Modeling Spatial Distribution of Snow Water Equivalent by Combining Meteorological and Satellite Data with Lidar Maps

UTKARSH MITAL,[a] DIPANKAR DWIVEDI,[a] ILHAN ÖZGEN-XIAN,[a,b] JAMES B. BROWN,[c,d] AND CARL I. STEEFEL[a]

[a] *Energy Geosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California*
[b] *Institute of Geoecology, Technische Universität Braunschweig, Braunschweig, Germany*
[c] *Environmental Genomics and System Biology, Lawrence Berkeley National Laboratory, Berkeley, California*
[d] *Department of Statistics, University of California, Berkeley, Berkeley, California*

(Manuscript received 11 February 2022, in final form 22 July 2022)

ABSTRACT: An accurate characterization of the water content of snowpack, or snow water equivalent (SWE), is necessary to quantify water availability and constrain hydrologic and land surface models. Recently, airborne observations (e.g., lidar) have emerged as a promising method to accurately quantify SWE at high resolutions (scales of ~100 m and finer). However, the frequency of these observations is very low, typically once or twice per season in the Rocky Mountains of Colorado. Here, we present a machine learning framework that is based on random forests to model temporally sparse lidar-derived SWE, enabling estimation of SWE at unmapped time points. We approximated the physical processes governing snow accumulation and melt as well as snow characteristics by obtaining 15 different variables from gridded estimates of precipitation, temperature, surface reflectance, elevation, and canopy. Results showed that, in the Rocky Mountains of Colorado, our framework is capable of modeling SWE with a higher accuracy when compared with estimates generated by the Snow Data Assimilation System (SNODAS). The mean value of the coefficient of determination R^2 using our approach was 0.57, and the root-mean-square error (RMSE) was 13 cm, which was a significant improvement over SNODAS (mean $R^2 = 0.13$; RMSE = 20 cm). We explored the relative importance of the input variables and observed that, at the spatial resolution of 800 m, meteorological variables are more important drivers of predictive accuracy than surface variables that characterize the properties of snow on the ground. This research provides a framework to expand the applicability of lidar-derived SWE to unmapped time points.

SIGNIFICANCE STATEMENT: Snowpack is the main source of freshwater for close to 2 billion people globally and needs to be estimated accurately. Mountainous snowpack is highly variable and is challenging to quantify. Recently, lidar technology has been employed to observe snow in great detail, but it is costly and can only be used sparingly. To counter that, we use machine learning to estimate snowpack when lidar data are not available. We approximate the processes that govern snowpack by incorporating meteorological and satellite data. We found that variables associated with precipitation and temperature have more predictive power than variables that characterize snowpack properties. Our work helps to improve snowpack estimation, which is critical for sustainable management of water resources.

KEYWORDS: Snow cover; Decision trees; Machine learning; Regression

1. Introduction

In vast areas of the world, close to 40% of the annual precipitation falls as snow (Sturm et al. 2017). Snowpack acts as a natural reservoir, providing flood control and storage by capturing precipitation in the cold months and releasing snowmelt in the warm months. The released snowmelt is the primary source of freshwater for close to 2 billion people worldwide (Mankin et al. 2015). The water content of a snowpack is commonly determined by the snow water equivalent (SWE), which is defined as the equivalent amount of water that would result if the snowpack were completely melted.

The spatial distribution of SWE is critical for understanding the timing and magnitude of snowmelt runoff (Luce et al. 1998). Characterizing the spatial distribution of SWE is also necessary to constrain and evaluate hydrologic and land surface models (Clark et al. 2011; Shuai et al. 2022) and to understand how shifting weather patterns and climate extremes affect snowpack and water quality (e.g., Chen et al. 2020). Since SWE is highly heterogeneous across the landscape (e.g., Lehning et al. 2011), it is difficult for field observations to characterize the spatial distribution of SWE. This has motivated the use of remote sensing approaches, which have been monitoring the snowpack for several decades (Peck et al. 1980). However, satellite estimates of mountain snowpack are either too coarse (~25-km resolution), or provide information about the snow cover area, which is an indirect measure of SWE (Brown et al. 2003; Dozier et al. 2016). Recently, airborne observations have emerged as a promising method to estimate SWE at resolutions of ~100 m and finer (e.g., Bühler et al. 2015; Kim et al. 2019; Painter et al. 2016). A prominent example is the Airborne Snow Observatory (ASO), which uses lidar to estimate SWE for drainage basins in the western United States (Painter et al. 2016).

Airborne observations have a high cost of data acquisition that constrains their spatial extent and revisit times (Ferraz et al. 2018; Oaida et al. 2019). The infrequent nature of

Corresponding author: Utkarsh Mital, umital@lbl.gov

DOI: 10.1175/AIES-D-22-0010.1 e220010

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

airborne observations has prompted some studies to use this data purely for model validation (e.g., Bair et al. 2018; Behrangi et al. 2018; Oaida et al. 2019). Despite the availability of only a few snapshots, airborne observations provide access to rich information about the spatial distribution of SWE. This information presents an opportunity to facilitate the development of a new generation of modeling approaches that seek to estimate SWE at unmapped time points.

Past studies that have sought to estimate SWE at unmapped time points fall under two broad categories: data assimilation (DA) and machine learning (ML) techniques. DA techniques introduce a coupling between statistical approaches and process-based models to reduce the uncertainty associated with pure process-based approaches (Largeron et al. 2020). A prominent dataset using this approach has been generated by the Snow Data Assimilation System (SNODAS), which combines forecast models, satellite data and in situ observations to generate daily estimates of snowpack (Carroll et al. 2001). However, SNODAS is yet to integrate high-resolution lidar data in its estimates due to which it lacks the detail to accurately model spatial distribution of SWE at high elevations (Clow et al. 2012). Other DA approaches seek to blend process-based models and/or satellite estimates with in situ observations using an optimal interpolation approach (Brasnett 1999; de Rosnay et al. 2014; Liu et al. 2015; Kongoli et al. 2019; Gan et al. 2021). Recently, DA approaches have been developed that incorporate lidar data in their workflow. These include the work of Hedrick et al. (2018), who inserted lidar-derived snow depth values into a process-based snow model to improve estimates of mountain snowpack in between airborne observations. Margulis et al. (2019) coupled a batch smoother with a land surface model to transform intermittent snow-depth images into space–time-continuous estimates of SWE. Malek et al. (2020) blended lidar data with in situ observations using ensemble optimal interpolation. Generally, the reliance of DA techniques on process-based models causes them to be (i) computationally expensive and (ii) influenced by prior assumptions made about the description of physical processes governing the spatial distribution of snowpack. If they do not incorporate process-based models, DA techniques resemble statistical approaches that, although useful, provide limited insights on the processes governing snowpack.

ML techniques, on the other hand, are computationally efficient and learn directly from snowpack data without needing to explicitly describe the governing physics. ML techniques also have the ability to provide important insights on the physical processes governing snowpack (Anderton et al. 2004; Fassnacht et al. 2003; Balk and Elder 2000). By leveraging the interannual consistency in the spatial distribution of snowpack (e.g., Deems et al. 2008; Erickson et al. 2005; Schirmer et al. 2011), recent ML studies have extracted spatial patterns of SWE from historical maps (Schneider and Molotch 2016; Zheng et al. 2018). These patterns are then used as predictors for interpolating in situ SWE measurements to estimate the spatial distribution of SWE. Studies (both ML- and DA-based) have demonstrated that the accuracy of SWE estimates improves when spatial patterns of SWE are extracted from historical lidar observations (Zheng et al. 2018; Malek et al. 2020). However, the success of current approaches (ML-based or otherwise) that seek to use lidar data as predictors to interpolate in situ observations hinges on the availability of a large database of historical lidar-derived maps of SWE, along with a dense network of in situ measurements to accurately estimate spatial distribution of SWE.

In this study, we consider a scenario where both the database of lidar-derived SWE as well as the network of in situ SWE measurements is sparse. This data sparsity not only makes it difficult to use historical lidar maps as predictors of SWE distribution at unmapped time points, but also makes spatial interpolation of point measurements challenging. In particular, our study area encompasses the Rocky Mountains in Colorado, where the frequency of lidar-derived observations is either once per season, or no more than once every two months. Furthermore, three out of five basins considered in our study area have fewer than two in situ SWE measurements. We seek to compensate for this data sparsity by learning relationships between lidar-derived SWE and gridded estimates of 15 predictor variables that are known to drive and capture snowpack variability. We consider meteorological variables (precipitation and temperature), topographic and vegetation variables, as well as surface reflectance variables. A recent study by Cartwright et al. (2022) used an ML framework to model snowpack as a function of topography and canopy in southwest Alberta. Their study investigated if strategically acquired lidar transects can serve as viable substitutes for wall-to-wall snowpack estimates, which can potentially aid in developing a time- and cost-efficient flight path for lidar data acquisition. In our study, we seek to investigate if existing lidar maps can serve as viable substitutes for snowpack at different time points.

Our central objective is to model the spatial distribution of SWE at unmapped time points. We do so by using an ML framework based on random forests (RF), which have been widely adopted in hydrological applications (e.g., Dwivedi et al. 2022; Mital et al. 2020; Wainwright et al. 2020; Z. Wang et al. 2015; Cartwright et al. 2022). We break down our central objective as follows: (i) What variables can help us to capture the processes of snow accumulation and melt? (ii) Can we train a model on a sparse dataset of lidar-derived maps of SWE that uses the above variables as predictors? (iii) How does the trained model compare with SNODAS (which is widely viewed as a reliable product)? (iv) Are there certain processes or characteristics that are more important to model SWE?

