Water Resources Research

RESEARCH ARTICLE
10.1029/2020WR029395

Investigating the Relationship Between Peak Snow-Water Equivalent and Snow Timing Indices in the Western United States and Alaska

A. Heldmyer[1], B. Livneh[1,2], N. Molotch[3,4,5], and B. Rajagopalan[1,2]

[1]Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder, Boulder, CO, USA, [2]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA, [3]Department of Geography, University of Colorado Boulder, CO, USA, [4]Institute of Arctic and Alpine Research, University of Colorado Boulder, Boulder, CO, USA, [5]NASA Jet Propulsion Laboratory, Pasadena, CA, USA

Key Points:

- Annual snow timing indices, primarily snow disappearance timing, explain most of the variance in peak SWE at a majority of SNOTEL locations
- The degree of explanatory power varies regionally and across years, and is affected more by local meteorology than by topography
- A diagnostic model for peak SWE shows a mean $R^2 = 0.76$ and percent bias = 3.6% across five mountainous US regions in a cross-validation

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. Heldmyer,
aaron.heldmyer@colorado.edu

Citation:

Heldmyer, A., Livneh, B., Molotch, N., & Rajagopalan, B. (2021). Investigating the relationship between peak snow-water equivalent and snow timing indices in the western United States and Alaska. *Water Resources Research*, *57*, e2020WR029395. https://doi.org/10.1029/2020WR029395

Received 11 DEC 2020
Accepted 20 APR 2021

Abstract Understanding the distribution of snow-water equivalent (SWE) is crucial for the prediction of water resources in the western United States. Backward running SWE reconstructions use satellite-observed, binary snow presence imagery to reconstruct SWE mass estimates. This approach relies on the connection between snow timing and peak SWE, yet few studies have directly examined this relationship. Here, we investigate the strength and spatiotemporal variation in this relationship across the western United States and Alaska. Within the SNOTEL network ($n = 611$ sites), we find that most variance in peak SWE is explained (median $R^2 = 0.64$, $\sigma = 0.18$) by, in order of explanatory skill, the timing of snow disappearance, onset, and cover duration, with variation in skill primarily related to climate conditions—like winter storm size and storm frequency—rather than topographical setting. We expand this analysis with a diagnostic model of peak SWE driven by remotely sensed snow timing indices applied to five hydrologically important regions in the western United States and Alaska. Uncertainties arising between blending point and 500 m grid-scale observations were found to influence model SWE bias, but a robust correlation (median $R = 0.88$) with observations persisted across all tested thresholds. Overall, this supports the viability of snow timing information for quantifying spatial patterns of peak SWE (mean $R^2 = 0.76$, percent bias = 3.6%) over the past two decades. These findings carry important implications for the development of SWE reanalysis products and for the evaluation of climate and hydrologic models.

Plain Language Summary Knowing how much snow is on the ground in the mountains of the US West and Alaska is important information for water resource managers. Currently, ground station networks that monitor the amount of water in the snowpack are too sparse to observe across landscapes, while most satellites can only observe the presence of snow, rather than water content. Snow models are often used to "fill in the gaps" by estimating snow water content using physical process knowledge about snow melt, and one example is the snow reconstruction. This type of model typically estimates historical snow water content by relating snow disappearance timing with energy inputs into the snow to reconstruct a quantity of snow water content. This relationship between snow timing and water content is known, but its variation has yet to be fully explored. Here, we examine this relationship in detail, showing how it varies in space and time, as well as what meteorological and topographical conditions influence it. We also use satellite-observed snow information to estimate snow water content across landscapes with encouraging results in coastal areas, but with less potential further inland and in Alaska. These results may be valuable for future development of snow models.

1. Introduction

In the Western United States, water from snow-dominated mountainous regions provide important ecosystem services such as irrigation water, hydroelectric power production, and mitigation of wildfire risk (Bales et al., 2006; Cayan et al., 2001; Knowles & Cayan, 2004; Mote et al., 2005). These regions are considered "water towers" because they produce disproportionately more water than adjacent lowlands (Viviroli et al., 2007), and represent roughly two-thirds of the inflow to reservoirs in the western United States (Li et al., 2017). Understanding the volume and spatial distribution of snowpack in mountainous headwaters across the West therefore remains a critical objective for regional water resource security.

Observations of snow via a ground-based network offer high-quality information such as snow-water equivalent (SWE) and have been made in some capacity in the western United States for decades. Since at least 1,906, snow course measurements of snow depth and SWE have been made at regular intervals over the cold season by expeditionary teams collecting snow cores along a transect (Palmer, 1988). The SNOTEL network, which began in 1978, was designed to supplement, and in some cases, replace snow course survey (SCSs), providing relatively cost-effective snow monitoring in regions of high snow accumulation (Dressler et al., 2006). However, while the collected data is of high quality, SNOTEL sampling is limited in geographical coverage (Blöschl, 1999) and spatial representativeness given siting constraints associated with accessibility and protection from public disturbance (Meromy et al., 2013; Molotch & Bales, 2006). Furthermore, SNOTEL stations are often placed in areas of relatively high accumulation, which limits representativeness across the landscape at the elevation in which they are located (Broxton et al., 2019). Thus, while SNOTEL and snow course networks provide valuable information on SWE, their sparse sampling limits direct observation of SWE across entire landscapes.

More recently available observations of snow through remote sensing do provide spatially continuous snow information. In 1999, the launch of the Moderate Resolution Imaging Spectroradiometer (MODIS) platform aboard the Terra satellite made available a series of snow cover products (500 m) shown to be more effective at capturing snow covered area (SCA), particularly in topographically complex forested areas, relative to other satellite products (Gafurov et al., 2013; Maurer et al., 2002; Rittger et al., 2013). However, with the exception of passive microwave sensing, most remote sensing instruments cannot directly estimate SWE, and even passive microwave-based measurements have been reported to provide relatively low accuracy in forests and regions that receive deeper snowpack (Cho et al., 2017; Vuyovich et al., 2014). The principal remotely sensed snow product that is available, SCA, contains only limited information as compared to SWE, which, as it stands currently, is still most reliably observed from ground-based measurements (Vuyovich et al., 2014).

A number of backward running SWE "reconstructions" leverage the strengths of remote-sensing datasets to produce spatially continuous estimates of SWE. The earliest SWE reconstructions followed an initial feasibility study produced by Cline et al. (1998), and typically use information about the DSD to initialize a backward running model that uses a spatially distributed energy balance representation to reconstruct the amount of melt that occurred over time, and thus how much total SWE was present at a given location. During winter season, the timing of complete snow ablation has been shown to be largely controlled by the total magnitude of peak snow accumulation, and this provides the basis for many SWE reconstructions (Durand et al., 2008a, 2008b; Molotch, 2009). However, reconstruction approaches carry a number of important drawbacks, perhaps most significant among them being a limitation to backward running operation due to their reliance on observations of snow disappearance, as well as high computational expense. Slater et al. (2013) detail a number of other sources of uncertainty, including inference of the final date of seasonal snow cover, errors arising from model forcing data, and issues associated with snow process equation fidelity and parameterization in the snow model. Yet, because of a foundation on satellite-derived observational data, snow reconstructions and reanalyses still provide a relatively high degree of accuracy (Bair et al., 2016; Guan et al., 2013; Margulis et al., 2016; Schneider & Molotch, 2016), although this accuracy is not consistent in space or time.

No studies to date have sought to quantify the spatiotemporal variation in the relationship between the timing of snow processes and depth in the context of SWE reconstructions, though several studies have examined timing-depth relationships in other contexts. For example, Lute and Luce (2017) develop models for April 1 SWE and snow residence time based on mean winter temperature and cumulative winter precipitation to assess tradeoffs between model transferability and complexity. Pflug and Lundquist (2020) leverage historical, recurring snow patterns to resolve spatial patterns in snow depth in the Tuolumne watershed of the Sierra Nevada mountain range. Other studies investigate long-term trends in snow cover onset timing, accumulation, persistence, disappearance timing, and SWE depth at continental scales (Broxton et al., 2016; Dawson et al., 2018; Zhang & Ma, 2018). These studies serve as the motivation for the parsimonious and observational approach that underpins this investigation.

Here we assess the relationship between peak SWE and snow timing variables—hereafter the 'SWE-timing' relationship, across the western United States and Alaska. We use the term 'snow timing' to refer to remotely

sensed snow onset, cover duration, and disappearance. The SWE-timing relationship is first analyzed using SNOTEL-derived variables, then through a diagnostic model which includes remotely sensed snow timing information to evaluate the predictive capability of snow timing for SWE depth across wider contexts. This study provides a data-driven analysis into these relationships, as well as the influence of explanatory factors across the breadth of spatiotemporal contexts available from observations, from a multidimensional (i.e., in space and time) perspective. It is unique from solely station-based analyses (e.g., Fassnacht et al., 2003; Harpold & Molotch, 2015; Lute & Luce, 2017; Serreze et al., 1999; Trujillo & Molotch, 2014) since we merge SNOTEL and snow course SWE information with remotely sensed snow cover information to evaluate the strength of the SWE-timing relationship at both point and regional scales. It is also unique from reconstruction studies and those using meteorological forcings (Broxton et al., 2016; Dawson et al., 2018), as this analysis does not rely on meteorological forcings. Rather, it is predicated primarily on snow timing variables obtained directly from observations to limit error sources, that is, from reconstructed melt rates or estimates of air temperature or radiation fluxes. Further, the parameterized model at the core of many reconstructions contains equations that are often of inadequate fidelity (sources of structural uncertainty) for snow process representation (Slater et al., 2013).

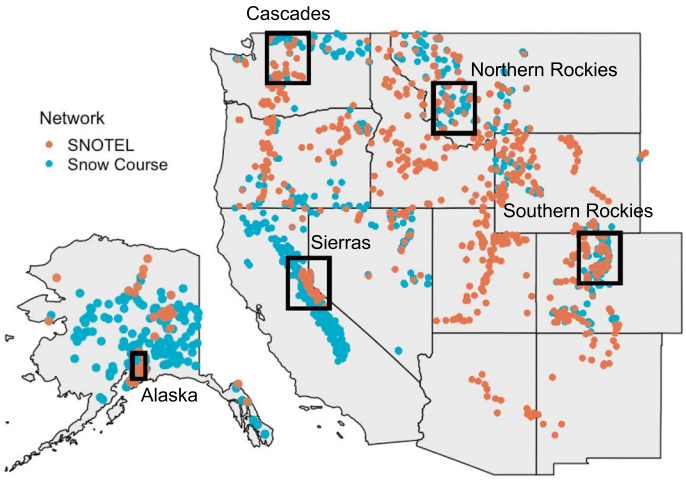
The objective of this study is to address the following questions: (1) How does the relationship between peak SWE and snow timing evolve through time? (2) How does this relationship vary across the western United States and Alaska? and (3) What meteorological and topographical conditions affect this relationship? We begin by examining the SWE-timing relationship solely within the SNOTEL record to minimize uncertainties arising from disparate observational datasets. Here, we explore auxiliary snow and terrain indices at SNOTEL locations in the context of explaining the strength of the relationship on a site-to-site basis. The analysis is subsequently expanded to the regional scale through the incorporation of remotely sensed SCA data. An observation-based hierarchical model is used to quantify the influences of terrain and climate conditions on estimated peak SWE distribution, and determine the predictive skill that may be derived from snow timing. Finally, we apply the data-driven model and evaluate the utility of snow timing variables in estimating SWE across five hydrologically unique regions in the western United States and Alaska, through both cross-validation procedures as well as via comparison to supplemental SCS observations. Peak SWE is estimated within this diagnostic model for individual grid cells, as opposed to a regionally determined peak SWE obtained from a uniform date of maximum SWE depth, to account for spatial variability in peak SWE timing.

