LANDFIRE dataset (LANDFIRE, 2013a,b), respectively. Remaining non-calibration parameters were computed based on available data, set to default values, or retained from the original USGS version of PRMS on the Feather River. Details of this process for both the GIS and GMM versions can be found in Appendix B.1.

In this study, we use a traditionally designed PRMS model as a benchmark for model performance. This model is henceforth referred to as the "GIS" model, in reference to the geographic information system framework typically used to delineate HRUs. The GIS model was based on the Feather River PRMS model (version 2) designed by the USGS in the early 2000s, in which HRUs were delineated using standard methods (see Koczot et al., 2005, for details). We updated the model to PRMS version 4.0.3; as part of this process, we made minor manual updates in the HRU boundaries to better reflect drainage divides. In addition, PRMS version 4 allows for greater functionality in terms of streamflow routing, which was introduced in lieu of straight summation of HRU outflows at each time step as was done in version 2. Including explicit streamflow routing is more reflective of current hydrologic models. Stream segments for routing were also selected using GIS tools. Details of the new model design, including how non-calibration parameter values were selected, are given in Appendix B.1.

Other than the process for selecting modeling locations, the PRMS set-up for the GMM models was largely the same as for the GIS model. Topographic and vegetation parameters were based on the values at the selected modeling locations. Unlike HRUs, which have each represent a different square area, each modeling location necessarily represents an equal portion of the headwater catchment based on the GMM feature rasters. The GMM models did not employ a streamflow-routing method since modeling locations are not coupled with drainage areas. Instead, runoff from each modeling location was summed for each time step.

For details on the model design, see Appendix B.1.

In order to explore the first two research questions (assessing the accuracy of the distribution methods and robustness to unobserved events), we trained four models for each subbasin with varying numbers of target modeling locations: one (lumped case, for comparison), 50, 100, and 200. This process allowed us to identify how the number of modeling locations changes the performance of the model. In order to address the third research question (identifying important drivers of predictive accuracy), we trained univariate-GMM models (i.e., each using only one of the five variables from the multivariate versions). This allowed identification of the most important variables for each subbasin. For this step, models were run for the best-performing number of modeling locations from the multivariate GMMs.

III.2.4 Calibration and Evaluation Strategy

We used a multi-step, multi-objective method to calibrate the models in order to avoid the overfitting that is common when calibrating on streamflow alone (Hay et al., 2006; Gupta et al., 1998). In order to isolate questions of basin partitioning from issues of distributed-parameter calibration, all calibration targets were basinwide; in other words, we did not calibrate to internal basin gauges or other internal targets. Based on the availability of data ($\sim$20-year period of record) and the dominant hydrologic processes in the Feather River, we chose to calibrate on SWE, ET, and basin outflow. Five calibration steps were used, each with a specific target variable, objective function, and set of calibration parameters. The order of the steps reflects the modeling order of hydrologic processes in PRMS. The objective functions and target variables for each step are as follows: daily RMSE of SWE; cumulative annual bias of ET; daily weighted sum of Kling-Gupta Efficiency (KGE, Kling et al., 2012; Gupta et al., 2009) and Log Nash-Sutcliffe Efficiency (LogNSE Nash and Sutcliffe, 1970) of full-natural flow (FNF); monthly weighted sum of KGE and LogNSE of FNF; and cumulative annual bias of FNF. For any objective functions using the KGE metric, all rows with missing observations were removed before the calculation. Daily FNF values were weighted between two metrics in order to capture the performance for both high and low flows. The objective functions for each step are listed in Table III.1.

Daily SWE maps published by Margulis et al. (2016) were masked to the Almanor and East Branch subbasins and then aggregated to catchment-wide values. Annual distributed evapotranspiration data were calculated following Roche et al. (2020) on a 30-m basis and were aggregated to catchment-wide values. Finally, daily runoff values for the outlet of each subbasin were computed from FNF values provided by PG&E. FNF is a reconstructed time series of runoff that would have occured in the absence of diversions or other human activity. Uncertainties in sensor readings can result in negative FNF values, so the values for the period of record were smoothed using a five-day moving-average window. Any FNF values that were still negative after the smoothing were masked to NaN.

PRMS, like many large-scale hydrologic models, has hundreds of parameters available for calibration, some of which may be individually calibrated for different months of the year or on a per-HRU-basis. We selected calibration parameters on the basis of previous sensitivity analyses of PRMS (Markstrom et al., 2016) and the availability of informative target data on which to calibrate them (Avanzi et al., 2020a). While some parameters are calibrated separately for different months, all are spatially lumped. The parameters calibrated at each

Table III.1. Calibration details for PRMS models

Step	Calibration variable	Parameters calibrated	Objective function
1	SWE	tmax_allsnow tmax_allrain (Jan-Dec) adjmix_rain (Oct-May) freeh2o_cap	Daily RMSE
2	ET	pref_flow_den soil_moist_max soil_rechr_max	Annual cumulative absolute bias
3	FNF (daily)	carea_max smidx_coef smidx_exp K_coef (per stream segement)*	0.75 × daily KGE + 0.25 × daily LogNSE
4	FNF (monthly)	fastcoef_lin fastcoef_sq sat_threshold slowcoef_lin slowcoef_sq soil2gw_max ssr2gw_exp ssr2gw_rate	0.75 × monthly mean KGE + 0.25 × monthly mean LogNSE
5	FNF (annual)	gwflow_coef gwsink_coef gwstor_min	Annual cumulative absolute bias

Asterisk indicates parameters that were only calibrated in the GIS PRMS model, as they do not appear in the GMM-based models. See Markstrom et al. (2016) for details on PRMS parameters.

step are presented in Table III.1. Phase partitioning parameters *tmax_allsnow*, *tmax_allrain*, and *adjmix_rain* (which determine the percentages of precipitation falling as rain and snow) were calibrated to basin-wide SWE. In addition, we calibrated *freeh2o_cap*, the free-water holding capacity of the snowpack. Subsurface parameters that are related to how much water is directed to the capillary soil layer were calibrated to ET. The capillary layer is the only soil layer from which ET can occur, so these parameters govern the largest source of ET in the basin (transpiration by plants). Though this neglects parameters related to evaporation from intercepted storage or sublimation from snow, these contributors to ET are much lower than transpiration by plants from soil storage. Finally, other subsurface parameters governing overland flow, interflow, and groundwater flow were calibrated to streamflow at various time steps (daily, monthly, and annual), reflecting the timescales over which we expect these processes to occur. For details on the use and physical meaning of these parameters, see Markstrom et al. (2015).

Based on the availability of input and calibration data, the total calibration period included water years 1998-2016. Instead of a traditional split-sample approach using continuous calibration and validation periods, a stochastic, multi-split process was employed to avoid biases that might arise from arbitrarily selecting a calibration period. Eleven of the nineteen water years in the period of record were randomly selected for calibration, a process that was repeated to give five 70%/30% calibration/validation splits (see Table B.2 in Appendix B.2 for specific years used for each split). Each GIS or GMM model was calibrated five times according to each of the splits and each calibration was assessed separately across its validation period.

All calibrations were performed using the Shuffled Complex Evolution (SCE) algorithm, which was specifically developed for large hydrologic models with many degrees of freedom (Duan et al., 1993, 1994). This algorithm was designed to handle arbitrary objective functions, differences in parameter sensitivities, and the presence of non-optimal local minima in the feasible space. In brief, SCE works as follows: randomly selected sample points are partitioned into complexes, which each evolve separately, allowing the parameter space to be explored more efficiently. Periodically, complexes are combined, shuffled, and re-partitioned into new complexes. This step allows for sharing of the information gained separately by each complex as it evolved. The algorithm stops when one of several possible specified convergence criteria is met. SCE requires the user to set a number of metaparameters, which may be individually set for each step. These were selected based on a combination of suggested default values (Hay and Umemoto, 2006) and trial-and-error. Table B.1 lists the metaparameters used for each step and a short explanation of how they were chosen. More details of the SCE algorithm can be found in Appendix B.2 and Duan et al. (1992, 1993, 1994).

The metrics used in model-performance assessment for each of the calibration/validation splits were daily absolute bias of SWE, cumulative-annual absolute bias of ET, and daily KGE, LogNSE, and root-mean-squared error (RMSE) of streamflow. RMSE gives more weight to accurately reproducing peaks in the time series, while LogNSE gives more weight to baseflow periods. It should be noted that SWE bias was calculated across all pixels, including those without snow, to be consistent with the PRMS model's calculation of basin-wide metrics. The performance of the models relative to observed values may therefore appear artificially good, but values are appropriate for comparison across models. In the

main text, metrics and scores are shown in figures as absolute values. Signed scores are reported in Appendix B, Tables B.3 and B.4.

To allow for comparison across metrics with different units, all twenty-five calibrations (five for each of the five models) were ranked for each metric. Then, the average rank across all five metrics for each calibration was calculated and the median average rank was the "score" of that model. Unless otherwise noted, all performance values reported are for the calibration with the median average rank.

Other performance metrics focusing on conditions that are of particular interest to forecasters and modelers, such as peak SWE and flood periods, are also presented, but were not used in scoring model performance. These include the Relative Error of High Flows (REHF) (Silvestro et al., 2018) to assess reproduction of peak flows, peak SWE error, date of peak SWE, and baseflow error. Peak flows were identified as the top 5% of flows across the period of record and days of peak flow were the days these occurred. Only the peaks occurring within a calibration's validation period were used to calculate REHF. All other metrics were calculated separately for each validation year and averaged.

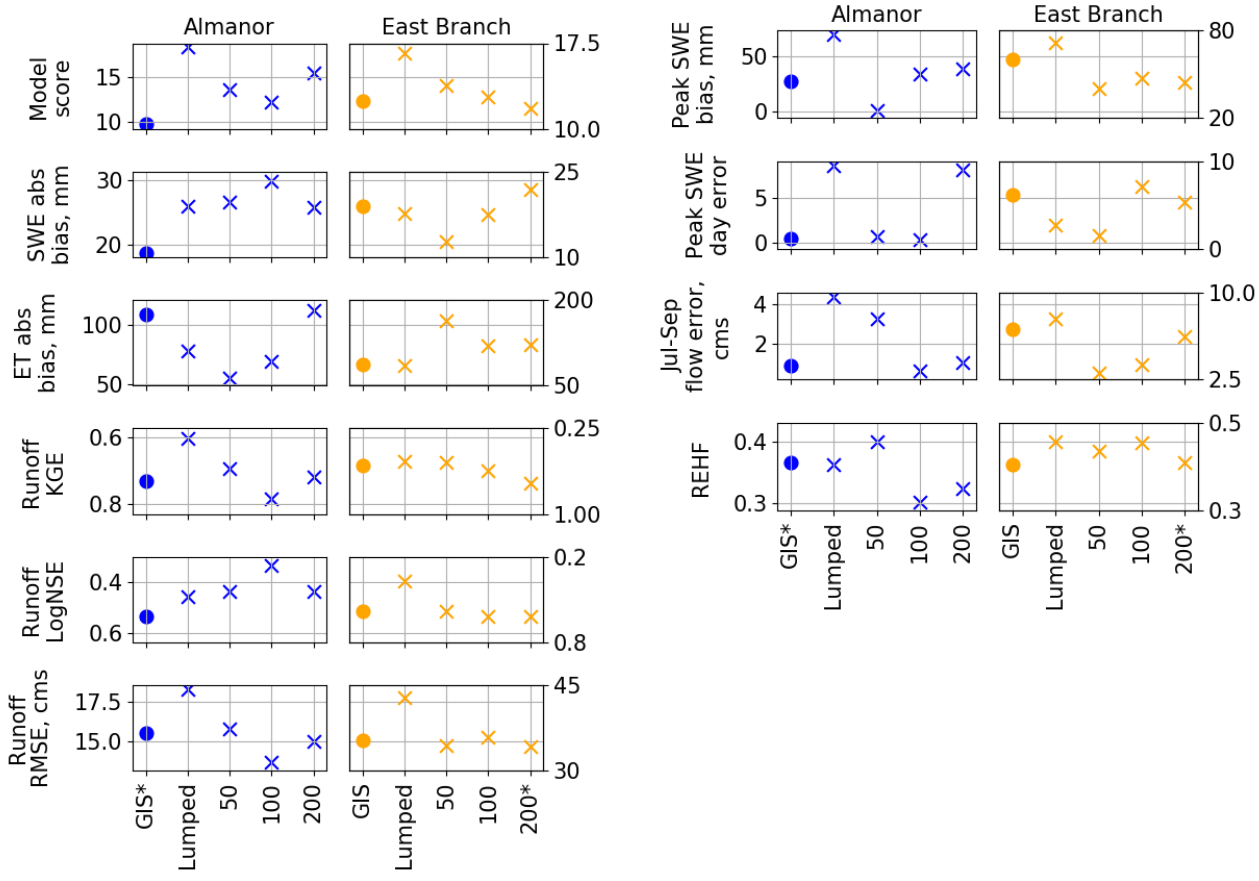
III.3 Results

This section presents results on GIS and GMM model performance. Section III.3.1 gives results under average conditions (i.e., metrics computed across validation periods), which were used to rank overall performance of the models. It also discusses computational resources required to run the GMM algorithm. Next, we present further analysis aimed at verifying GMM performance during periods of extreme conditions (Section III.3.2). Finally, we present the results of the univariate GMM models, demonstrating how these can inform feature selection for different parts of the water balance (Section III.3.3).

III.3.1 GMM model set-up and performance

In both subbasins, the GMM model performance is comparable to and, in some cases, better than the GIS models with respect to all metrics calculated for the validation periods (Figure III.3). The median average rank (Figure III.3a) shows that the GIS model is best-performing (lowest ranked) in Almanor and the 200-location GMM model is best in the East Branch. The best-performing GMM model in Almanor is the 100-location one.

Within a subbasin, there is little variability in performance across models, including both GIS and GMM models. SWE and ET bias in both catchments, for example, vary by less than 20 mm. Runoff RMSE is variable between the two catchments, reflecting the difference is average flow, but models for a given catchment show similar performance. However, no single model consistently performs best for all components of the water balance. Performance rankings are more consistent in the East Branch than Almanor, particularly with respect to streamflow metrics. Here, the 200-location GMM generally performs best with the exception of SWE. In both subbasins, the lumped GMM model, which represents a baseline from which to assess improvements due to any type of spatial distribution, ranks somewhere in the middle with respect to SWE and ET performance, but consistently performs poorly with respect to streamflow.



(a) All validation periods

(b) Extreme and peak periods

Figure III.3: Model performance metrics Almanor and the East Branch (note differing axis extents for the subbasins). Axes are oriented such that the best performance will appear lowest on the plot. × denotes GMM models. The best-performing model (lowest average median rank across calibrations) in each subbasin is marked with an asterisk (*).

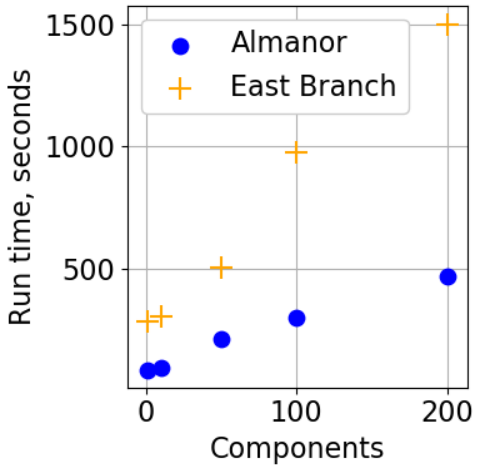


Figure III.4: GMM algorithm runtimes by subbasin and number of components

In the East Branch, overall ranking consistently improves with higher numbers of GMM modeling locations, but in Almanor, this is only true up to 100 modeling locations. Thus, the best-performing GMM uses half the number of modeling locations in Almanor than the East Branch (Figure III.3a), possibly reflecting the difference in catchment area (the area of the East Branch is approximately double that of Almanor). The optimal number of modeling locations may also be influenced by the input-data raster resolutions.

The computational times required to run the GMM algorithm were between 80 and 1500 seconds on a single core of a high-performance computing cluster (3.9 GHz; Figure III.4). This time includes all steps from raster sub-sampling, scaling, fitting the GMM model, and the nearest-neighbor search for the physical modeling locations. Times varied by subbasin (East Branch models took longer to run than Almanor due to the larger raster size) and number of components.

III.3.2 Extreme and peak periods

As with average-condition metrics, the GMM-based approach yields comparable, if not better, performance than the GIS method during extreme periods (Figure III.3b). In particular, the GMM models that perform best in each subbasin under average conditions (100-location in Almanor and 200-location in the East Branch; Figure III.3a) also match or exceed GIS performance during extreme periods. We note that the lumped model performs poorly in both subasins with regard to extreme periods.

Peak SWE marks the transition from accumulation to ablation season in the Sierra; both timing and magnitude of peak SWE are important seasonal benchmarks in snow-dominated basins. On average, peak SWE was better simulated in Almanor, but day of peak SWE was better simulated in the East Branch. Peak SWE tended to be overestimated in both basins (Table B.4 in Appendix B). Day of peak SWE was estimated later than observed in the East Branch; there was no consistent pattern in Almanor.

July-September flow was used to capture baseflow performance, and REHF was used to assess peak flows. Like the SWE metrics, performance was better for both in Almanor.

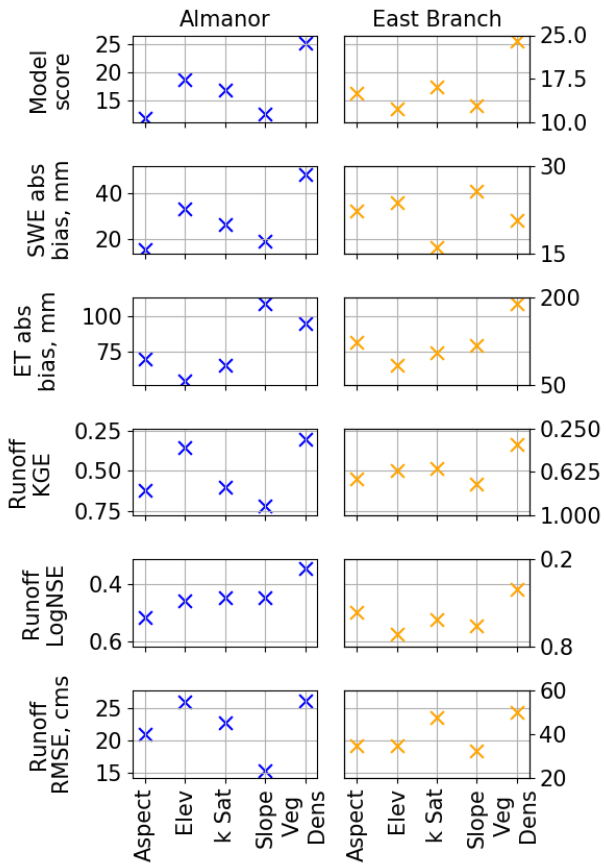


Figure III.5: Performance metrics in Almanor and the East Branch for univariate models (note differing axis extents for the subbasins). Axes are oriented such that the best performance will appear lowest on the plot.

For baseflow in particular, observed flows are lower on average in the East Branch, making the performance in Almanor even more comparatively strong. Peak flows were generally challenging to capture in both subbasins, with minimum REHF of 0.3 (100-location GMM in Almanor). Though there was little consistency in model rankings between baseflow and peak flows, the best-performing GMM models were again able to meet or exceed GIS performance.

III.3.3 Univariate GMMs

Univariate GMM performance was generally worse than multivariate models (Figures III.3a and III.5), though metrics fell within the same order of magnitude. In the East Branch, the elevation-driven GMM performed best (as based on median average rank) followed by (in order) the models driven by slope, aspect, saturated hydraulic conductivity, and vegetation, with median rank of slope-driven model only slightly larger than elevation-based model. In Almanor, models in order of performance from best to worst were based on slope, aspect, saturated hydraulic conductivity, elevation, and vegetation. Slope performed particularly well with respect to runoff RMSE and KGE, while aspect performed better with respect to SWE bias and runoff LogNSE.

III.4 Discussion

The GMM method for selecting modeling locations has multiple benefits compared to GIS models, which are discussed below.

III.4.1 Model performance

The GMM method provides a sound, objective basis for spatially distributing hydrologic models. GMM-based models match the performance of the GIS models under both average and extreme conditions for metrics related to both streamflow and other model variables such as SWE and ET. They accurately simulate water-balance components over time and provide more-accurate spatial distribution of streamflow generation than do GIS models.

The best-performing GMM model exceeds the overall performance of the GIS model in the East Branch, but not in Almanor. Thus the variables selected for GMM prediction may be more relevant for and comprehensive of water-balance partitioning in the East Branch than Almanor due to the particular hydrology of the catchments. Almanor is more subsurface-dominated than the East Branch, and saturated hydraulic conductivity, the only predictor used here related to subsurface conditions, is a relatively limited characterization of soil and groundwater flow. Another factor may be the relative importance of the GMM variables: the most informative variables in Almanor, based on the performance of the univariate models, were aspect and slope (Figure III.5). The GIS and GMM distributions of these factors were similar and largely consistent with the raster values (Figure B.4 in Appendix B). The most informative variable in the East Branch was elevation. Here, the GMM models span a greater range than the GIS distributions, which may have contributed to improved performance. In addition, the overall lower elevation of the East Branch and the fact that its highest elevations are rain-shadowed means that a greater proportion of the precipitation in East Branch falls in the rain/snow transition than in Almanor (40% versus 33%, where the rain/snow transition is defined between 1300 and 2200 m; Cui et al., 2020). This means that uncertainties in modeling precipitation phase (well reported in the literature; e.g. Harpold et al., 2017; Jennings and Molotch, 2019; Feiccabrino et al., 2015) will affect the East Branch more than Almanor. Thus, the East Branch may be more sensitive to tuning the elevational distribution of modeling locations than Almanor; in other words, there may be more potential for the GMM approach to improve results.

In addition to capturing temporally averaged metrics, the GMM models demonstrated the ability to accurately reproduce periods of extreme or peak conditions. Good performance during extreme flow periods is particularly important for applications like flood forecasting, but increasingly necessary for all streamflow modeling as climate change increases year-to-year variability and induces more severe weather events. Day of peak SWE, for example, has traditionally been estimated as April 1st (Montoya et al., 2014). This estimate has always been uncertain due to seasonal weather characteristics and elevation effects, but is becoming increasingly inexact due to climate-change-induced shifts in precipitation (Margulis et al., 2016). Thus, it is valuable for forecasters to be able to model the date of peak SWE rather than relying on the April 1st estimate. The ability to reproduce baseflows – and, correspondingly, low-flow periods – is also of greater concern as length and severity of dry periods in arid regions are projected to increase (Williams et al., 2020; Woodhouse et al., 2010; Cayan

et al., 2010). In each of these cases, the best-performing GMM model is able to meet or exceed the performance of the GIS model. Furthermore, there was consistency between GMM models that performed well under average conditions and those that performed well under peak conditions, meaning that forecasters would not need to rely on a separate model for extreme periods. Since models calibrated to average periods do not always work well under extreme conditions (see, e.g., Vaze et al., 2010), the consistent performance of the GMM-based models is a significant advantage.

Finally, in both subbasins the lumped model had the worst overall performance, particularly with regards to streamflow, which may be partly the result of uncertainties in the simulation of subsurface processes, upon which runoff predictions are dependent, due to model structure and lack of data. Moreover, the lumped models perform even worse with respect to other models during extreme periods. This result is an argument in favor of performing the additional steps required to run the GMM algorithm and so obtain a spatially distributed model for montane regions. The spatially distributed models do not show this uniform drop in performance. For forecasters in particular, accurate modeling of streamflow and the ability to capture both flood and drought conditions is imperative to optimize dam operations and to protect infrastructure and communities downstream. Our results are consistent with other studies (e.g., Lobligeois et al., 2014) that have shown that spatially distributing hydrologic models can yield significant improvements over lumped models in basins with heterogeneous climatic inputs. This finding supports further attention being given to the distribution step (i.e., selection of modeling locations) of model set-up.

III.4.2 Modeling set-up

Another key advantage of the GMM method is its efficiency and repeatability. GMM requires only rasters of input variables, thus combining the data-processing advantages of pixel-based models while still being based in physical basin characteristics. Once the rasters are prepared, running the GMM algorithm from start to finish, including subsampling and saving the outputs, required less than half an hour on a high-performance-computing core. As long as a seed is set in the random number generator for the GMM optimization, the process is also repeatable. While the GMM method does not address questions of scalability of parameters or automatically identify optimal resolutions, this efficiency can be leveraged to test multiple spatial resolutions and allow modelers to understand how the resolution influences their results. The ease of setup also allows modelers to test different combinations of model input variables and understand the drivers of hydrologic processes in their basin. This can help inform which variables to use as inputs to the GMM. The GMM algorithm requires no specific software and can be implemented through open-source products, as was done for this study. There are no theoretical limits on modeling locations using GMM and all locations necessarily represent equal areas to comply with the multivariate selection process. This removes questions of relative HRU size and decisions about the maximum range of HRU areas. Finally, since the GMM method is separate from calibration, it can be applied for any number of calibration designs, including different optimization algorithms; single- or multiobjective functions; and semi- or fully distributed parameters.

Traditional HRU delineation, on the other hand, is necessarily subjective: delineation usually begins by identifying areas with similar topography using a digital elevation model

(DEM), but there are few norms or guidelines to selecting the number of HRUs to use (and, by extension, their average size) other than the resolution of the input data and the computational power available to run the model. Once initial HRUs are delineated, smaller HRUs are generally merged into neighboring larger ones so sizes fall within a similar range; which HRUs to merge and where is entirely subjective. Some common software tools including GIS Weasel automatically and randomly merge smaller HRUs, but do not contain the ability to set the seed of their random number generators, making this process impossible to replicate (Viger and Leavesley, 2007). Though the time required for GIS-based HRU delineation is not consistent, since it depends on the size and topography of each basins, our experience in this study and conversations with modelers and forecasters suggest the process is on the order of days to weeks. Since the spatial-distribution process is so labor intensive, the assumptions made during this process cannot be easily tested by creating alternative versions of the model.

GIS-based HRU delineation also presents theoretical problems, including the fact that hydrologic processes are simulated at the geometric center of a supposedly homogeneous HRU. This means that extreme elevations will never be represented by the model, potentially missing areas that are significant contributors to runoff production. Another issue is that if an HRU is not convex, its geometric center is not guaranteed to fall within the HRU or even within the river basin itself.

III.4.3 Spatially distributed performance

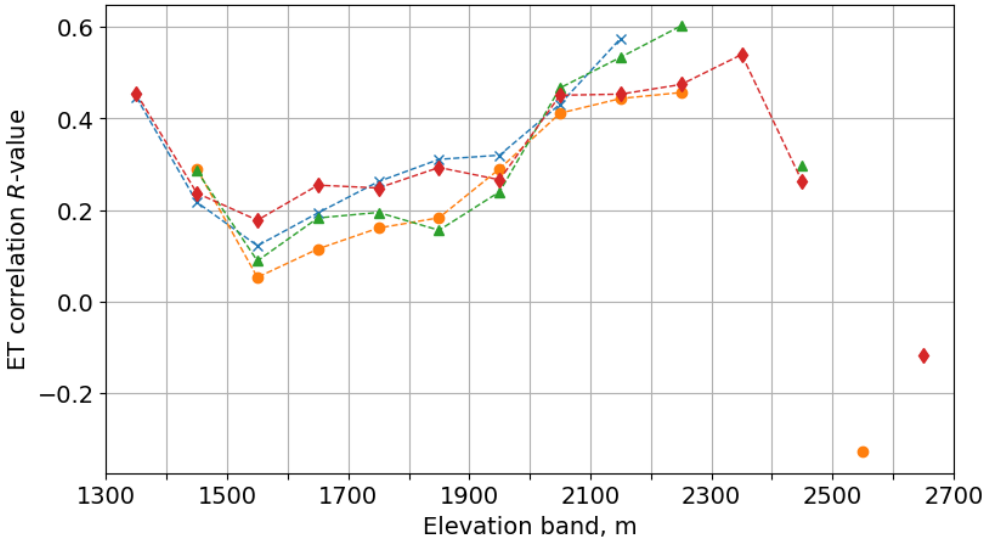
Optimized GMM modeling locations also lead to more realistic spatial representation of the basin in PRMS, which has implications for model interpretation and distributed performance. For example, the extent of the elevations represented in the GIS model is less than half the true range (the GIS model covers 1391 to 2155 m in Almanor and 1103 to 2001 m in the East Branch, while the range of the DEMs is 1365 to 2950 m in Almanor and 700 to 2550 m in the East Branch). Due to this limited elevational range, all hydrologic processes in the GIS model can occur only up to 2200 m in Almanor and between 1100 and 2100 m in the East Branch. Compared to the GIS models, the best-performing GMM models cover a 42% greater range in Almanor (1406 to 2492 m) and 38% greater range in the East Branch (1016 to 2263 m). Though elevation gradients are only one of many types of spatial heterogeneity, they are particularly relevant due to the strong orographic influence on precipitation in our study area (Roe, 2005; Roe and Baker, 2006). Their greater range means the GMM models are better positioned to capture processes with strong elevational dependence, including SWE distribution, vegetation, and timing of runoff generation. This finding applies to these study sites in particular, but based on the theoretical limits on HRU elevations as discussed in Section III.4.2, we expect the GMM method to give broader elevational representation than the GIS method in any other montane catchment.

The implications of using GMM versus GIS for spatial distribution are clear when we examine elevational trends in model performance for ET and runoff (Figures III.6 and III.7, respectively). In Figure III.6, average daily volumetric bias shows how well the models match overall ET volume at different elevations, while the Pearson correlation coefficient shows how well temporal patterns are simulated. Bias here may be driven by two factors: 1) errors in data or modeling assumptions or 2) biased elevational distribution of area in the model. We

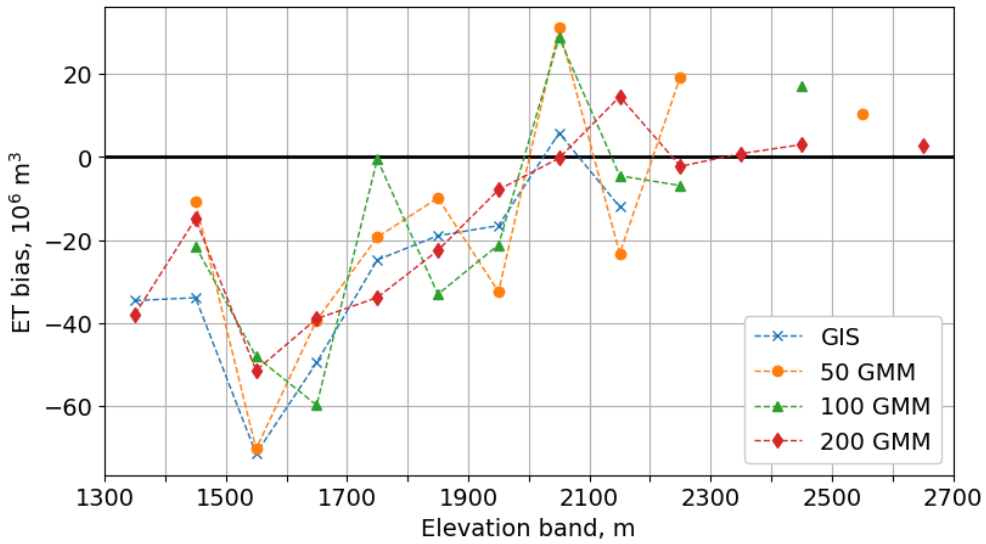
see that while ET correlation in Almanor shows a clear "U-shaped" pattern across all models, only the GMM models that cover elevations above 2300 m reveal a significant drop-off at the higher elevations. This pattern is not captured by the GIS model, nor is it seen in the East Branch, where elevations do not exceed 2300 m (Figure B.3 in Appendix B). Notably, this drop in correlation performance occurs at about the same elevation above which the ET bias of the Almanor models becomes consistently positive (Figure III.6b). We hypothesize that these patterns are related to ET modeling above and below the tree line, which since the highest portions of Almanor, including Mount Lassen above about 2400 m, are largely free of vegetation. Elevations below the tree line are transpiration dominated, while those above the tree line are evaporation dominated; thus, ET calculations in PRMS appears to underestimate the transpiration component and overestimate the evaporation component. These errors could be related to problems with input data (e.g. inaccurate vegetation densities) and/or to model-structural error. In the evaporation-dominated higher elevations, structural issues may include estimating sublimation from the snowpack or evaporation from soil storage. Below the tree line, bias may be related to underestimation of the depth of the root zone or other problems with subsurface modeling. The trend of ET error with elevation may compound or be related to errors in timing of SWE accumulation and runoff. These patterns would not be clear using only the GIS model, which does not capture any location above the tree line.