The remainder of the paper is organized as follows: We start by outlining the study area and the available lidar data and in situ measurements. We then summarize the various datasets from which we obtain and derive the predictor variables. We then present our ML framework and contrast it with SNODAS and simple interpolation of in situ measurements. This is followed by a description of the results in which we compare SWE estimates obtained using different approaches and investigate the relative importance of different input variables. We then discuss the implications and limitations of our results and present some conclusions.

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

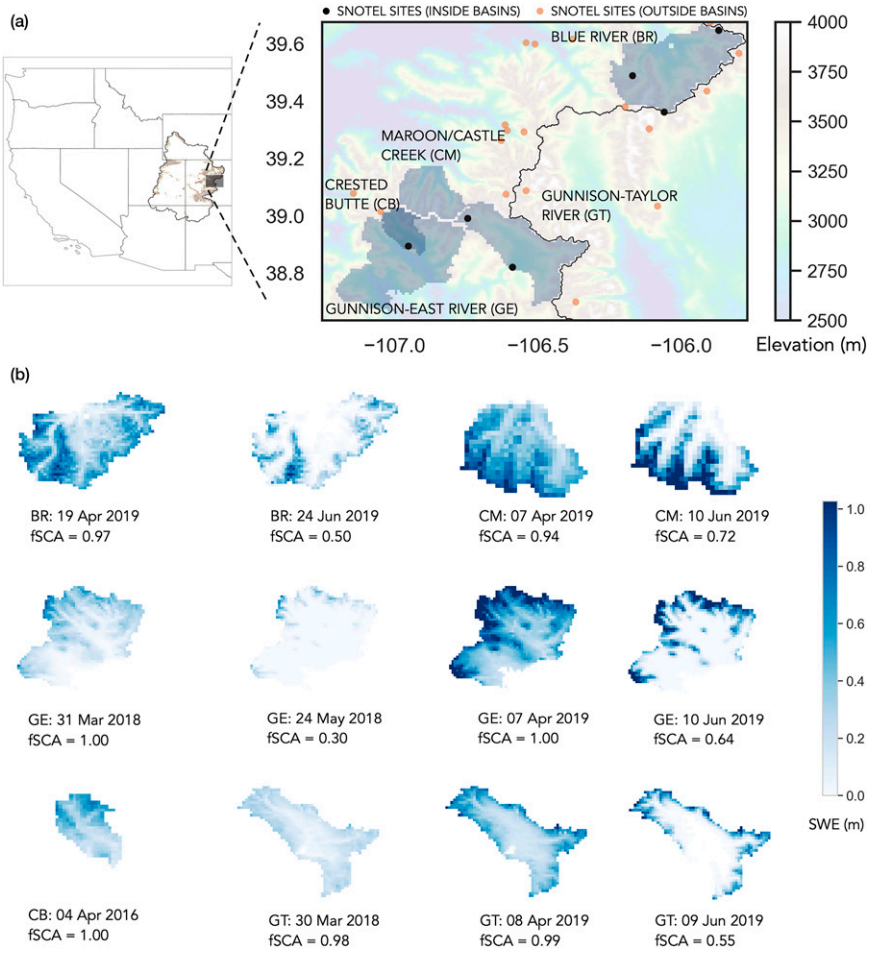

FIG. 1. Lidar-derived SWE maps within UCRB obtained via ASO: (a) spatial extent of the basins covered by the maps along with SNOTEL stations, and (b) actual maps (upscaled to 800-m resolution). Each map is labeled by its basin, date of acquisition, and fraction of snow-covered area at 50-m resolution.

2. Data

a. Study area and lidar data

We conducted this study in the Rocky Mountains that are part of the Upper Colorado River basin (UCRB). UCRB drains into the Colorado River, which is the principal source of water in the southwestern United States (James et al. 2014). We obtained lidar-derived maps of SWE generated by the ASO at a spatial resolution of 50 m (Painter 2018; https://nsidc.org/data/ASO_50M_SWE/). ASO derives maps of SWE by combining lidar estimates of snow depth with simulations of snow density distribution (Painter et al. 2016). These maps are accurate proxies of SWE and, therefore, are used as ground truth SWE values.

Across the UCRB, ASO maps quantify SWE across Blue River (BR), Crested Butte (CB), Maroon/Castle Creek (CM), Gunnison–East River (GE), and Gunnison–Taylor River (GT) basins. These basins are located on the eastern edge of UCRB and were mapped in 2018 and 2019 (with one exception for Crested Butte; Fig. 1). There are seven unique maps corresponding to the early-melt period (March/April); five of these have a corresponding map obtained approximately 2 months later during the late-melt period (May/June), for a total of 12 maps. For each map, we computed the fraction of snow-covered area (fSCA) by obtaining the number of snow-covered pixels (where SWE $>$ 0) and normalizing them by the total number of pixels (Fig. 1). Maps corresponding to the late-melt period have a patchy snow cover, with fSCA values significantly lower than 1. The fSCA computations were carried out at the original 50-m resolution of the SWE maps. We clarify that the fSCA values are used only for labeling the various maps and are not used in our modeling framework.

Figure 1 also shows the locations of Snowpack Telemetry (SNOTEL) stations that provide daily in situ measurements of SWE (available from https://www.wcc.nrcs.usda.gov). There are 24 stations within the displayed bounding box (−107.25°E, 38.63°N, −105.77°E, 39.68°N). However, there are only six stations that overlap with the entire area mapped by ASO snapshots. Blue River basin has the highest overlap

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

with three stations, while Maroon/Castle Creek basin has no overlapping stations. This extreme sparsity of in situ measurements, coupled with sparse lidar data, suggests that modeling SWE by blending in situ measurements and lidar data is not a viable option for our study area.

b. *Quality control on ASO data*

The maps corresponding to Gunnison–East River were subjected to additional quality control measures. A visualization of the raw maps revealed a number of pixels that appeared to be numerical artifacts that had to be manually removed. In addition, the spatial extent of these maps was not consistent across snapshots. By spatial extent, we refer to the spatial extent of the basin that includes both snow-covered and snow-free pixels. Therefore, we compared the various GE maps and manually masked pixels near the boundary to ensure a consistent spatial extent. This additional quality control was carried out at 800-m resolution.

Further, the original dataset contains an additional map in the Crested Butte basin (CB: 30 March 2018). We decided to exclude that map from our study, since the map labeled GE: 31 March 2018 encapsulates the spatial extent of Crested Butte and is only one day apart from the excluded map.

c. *Datasets for modeling SWE*

To model SWE at unmapped time points, we considered several modeled and observed variables. The datasets underlying the variables are described in the following sections.

1) METEOROLOGICAL DATA

We obtained gridded estimates of daily precipitation and temperature from the Parameter–Elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 2008). PRISM data at 800-m spatial resolution are a proprietary dataset and were purchased from the PRISM Climate Group at Oregon State University (https://prism.oregonstate.edu, created 3 August 2020).

2) TOPOGRAPHY AND VEGETATION DATA

We obtained elevation maps from the National Elevation Dataset (NED; U.S. Geological Survey 2019; https://apps.nationalmap.gov) at a spatial resolution of 10 m. We also obtained tree canopy maps from the National Land Cover Database (NLCD; https://www.mrlc.gov/data) at a spatial resolution of 30 m. Tree canopy maps estimate percent tree canopy for a given pixel and are generated by the U.S. Forest Service.

3) SURFACE REFLECTANCE DATA

We also obtained the Moderate Resolution Imaging Spectroradiometer (MODIS) *Terra* surface reflectance data (MOD09GA, version 6; Vermote and Wolfe 2015). In particular, we obtained bands 1–7, which correspond to land properties. Daily estimates of surface reflectance in these bands are available at a spatial resolution of 500 m and were downloaded from NASA's Earth Science Data Systems (https://earthdata.nasa.gov).

4) IN SITU SWE MEASUREMENTS

We obtained in situ SWE observations from SNOTEL stations inside the mapped basins (six SNOTEL stations; https://www.wcc.nrcs.usda.gov). We acquired SWE data corresponding to dates of ASO snapshots.

5) SNODAS DATA

We also obtained gridded model estimates of daily SWE at a spatial resolution of 1 km from the SNODAS dataset (National Operational Hydrologic Remote Sensing Center 2004; https://nsidc.org/data/G02158/versions/1). SNODAS integrates satellite data and in situ SWE measurements with physically based model estimates of snow cover (Carroll et al. 2001). It ingests output from numerical weather prediction models and attempts to capture the dynamics of snowpack across the conterminous United States (Barrett 2003). We use the SNODAS dataset to compare the performance of our models. Although other data products exist that also provide daily estimates of SWE in upper Colorado during the time period of our study [e.g., Gan et al. (2021) use optimal interpolation to generate estimates at 0.125° resolution], we use SNODAS since it is widely viewed as an accurate product with high-resolution daily estimates.

6) DATASET REPROJECTION

We reprojected all the gridded datasets described above to NAD83 (EPSG:4269) to have a consistent coordinate system. To make the spatial resolutions of various datasets consistent, we used bilinear interpolation to resample the ASO, elevation, canopy, MODIS, and SNODAS datasets to a resolution of 800 m. This resolution was chosen to ensure consistency with the PRISM grid.