2. Methods

We first investigate links between peak SWE and snow timing variables, that is, the SWE-timing relationship, across a large sample of SNOTEL locations across the West (Section 2.1). An analysis into the explanatory power of temporally varying snow indices and spatially varying local topographic characteristics contributing to the strength of the SWE-timing relationship is described in Section 2.2. Next, an application of the SWE-timing relationship is conducted over five densely gaged and hydrologically important regions within the SNOTEL network (Figure 1) in order to assess the predictive capability of snow timing over larger areas. Here, remotely sensed snow cover information is combined with SNOTEL SWE into a data-driven hierarchical peak SWE model (Section 2.3). In acknowledgment of the sparsity of SNOTEL sampling, we validate the model configurations through both cross-validation as well as through comparison with distinct snow course observations within the five study regions (Section 2.4). A description of all data sources and data screening procedures is provided in Section 2.5.

2.1. Station-Based Analysis of the Relationship Between SWE and Snow Timing Variables

For each SNOTEL site with continuously available data over the past two decades (full details of screening procedure in Section 2.5), daily observations were processed into four annually aggregated variables of primary interest for this analysis. These include annual peak SWE depth, as well as three snow timing variables that would typically be available from remote sensing: date of snow onset (DSO), DSD, and snow cover duration (SCD). We analyze the spatiotemporal variability in the strength of the relationship between these timing variables and peak SWE depth. An ordinary least squares (OLS) regression between log-transformed



Domain	SNOTEL count	SCS count
Cascades	26	33
Sierras	26	99
Northern Rockies	19	34
Southern Rockies	39	59
Alaska	7	46

Figure 1. Locations of SNOTEL network sites (*red*) used in the evaluation of the SWE-timing relationship across the West. The five regional study domains used for analysis of the peak SWE-snow timing relationship are shown as *black* boxes (clockwise from top-left: Cascades, Northern Rockies, Southern Rockies, Alaska, Sierras), with SCS network locations (*blue*) used for validation. Only the SCS locations within each region were used in this study. Data sources and screening procedures are provided in Section 2.5.

(for linearization of the relationship) annual peak SWE and timing variables was performed for each SNOTEL site $s = 1...N_s$, where N_s is the number of SNOTEL stations during the study time period, t:

$$\log\left(SWE_{peak}\left(s,t\right)\right) = \beta_0\left(s\right) + \sum_{i=1}^{n} \beta_i\left(s\right) ST_i\left(s,t\right) + \epsilon\left(s,t\right) \quad (1)$$

Regressing the snow variables in this manner yielded a set of $N_s = 611$ y-intercept coefficients (β_0), regression coefficients (β_i), and coefficients of determination (R^2), along with an error time series ϵ at each SNOTEL station, for each snow timing variable, ST_i. The coefficient of determination, R^2, describes the variance in peak SWE explained by the snow timing variable(s). The R^2 in a regression incorporating all snow timing variables represents a theoretical maximum relationship strength between SWE and snow timing at a site, while comparing R^2 values between regressions using each timing variable in isolation yields a site-specific "dominant index"; that is, the timing variable that explains the greatest amount of variability in annual peak SWE. The relative contribution among timing variables in explaining peak SWE provides context for the unique conditions that govern accumulation and ablation at a given site. Comparisons across sites were made to uncover the broader spatial pattern of these site dynamics, as well as consistency among sites within each region.

2.2. Analysis of Snow Timing Skill With Auxiliary Indices

To explain the site-to-site variation in SWE-timing relationship strength among the snow timing variables, temporally varying snow indices (I_t) identified by Fassnacht and López-Moreno (2020) were collected along with spatially varying topographical indices (I_s) developed from a 30 m shuttle radar topography (SRTM) digital elevation model (DEM; NASA LP DAAC, 2013) at each SNOTEL site. Conditions affecting snowpack, such as the frequency and magnitude of winter storms, snowmelt during the period of accumulation, and topographical complexity such as slope, aspect, and elevation are largely described by I_t and I_s, and summarized below in Table 1. Full descriptions of the indices are provided in Section 2.5.

To broadly characterize subsets of gages where snow timing variables provide high or low explanatory skill in explaining peak SWE variability, differences were computed between the full population of gages and a subset of gages that fall below the 10th percentile of R^2 values, and those that are above the 90th percentile. A one-sample, two-sided t-test between these samples and the SNOTEL network as a whole was performed at the significance level $p = 0.05$. Defining the characteristics of sites that fall in one of these two groups was then determined by identifying the indices in which the samples significantly differ from the population.

Table 1
Description of Temporal (I_t) and Spatial (I_s) Indices Used in the Station-Based Analysis

Temporal indices	Description	References	Spatial indices (from SRTM 30-m DEM)	Description	References
Peak date	Date of peak SWE	Fassnacht and López-Moreno (2020)	Latitude	WGS84 latitudinal coordinate	NASA LP DAAC (2013)
Peakedness	Days with SWE >50% peak SWE		Longitude	WGS84 longitudinal coordinate	
Snow days	Days when SWE increase >5 mm		Elevation	SRTM-derived elevation (m)	
Accum snow days	As above, but only during accumulation period		Slope	SRTM-derived slope (degrees from horizontal)	
Avg accum	Peak SWE/days with snowfall		Aspect	SRTM-derived aspect (degrees from North)	
Melt days accum	Days when SWE decrease >5 mm during accumulation season		Topographical position index (TPI)	Difference between the value of a cell and the mean value of its eight surrounding cells (m)	Hijmans (2020)
Melt during accum	Sum of SWE decrease >5 mm during accumulation period		Terrain ruggedness index (TRI)	Mean of the absolute differences between the value of a cell and the value of its eight surrounding cells (m)	
Cumulative SWE	Peak SWE + amount of melt during accumulation period		Roughness	Difference between the maximum and the minimum value of a cell and its eight surrounding cells (m)	
Melt days melt	Days when SWE decrease >5 mm during ablation period		Exposure index, filter length = 0.01°	Similar to TPI, using neighbor cells within 0.01° (m)	Anderton, et al. (2004)
Accum days melt	Days when SWE increase >5 mm during ablation period		Exposure index, filter length = 0.05°	Within 0.05° (m)	
Melt rate	Peak SWE/ablation period length		Exposure index, filter length = 0.1°	Within 0.1° (m)	

The residual series, ϵ, produced from each site-specific regression (Equation 1) represents the variability in annual peak SWE not explained by the variability in the snow timing variables. Typically for a site with a relatively strong relationship, these errors are stochastic with relatively low magnitude, and randomly distributed. However, correlation between the error series and another index may represent additional skill attributable to the index, beyond that which is achievable through the use of snow timing variables alone. Consequentially, sites may be characterized through this method by relating the strength of the SWE-timing relationship to other indices. We explore diverging site characteristics between strong and weak SWE-timing relationship sites by analyzing the Pearson correlation between the residual series ϵ generated from the regression (Equation 1) that includes all timing variables, and I_t through time at each SNOTEL site. Toward the same end, we also analyze the Pearson correlation between the regression R^2 and I_s across all SNOTEL sites.

2.3. Evaluating the Predictive Power of Snow Timing Through a Data-Driven Model

An application of the above site-based analyses is a spatially continuous framework for estimating peak SWE, developed for WY2001-2019. SNOTEL-derived peak SWE and snow timing data are blended with remotely sensed snow timing data, as well as a subset of auxiliary variables from Table 1, to generate maps of peak SWE across five regions in CONUS and Alaska (Figure 1). The purpose is to evaluate the predictive potential of the SWE-timing relationship through the prediction of peak SWE at locations outside of those

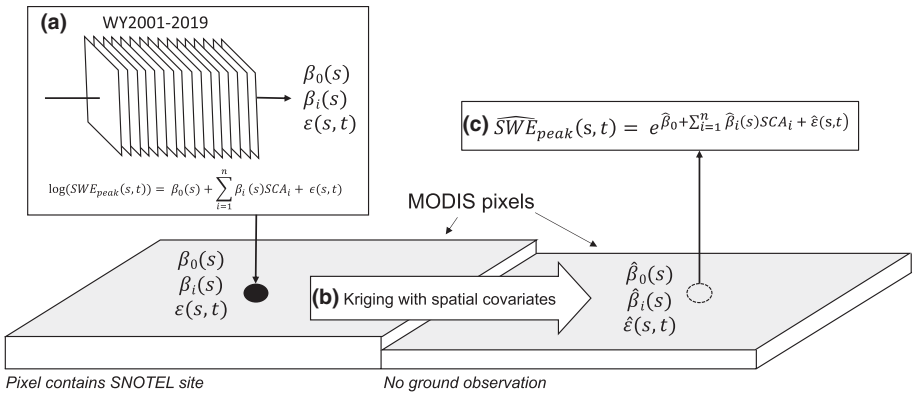


Figure 2. A schematic of the hierarchical model used to test the predictive skill of snow timing (day of snow disappearance) for annual peak SWE. (a) Initially, a log-linear relationship is developed between SNOTEL-observed annual peak SWE and SNOTEL-observed day of snow disappearance for all MODIS pixels containing a SNOTEL gage. A set of regression coefficients and a residual time series are produced. (b) These are then kriged using pixel geographic position and terrain coefficients to obtain model-estimated regression coefficients at ungauged locations. (c) The coefficients are then used to back-calculate a model-estimated peak SWE for each year at ungauged locations.

used as training data, that is, those locations withheld from training. The utility of snow timing as a predictor variable across a variety of sampling densities and regions, as well as the influence of additional indices from the site-based analysis, is explicitly examined.