Simulation of the distribution of runoff production across elevations also benefits from the broader spatial range in the GMM models (Figure III.7). Due to misrepresentation of area per elevation band, all models tend to over-produce runoff at mid-to-low elevations (1200 to 1600 m in the East Branch) as compared to observed precipitation minus evapotranspiration. (P-ET is a first-order estimate of runoff production, which is not directly observable by elevation band.) The overestimation at mid-elevations compensates for underestimations at higher elevations (particularly elevations not represented at all where, by default, runoff production is zero), in order to match overall runoff volume. This error is greatest in the GIS model, which represents the narrowest range of elevations of any model.

This misrepresentation of contributing area may lead simulated runoff to interact with other water-balance components in non-physical ways. Over- or under-generation of runoff may lead to errors in partitioning infiltration versus runoff, potentially impacting ET simulation since the majority of ET in vegetated areas is transpiration from soil storage and generally receives priority allocation of runoff over streamflow (Bales et al., 2018a). Moreover, misrepresentation of contributing area may lead to particularly poor representation during extreme periods like drought. It has been shown that vegetation growth at lower elevations of some Northern Sierra basins may become water limited during droughts, even as vegetation growth in the basin as a whole is energy limited. Thus, failure to simulate these lower elevations and corresponding vegetation water demand may lead to the models overestimating runoff during droughts; on the other hand, failure to capture the higher elevations may mean the models miss an important drought mitigation factor (Bales et al., 2018a).



(a) Correlation over time



(b) Volumetric bias

Figure III.6: ET performance for each elevation band in Almanor, averaged across calibrations. Tick marks indicate upper and lower bounds of 100-m elevation bands. The lumped GMM model is not shown.

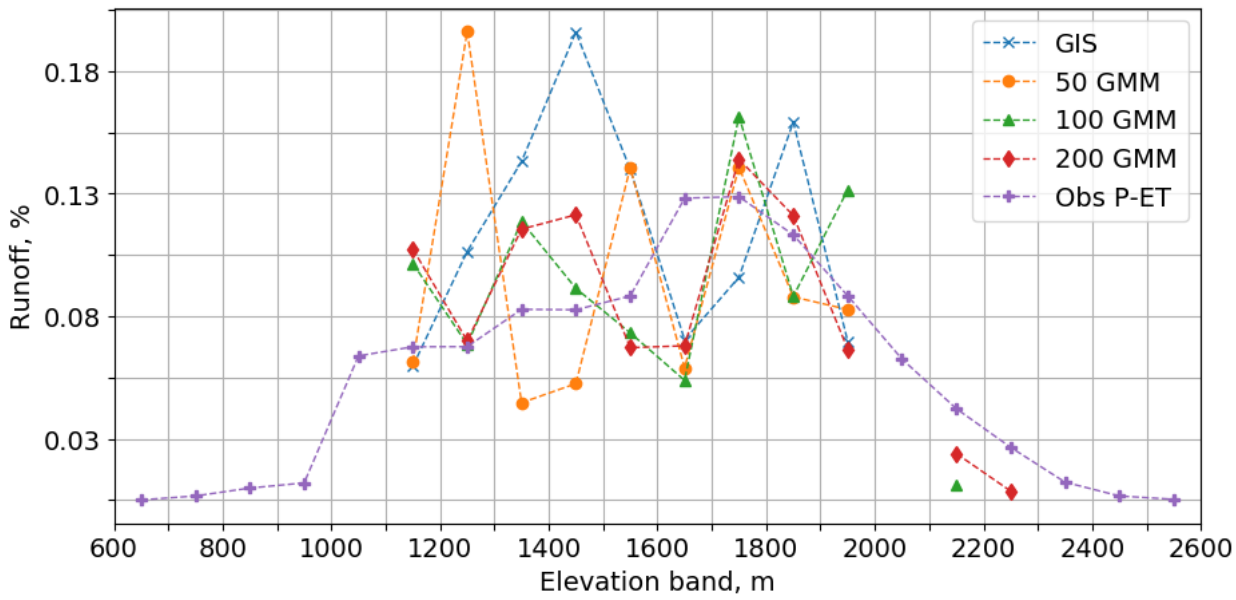


Figure III.7: Percent runoff by elevation band in the East Branch. Tick marks indicate upper and lower bounds of 100-m elevation bands. The lumped GMM model is not shown.

III.4.4 Univariate models

The univariate models generally perform worse than the multivariate models, which capture more of the factors that control the water balance in montane catchments (Figure III.5). However, the univariate models provide insights into the most important drivers for different hydrologic processes in the basins. The top-performing univariate model in the East Branch is elevation-based. Since the East Branch sits largely in the rain/snow transition zone, elevation is a critical factor for determining runoff timing by way of precipitation phase. However, both the aspect and slope models performed better than the elevation model with regard to overall SWE bias, suggesting that the rain-shadowed nature of the East Branch and strong directional precipitation patterns are important factors in influencing accumulation and ablation. In addition, slope and aspect may influence the timing and shape of the SWE ablation curve, since they affect the amount of incident solar radiation a site will receive (Maxwell et al., 2019).

In Almanor, the top-performing GMM models were based on slope and aspect, followed by those based on saturated hydraulic conductivity, elevation, and vegetation. Almanor is a higher-elevation basin than East Branch with more area above the rain/snow transition zone. Thus, elevation may be less informative since precipitation phase is more consistent than in the East Branch. Instead, slope and aspect are greater controls on the timing of snow accumulation and melt and, by extension, runoff. In addition, baseflow fed by groundwater is a larger component of streamflow in Almanor than the East Branch, so soil characteristics (i.e., saturated hydraulic conductivity) may be more relevant for determining streamflow.

The relatively good performance of the aspect-driven models in both catchments may be due to rain-shadowing effects: direction of slope matters not only for snow ablation due to solar radiation but also for snow accumulation, since the wettest parts of both basins are

the western-facing, non-rain-shadowed portions along the main stem of the North Fork. The saturated-hydraulic-conductivity-based model performed reasonably well in both subbasins, but, as expected, topographic features were still overall most relevant for runoff generation. The vegetation-density model gave poor results across both basins, indicating that subsurface conditions may be stronger drivers of ET variation in the Feather River. Since the majority of the land cover in both subbasins is forest, vegetation density may be less informative due to relatively little variation across the landscape.

The univariate models and their differing importance in each subbasin demonstrate the physical basis for GMM-based models and their ability to provide hydrologic processes understanding in headwater catchments. We suggest that univariate GMMs could be used in practice to assess the most relevant input features before running a multivariate GMM to distribute a new model. This is relevant for both modelers and scientist seeking to improve forecasting performance, prioritize data collection, and better understand the hydrologic cycle.

III.5 Conclusion

We introduce a new method for spatial distribution of hydrologic models using the Gaussian Mixture Models (GMM) algorithm and demonstrate its use in two geologically distinct headwater catchments of the Sierra Nevada. Unlike traditional GIS-based methods, the GMM method is objective, repeatable, and computationally fast (on the order of minutes). The method identifies the set of modeling locations that best represent the basin as a whole, leveraging an efficient machine-learning tool while being grounded in physical basin properties. Analysis shows that GMM-based models are able to match or exceed the performance of traditional, GIS-based models with respect to both average and extreme conditions for both streamflow and other water-balance components. Furthermore, we show that the modeling locations selected using GMM better represent the geometry of the basin and thus more accurately reproduce the spatial distribution of processes such as runoff production. Finally, we show how the method can be adapted to test multiple feature combinations and identify the relative importance of a basin's hydrologic drivers. An elevation-based GMM model performed best in the study basin that sits primarily on the rain/snow transition zone, while slope- and aspect-based models performed best for the higher-elevation catchment. Further research should investigate how different input components, climates, and topographies influence GMM performance.

The improved spatial representation of GMM-based hydrologic models creates a more-robust decision-making and process-understanding tool for water-supply agencies, utility companies, and flood-control operators, especially in topographically heterogeneous basins. In addition to enhancing models directly, the efficiency of the GMM method would encourage more-regular model updates than is usually possible for operators given time and labor constraints when using traditional methods. This allows agencies to stay abreast of changes to their basins, such as land use or vegetation coverage, and advances in model structure and data collection technology. Leveraging advances in machine learning, GMM is a powerful and promising new tool for hydrologic modeling.

Chapter IV

Drought-induced shifts in the allocation of available water: a Budyko approach

Based on the paper of the same name submitted for publication by:
Tessa Maurer, Francesco Avanzi, Steven D. Glaser, Martha Conklin, Roger C. Bales

Abstract

An inconsistent statistical relationship between precipitation and runoff has been observed between drought and non-drought periods, with less runoff usually recorded during droughts than would be predicted using non-drought relationships. Most studies have examined these shifts using multi-linear regression models, which can identify correlations but are less appropriate for analyzing underlying hydrologic mechanisms. In this analysis, we show how the Budyko framework can be leveraged to quantify the impact of shifts in water allocation during drought using thirty years of spatial data maps for fourteen basins in California. We distinguish "regime" shifts, which result from changes in the aridity index along the same Budyko curve, from "partitioning shifts", which imply a change in the Budyko parameter ω and thus to the relationship among water-balance components that governs partitioning of available water. Regime shifts are primarily due to measurable climatic changes, making them predictable based on drought conditions. Partitioning shifts are related to nonlinear and indirect catchment feedbacks to drought conditions and are thus harder to predict *a priori*. We show that regime shifts dominate changes in absolute runoff during droughts, but that gains or losses due to partitioning shifts are still significant. We further discuss how basin characteristics and feedbacks correlate and may influence these shifts, finding that low aridity, high baseflow, a shift from snow to rain, and resilience of high-elevation runoff correlate to an increase in runoff as a fraction of precipitation during droughts. This new application of the Budyko framework can help identify mechanisms influencing catchment

response to drought, with implications for water management in arid and drought-prone regions.

IV.1 Introduction

Droughts can threaten human and natural systems worldwide, accounting for more than 50% of all natural hazard deaths over the course of the 20th and early 21st century (Van Loon, 2015; Maskey and Trambauer, 2015). As baseline water stress intensifies globally due to growing populations and land-use changes (Hofste et al., 2019), the impact of meteorological and hydrologic droughts may become more severe (Masih et al., 2014). In Mediterranean climates with highly seasonal precipitation, droughts exacerbate already significant water-management challenges, as these basins typically rely on intricate systems of natural and built water storage to maintain water supply across regularly occurring seasonal and multi-year dry periods (He et al., 2017; Woodhouse et al., 2010).

The need to adequately understand and predict the water-balance implications of droughts is becoming more acute as climate change makes basins susceptible to more severe and prolonged droughts (Dai, 2013; Trenberth et al., 2014; Woodhouse et al., 2010). Currently, it is understood that relationships between water-balance components are not consistent between drought and non-drought periods. A particular focus is the change or shift in the precipitation-runoff relationship during droughts, which usually results in less observed runoff per unit of precipitation than would be predicted using non-drought relationships. These drought-induced shifts have been observed in basins around the world (e.g., Saft et al., 2016; Avanzi et al., 2020b; Tian et al., 2020b), compounding water shortages for municipal, industrial, and agricultural systems.

Despite documentation of these shifts and their implications for human water supply, it is not fully understood which hydrologic mechanisms trigger them nor whether these mechanisms are consistent across different droughts and basins (Bales et al., 2018a). Most studies of drought-induced changes to the precipitation-runoff relationship have used statistical models to identify and analyze shifts. Saft et al. (2016) used multi-linear regression to identify factors associated with shifts during the decade-long Millennium Drought in Australia, finding that shifts are correlated to pre-drought catchment characteristics that make certain basins more susceptible to shifts, including aridity, rainfall seasonality, and interannual variability of groundwater storage. Potter et al. (2011) also looked at the Millennium Drought, using a regression approach to calculate the sensitivity of streamflow to anomalies in rainfall and maximum daily air temperature. These factors accounted for 73% of the reduction in streamflow, but left the remainder unexplained. Avanzi et al. (2020b) followed the same statistical approach as Saft et al. (2016) for California's Sierra Nevada, identifying shifts in the majority of basins in that mountain range and using a physically based model to identify interactions between evapotranspiration (ET) and subsurface water storage as the source of the shift. Examining a slightly wetter, monsoon region in China, Tian et al. (2020b) developed a multivariate generalized additive model to identify basin characteristics, including climate characteristics, most associated with large negative shifts in the runoff coefficient (ratio of runoff to precipitation). They found that drier and lower catchments, based on the aridity index and mean catchment elevation, respectively, were more susceptible to large shifts.

Despite ongoing uncertainty in the mechanisms causing shifts, these are understood to relate at least in part to catchment-related feedbacks between vegetation water demand (ET) and subsurface storage. During wet periods when vegetation water demand is easily

satisfied, ET becomes a more or less constant term, bounded by the energy availability (Bales et al., 2018a; Budyko, 1974), and the relationship between precipitation and runoff appears linear. During droughts, however, catchment feedbacks between ET and subsurface storage can trigger a hysteretic response in the water balance that appears to shift in the precipitation-runoff relationship (Bales et al., 2018a; Avanzi et al., 2020b; Goulden and Bales, 2019). First, soil water storage can decouple ET from precipitation by allowing vegetation to withstand periods of mild to moderate drought. The extent to which soil storage acts to buffer the precipitation deficit depends on lithology and pre-drought water content of the soil as well as the vegetation type and rooting depth, which can vary significantly among and within basins (Bales et al., 2018a; Oroza et al., 2018; Hahm et al., 2019b; Tague and Grant, 2009). For example, in the Sierra Nevada of California, deep regolith storage can support mixed conifer and evergreen needle-leaf species with deep rooting depths, but ET in grasslands and pine-oak forests is more responsive to yearly changes in precipitation (Bales et al., 2018a; Klos et al., 2018). In areas like California with highly seasonal precipitation, dry-season baseflows can offer an approximation of the baseline soil water storage in a basin. Basin aridity may influence overall water available to support vegetation through dry periods (Avanzi et al., 2020b; Saft et al., 2016; Tian et al., 2020b). The second major factor is the vegetation feedback to reduced water availability. Even in areas with high soil water storage, prolonged and/or severe droughts can deplete subsurface storage, eliciting vegetation stress responses such as stomatal closure (Avanzi et al., 2020b; Goulden and Bales, 2019) or, in extreme cases, tree die-offs (Bales et al., 2018a). These changes to transpiration demand also influence water allocation between ET and runoff, even after the dry period is over. Finally, climate-induced changes other than precipitation deficit may contribute to shifts in the water balance. Changes in temperature may increase evaporative demand and thus increase ET in areas with sufficient water (Teuling et al., 2013; Mastrotheodoros et al., 2020). In basins with significant snowfall, temperature also influences precipitation phase and the elevation of the snow line (Zhang et al., 2017a). These changes, in turn, influence the timing of available water (Rungee et al., 2019; Avanzi et al., 2020b) and the spatial distribution of runoff production in the basin (Avanzi et al., 2020b; Bales et al., 2018a).

Here, we revisit the question of drought-induced shifts in the precipitation-runoff relationship through the lens of the Budyko framework in California's Sierra Nevada. The Budyko hypothesis (Budyko, 1974) is a conceptual water-balance model that has been used in numerous catchments around the world to characterize the long-term water balance as a trade-off between supply (precipitation) and demand (PET; e.g., Li et al., 2013; Zhang et al., 2008, 2001; Greve et al., 2016; Moussa and Lhomme, 2016; Shen et al., 2017; O'Grady et al., 2011; Gnann et al., 2019). As a conceptual model, the Budyko framework can provide a macroscale understanding of the relationship between water-balance components across a catchment, while minimizing the need for high-resolution data or large parameter sets (Hrachowitz and Clark, 2017). While the Budyko approach has been leveraged to examine the water-balance impacts of general climatic changes (Li et al., 2019; Wang and Alimohammadi, 2012), vegetation, (Zhang et al., 2016; Ning et al., 2019; Oudin et al., 2008), and land-use changes or other human activity (Liu et al., 2017; Shen et al., 2017), its application to drought impacts specifically has been limited (see, e.g., Huang et al., 2017).

We apply the Budyko framework to the question of drought-induced shifts in the precipitation-runoff relationship for the first time, characterizing the water balance across

three droughts in fourteen basins in the Sierra Nevada. We distinguish "regime" shifts, which result from changes in the aridity index along the same Budyko curve, from "partitioning shifts", which imply a change in the Budyko calibration parameter and thus to the relationships between evaporative demand, precipitation, and ET that govern partitioning of available water. We use this new framing to answer the following questions: 1) Are changes to the precipitation-runoff relationship during droughts captured in the Budyko framework?; 2) What is the impact of these changes on partitioning of available water during drought?; and 3) Can any correlation be identified between partitioning shifts and known basin drought response mechanisms?

IV.2 Methods

IV.2.1 Study area

Our study area comprises the fourteen major river basins draining into the Sacramento-San Joaquin Valley of California (Figure IV.1). All basins in the study area have a Mediterranean climate, with seasonal precipitation that falls largely between October and May. The wet season is offset from the peak growing period, which occurs in the warmer summer months. Most basins have headwaters on the eastern edge, with elevations decreasing smoothly to the west. The exceptions are the Shasta, which has headwaters to the east, north, and far western edges and drains to the south; the Feather, the eastern two-thirds of which are lower and rain-shadowed; and the Kern, which has headwaters in the northern portion of the basin and drains to the south. Elevations generally increase from north to south in the Sierra Nevada, from an average elevation of 1530 m in the Feather to 2200 m in the Kern. Shasta has a high peak elevation (4300 m), but little surface area above 2400 m. For ease of reference, we refer to all study basins collectively as the Sierra Nevada. The northern basins or Northern Sierra Nevada includes the Shasta, Feather, Yuba, American, and Cosumnes basins; the Central Sierra Nevada includes the Mokelumne, Stanislaus, Tuolumne, and Merced; and the Southern Sierra Nevada includes the San Joaquin, Kings, Kaweah, Tule, and Kern.

IV.2.2 Data

We used gridded data products of precipitation, temperature, and evapotranspiration to estimate water-balance components for this study. Precipitation (P) and minimum and maximum temperature on the daily timestep were obtained from the Parameter-elevation Regressions on Independent Slopes Model (PRISM; Daly et al., 2008). Using the gridded temperature products, potential evapotranspiration (PET) was calculated with the Hamon method (Hamon, 1963) on a daily, pixel-by-pixel basis using mean daily PRISM temperatures. ET datasets were available on an annual (water year) basis, calculated for the Sierra Nevada following Roche et al. (2020). PRISM data have a pixel size of 800 m and were downscaled using a nearest-neighbor algorithm to match the 30-m pixel size of the ET data. Finally, runoff (Q) was obtained in the form of monthly reconstructed unimpaired flow values at the outlet of each river basin from the California Data Exchange Center (`http://cdec.water.ca.gov/index.html`); see Appendix C.1 for the gauges used. Raster

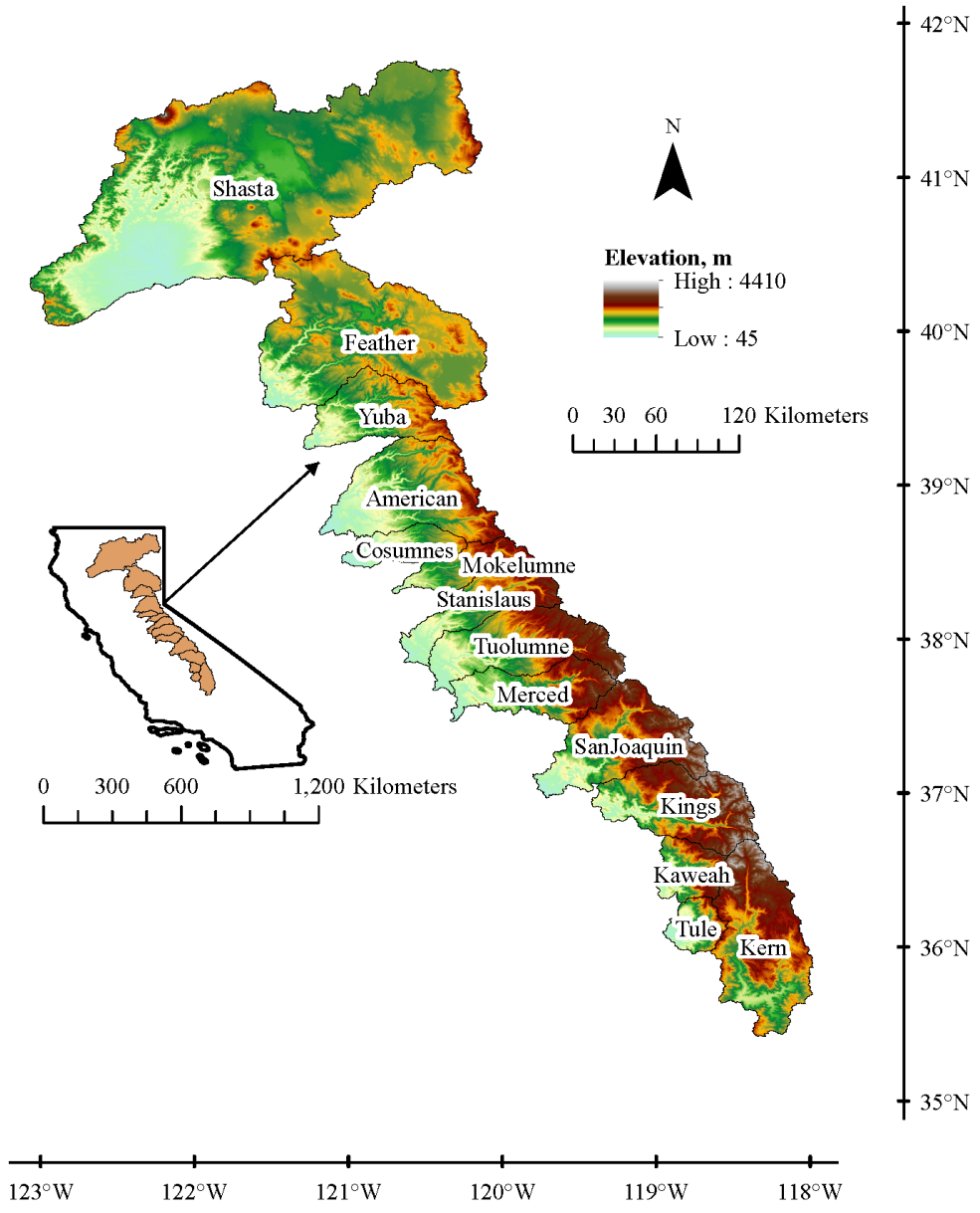


Figure IV.1: Map of California, indicating extent of the river basins used in this study. The Northern Sierra extends from the Shasta to Cosumnes, the Central Sierra from the Mokelumne to the Merced, and the Southern Sierra from the San Joaquin to the Kern.

data were binned to two spatial scales we considered in this study: basin-wide and by 100-m elevation bands. All data were obtained for water years 1985–2018 and aggregated from their original timesteps to the annual (water year) timescale. (The water year in California runs from October 1st through September 30th and is referred to by the latter of the two calendar years that it spans.) Finally, annual precipitation data were adjusted by the long-term average residual of $P - ET - Q$ so total basin storage over the period of record was zero.

IV.2.3 Extended Budyko framework

The original Budyko formulation conceived of the water balance as a trade-off between supply, in the form of water from precipitation, and demand, in other words, potential evapotranspiration. Their mutual availability determines the partition of water between evapotranspiration and runoff. The aridity index, PET/P, is plotted against the fraction of precipitation that goes to ET (evaporative index). An aridity index less than one indicates an energy-limited area, where vegetation productivity is limited by potential evapotranspiration, while an aridity index greater than one indicates a water-limited area, where water availability is the limiting factor. This formulation was applied strictly to the long-term (i.e., 10+ years) water balance, a timescale over which change in storage could be assumed to average out to zero. To use the Budyko framework on a shorter timescale, we adopt the approach of Du et al. (2016), who introduced an "extended" Budyko framework in which precipitation values are adjusted to include plant-accessible soil storage change, essentially expanding the available water supply (P$-\Delta$S; Figure IV.2a). Annual soil storage is estimated using a second conceptual mass-conservation approach, the *abcd* model. The *abcd* model, named for its four parameters, is an explicit water balance model developed by Thomas (1981) that provides estimates of direct and indirect runoff, soil water and groundwater storage, and actual ET, calibrated to streamflow at the basin outlet. This allows for isolation of the change in plant-accessible soil water storage from deep subsurface storage changes. The *abcd* model assumes that "ET opportunity", the sum of actual ET over the timestep and soil water storage at the end of the timestep, is a function of available water. This relationship is parameterized by a ($0\sim1$), representing the tendency for runoff to occur before soil is saturated, and b, the maximum ET opportunity. The rate that ET occurs from soil storage is assumed to be proportional to the ET opportunity, and thus soil storage is also a function of b. The remaining two parameters in the model, c and d, control the partitioning of direct runoff from groundwater recharge and discharge. However, since we are interested only in change in soil storage, the last two parameters and related calculations were not used in this study. For full details of the extended Budyko model, see Du et al. (2016); for more information on the *abcd* model, see Wang and Tang (2014). Results of the *abcd* calibration are presented in Appendix C.2. Note that using one year of model spin-up for the *abcd* model and calculating change in storage eliminates two years from the period of record.

Various mathematical models exist to represent data plotted in a Budyko framework; one of the most versatile is the Fu equation, in which the ET fraction of available water (evaporative index) is a function of the aridity index (PET/P) and the parameter ω, a constant of integration (Fu, 1981; Zhang et al., 2004). The Fu model, modified for the

extended Budyko framework following Du et al. (2016), is given in Equation IV.1.

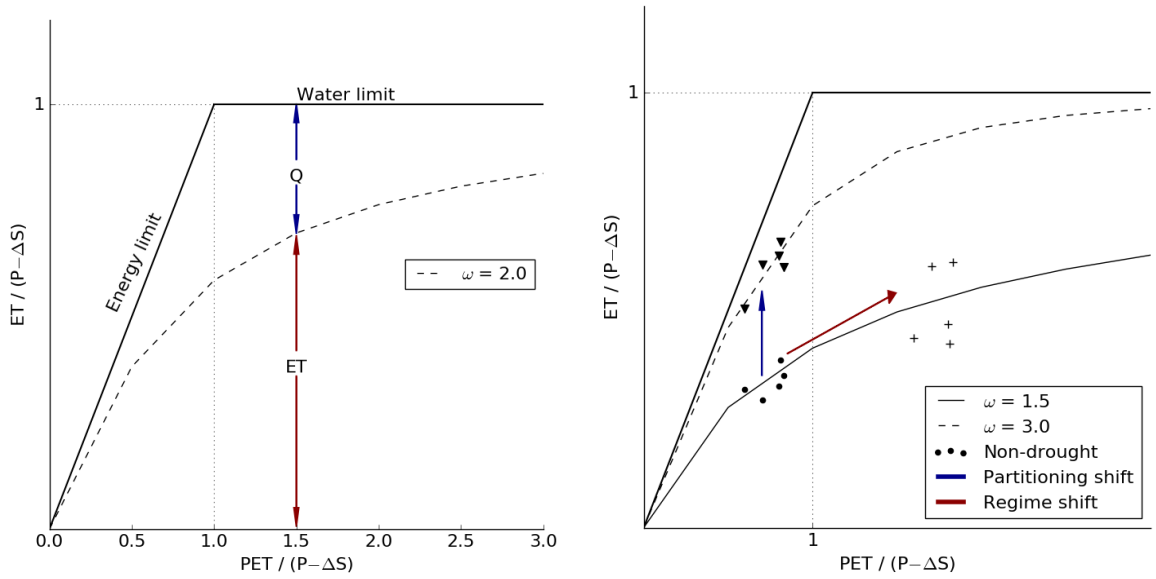
$$\frac{ET}{P-\Delta S} = 1 + \frac{PET}{P-\Delta S} - [1 + (\frac{PET}{P-\Delta S})^{\omega}]^{1/\omega} \tag{IV.1}$$

The value of ω will determine how close or far from the theoretical limit lines the data fall; the higher ω, the closer the curve comes to the energy and water limit lines. Thus, for a given $\frac{PET}{P-\Delta S}$ value, ω reflects the partitioning of available water between ET and runoff (Figure IV.2a). The physical meaning of ω has been connected to various basin characteristics, including vegetation coverage type and density, average slope, and relative soil infiltration capacity (Zhang et al., 2001, 2016; Yang et al., 2007; Jaramillo et al., 2018) as well as climate characteristics such as the seasonal offset between peak precipitation and potential evapotranspiration (Ning et al., 2019). In the context of droughts, changes to the water balance in the Budyko framework can occur in one of two ways: 1) data can shift along the same curve, changing water-balance components due to changes in the water or energy limitations and 2) data can shift to a new curve with a different omega value (Figure IV.2b). We refer to the former as a regime shift, since the basin becomes more or less energy or water limited, and to the latter as a partitioning shift.

For each basin in our study area, we calibrated the Fu equation twice, once for drought years and another for non-drought years, allowing us to assess the changes due to one factor or the other and the implications for ET and runoff. The difference between the two ω values indicates the direction and intensity of the partitioning shift. In order to understand the effect of the two shift types on ET and runoff, we first calculated the hypothetical drought evaporative indices that would have been seen if only a regime shift had occured (no change in ω; see "+" data points in Figure IV.2b). This was by applying Equation IV.1 to the annual observed drought values of $\frac{PET}{P-\Delta S}$ and the non-drought ω. We were then able to compare the hypothetical values to the non-drought values (black circles in Figure IV.2b). These two sets of data points were converted to absolute values of ET and runoff based on annual precipitation and change in storage values; the difference between their averages was the impact due to a regime shift. To calculate the impact due to partitioning shifts, we subtracted the regime shift impacts from the total observed impacts.

IV.2.4 Identifying mechanisms of water balance shifts

The shifts in the partitioning of available water can be related to feedback mechanisms between climatic conditions and catchment characteristics that either exacerbate or mitigate drought (Bales et al., 2018a; Teuling et al., 2013; Avanzi et al., 2020b). (In this study, "exacerbation" and "mitigation" are used with respect to runoff.) We examined four basin characteristics and responses to drought that may relate to observed shifts in the precipitation-runoff relationship. These are mechanisms that have previously been associated with drought-induced shifts in the water balance: 1) amount of available water storage (Avanzi et al., 2020b; Rungee et al., 2019; Oroza et al., 2018; Hahm et al., 2019a); 2) timing of water availability, which is related to precipitation phase (Avanzi et al., 2020b; Rungee et al., 2019; Berghuijs et al., 2014); 3) catchment aridity, which has been correlated with



(a) (b)

Figure IV.2: Conceptual plots of the Budyko framework used in this study. (a) Extended Budyko framework (see Du et al., 2016), illustrating how the calibrated Fu equation dictates partitioning of available water. (b) Illustration of drought-induced water balance shifts. ET is evapotranspiration, P is precipitation, ΔS is change in plant-accessible soil storage, and Q is runoff.

sensitivity to interannual changes in precipitation and departures from the historic mean precipitation (Berghuijs et al., 2014; Saft et al., 2016; Tian et al., 2020b); and 4) high-elevation runoff, related to basin spatial heterogeneity that can serve to mitigate drought (Bales et al., 2018a). Since not all of these mechanisms are directly measured across the Sierra, we use proxies to estimate their effects. Available soil water storage is estimated using average dry-season flow (July–September). Due to the highly seasonal precipitation in the Sierra Nevada, flow during this period almost exclusively reflects outflow from storage rather than surface runoff. Changes to timing of water availability during drought was estimated by looking at changes to precipitation phase (rain versus snow; Avanzi et al., 2020b; Rungee et al., 2019). Phase was estimated following Berghuijs et al. (2014) using a single-threshold temperature index method on a per-pixel basis. For each day, precipitation in pixels with an average temperature of 1°C or above was assumed to be rain; otherwise it was assumed to be snow. Catchment aridity (PET/P) was calculated directly, not including soil storage in order to isolate the effects of climate, and averaged over the study period. Finally, high-elevation runoff was estimated as the average annual precipitation minus ET for elevations above 2000 m. This was compared to the area-normalized annual flow at the basin outlet to estimate the proportion of annual runoff from high elevations.