3. Methods

Our first objective is concerned with determining appropriate variables to help capture snow accumulation and melt processes. We achieve that by obtaining 15 different predictor variables from the above datasets (sections 3a–3c). Our second objective involves using the above variables as predictors to model SWE. We investigated this using random forests (section 3d) and compared the model performance with SNODAS and spatial interpolation of in situ measurements (section 3e; third objective). Our final objective is concerned with identifying if certain sets of variables are more important than others. We investigated this by training three different ML models, each of which considered a different class of model inputs (section 3f).

a. *Meteorological variables*

Precipitation and temperature are strongly correlated to snowpack variability (Broxton et al. 2016; Hamlet et al. 2005; Scalzitti et al. 2016). We use gridded estimates of precipitation and temperature to extract variables that are more closely associated with SWE—snowfall and positive degree-days (PDD).

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

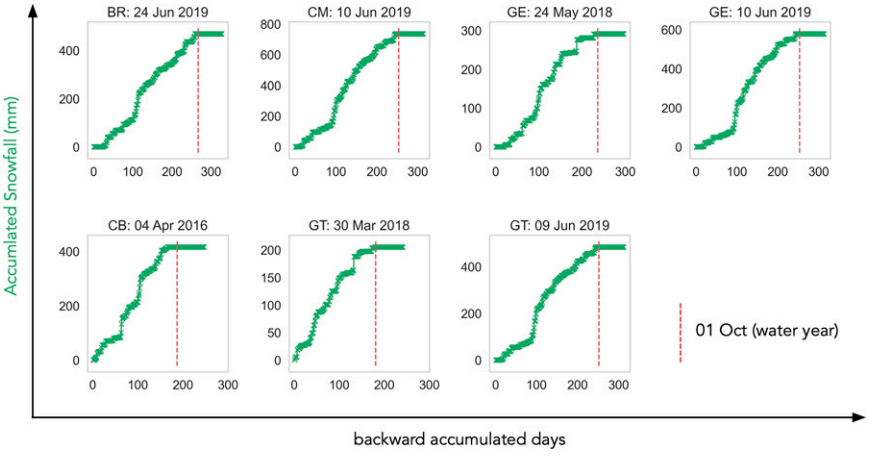

FIG. 2. Accumulated snowfall (averaged over the spatial extent of an ASO snapshot) as a function of backward accumulated days. This corresponds to backward time accumulation of snowfall, starting from the date of an ASO snapshot and accumulating backward in time to encompass the entire snow season. The dotted line corresponds to 1 Oct of the water year and approximately corresponds to the date when backward time accumulation of snowfall levels out. For brevity, plots for snapshots corresponding to the early-melt period are not shown if they have a corresponding late-melt snapshot. Those plots are presented in appendix Fig. A1.

1) ACCUMULATED SNOWFALL

Snowfall, rather than precipitation (which includes rainfall), is the primary mechanism behind snow accumulation. We considered precipitation on a given day to be snow if the mean air temperature was less than or equal to 0°C. Although this threshold corresponds to the freezing point of water, studies have shown that other factors (such as relative humidity) can cause precipitation to comprise both rain and snow for a wide range of air temperatures (Harpold et al. 2017; Jennings and Molotch 2019; Marks et al. 2013). In our study, we found that considering different temperature thresholds had a minimal effect on our results. As partitioning thresholds suggested in literature [e.g., Jennings et al. (2018) suggested 3.8°C] are largely an outcome of empirical studies that ingest measurements at low elevations, we chose the 0°C air temperature threshold to retain objectivity.

Furthermore, since snow accumulation takes place over the entire snow season, we did a backward time accumulation of snowfall. We started from the date of the ASO snapshot and accumulated snowfall backward in time to encompass the entire snow season (Fig. 2, along with appendix Fig. A1). We see that for all cases, accumulated snowfall continues to increase as the number of backward accumulated days increases, and eventually levels out around 1 October of the water year. Therefore, we considered accumulated snowfall starting from 1 October.

2) PDD SUM

PDD sum is the sum of mean daily temperatures above 0°C in a given time period. It constitutes a convenient variable to approximate melt processes. Its use is motivated by the development of degree-day models that relate the amount of snowmelt during a time period to the PDD sum for the same period (Hock 2003; Ohmura 2001). Snowmelt in midlatitude mountains is primarily driven by net radiation at the surface (Bales et al. 2006; Marks and Dozier 1992; Mott et al. 2018). An accurate evaluation of radiation-driven melt processes requires an energy balance approach that takes into account net radiation (shortwave and longwave) as well as various heat fluxes (Mott et al. 2018). Gridded datasets of various components of the energy balance equation are typically unavailable at high spatial resolutions (i.e., finer than 1 km), as opposed to gridded datasets of air temperature, which tend to be widely available. Since air temperature is highly correlated with several energy balance components (Hock 2003; Ohmura 2001), we resort to using PDD sum.

To determine an appropriate time period of melt, we did a backward accumulation of PDD starting from the date of an ASO snapshot and going backward in time to encompass the melt season (Fig. 3, along with appendix Fig. A2). We observe that, for ASO snapshots, PDD sum increases with the number of backward accumulated days, eventually leveling out around 15 March. Therefore, for each ASO snapshot, we considered the PDD sum over a time period starting from 15 March till the day of the snapshot.

3) ACCUMULATED PRECIPITATION AND MEAN SEASONAL TEMPERATURE

We considered two additional meteorological variables. First, since the method for extracting snowfall from precipitation is an approximate one, we also consider accumulated precipitation; this is accumulated over the same time period as accumulated snowfall. In addition, we also consider the mean seasonal air temperature (hereinafter referred to as T_{mean}), which is computed by averaging mean daily temperatures across the snow season (i.e., from 1 October till the date of an ASO snapshot).

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

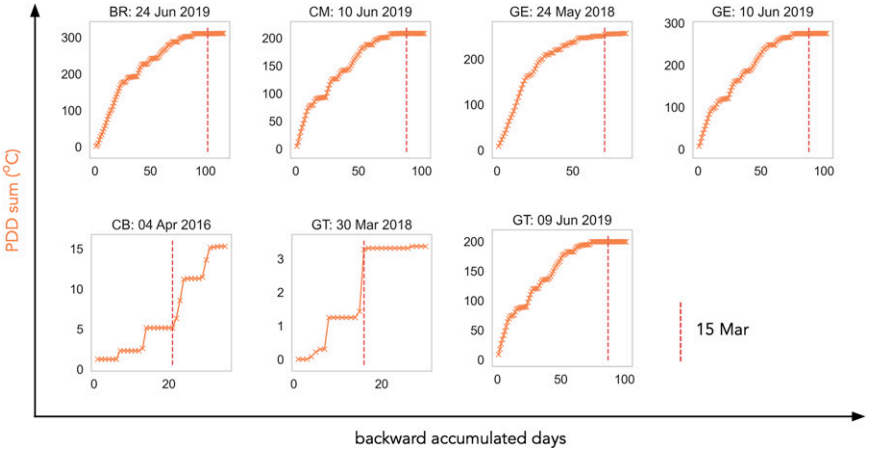

FIG. 3. PDD sum (averaged over the spatial extent of an ASO snapshot) as a function of backward accumulated days. This corresponds to backward time accumulation of PDD, starting from the date of an ASO snapshot and accumulating backward in time to encompass the entire melt season. The dotted line corresponds to 15 Mar of the melt season and approximately corresponds to the date when backward time accumulation of PDD levels out. For brevity, plots for snapshots corresponding to the early-melt period are not shown if they have a corresponding late-melt snapshot. Those plots are presented in appendix Fig. A2. Note that for CB: 4 Apr 2016, PDD sum continues to increase but the magnitude is small (relative to late-melt-period dates), suggesting a negligible melt effect.

The T_{mean} helps to capture the averaged spatial heterogeneity of temperature across a basin, which may be related to the spatial heterogeneity of snow accumulation and melt processes.

b. *Surface reflectance variables*

Surface reflectance data measure the fraction of incoming solar radiation reflected from Earth's surface across the electromagnetic spectrum. Several indices can be derived from these data that provide information about the characteristics of snow cover (Hall et al. 1995; Painter et al. 2012; Salomonson and Appel 2006; X.-Y. Wang et al. 2015). We use the following indices:

1) NDSI

Normalized difference snow index (NDSI) is a spectral band ratio that leverages the fact that snow reflectance is high in visible wavelengths and low in shortwave infrared wavelengths (Hall et al. 1995; Salomonson and Appel 2004). It helps to distinguish between snow-covered and snow-free areas and is defined as

$$\text{NDSI} = \frac{\text{b4} - \text{b6}}{\text{b4} + \text{b6}},$$

where b4 refers to MODIS band 4 (visible green band; 0.555 μm) and b6 refers to MODIS band 6 (shortwave infrared band; 1.640 μm).

2) NDSI7

New normalized difference snow index (NDSI7) is similar to NDSI, with the only difference being that MODIS band 6 is replaced by MODIS band 7 (shortwave infrared band; 2.130 μm). It was initially proposed by Salomonson and Appel (2006) because the MODIS instrument aboard the *Aqua* satellite has a nonfunctional band 6. We include NDSI7 in our study since it provides another measure to distinguish between snow-covered and snow-free areas.