The statistical model follows a two-level hierarchical approach similar to Suchetana et al. (2016) and Verdin et al. (2015), blending sparsely sampled but serially complete ground observations (i.e., direct measurements of SWE) with spatially continuous yet indirect observations of snow cover via satellite (Figure 2). The first hierarchy estimates annual peak SWE at SNOTEL sites through a linear regression between annual peak SWE and two snow timing variables developed from satellite observations, DSO, and DSD, following the regression framework described in Equation 1. The timing variable SCD was excluded because it was a linear combination of DSO and DSD. The second hierarchy spatially interpolates the regression coefficients developed at each site, along with a residual time-series, thus allowing estimates of peak SWE to be made at ungauged locations. A spatial Gaussian process (i.e., co-Kriging) is used to describe the spatial pattern of the coefficients and residual series, characterized by the exponential variogram function:

$$\tilde{\gamma}(h) = \tau^2 + \sigma^2 \left(1 - e^{-\frac{|h|}{\phi}}\right) \tag{2}$$

where τ^2 is the nugget effect, which describes the level of dissimilarity between points at infinitesimally small lag; σ^2 is the sill, or the apparent total variance as the variogram approaches infinity; and ϕ is the effective range, or the distance at which covariance between points plateaus.

Three coefficients were identified by applying the multivariate regression equation at each site location: one describing the intercept, and two for each timing variable DSO and DSD. A combination of spatial predictors was implemented in the second model hierarchy for the theoretical variograms (Equation 2) fit using the 'gstat' package in *R* (Pebesma, 2004) for each coefficient and the residual series. These spatial predictors include I_s identified as relatively influential (highest ranking correlations) in the site-based analysis: terrain ruggedness index (TRI), roughness, slope, elevation, latitude, and longitude. The coefficients and residual series were co-Kriged, and then used to estimate peak SWE at ungauged locations from Equation 1, yielding a spatially continuous estimate of peak SWE for each year. Model estimates are in the same horizontal resolution as the MODIS satellite data (i.e., 500 m) and are calculated for each year within the study period (WY2001–2019) across all five domains outlined in Figure 1.

2.4. Model Validation

Validation of the estimated peak SWE was done through cross-validation of modeled versus SNOTEL SWE, as well as a comparison to SCS SWE observations within each of the five test regions. Cross-validations were applied to test model performance under a leave-one-out cross-validation (LOOCV) framework, in which one SNOTEL point is systematically dropped from the training set, and predicted over using the remaining data. This was done for each SNOTEL gage located in a test region. SCS SWE observations were used as an additional validation dataset. SCS data are generally only available on the first day of months in the latter part of the snow season (approximately January through June). Therefore, the maximum first-of-month SCS SWE observation within a given year is taken as an approximator for peak SWE, and expected to have a negative bias relative to model-estimated SWE because peak SWE is unlikely to occur at the start of the month. This bias is estimated using SNOTEL data within each region, by comparing the bias between SNOTEL peak SWE and SNOTEL maximum first-of-month SWE. SCS measurements taken at the same location as a SNOTEL site were discarded to avoid duplications of sampled data in both the validation and training datasets. Standard statistical metrics such as bias, percent bias, and R^2 are applied on a yearly basis to quantify performance.

2.5. Data Sources and Screening Procedures

SNOTEL and SCS screening, selection of the snow disappearance threshold, and selection and calculation of additional indices (I_t and I_s), are described in Supporting Information S1. MODIS data selection and processing are described below.

2.5.1. MODIS Data Processing

Images from the MOD10A1 (Hall & Riggs, 2016) platform aboard the MODIS Terra satellite were collected for each water year over a region spanning the western United States and Alaska, defined by the bounding box extending from 152°W to 100°W and 32°N to 68°N. MOD10A1 was chosen over MODSCAG (Painter et al., 2009) for this application, given the necessity for expanded geographical coverage, availability within Google Earth Engine, as well as a comparative analysis described in supporting information S2. Following Langlois et al. (2004), a moving median window (i.e., a low-pass filter) of filter length $k = 25$ days was applied to the daily fractional snow covered area (fSCA) series to obtain a binary snow cover series using a fractional cover threshold such that I_t may be extracted. The filter length smoothes out small-scale short-lived snow deposition events. Frequently, an adjustment to remotely sensed fSCA data for the viewable gap fraction (VGF) is applied to address issues of canopy masking snow cover (J. Liu et al., 2008; Molotch & Margulis, 2008; Raleigh et al., 2013). To evaluate uncertainties associated with threshold selection including the known impacts of canopy coverage, five different fSCA thresholds (1%, 5%, 10%, 20%, and 30%) were tested, resulting in five unique sets of MODIS-derived snow timing variable images for use in the diagnostic model. Pixels with persistent snow cover before the beginning, or after the end of the water year, were omitted from the analysis for that year.

3. Results

3.1. Station-Based Analysis of the Relationship Between SWE and Snow Timing Variables

Snow timing variables (DSO, SCD, and DSD) explain 64% of peak SWE variability (median $R^2 = 0.64$; $\sigma = 0.18$) across 611 SNOTEL sites (Figure 3). The strongest contributor to the overall R^2 among timing variables varied from gage to gage. However, the majority of the skill was most frequently attributable to DSD with median R^2 of 0.56 ($\sigma = 0.20$) using DSD as the lone independent variable, followed by SCD ($R^2 = 0.36$, $\sigma = 0.20$), and DSO ($R^2 = 0.05$, $\sigma = 0.10$). An exception was found in the Northern Rockies, for example, where the peak-SWE from several gages was predicted most strongly by the DSO. Gages in this region also tended to have the lowest overall R^2.

Alaska and the Northern Rockies show a high degree of variation in relative explanatory power among timing variables, while the Sierras and Cascades, for example, are much more uniform (Figure 4). Among the snow timing variables, DSD shows the strongest correspondence with peak SWE. Although this is a well-known relationship that underpins SWE reconstruction models (Déry et al., 2005; Guan et al., 2013;

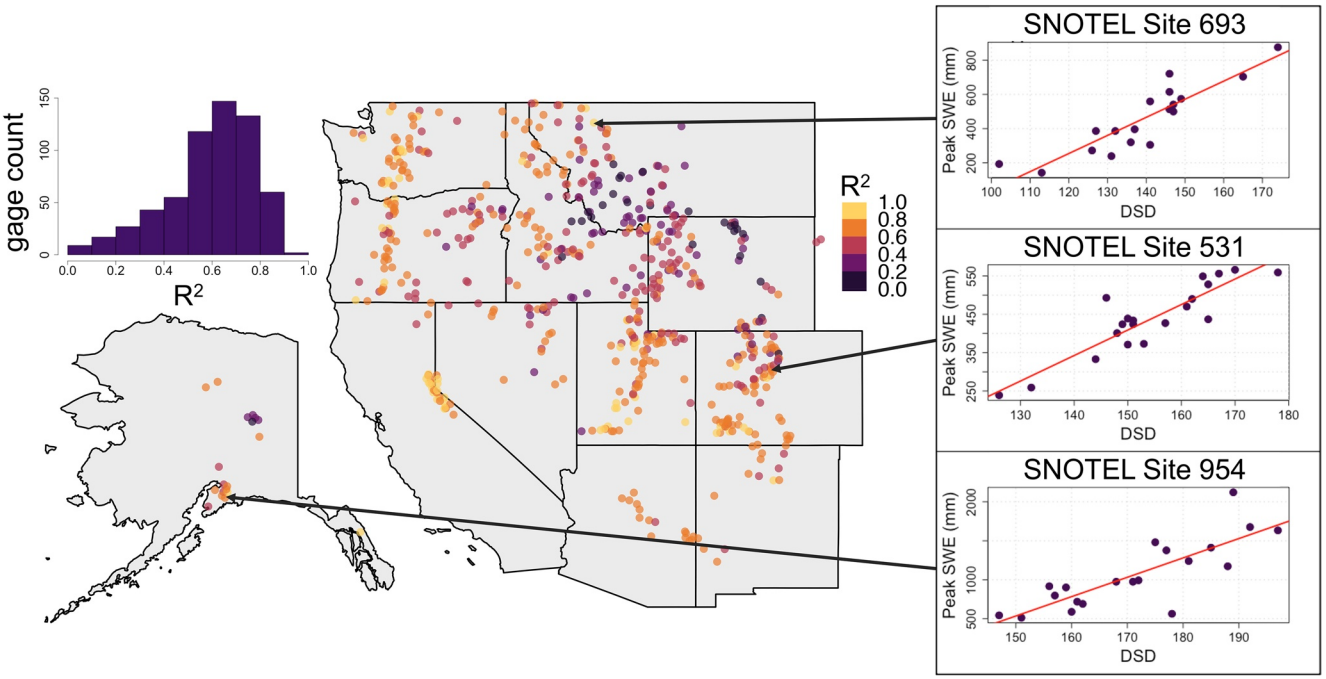


Figure 3. Geographical variation in the coefficient of determination (R2) across 611 SNOTEL sites for the multivariate regression of peak SWE as a function of the three snow timing variables: day of snow onset (DSO), snow cover duration (SCD), and day of snow disappearance (DSD). A histogram of R2 values across all gages is shown in the top-left. A panel of three representative SNOTEL gages (R^2 from top to bottom: 0.82, 0.81, 0.69) demonstrate the relationship between peak SWE and one timing variable, DSD.

Molotch, 2009; Schneider & Molotch, 2016; Sturm et al., 2010), it has not been rigorously explored as a direct predictor of peak SWE as it is here. The physical connection of DSD to peak SWE can be understood by the higher energy requirement for ablation of deeper snow since it contains more mass than shallower snow. Deeper snowpack has a higher cold content (energy deficit), and thus more energy and time are needed to raise the internal temperature to 0°C and for melt to occur (Marks et al., 1999), altogether resulting in a later DSD. By contrast, DSO has the weakest relationship with peak SWE volume by a substantial margin. This result is expected, as early winter storms that deposit snow dictate the timing of snow onset, and the natural variation in timing and magnitude of these local weather events shows little connection with peak SWE depth at most SNOTEL sites. SCD, as a variable developed from snow onset and disappearance, generally shows a relationship strength that falls between the two.

Despite this general hierarchy of explanatory power (DSD > SCD > DSO), certain gages offer exceptions, and in many cases are spatially clustered. Most gages are dominated by DSD among timing variables in terms of explanatory power, and even gages that are not dominated by DSD are explained by a reduced ability to accurately determine DSD, rather than an apparent increase in explanatory power from DSO or SCD. This is evident from the coincidence of low R^2 from all timing indices and from DSD alone. For example, the majority of DSO-dominated gages are concentrated in the Northern Rockies. Of the 24 gages dominated by DSO, 21 are located in either ID, MT, or WY. These gages also tend to be those with the weakest SWE-timing relationship overall. Twenty-one of the 24 DSO-dominated gages are in the lowest 10% of gages in terms of overall R^2. By contrast, a total of 97 gages across the West are dominated by SCD, showing a mean $R^2 = 0.53$, and are found in every state, though a majority ($n = 31$) still exists in ID, WY, or MT.