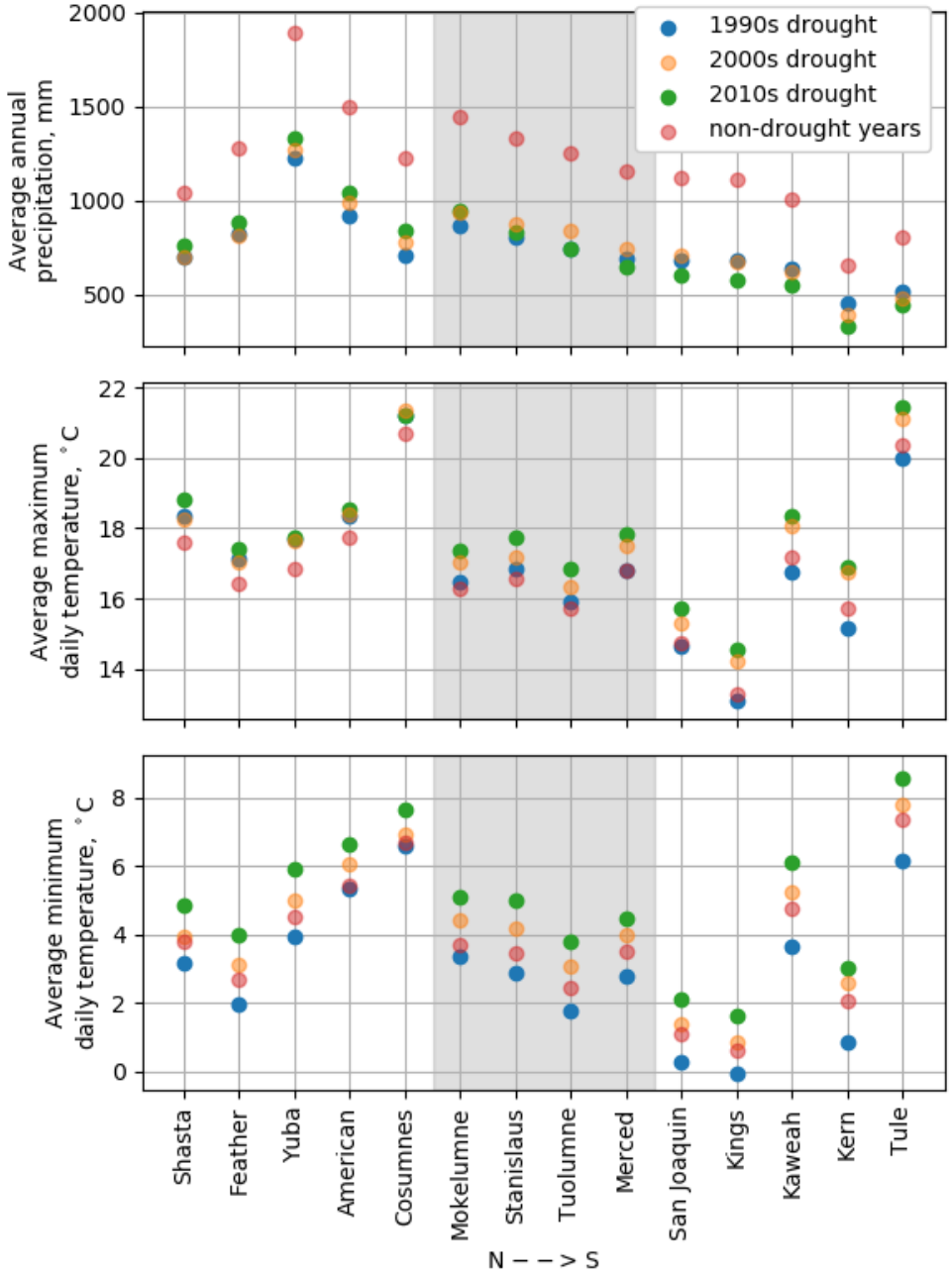


Figure IV.3: Climatic conditions during drought periods. The Central Sierra is shaded in gray, with the Northern and Southern basins to the left and right, respectively. The most recent drought (2010s) was the wettest drought in the Northern Sierra, but the driest in the Southern Sierra (top panel). Maximum temperatures only increase in the Central and Southern Sierra (middle panel), but minimum temperatures increase across the whole study area (bottom panel).

IV.3 Results

IV.3.1 Drought characterization

The period of record of the available data covers three drought periods, as defined by the State of California (see `https://water.ca.gov/Water-Basics/Drought`; accessed 29 July 2020): 1987–1992, 2007–2009, and 2012–2016. These droughts are referred to hereafter by the decade in which they ended (1990s, 2000s, and 2010s drought respectively). Average conditions varied across basins and droughts (Figure IV.3). Average maximum daily temperature shows no significant change between droughts in the northern basins (average increase of 0.21°C), but droughts in the Central and Southern Sierra basins are progressively warmer (average increases of 0.94 and 1.46 °C, respectively). In contrast, average minimum daily temperature shows increases across all droughts and basins (increases of 1.62, 1.88, and 2.11°C for the northern, central, and southern basins respectively). Average precipitation during the droughts decreases from north to south across the Sierra Nevada, reflecting similar variability in long-term average conditions (average annual precipitation across the period of record was 1245, 1122, and 799 mm in the northern, central, and southern Sierra, respectively). In the northern Sierra, the earlier two droughts were the driest (1990s and 2000s), but but the 2010s drought was driest in the Southern Sierra. Thus, droughts in the Northern Sierra were progressively wetter with higher minimum temperatures, but droughts in the Southern Sierra are progressively drier and hotter (Figure IV.3).

IV.3.2 Water balance during droughts

The extended Budyko model simulated runoff well in the study basins and was thus suitable for adjusting available water for the annual timestep (Appendix C.2). While a handful of years, amounting to 2.7% of all basin years, still lie above the water limit line, the model allowed for stable calibration in all basins of the Fu equation parameter ω. Both drought and non-drought ω values are on the order of values reported in the literature ($1 \sim 10$; Zhang et al., 2004; Du et al., 2016; Li et al., 2013) for all basins except the Yuba, where extreme energy limitation resulted in very high ω values (Figure IV.4). As the wettest basin in the Sierra Nevada (average annual precipitation of more than 1720 mm), these conditions are consistent with basin climate. However, both ω values in the Yuba are far outside the normal range, to the point where they are effectively infinity (note that the two lines are indistinguishable in Figure IV.4). As a result, we do not consider the direction or magnitude of the shift to carry significance and exclude the basin from further analysis. The changes along both Budyko axes (extended aridity index on the x-axis and extended evaporative index on the y-axis) between droughts and non-drought periods were significant in all basins to the $\alpha = 0.01$ level based on a Kolmogorov-Smirnoff test, with the exception of change in the evaporative index on the Feather, which was significant to the $\alpha = 0.05$ level (Table C.4).

In general, northern basins saw a shift in favor of runoff (decrease in ω), while the southern basins saw a shift in favor of ET (increase in ω), with the exception of the Cosumnes in the north and the San Joaquin and the Kings in the south (Figure IV.4). Note that a shift in favor of ET or runoff does not guarantee that the quantity will increase in absolute terms.

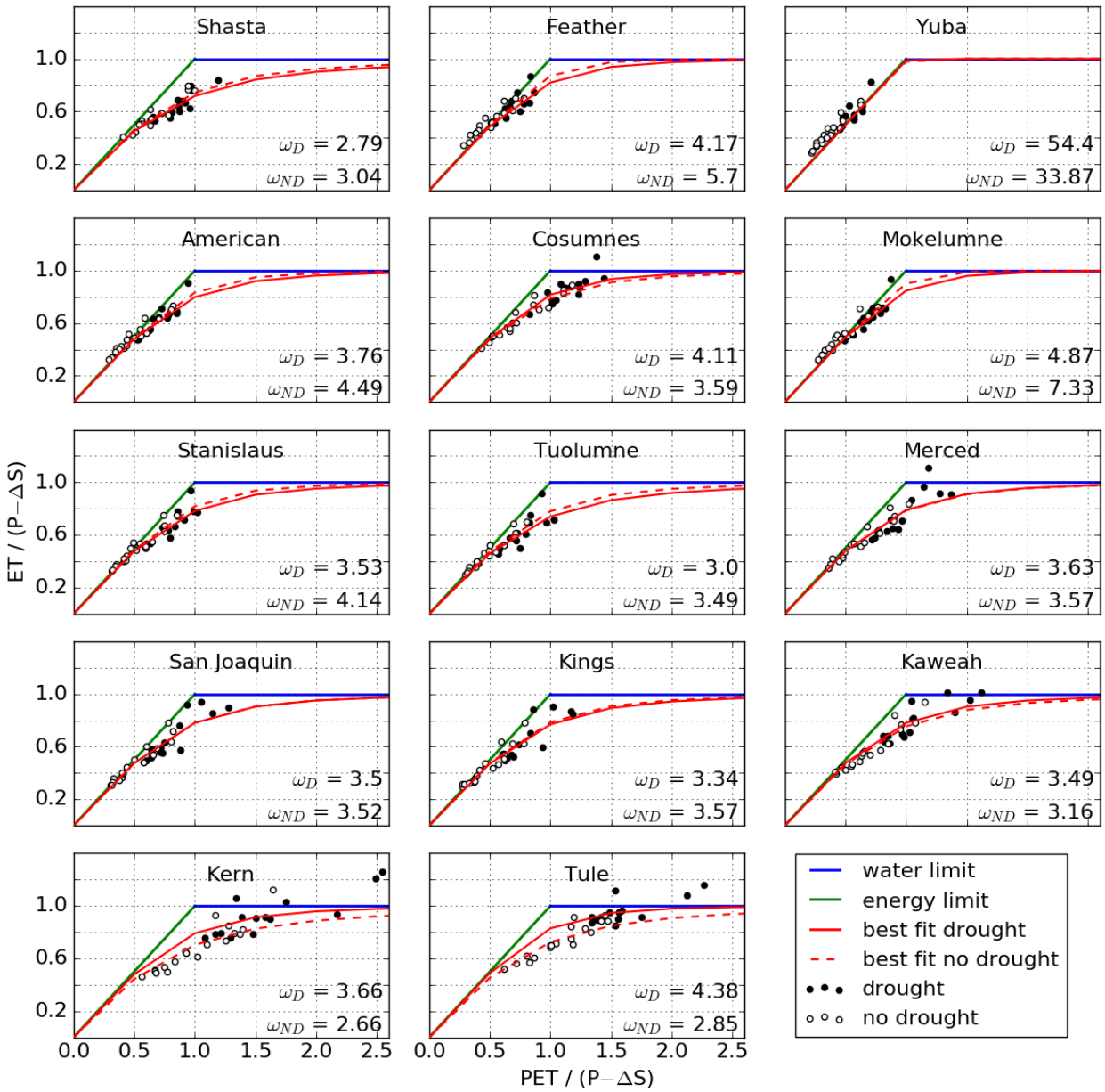


Figure IV.4: Annual (water year) water balances plotted in the extended Budyko framework, with calibrated best-fit lines and corresponding parameter values for drought (ω_D) and non-drought periods (ω_{ND}).

Likewise, movement to the right along the same Budyko curve will result in an increase in ET *as a fraction of available water*, but not necessarily in an increase in absolute ET, due to the drop in precipitation during droughts. The Tule and Kern basins in the south see a particularly strong shift in favor of ET (towards a higher ω value) while the Feather and Mokelumne further north see the opposite shift (partitioning changes in favor of runoff). In other words, drought may imply increases or decreases in the absolute quantities of ET and runoff.

The absolute changes in ET and runoff due to the partitioning shifts varied both in sign and magnitude, while regime changes were more consistent (Table IV.1 and Figure IV.4). With respect to runoff, the magnitudes of regime-related changes dominate those of partitioning-related changes, with the former always at least ten times higher than the latter (Table IV.1 and Figure IV.5). This results in an overall drop in runoff across the study area, since runoff regime changes are always negative (Table IV.1). However, partitioning shifts still account for significant change in the Southern Sierra, where regime-related changes are lower. In the case of ET, changes due to regime shifts still tend to be higher magnitude than partitioning shifts, but not exclusively. As a result, one type of shift can offset the other in basins where they have opposite signs. For example, ET is almost always reduced during droughts from regime shifts alone, but the Feather and Mokelumne would have seen an increase in overall ET if it were not for the curve shift downwards in favor of runoff (regime shift values are positive, but ω decreases).

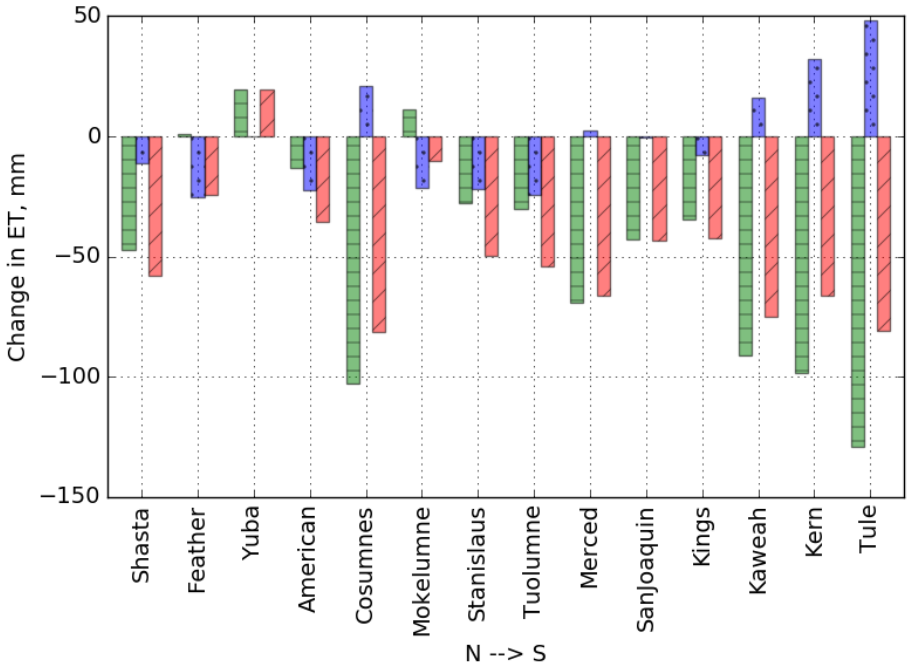
IV.3.3 Drought feedback mechanisms

Catchment aridity

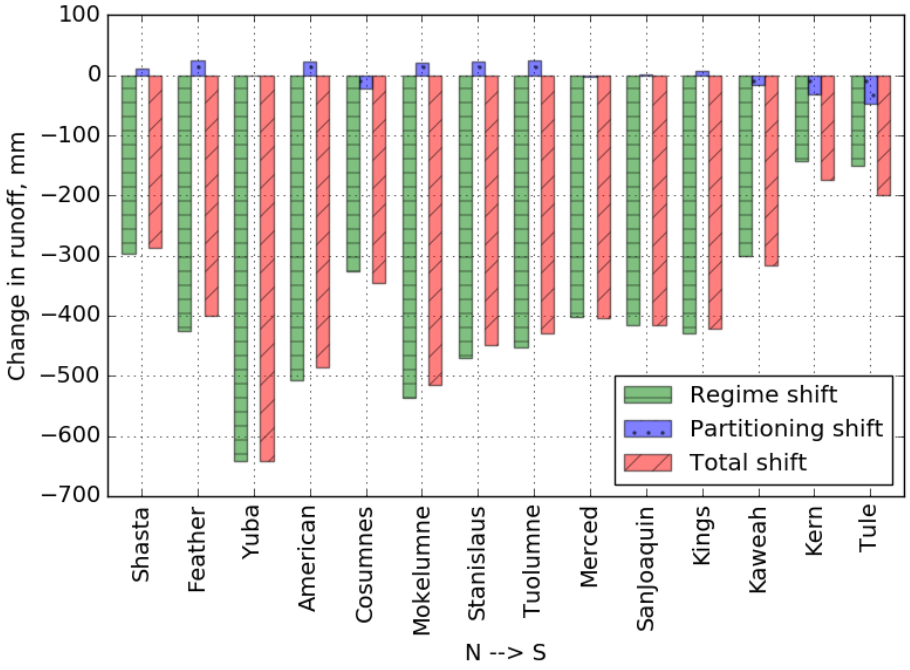
Catchment aridity was higher in basins that saw a shift in favor of ET ($PET/P \geq 0.766$) and vice versa ($PET/P \leq 0.749$). Shift magnitude was highly correlated with average aridity ($r = 0.83$, $p < 0.001$). The threshold dividing the two categories is notable, as $PET/P = 0.76$ has previously been identified as the cutoff between energy-limited water balance regimes and drier regimes, equitant and water-limited (McVicar et al., 2012). Thus, basins where more water than energy is available for evapotranspiration see a shift towards runoff, while those where water and energy availability are more or less equal or where energy is more plentiful see a shift towards ET.

Dry-season baseflow

Baseflow was generally higher in basins that saw a partitioning shift in favor of runoff, with an average baseflow of 14.5 mm in those basins versus 6.9 mm in those that shifted in favor of ET (Figure IV.6). The only basin that showed a significant departure from other basins displaying similar partitioning behavior was the American, which had relatively low baseflow. Notably, the basins where shifts were the opposite of what would be expected geographically (the Cosumnes in the north shifting towards ET versus the San Joaquin and Kings further south shifting toward runoff) showed the most extreme baseflow values. The Cosumnes had the lowest flows at 2.12 mm and the San Joaquin and Kings had the highest, at 20.9 and 21.9 mm, respectively.



(a) ET



(b) Runoff

Figure IV.5: Changes in absolute ET and Q values during droughts that can be attributed to regime versus partitioning shifts.

Table IV.1. Change in evapotranspiration and runoff during drought attributable to regime and partitioning shifts

Basin	**Evapotranspiration change, mm**			**Runoff change, mm**		
(N→S)	Total[a]	Regime	Partitioning[b]	Total[a]	Regime	Partitioning[b]
Shasta	-58.1	-47.1	-11.1	-287	-298	11.1
Feather	-24.2	1.0	-25.2	-400	-425	25.2
Yuba[c]	19.5	19.5	0.0	-642	-642	0.0
American	-35.5	-13.1	-22.5	-485	-507	22.5
Cosumnes	-81.4	-102.6	21.2	-346	-325	-21.2
Mokelumne	-10.4	11.1	-21.5	-515	-537	21.5
Stanislaus	-49.7	-27.7	-22.0	-448	-470	22.0
Tuolumne	-54.3	-30.0	-24.2	-429	-453	24.2
Merced	-66.5	-69.0	2.5	-405	-402	-2.5
San Joaquin	-43.3	-42.8	-0.4	-416	-416	0.4
Kings	-42.3	-34.8	-7.5	-421	-429	7.5
Kaweah	-74.9	-91.0	16.1	-317	-301	-16.1
Kern	-66.2	-98.4	32.2	-175	-142	-32.2
Tule	-80.7	-129.0	48.3	-199	-151	-48.3

[a]Totals for each variable are the sum of Regime and Partitioning values.
[b]Partitioning values for evapotranspiration and runoff are the negative of each other.
[c]Since the ω values in Yuba are both effectively infinite, the partitioning shift has no effect.

Precipitation phase

Percent of precipitation falling as snow decreased during drought in all basins except Tule. Northern Sierra Nevada basins saw greater percent decreases than the central and southern Sierra (-2.3%, -1.95%, and -0.52%, respectively), despite the latter having seen greater temperature increases. The northern basins are overall lower elevation, so more area lies in the rain-snow transition where precipitation phase is susceptible to increases in temperature. For the most part, basins that saw a shift towards runoff (decrease in ω) saw a stronger decrease in percent snow. The Pearson correlation coefficient of $r = 0.62$ ($p < 0.05$) between change in ω and percent change in snow shows a moderate relationship between the two.

Precipitation excess above 2000 m

Using P-ET above 2000 m as an index for high-elevation runoff and expanding the analysis to the rest of the study site, we find that most basins in the Sierra rely substantially on high-elevation runoff. Nine of the thirteen basins analyzed (excluding the Yuba) saw an average high-elevation runoff fraction above 0.33 (those that did not were the Shasta, Feather, American, and Cosumnes). However, overall fraction of runoff from high elevations was not significantly correlated with changes in ω. Instead, we found that changes in high-elevation runoff between drought and non-drought periods was moderately negatively correlated with partitioning shift ($r = -0.55$, $p < 0.05$). In other words, strong decreases in high-elevation runoff during drought were associated with strong shifts in favor of ET and vice versa. Specifically, basins that see a significant decrease ($> 9\%$) in high-elevation runoff during drought see strong shift towards ET (Tule, Kaweah, Kern). All other basins, including the Cosumnes and Merced, which shifted in favor of ET, saw a positive or small negative percent changes in high-elevation runoff (Figure IV.6).

IV.4 Discussion

The Budyko approach used here has allowed us to distinguish two types of drought-induced shifts, regime and partitioning, for the first time. To fully explore how this new framework can be leveraged to better understand drought implications for the water balance, we present the discussion in three sections. We begin with an explanation of how regime and partitioning shifts primarily relate to climate and basin feedbacks, respectively (Section IV.4.1). Next, we discuss the relative impact of these shifts on absolute values of ET and runoff in the Sierra Nevada during drought (Section IV.4.2). Finally, in Section IV.4.3, we offer an interpretation of how partitioning shifts may relate to hydrologic processes by analyzing correlations between shifts and the four basin drought responses enumerated in Section IV.3.3.

IV.4.1 Interpreting regime and partitioning shifts

Due to the nonlinear relationship between the aridity and evaporative indices in the Budyko framework (Figure IV.2), both regime and partitioning shifts result in changes in the precipitation-runoff relationship as observed in other studies (e.g., Avanzi et al., 2020b; Tian et al., 2020b; Saft et al., 2016; Petheram et al., 2011). The primary difference, however, is that regime shifts – movement along the same Budyko curve (Figure IV.2b) – are reflective of predictable climatic variability during drought, while partitioning shifts represent a change to a new equilibrium state that cannot be easily forecast *a priori*. Regime shifts are almost exclusively controlled by measurable climatic factors through PET (a function of temperature) and precipitation. Endogenous basin characteristics (i.e., factors influencing available subsurface water storage) are a secondary influence, since even during drought withdrawals from the subsurface were ten times less than precipitation rates. Thus, readily available observations of climate patterns are mostly sufficient to predict regime shifts and their impact on water resources during drought. Partitioning shifts, on the other hand, are a function of nonlinear and indirect catchment feedbacks to climatic changes during drought. While there is understanding that these mechanisms relate at least in part to vegetation and subsurface water storage interactions (Avanzi et al., 2020b), a relative dearth of data related to both has so far prevented a full enumeration of these mechanisms and how they interact. This makes the impact of partitioning shifts on drought water supply largely unpredictable and highlights the need for future research focused on process understanding of these shifts.

The ability to distinguish these types of shifts while allowing for each to induce nonlinear changes in the water balance is an advantage of the Budyko framework. Previous studies have used linear models to relate precipitation, Box-Cox transformed runoff, and a dummy variable to account for drought (Saft et al., 2016; Avanzi et al., 2020b). This statistical framing is primarily concerned with the direct impact of precipitation on runoff. The Budyko framework, however, considers allocation of water relative to the aridity index, a combination of two major water-balance drivers (PET and precipitation), rather than precipitation alone. Moreover, the Budyko framework governs available water partitioning by physical behavior under limit conditions (when the aridity index is zero, all water goes to runoff; when the aridity index is one, all water goes to ET). This framework allows for the possibility that even expected and predictable water balance changes during drought may be nonlinear and that some shifts observed in other studies may be the result of factors that are not captured

in a two-dimensional precipitation-runoff plane. This critical difference may explain why previous studies have observed less runoff than expected without a shift in relationship (Avanzi et al., 2020b; Tian et al., 2020b; Saft et al., 2016), while most study basins under the Budyko framework show a shift towards *more* runoff as a fraction of available water than would be expected using non-drought relationships (decrease in ω; Figure IV.4). The direct precipitation-runoff relationship and the Budyko framework are complementary approaches, but the understanding that water balance shifts during droughts are due to many interacting factors (see Avanzi et al. (2020b) and Saft et al. (2016)) argues for expanding the tools used to analyze this phenomenon. These and new approaches should be the subject of further study.

IV.4.2 Impact of regime and partitioning shifts

Nonlinearities in the relationship between the aridity index, ω, and the evaporative index also mean that regime and partitioning shifts are not equally responsible for changes in ET and runoff during drought (Figure IV.5 and Table IV.1). Regime shifts accounted for at least 75% of runoff reductions across the study area and also dominated changes in absolute ET in most basins. This suggests that most reductions in runoff during drought may be predictable from precipitation and other climate factors; however, the relatively small impacts due to partitioning shifts still represent significant volumes of water. For example, partitioning shifts in the Feather River provide 25.2 mm of additional runoff annually during droughts (4.6% of average annual runoff). Over the approximately 9400 km^2 basin, this amounts to more than 225×10^6 m^3 of water. In the Kern, with an area of approximately 5300 km^2, a loss of 32.2 mm per year (22% of average annual runoff) due to partitioning shifts translates into nearly 290×10^6 m^3.

It is important to note that movement in the Budyko space due to regime shifts does not necessarily indicate whether absolute values of ET and runoff will increase or decrease. Since the aridity index ($PET/P - \Delta S$) typically increases during droughts, regime shifts result exclusively in an increase in ET *as a fraction of precipitation*. This results in a decrease to absolute runoff across all basins in the study area, but usually does not translate into an increase in absolute ET (Table IV.1) due to the available water decreasing significantly during drought. Only in the Feather and Mokelumne basins did ET increase in absolute terms (1 and 11.1 mm respectively), indicating that available water was sufficient to support vegetation. Other than the Yuba, the Feather and Mokelumne basins are the wettest in the Sierra Nevada (average annual precipitation of 1180 and 1290 mm, respectively), while the water availability in the Feather may also be partly supported by the greater groundwater storage in parts of the basin (Avanzi et al., 2020b). An increase in ET during droughts has also been observed or predicted in the overall wetter and colder European Alps (Teuling et al., 2013; Mastrotheodoros et al., 2020).

The direction of a partitioning shift, on the other hand, is a direct indicator of the sign of the change in absolute ET or runoff. This is because the partitioning shift relates to change in evaporative index for a given aridity index; in other words, assuming a constant amount of available water. Furthermore, because the derivative of the evaporative index with respect to ω is nonlinear (see Equation IV.1), the same unit change starting on the higher end of the ω spectrum will have less impact on the evaporative index than changes

Table IV.2. Summary of basin response mechanisms influencing water partitioning relative to stated threshold

Basin[a] (N→S) Threshold:	**Average aridity**[b] 0.76	**Average baseflow** 10 mm	**Decrease in snow** -2.5%	**High-elevation runoff fraction change** -0.1
Shasta	−	+	+	+
Feather	−	+	−	+
American	−	−	−	+
Cosumnes	+	−	+	+
Mokelumne	−	+	−	+
Stanislaus	−	+	−	+
Tuolumne	−	+	+	+
Merced	+	+	+	+
San Joaquin	−	+	+	+
Kings	−	+	+	+
Kaweah	+	+	+	−
Kern	+	−	−	−
Tule	+	−	+	−

[a]Blue indicates negative shift in ω; red indicates positive shift. Yuba is not reported.
[b]Mechanisms colors reflect expected effect:
Blue = shift towards runoff; red = shift towards ET; black = no expected effect.

on the lower end (Figure IV.5b). For example, the Feather and Tule see the same magnitude shift in ω ($|\omega| = 1.52$) but in different directions and starting from different non-drought values ($\omega_{ND} = 5.7$ and 2.85, respectively). In the wetter Feather, the increase in runoff due to partitioning is 25.2 mm, but in the more southern Tule, the decrease in runoff is nearly twice as large at 48.3 mm (Table IV.1). This is further demonstrated in the Kern and Tule, which had the lowest non-drought omega values (2.66 and 2.85, respectively) and where runoff was most impacted. This shows that even basins within the same mountain range or region may have high variability in their vulnerability to drought. It further suggests that water agencies that rely on multiple headwater basins (not uncommon in areas like California with highly interconnected water systems) should consider their management strategies on a per-catchment basis.

IV.4.3 Mechanisms of partitioning shifts during drought

In this section, we discuss how the relationships observed between the change in ω and four basin response mechanisms (see Section IV.3.3) may inform our understanding of processes that drive partitioning shifts. Such processes are related to endogenous basin characteristics that dictate the response of the catchment's water balance to drought climate conditions (Troch et al., 2015). Under the Budyko framework, these responses primarily impact the water balance through partitioning shifts, and their effects are captured by the ω parameter, which controls the distribution of available water between ET and runoff. Previous literature has related ω or similar Budyko parameters to basin characteristics including vegetation type (Zhang et al., 2001; Ning et al., 2017, 2019, 2020; Roderick and Farquhar, 2011), topographic features like average slope (Yang et al., 2007, 2009; Ning et al., 2019), and soil characteristics like infiltration capacity and soil water storage (Ning et al., 2019; Yang et al., 2007). Since

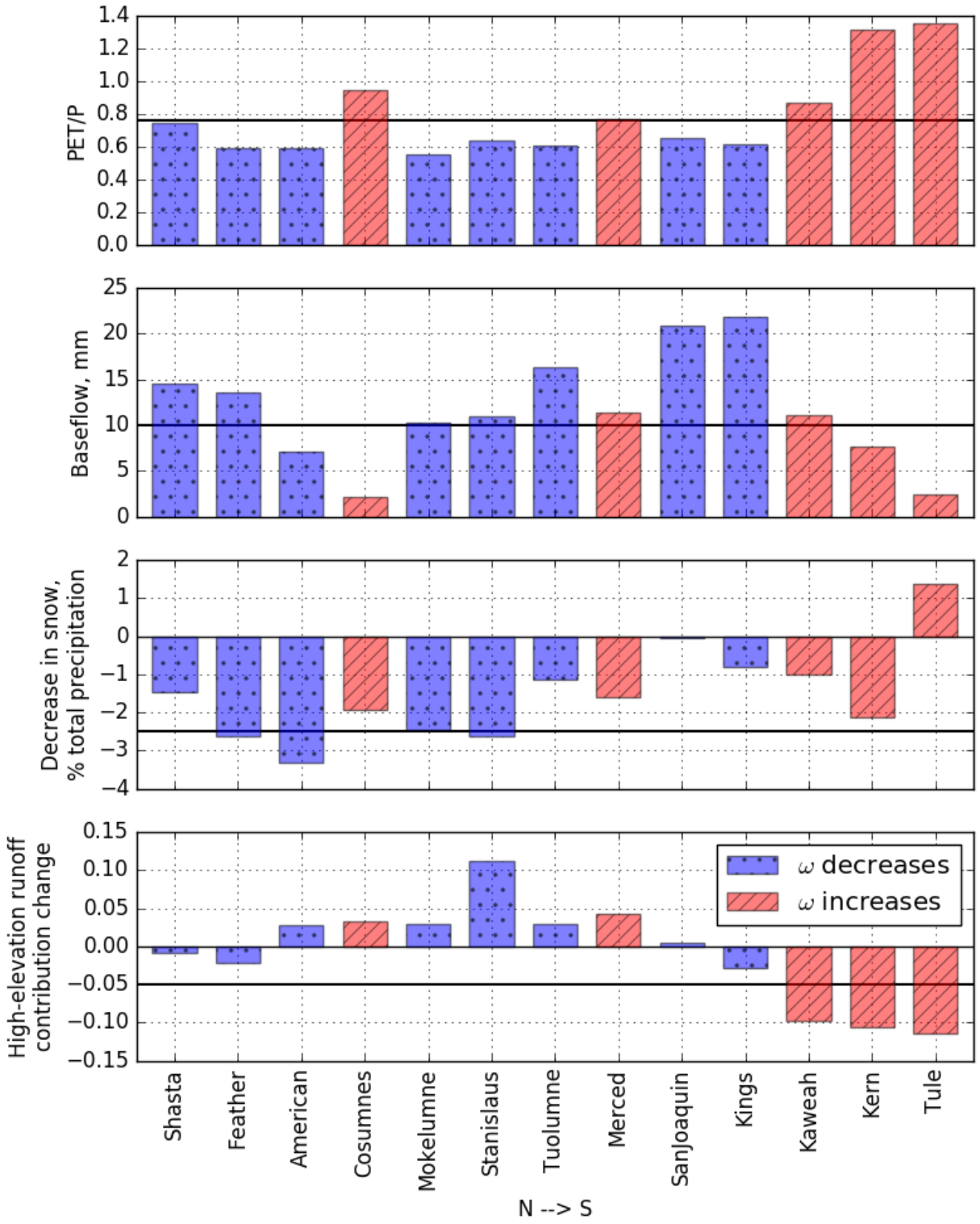


Figure IV.6: Basin responses driving partitioning shifts. Increases in ω reflect a shift in favor of ET; decreases reflect a shift in favor of runoff. Thresholds for each mechanism discussed in Section IV.3.3 are marked with bolded black lines.

the four mechanisms we examine are all fundamentally related to these basin features, our work is broadly consistent with previous literature on the interpretation of this parameter.