3) NDFSI

The presence of forests can lead to mass loss (and subsequent sublimation) due to canopy interception. When the underlying surface consists of vegetation, NDSI and NDSI7 are less effective at identifying snow (Rittger et al. 2013). Based on the observation that in snow-covered forests, the reflectance in the near-infrared spectrum is higher than that in the visible spectrum, X.-Y. Wang et al. (2015) proposed normalized difference forest snow index (NDFSI) to differentiate between snow-covered and snow-free forests:

$$\text{NDFSI} = \frac{\text{b2} - \text{b6}}{\text{b2} + \text{b6}},$$

where b2 refers to MODIS band 2 (near-infrared band; 0.858 μm).

4) NDVI

Normalized difference vegetation index (NDVI) helps to distinguish between forest-covered and forest-free areas. It leverages the fact that when vegetation is present, reflectance is high in the near-infrared spectrum and low in the visible spectrum. Using MODIS surface reflectance data, it can be defined as

$$\text{NDVI} = \frac{\text{b2} - \text{b1}}{\text{b2} + \text{b1}},$$

where b1 refers to MODIS band 1 (visible red band; 0.645 μm). We use NDVI because using NDFSI alone is

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

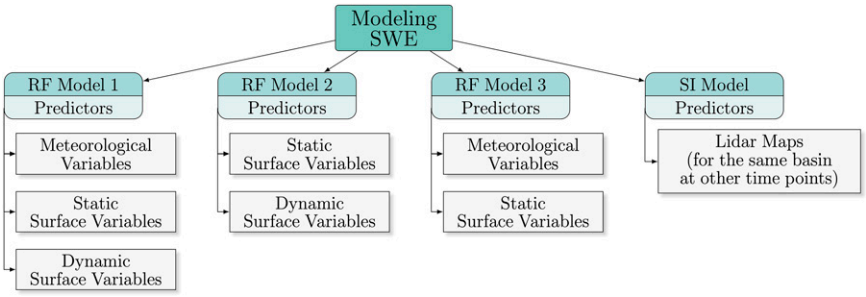

FIG. 4. The predictor space of each model to estimate SWE.

not sufficient to segregate forest-covered and forest-free areas. Wang et al. (2020) noted that the threshold NDFSI value is different for deciduous and coniferous forests and proposed that this threshold can be set according to the NDVI values of the forest. The use of NDVI to account for the influence on vegetation on the snowpack was also proposed by Klein et al. (1998).

5) NDGSI

Dust radiative forcing has been shown to be a significant driver of snowmelt in UCRB (Painter et al. 2007). The spectral albedo of pure snow is sensitive to changes in snow optical grain radius, which in turn has a logarithmic relationship with normalized difference grain size index (NDGSI; Painter et al. 2012). NDGSI is defined as

$$\text{NDGSI} = \frac{\text{b2} - \text{b5}}{\text{b2} + \text{b5}},$$

where b5 refers to MODIS band 5 (shortwave infrared band; 1.240 μm). Because it is the only band ratio that is related to the spectral albedo of pure dust-free snow, NDGSI may help to approximate the effect of dust in an ML framework.

c. *Topographic and vegetation variables*

Topography and vegetation are known to affect snow accumulation and melt at the watershed scale (e.g., Jost et al. 2007). Following the analysis by Zheng et al. (2016), we accounted for (i) topographic factors using elevation, slope, aspect, latitude and longitude, and (ii) vegetation factors using tree canopy. The topographic factors were obtained from elevation maps while tree canopy was obtained from tree canopy maps (section 2c). Tree canopy serves as an additional variable to quantify vegetation cover since the correlation of NDVI with vegetation becomes weaker in the presence of snow (Dye and Tucker 2003).

d. *ML framework: RF*

Our modeling framework uses RF, which is an ML method based on an ensemble of regression trees (Breiman 2001). Regression trees are flowchart-like structures that seek to minimize the mean-squared error between the model and the observations by recursively partitioning the input into smaller subspaces. The RF model trains each regression tree on a bootstrapped set of data points and gives each tree access to a random subset of features for training. The final output of the RF model is obtained by taking the average prediction across all regression trees in the ensemble. The RF model also provides measures of the relative "feature importance" of each predictor variable. This is obtained by calculating the mean decrease in mean square error (MSE), which is based on the total decrease in node MSE from splitting on a variable, averaged over all trees (Biau and Scornet 2016).

We implemented RF using Python's *scikit-learn* module (Pedregosa et al. 2011) and used the default hyperparameter values specified by the inventors of RF (as documented at https://CRAN.R-project.org/package=randomForest). Therefore, we used 500 trees, considered $p/3$ features when looking for the best split (where p is the number of predictor variables), and specified a node size of 5 (minimum number of samples in a leaf node). RF models are known to work well with these default values and are relatively insensitive to hyperparameter tuning (Probst et al. 2019; Schratz et al. 2019). We verified this by implementing a randomized search algorithm (Bergstra and Bengio 2012) based on threefold cross validation (section 4).

We evaluated the ability of RF to predict spatially distributed SWE using block or spatial leave-one-out cross validation. Specifically, for each map, we trained a model using the remaining 11 maps, minimizing dependence between the training and test sets. This also meant that each model had access to information from multiple spatial domains across multiple time points. Combining maps from multiple spatial domains expands the database and helps to model SWE in a given basin using information from neighboring basins. Our premise is that since the different spatial domains are roughly adjacent to each other, there should be similarity in the processes governing the spatial distribution of snowpack.

It is important to note that geospatial data inherently exhibit a degree of spatial autocorrelation, wherein nearby points in space are more similar to each other. Such structures imply a dependence within sampled data and can lead to overoptimistic generalization estimates if data are split randomly into training and validation/test sets. Spatial or block cross validation, wherein data obtained at approximately the same time point are spatially segregated, helps to mitigate this dependence and provides truer error estimates (Lyons et al. 2018; Roberts et al. 2017; Karasiak et al. 2022; Schratz et al. 2019).

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

TABLE 1. Number of gridded pixels for each basin. The input dimension of the predictor space is number of predictor variables × number of pixels. The output dimension of the target SWE is number of pixels × 1.

Basin	No. of pixels (basin)[a]	No. of pixels (training)[b]
Blue River (BR)	1125	10 417
Crested Butte (CB)	243	11 299
Maroon/Castle Creek (CM)	427	11 115
Gunnison–East River (GE)	1388	10 154
Gunnison–Taylor River (GT)	881	10 661

[a] Corresponds to the number of pixels in the basin (for one time point).
[b] Corresponds to the cumulative number of pixels summed across all basins, across all time points, minus the pixels reserved for block (or spatial) leave-one-out cross validation.

It is also worth pointing out that a trained RF model (or any other ML model) is an executable object that can be easily deployed for an operational application. The end-user need only build a data acquisition pipeline to acquire model inputs, which can be fed to the trained model to generate estimates of snowpack.

e. SI estimates

Recent studies (Malek et al. 2020; Zheng et al. 2018) have demonstrated that spatial interpolation of in situ measurements can yield reliable estimates of SWE when spatial patterns are extracted using lidar data at different time points. Therefore, we develop a simple interpolation (SI) model based on univariate linear interpolation of SNOTEL data, which uses lidar data as the only predictor. A state-of-the-art approach involves the use of optimal interpolation where remotely sensed data or forecast model estimates are used to generate a background field, which is then updated by incorporating in situ measurements that help account for various background errors using a weighted sum (see references in section 1). Here, we use the univariate SI model as we are interested in investigating the efficacy of using lidar data as predictors in our study area, and not in generating state-of-the-art interpolation estimates.

In contrast with the RF framework, the use of a lidar map as a predictor implies that an interpolation approach cannot incorporate information that is outside the spatial domain of the map. This means that SNOTEL measurements outside the spatial domain cannot be used. Even the simplest form of spatial interpolation, that is, univariate linear interpolation, requires at least two data points (here, SNOTEL measurements) for uniqueness. Therefore, we implement the SI model only for BR and GT basins (or two out of five basins as shown in Fig. 1).

For both basins, we implement a spatial leave-one-out approach. This means that, for a given time point, the actual lidar map serves as the ground truth while other lidar maps are used as predictors for interpolation. If multiple lidar maps are available for use as predictors (which is the case for GT basin), the lidar map that has the highest correlation with the ground truth map is used.

f. Overview of RF models

We classify the 15 variables described in sections 3a–3c into three broad categories:

(i) meteorological variables: accumulated snowfall, PDD sum, accumulated precipitation, and T_{mean},
(ii) dynamic surface variables: NDSI, NDSI7, NDFSI, NDVI, NDGSI, and Canopy, and
(iii) static surface variables: elevation, slope, aspect, latitude, and longitude.

Note that meteorological variables were obtained via one distinct data source (PRISM), static surface variables were obtained via another distinct data source (NED), and dynamic surface variables were obtained via two separate data sources (MODIS and NLCD).

TABLE 2. R^2 values [Eq. (1)] for modeling SWE.

Basin	Date	RF model 1	RF model 2	RF model 3	SI model	SNODAS
BR	19 Apr 2019	0.62	0.65	0.51	−0.61	0.32
	24 Jun 2019	0.66	0.48	0.74	−0.64	0.15
CB	4 Apr 2016	0.77	0.49	0.86	—	−0.49
CM	7 Apr 2019	0.13	0.41	0.29	—	0.18
	10 Jun 2019	0.83	0.75	0.80	—	0.05
GE	31 Mar 2018	0.67	−1.70	0.59	—	0.28
	24 May 2018	0.28	−0.69	−1.09	—	−0.13
	7 Apr 2019	0.16	−0.39	0.17	—	0.52
	10 Jun 2019	0.88	0.77	0.78	—	0.31
GT	30 Mar 2018	0.37	−3.41	0.29	0.35	0.13
	8 Apr 2019	0.71	0.70	0.72	0.79	−0.07
	9 Jun 2019	0.78	0.82	0.69	0.41	0.33
Mean		0.57	−0.09	0.45	—	0.13

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

TABLE 3. RMSE values (cm) for modeling SWE.