3.2. Analysis of Snow Timing Skill With Auxiliary Indices

The sites with the strongest and weakest SWE-timing relationship relative to the population are shown in Figure 5. For sites with a strong SWE-timing relationship, statistically significant (p <0.05) differences in I_t and I_s relative to the entire population include longitude (higher R^2 sites are generally more West than the population), slope (at steeper slopes), roughness (in more rough terrain), and average accumulation (more

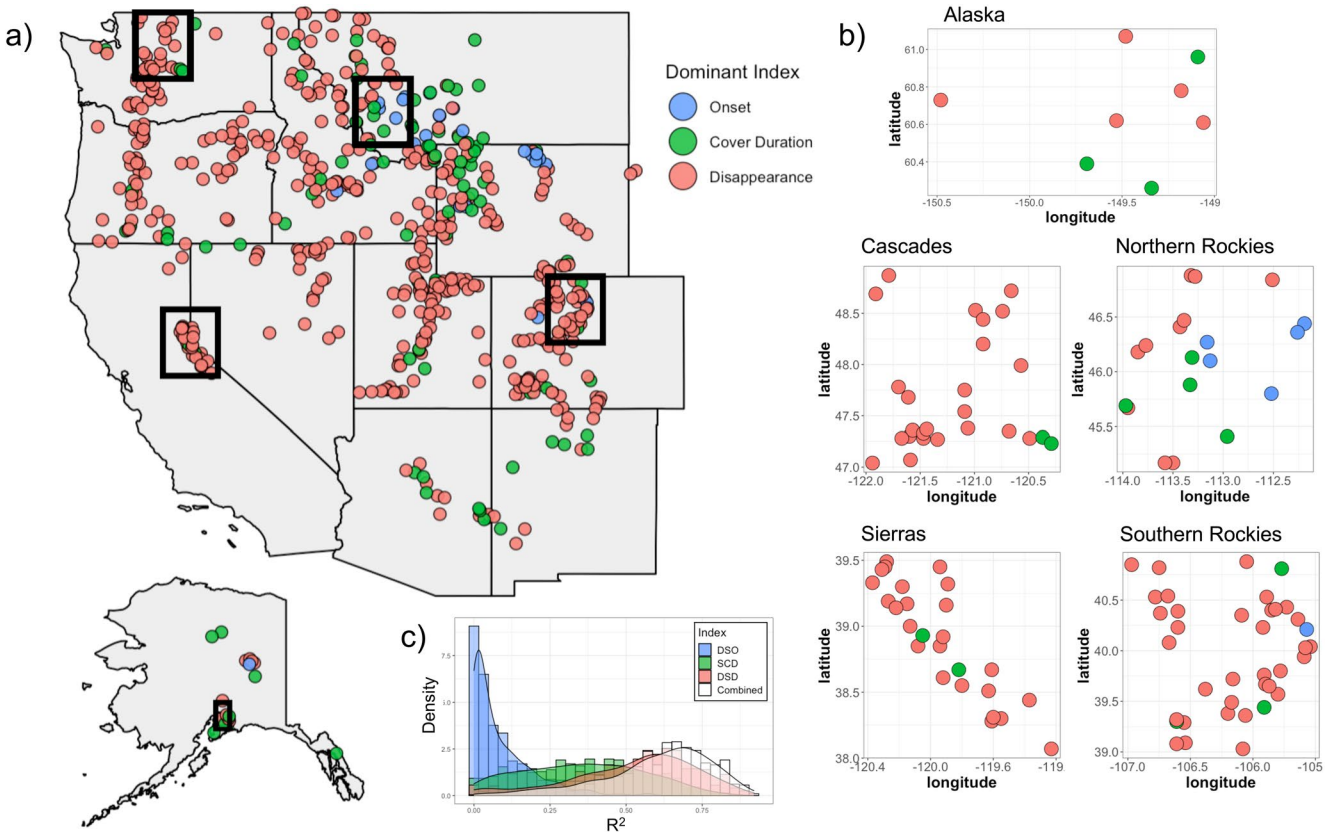


Figure 4. Coefficients of determination (R^2) across 611 SNOTEL sites (WY2000–2019) for 3 annual snow timing variables predicting annual peak SWE: day of snow onset (DSO), snow cover duration (SCD), and day of snow disappearance (DSD). (a) This shows the dominant (highest R^2) timing index by color, with the five study regions outlined. Individual panels for each study region are displayed in (b), and (c) shows a histogram and kernel density function of the distributions of R^2 values across sites for predicting peak SWE as a function of each timing index in isolation as well as combined.

snowfall). For weak SWE-timing relationship sites, longitude (more East than the population), latitude (more North), elevation (gages are at lower elevations), and exposure indices at filter lengths 0.05° and 0.1° (greater exposure at these scales) were found to be significant. Notably, while high R^2 sites were often found

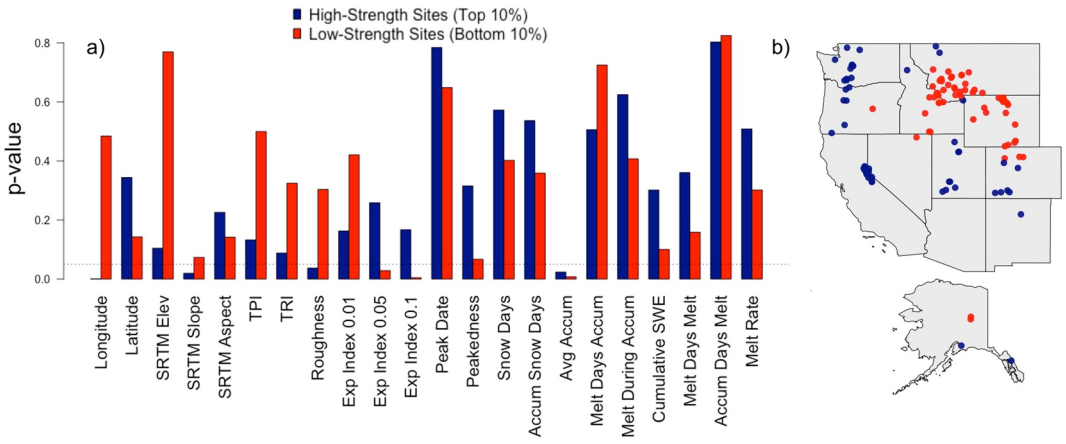


Figure 5. (a) This shows the statistical significance in differences between the population of SNOTEL sites and samples representing gages with a strong SWE-timing relationship (*blue*) and weak relationship (*red*), as characterized by I_t and I_s from Table 1. The horizontal line represents a significance level of 0.05. In (b), color-corresponding map shows the geographical distribution of gages in each group highlighting the uniqueness of interior gages.

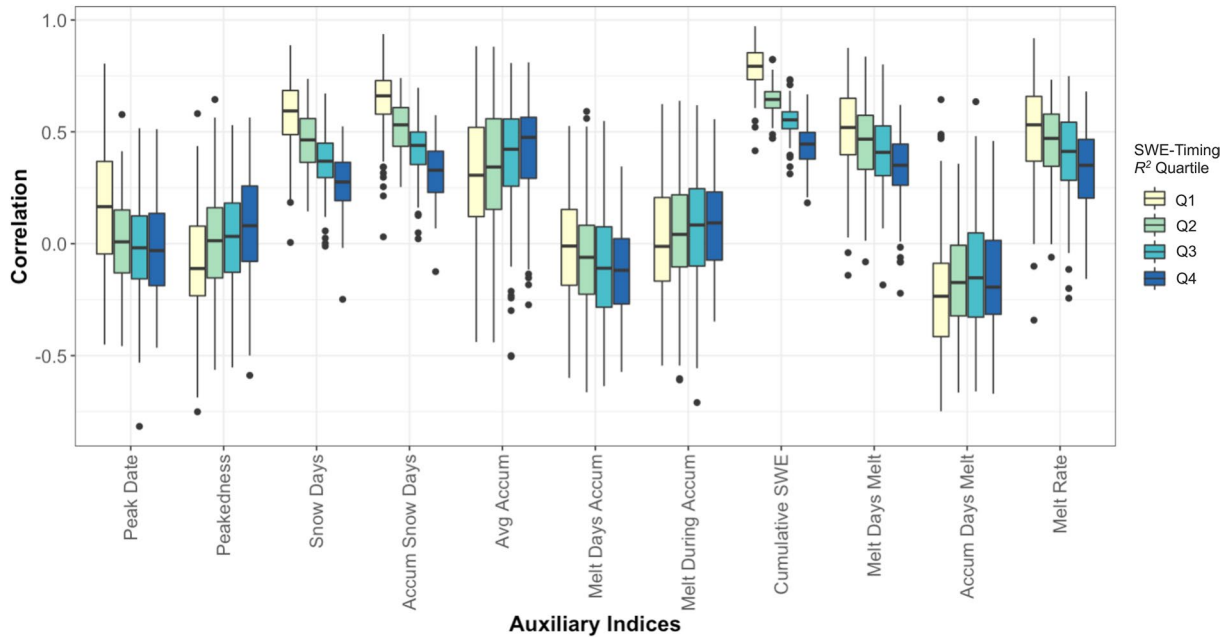


Figure 6. Boxplots of Pearson correlations between regression residuals from Equation 1 and snow indices from Table 1, disaggregated into quartiles based on the R^2 value of the SWE-timing relationship fit. Q1 describes R^2 values below the 25th percentile, Q2 between the 25th and 50th percentiles, Q3 between the 50th and 75th percentiles, and Q4 above the 75th percentile.

near the coast, low R^2 sites were concentrated in interior regions, such as the Northern Rockies (MT, ID, WY) and the interior of Alaska.

The site-based residuals from Equation 1 were regressed against I_t (Table 1) in Figure 6. Correlations were disaggregated into quartiles based on the R^2 of the fit between peak SWE and timing indices which generated the residuals. We find stronger correlations at low quartile (Q1) sites and weaker correlations at higher quartile sites for peak date, number of snow days, number of snow days during accumulation, number of melt days during accumulation, cumulative SWE, melt days during melt, and melt rate. We find the opposite pattern for peakedness, average accumulation, and melt during accumulation. We find 6 of the 11 tested indices describe a relatively high correlation (R >0.4) with residual series across SNOTEL sites. Among these, the number of days with snowfall (snow days), the number of days with snowfall during the accumulation period (accum snow days), and the number of days where SWE decreases >5 mm during the melt season (melt days melt) are not directly calculated using peak SWE volume. This is important to note because the magnitude of the residuals in any regressed relationship naturally corresponds with the magnitude of the regressand. Therefore, indices calculated using peak SWE volume (the regressand) will show some degree of baseline correspondence with residuals.