Regarding the first metric, aridity, there is a clear pattern in partitioning shifts where the wet catchments see a shift in favor of runoff while the arid basins shift in favor of ET. This reflects both the greater average aridity of the southern basins as well as the more severe drought conditions (higher temperatures and lower precipitation; Goulden and Bales, 2019). As a lower-elevation basin, the Cosumnes is also more arid and sees a shift toward ET. The high correlation between average PET and shift in ω suggests that overall climate may predispose basins to a certain drought response through long-term co-evolution of landscapes and climate (Troch et al., 2015). This agrees with previous findings that catchment aridity is a strong predictor of shifts in the runoff coefficient (Saft et al., 2016; Tian et al., 2020b). Aridity is both a key indicator of catchment climate (Budyko, 1974) as well as being correlated with vegetation and water storage (Saft et al., 2016), both of which also influence the intensity of the feedback cycle between precipitation deficit and vegetation response. Our findings again suggest that dry basins are likely to become drier and that this impact is likely to have a disproportionate impact on runoff compared to wetter basins (Section IV.4.2).

The second metric, amount of dry-season baseflow, provides an estimate of the baseline amount of subsurface storage in a catchment, thus serving as a proxy for a basin's potential for buffering the precipitation deficit with soil storage. Higher baseflows were associated with shifts in favor of runoff, reflecting one or more basin mechanisms supporting streamflow during drought. They may relate to deep groundwater contributions to streams, which are less vulnerable to plant water use, particularly on shorter timeframes, and can thus sustain flows during periods when vegetation is more heavily reliant on near-surface storage. If baseflows are indicative of higher groundwater tables, these soils may become saturated more quickly during a rainfall event, thus leading to saturation-excess runoff (Petheram et al., 2011). The more and higher groundwater tables would make a basin less susceptible to losing this mechanism over large areas during drought (Saft et al., 2016). Finally, areas with higher average baseflow levels are less likely to see storage severely depleted by vegetation over the course of a multi-year drought and are able to continue sustaining streamflow (Rungee et al., 2019). The fact that geographically anomalous basins (Cosumnes, San Joaquin, and Kings) showed the most extreme baseflows suggests that subsurface storage can be a significant factor in basin response, both mitigating and exacerbating drought conditions.

Third, higher temperatures during droughts may induce a shift in precipitation phase from rain to snow, changing the timing of water availability to earlier in the season. Other analyses of the Sierra Nevada water balance during droughts (e.g. Rungee et al., 2019) suggest that snowpack augments plant-accessible subsurface storage by 1) increasing infiltration efficiency, as snowmelt is slow as compared to intense rainfall events, and 2) shortening the length of the dry season by delaying infiltration. As was suggested by Avanzi et al. (2020b) and Shao et al. (2012), this implies that shifts from snow to rain may favor runoff rather than ET, at least on the seasonal timescale, since more water is able to runoff or infiltrate to deep groundwater in periods of low vegetation productivity. Our findings are consistent with this hypothesis: changes in percentage of precipitation that fell as snow tended to be stronger ($> 2.5\%$) mostly overlap with basins that shift in favor of runoff during droughts (the only exception being the low-elevation Cosumnes basin). Basins where there was little

change to snow percentage did not necessarily see a shift in favor of ET, but loss of SWE may be a predictor of greater runoff.

Finally, the generation of high-elevation runoff, which is more resilient to increases in PET due to overall lower temperatures and sparser vegetation, can help mitigate runoff losses elsewhere in the basin (Goulden and Bales, 2019). Given the orographic effect of the Sierra Nevada, high elevations may also be less susceptible to decreases in precipitation. Our findings on the importance of high-elevation runoff broadly agree with Goulden and Bales (2019), who identified high-elevation runoff as a drought mitigation factor in the Kings River during the 2012–2016 drought. Here, we find that resilient high-elevation runoff is not guaranteed to mitigate drought so much as decreases in high-elevation runoff act to exacerbate drought. Both the Merced and Cosumnes basins saw slight increases in high-elevation-runoff fraction during drought, but saw a shift in favor of ET. However, all basins that saw a strong decrease in contribution of high-elevation runoff (> 0.5) also saw a shift in favor of ET (Kaweah, Kern, Tule). Thus, high-elevation runoff may not always offset other factors like high aridity and low baseflow, but loss of this important runoff source may shift water allocation towards ET. Alternatively, loss of high-elevation runoff may be correlated with other changes that cause a shift towards ET, such as temperature increases driving increases in ET demand at high elevations or lateral redistribution of precipitation excess from higher elevations to unsaturated soil at lower elevations.

IV.5 Conclusion

Applying the Budyko framework to assess the impact of droughts on the water balance, we identify two distinct types of shifts in water allocation, regime and partitioning shifts. We show how regime shifts are primarily due to predictable climatic variability during droughts, but partitioning shifts imply a change in the Budyko parameter ω and are related to less-predictable nonlinear basin feedback mechanisms. We show that regime shifts dominate changes in absolute runoff during droughts, but that gains or losses due to partitioning shifts are still significant. Changes in absolute ET are influenced by both types of shifts.

Finally, we examine the correlation between partitioning shifts and known basin response mechanisms. We find that a low aridity index, high baseflow, shift from snow to rain, and the resilience of high-elevation runoff correlate to increased runoff as a fraction of precipitation during droughts. These findings help characterize how different basins will respond to drought conditions, with implications for natural and human systems in drought-prone regions.

Chapter V

Conclusions and Future Work

Each of the analyses presented here demonstrates how data can be leveraged with a specific class of hydrologic model. Ground-based wireless-sensor networks have the potential to greatly expand the number of data points available, enabling use of data-driven machine learning algorithms, as presented in Chapter II. Spatially distributed data products derived from remote sensing or from remote sensing blended with ground-based sensors can also be leveraged to improve models. By providing full characterization across the landscape of basin features that influence runoff production, such as elevation, slope, and aspect derived from digital elevation models, spatial maps can provide the basis for optimal spatial distribution of input data for a physically based model (Chapter III). Finally, these novel blended products, particularly spatial evapotranspiration maps, expand the potential for leveraging conceptual models and connecting results with physical principles (Chapter IV).

Several lessons for leveraging novel data sources for hydrologic modeling can be drawn from this work and may help guide future research. Firstly, data that are gathered strategically and across multiple water-balance components are most helpful in improving or developing new models. For example, ground-based wireless sensors used in Chapter II were placed intentionally across vegetation densities, enabling the Random Forest rain-on-snow analysis. Having new spatial data products of evapotranspiration (ET) and snow water equivalent (SWE), two key but relatively under-measured water-balance components, was critical to constraining the physically based Precipitation Runoff Modeling System model in Chapter III. Secondly, remotely sensed data are a powerful tool for analyzing the water balance in a spatially distributed way, but are most effective when calibrated or coupled with ground-based sensors. For example, the PRISM-based precipitation inputs to the PRMS model in Chapter III were adjusted daily based on ground sensors; preliminary analysis for that work showed significant improvement using ground-based measurements across the basin and also enabled the maps to be used in real time, even if the product had not yet been developed for the date the model was being run. Both the ET and SWE maps used to constrain PRMS in Chapter III and to drive the Budyko model in Chapter IV were developed by coupling remotely sensed with ground-based data. Thirdly, scale continues to be a challenge both on the modeling and data sides, whether scaling up physically based models or scaling down conceptual models. While avenues for future research on this topic are extensive, one pathway explored in this work is combining model types in order to leverage the strengths of each. For example, Chapter III demonstrates how a machine-learning algorithm can be

leveraged to optimally identify spatial modeling locations for a physically based model and thus minimize the overparameterization issues that come with scaling such models across the landscape. Another potential path for future work includes more explicit comparison of the results from different classes of models, including questions such as: for what processes and at what scales do the results of physically based and conceptual models agree? For which do they not? What shortcomings in our understanding of hydrologic processes does this reveal, and what data are needed to improve each model type?

Finally, data collection is important, but it is equally important to make data usable and easily accessible. This is particularly critical to the process of incorporating new scientific understanding of hydrologic processes and/or modeling shortcomings into the decision-making process for water-resources managers. Clean, timely, and well-formatted data can encourage model use, expansion, and updates, for example by enabling new, easier model design (Chapter III), but challenges remain. Thus, an avenue of future, possibly interdisciplinary, research in hydrology should be a greater focus on the barriers to operationalizing scientific advances. This work should aim to answer questions such as: what steps are needed to incorporate advances in process understanding into widely used operational models? How can data be leveraged to design models that are easier to update and calibrate? How are data used operationally, and how can database managers make them accessible and usable by water resource managers? Given that many barriers to model implementation and updates are related to human behaviors and institutional structure, capacity, and culture, such work may benefit from collaboration between engineers and social scientists.

Novel data bring new opportunities to each class of hydrologic model examined here. Rapid expansion of remote and ground-based sensors has enabled the use of machine-learning algorithms in hydrology for the first time. The application of these methods to natural systems may not always be appropriate when thorough process understanding is required, but hold promise for identifying previously unknown and/or unanticipated correlations among processes. Future research should aim to identify under what circumstances machine-learning algorithms are appropriate for hydrologic applications and what data (type and quantity) are needed for such uses. For physically based models, new spatial data sources are poised to make multi-objective calibration standard practice, which will help relieve issues of overparameterization and improve representation of internal water balance states. Lastly, conceptual models benefit from data not only as inputs and validation, but to be able to connect the results to process understanding, as presented in Chapter IV. This can help overcome the limited ability of conceptual models to describe the specific mechanics of modeled phenomena. Possibly the most novel and powerful strategy in these analyses is the blending of model types, as in Chapter III where the Gaussian Mixture Model, a machine-learning algorithm, was used to identify the optimal modeling locations for the physically based Precipitation-Runoff Modeling System model. As discussed, combining model types has the potential to help address scale issues in hydrologic modeling, but other applications exist as well. For example, more process-based physical models could be coupled with either machine-learning or conceptual models to connect results to specific causal mechanisms. Future research should more fully explore the scales and processes over which these model classes overlap, complement, and contradict each other.

Novel data sources are a key resource for the field of hydrology to confront large-scale challenges such as climate change, population growth, and land-use changes. This disser-

tation has presented examples of how these data can provide new insights into hydrologic process understanding as well as help enhance hydrologic models. The results support a continued focus on how the two interdependent tools of models and data can be jointly leveraged, highlighting areas in which additional strategic investment is needed to manage the unprecedented strain on both human and natural water systems in the twenty-first century.

References

American Society of Civil Engineers. Artificial Neural Networks in Hydrology. I: Preliminary Concepts. *Journal of Hydrologic Engineering*, 2000. ISSN 1084-0699. doi: 10.1061/(asce)1084-0699(2000)5:2(115).

E. A. Anderson. National Weather Service river forecast system-snow accumulation and ablation model. Technical report, National Oceanic and Atmospheric Administration, Washington, D.C., 1973. URL `http://archive.org/details/nationalweathers00ande`.

E. A. Anderson. A point energy and mass balance model of a snow cover. Technical report, National Oceanic and Atmospheric Administration, Washington, D.C., 1976. URL `https://repository.library.noaa.gov/view/noaa/6392`.

V. Andréassian, N. Le Moine, C. Perrin, M.-H. Ramos, L. Oudin, T. Mathevet, J. Lerat, and L. Berthet. All that glitters is not gold: the case of calibrating hydrological models. *Hydrological Processes*, 26(14):2206–2210, 2012. doi: 10.1002/hyp.9264.

N. W. Arnell and S. N. Gosling. The impacts of climate change on river flood risk at the global scale. *Climatic Change*, 134(3):387–401, 2016. ISSN 15731480. doi: 10.1007/s10584-014-1084-5.

R. Arsenault and F. P. Brissette. Continuous streamflow prediction in ungauged basins: The effects of equifinality and parameter set selection on uncertainty in regionalization approaches. *Water Resources Research*, 50(7):6135–6153, 2014. doi: 10.1002/2013WR014898.

F. Avanzi, C. De Michele, A. Ghezzi, C. Jommi, and M. Pepe. A processing-modeling routine to use SNOTEL hourly data in snowpack dynamic models. *Advances in Water Resources*, 73:16–29, 2014. ISSN 03091708. doi: 10.1016/j.advwatres.2014.06.011.

F. Avanzi, S. Yamaguchi, H. Hirashima, and C. De Michele. Bulk volumetric liquid water content in a seasonal snowpack: Modeling its dynamics in different climatic conditions. *Advances in Water Resources*, 86:1–13, 2015. ISSN 03091708. doi: 10.1016/j.advwatres.2015.09.021.

F. Avanzi, G. Petrucci, M. Matzl, M. Schneebeli, and C. De Michele. Early formation of preferential flow in a homogeneous snowpack observed by micro-CT. *Water Resources Research*, 53(5):3713–3729, 2017. ISSN 19447973. doi: 10.1002/2016WR019502.

F. Avanzi, T. Maurer, S. Malek, S. D. Glaser, R. C. Bales, and M. H. Conklin. Feather River Hydrologic Observatory: Improving Snowpack Forecasting for Hydropower Generation using Intelligent Information Systems. In *California's Fourth Climate Change Assessment*. California Energy Commission. Publication number: CNRA-CEC-2018-001, 2018.

F. Avanzi, R. C. Johnson, C. A. Oroza, H. Hirashima, T. Maurer, and S. Yamaguchi. Insights Into Preferential Flow Snowpack Runoff Using Random Forest. *Water Resources Research*, 2019. ISSN 19447973. doi: 10.1029/2019WR024828.

F. Avanzi, T. Maurer, S. D. Glaser, R. C. Bales, and M. H. Conklin. Information content of spatially distributed ground-based measurements for hydrologic-parameter calibration in mixed rain-snow mountain headwaters. *Journal of Hydrology*, 582:124478, 2020a. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2019.124478.

F. Avanzi, J. Rungee, T. Maurer, R. Bales, Q. Ma, S. Glaser, and M. Conklin. Climate elasticity of evapotranspiration shifts the water balance of Mediterranean climates during multi-year droughts. *Hydrology and Earth System Sciences Earth Science Systems*, 24: 4317–4337, 2020b. doi: 10.5194/hess-24-4317-2020.

Y. Bai, T. Wagener, and P. Reed. A top-down framework for watershed model evaluation and selection under uncertainty. *Environmental Modelling & Software*, 24(8):901–916, 2009. ISSN 1364-8152. doi: 10.1016/j.envsoft.2008.12.012.

R. Bales, M. Matt, E. Stacy, M. Conklin, and X. Meng. Southern Sierra Critical Zone Observatory (SSCZO), Providence Creek meteorological data, soil moisture and temperature, snow depth and air, v3 temperature, 2013. URL `https://doi.org/10.6071/Z7WC73`.

R. Bales, E. Stacy, M. Meadows, P. Kirchner, M. Conklin, and X. Meng. Southern Sierra Critical Zone Observatory (SSCZO), Wolverton Creek meteorological data, soil moisture and temperature, v3, 2017.

R. C. Bales, N. P. Molotch, T. H. Painter, M. D. Dettinger, R. Rice, and J. Dozier. Mountain hydrology of the western United States. *Water Resources Research*, 42(8):1–13, 2006. ISSN 00431397. doi: 10.1029/2005WR004387.

R. C. Bales, M. L. Goulden, C. T. Hunsaker, M. H. Conklin, P. C. Hartsough, A. T. O'Geen, J. W. Hopmans, and M. Safeeq. Mechanisms controlling the impact of multi-year drought on mountain hydrology. *Scientific Reports*, 8(1):1–8, 2018a. ISSN 20452322. doi: 10.1038/s41598-017-19007-0.

R. C. Bales, E. M. Stacy, X. Meng, M. H. Conklin, P. B. Kirchner, and Z. Zheng. Spatially distributed water-balance and meteorological data from the Wolverton catchment, Sequoia National Park, California. *Earth System Science Data*, 10(4):2115–2122, 2018b. ISSN 18663516. doi: 10.5194/essd-10-2115-2018.

P. Bartelt and M. Lehning. A physical SNOWPACK model for the Swiss avalanche warning Part I: Numerical model. *Cold Regions Science and Technology*, 2002. ISSN 0165232X. doi: 10.1016/S0165-232X(02)00074-5.

P. Beaudry and D. L. Golding. Snowmelt during rain-on-snow in coastal British Columbia. In *Proceedings of the 51st Annual Western Snow Conference*, Vancouver, Washington, 1983. Western Snow Conference. URL `sites/westernsnowconference.org/PDFs/1983Beaudry.pdf`.

N. Berg, R. Osterhuber, and J. Bergman. Rain-induced outflow from deep snowpacks in the central sierra nevada, California. *Hydrological Sciences Journal*, 36(6):611–629, 1991. ISSN 21503435. doi: 10.1080/02626669109492547.

W. R. Berghuijs, R. A. Woods, and M. Hrachowitz. A precipitation shift from snow towards rain leads to a decrease in streamflow. *Nature Climate Change*, 4(7):583–586, 2014. ISSN 17586798. doi: 10.1038/nclimate2246.

S. N. Berris and R. D. Harr. Comparative snow accumulation and melt during rainfall in forested and clear-cut plots in the Western Cascades of Oregon. *Water Resources Research*, 23(1):135–142, 1987. ISSN 19447973. doi: 10.1029/WR023i001p00135.

K. Beven. Changing ideas in hydrology - the case of physically-based models. *Journal of Hydrology*, 105:157–172, 1989.

K. Beven. A manifesto for the equifinality thesis. In *Journal of Hydrology*, 2006. doi: 10.1016/j.jhydrol.2005.07.007.

K. Beven, E. Wood, and M. Sivapalan. On hydrological heterogeneity - catchment morphology and catchment response. *Journal of Hydrology*, 100:353–375, 1988.

C. Bishop. *Pattern recognition and machine learning*. Springer, Berlin, 2006. ISBN 0-387-31073-8.

G. Blöschl and M. Sivapalan. Scale issues in hydrological modelling: a review. *Hydrological Processes*, 9(3-4):251–290, 1995. ISSN 10991085. doi: 10.1002/hyp.3360090305.

G. Blöschl, R. B. Grayson, and M. Sivapalan. On the representative elementary area (REA) concept and its utility for distributed rainfall-runoff modelling. *Hydrological Processes*, 9 (3-4):313–330, 1995. ISSN 10991085. doi: 10.1002/hyp.3360090307.

G. F. Boisramé, S. E. Thompson, M. Kelly, J. Cavalli, K. M. Wilkin, and S. L. Stephens. Vegetation change during 40 years of repeated managed wildfires in the Sierra Nevada, California. *Forest Ecology and Management*, 402(July):241–252, 2017. ISSN 03781127. doi: 10.1016/j.foreco.2017.07.034.

M. Bongio, F. Avanzi, and C. De Michele. Hydroelectric power generation in an Alpine basin: future water-energy scenarios in a run-of-the-river plant. *Advances in Water Resources*, 94:318–331, 2016. ISSN 0309-1708. doi: 10.1016/j.advwatres.2016.05.017.

L. Breiman. Random Forest. *Machine Learning*, 45:5–32, 2001. doi: 10.1023/A:1010933404324.

L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. *Classification and Regression Trees*. Taylor & Francis, 1984. ISBN 9780412048418.

P. D. Broxton, A. A. Harpold, J. A. Biederman, P. A. Troch, N. P. Molotch, and P. D. Brooks. Quantifying the effects of vegetation structure on snow accumulation and ablation in mixed-conifer forests. *Ecohydrology*, 8(6):1073–1094, 2015. ISSN 19360592. doi: 10.1002/eco.1565.

M. Budyko. *Climate and Life*. Academic Press, Inc., 1974. ISBN 0-12-139450-6.

N. Burley and A. Fabbiani-Leon. PRMS San Joaquin Model Update. In *California Snow Surveys 64th Annual Meeting of Cooperators*, Kings Beach, CA, 2018. California Cooperative Snow Survey Program. URL `http://cdec.water.ca.gov/snow/meeting/2018/Wednesday/17{_}PRMS{_}SanJoaquin{_}Model{_}Update.pdf`.

California Department of Water Resources. History of Snow Surveying, 2012. URL `https://cdec.water.ca.gov/snow/info/HistSnowSurvey.html`.

N. Cannone, S. Sgorbati, and M. Guglielmin. Unexpected impacts of climate change on alpine vegetation. *Frontiers in Ecology and the Environment*, 5(7):360–364, 2007. ISSN 15409295. doi: 10.1890/1540-9295(2007)5[360:UIOCCO]2.0.CO;2.

T. M. Carpenter and K. P. Georgakakos. Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales. *Journal of Hydrology*, 329(1):174–185, 2006. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2006.02.013.

D. R. Cayan, T. Das, D. W. Pierce, T. P. Barnett, M. Tyree, and A. Gershunov. Future dryness in the southwest US and the hydrology of the early 21st century drought. *Proceedings of the National Academy of Sciences*, 107(50):21271–21276, 2010. ISSN 0027-8424. doi: 10.1073/pnas.0912391107.

M. Čeh, M. Kilibarda, A. Lisec, and B. Bajat. Estimating the performance of random forest versus multiple regression for predicting prices of the apartments. *ISPRS International Journal of Geo-Information*, 7(5), 2018. ISSN 22209964. doi: 10.3390/ijgi705168.

J. M. Chen, Black, and T. A. Defining leaf area index for non-flat leaves. *Plant, Cell & Environment*, 15(4):421–429, 1992. ISSN 13653040. doi: 10.1111/j.1365-3040.1992.tb00992.x.

Copernicus. Sentinel data, 2020. URL `https://sentinel.esa.int/web/sentinel/home`.

G. Cui, R. Bales, R. Rice, M. Anderson, F. Avanzi, P. Hartsough, and M. Conklin. Detecting rain-snow transition elevations in mountain basins using wireless-sensor networks. *Journal of Hydrometeorology*, 21(9):2061–2081, 2020. doi: 10.1175/JHM-D-20-0028.1.

A. Dai. Increasing drought under global warming in observations and models. *Nature Climate Change*, 3(1):52–58, 2013. ISSN 1758-6798. doi: 10.1038/nclimate1633.

C. Daly, M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, March, 2008. doi: 10.1002/joc.1688.

C. De Michele, F. Avanzi, A. Ghezzi, and C. Jommi. Investigating the dynamics of bulk snow density in dry and wet conditions using a one-dimensional model. *Cryosphere*, 7(2): 433–444, 2013. ISSN 19940424. doi: 10.5194/tc-7-433-2013.

M. D. Dettinger, F. M. Ralph, T. Das, P. J. Neiman, and D. R. Cayan. Atmospheric Rivers, Floods and the Water Resources of California. *Water*, 3(2):445–478, 2011. doi: 10.3390/w3020445.

R. J. Donohue, M. L. Roderick, and T. R. McVicar. On the importance of including vegetation dynamics in Budyko's hydrological model. *Hydrology and Earth System Sciences*, 11(2):983–995, 2007. ISSN 16077938. doi: 10.5194/hess-11-983-2007.

J. M. Donovan and K. M. Koczot. User's Manual for the Draper Climate-Distribution Software Suite with Data-Evaluation Tools. Technical report, U.S. Geological Survey Techniques and Methods 7-C22, 2019. URL `https://doi.org/10.3133/tm7C22`.

J. Dozier. Mountain hydrology, snow color, and the fourth paradigm. *Eos*, 2011. ISSN 00963941. doi: 10.1029/2011EO430001.

C. Du, F. Sun, J. Yu, X. Liu, and Y. Chen. New interpretation of the role of water balance in an extended Budyko hypothesis in arid regions. *Hydrology and Earth System Sciences*, 20(1):393–409, 2016. ISSN 16077938. doi: 10.5194/hess-20-393-2016.

Q. Duan, S. Sorooshian, and V. Gupta. Effective and Efficient Global Optimization. *Water Resources Research*, 28(4):1015–1031, 1992. doi: 10.1029/91WR02985.

Q. Duan, S. Sorooshian, and V. K. Gupta. Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology*, 158(3-4):265–284, 1994. ISSN 00221694. doi: 10.1016/0022-1694(94)90057-4.

Q. Y. Duan, V. K. Gupta, and S. Sorooshian. Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, 76(3):501–521, mar 1993. ISSN 1573-2878. doi: 10.1007/BF00939380.

E. Dunford and C. Niederhof. Influence of Aspen, Young Lodgepole Pine, and Open Grassland Types Upon Factors Affecting Water Yield. *Journal of Forestry*, 42(9):673–677, 1944. ISSN 0022-1201. doi: 10.1093/jof/42.9.673.

EROS Data Center. National Elevatoin Dataset, 1999. URL `https://www.usgs.gov/core-science-systems/national-geospatial-program/national-map`.

J. Fan, Y. Xu, H. Ge, and W. Yang. Vegetation growth variation in relation to topography in Horqin Sandy Land. *Ecological Indicators*, 113:106215, 2020. ISSN 1470-160X. doi: 10.1016/j.ecolind.2020.106215.

M. Fehlmann, E. Gascón, M. Rohrer, M. Schwarb, and M. Stoffel. Improving Medium-Range Forecasts of Rain-on-Snow Events in Prealpine Areas. *Water Resources Research*, 55, 2019. doi: 10.1029/2018WR024644.

J. Feiccabrino, W. Graff, A. Lundberg, N. Sundström, and D. Gustafsson. Meteorological Knowledge Useful for the Improvement of Snow Rain Separation in Surface Based Models. *Hydrology*, 2(4):266–288, 2015. doi: 10.3390/hydrology2040266.

J. Fiddes and S. Gruber. TopoSUB: A tool for efficient large area numerical modelling in complex topography at sub-grid scales. *Geoscientific Model Development*, 5(5):1245–1257, 2012. ISSN 1991959X. doi: 10.5194/gmd-5-1245-2012.

W.-a. Flügel. Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the river Bröl, Germany. *Hydrological Process*, 9(September 1994):423–436, 1995. doi: 10.1002/hyp.3360090313.

W.-a. Flügel. Combining GIS with regional hydrological modelling using hydrological response units (HRUs): An application from Germany. *Mathematics and Computers in Simulation*, 43:297–304, 1997. doi: 10.1016/S0378-4754(97)00013-X.

G. J. Freeman. Runoff Impacts of Climate Change on Northern California's Watersheds as Influenced by Geology and Elevation. In *Proceedings of 76th Annual Western Snow Conference*, pages 23–34, Hood River, OR, 2008. Western Snow Conference. URL `sites/westernsnowconference.org/PDFs/2008Freeman.pdf`.

G. J. Freeman. Climate change and the changing water balance for California's North Fork Feather River. In *Proceedings of the 79th Annual Western Snow Conference*, pages 71–82, Stateline, NV, 2011. Western Snow Conference. URL `sites/westernsnowconference.org/PDFs/2011Freeman.pdf`.

M. Friedl and D. Sulla-Menashe. MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006 [Data set], 2019. URL `https://doi.org/10.5067/MODIS/MCD12Q1.006`.

B. Fu. On the Calculation of the Evaporation from Land Surface. *Chinese Journal of Atmospheric Sciences*, 5(23-31), 1981.

T. Y. Gan and G. F. Biftu. Automatic calibration of conceptual rainfall-runoff models: Optimization algorithms, catchment conditions, and model structure. *Water Resources Research*, 1996. ISSN 00431397. doi: 10.1029/95WR02195.

J. M. García-Ruiz, I. I. López-Moreno, S. M. Vicente-Serrano, T. Lasanta-Martínez, and S. Beguería. Mediterranean water resources in a global change scenario. *Earth-Science Reviews*, 105(3-4):121–139, 2011. doi: 10.1016/j.earscirev.2011.01.006.

J. Garvelmann, S. Pohl, and M. Weiler. Spatio-temporal controls of snowmelt and runoff generation during rain-on-snow events in a mid-latitude mountain catchment. *Hydrological Processes*, 29:3649–3664, 2015. doi: 10.1002/hyp.10460.

M. Ghestem, R. C. Sidle, and A. Stokes. The Influence of Plant Root Systems on Subsurface Flow: Implications for Slope Stability. *BioScience*, 61(11):869–879, nov 2011. ISSN 0006-3568. doi: 10.1525/bio.2011.61.11.6.

K. E. Gleason and A. W. Nolin. Charred forests accelerate snow albedo decay: parameterizing the post-fire radiative forcing on snow for three years following fire. *Hydrological Processes*, 30(21):3855–3870, 2016. ISSN 10991085. doi: 10.1002/hyp.10897.

S. J. Gnann, R. A. Woods, and N. J. Howden. Is There a Baseflow Budyko Curve? *Water Resources Research*, 55(4):2838–2855, 2019. ISSN 19447973. doi: 10.1029/2018WR024464.

M. L. Goulden and R. C. Bales. California forest die-off linked to multi-year deep soil drying in 2012–2015 drought. *Nature Geoscience*, 12(8):632–637, 2019. ISSN 17520908. doi: 10.1038/s41561-019-0388-5.

B. Gregorutti, B. Michel, and P. Saint-Pierre. Correlation and variable importance in random forests. *Statistics and Computing*, 10, 2013. doi: 10.1007/s11222-016-9646-1.

P. Greve, L. Gudmundsson, B. Orlowsky, and S. I. Seneviratne. A two-parameter Budyko function to represent conditions under which evapotranspiration exceeds precipitation. *Hydrology and Earth System Sciences*, 20(6):2195–2205, 2016. ISSN 16077938. doi: 10.5194/hess-20-2195-2016.

B. Guan, N. P. Molotch, D. E. Waliser, S. M. Jepsen, T. H. Painter, and J. Dozier. Snow water equivalent in the Sierra Nevada: Blending snow sensor observations with snowmelt model simulations. *Water Resources Research*, 49(8):5029–5046, 2013. ISSN 00431397. doi: 10.1002/wrcr.20387.

B. Guan, D. E. Waliser, F. M. Ralph, E. J. Fetzer, and P. J. Neiman. Hydrometeorological characteristics of rain-on-snow events associated with atmospheric rivers. *Geophysical Research Letters*, 43:2964–2973, 2016. doi: 10.1002/2016GL067978.