Basin	Date	RF model 1	RF model 2	RF model 3	SI model	SNODAS
BR	19 Apr 2019	13	13	15	27	17
	24 Jun 2019	11	14	10	25	18
CB	4 Apr 2016	7	11	6	—	18
CM	7 Apr 2019	21	17	19	—	20
	10 Jun 2019	17	20	18	—	39
GE	31 Mar 2018	9	25	10	—	13
	24 May 2018	7	11	12	—	9
	7 Apr 2019	29	37	29	—	22
	10 Jun 2019	13	18	17	—	31
GT	30 Mar 2018	7	18	7	7	8
	8 Apr 2019	10	10	9	8	18
	9 Jun 2019	12	11	14	20	21
Mean		13	17	14	—	20

We considered three different RF models in our study. Each RF model was trained to predict spatially distributed SWE and considered a different set of predictor variables. In addition, we also documented the performance of SNODAS and the SI model for comparison. The predictor space of the RF and SI models is outlined in Fig. 4.

RF model 1 considered all 15 variables and is the most sophisticated model in this study. All three RF models considered the five static surface (or topographic) variables since they characterize the underlying terrain, and their significance has been well documented in literature (Anderton et al. 2004; Balk and Elder 2000; Fassnacht et al. 2003). RF model 2 considered the six dynamic surface variables, and RF model 3 considered the four meteorological variables. The objective of RF models 2 and 3 was to help investigate how our primary model—RF model 1—breaks down by considering the relative importance of dynamic surface and meteorological variables, respectively. Dynamic surface variables help to quantify the characteristics of snow cover, while meteorological variables seek to capture the physical mechanisms driving the processes of snow accumulation and melt. Table A1 (in the appendix) lists the precise order of predictor variables for each RF model, which was maintained for repeatability. Table 1 outlines the number of gridded pixels for each basin, which informs the model dimensions during the block leave-one-out cross validation.

g. *Model evaluation metrics*

To evaluate the performance of a model to estimate the spatial distribution of SWE, we computed the coefficient of determination R^2 on test data; R^2 is defined as

$$R^2 = 1 - \frac{\text{MSE}}{\text{MST}}, \tag{1}$$

where MSE is the mean-square error between observed and modeled values in the test set and MST is the mean total sum of squares of observed values in the test set. The R^2 is highly informative for evaluating a regression model and can be interpreted as the proportion of variance of the observed values that can be predicted from the modeled values (Chicco et al. 2021). It normalizes the spread of prediction errors (i.e., MSE) by the spread of observed values (i.e., MST) and is invariant to linear scaling; R^2 is dimensionless and ranges from $-\infty$ to 1. Higher and positive values are desirable, with $R^2 = 1$ indicating that the modeled values perfectly match the observations. Negative values indicate a poor model performance.

TABLE 4. Relative bias percentages [Eq. (2)] for modeling SWE (negative value implies overprediction, and positive value implies underprediction).

Basin	Date	RF model 1	RF model 2	RF model 3	SI model	SNODAS
BR	19 Apr 2019	13	4	21	−32	11
	24 Jun 2019	16	13	−20	113	28
CB	4 Apr 2016	−7	−16	−4	—	10
CM	7 Apr 2019	−33	−18	−28	—	−6
	10 Jun 2019	4	20	−6	—	44
GE	31 Mar 2018	−25	−112	−17	—	−5
	24 May 2018	−97	−158	−246	—	94
	7 Apr 2019	36	54	37	—	5
	10 Jun 2019	5	31	3	—	16
GT	30 Mar 2018	−3	−76	−7	−27	−19
	8 Apr 2019	−12	6	−9	−2	−13
	9 Jun 2019	−36	−22	−39	80	21
Mean[a]		24	44	36	—	23

[a] Mean of absolute values of relative bias.

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

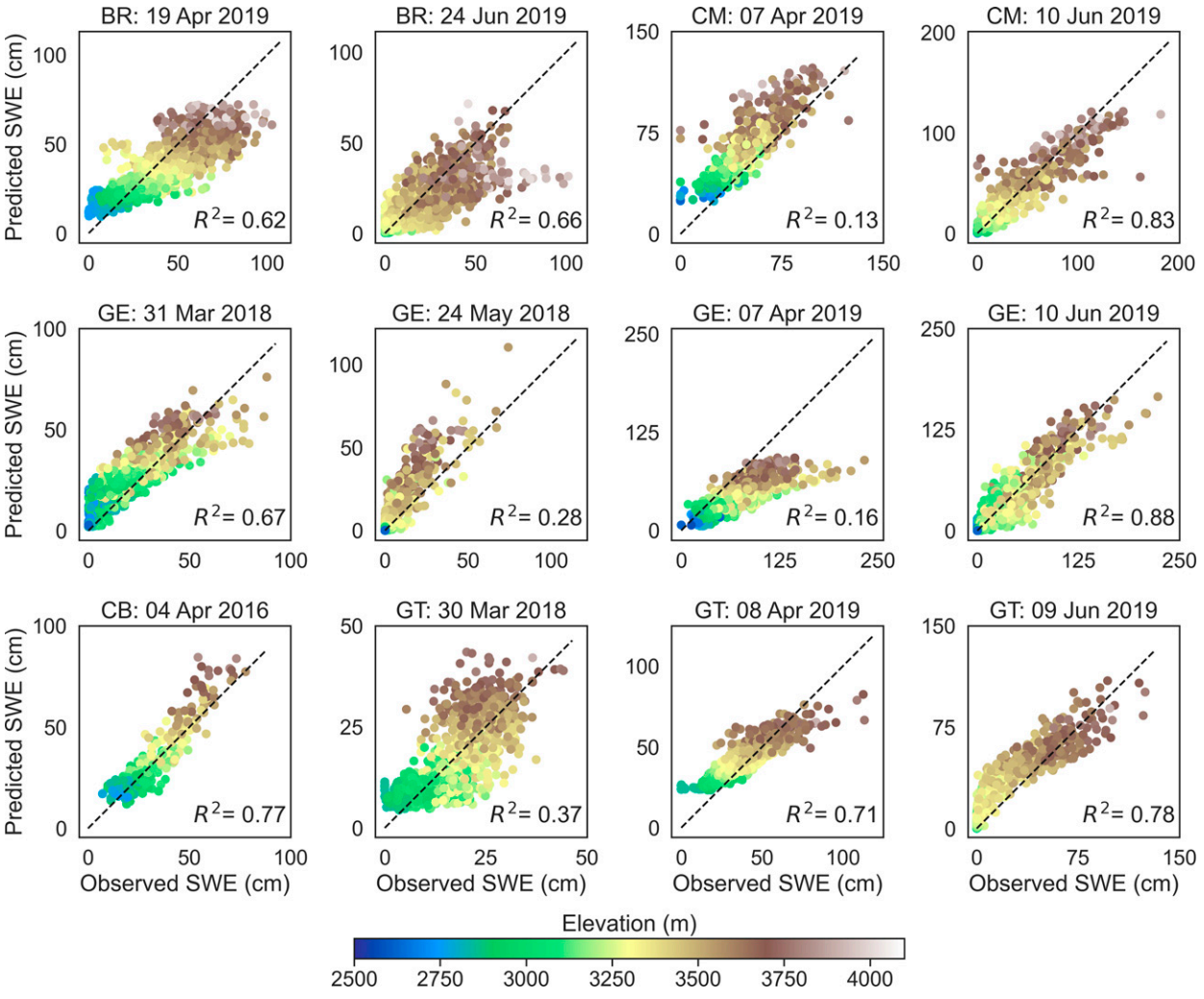

FIG. 5. Scatterplots to predict SWE using RF model 1. The individual points are color coded by their elevation. We have also plotted the 1:1 line for reference, which helps to visualize bias in the model. A majority of points above the 1:1 line implies overprediction ($b < 0$), whereas a majority of points under the 1:1 line implies underprediction ($b > 0$), as documented in Table 4.

In addition, we also computed root-mean-square error (RMSE) and relative bias b. RMSE helps us to evaluate the spread of prediction errors across a given basin. Relative bias helps to evaluate the relative deviation of the mean value of predicted SWE from the mean value of observed SWE and is expressed as a percentage:

$$b = 100 \times (\mu_o - \mu_p)/\mu_o, \tag{2}$$

where μ_p is the mean value of predicted SWE and μ_o is the mean value of observed SWE; $b = 0$ implies zero relative bias in the model, $b < 0$ implies overprediction, and $b > 0$ implies underprediction.

h. Finding a parsimonious feature space for RF model 1

At the end of our analysis, we investigated if the feature space of RF model 1 (consisting of 15 variables) can be made parsimonious. We used the sequential feature selection algorithm (Pudil et al. 1994; Chandrashekar and Sahin 2014). The algorithm starts with an empty set of features and adds features one at a time such that the model score (here, the five-fold cross-validation score) is maximized. This process continues till the model score stops improving. We ran this algorithm 12 times, wherein each time the training data consisted of 11 (of 12) snapshots and yielded a different set of features. The final set of parsimonious features was obtained by taking the union of the 12 sets of features obtained above.