The analysis between residuals and I_s yielded little correspondence and was hence excluded from Figure 6. None of the indices, with the exception of latitude ($R = -0.24$) and longitude ($R = -0.25$), exhibited a correlation of magnitude R >0.20. This implies a weak influence of the tested indices on SWE-timing relationship strength, and thus the majority of relationship strength is likely attributable to meteorology, rather than topography.

3.3. Evaluating the Predictive Power of Snow Timing Through a Data-Driven Model

For simplicity, the hierarchical model (Section 2.3) variogram for all described model experiments was constructed using a single configuration. Only the six I_s variables with the strongest correlation with Equation 1 residuals—latitude, longitude, elevation, roughness, TRI, slope, and elevation—were used. Low correspondence was found between most spatial covariates and SWE-timing relationship strength such that combinations of these indices yielded negligible difference in the variogram parameters. For example, the

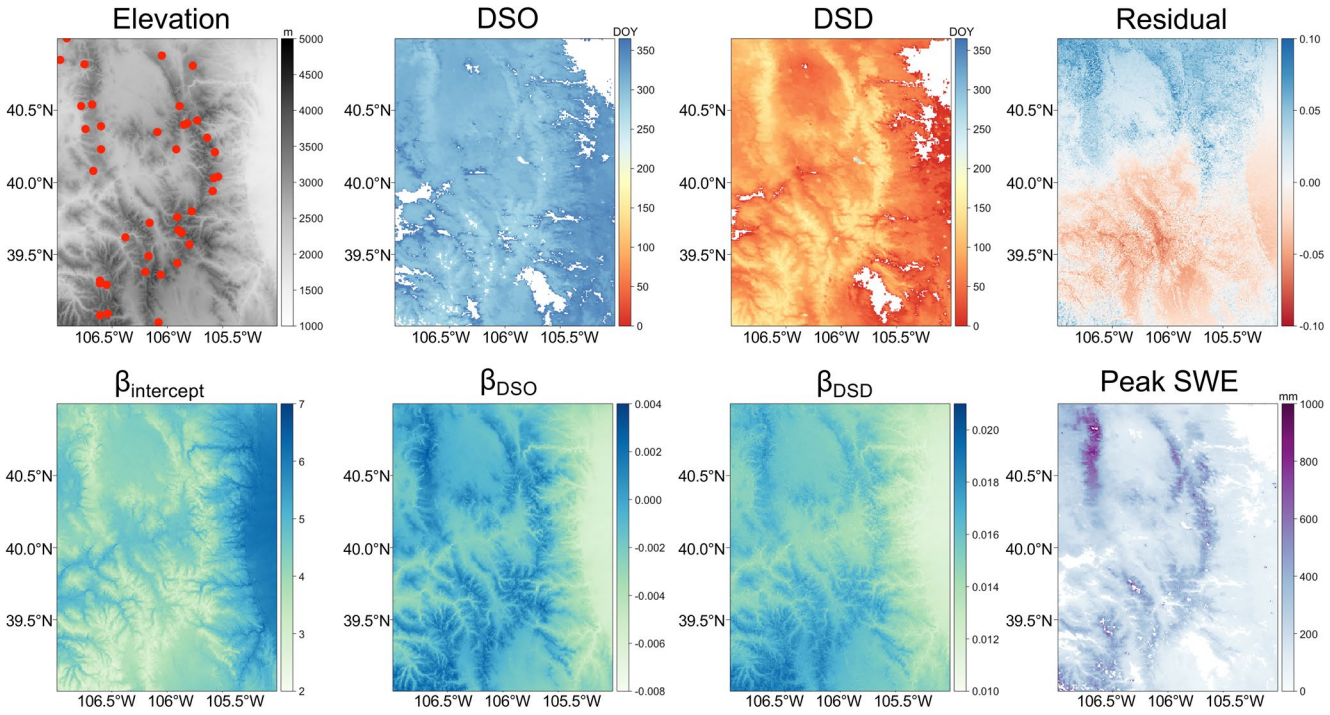


Figure 7. Example spatial maps for components of the hierarchical model applied in the Southern Rockies for 2012. The elevation map also shows SNOTEL locations within the region as *red* points. SO and DSD show MODIS-derived timing variables. These are multiplied by their respective co-Kriged betas, shown below each, added to the residuals and intercept beta, and exponentiated to reverse the log transformation and arrive at an estimated peak SWE. Masked pixels seen in the SO, DSD, and peak SWE maps are attributable to glaciers (persistent snow usually at high elevations), or did not receive persistent snow within the appropriate seasonal timeframe.

nugget effect τ^2 from the model fit using only elevation as a covariate was approximately 2.83; using all six covariates, $\tau^2 \approx 2.86$. This result is consistent with previous studies describing the outsized explanatory power of elevation relative to other physiographic variables for snow cover persistence (Molotch & Bales, 2006; Molotch & Meromy, 2014). Despite only a modest increase to τ^2, we choose to include all six variables within the model to account for locations where these indices have a greater influence on the SWE-timing relationship.

To understand the influence of the choice of fSCA threshold (described in Section 2.5) on model-estimated peak SWE, five unique thresholds were evaluated (1%, 5%, 10%, 20%, and 30%) relative to SNOTEL observations. In Alaska, a 5% threshold minimized percent bias (median = −11.9%, σ = 68.4% among gage locations in this region); in the Cascades a 10% threshold (percent bias of −3.1%, σ = 36.0%); in the Sierras, it was 30% (percent bias of 1.5%, σ = 19.5%); in the Northern Rockies, it was 20% (percent bias of 0.3%, σ = 30.3%); and in the Southern Rockies, a 30% threshold minimized percent bias (−0.9%, σ = 22.4%). Across the four regions within CONUS, relatively little variation in percent bias (median regional σ = 24.88%) and R^2 (median regional σ = 0.12) among fSCA thresholds was identified. However in Alaska, major differences in model performance across most gages were found (median percent bias σ = 84.12%, median R^2 σ = 0.17), largely due to indefinite boundaries between seasons leading to the rejection of varying subsets of satellite-derived data. For purposes of standardization through space, a single fSCA threshold of 10%, which resulted in the lowest overall mean absolute percent bias of 5% across all 5 regions, was selected for the final model configuration used in the remainder of this study. Kernel density distributions of percent bias and R^2 are provided in the Supporting Information Figure S3.

Maps of components of the hierarchical model, and estimated peak SWE are shown below in Figure 7, as an example for the Southern Rockies region in 2012. The spatial structure of model β values appears to closely follow the elevational gradient due to the outsized influence of elevation on the SWE-timing relationship relative to the other I_s used to construct the model variograms. Areas with snow that either does not fall

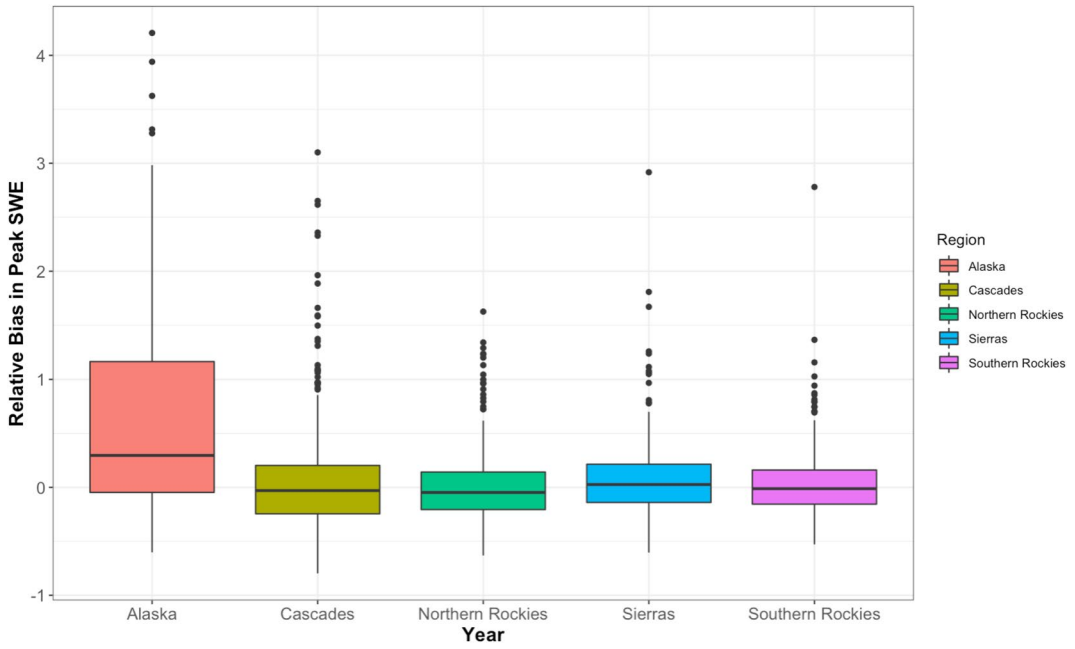


Figure 8. LOOCV percent bias between model predicted peak SWE and SNOTEL peak SWE for each of the five regions for WY2001–2019, using a MODIS fSCA threshold of 10%.

before January 1 or does not melt before October 1 are masked to avoid misleading and erroneous estimates of peak SWE from being made in locations with an unclear seasonal snow pattern.

3.4. Model Validation

LOOCV results describing the relative bias of the model-estimated peak SWE compared to SNOTEL within each region are shown below in Figure 8. Median percent bias in Alaska, the Cascades, the Northern Rockies, the Sierras, and the Southern Rockies are 14.7%, −3.1%, −0.5%, 3.2%, and −3.2%, respectively. Correlations between modeled SWE and SNOTEL observations were largely high. Among all gages and threshold values, Alaska had median $R = 0.75$, $\sigma = 0.12$; Cascades $R = 0.88$, $\sigma = 0.09$; Sierras $R = 0.96$, $\sigma = 0.05$; Northern Rockies $R = 0.86$, $\sigma = 0.06$; Southern Rockies $R = 0.88$, $\sigma = 0.11$. Across the regions in CONUS, model performance through the time period remains relatively consistent (mean $R^2 = 0.76$), suggesting that annually changing climate patterns are largely captured using snow timing variables. In Alaska, however, the agreement with SNOTEL observations varies much more widely, particularly in climatologically outlying dry years such as 2015.

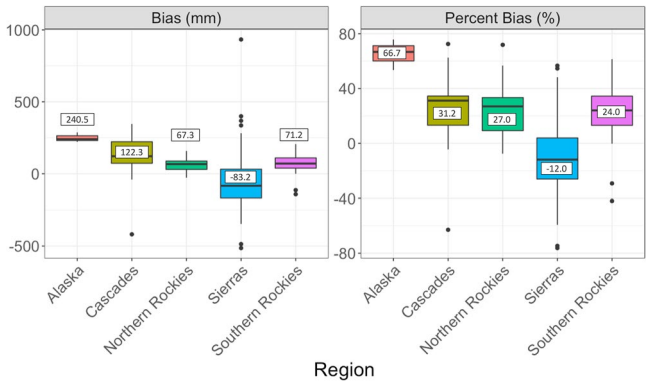


Figure 9. Distribution of model-estimated peak SWE bias against snow course survey (SCS) maximum first-of-month SWE at SCS locations within each region. Bias is shown in (a), and percent bias in (b). Median values are overlaid.