H. V. Gupta, S. Sorooshian, and P. O. Yapo. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, 34(4):751–763, 1998. ISSN 00431397. doi: 10.1029/97WR03495.

H. V. Gupta, H. Kling, K. K. Yilmaz, and G. F. Martinez. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1-2):80–91, 2009. ISSN 00221694. doi: 10.1016/j.jhydrol.2009.08.003.

W. J. Hahm, D. N. Dralle, D. M. Rempe, A. B. Bryk, S. E. Thompson, T. E. Dawson, and W. E. Dietrich. Low Subsurface Water Storage Capacity Relative to Annual Rainfall Decouples Mediterranean Plant Productivity and Water Use From Rainfall Variability. *Geophysical Research Letters*, 2019a. ISSN 19448007. doi: 10.1029/2019GL083294.

W. J. Hahm, D. M. Rempe, D. N. Dralle, T. E. Dawson, S. M. Lovill, A. B. Bryk, D. L. Bish, J. Schieber, and W. E. Dietrich. Lithologically Controlled Subsurface Critical Zone Thickness and Water Storage Capacity Determine Regional Plant Community

Composition. *Water Resources Research*, 55(4):3028–3055, 2019b. ISSN 19447973. doi: 10.1029/2018WR023760.

W. Hamon. Computation of direct runoff amounts from storm rainfall. *International Association of Hydrological Sciences Publication*, 63:52–62, 1963.

A. A. Harpold, M. L. Kaplan, P. Zion Klos, T. Link, J. P. McNamara, S. Rajagopal, R. Schumer, and C. M. Steele. Rain or snow: Hydrologic processes, observations, prediction, and research needs. *Hydrology and Earth System Sciences*, 21(1):1–22, 2017. ISSN 16077938. doi: 10.5194/hess-21-1-2017.

R. D. Harr. Some characteristics and consequences of snowmelt during rainfall in western Oregon. *Journal of Hydrology*, 53(3-4):277–304, 1981. ISSN 00221694. doi: 10.1016/0022-1694(81)90006-8.

R. D. Harr and S. N. Berris. Snow Accumulation and subsequent melt during rainfall in forested and clearcut plots in Western Oregon. In *Western Snow Conference*, pages 38–44, Vancouver, Washington, 1983. Western Snow Conference. URL `sites/westernsnowconference.org/PDFs/1983Harr.pdf`.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer-Verlag, 2nd edition, 2009. ISBN 0387848576. URL `web.stanford.edu/{~}hastie/ElemStatLearn/`.

L. E. Hay and M. Umemoto. Multiple-objective stepwise calibration using Luca. *U. S. Geological Survey Open-File Report*, page 25, 2006. URL `https://pubs.usgs.gov/of/2006/1323/`.

L. E. Hay, G. H. Leavesley, M. P. Clark, S. L. Markstrom, R. J. Viger, and M. Umemoto. Step wise, multiple objective calibration of a hydrologic model for a snowmelt dominated basin. *Journal of the American Water Resources Association*, 42(4):877–890, 2006. ISSN 1093474X. doi: 10.1111/j.1752-1688.2006.tb04501.x.

M. He, M. Russo, and M. Anderson. Hydroclimatic characteristics of the 2012-2015 California drought from an operational perspective. *Climate*, 5(1):1987–1992, 2017. ISSN 22251154. doi: 10.3390/cli5010005.

R. W. Hofste, P. Reig, and L. Schleifer. 17 Countries, Home to One-Quarter of the World's Population, Face Extremely High Water Stress. Technical report, World Resources Institute, Washington, D.C., 2019. URL `https://www.wri.org/blog/2019/08/17-countries-home-one-quarter-world-population-face-extremely-high-water-stress`.

M. Hrachowitz and M. P. Clark. The complementary merits of competing modelling philosophies in hydrology. *Hydrology and Earth System Sciences*, 21(8):3953–3973, 2017. ISSN 16077938. doi: 10.5194/hess-21-3953-2017.

T. Hu, Q. Ma, Y. Su, J. J. Battles, B. M. Collins, S. L. Stephens, M. Kelly, and Q. Guo. A simple and integrated approach for fire severity assessment using bi-temporal airborne

LiDAR data. *International Journal of Applied Earth Observation and Geoinformation*, 78:25–38, 2019. doi: 10.1016/j.jag.2019.01.007.

S. Huang, P. Li, Q. Huang, G. Leng, B. Hou, and L. Ma. The propagation from meteorological to hydrological drought and its potential influence factors. *Journal of Hydrology*, 2017. ISSN 00221694. doi: 10.1016/j.jhydrol.2017.01.041.

J. Hughes, C. Langevin, and E. Banta. Documentation for the MODFLOW 6 Framework. *USGS: Techniques and Methods 6-A57*, 2017. ISSN 1600-051X; 0303-6979.

Y. Hundecha, B. Arheimer, C. Donnelly, and I. Pechlivanidis. A regional parameter estimation scheme for a pan-European multi-basin model. *Journal of Hydrology: Regional Studies*, 6:90–111, 2016. ISSN 2214-5818. doi: 10.1016/j.ejrh.2016.04.002.

C. T. Hunsaker and M. Safeeq. Kings River Experimental Watersheds meteorology data, 2018.

W. W. Immerzeel, A. F. Lutz, M. Andrade, A. Bahl, H. Biemans, T. Bolch, S. Hyde, S. Brumby, B. J. Davies, A. C. Elmore, A. Emmer, M. Feng, A. Fernández, U. Haritashya, J. S. Kargel, M. Koppes, P. D. A. Kraaijenbrink, A. V. Kulkarni, P. A. Mayewski, S. Nepal, P. Pacheco, T. H. Painter, F. Pellicciotti, H. Rajaram, S. Rupper, A. Sinisalo, A. B. Shrestha, D. Viviroli, Y. Wada, C. Xiao, T. Yao, and J. E. M. Baillie. Importance and vulnerability of the world's water towers. *Nature*, 577(7790):364–369, 2020. ISSN 1476-4687. doi: 10.1038/s41586-019-1822-y.

F. Jaramillo, N. Cory, B. Arheimer, H. Laudon, Y. Van Der Velde, T. B. Hasper, C. Teutschbein, and J. Uddling. Dominant effect of increasing forest biomass on evapotranspiration: Interpretations of movement in Budyko space. *Hydrology and Earth System Sciences*, 22(1):567–580, 2018. ISSN 16077938. doi: 10.5194/hess-22-567-2018.

K. Jennings and N. Molotch. The sensitivity of modeled snow accumulation and melt to precipitation phase methods across a climatic gradient. *Hydrology and Earth System Sciences*, 23:3765–3786, 2019. doi: 10.5194/hess-23-3765-2019.

K. S. Jennings, T. G. Kittel, and N. P. Molotch. Observations and simulations of the seasonal evolution of snowpack cold content and its relation to snowmelt and the snowpack energy budget. *Cryosphere*, 12(5):1595–1614, 2018. ISSN 19940424. doi: 10.5194/tc-12-1595-2018.

M. M. Kalcic, I. Chaubey, and J. Frankenberger. Defining Soil and Water Assessment Tool (SWAT) hydrologic response units (HRUs) by field boundaries. *International Journal of Agricultural and Biological Engineering*, 8(3):1–12, 2015. ISSN 19346352. doi: 10.3965/j.ijabe.20150803.951.

T. Karpouzoglou and S. Vij. Waterscape: a perspective for understanding the contested geography of water. *Wiley Interdisciplinary Reviews: Water*, 4(3):e1210, 2017. doi: 10.1002/wat2.1210.

R. Kattelmann. Water release from a forested snowpack during rainfall. In *Forest Hydrology and Watershed Management*, number August, page Pub. no. 167, Vancouver, 1987. URL `http://hydrologie.org/redbooks/a167/167026.pdf`.

R. Kattelmann. Flooding from rain-on-snow events in the Sierra Nevada. In *Destructive Water: Water-Caused Natural Disasters, their Abatement and Control*, number 239, pages 59–65, Anaheim, California, 1997. International Association of Hydrological Sciences. URL `http://hydrologie.org/redbooks/a239/iahs{_}239{_}0059.pdf`.

B. Khakbaz, B. Imam, K. Hsu, and S. Sorooshian. From lumped to distributed via semi-distributed: Calibration strategies for semi-distributed hydrologic models. *Journal of Hydrology*, 418-419:61–77, 2012. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2009.02.021.

U. Khan, N. K. Tuteja, and A. Sharma. Delineating hydrologic response units in large upland catchments and its evaluation using soil moisture simulations. *Environmental Modelling & Software*, 46:142–154, 2013. ISSN 1364-8152. doi: 10.1016/j.envsoft.2013.03.005.

U. Khan, N. K. Tuteja, A. Sharma, S. Lucas, B. Murphy, and B. Jenkins. Applicability of Hydrologic Response Units in low topographic relief catchments and evaluation using high resolution aerial photograph analysis. *Environmental Modelling and Software*, 81:56–71, 2016. ISSN 13648152. doi: 10.1016/j.envsoft.2016.03.010.

T. Kim, J. Y. Shin, H. Kim, and J. H. Heo. Ensemble-Based Neural Network Modeling for Hydrologic Forecasts: Addressing Uncertainty in the Model Structure and Input Variable Selection. *Water Resources Research*, 2020. ISSN 19447973. doi: 10.1029/2019WR026262.

J. W. Kirchner. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, 42 (3):1–5, 2006. ISSN 00431397. doi: 10.1029/2005WR004362.

V. Klemeš. Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, 31(1):13–24, 1986. doi: 10.1080/02626668609491024.

H. Kling, M. Fuchs, and M. Paulin. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424-425:264–277, 2012. ISSN 00221694. doi: 10.1016/j.jhydrol.2012.01.011.

P. Z. Klos, M. L. Goulden, C. S. Riebe, C. L. Tague, A. T. O'Geen, B. A. Flinchum, M. Safeeq, M. H. Conklin, S. C. Hart, A. A. Berhe, P. C. Hartsough, W. S. Holbrook, and R. C. Bales. Subsurface plant-accessible water in mountain ecosystems with a Mediterranean climate. *Wiley Interdisciplinary Reviews: Water*, 5(3):e1277, 2018. ISSN 2049-1948. doi: 10.1002/wat2.1277.

E. E. Knapp, C. N. Skinner, M. P. North, and B. L. Estes. Long-term overstory and understory change following logging and fire exclusion in a Sierra Nevada mixed-conifer forest. *Forest Ecology and Management*, 310:903–914, 2013. ISSN 03781127. doi: 10.1016/j.foreco.2013.09.041.

K. M. Koczot, A. E. Jeton, B. J. Mcgurk, and M. D. Dettinger. Precipitation-Runoff Processes in the Feather River Basin , Northeastern California , with Prospects for Streamflow Predictability, Water Years 1971 – 97 U . S . Department of the Interior. Technical report, U.S. Geological Survey Scientific Investigations Report, Reston, VA, 2005. URL https://pubs.usgs.gov/sir/2004/5202/sir2004-5202.pdf.

L. Korhonen, I. Korpela, J. Heiskanen, and M. Maltamo. Airborne discrete-return LIDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. *Remote Sensing of Environment*, 115(4):1065–1080, 2011. ISSN 00344257. doi: 10.1016/j.rse.2010.12.011.

M. A. Lamjiri, M. D. Dettinger, F. M. Ralph, N. S. Oakley, and J. J. Rutz. Hourly Analyses of the Large Storms and Atmospheric Rivers that Provide Most of California' s Precipitation in Only 10 to 100 Hours per Year. *San Francisco Estuary & Watershed Science*, 16(4): 1–17, 2018. doi: 10.15447/sfews.2018v16iss4art1.

LANDFIRE. Existing Vegetation Type, 2013a. URL https://landfire.gov/evt.php.

LANDFIRE. Existing Vegetation Cover, 2013b. URL https://landfire.gov/evc.php.

C. D. Langevin, J. D. Hughes, E. Banta, R. G. Niswonger, S. Panday, and A. Provost. Documentation For The MODFLOW 6 Groundwater Flow Model. *U.S. Geological Survey*, 2017. ISSN 2328-7055.

D. J. Leathers, D. R. Kiuck, and S. Kroczynski. The Severe Flooding Event of January 1996 across North-Central Pennsylvania. *Bulletin of the American Meteorological Society*, 79 (5):785–798, 1998. doi: 10.1175/1520-0477(1998)079⟨0785:TSFEOJ⟩2.0.CO;2.

G. H. Leavesley, R. Lichty, B. Troutman, and L. Saindon. Precipitation-Runoff Modeling System: User's manual. Technical report, U.S. Geological Survey, Denver, CO, 1983. URL https://pubs.usgs.gov/wri/1983/4238/report.pdf.

M. Lehning, P. Bartelt, B. Brown, and C. Fierz. A physical SNOWPACK model for the Swiss avalanche warning Part III: meteorological forcing, thin layer formation and evaluation. *Cold Regions Science and Technology*, 2002a. ISSN 0165232X.

M. Lehning, P. Bartelt, B. Brown, C. Fierz, and P. Satyawali. A physical SNOWPACK model for the Swiss avalanche warning Part II. Snow microstructure. *Cold Regions Science and Technology*, 2002b. ISSN 0165232X. doi: 10.1016/S0165-232X(02)00073-3.

D. Li, M. Pan, Z. Cong, L. Zhang, and E. Wood. Vegetation control on water and energy balance within the Budyko framework. *Water Resources Research*, 49(2):969–976, 2013. ISSN 00431397. doi: 10.1002/wrcr.20107.

Y. Li, C. Liu, W. Yu, D. Tian, and P. Bai. Response of streamflow to environmental changes: A Budyko-type analysis based on 144 river basins over China. *Science of the Total Environment*, 664:824–833, 2019. ISSN 18791026. doi: 10.1016/j.scitotenv.2019.02.011.

J. Liu, Q. Zhang, V. P. Singh, and P. Shi. Contribution of multiple climatic variables and human activities to streamflow changes across China. *Journal of Hydrology*, 545:145–162, 2017. ISSN 00221694. doi: 10.1016/j.jhydrol.2016.12.016.

F. Lobligeois, V. Andréassian, C. Perrin, P. Tabary, and C. Loumagne. When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events. *Hydrology and Earth System Sciences*, 18(2):575–594, 2014. ISSN 10275606. doi: 10.5194/hess-18-575-2014.

J. D. Lundquist, S. E. Dickerson-Lange, J. A. Lutz, and N. C. Cristea. Lower forest density enhances snow retention in regions with warmer winters: A global framework developed from plot-scale observations and modeling. *Water Resources Research*, 49(10):6356–6370, 2013. ISSN 00431397. doi: 10.1002/wrcr.20504.

Q. Ma, Y. Su, and Q. Guo. Comparison of Canopy Cover Estimations From Airborne LiDAR, Aerial Imagery, and Satellite Imagery. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(9):4225–4236, 2017. doi: 10.1109/JSTARS.2017.2711482.

S. A. Margulis, G. Cortés, M. Girotto, and M. Durand. A Landsat-Era Sierra Nevada Snow Reanalysis (1985–2015). *Journal of Hydrometeorology*, 17(4):1203–1221, 2016. ISSN 1525-755X. doi: 10.1175/JHM-D-15-0177.1.

D. Marks, J. Kimball, D. Tingey, and T. Link. The sensitivity of snowmelt processes to climate conditions and forest cover during rain-on-snow: a case study of the 1996 Pacific Northwest flood. *Hydrological Processes*, 12(10-11):1569–1587, 1998. ISSN 08856087. doi: 10.1002/(SICI)1099-1085(199808/09)12:10/11⟨1569::AID-HYP682⟩3.0.CO;2-L.

D. Marks, A. Winstral, M. Reba, J. Pomeroy, and M. Kumar. An evaluation of methods for determining during-storm precipitation phase and the rain/snow transition elevation at the surface in a mountain basin. *Advances in Water Resources*, 55:98–110, 2013. ISSN 03091708. doi: 10.1016/j.advwatres.2012.11.012.

S. L. Markstrom. No Title. personal communication, 2019.

S. L. Markstrom, R. S. Regan, L. E. Hay, R. J. Viger, R. M. Webb, R. A. Payn, and J. H. LaFontaine. PRMS-IV, the precipitation-runoff modeling system, version 4. In *U.S. Geological Survey Techniques and Methods*, chapter B7, page 158. U.S. Geological Survey, 2015. doi: 10.3133/tm6B7.

S. L. Markstrom, L. E. Hay, and M. P. Clark. Towards simplification of hydrologic modeling : identification of dominant processes. *Hydrology and Earth System Sciences*, 20:4655–4671, 2016. doi: 10.5194/hess-20-4655-2016.

A. M. Marshall, J. T. Abatzoglou, T. E. Link, and C. J. Tennant. Projected changes in interannual variability of peak snowpack amount and timing in the Western United States. *Geophysical Research Letters*, 2019. ISSN 0094-8276. doi: 10.1029/2019GL083770.

I. Masih, S. Maskey, F. E. Mussá, and P. Trambauer. A review of droughts on the African continent: A geospatial and long-term perspective. *Hydrology and Earth System Sciences*, 18(9):3635–3649, 2014. ISSN 16077938. doi: 10.5194/hess-18-3635-2014.

S. Maskey and P. Trambauer. *Hydrological Modeling for Drought Assessment*. Elsevier Inc., Delft, The Netherlands, 2015. ISBN 9780123964700. doi: 10.1016/B978-0-12-394846-5.00010-2.

T. Mastrotheodoros, C. Pappas, P. Molnar, P. Burlando, G. Manoli, J. Parajka, R. Rigon, B. Szeles, M. Bottazzi, P. Hadjidoukas, and S. Fatichi. More green and less blue water in the Alps during warmer summers. *Nature Climate Change*, 10(2):155–161, 2020. ISSN 17586798. doi: 10.1038/s41558-019-0676-5.

J. D. Maxwell, A. Call, and S. B. St. Clair. Wildfire and topography impacts on snow accumulation and retention in montane forests. *Forest Ecology and Management*, 432: 256–263, 2019. ISSN 0378-1127. doi: https://doi.org/10.1016/j.foreco.2018.09.021.

A. B. Mazurkiewicz, D. G. Callery, and J. J. McDonnell. Assessing the controls of the snow energy balance and water available for runoff in a rain-on-snow environment. *Journal of Hydrology*, 354(1-4):1–14, 2008. ISSN 00221694. doi: 10.1016/j.jhydrol.2007.12.027.

G. J. McCabe, M. P. Clark, and L. E. Hay. Rain-on-snow events in the western United States. *Bulletin of the American Meteorological Society*, 88(3):319–328, 2007. ISSN 00030007. doi: 10.1175/BAMS-88-3-319.

G. McLachlan and D. Peel. *Finite Mixture Models*. John Wiley & Sons, New York, 2004. ISBN 0471006262.

T. R. McVicar, M. L. Roderick, R. J. Donohue, and T. G. Van Niel. Less bluster ahead? ecohydrological implications of global trends of terrestrial near-surface wind speeds. *Ecohydrology*, 5(4):381–388, 2012. ISSN 19360584. doi: 10.1002/eco.1298.

J. D. Miller, B. M. Collins, J. A. Lutz, S. L. Stephens, J. W. van Wagtendonk, and D. A. Yasuda. Differences in wildfires among ecoregions and land management agencies in the Sierra Nevada region, California, USA. *Ecosphere*, 3(9):art80, 2012. ISSN 2150-8925. doi: 10.1890/es12-00158.1.

P. C. Milly, J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer. Climate change: Stationarity is dead: Whither water management?, 2008. ISSN 00368075.

N. P. Molotch and R. C. Bales. Scaling snow observations from the point to the grid element: Implications for observation network design. *Water Resources Research*, 2005. ISSN 00431397. doi: 10.1029/2005WR004229.

N. P. Molotch and S. A. Margulis. Estimating the distribution of snow water equivalent using remotely sensed snow cover data and a spatially distributed snowmelt model: A multi-resolution, multi-sensor comparison. *Advances in Water Resources*, 2008. ISSN 03091708. doi: 10.1016/j.advwatres.2008.07.017.

E. L. Montoya, J. Dozier, and W. Meiring. Biases of April 1 snow water equivalent records in the Sierra Nevada and their associations with large-scale climate indices. *Geophysical Research Letters*, 41(16):5912–5918, 2014. ISSN 19448007. doi: 10.1002/2014GL060588.

Mountain Partnership. Mountains as the water towers of the world: a call for action on the sustainable development goals. Technical report, Food and Agriculture Organization of the United Nations, 2014. URL `http://www.fao.org/fileadmin/templates/mountain{_}partnership/doc/POLICY{_}BRIEFS/SDGs{_}and{_}mountains{_}water{_}EN.pdf`.

R. Moussa and J. P. Lhomme. The Budyko functions under non-steady-state conditions. *Hydrology and Earth System Sciences*, 20(12):4867–4879, 2016. ISSN 16077938. doi: 10.5194/hess-20-4867-2016.

K. N. Musselman, N. P. Molotch, and P. D. Brooks. Effects of vegetation on snow accumulation and ablation in a mid-latitude sub-alpine forest. *Hydrological Processes*, 22: 2767–2776, 2008. doi: 10.1002/hyp.7050.

K. N. Musselman, F. Lehner, K. Ikeda, M. P. Clark, A. F. Prein, C. Liu, M. Barlage, and R. Rasmussen. Projected increases and shifts in rain-on-snow flood risk over western North America. *Nature Climate Change*, 3, 2018. ISSN 1758-678X. doi: 10.1038/s41558-018-0236-4.

J. E. Nash and J. V. Sutcliffe. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3):282–290, 1970. ISSN 0022-1694. doi: 10.1016/0022-1694(70)90255-6.

National Oceanic and Atmospheric Administration. Sacramento-Soil Moisture Accounting Model, 2002. URL `https://www.nws.noaa.gov/oh/hrl/nwsrfs/users{_}manual/part2/{_}pdf/23sacsma.pdf?referrer=Baker`.

T. Ning, Z. Li, and W. Liu. Vegetation dynamics and climate seasonality jointly control the interannual catchment water balance in the Loess Plateau under the Budyko framework. *Hydrology and Earth System Sciences*, 21(3):1515–1526, 2017. ISSN 16077938. doi: 10.5194/hess-21-1515-2017.

T. Ning, S. Zhou, F. Chang, H. Shen, Z. Li, and W. Liu. Interaction of vegetation, climate and topography on evapotranspiration modelling at different time scales within the Budyko framework. *Agricultural and Forest Meteorology*, 275(January):59–68, 2019. ISSN 01681923. doi: 10.1016/j.agrformet.2019.05.001.

T. Ning, Z. Li, Q. Feng, W. Chen, and Z. Li. Effects of forest cover change on catchment evapotranspiration variation in China. *Hydrological Processes*, 34(10):2219–2228, 2020. ISSN 10991085. doi: 10.1002/hyp.13719.

N. Obojes, M. Bahn, E. Tasser, J. Walde, N. Inauen, E. Hiltbrunner, P. Saccone, J. Lochet, J. C. Clément, S. Lavorel, U. Tappeiner, and C. Körner. Vegetation effects on the water balance of mountain grasslands depend on climatic conditions. *Ecohydrology*, 8(4):552–569, 2015. ISSN 19360592. doi: 10.1002/eco.1524.

A. T. O'Geen, M. Safeeq, J. Wagenbrenner, E. Stacy, P. Hartsough, S. Devine, Z. Tian, R. Ferrell, M. Goulden, J. W. Hopmans, and R. Bales. Southern Sierra Critical Zone Observatory and Kings River Experimental Watersheds: A synthesis of measurements, new insights, and future directions. *Vadose Zone Journal*, 17(1), 2018. ISSN 15391663. doi: 10.2136/vzj2018.04.0081.

A. P. O'Grady, J. L. Carter, and J. Bruce. Can we predict groundwater discharge from terrestrial ecosystems using existing eco-hydrological concepts? *Hydrology and Earth System Sciences*, 15(12):3731–3739, 2011. ISSN 10275606. doi: 10.5194/hess-15-3731-2011.

C. A. Oroza, Z. Zheng, S. D. Glaser, D. Tuia, and R. C. Bales. Optimizing embedded sensor network design for catchment-scale snow-depth estimation using LiDAR and machine learning. *Water Resources Research*, 52(10):8174–8189, oct 2016. ISSN 00431397. doi: 10.1002/2016WR018896.

C. A. Oroza, R. C. Bales, E. M. Stacy, Z. Zheng, and S. D. Glaser. Long-Term Variability of Soil Moisture in the Southern Sierra: Measurement and Prediction. *Vadose Zone Journal*, 2018. ISSN 1539-1663. doi: 10.2136/vzj2017.10.0178.

L. Oudin, V. Andréassian, J. Lerat, and C. Michel. Has land cover a significant impact on mean annual streamflow? An international assessment using 1508 catchments. *Journal of Hydrology*, 357(3-4):303–316, 2008. ISSN 00221694. doi: 10.1016/j.jhydrol.2008.05.021.

O. Oyebode and D. Stretch. Neural network modeling of hydrological systems: A review of implementation techniques, 2019. ISSN 19397445.

T. H. Painter, D. F. Berisford, J. W. Boardman, K. J. Bormann, J. S. Deems, F. Gehrke, A. Hedrick, M. Joyce, R. Laidlaw, D. Marks, C. Mattmann, B. McGurk, P. Ramirez, M. Richardson, S. M. K. Skiles, F. C. Seidel, and A. Winstral. The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. *Remote Sensing of Environment*, 2016. ISSN 00344257. doi: 10.1016/j.rse.2016.06.018.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning*, 12:2825–2830, 2011. URL `http://jmlr.org/papers/v12/pedregosa11a.htm`.

C. Petheram, N. Potter, J. Vaze, F. Chiew, and L. Zhang. Towards better understanding of changes in rainfall-runoff relationships during the recent drought in south-eastern Australia. In *MODSIM 2011 - 19th International Congress on Modelling and Simulation - Sustaining Our Future: Understanding and Living with Uncertainty*, number December, pages 3622–3628, 2011. ISBN 9780987214317.

F. Pianosi, F. Sarrazin, and T. Wagener. A Matlab toolbox for Global Sensitivity Analysis. *Environmental Modelling & Software*, 70:80–85, 2015. ISSN 1364-8152. doi: 10.1016/j.envsoft.2015.04.009.

M. R. Pickard. *Influence of within-stand tree spatial arrangement on snowpack distribution and ablation in the Sierra Nevada, CA*. PhD thesis, UC Merced, 2015. URL `https://escholarship.org/uc/item/9p5453jj`.

J. W. Pomeroy, X. Fang, and D. G. Marks. The cold rain-on-snow event of June 2013 in the Canadian Rockies — characteristics and diagnosis. *Hydrological Processes*, 2016a. doi: 10.1002/hyp.10905.

J. W. Pomeroy, R. E. Stewart, and P. H. Whitfield. The 2013 flood event in the South Saskatchewan and Elk River basins: Causes, assessment and damages. *Canadian Water Resources Journal*, 41(1-2):105–117, 2016b. ISSN 07011784. doi: 10.1080/07011784.2015.1089190.

S. Pool, M. Vis, and J. Seibert. Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrological Sciences Journal*, 63(13-14):1941–1953, 2018. ISSN 21503435. doi: 10.1080/02626667.2018.1552002.

N. J. Potter, C. Petheram, and L. Zhang. Sensitivity of streamflow to rainfall and temperature in south-eastern Australia during the Millennium drought. In *MODSIM 2011 - 19th International Congress on Modelling and Simulation - Sustaining Our Future: Understanding and Living with Uncertainty*, number November 2014, pages 3636–3642, 2011. ISBN 9780987214317.

PRISM Climate Group. Parameter-elevation Relationships on Independent Slopes Mode, 2004. URL `http://prism.oregonstate.edu`.

A. M. Provost, C. D. Langevin, and J. Hughes. Documentation for the " XT3D " Option in the Node Property Flow (NPF) Package of MODFLOW 6. *U.S. Geological Survey*, 2017. URL `https://pubs.usgs.gov/tm/06/a56/tm6a56.pdf`.

J. Qi, S. Li, Q. Yang, Z. Xing, and F. R. Meng. SWAT Setup with Long-Term Detailed Landuse and Management Records and Modification for a Micro-Watershed Influenced by Freeze-Thaw Cycles. *Water Resources Management*, 31(12):3953–3974, 2017. ISSN 15731650. doi: 10.1007/s11269-017-1718-2.

Y. Qiu, B. Fu, J. Wang, and L. Chen. Soil moisture variation in relation to topography and land use in a hillslope catchment of the Loess Plateau, China. *Journal of Hydrology*, 240 (3):243–263, 2001. ISSN 0022-1694. doi: 10.1016/S0022-1694(00)00362-0.

F. M. Ralph and M. D. Dettinger. Storms, Floods, and the Science of Atmospheric Rivers. *Eos*, 92(32):265–272, 2011. doi: 10.1029/2011EO320001.

S. Reed, V. Koren, M. Smith, Z. Zhang, F. Moreda, D.-J. Seo, and A. DMIP Participants. Overall distributed model intercomparison project results. *Journal of Hydrology*, 298(1): 27–60, 2004. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2004.03.031.

P. Reggiani and T. H. Rientjes. Flux parameterization in the representative elementary watershed approach: Application to a natural basin. *Water Resources Research*, 41(4): 1–18, 2005. ISSN 00431397. doi: 10.1029/2004WR003693.

P. Reggiani, M. Sivapalan, and S. M. Hassanizadeh. Conservation equations governing hillslope responses: Exploring the physical basis of water balance. *Water Resources Research*, 36(7):1845–1863, 2000. ISSN 00431397. doi: 10.1029/2000WR900066.

J. Revuelto, J. López-Moreno, C. Azorin-Molina, and S. Vincente-Serrano. Canopy influence on snow depth distribution in a pine stand determined from terrestrial laser data J. *Water Resources Research*, 51, 2015. doi: 10.1002/2014WR016496.

K. Richards. Improving snow & streamflow modeling on the Feather River using intelligent information systems. In *California Snow Surveys 64th Annual Meeting of Cooperators*, Kings Beach, CA, 2018. California Cooperative Snow Survey Program. URL `http://cdec.water.ca.gov/snow/meeting/2018/Wednesday/18{_}CCSS{_}Richards{_}20181114.pdf`.

J. Richardson, L. Moskal, and S.-H. Kim. Modeling approaches to estimate effective leaf area index from aerial discrete-return LIDAR. *Agricultural and Forest Meteorology*, 149 (6-7):1152–1160, 2009. doi: 10.1016/j.agrformet.2009.02.007.

B. D. W. Roberts, T. I. Dowling, and J. Walker. FLAG : A Fuzzy Landscape Analysis GIS Method for Dryland Salinity Assessment. Technical Report 8, CSIRO Water and Land, Canberra, Australia, 1997. URL `http://www.clw.csiro.au/publications/technical97/tr8-97.pdf`.

J. W. Roche, Q. Ma, J. Rungee, and R. C. Bales. Evapotranspiration mapping for forest management in California's Sierra Nevada. *Frontiers for Global Change*, 2020. doi: 10.3389/ffgc.2020.00069.

M. L. Roderick and G. D. Farquhar. A simple framework for relating variations in runoff to variations in climatic conditions and catchment properties. *Water Resources Research*, 47 (6):1–11, 2011. ISSN 00431397. doi: 10.1029/2010WR009826.

G. H. Roe. Orographic Precipitation. *Annual Review of Earth and Planetary Sciences*, 33 (1):645–671, 2005. doi: 10.1146/annurev.earth.33.092203.122541.

G. H. Roe and M. B. Baker. Microphysical and Geometrical Controls on the Pattern of Orographic Precipitation. *Journal of the Atmospheric Sciences*, 63(3):861–880, 2006. doi: 10.1175/JAS3619.1.