4. Results

Tables 2–4 present the R^2, RMSE, and normalized bias values, respectively, for all four models. We also compare our models with SNODAS. Note that the SI model can only predict values for BR and GT basins as discussed in section 3e, which precludes computation of its overall mean.

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

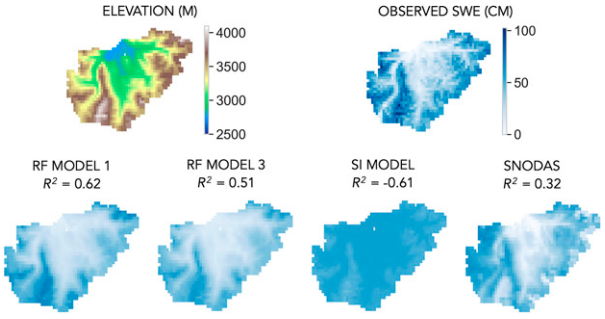

FIG. 6. Comparison between observed and predicted SWE for BR: 19 Apr 2019. Note that RF models seem to capture the overall variability of SWE better than SNODAS in high-elevation areas.

Note that Table A2 (in the appendix) presents R^2 values for an alternative RF model 1, which was trained using tuned hyperparameter values (section 3d) but yielded similar results as Table 1. Since the tuning process imposes an additional computational burden, we resorted to the default hyperparameter values. Additionally, we report that the RF and SI models were developed on a MacBook Pro with a six-core Intel i9 processor and 2.9-GHz processor speed. The training of each RF model was parallelized on all six cores, taking ~3 s to train. The SI model took ~0.8 ms to train.

We note that RF model 1, which uses all 15 predictor variables, has the best overall performance. It has the highest mean R^2, lowest mean RMSE, and the second-lowest bias. While RF model 1 is not necessarily the best model for a given ASO snapshot, it is the most robust model overall. RF model 3, which uses meteorological and static surface variables as predictors, is the next best model. RF model 2, which uses dynamic and static surface variables as predictors, is not very robust as characterized by multiple negative values of R^2 (which are also accompanied by high biases). The SI model performs well for two GT snapshots but exhibits inferior performance for other snapshots. Although SNODAS has the lowest bias, it has the highest mean RMSE and the second-lowest mean R^2, making it less reliable overall.

Figure 5 shows the scatterplots between observed and predicted (modeled) SWE for all ASO snapshots as obtained via the most sophisticated (and best performing) model considered in this study—RF model 1. All points are color coded by

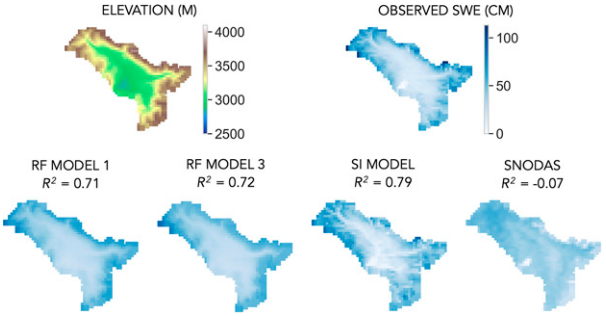

FIG. 7. Comparison between observed and predicted SWE for GT: 8 Apr 2019. Discussion is as in Fig. 6.

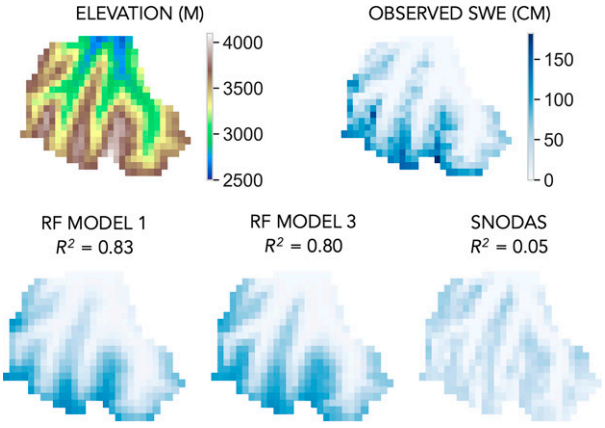

FIG. 8. Comparison between observed and predicted SWE for CM: 10 Jun 2019. Discussion is as in Fig. 6.

their elevation, which also helps to visualize the elevation dependence of the model. We see that the model succeeds in predicting lower SWE at lower elevations and higher SWE at higher elevations.

Figures 6–9 show a relative comparison between the spatial distribution of observed and predicted SWE obtained via RF model 1, RF model 3, SI model (where applicable), and SNODAS. We do not present the distribution obtained via RF model 2 as it is not robust. For brevity, we show comparisons for only a few basins that help to highlight some key aspects of the various models. First, we observe that the ability of the SI model to predict SWE at unmapped time points is not reliable in our study area. While it seems to capture the spatial distribution of SWE very well for GT: 8 April 2019 (Fig. 7), it produces a nearly homogenous distribution for BR: 19 April 2019 (Fig. 6). Second, we observe that although the RF models seem to capture the overall spatial variability of SWE, there is a tendency for some of the small-scale spatial features to be smoothed out. Finally, we observe that SNODAS estimates seem to capture the spatial variability of SWE better at

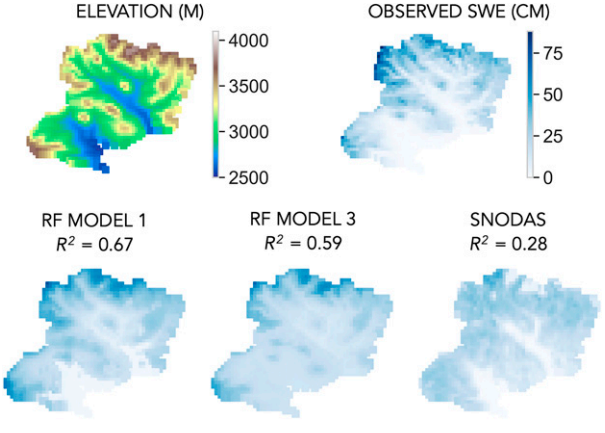

FIG. 9. Comparison between observed and predicted SWE for GE: 31 Mar 2018. Discussion is as in Fig. 6.

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

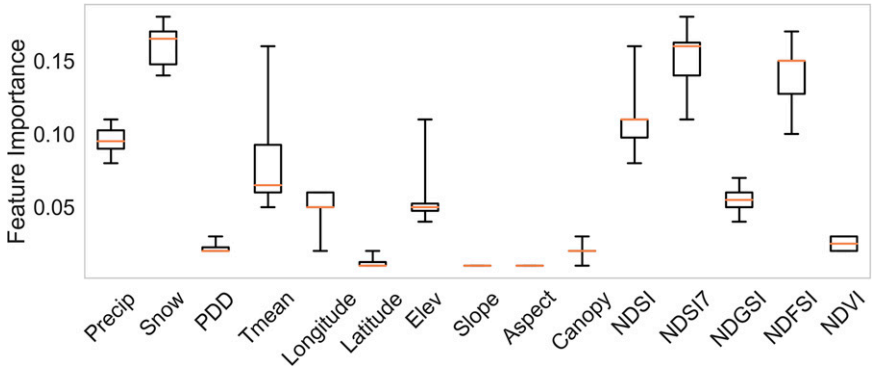

FIG. 10. Feature importances for RF model 1. The horizontal orange bar inside the box represents the median, and the "whiskers" extend to the range of the values. Variable names have been abbreviated: Precip is Accumulated precipitation, Snow is Accumulated snowfall, PDD is PDD sum, and Elev is Elevation.

lower elevations but are less accurate at higher elevations as documented previously (Clow et al. 2012).

To investigate the inner workings of the various RF models, we also computed the feature importances corresponding to each model. Feature importance essentially tells how much reduction in model MSE can be attributed to a given variable during the training process. RF models were trained using a leave-one-out approach wherein for each map we trained a model using the remaining 11 maps. As a result, each RF model is represented by 12 submodels—one for each map. This provided us with a statistical distribution of feature importances for each RF model. We visualized this distribution using box plots Figs. 10 and 11).

In RF model 1 (Fig. 10), the three variables with the highest median values of importance are accumulated snowfall, NDSI7, and NDFSI. In RF model 2 (Fig. 11a), the corresponding variables are NDSI7, NDFSI, and elevation. In RF model 3 (Fig. 11b), these are accumulated snowfall, T_{mean}, and PDD sum.

Last, the feature selection algorithm (section 3h) helped us arrive at a parsimonious set of eight features: (i) accumulated snowfall, (ii) T_{mean}, (iii) longitude, (iv) elevation, (v) NDSI, (vi) NDSI7, and (viii) NDVI. Note that the features are a combination of meteorological, static surface and dynamic surface variables. We document the model performance in Table A3 (in the appendix) and find that the mean R^2 (=0.54) is only slightly inferior to that of RF model 1, while the mean values of RMSE and relative bias are almost identical to RF model 1.