Results from comparing model-estimated peak SWE against SCS maximum first-of-month SWE are shown below in Figure 9. Because SCS data are recorded for the first of every month, rather than daily, an inherent bias between model-estimated peak SWE and SCS maximum first-of-month SWE is expected. This bias is estimated using SNOTEL data within each region, by comparing the bias between SNOTEL peak SWE and SNOTEL maximum first-of-month SWE. Across all regions, the mean expected bias between SNOTEL observations is approximately 32 mm ($\sigma = 13$ mm), and mean expected percent bias is 6.8% ($\sigma = 2.8\%$). Comparatively, the bias between model-estimated and SCS peak SWE varies more widely. Across all regions, the mean bias between model and SCS observations is approximately 12 mm ($\sigma = 184$ mm), and mean percent bias is approximately 6.2% ($\sigma = 32\%$).

4. Discussion

A retrospective analysis was conducted to diagnose the explanatory power of snow timing variables in estimating peak SWE. This work was motivated by the premise of incomplete in situ observational data for the adequate measurement of peak SWE. While SWE reconstruction methods reassemble melt energy forcings together with remotely sensed snow timing variables, the energy budget terms are rarely known without uncertainty. Snow timing, while also uncertain, has the advantage of being generally continuously observed and readily available from remote sensing sources. Snow timing metrics provide utility for estimating peak SWE in their relative parsimony, transferability, and simplicity, at the cost of poorer representation of tertiary, mechanistic physical connections present in most physical models. Given these limitations, such variables find usefulness primarily as a diagnostic tool: the spatiotemporal patterns of SWE-timing relationship strength (and its diminishment) offer insight into the strength and scale of the various meteorological and topographical conditions influencing melt and accumulation patterns.

A number of implications for SWE reconstructions and other snow models arise from this analysis. Central to many reconstructions is the relationship between the timing of complete snow ablation and the total magnitude of peak snow accumulation (Durand et al., 2008a, 2008b; Molotch, 2009), which we investigate in this analysis as DSD and peak SWE, respectively. Some local-scale processes, such as variability in wind speed and multilayer snowpack processes, are not explicitly included in reconstructions (e.g., Molotch, 2009), and generate a spatiotemporally variable field of uncertainty. Naturally, this notion extends to the limitations of other snow models in process representation, introduced by parameterization, scale uncertainties, and other factors. Our analysis seeks to provide a "diagnosis" of these uncertainties across key headwaters regions in the West, and in doing so provide important information for understanding potential locality-specific sources of model error. For example, the relatively poor SWE-timing relationship strength in the Northern Rockies suggests caution when interpreting snow reconstructions in this region. In this region, for example, the skill of reconstruction-based SWE will rely more heavily on energy-balance calculations than on the date of complete ablation. In contrast with the maritime climate of the Cascades and Sierras where SWE-timing relationship strength was stronger, winter wind speeds in the Northern Rockies are considerably higher (Klink, 1999), and wind-borne redistribution of the lower density continental snowpack may present a challenge for snow reconstructions which do not explicitly take this process into account.

The initial comparison between sites of strong and weak SWE-timing relationship strength across the western United States and Alaska (Figure 5) yields several persistent site characteristics that describe each group. Sites with relatively high SWE-timing strength differ from the population in that they are generally located nearer to the ocean (lower longitude), have a steeper slope, greater terrain roughness, and show a greater average accumulation (high ratio between peak SWE and number of days with snowfall). The majority of strong-relationship sites are found in the Sierras and other coastal locations under maritime snow regimes, which suggest that meteorology could be the primary driver of relationship strength: the shorter accumulation periods and higher maximum snow accumulations (i.e., fewer, larger snowstorms) common for these locations (Trujillo & Molotch, 2014) intuitively suggest greater homogeneity among SNOTEL sites and a greater correspondence overall between peak SWE and snow timing. By contrast, low-strength sites are most often located further inland (as seen in Figure 5), comparatively lower in elevation, and much more exposed at spatial scales on the order of 0.05°–0.1° (shown in Figure S8 in Supporting Information). These sites are typified by longer accumulation and ablation periods (slower rates of accumulation and melt); the total number of days with snowfall, days with snowfall during accumulation only, and the number of days where SWE decreases >5 mm during melt season are all relatively high compared to the population. The high number of days receiving snow may be characterized by the presence of smaller magnitude storms and/or a higher frequency of storms (a lower ratio between peak SWE and the number of days with snowfall). This finding is also consistent with Trujillo and Molotch (2014) who noted that the continental snow regimes are often characterized by lower maximum snow accumulations and longer accumulation periods, with the opposite conditions for maritime regimes. Furthermore, western mountain ranges (e.g., Cascades, Sierras) typically experience storms that come from persistent directions, that is, from the West, while the interior Rockies experience both down and upslope storms, which may strongly influence snow evolution at a point (e.g., wind redistribution, preferential deposition, saltation) and at larger scales (e.g.,

orographic effects) (Winstral et al., 2002). This may at least partially explain the overall difference in skill between SNOTEL sites in continental and maritime snow regimes.

The spatial analysis between SWE-timing strength and I_s showed that none of the terrain variables have a particularly strong correspondence with SWE-timing relationship strength, attributable to several possible explanations which may be worthwhile to explore in future studies. First, a lack of variability in terrain characteristics among SNOTEL locations diminishes any potentially strong correspondence with SWE-timing strength. Also, the location bias of SNOTEL gages tends toward open clearings, which preclude the opportunity to test influences of vegetation which have well-documented impacts on snow evolution (Hiemstra et al., 2002; Musselman et al., 2008; Rasmus et al., 2011) through canopy interception, wind blocking effects, and long-wave radiation emitted from vegetation. Second is the mismatch between process and DEM scales (Blöschl, 1999). For example, rocky outcroppings, small valleys, and other topographical complexities relevant at sub-30 m resolution (the resolution of the DEM used in this study), and by consequence the local processes they affect, would not be accounted for at the scale of the DEM. Wind advection, particularly for less dense continental snowpack, may also play a large role in determination of SWE-timing relationship strength (Winstral et al., 2002). There also appears to be a tendency for SNOTEL sites to be situated in relatively exposed locations in the Northern Rockies, as evidenced by statistically significant high exposure index values at 0.05° and 0.1° filter lengths among these sites relative to the overall population, and this exposure in turn appears to be significantly correlated with poor SWE-timing relationship strength. Other untested variables are described below.

An fSCA 10% threshold was found to minimize the average percent bias in model versus SNOTEL peak SWE across all regions, despite notable spatial variation. Large standard deviations between gage locations within regions, as well as differences between regions themselves, may reflect known local-scale variation in grid- versus point-scale representation of snow depletion (Cristea et al., 2017; Homan et al., 2011), as well as the implicit effects of fractional vegetation coverage. The issue of representativeness between point- and grid-scales has been discussed by a number of studies (Y. Liu et al., 2013, 2015; Livneh et al., 2014; Meromy et al., 2013; Molotch & Bales, 2006), but specific to this application, the role of vegetation is expected to have a large effect on point-to-grid representativeness, given SNOTEL siting within clearings surrounded by sometimes dense vegetation (Bair et al., 2016; Molotch & Bales, 2006; Raleigh et al., 2013). SNOTEL stations are also preferentially located in flat, midelevation regions of easy access, and as a result may misrepresent conditions when spatially interpolated (Lundquist et al., 2019; Wayand et al., 2013), as is done for the at-a-station coefficients and residuals developed from Equation 1 and Kriged over the MODIS grid. Further, the potentially poor representation of snow cover from the MOD10A1 product at low (fSCA <0.10) cover fractions (Rittger et al., 2013; Salomonson & Appel, 2004) may limit the detection of small persistent snow at the SNOTEL site and thus results in differing dates of meltout. The effects of vegetation on VGF and uncertainties fSCA were not directly quantified, but instead were implicitly evaluated by exploring a range of fSCA (1%–30%). Overall, this range is comparable to the satellite fSCA errors reported in observational studies: +/− 11% fSCA biases (Bair et al., 2016), and 9%–22% in meadows and 9%–37% in forests by Raleigh et al. (2013). Yet, it is possible that additional VGF and fSCA errors may exist, which are outside the scope of this study.

In contrast with bias, correlation of modeled versus SNOTEL peak SWE was relatively strong across regions and threshold values, and did not significantly vary between gages. Among all gages and threshold values, Alaska had median $R = 0.75$, $\sigma = 0.12$; Cascades $R = 0.88$, $\sigma = 0.09$; Sierras $R = 0.96$, $\sigma = 0.05$; northern Rockies $R = 0.86$, $\sigma = 0.06$; southern Rockies $R = 0.88$, $\sigma = 0.11$. This result suggests interannual spatial patterns in relative snow volume being reasonably well captured through the use of timing variables, even in the presence of threshold-dependent bias.

The diagnostic model used in this study follows a hierarchical structure, where time-varying relationships between SWE and snow timing are modeled to develop regression coefficients in the first level, and then interpolated with a spatial structure developed with I_s. This hierarchical form of analysis allows for a clearer, organized conceptualization of the multidimensional variables which influence the SWE-timing relationship in both space and time. While model performance was found to be adequate for this application, there is potential for extending this analysis into a Bayesian framework, similar to previous studies which unite remote sensing and ground-based observations (Bracken et al., 2016, 2018; Verdin et al., 2015). As

opposed to the traditional OLS-based parameter estimation used here, Bayesian kriging defines and routinely updates a set of prior distributions for parameters through a Markov Chain Monte Carlo (MCMC) simulation to produce a posterior distribution, effectively quantifying parameter uncertainty. This quantification would represent added analytical value. Additionally, outside datasets such as annual research and operational snow surveys, as well as CZO LiDAR data (Harpold et al., 2014), would provide value as additional validation of the hierarchical model, which is currently limited by greater uncertainty at unsampled locations with topographical characteristics outside those of the SNOTEL dataset used to train the model.