O. Rössler, P. Froidevaux, U. Börst, R. Rickli, O. Martius, and R. Weingartner. Retrospective analysis of a nonforecasted rain-on-snow flood in the Alps-A matter of model limitations or unpredictable nature? *Hydrology and Earth System Sciences*, 18(6):2265–2285, 2014. ISSN 16077938. doi: 10.5194/hess-18-2265-2014.

J. Rungee, R. Bales, and M. Goulden. Evapotranspiration response to multiyear dry periods in the semiarid western United States. *Hydrological Processes*, 33(2):182–194, 2019. ISSN 10991085. doi: 10.1002/hyp.13322.

J. J. Rutz, W. James Steenburgh, and F. Martin Ralph. Climatological characteristics of atmospheric rivers and their inland penetration over the western united states. *Monthly Weather Review*, 142(2):905–921, 2014. ISSN 00270644. doi: 10.1175/MWR-D-13-00168.1.

M. Safeeq, G. E. Grant, S. L. Lewis, and B. Staab. Predicting landscape sensitivity to present and future floods in the Pacific Northwest, USA. *Hydrological Processes*, 29(26): 5337–5353, 2015. doi: 10.1002/hyp.10553.

M. Saft, A. W. Western, L. Zhang, M. C. Peel, and N. J. Potter. The influence of multiyear drought on the annual rainfall-runoff relationship: An Australian perspective. *Water Resources Research*, 51:2444–2463, 2016. ISSN 1093-474X. doi: :10.1002/2014WR015348.

P. C. Saksa, M. H. Conklin, J. Battles, C. Tague, and R. C. Bales. Forest thinning impacts on the water balance of Sierra Nevada mixed-conifer headwater basins. *Water Resources Research*, 53, 2017. doi: 10.1002/2016WR019240.

L. Samaniego, R. Kumar, and S. Attinger. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, 46(5), 2010. doi: 10.1029/2008WR007327.

L. Schmidt, F. Heße, S. Attinger, and R. Kumar. Challenges in Applying Machine Learning Models for Hydrological Inference: A Case Study for Flooding Events Across Germany. *Water Resources Research*, 2020. ISSN 19447973. doi: 10.1029/2019WR025924.

J. Seel, A. Mills, R. Wiser, S. Deb, A. Asokkumar, M. Hassanzadeh, and A. Aarabali. Impacts of High Variable Renewable Energy Futures on Wholesale Electricity Prices, and on Electric-Sector Decision Making. Technical report, Lawrence Berkeley National Laboratory, Berkeley, 2018. URL `https://emp.lbl.gov/publications/impacts-high-variable-renewable`.

Q. Shao, A. Traylen, and L. Zhang. Nonparametric method for estimating the effects of climatic and catchment characteristics on mean annual evapotranspiration. *Water Resources Research*, 48(3):1–13, 2012. ISSN 00431397. doi: 10.1029/2010WR009610.

C. Shen. A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists, 2018. ISSN 19447973.

Q. Shen, Z. Cong, and H. Lei. Evaluating the impact of climate and underlying surface change on runoff within the Budyko framework: A study across 224 catchments in China. *Journal of Hydrology*, 554:251–262, 2017. ISSN 00221694. doi: 10.1016/j.jhydrol.2017.09.023.

F. Silvestro, A. Parodi, L. Campo, and L. Ferraris. Analysis of the streamflow extremes and long-term water balance in the Liguria region of Italy using a cloud-permitting grid spacing reanalysis dataset. *Hydrology and Earth System Sciences*, 22(10):5403–5426, 2018. ISSN 16077938. doi: 10.5194/hess-22-5403-2018.

M. Sivapalan and J. D. Kalma. Scale problems in hydrology: Contributions of the Robertson workshop. *Hydrological Processes*, 9(3-4):243–250, 1995. ISSN 10991085. doi: 10.1002/hyp.3360090304.

T. H. Snelder, T. Datry, N. Lamouroux, S. T. Larned, E. Sauquet, H. Pella, and C. Catalogne. Regionalization of patterns of flow intermittence from gauging station records. *Hydrology and Earth System Sciences*, 17(7):2685–2699, 2013. ISSN 10275606. doi: 10.5194/hess-17-2685-2013.

Soil Conservation Service. National Engineering Handbook, Section 4, Hydrology. Technical report, U.S. Department of Agriculture, Washington, D.C., 1972.

Soil Survey Staff Natural Resources Conservation Service. Web Soil Survey, 2019. URL `http://websoilsurvey.nrcs.usda.gov/`.

M. Sturm and J. Holmgren. A seasonal snow cover classification system for local to global applications. *Journal of Climate*, 8(5):1261–1282, 1995. doi: 10.1175/1520-0442(1995)008⟨1261:ASSCCS⟩2.0.CO;2.

G. K. Summerell, J. Vaze, N. K. Tuteja, R. B. Grayson, G. Beale, and T. I. Dowling. Delineating the major landforms of catchments using an objective hydrological terrain analysis method. *Water Resources Research*, 41(12):1–12, 2005. ISSN 00431397. doi: 10.1029/2005WR004013.

D. L. Swain, B. Langenbrunner, J. D. Neelin, and A. Hall. Increasing precipitation volatility in twenty-first-century California. *Nature Climate Change*, 8(5):427–433, 2018. ISSN 17586798. doi: 10.1038/s41558-018-0140-y.

C. Tague and L. Band. RHESSys: Regional Hydro-Ecologic Simulation System—An object-oriented approach to spatially distributed modeling of carbon, water, and nutrient cycling. *Earth Interactions*, 8(19):1–42, 2004. doi: 10.1175/1087-3562(2004)8⟨1:RRHSSO⟩2.0.CO;2.

C. Tague and G. E. Grant. Groundwater dynamics mediate low-flow response to global warming in snow-dominated alpine regions. *Water Resources Research*, 45(7):1–12, 2009. ISSN 00431397. doi: 10.1029/2008WR007179.

A. D. Teshager, P. W. Gassman, S. Secchi, J. T. Schoof, and G. Misgna. Modeling Agricultural Watersheds with the Soil and Water Assessment Tool (SWAT): Calibration and Validation with a Novel Procedure for Spatially Explicit HRUs. *Environmental Management*, 57(4):894–911, 2016. ISSN 14321009. doi: 10.1007/s00267-015-0636-4.

A. J. Teuling, A. F. Van Loon, S. I. Seneviratne, I. Lehner, M. Aubinet, B. Heinesch, C. Bernhofer, T. Grünwald, H. Prasse, and U. Spank. Evapotranspiration amplifies European summer drought. *Geophysical Research Letters*, 40(10):2071–2075, 2013. ISSN 00948276. doi: 10.1002/grl.50495.

H. A. Thomas. Improved methods for national water assessment, water resources contract: WR15249270. Technical report, U.S. Geological Survey, 1981. URL `https://pubs.er.usgs.gov/publication/70046351`.

Q. Tian, D. Wang, D. Li, L. Huang, M. Wang, C. Liao, and F. Liu. Variation of soil carbon accumulation across a topographic gradient in a humid subtropical mountain forest. *Biogeochemistry*, 2020a. ISSN 1573-515X. doi: 10.1007/s10533-020-00679-2.

W. Tian, P. Bai, K. Wang, K. Liang, and C. Liu. Simulating the change of precipitation-runoff relationship during drought years in the eastern monsoon region of China. *Science of the Total Environment*, 723:138172, 2020b. ISSN 18791026. doi: 10.1016/j.scitotenv.2020.138172.

W. T. Tinkham, A. M. Smith, H. P. Marshall, T. E. Link, M. J. Falkowski, and A. H. Winstral. Quantifying spatial distribution of snow depth errors from LiDAR using Random Forest. *Remote Sensing of Environment*, 141:105–115, 2014. ISSN 00344257. doi: 10.1016/j.rse.2013.10.021.

Q. Q. Tran, J. De Niel, and P. Willems. Spatially Distributed Conceptual Hydrological Model Building: A Generic Top-Down Approach Starting From Lumped Models. *Water Resources Research*, 54(10):8064–8085, 2018. ISSN 19447973. doi: 10.1029/2018WR023566.

K. E. Trenberth, A. Dai, G. van der Schrier, P. D. Jones, J. Barichivich, K. R. Briffa, and J. Sheffield. Global warming and changes in drought. *Nature Climate Change*, 4(1):17–22, 2014. ISSN 1758-6798. doi: 10.1038/nclimate2067.

P. A. Troch, T. Lahmers, A. Meira, R. Mukherjee, J. W. Pedersen, T. Roy, and R. Valdés-Pineda. Catchment coevolution: A useful framework for improving predictions of hydrological change? *Water Resources Research*, 51(7):4903–4922, 2015. ISSN 19447973. doi: 10.1002/2015WR017032.

U.S. Geological Survey. Landsat–Earth observation satellites (ver. 1.2, April 2020). Technical report, U.S. Geological Survey, 2016. URL `https://doi.org/10.3133/fs20153081`.

A. Valéry, V. Andréassian, and C. Perrin. 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 1 – Comparison of six snow accounting routines on 380 catchments. *Journal of Hydrology*, 517:1166–1175, 2014. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2014.04.059.

A. F. Van Loon. Hydrological drought explained. *Wiley Interdisciplinary Reviews: Water*, 2(4):359–392, 2015. doi: 10.1002/wat2.1085.

A. Varhola, N. C. Coops, M. Weiler, and R. D. Moore. Forest canopy effects on snow accumulation and ablation: An integrative review of empirical results. *Journal of Hydrology*, 392:219–233, 2010. doi: 10.1016/j.jhydrol.2010.08.009.

J. Vaze, D. A. Post, F. H. S. Chiew, J.-M. Perraud, N. R. Viney, and J. Teng. Climate non-stationarity – Validity of calibrated rainfall–runoff models for use in climate change studies. *Journal of Hydrology*, 394(3):447–457, 2010. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2010.09.018.

W. Veatch, P. Brooks, J. Gustafson, and N. P. Molotch. Quantifying the effects of forest canopy cover on net snow accumulation at a continental, mid-latitude site. *Ecohydrology*, 2:115–128, 2009. doi: 10.1002/eco.45.

R. Viger and G. H. Leavesley. The GIS Weasel User's Manual. In *Techniques and Methods Book 6*, chapter B4, page 201. U.S. Geological Survey, 2007. URL `https://pubs.usgs.gov/tm/2007/06B04/`.

D. Viviroli, H. H. Dürr, B. Messerli, M. Meybeck, and R. Weingartner. Mountains of the world, water towers for humanity: Typology, mapping, and global significance. *Water Resources Research*, 43(7):1–13, 2007a. ISSN 00431397. doi: 10.1029/2006WR005653.

D. Viviroli, J. Gurtz, and M. Zappa. The Hydrological Modelling System PREVAH. Part I - Overview and Selected Applications. Technical report, Institute of Geography, University of Berne, Berne, Switzerland, 2007b. URL `https://dviviroli.github.io/documents/Viviroli{_}et{_}al{_}2007{_}GB{_}P40{_}I.pdf`.

D. Wang and N. Alimohammadi. Responses of annual runoff, evaporation, and storage change to climate variability at the watershed scale. *Water Resources Research*, 48(5), 2012. ISSN 00431397. doi: 10.1029/2011WR011444.

D. Wang and Y. Tang. A one-parameter Budyko model for water balance captures emergent behavior in darwinian hydrologic models. *Geophysical Research Letters*, 41:4569–4577, 2014. doi: 10.1002/2014GL060509.

Z. Wang, C. Lai, X. Chen, B. Yang, S. Zhao, and X. Bai. Flood hazard risk assessment model based on random forest. *Journal of Hydrology*, 527:1130–1141, 2015. ISSN 00221694. doi: 10.1016/j.jhydrol.2015.06.008.

N. E. Wayand, J. D. Lundquist, and M. P. Clark. Modeling the influence of hypsometry, vegetation, and stormenergy on snowmelt contributions to basins during rain-on-snow floods. *Water Resources Research*, 51:8551–8569, 2015. doi: 10.1002/2014WR016576.

N. Wever, T. Jonas, C. Fierz, and M. Lehning. Model simulations of the modulating effect of the snow cover in a rain-on-snow event. *Hydrology and Earth System Sciences*, 18(11): 4657–4669, 2014. ISSN 16077938. doi: 10.5194/hess-18-4657-2014.

W. Wilcke, J. Boy, R. Goller, K. Fleischbein, C. Valarezo, and W. Zech. Effect of topography on soil fertility and water flow in an Ecuadorian lower montane forest. In F. N. Scatena, L. A. Bruijnzeel, and L. S. Hamilton, editors, *Tropical Montane Cloud Forests: Science for Conservation and Management*, International Hydrology Series, pages 402–409. Cambridge University Press, Cambridge, 2011. ISBN 9780521760355. doi: 10.1017/CBO9780511778384.045.

A. P. Williams, E. R. Cook, J. E. Smerdon, B. I. Cook, J. T. Abatzoglou, K. Bolles, S. H. Baek, A. M. Badger, and B. Livneh. Large contribution from anthropogenic warming to an emerging North American megadrought. *Science*, 368(6488):314–318, 2020. ISSN 0036-8075. doi: 10.1126/science.aaz9600.

E. F. Wood, M. Sivapalan, K. Beven, and L. Band. Effects of spatial variability and scale with implications to hydrologic modeling. *Journal of Hydrology*, 102(1-4):29–47, 1988. ISSN 00221694. doi: 10.1016/0022-1694(88)90090-X.

C. A. Woodhouse, D. M. Meko, G. M. MacDonald, D. W. Stahle, and E. R. Cook. A 1,200-year perspective of 21st century drought in southwestern North America. *Proceedings of the National Academy of Sciences of the United States of America*, 107(50):21283–21288, 2010. ISSN 00278424. doi: 10.1073/pnas.0911197107.

S. Würzer and T. Jonas. Spatio-temporal aspects of snowpack runoff formation during rain-on-snow. *Hydrological Processes*, 2018. ISSN 08856087. doi: 10.1002/hyp.13240.

S. Würzer, T. Jonas, N. Wever, and M. Lehning. Influence of Initial Snowpack Properties on Runoff Formation during Rain-on-Snow Events. *Journal of Hydrometeorology*, 17(6): 1801–1815, 2016. ISSN 1525-755X. doi: 10.1175/JHM-D-15-0181.1.

S. Würzer, N. Wever, R. Juras, M. Lehning, and T. Jonas. Modelling liquid water transport in snow under rain-on-snow conditions – considering preferential flow. *Hydrology and Earth System Sciences*, 21(April):1741–1756, 2017. doi: 10.5194/hess-21-1741-2017.

D. Yang, F. Sun, Z. Liu, Z. Cong, G. Ni, and Z. Lei. Analyzing spatial and temporal variability of annual water-energy balance in nonhumid regions of China using the Budyko hypothesis. *Water Resources Research*, 43(4):1–12, 2007. ISSN 00431397. doi: 10.1029/2006WR005224.

D. Yang, W. Shao, P. J. Yeh, H. Yang, S. Kanae, and T. Oki. Impact of vegetation coverage on regional water balance in the nonhumid regions of China. *Water Resources Research*, 45(7):1–13, 2009. ISSN 00431397. doi: 10.1029/2008WR006948.

K. L. Young, M. K. Woo, and S. A. Edlund. Influence of local topography, soils, and vegetation on microclimate and hydrology at a High Arctic site, Ellesmere Island, Canada. *Arctic and Alpine Research*, 29(3):270–284, 1997. ISSN 00040851. doi: 10.2307/1552141.

L. Zhang, W. R. Dawes, and G. R. Walker. Response of mean annual evapotranspiration to vegetation changes at catchment scale. *Water Resources Research*, 37(3):701–708, 2001. ISSN 00431397. doi: 10.1029/2000WR900325.

L. Zhang, K. Hickel, W. R. Dawes, F. H. Chiew, A. W. Western, and P. R. Briggs. A rational function approach for estimating mean annual evapotranspiration. *Water Resources Research*, 40(2):1–14, 2004. ISSN 00431397. doi: 10.1029/2003WR002710.

L. Zhang, N. Potter, K. Hickel, Y. Zhang, and Q. Shao. Water balance modeling over variable time scales based on the Budyko framework - Model development and testing. *Journal of Hydrology*, 360(1-4):117–131, 2008. ISSN 00221694. doi: 10.1016/j.jhydrol.2008.07.021.

S. Zhang, X. Zhang, T. Huffman, X. Liu, and J. Yang. Influence of topography and land management on soil nutrients variability in Northeast China. *Nutrient Cycling in Agroecosystems*, 89(3):427–438, 2011. ISSN 1573-0867. doi: 10.1007/s10705-010-9406-0.

S. Zhang, H. Yang, D. Yang, and A. W. Jayawardena. Quantifying the effect of vegetation change on the regional water balance within the Budyko framework. *Geophysical Research Letters*, 43(3):1140–1148, 2016. ISSN 19448007. doi: 10.1002/2015GL066952.

Z. Zhang, S. Glaser, R. Bales, M. Conklin, R. Rice, and D. Marks. Insights into mountain precipitation and snowpack from a basin-scale wireless-sensor network. *Water Resources Research*, 53(8):6626–6641, 2017a. ISSN 19447973. doi: 10.1002/2016WR018825.

Z. Zhang, S. D. Glaser, R. C. Bales, M. Conklin, R. Rice, and D. G. Marks. Technical report: The design and evaluation of a basin-scale wireless sensor network for mountain hydrology. *Water Resources Research*, 2017b. ISSN 19447973. doi: 10.1002/2016WR019619.

Z. Zheng, Q. Ma, K. Qian, and R. C. Bales. Canopy effects on snow accumulation: Observations from Lidar, canonical-view photos, and continuous ground measurements from sensor networks. *Remote Sensing*, 10(11), 2018. ISSN 20724292. doi: 10.3390/rs10111769.

Appendix A

Supplementary Information: Chapter II

A.1 Data collection & cleaning

For LiDAR-derived data, rasters were computed from the raw point clouds (∼1.4 billion classified points and a point density of 11.65 pts/m^2) using a binning interpolation type with average cell assignment and a linear void-fill method. Topographic information (elevation, slope, and aspect) was obtained by converting the point cloud to a digital terrain model (DTM) to derive elevation and then applying standard topographic tools to derive slope and aspect. LAI was computed by using the LiDAR360 toolbox available at `https://greenvalleyintl.com/software/lidar360/` (see Chen et al., 1992; Richardson et al., 2009, for calculation details). Vegetation height was obtained by filtering the cloud for first returns, which were converted to a Digital Surface Model (DSM). Subtracting the DTM from the DSM was assumed as a proxy of vegetation height. Open fraction was estimated as the ratio between a raster of ground-pulse counts and another one of all-pulse counts (Zheng et al., 2018). LiDAR data are available at `https://doi.org/10.5069/G9BP00QB`.

The Wolverton wireless-sensor network data were supplemented with relative humidity values from two weather stations installed in the same catchment (also described in Bales et al., 2018b), and precipitation was taken from the Giant Forest meteorological station (available through the California Data Exchange Center, `cdec.water.ca.gov`, station code GNF). Certain periods of the precipitation record from the GNF station in Water Year 2011 were deemed unreliable when visually screened. These periods were discarded and filled by matching the cumulative precipitation pattern from the nearby Big Meadows (BIM) station.

At UPM, LOM, and P301, relative humidity, solar radiation, and precipitation data were provided by the Kings River Experimental Watersheds (KREW) meteorological sites (O'Geen et al., 2018, `https://www.fs.fed.us/psw/topics/water/kingsriver/`). Hourly KREW level 2 (i.e., quality controlled and gap filled) precipitation data from the Lower Providence (LP) and Upper Providence (UP) sites were not available and required additional processing. First, hourly precipitation measurements from KREW gauges (Hunsaker and Safeeq, 2018; O'Geen et al., 2018) were processed for outliers and range control. Second, level 2 daily precipitation data were obtained from the US Forest Service data archive

(https://www.fs.usda.gov/rds/archive/catalog/RDS-2018-0028). We then proportionally adjusted the hourly precipitation values for the day to ensure that the daily totals between the two datasets are identical. An independent quality control on hourly precipitation measurement was not possible due to a lack of reference gauges in the proximity. Atmospheric-river classification was derived based on 3-hour MERRA classifications (http://www.inscc.utah.edu/~rutz/ar_catalogs/merra_0.5/) at the location closest to each WSN. Hourly values were forward-filled from each three-hour timestamp.

A.1.1 Alternative correlated features

A Random Forest model trained on the alternative correlated features performed comparably to the baseline model.

Table A.1. Alternative correlated variables[†]

Climate & snow
Initial snow depth
Storm length
Total precipitation
Average temperature
Change in air temperature
Average wind speed
Average dewpoint temperature*
Standard deviation of dewpoint temp
Coefficient of variation of dewpoint temp
Autocorrelation of dewpoint temp
Maximum duration of stable dewpoint temp
Atmospheric river or not
Topography & vegetation
Aspect (10 m)
Aspect (1 m)
Canopy height (10 m)
Canopy height (1 m)
Elevation (10 m)
Elevation (1 m)*
Leaf Area Index (LAI) (10 m)
Open fraction (10 m)
Open fraction (1 m)
Sheltering coefficient
Slope (10 m)
Slope (1 m)*
Event timing
Day of water year (DOWY)
Number of previous events
Start time

[†] In this model, average dewpoint temperature, elevation (1 m), and slope (1 m) were dropped instead of average air temperature, elevation (10 m), and slope (10 m) as indicated by asterisks.

Table A.2. Summary of data sources[†]

Site	Snow depth	Temperature	Precipitation	Relative humidity	Wind speed
LOM	WSN	WSN	KREW: LP	KREW: LP	KREW: LP
UPM	WSN	WSN	KREW: UP	KREW: UP	KREW: UP
P301	WSN	WSN	KREW: UP	KREW: UP	KREW: UP
SK1	WSN	WSN	CDEC: GNF	WTN	WTN
SK2	WSN	WSN	CDEC: GNF	WTN	WTN
SK3	WSN	WSN	CDEC: GNF	PTM (2008–12), WTN (2012–16)	PTM
SK4	WSN	WSN	CDEC: GNF	PTM (2008–12), WTN (2012–16)	PTM

[†] Note on abbreviations: WSN indicates the data were measured at each node of the wireless-sensor network and CDEC refers to the California Data Exchange Center with site codes WTN (Wolverton), PTM (Panther Meadows), and GNF (Giant Forest). KREW refers to the Kings River Experimental Watershed with LP and UP referring specifically to the KREW meteorological stations co-located with the LOM and UPM networks, respectively.

Table A.3. Validation performance of Random Forest model trained with alternative correlated features

Random Forest model	RMSE, cm	Absolute bias, cm	Kling-Gupta efficiency
Baseline	3.91	2.98	0.72
Alternative correlated features	3.91	3.01	0.77

A.1.2 Correlation

Figures A.1 to A.5 plot the correlations between snow depth and all continuous features used to drive the Random Forest model.

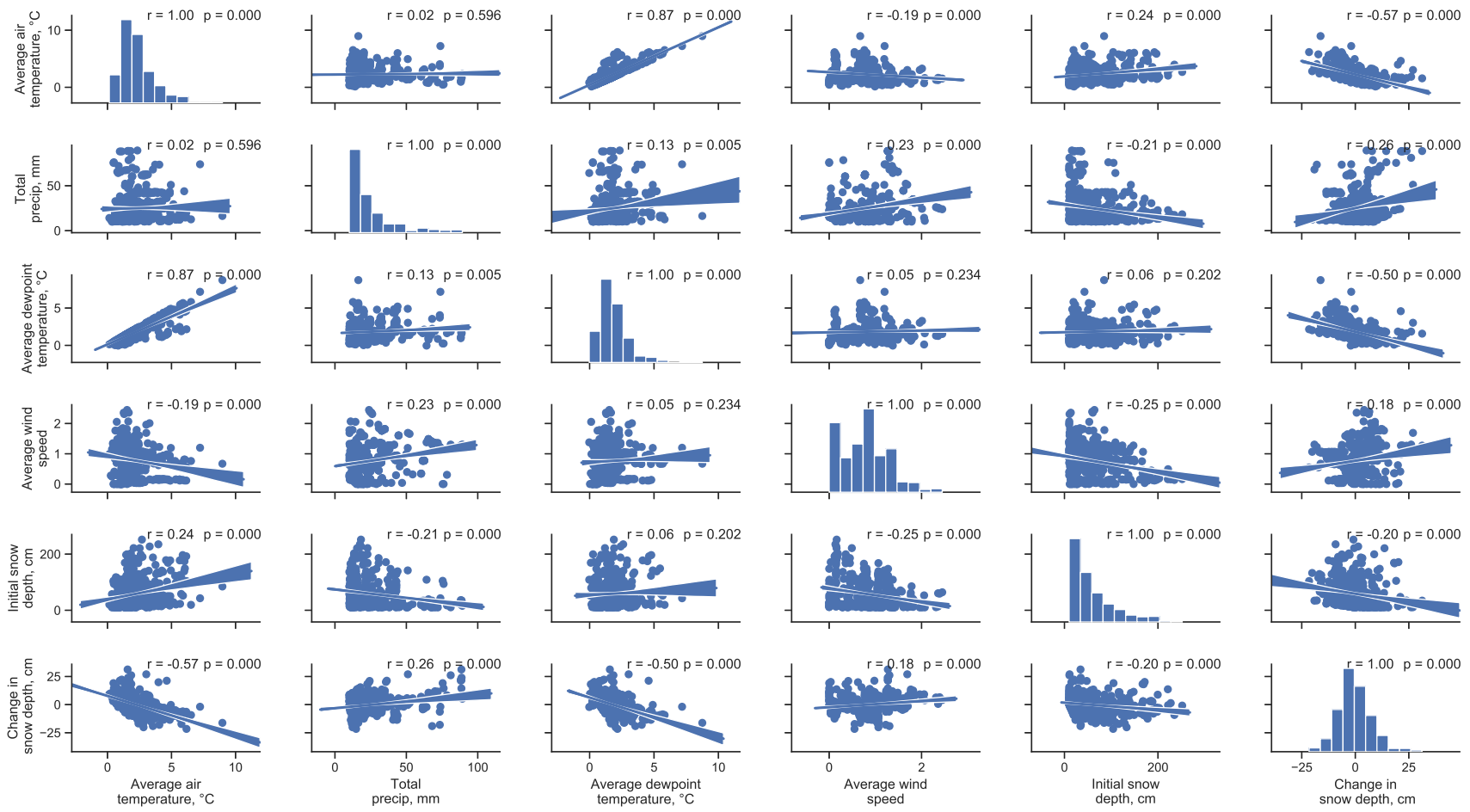


Figure A.1: Correlation-distribution plots between change in snow depth and climate features. Plots on the diagonal show the distribution of the indicated variable. Off diagonals show the correlation between two features, with correlation coefficients and p-values reported.

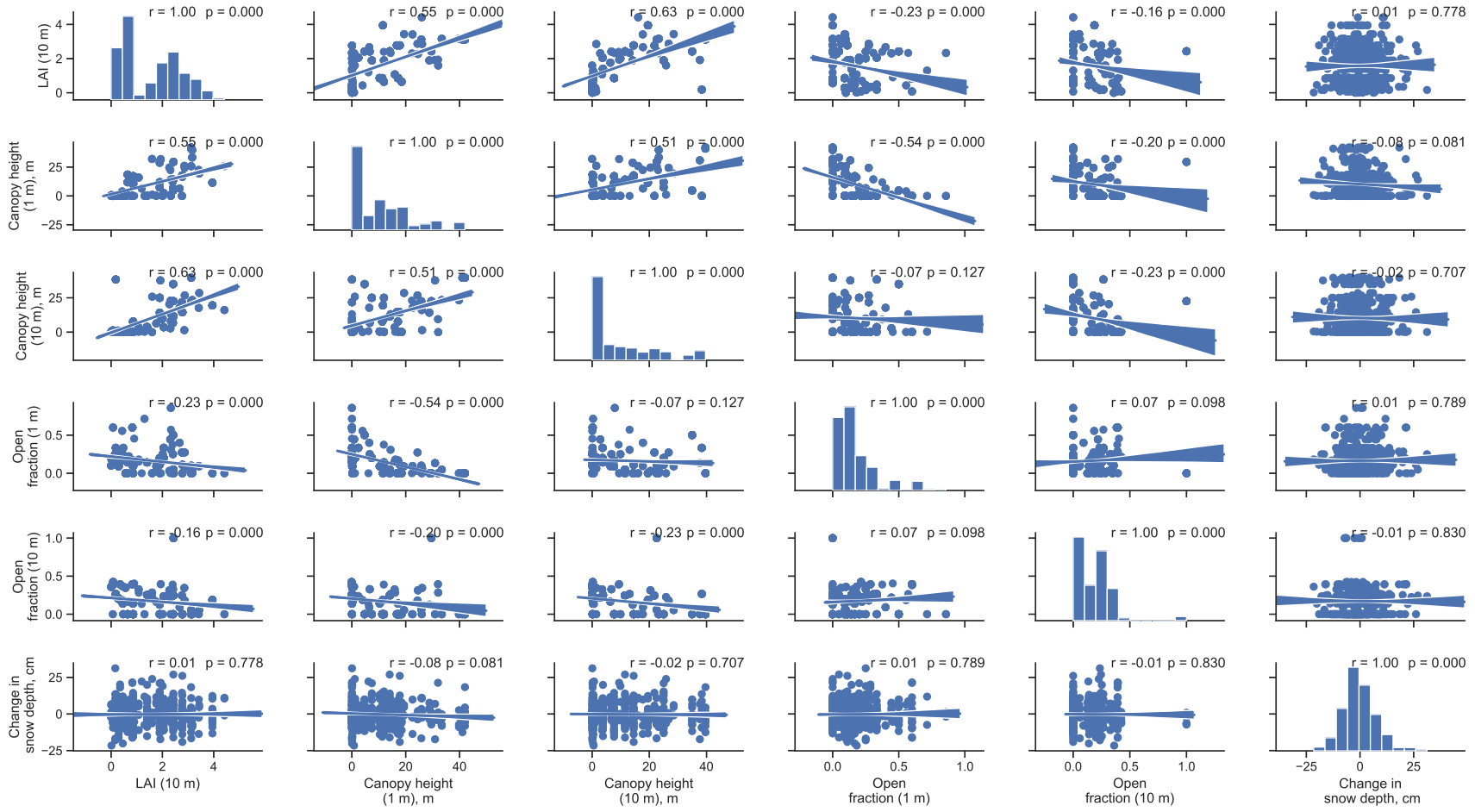


Figure A.2: Correlation-distribution plots between change in snow depth and vegetation features. Plots on the diagonal show the distribution of the indicated variable. Off diagonals show the correlation between two features, with correlation coefficients and p-values reported.