5. Discussion

a. *Benefits of the RF-based framework*

Our central objective was to develop an ML framework to model spatial distribution of SWE at unmapped time points. Although SNODAS provides daily estimates of SWE, it lacks accuracy in high-elevation or alpine areas (Clow et al. 2012). A sparse database of lidar maps, coupled with sparse in situ measurements, in the study area makes it challenging to obtain reliable estimates of SWE using the SI model. In the BR basin, the SI model performs poorly since early-melt SWE is used as a predictor for late-melt SWE and vice versa. Figure 12 (left) shows that although the SWE values for the two BR snapshots are strongly correlated at high elevations, they are poorly correlated at low elevations. Given that in situ measurements were available at low elevations (Fig. 1), interpolation used the relationship at low elevations to estimate SWE values at high elevations, resulting in a poor performance. However, the SI model compares well to (or even outperforms) RF model 1 for the two early-melt

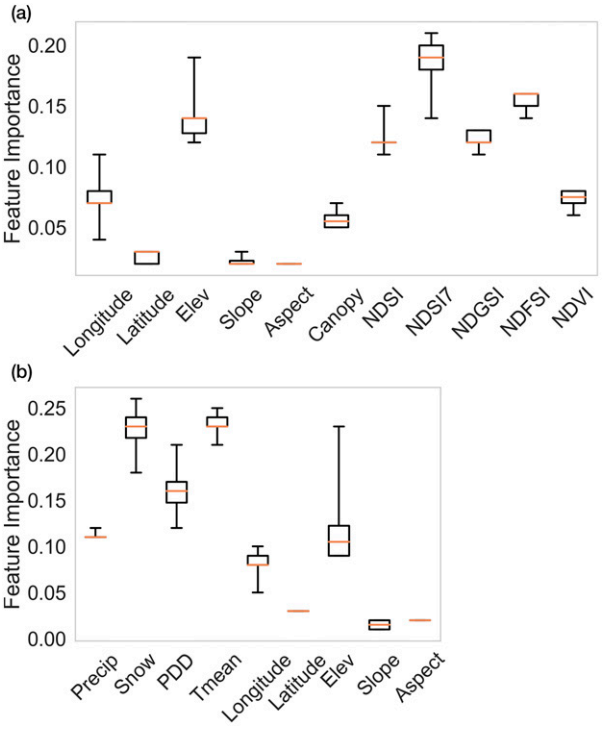

FIG. 11. Feature importances for (a) RF model 2 and (b) RF model 3 (variable names have been abbreviated as in Fig. 10).

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

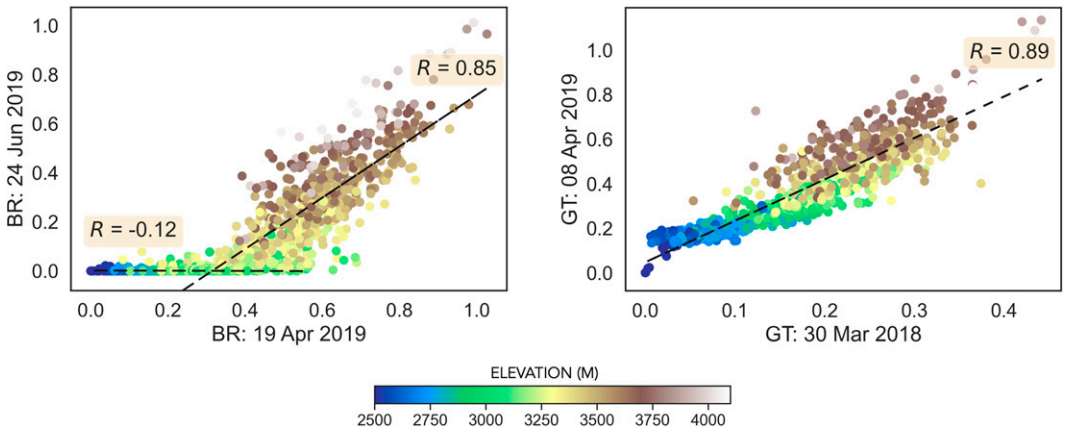

FIG. 12. Comparison of SWE values (m) for (left) BR snapshots that are 2 months apart and (right) two early melt snapshots of GT; *R* refers to the Pearson's correlation between SWE values of snapshots. The BR snapshots correspond to early-melt (April) and late-melt (June) time points, which causes them to be strongly correlated at high elevation and uncorrelated at low elevations. The GT snapshots show similar patterns of SWE across all elevations.

predictions in the GT basin. Figure 12 (right) shows that SWE values for the two early-melt snapshots of GT are highly correlated with each other at all elevations, which causes the two snapshots to be effective predictors for one another. These results show that, for a given basin, additional lidar campaigns that capture spatial patterns of SWE at desired periods in the melt season can significantly improve the performance of the SI model. Although optimal interpolation approaches are likely to yield superior SWE estimates than the SI model, they will be similarly constrained by the sparsity of lidar maps and in situ measurements in the study area.

The RF models have a distinct advantage over SI models on account of their ability to incorporate information from basins other than the one being modeled. Additionally, the RF models consider several predictor variables that seek to approximate the underlying processes governing snow accumulation and melt. By learning relationships between the predictors and SWE from multiple spatial domains across multiple time points, the RF models exhibit a degree of transferability across space and time (within the study area). For instance, we see that the RF models could predict SWE for both early-melt and late-melt time points in BR and CM basins (Tables 1–3), even though the training set did not have a SWE map from the same basin for the same melt period.

When compared with SNODAS, the RF-based framework performs better by virtue of yielding more accurate estimates at higher elevations (as shown in Fig. 13). This is possibly because SNODAS is designed to model snow dynamics across the conterminous United States (Schneider and Molotch 2016), while the RF models are trained specifically for mountains in the Upper Colorado River basin. The ability of SNODAS to capture snow variability at lower elevations is likely why it has a slightly lower bias when compared with RF model 1. Note that outside the study area, the RF models may yield inferior estimates of SWE since the relationship between input variables and SWE may be different.

It is also worth emphasizing that the ML framework constitutes an easy way to capture relationships between snowpack and different sets of input variables. Interpretable ML techniques (such as feature importance) help provide insight on what variables are important for modeling spatial distribution of snowpack. We discuss this in more detail next.

b. *Comparison of different RF models*

As outlined in section 3f, the objective behind considering RF models 2 and 3 was to investigate how our primary model, RF model 1, breaks down. We present the feature importances for RF model 1 in Fig. 10. However, a number of these features are correlated as shown in Fig. 14, which can severely affect the feature importances (Strobl et al. 2008). By segregating the predictors into different categories, we separately

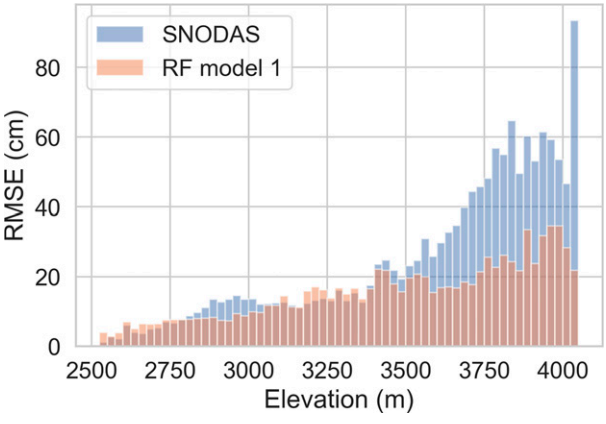

FIG. 13. Distribution of RMSE values (cm) as a function of elevation. Note that, at higher elevations, RMSE for RF model 1 is substantially lower than that for SNODAS.

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

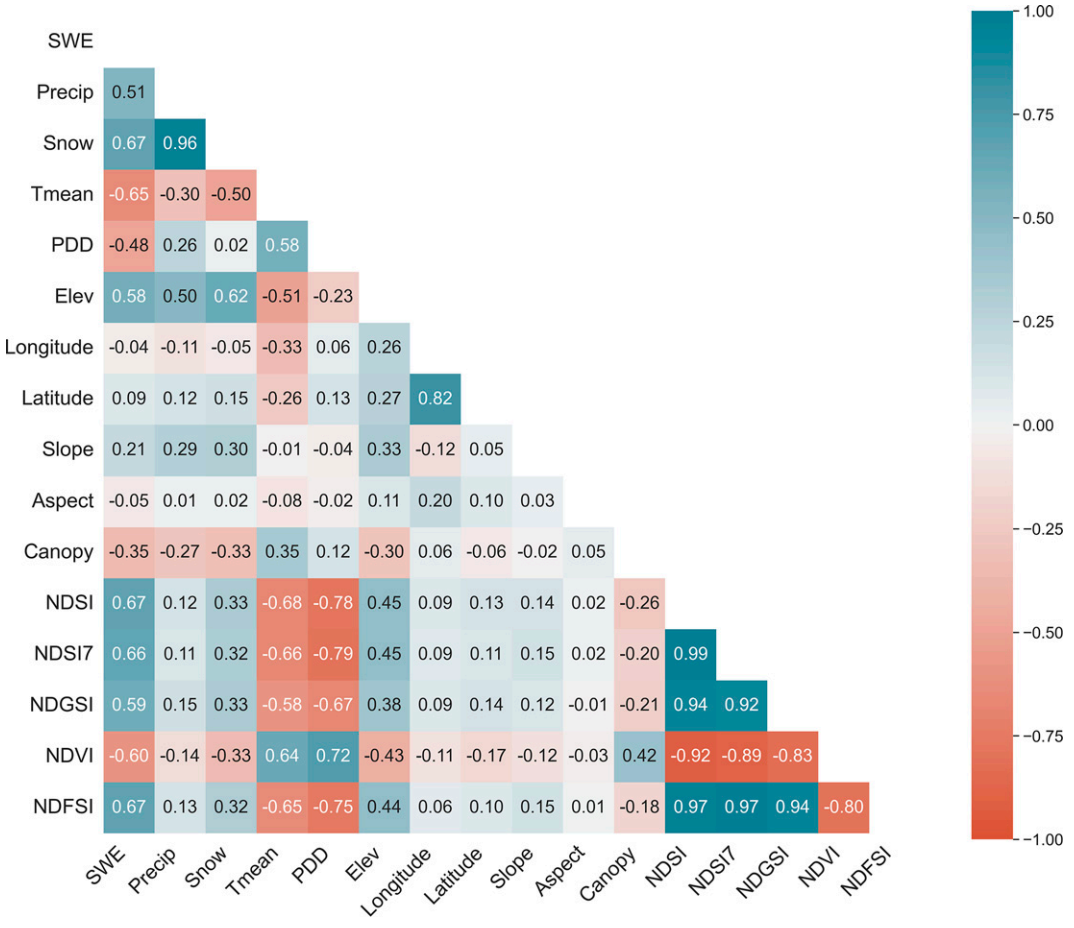

FIG. 14. Lower triangle of heat map showing correlations between predictor variables and SWE. Variable names have been abbreviated as in Fig. 10.

investigated the influence of meteorological and dynamic surface variables on the RF models.