Overall, the SWE-timing relationship applied with derived timing variables resulted in low median peak SWE biases and high R^2 values in the LOOCV analysis relative to SNOTEL observations, yet high variability among gage-years overall (median bias = −2.4 mm, σ = 213.8 mm; median percent bias = −0.5%, σ = 33.4%; median R^2 = 0.78, σ = 0.16). The strongest model performance in the Sierras region (bias = 11.4 mm, percent bias = 2.6%, R^2 = 0.93) is likely reflective of the meteorology within the Sierras region being more seasonally consistent, as evident from the uniform ratios of timing index explanatory power among gages in the region in Figure 4, and most individual sites within the Sierra region showing a SWE-timing relationship strength in the top 10% of all gages analyzed (Figure 5). Marks and Dozier (1992) affirm the consistent dominance of solar radiation for spring snowmelt in the region leading to repeating melt patterns year over year. Conversely, performance was relatively poor in Alaska, a region at much higher latitude, with a unique climatology compared to CONUS such that it had poorer representation by the model training dataset and much sparser gage representation (only 19 gages in all of the state of Alaska contained sufficient data). This region was further beset by a high degree of cloud cover and polar darkness as was common at higher latitudes (Lindsay et al., 2015) and characterized by a heterogeneous mixture of intermittent and persistent snowpack patterns (similar to low-lying SNOTEL sites in the southwestern United States), the latter of which hindered collection of snow timing data points from MODIS for use in the model.

A wider range of biases between model-estimated peak SWE and SCS observations relative to the LOOCV SNOTEL validation was found. One source of the marginally higher bias in comparing the model to SCS data as opposed to SNOTEL is that SCS sampling locations span a wider, and often lower, elevation range than SNOTEL. The mean elevation of SNOTEL gages is approximately 2,225 m above sea level (spanning 91–3,536 m) and the mean elevation of SCS observations is 1,730 m (spanning 15–3,535 m). As evidenced by Figure 6, elevation is a significant factor in determining SWE-timing relationship strength at low strength sites. The remainder of unaccounted-for bias is likely attributable to the various factors described above which affect peak SWE-snow timing relationship skill. These appear to be relatively influential at local scales, considering the generally high median bias within each region, except the Sierras, which continues to be the region of highest model skill.

5. Conclusions

An analysis into the utility of SCA-derived snow timing variables for estimating peak SWE was presented with the goal of quantifying relationships between the hydrologically useful, yet sparsely sampled, SNOTEL peak SWE observations and more spatially continuous satellite-based SCA retrievals. A regression analysis across 611 SNOTEL locations revealed the relatively strong explanatory strength of snow timing variables (median R^2 = 0.64, median bias = −7.25 mm, median percent bias = 1.96%), in particular the DSD (median R^2 = 0.56, median bias = −8.85 mm, median percent bias = 2.35%), in explaining peak SWE variability over a given water year.

The SNOTEL sites where the explanatory relationship was strongest were characterized by proximity to the Pacific coast, a steeper average slope in the surrounding area, a greater topographical roughness, and a higher ratio between peak SWE and number of days with snowfall (i.e., larger-scale snowstorm events). Sites with a weaker relationship were generally located in the Northern Rocky Mountain region, at lower elevation, and more topographically exposed. A statistical model was implemented in five hydrologically diverse regions across the western United States and Alaska for WY2001–2019, and leveraged this relationship to understand the capability for estimating peak SWE as a function of snow timing (DSO and day of snow disappearance).

Through an LOOCV procedure, the overall performance of this simplified, observationally based model was found to be relatively strong at SNOTEL sites, with a median $R^2 = 0.78$ between model peak SWE and SNOTEL peak SWE across 118 locations within the five regions. Comparable biases were found when comparing model peak SWE to both SNOTEL (−2.4 mm, −0.5%) and SCS (39.6 mm, 8.3%) networks despite a more limited sampling rate in SCS data, indicating performance consistency at locations beyond the SNOTEL network.

At this stage, the model is limited in its application as a diagnostic tool. Future implementation of this approach would require an additional effort to investigate the effects of vegetation and other local-scale influences on the SWE-timing relationship, which may be accomplished using new and emerging datasets from, for example, LiDAR and ground-based surveys. Despite this, consistently high correlation with observations across years offers promising potential for snow timing variables to capture changing spatial patterns in peak snow (Zeng et al., 2018), critical for water resources derived from mountainous headwater regions.

Data Availability Statement

Maps of snow timing variables developed for use in the diagnostic model are archived on Zenodo and available at http://doi.org/10.5281/zenodo.4327643 (Heldmyer & Livneh, 2020). SNOTEL, SCS, and MODIS observational data used in this study are available publicly, and we refer readers to the provided citations within the Methods section.

Acknowledgments
This research is supported by funding from the Center for Water, Earth Science, and Technology (CWEST) Graduate Research Fellowship. This work utilized resources from the University of Colorado Boulder Research Computing Group, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University.

References

Anderton, S. P., White, S. M., & Alvera, B. (2004). Evaluation of spatial variability in snow water equivalent for a high mountain catchment. *Hydrological Processes*, *18*(3), 435–453. https://doi.org/10.1002/hyp.1319

Bair, E. H., Rittger, K., Davis, R. E., Painter, T. H., & Dozier, J. (2016). Validating reconstruction of snow water equivalent in California's Sierra Nevada using measurements from the NASA Airborne Snow Observatory. *Water Resources Research*, *52*(11), 8437–8460. https://doi.org/10.1002/2016WR018704

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

Blöschl, G. (1999). Scaling issues in snow hydrology. *Hydrological Processes*, *13*(14–15), 2149–2175. https://doi.org/10.1002/(sici)1099-1085(199910)13:14/15<2149::aid-hyp847>3.0.co;2-8

Bracken, C., Holman, K. D., Rajagopalan, B., & Moradkhani, H. (2018). A Bayesian hierarchical approach to multivariate nonstationary hydrologic frequency analysis. *Water Resources Research*, *54*(1), 243–255. https://doi.org/10.1002/2017WR020403

Bracken, C., Rajagopalan, B., Cheng, L., Kleiber, W., & Gangopadhyay, S. (2016). Spatial Bayesian hierarchical modeling of precipitation extremes over a large domain. *Water Resources Research*, *52*(8), 6643–6655. https://doi.org/10.1002/2016WR018768

Broxton, P. D., van Leeuwen, W. J. D., & Biederman, J. A. (2019). Improving Snow Water Equivalent Maps With Machine Learning of Snow Survey and Lidar Measurements *Water Resources Research*, *55*(5), 3739–3757. https://doi.org/10.1029/2018WR024146

Broxton, P. D., Zeng, X., & Dawson, N. (2016). Why do global reanalyses and land data assimilation products underestimate snow water equivalent? *Journal of Hydrometeorology*, *17*(11), 2743–2761. https://doi.org/10.1175/JHM-D-16-0056.1

Cayan, D. R., Dettinger, M. D., Kammerdiener, S. A., Caprio, J. M., & Peterson, D. H. (2001). Changes in the onset of spring in the Western United States. *Bulletin of the American Meteorological Society*, *82*(3), 399–415. https://doi.org/10.1175/1520-0477(2001)082<0399:CITOOS>2.3.CO;2

Cho, E., Tuttle, S. E., & Jacobs, J. M. (2017). Evaluating consistency of snow water equivalent retrievals from passive microwave sensors over the North Central U. S.: SSM/I vs. SSMIS and AMSR-E vs. AMSR2. *Remote Sensing*, *9*(5), 465. https://doi.org/10.3390/rs9050465

Cline, D. W., Bales, R. C., & Dozier, J. (1998). Estimating the spatial distribution of snow in mountain basins using remote sensing and energy balance modeling. *Water Resources Research*, *34*(5), 1275–1285. https://doi.org/10.1029/97WR03755

Cristea, N. C., Breckheimer, I., Raleigh, M. S., HilleRisLambers, J., & Lundquist, J. D. (2017). An evaluation of terrain-based downscaling of fractional snow covered area data sets based on LiDAR-derived snow data and orthoimagery. *Water Resources Research*, *53*(8), 6802–6820. https://doi.org/10.1002/2017WR020799

Dawson, N., Broxton, P., & Zeng, X. (2018). Evaluation of remotely sensed snow water equivalent and snow cover extent over the contiguous United States. *Journal of Hydrometeorology*, *19*(11), 1777–1791. https://doi.org/10.1175/JHM-D-18-0007.1

Déry, S. J., Salomonson, V. V., Stieglitz, M., Hall, D. K., & Appel, I. (2005). An approach to using snow areal depletion curves inferred from MODIS and its application to land surface modelling in Alaska. *Hydrological Processes*, *19*(14), 2755–2774. https://doi.org/10.1002/hyp.5784

Dressler, K. A., Fassnacht, S. R., & Bales, R. C. (2006). A comparison of snow telemetry and snow course measurements in the Colorado River Basin. *Journal of Hydrometeorology*, *7*(4), 705–712. https://doi.org/10.1175/JHM506.1

Durand, M., Molotch, N. P., & Margulis, S. A. (2008a). A Bayesian approach to snow water equivalent reconstruction. *Journal of Geophysical Research*, *113*(D20). https://doi.org/10.1029/2008JD009894

Durand, M., Molotch, N. P., & Margulis, S. A. (2008b). Merging complementary remote sensing datasets in the context of snow water equivalent reconstruction. *Remote Sensing of Environment*, *112*(3), 1212–1225. https://doi.org/10.1016/j.rse.2007.08.010

Fassnacht, S. R., Dressler, K. A., & Bales, R. C. (2003). Snow water equivalent interpolation for the Colorado River Basin from snow telemetry (SNOTEL) data. *Water Resources Research*, *39*(8). https://doi.org/10.1029/2002WR001512

Fassnacht, S. R., & López-Moreno, J. I. (2020). Patterns of trends in niveograph characteristics across the western United States from snow telemetry data. *Frontiers of Earth Science*, *14*(2), 315–325. https://doi.org/10.1007/s11707-020-0813-5

Gafurov, A., Kriegel, D., Vorogushyn, S., & Merz, B. (2013). Evaluation of remotely sensed snow cover product in Central Asia. *Hydrology Research*, *44*(3), 506–522. https://doi.org/10.2166/nh.2012.094

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

Hall, D. K., & Riggs, A. (2016). MODIS/Terra snow cover daily L3 global 500m SIN grid, version 6. NASA National Snow and Ice Data Center Distributed. *Active Archive Center*. https://doi.org/10.5067/MODIS/MOD10A1.006

Harpold, A. A., Guo, Q., Molotch, N., Brooks, P. D., Bales, R., Fernandez-Diaz, J. C., et al. (2014). LiDAR-derived snowpack data sets from mixed conifer forests across the Western United States. *Water Resources Research*, *50*(3), 2749–2755. https://doi.org/10.1002/2013WR013935

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

Heldmyer, A., & Livneh, B. (2021). *Annual snow timing index rasters for the Western US and Alaska, WY2001-2019 [Data set]*. Zenodo.