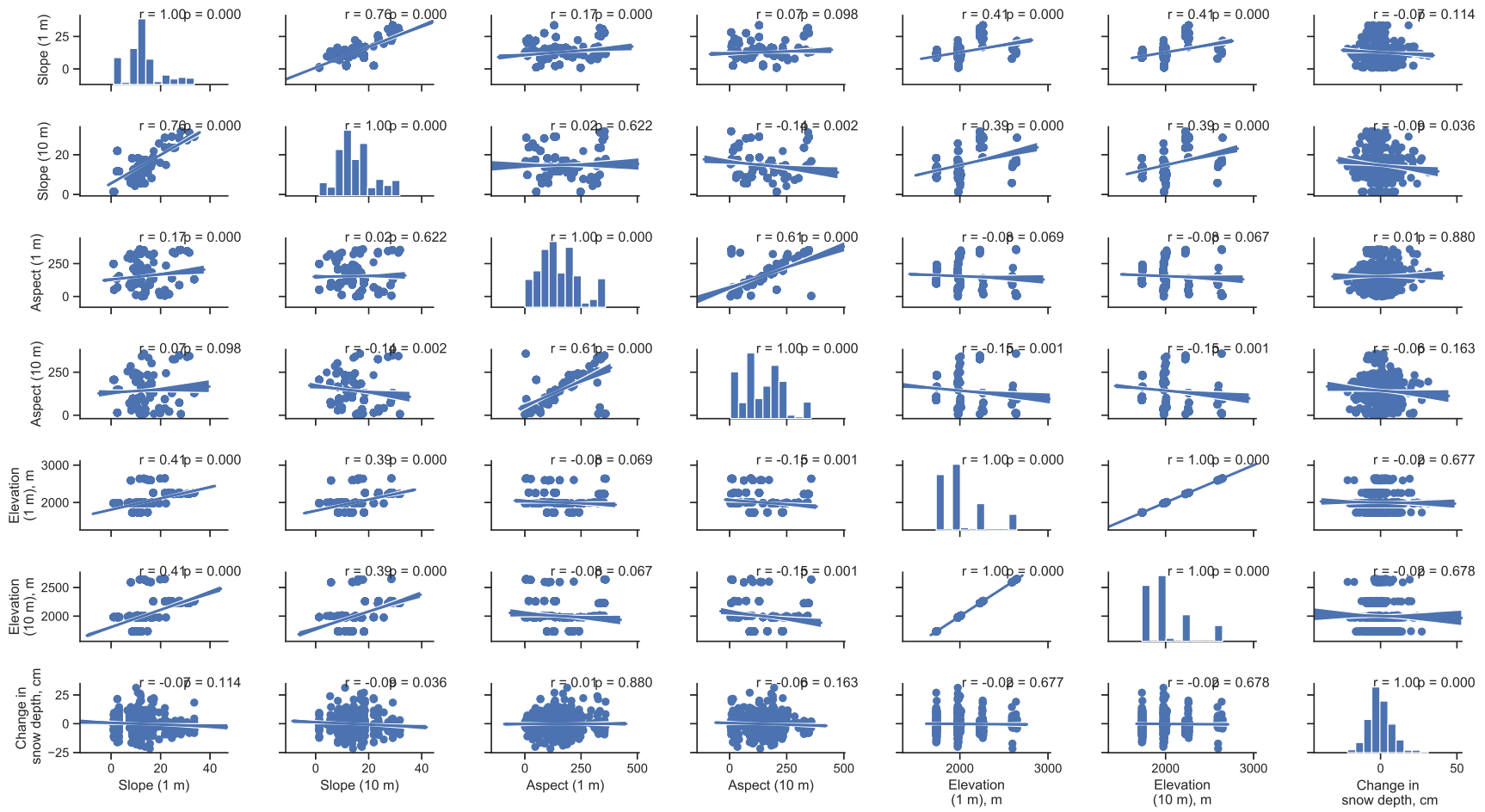


Figure A.3: Correlation-distribution plots between change in snow depth and topographic features. Plots on the diagonal show the distribution of the indicated variable. Off diagonals show the correlation between two features, with correlation coefficients and p-values reported.

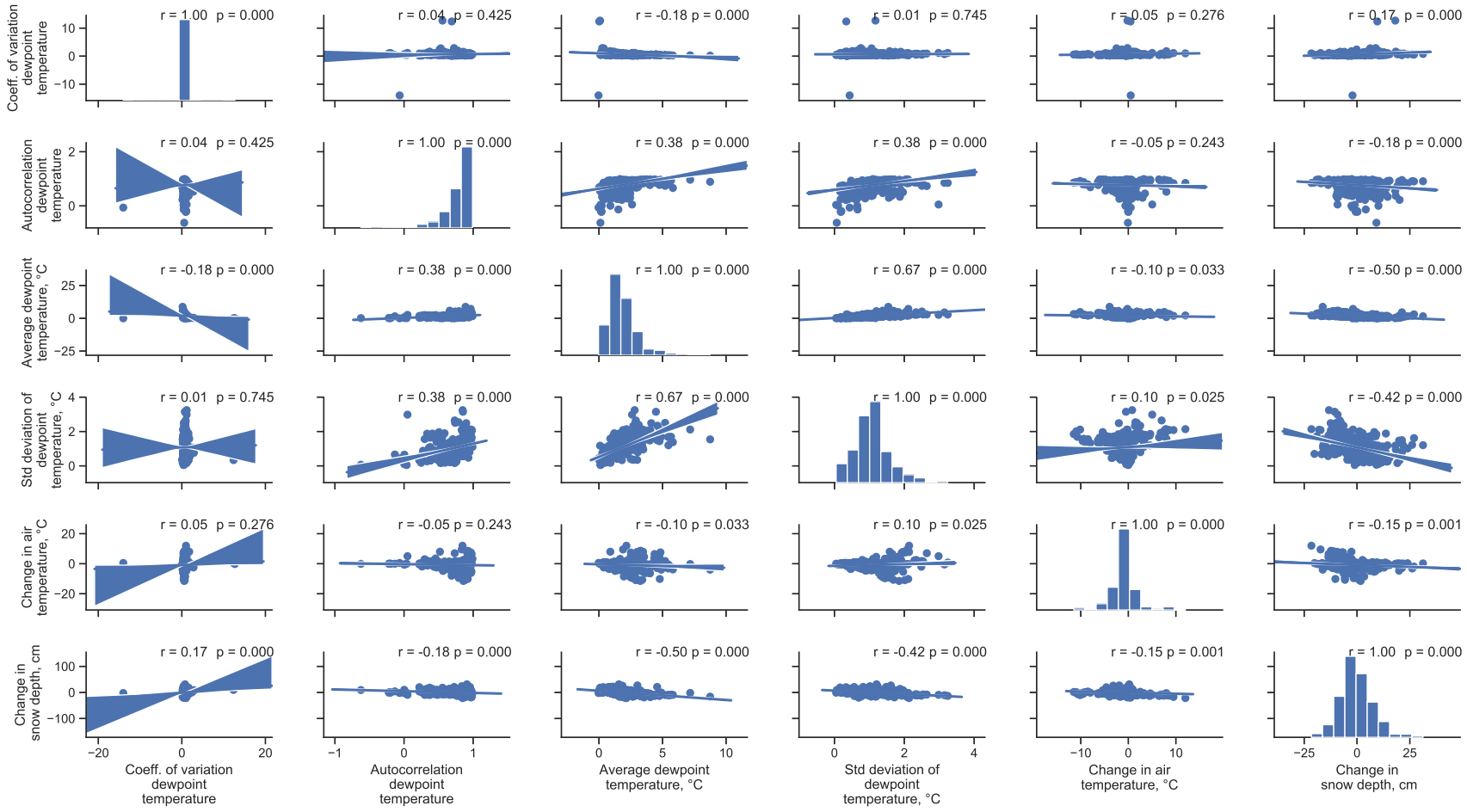


Figure A.4: Correlation-distribution plots between change in snow depth and dewpoint-temperature-derived features. Plots on the diagonal show the distribution of the indicated variable. Off diagonals show the correlation between two features, with correlation coefficients and p-values reported.

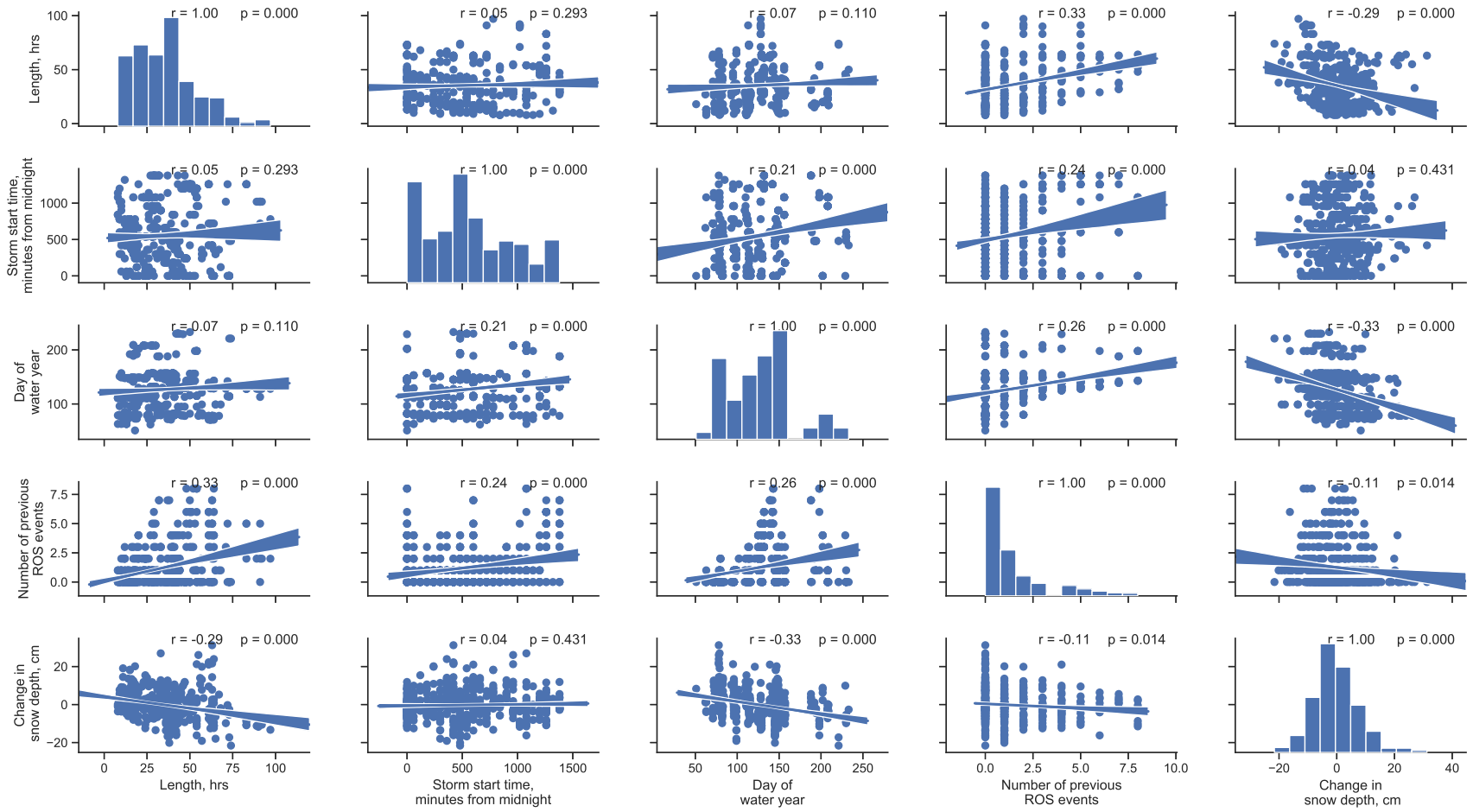


Figure A.5: Correlation-distribution plots between change in snow depth and timing features. Plots on the diagonal show the distribution of the indicated variable. Off diagonals show the correlation between two features, with correlation coefficients and p-values reported.

A.1.3 Climatology of ROS events by site

ROS storms were concentrated in the period between January and March (69% of events), when the Sierra most regularly sees a combination of deep, continuous snowpack and precipitation events. 20% of events occurred between October and December and 11% between April and June. The average storm length was comparable across these three periods (32.1, 36.9, and 32.9 hours, respectively), and 73% of all events lasted at least 20 hours.

Of the climate characteristics, precipitation, wind speed, and average air temperature during storms showed more site-to-site variations than did average dewpoint temperature or change in snow depth (Figure A.6). Median dewpoint temperature range during events showed less variation across sites than air temperature (1.4 to 1.8 °C versus 1.7 to 2.5 °C, respectively). The ranges of the first and third quartiles of dewpoint temperature (0.40 °C and 0.60 °C, respectively) were also less than air temperature (ranging by 0.87 °C and 1.4 °C, respectively). This is expected since dewpoint temperature is considered more indicative of ROS conditions than air temperature (Marks et al., 2013). Wind speed was relatively variable across sites (medians ranging from 0.13 m/s to 0.98 m/s). Despite these differences in weather characteristics during ROS events, resulting changes in snow depth were similar across sites, with mean change ranging from -0.52 to -2.26 cm and the middle quartiles not exceeding +/-10 cm change (Figure A.6). No strong trend with elevation emerged (R^2 = -0.02, p = 0.677), which points to additional controlling factors such as topography and microclimates.

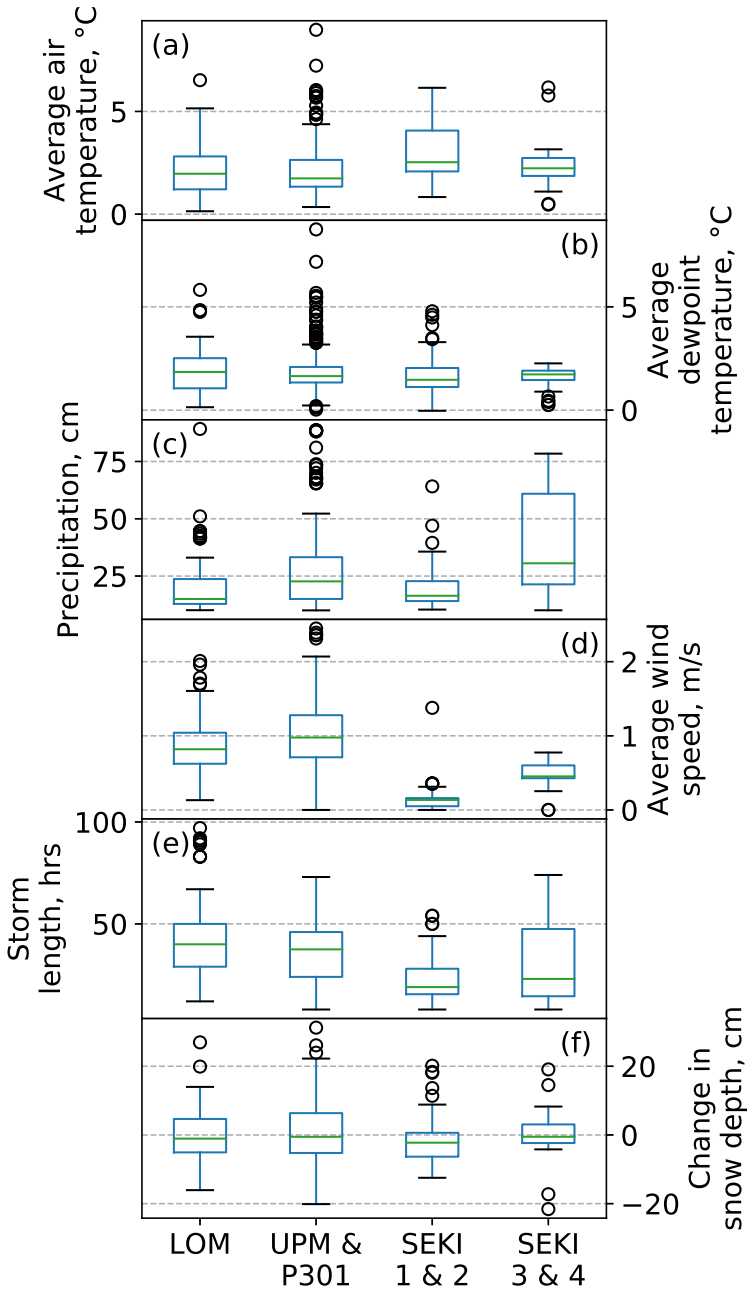


Figure A.6: Distributions of storm characteristics during rain-on-snow events. Solid green lines indicate medians, with box edges indicating quartiles 1 and 3. Whiskers extend to 150% of the interquartile range, with outliers beyond that range indicated by empty circles.

A.1.4 Random Forest performance

Validation results for baseline Random Forest model broken down by quintile (or tercile, if quintile could not be calculated due to duplicate bin edges) of vegetation and timing features. Open fraction at 10 m and number of previous ROS events are not shown, since quintiles/terciles could not be computed.

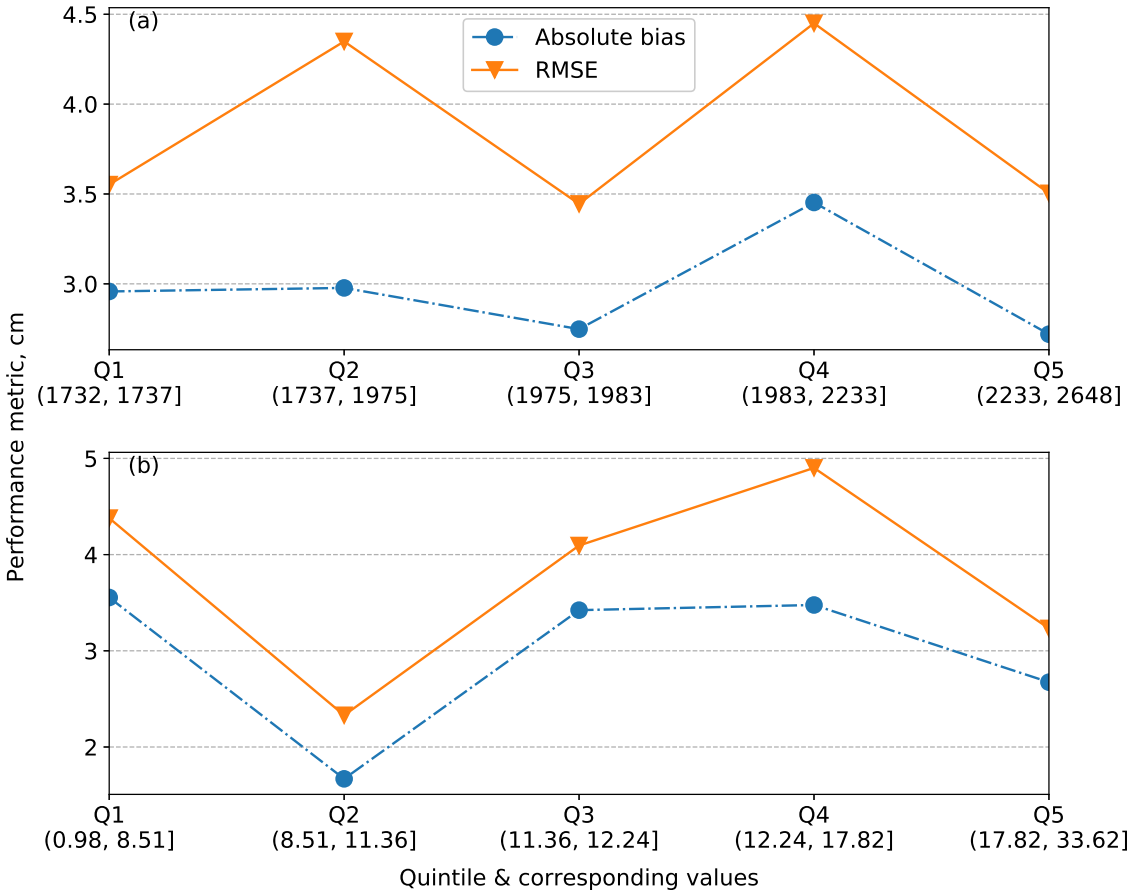


Figure A.7: Model performance calculated by quintile of (a) elevation at 1-m resolution and (b) slope at 1-m resolution.

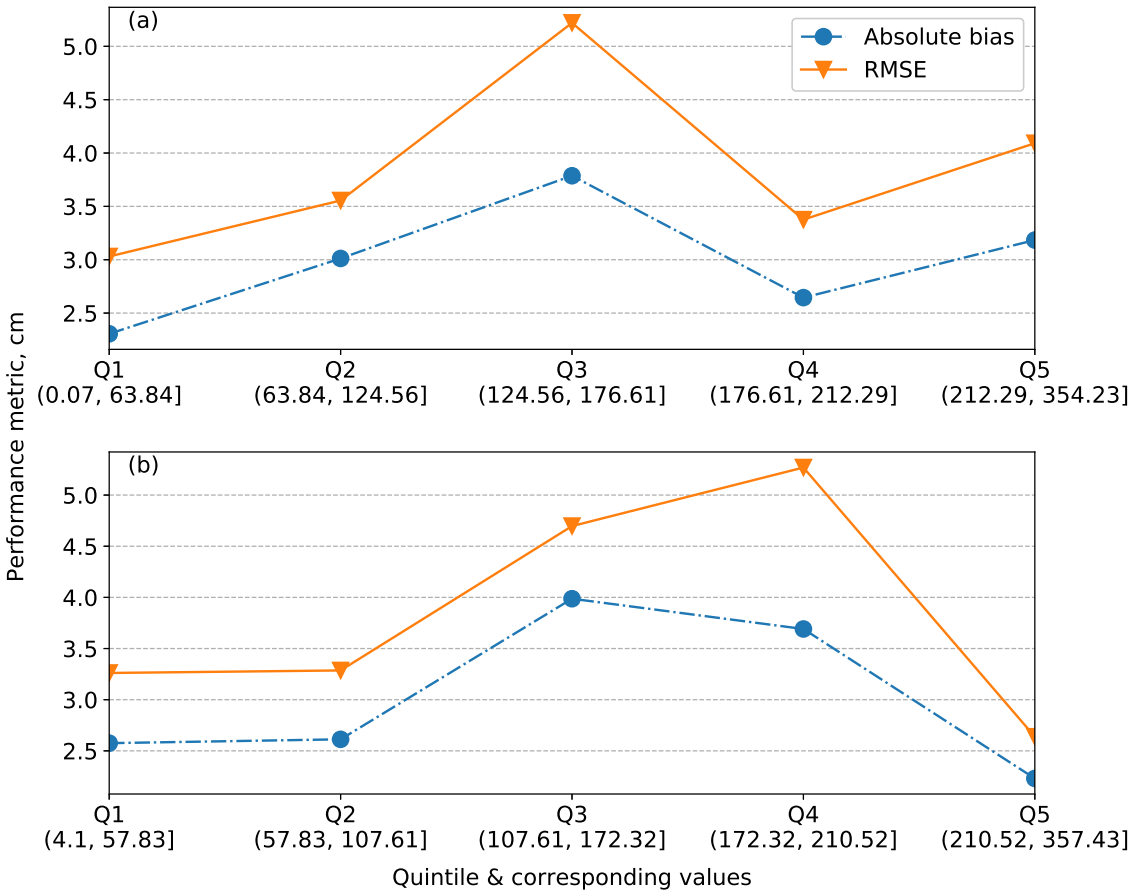


Figure A.8: Model performance calculated by quintile of (a) aspect at 1-m resolution and (b) aspect at 10-m resolution.

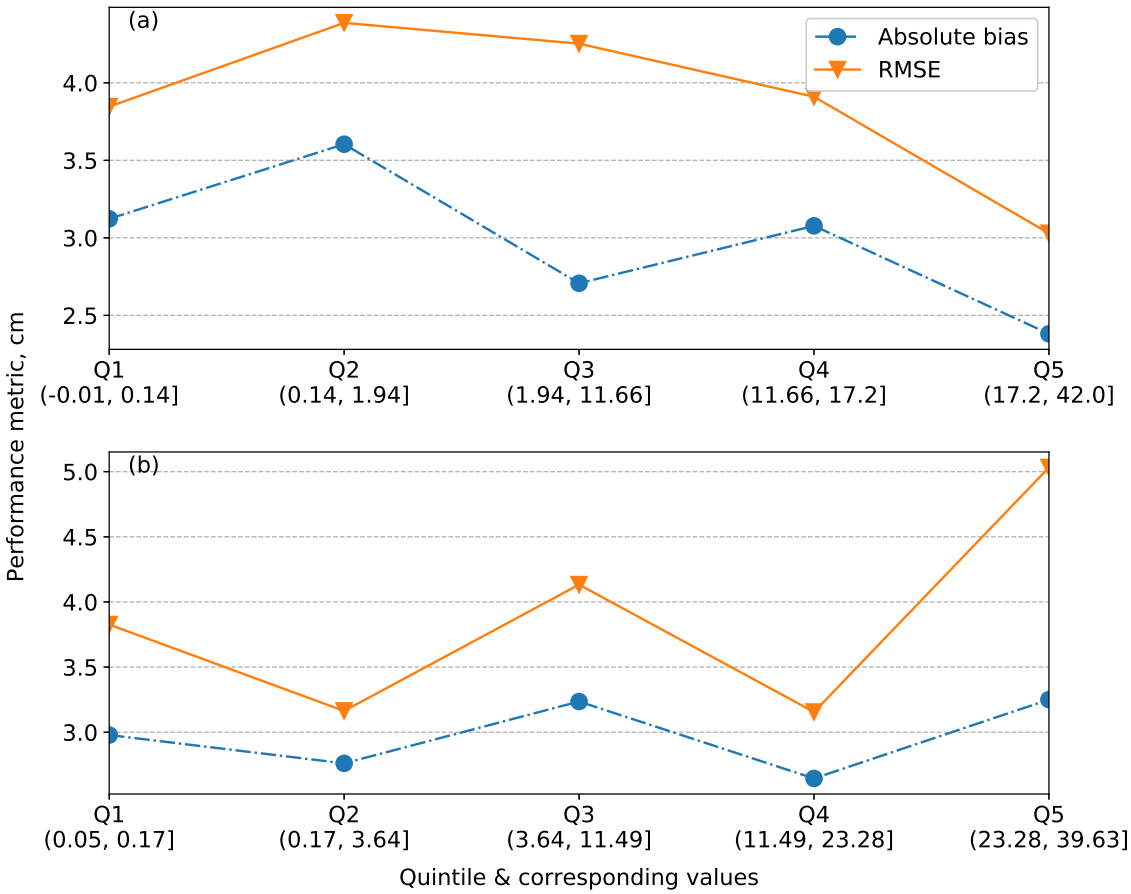


Figure A.9: Model performance calculated by quintile of (a) canopy height at 1-m resolution and (b) canopy height at 10-m resolution.

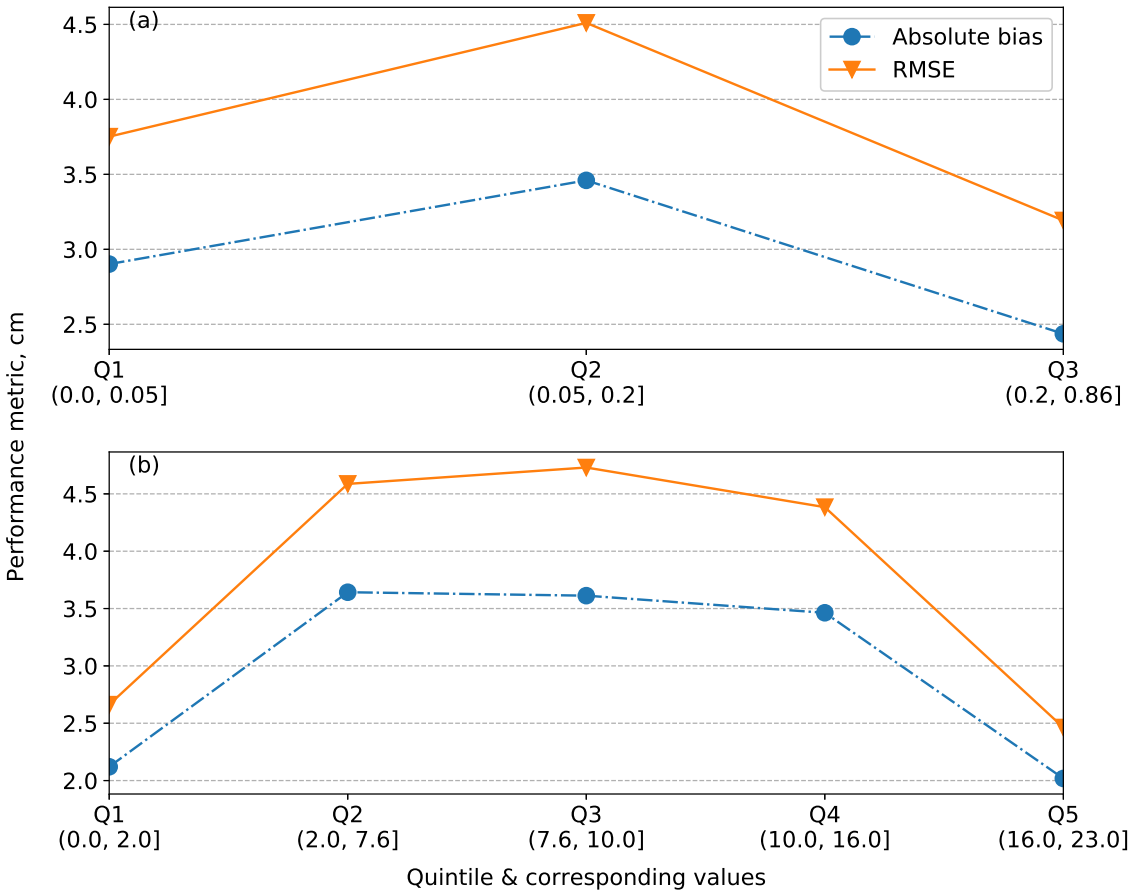


Figure A.10: Model performance calculated by (a) tercile of open fraction at 1-m resolution and (b) storm start time.

The components of the KGE metric for the baseline Random Forest model and the models trained on subsets of data are presented in Table A.4.

Table A.4. Random Forest validation performance

Random Forest model	Correlation coefficient	Ratio of means	Ratio of coeff. of variation	Overall KGE
Baseline	0.88	0.75	1.00	0.72
Climate features only	0.87	0.66	1.15	0.61
Climate + topography & vegetation	0.87	0.72	1.03	0.69
Climate + timing	0.87	0.63	1.21	0.56

Validation performance results of the leave-one-out Random Forest models, each calibrated on all but one water year, are shown in Figure A.11. These "leave-one-out" models assess the transferability of information of snowpack response across different years and year types, a common (albeit somewhat simplistic) validation method in hydrology (Klemeš, 1986). The results of these Random Forest models show higher RMSE and bias in wet years than the baseline calibration-validation model, while dry years tend to show less (Figure A.11). The low RMSE and bias in dry years is due to the overall shallower snowpack in these seasons.

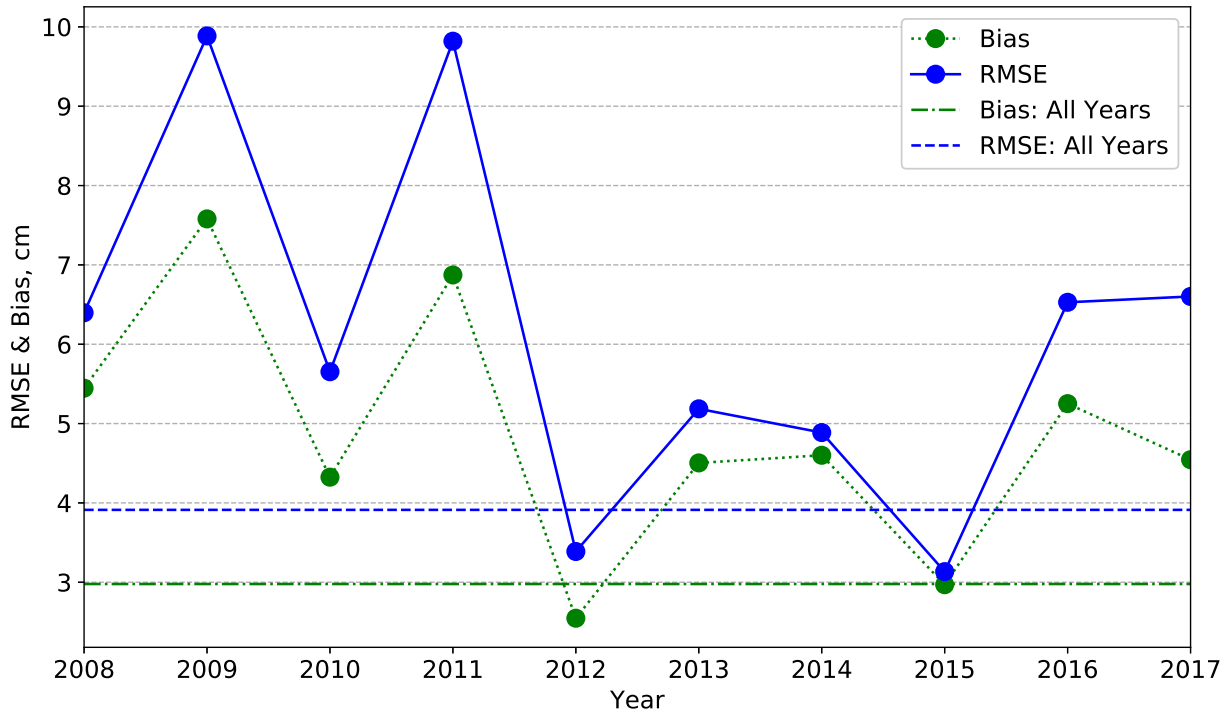


Figure A.11: Results of the leave-one-out Random Forest models. Metrics for the baseline randomized calibration-validation model are also presented.