Hiemstra, C. A., Liston, G. E., & Reiners, W. A. (2002). Snow redistribution by wind and interactions with vegetation at upper treeline in the medicine bow mountains, Wyoming, U.S.A. *Arctic, Antarctic, and Alpine Research*, *34*(3), 262–273. https://doi.org/10.1080/15230430.2002.12003493

Hijmans, R. (2020). Raster: Geographic data analysis and modeling (version R package version 3.0-12). Retrieved from https://CRAN.R-project.org/package=raster

Homan, J. W., Luce, C. H., McNamara, J. P., & Glenn, N. F. (2011). Improvement of distributed snowmelt energy balance modeling with MODIS-based NDSI-derived fractional snow-covered area data. *Hydrological Processes*, *25*(4), 650–660. https://doi.org/10.1002/hyp.7857

Klink, K. (1999). Climatological mean and interannual variance of United States surface wind speed, direction and velocity. *International Journal of Climatology*, *19*(5), 471–88. https://doi.org/10.1002/(SICI)1097-0088(199904)19:5<471::AID-JOC367>3.0.CO;2-X

Knowles, N., & Cayan, D. R. (2004). Elevational dependence of projected hydrologic changes in the San Francisco estuary and watershed. *Climatic Change*, *62*(1), 319–336. https://doi.org/10.1023/B:CLIM.0000013696.14308.b9

Langlois, A., Royer, A., Fillol, E., Frigon, A., & Laprise, R. (2004). Evaluation of the snow cover variation in the Canadian Regional Climate Model over eastern Canada using passive microwave satellite data. *Hydrological Processes*, *18*(6), 1127–1138. https://doi.org/10.1002/hyp.5514

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

Lindsay, C., Zhu, J., Miller, A., Kirchner, P., & Wilson, T. (2015). Deriving snow cover metrics for Alaska from MODIS. *Remote Sensing*, *7*(10), 12961–12985. https://doi.org/10.3390/rs71012961

Liu, J., Woodcock, C. E., Melloh, R. A., Davis, R. E., McKenzie, C., & Painter, T. H. (2008). Modeling the view angle dependence of gap fractions in forest canopies: Implications for mapping fractional snow cover using optical remote sensing. *Journal of Hydrometeorology*, *9*(5), 1005–1019. https://doi.org/10.1175/2008JHM866.1

Liu, Y., Peters-Lidard, C. D., Kumar, S., Foster, J. L., Shaw, M., Tian, Y., & Fall, G. M. (2013). Assimilating satellite-based snow depth and snow cover products for improving snow predictions in Alaska. *Advances in Water Resources*, *54*, 208–227. https://doi.org/10.1016/j.advwatres.2013.02.005

Liu, Y., Peters-Lidard, C. D., Kumar, S. V., Arsenault, K. R., & Mocko, D. M. (2015). Blending satellite-based snow depth products with in situ observations for streamflow predictions in the Upper Colorado River Basin. *Water Resources Research*, *51*(2), 1182–1202. https://doi.org/10.1002/2014WR016606

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

Lundquist, J., Hughes, M., Gutmann, E., & Kapnick, S. (2019). Our skill in modeling mountain rain and snow is bypassing the skill of our observational networks. *Bulletin of the American Meteorological Society*, *100*(12), 2473–2490. https://doi.org/10.1175/BAMS-D-19-0001.1

Lute, A. C., & Luce, C. H. (2017). Are model transferability and complexity antithetical? Insights from validation of a variable-complexity empirical snow model in space and time. *Water Resources Research*, *53*(11), 8825–8850. https://doi.org/10.1002/2017WR020752

Margulis, S. A., Cortés, G., Girotto, M., & Durand, M. (2016). A Landsat-Era Sierra Nevada Snow Reanalysis (1985–2015). *Journal of Hydrometeorology*, *17*(4), (1203–1221). JSTOR. https://doi.org/10.1175/jhm-d-15-0177.1

Marks, D., Domingo, J., Susong, D., Link, T., & Garen, D. (1999). A spatially distributed energy balance snowmelt model for application in mountain basins. *Hydrological Processes*, *13*(12–13), 1935–1959. https://doi.org/10.1002/(SICI)1099-1085(199909)13:12/13<1935::AID-HYP868>3.0.CO;2-C

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

Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P., & Nijssen, B. (2002). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *Journal of Climate*, *15*(22), 3237–3251. https://doi.org/10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2

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

Molotch, N. P. (2009). Reconstructing snow water equivalent in the Rio Grande headwaters using remotely sensed snow cover data and a spatially distributed snowmelt model. *Hydrological Processes*, *23*(7), 1076–1089. https://doi.org/10.1002/hyp.7206

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

Molotch, N. P., & Margulis, S. A. (2008). 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*, *31*(11), 1503–1514. https://doi.org/10.1016/j.advwatres.2008.07.017

Molotch, N. P., & Meromy, L. (2014). Physiographic and climatic controls on snow cover persistence in the Sierra Nevada Mountains. *Hydrological Processes*, *28*(16), 4573–4586. https://doi.org/10.1002/hyp.10254

Mote, P. W., Hamlet, A. F., Clark, M. P., & Lettenmaier, D. P. (2005). Declining mountain snowpack in western North America. *Bulletin of the American Meteorological Society*, *86*(1), 39–50. https://doi.org/10.1175/BAMS-86-1-39

Musselman, K. N., Molotch, N. P., & Brooks, P. D. (2008). Effects of vegetation on snow accumulation and ablation in a mid-latitude sub-alpine forest. *Hydrological Processes*, *22*(15), 2767–2776. https://doi.org/10.1002/hyp.7050

NASA, LP DAAC. (2013). *Land Process Distributed Active Archive Center*. NASA Shuttle Radar Topography Mission (SRTM) Version 3.0 (SRTM Plus) Product Release. Retrieved from https://lpdaac.usgs.gov/

Painter, T. H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R. E., & Dozier, J. (2009). Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. *Remote Sensing of Environment*, *113*(4), 868–79. https://doi.org/10.1016/j.rse.2009.01.001

Palmer, P. (1988). The SCS snow survey water supply forecasting program: Current operations and future directions. *Proceedings of the 56th Annual Western Snow Conference*, 43–51.

Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computers and Geosciences*, *30*(7), 683–691. https://doi.org/10.1016/j.cageo.2004.03.012

Pflug, J. M., & Lundquist, J. D. (2020). Inferring distributed snow depth by leveraging snow pattern repeatability: Investigation using 47 lidar observations in the Tuolumne Watershed, Sierra Nevada, California. *Water Resources Research*, *56*(9), e2020WR027243. https://doi.org/10.1029/2020WR027243

Raleigh, M. S., Rittger, K., Moore, C. E., Henn, B., Lutz, J. A., & Lundquist, J. D. (2013). Ground-based testing of MODIS fractional snow cover in subalpine meadows and forests of the Sierra Nevada. *Remote Sensing of Environment*, *128*, 44–57. https://doi.org/10.1016/j.rse.2012.09.016

Rasmus, S., Lundell, R., & Saarinen, T. (2011). Interactions between snow, canopy, and vegetation in a boreal coniferous forest. *Plant Ecology & Diversity*, *4*(1), 55–65. https://doi.org/10.1080/17550874.2011.558126

Rittger, K., Painter, T. H., & Dozier, J. (2013). Assessment of methods for mapping snow cover from MODIS. *Advances in Water Resources*, *51*, 367–380. https://doi.org/10.1016/j.advwatres.2012.03.002

Salomonson, V. V., & Appel, I. (2004). Estimating fractional snow cover from MODIS using the normalized difference snow index. *Remote Sensing of Environment*, *89*(3), 351–360. https://doi.org/10.1016/j.rse.2003.10.016

Schneider, D., & Molotch, N. P. (2016). Real-time estimation of snow water equivalent in the Upper Colorado River Basin using MODIS-based SWE Reconstructions and SNOTEL data. *Water Resources Research*, *52*(10), 7892–7910. https://doi.org/10.1002/2016WR019067

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

Slater, A. G., Barrett, A. P., Clark, M. P., Lundquist, J. D., & Raleigh, M. S. (2013). Uncertainty in seasonal snow reconstruction: Relative impacts of model forcing and image availability. *Advances in Water Resources*, *55*, 165–177. https://doi.org/10.1016/j.advwatres.2012.07.006

Sturm, M., Taras, B., Liston, G. E., Derksen, C., Jonas, T., & Lea, J. (2010). Estimating snow water equivalent using snow depth data and climate classes. *Journal of Hydrometeorology*, *11*(6), 1380–1394. https://doi.org/10.1175/2010JHM1202.1

Suchetana, B., Rajagopalan, B., & Silverstein, J. (2016). Hierarchical modeling approach to evaluate spatial and temporal variability of wastewater treatment compliance with biochemical oxygen demand, total suspended solids, and ammonia limits in the United States. *Environmental Engineering Science*, *33*(7), 514–524. https://doi.org/10.1089/ees.2016.0116

Trujillo, E., & Molotch, N. P. (2014). Snowpack regimes of the Western United States. *Water Resources Research*, *50*(7), 5611–5623. https://doi.org/10.1002/2013WR014753

Verdin, A., Rajagopalan, B., Kleiber, W., & Funk, C. (2015). A Bayesian kriging approach for blending satellite and ground precipitation observations. *Water Resources Research*, *51*(2), 908–921. https://doi.org/10.1002/2014WR015963

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

Vuyovich, C. M., Jacobs, J. M., & Daly, S. F. (2014). Comparison of passive microwave and modeled estimates of total watershed SWE in the continental United States. *Water Resources Research*, *50*(11), 9088–9102. https://doi.org/10.1002/2013WR014734

Wayand, N. E., Hamlet, A. F., Hughes, M., Feld, S. I., & Lundquist, J. D. (2013). Intercomparison of meteorological forcing data from empirical and mesoscale model sources in the North Fork American River Basin in Northern Sierra Nevada, California. *Journal of Hydrometeorology*, *14*(3), 677–699. https://doi.org/10.1175/JHM-D-12-0102.1

Winstral, A., Elder, K., & Davis, R. E. (2002). Spatial snow modeling of wind-redistributed snow using terrain-based parameters. *Journal of Hydrometeorology*, *3*(5), 524–538. https://doi.org/10.1175/1525-7541(2002)003<0524:SSMOWR>2.0.CO;2

Zeng, X., Broxton, P., & Dawson, N. (2018). Snowpack change from 1982 to 2016 over conterminous United States. *Geophysical Research Letters*, *45*(23), 12940–12947. https://doi.org/10.1029/2018GL079621

Zhang, Y., & Ma, N. (2018). Spatiotemporal variability of snow cover and snow water equivalent in the last three decades over Eurasia. *Journal of Hydrology*, *559*, 238–251. https://doi.org/10.1016/j.jhydrol.2018.02.031

Reference From the Supporting Information

Dozier, J., Painter, T. H., Rittger, K., & Frew, J. E. (2008). Time-Space Continuity of Daily Maps of Fractional Snow Cover and Albedo from MODIS. *Advances in Water Resources*, Hydrologic Remote Sensing, *31*(11), 1515–26. https://doi.org/10.1016/j.advwatres.2008.08.011