Appendix B

Supplementary Information: Chapter III

B.1 PRMS data and structure

B.1.1 Topography parameters

HRU delineations were largely imported from the USGS PRMS version 2 with some important adjustments. The HRUs from version 2 were compared to the DEM and certain HRUs were adjusted to better reflect the underlying topography.

PRMS requires inputs of latitude, longitude, elevation, slope, and aspect for each HRU. Latitude and longitude were defined at the HRU centroid. However, because HRUs are delineated based on topography and may not be convex, in several cases the geometric centroid of the HRU fell outside the boundaries of the HRU itself. In most cases, the centroid fell in a nearby body of water. In order to avoid non-physical behavior in the model and because the exact coordinates of the centroid are unimportant except in precipitation distribution, these centroids were left as calculated. In one case, however, the centroid fell outside the boundaries of the Feather River basin entirely. In this case, the centroid was taken from PRMS 2.

The remaining topographic parameters were sampled from the USGS NED at 30-m resolution. For the PRMS GIS model, these values were weighted across the spatial extent of the HRU. Simple averages were taken across the Landsat pixels in the HRU for elevation and slope. For aspect, flat areas were first set to NaN. The remaining pixels were grouped into classes by cardinal or intercardinal direction (N, NE, E, SE, S, SW, W, NW). The mean aspect of the modal class (most common class) for each HRU was taken as the aspect. For the GMM-based PRMS models, all topographic featuers were selected from the 30-m NED DEM pixels in which the modeling points selected by the GMM algorithm fell. Each modeling point was given equal area weight across the subbasin; in other words, the *hru_area* parameter was calculated as the total basin area divided by the number of modeling locations.

The PRMS GIS model was updated from version 2 with defined streamflow segments in order to introduce more-sophisticated stream-routing techniques than simply summing the outflow of all HRUs at each time step. These stream segments were delineated based

on segments defined (but not used in the model) in the USGS release of PRMS v2 for the Feather River. In some cases, multiple streamflow segments flowed through or merged in a single HRU. Where the association of a particular HRU's runoff with a particular stream segment (in other words, the segment where the HRU's runoff would enter the stream) was ambiguous, the runoff was routing to the most downstream segment.

B.1.2 Vegetation parameters

Vegetation inputs for the basin were updated based on the most recent available (year 2013, version 1.4.0) LANDFIRE dataset (https://landfire.gov/vegetation.php), a joint endeavor of the U.S. Forest Service and the U.S. Department of the Interior, with several other partners (LANDFIRE, 2013a,b). The LANDFIRE datasets are dated with a particular year, but are the compilation of several vegetation surveys and sources collected over the course of multiple years. The date given to a dataset is generally the median year of the input data. While this may introduce some uncertainty in the PRMS vegetation inputs, we believe it to be comparable to the uncertainty in other data sources and the model structure. Moreover, though input data were selected to best reflect current conditions and thus near-short-term forecasting, the version 4 was calibrated on a period of record beginning in water year 1998, so there is an inherent mismatch between the input vegetation data and the calibration period. However, we believe our method was the best option given that PRMS version 4 does not allow vegetation data to be updated mid-simulation.

The three PRMS parameters that were derived from the LANDFIRE data were *cov_type*, a categorical variable indicating the dominant vegetation type in each HRU, and *covden_sum* and *covden_win*, decimal fractions between zero and one indicating the coverage density of each HRU's dominant vegetation type in summer and winter, respectively. The *cov_type* values were derived from the LANDFIRE Existing Vegetation Type (EVT) gridded data at 30-m resolution. The frequency distributions of EVT data across each HRU were computed and dominant vegetation type for the HRU was selected for *cov_type*. The EVT vegetation types were translated into PRMS categories as follows: "Developed", "Sparse", or "Open" EVT types were considered bare ground in PRMS; "Herbs" were considered grasses; "Shrubs" were considered shrubs; and "Trees" considered trees.

The EVT data also provided information on whether the pixel contained evergreen, deciduous, or mixed vegetation types. This information was coupled with data from the LANDFIRE Existing Vegetation Coverage (EVC) gridded dataset to determine *covden_sum* and *covden_win* values for each HRU. EVC data were extracted for each HRU in the same way as EVT data. The EVC data provided information on coverage density in each pixel in the form of a percent range (e.g., "Tree Cover >= 10 and < 20%"). Two assumptions were made about these data: 1) that they reflected summer, not winter, conditions and 2) the coverage density was taken as the mean of the range given (15% in the previous example). The coverage density of a given HRU was taken as the average coverage density of the most frequent class in that HRU. This was used as the *covden_sum* value. For the *covden_win* value, this process was repeated, but it was assumed that coverage density in the pixels with deciduous vegetation went to zero and the coverage density in the pixels with mixed vegetation was halved. Coverage density in the pixels with evergreen vegetation was assumed to stay the same across seasons. The same area-weighting process across the entire

HRU was then performed. Thus, all *covden_sum* values are greater than or equal to their corresponding *covden_win* values for a given HRU.

B.1.3 Other parameters

Most parameters that were not derived from climate or topographic data and were not selected for calibration were lifted directly from the USGS PRMS version 2 values. If the parameter value was different for each HRU, the values for any new HRUs in version 4 were taken from the nearest old HRU, where nearest was defined as distance between centroids. Parameters that were new in PRMS 4 were set to the default value. An exception to these rules were the parameters *jh_coef_hru* and *x_coef*. The former, *jh_coef_hru*, which governs potential evapotranspiration calculations, was calculated per HRU based on HRU elevation using Equation 1-53 in the PRMS 4 manual (Markstrom et al., 2015). *x_coef*, a parameter set for each stream segment related to flow attenuation in the Muskingum routing, was calculated dynamically during recalibration based on the value of K_coef, related to flow velocity in the stream segments. This was done to minimize the possibility of creating non-physical flow behavior such as negative flow, which can occur depending on the relationship between the Muskingum routing parameters (Markstrom, 2019).

B.1.4 Input Data

PRMS v4 includes an updated DRAPER model for distributed precipitation to the HRU centroid. For details on the implementation of DRAPER, see Koczot et al. (2005). The algorithm for DRAPER in PRMS v4 is the same as previous versions, but the model now takes input data from three new stations: Antelope Lake (ANT), Frenchman Dam (FRD), and Sierraville (SVL). These replace data from PG&E's Desabla station (DSB), USFS's Quincy-Ranger station (QYR), and National Weather Service's Strawberry Valley (STV) station. Stations used in both versions are Brush Creek (BRS); DeSabla (DES); Quincy (QCY); and Strawberry Valley (SBY), run by DWR, and Bucks Creek Powerhouse (BUP); Canyon Dam (CNY); and Caribou Powerhouse (CBO), run by PG&E. All new stations are run by the Department of Water Resources, and data from all stations are available on CDEC (`https://cdec.water.ca.gov/`; site codes noted above). This change was made to better capture behavior of incoming storms in the eastern portion of the basin, which was underrepresented in the previous version, resulting in a systematic over- or under-estimation of precipitation in that region, depending on which direction the incoming storm arrived from. The new stations were selected for their locations, reliability (they are located in valleys and thus well maintained at all times of year), and the length and quality of their data record. The discarded stations all represented spatial duplicates; that is, they were one of two stations in very close proximity.

Temperature data in the PRMS v4 model are pre-distributed to each HRU and inputted directly using the *climate_by_hru* module. This option was selected based on an analysis of temperature variability across the Feather River basin that showed both temporal (sub-monthly) and spatial (intra-subbasin) variability in lapse rates. Because not every HRU was well-represented by a pair of temperature stations that would allow the lapse rates to be calculated internally by the model, we developed a separate distribution method run

prior to PRMS model execution. Temperature distribution was performed separately for the main North Fork of the Feather River (Almanor, Butt Creek, and Lower North Fork subbasins) and the East Branch subbasin due to substantial differences in topography and climate between the latter and the former.

The distribution was a two-step process: first, we regressed elevation against minimum and maximum temperature values for five (six) stations in the East Branch (main branch of the North Fork) to establish basin-wide lapse rates on a daily basis. These stations were selected based on length of record and quality of data. They were Quincy Ranger Station (CDEC code QYR), Antelope Lake (ANT), Cashman (CSH), Grizzly Ridge (GRZ), and Kettle Rock (KTL) in the East Branch and Canyon Dam (CNY), Bucks Creek Powerhouse (BUP), Caribou Powerhouse (CBO), Chester (CHS), Mineral (MIN), and Westwood (WWD) on the main stem of the North Fork. For days during which fewer than 80% of stations were available, this daily lapse rate was not computed and was replaced by a mean monthly lapse rate computed over the whole period of the record.

Second, these daily lapse rates were used to predict daily maximum and minimum temperature at a number of validation stations that cover a range of elevations, starting from two different seed stations (QYR for the East Branch and BUP for the main North Fork). in the East Branch, these validation stations were (DOY), (GRE), (JDP), (RTL), (TAY), and (TVL) and in the Almanor basin, these were (HAM), (HRK), and (LLP). Residuals between predicted and observed temperature at these evaluation stations were computed on a daily basis and distributed using a multilinear regression against elevation and the temperature of the seed station. The final values of maximum and minimum temperature to be inputted in the model were thus obtained by subtracting the residual computed during step 2 from the first-guess temperature obtained using the lapse rates computed at step 1.

B.2 Model calibration

Table B.1 indicates the parameters used for the SCE algorithm when calibrating each model. Several parameters were set to the default values recommended by the USGS implementation of SCE, called "Let Us CAlibrate" (LUCA; see Hay and Umemoto, 2006). For more information on SCE and the required metaparameters, see Duan et al. (1992, 1993, 1994) and Hay and Umemoto (2006). Table B.2 lists the water years included in each of the five calibration/validation splits.

Table B.1. SCE algorithm parameter values

SCE Parameter	**Value**	**Notes**
Number of complexes[a]	24	∼1 complex per three parameters extrapolated from Duan et al. (1994)
Number of calibration rounds	5	Following Hay and Umemoto (2006)
Maximum number of model simulations per step*	10000	LUCA default
Shuffling loops over which objective function must change by more than a given percent to avoid termination (L)*	5	LUCA default
Percent change allowed in L loops before convergence*	0.01	LUCA default
Maximum percent of parameter space to which points may converge without triggering termination*	0.001	LUCA default

[a] "Complexes" begin as randomized groups of sample points and are evolved to explore the parameter space for optimal convergence. Parameters marked with asterisks are related to a termination criterion of SCE.

Table B.2. Water years used for each calibration and validation split

Split	**Calibration years**	**Validation years**[a]
1	[1998 1999 2000 2002 2003 2005 2008 2009 2010 2011 2013 2014 2016]	[2001 2004 2006 2007 2012 2015]
2	[1998 1999 2000 2001 2002 2003 2005 2011 2012 2013 2014 2015 2016]	[2004 2006 2007 2008 2009 2010]
3	[1998 1999 2002 2005 2006 2007 2008 2009 2010 2012 2013 2014 2015]	[2000 2001 2003 2004 2011 2016]
4	[1998 2000 2002 2003 2004 2006 2007 2008 2009 2010 2012 2014 2016]	[1999 2001 2005 2011 2013 2015]
5	[1998 1999 2000 2002 2003 2004 2005 2008 2009 2010 2011 2012 2014]	[2001 2006 2007 2013 2015 2016]

[a] Streamflow metrics were also validated on water years 2017 and 2018, for which ET and SWE data were not available.

B.3 Supplementary results and figures

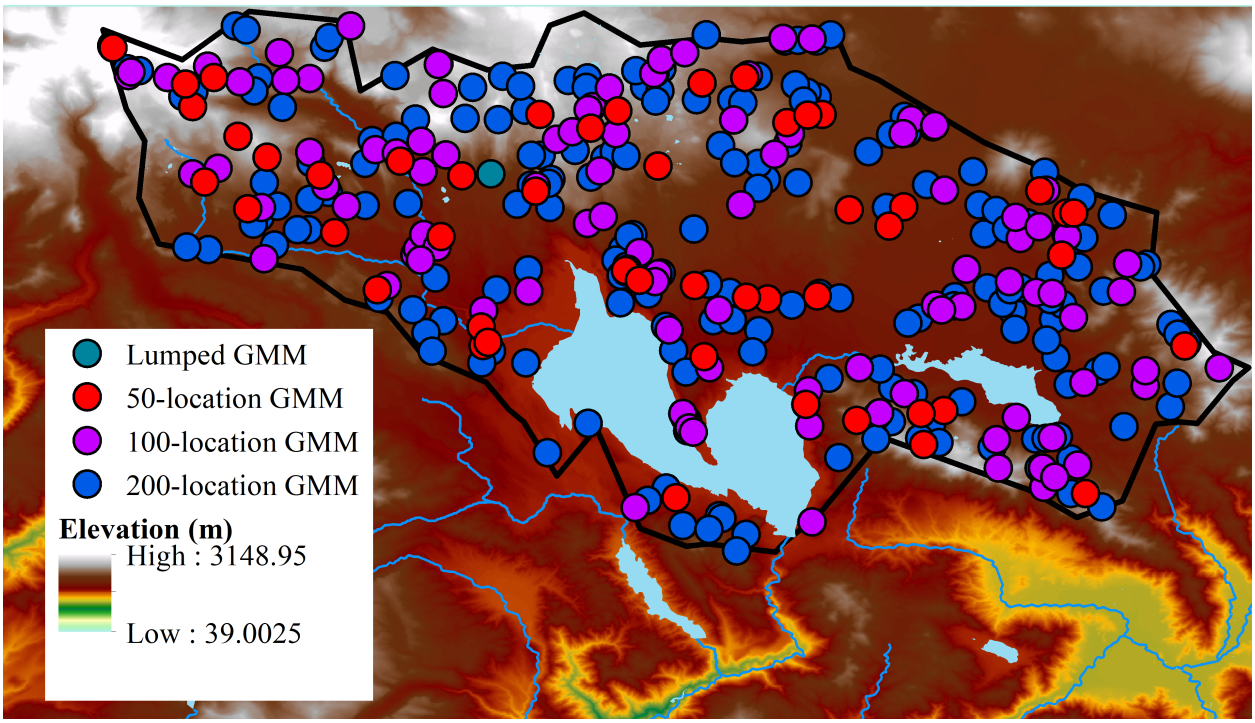


Figure B.1: Modeling locations selected for each GMM model in Almanor.

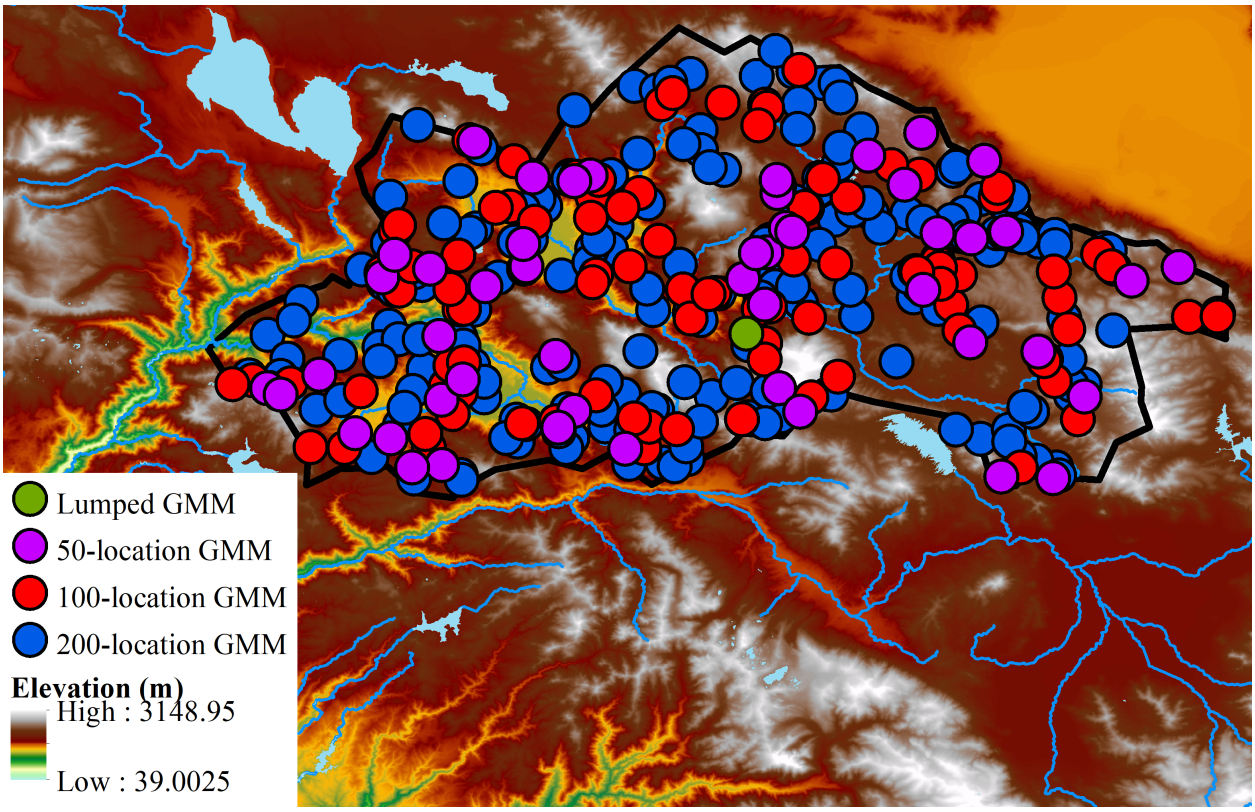


Figure B.2: Modeling locations selected for each GMM model in the East Branch.

Table B.3. Performance metrics measured for validation periods

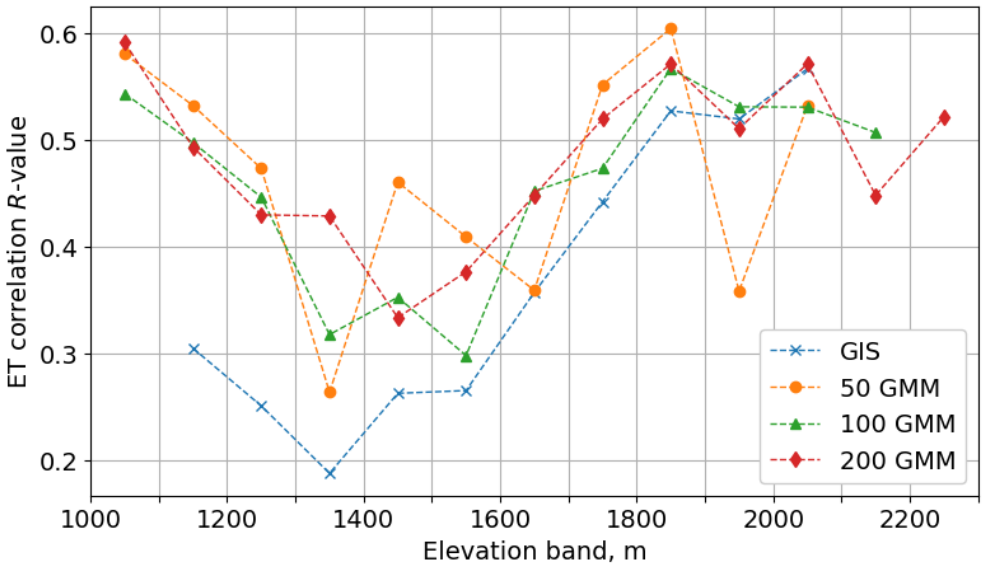
Basin	Model	**Median calibration split**[a]	Model score	SWE bias, mm	ET bias, mm	Runoff KGE	Runoff LogNSE	Runoff RMSE, cms
Almanor	GIS	1	9.8	18.8	108.6	0.73	0.53	15.5
	Lumped GMM	5	18.2	25.9	77.5	0.60	0.46	18.3
	50-location GMM	3	13.6	26.6	54.7	0.70	0.44	15.8
	100-location GMM	2	12.2	29.9	68.9	0.79	0.34	13.6
	200-location GMM	1	15.4	25.8	111.8	0.72	0.44	15.0
East Branch	GIS	5	12.4	18.9	86.8	0.58	0.58	35.3
	Lumped GMM	5	16.6	17.7	85.5	0.54	0.37	42.8
	50-location GMM	1	13.8	12.7	163.0	0.55	0.58	34.3
	100-location GMM	5	12.8	17.6	120.8	0.62	0.62	35.9
	200-location GMM	4	11.8	21.8	121.6	0.73	0.62	34.2

[a]Scores are reported for the median-ranking calibration of each model.

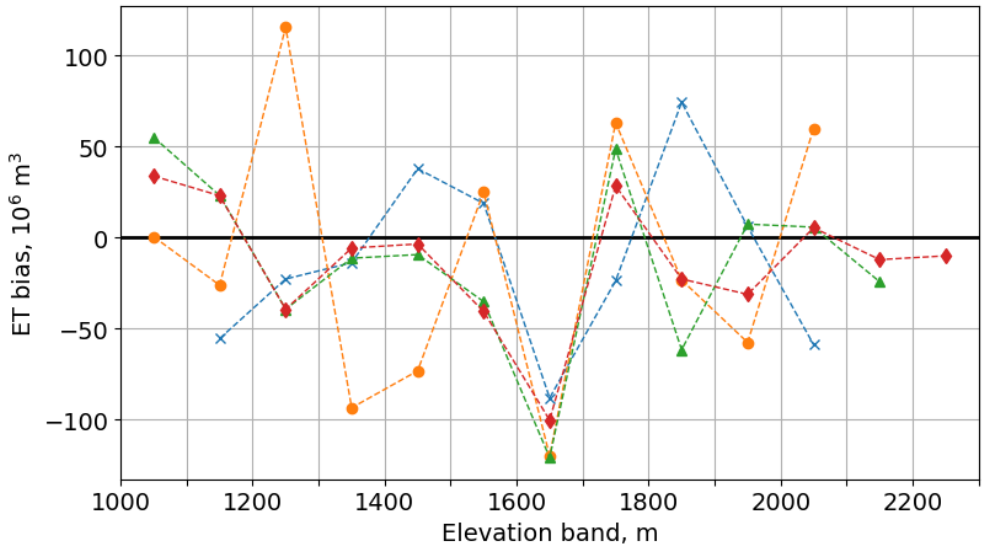
Table B.4. Performance metrics for extreme periods

Basin	Model	**Median calibration split**[a]	Peak SWE bias, mm	Day of peak SWE error, days	Jul-sep flow error, cms	REHF
Almanor	GIS	1	27.3	-0.5	0.9	0.37
	Lumped GMM	5	70.7	-8.7	4.4	0.36
	50-location GMM	3	0.6	0.7	3.3	0.40
	100-location GMM	2	-34.6	-0.3	-0.6	0.30
	200-location GMM	1	39.2	8.2	1.0	0.32
East Branch	GIS	5	60.3	6.2	6.9	0.40
	Lumped GMM	5	71.1	2.7	7.7	0.46
	50-location GMM	1	39.8	1.5	3.0	0.44
	100-location GMM	5	46.8	7.2	3.8	0.46
	200-location GMM	4	-44.2	-5.3	6.2	0.41

[a]Scores are reported for the median-ranking calibration of each model.



(a) Correlation over time



(b) Volumetric bias

Figure B.3: ET performance for each elevation band in the East Branch, averaged across calibrations. Tick marks indicate upper and lower bounds of 100-m elevation bands. The lumped GMM model is not shown.

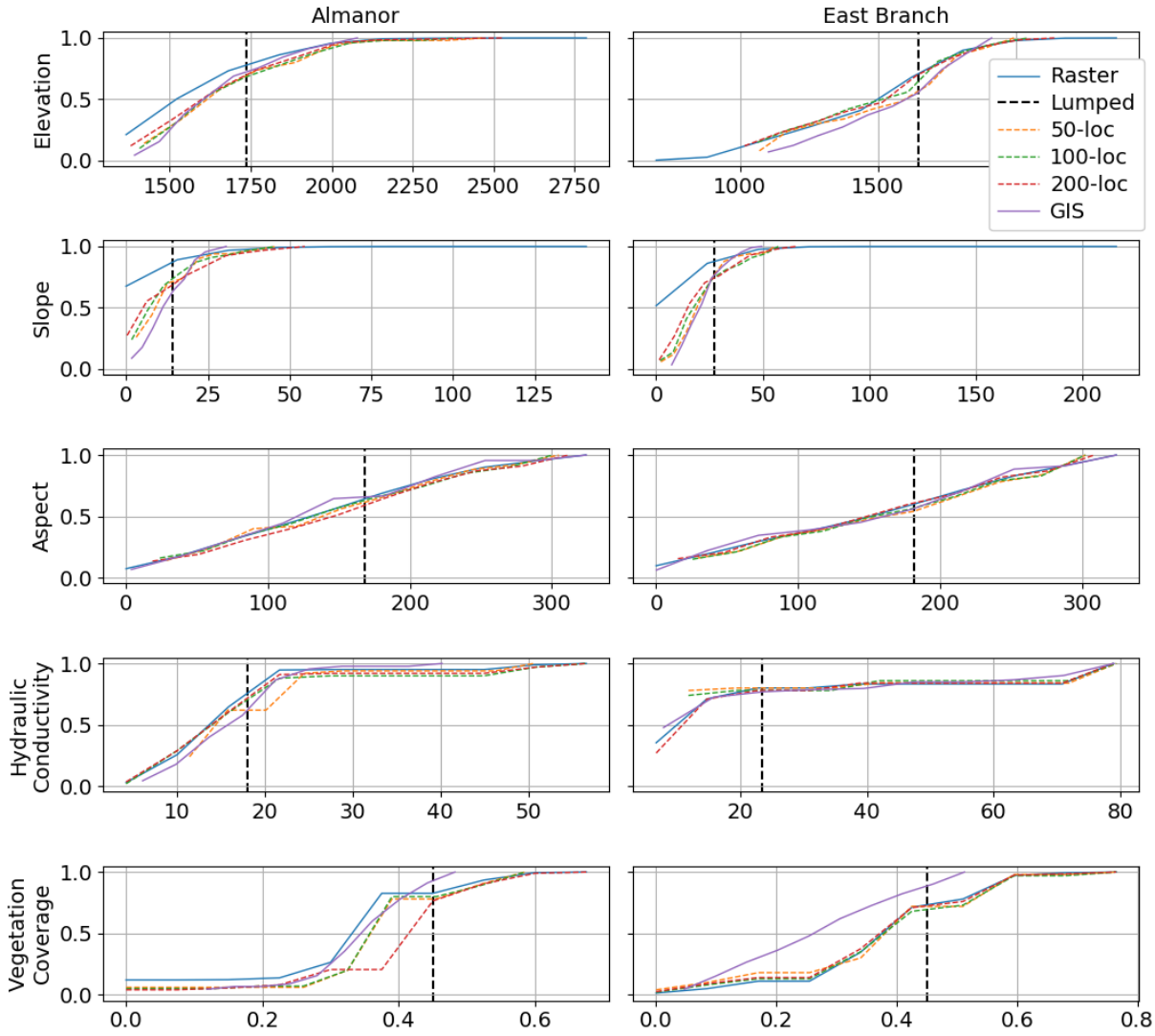


Figure B.4: Empirical cumulative distributions of each GMM input variable for each model and observed raster data.

Appendix C

Supplementary Information: Chapter IV

C.1 Data

Table C.1. Full-natural flow gauges

Basin	Gauge name	CDEC gauge code
Shasta	Sacramento River above Bend Bridge	SBB
Feather	Feather River at Oroville	FTO
Yuba	Yuba River near Smartville	YRS
American	American River at Folsom	AMF
Cosumnes	Cosumnes River at Michigan Bar	CSN
Mokelumne	Mokelumne - Mokelumne Hill	MKM
Stanislaus	Stanislaus River - Goodwin	SNS
Tuolumne	Tuolumne River - La Grange Dam	TLG
Merced	Merced River near Merced Falls	MRC
San Joaquin	San Joaquin River below Friant	SJF
Kings	Kings River - Pine Flat Dam	KGF
Kaweah	Kaweah River - Terminus Dam	KWT
Kern	Kern River - below Isabella	KRI
Tule	Success Dam	SCC

C.2 *abcd* model results

The *abcd* model (Thomas, 1981) can be understood under a generalized proportionality hypothesis framework (Wang and Tang, 2014). The primary equation assumes PET or "evaporation opportunity", Y, for each time step is a function of available water and two parameters, a and b. The former ranges from zero to one and can be understood physically as the tendency for runoff to occur in the basin before the soil is saturated. The latter is the maximum evaporation opportunity, measured in depth. Soil storage in the model is calculated under the assumption that actual ET from the soil occurs in proportion to

PET, Y. The model goes on to separate direct runoff from groundwater recharge based on parameters c and d, allowing total streamflow and baseflow to be calculated as well. However, as the primary goal of using the *abcd* model here was calculating change in soil storage, we did not use parameters c or d. For more details on the *abcd* model and its use in conjunction with the Budyko model, see Du et al. (2016).

A basic sensitivity test was performed for the initial conditions for soil and groundwater storage, which were tested one at a time. The value and direction of shift in ω are robust to initial values ranging between 5 and 500 mm, to reflect the order of magnitude of maximum dry-season storage water draw down that has been reported in the Sierra (Roche et al., 2020). Only in one basin, the San Joaquin, did ω show a shift in the opposite direction for initial soil storage values of 100 and 500 mm.

Table C.2. *abcd* model final calibrated parameters

Basin	**Parameter *a*[a]**	**Parameter *b*, mm[b]**
American	0.979	1355
Cosumnes	0.956	1450
Feather	0.944	1402
Kaweah	0.996	782
Kern	0.993	759
Kings	0.993	670
Merced	0.999	1396
Mokelumne	1.000	1076
SanJoaquin	0.983	780
Shasta	1.000	806
Stanislaus	0.985	1030
Tule	0.985	1442
Tuolumne	0.995	1122
Yuba	0.991	1450

[a]Parameter a reflects the propensity of a basin to generate runoff before the soil is saturated.

[b]Parameter b is the maximum possible evapotranspiration per time step.

Table C.3. *abcd* model performance with respect to runoff

Basin	**Nash-Sutcliffe Efficiency**	**Runoff relative error,** 10^{-4} **mm**
Shasta	0.772	2.04
Feather	0.843	-6.88
Yuba	0.883	-0.11
American	0.870	-0.31
Cosumnes	0.871	-6.46
Mokelumne	0.902	8.96
Stanislaus	0.911	-3.89
Tuolumne	0.894	-0.27
Merced	0.725	-325.19
SanJoaquin	0.916	0.06
Kings	0.883	0.88
Kaweah	0.899	13.34
Kern	0.489	-29.18
Tule	0.749	28.86

C.3 Statistical significance of water balance shifts

Table C.4. Results of Kolmogorov-Smirnoff tests comparing drought and non-drought distributions of the aridity and evaporative indices

Basin	**PET/(P-ΔS)** p**-values**	**ET/(P-ΔS)** p**-values**
Shasta	8.69E-05	0.001987912
Feather	0.000456505	0.024630349
Yuba	0.001666759	0.007511255
American	0.000456505	0.00545947
Cosumnes	0.000309095	0.000309095
Mokelumne	0.000805859	0.007511255
Stanislaus	0.000207438	0.007511255
Tuolumne	0.000207438	0.00393314
Merced	0.000207438	0.000805859
San Joaquin	0.000207438	0.002808533
Kings	0.000207438	0.000805859
Kaweah	0.000805859	0.000805859
Kern	0.00393314	0.000805859
Tule	9.92E-06	4.79E-05