A number of dynamic surface variables (specifically, most satellite-based surface indices) ranked as important for RF model 1. This is because surface properties get affected by the presence or absence of snow. However, excluding them (see results of RF model 3) did not have an adverse effect on SWE modeling. On the other hand, excluding meteorological variables (see results of RF model 2) greatly affected the robustness of SWE modeling; multiple SWE predictions had negative R^2 values.

From a data-driven perspective, a probable cause for the relative success of RF model 3 (based on meteorological variables) can be ascertained from the feature correlation heatmap (Fig. 14). The T_{mean} and PDD sum are highly correlated with the various spectral indices. This means that excluding spectral indices does not adversely impact the SWE model. We note that the feature importance of PDD sum increases significantly in RF model 3 when spectral indices are excluded. However, the same indices have a weak correlation with accumulated precipitation and snowfall (~0.1 and ~0.3, respectively). Given that accumulated snowfall is a particularly important feature in RF model 1, excluding it seems to have an adverse impact on SWE modeling in RF model 2.

From a physical perspective, both temperature and spectral indices are driven by changes in incoming solar radiation. We noted in section 3a that air temperature is correlated with several energy balance components, which are in turn driven by solar radiation. Surface reflectance, by definition, measures the fraction of incoming solar radiation reflected from Earth's surface. In fact, its dependence on solar radiation has been exploited by some studies to model the actual incoming solar radiation (Qin et al. 2011; López and Batlles 2014). Although precipitation is also driven by solar radiation, the effect is felt over larger time scales as part of the water cycle. Consequently, daily surface reflectance may be insufficient to approximate the effect of radiation on precipitation.

Our study suggests that despite the presence of biases in gridded estimates of precipitation and temperature (e.g., Lundquist et al. 2015; Strachan and Daly 2017), they still provide useful information to model SWE at unmapped time points. Gridded meteorological products from PRISM incorporate factors such as coastal proximity, topographic facet

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

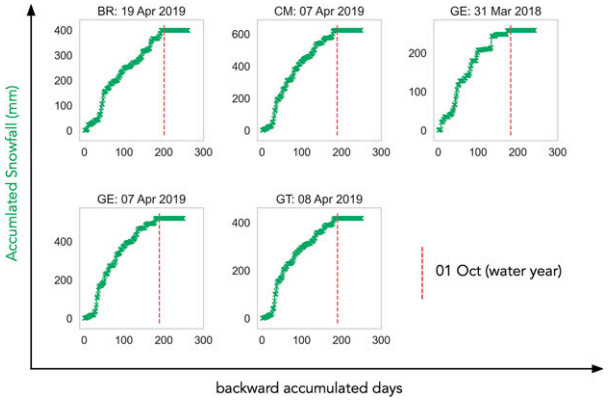

FIG. A1. Accumulated snowfall (averaged over the spatial extent of an ASO snapshot) as a function of backward accumulated days, for snapshots not shown in Fig. 2.

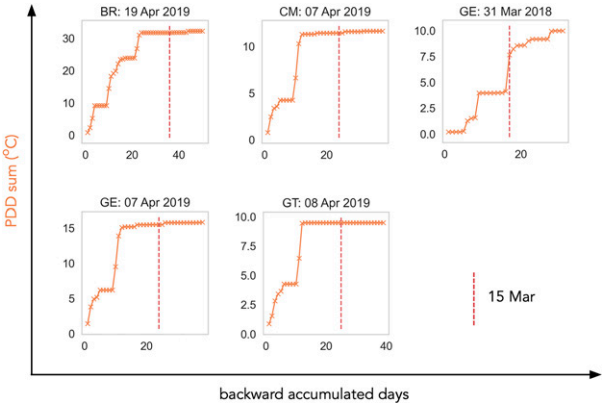

FIG. A2. PDD sum (averaged over the spatial extent of an ASO snapshot) as a function of backward accumulated days, for snapshots not shown in Fig. 3.

weighting, vertical atmospheric layer, and orographic effectiveness of terrain, in addition to elevation (Daly et al. 2008). This specialized knowledge about physiography and climate patterns likely contributes to their successful use as predictors of snowpack.

We highlight that although meteorological variables (RF model 3) are more robust predictors than dynamic surface variables (RF model 2), optimizing the predictive performance needs both types of variables. We see further evidence for this when we seek a parsimonious feature space, which turns out to be a combination of meteorological, static surface and dynamic surface variables. Note that the exclusion of a variable does not imply that its effect is not significant, rather this means that its effect is correlated with that of another variable. Depending on the training data and the feature selection algorithm, we may obtain a different set of parsimonious features. Therefore, if truncation of feature space is desired, care must be taken to evaluate the feature space using the available training data.

c. *Limitations and future directions*

Despite a good overall performance, Figs. 5–9 suggest that RF models have their limitations. First, we note limitations involving the temporal transferability of RF models. For instance, the SWE predictions for GE: 24 May 2018 suffer from a high overprediction bias. A likely reason is that the snapshot in question has the lowest fractional snow cover area relative to all others' snapshots (fSCA = 0.3; see Fig. 1). For such a patchy snow cover, it is likely that the relationships between predictors and SWE are distinct when compared with other snapshots and are therefore not entirely generalizable. A similar argument applies to SWE predictions for GE: 7 April 2019, which suffer from an underprediction bias. Here, it is possible that the snowpack is deeper/denser than that in other early-melt snapshots, again implying distinct relationships between predictors and SWE. It is also worth noting that ML-based approaches are most effective when the values of input variables and outputs do not exceed the extremes encountered during the training process.

In addition, we note that the RF models possess a tendency to smooth out the small-scale spatial variability of SWE. This tendency seems more prominent for RF model 3 (e.g., Fig. 9), which has a greater reliance on gridded meteorological variables. While gridded datasets generally assume an increase in precipitation with elevation, studies have shown that snowpack at higher elevations does not necessarily follow the same trend (Grünewald et al. 2014; Kirchner et al. 2014). This inconsistency is typically attributed to wind effects, with snow on the ground being affected by wind-driven predepositional and postdepositional processes (Mott et al. 2018). The presence of winds can also lead to undercatchment of precipitation at gauge locations, which in turn introduce biases in gridded estimates of precipitation (Lundquist et al. 2015). Furthermore, as elevation increases, increases in precipitation are influenced by seeder–feeder mechanisms (Bergeron 1965)—these may not be adequately captured in gridded datasets.

Further, we acknowledge that although air temperature variables (PDD sum and T_{mean}) serve as convenient proxies to capture radiation-driven snowmelt, they are likely insufficient. A general lack of high-resolution gridded estimates of energy balance components makes it challenging to approximate melt processes. Another important limitation of air temperature variables is their inability to adequately capture radiation variability due to topographic shading. Although PRISM partially accounts for this effect via topographic facet weighting (Daly et al. 2008), a lack of source stations at high elevations makes biases inevitable (Strachan and Daly 2017). A more accurate accounting of topographic shading may require computation of several additional terrain parameters such as azimuth, solar illumination angle, horizons, and radiation view factors (Dozier and Frew 1990), which would then need to be coupled with high-resolution radiation data (Aguilar et al. 2010).

We point out that the RF models were developed for application within the Rocky Mountains of Colorado. If it is desired to estimate SWE outside the study area, the RF models could be retrained after excluding geographic coordinates (i.e., longitude and latitude) from the feature space. An important future direction involves evaluating the ability of RF models to estimate the temporal variation of in situ SNOTEL

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

TABLE A1. Order of predictor variables for each RF model (documented for repeatability). Variable names have been abbreviated as in Fig. 10, and hyperparameter values are documented in section 3d. We used a seed value of 42.

Model	Predictor variables
RF model 1	Precip, Snow, PDD, Tmean, Longitude, Latitude, Elev, Slope, Aspect, Canopy, NDSI, NDSI7, NDGSI, NDFSI, and NDVI
RF model 2	Longitude, Latitude, Elev, Slope, Aspect, Canopy, NDSI, NDSI7, NDGSI, NDFSI, and NDVI
RF model 3	Precip, Snow, PDD, Tmean, Longitude, Latitude, Elev, Slope, and Aspect

observations for the entire snow season. This may require (i) expanding the modeling pipeline to incorporate SNOTEL data for training to capture how snowpack evolves during early season (i.e., October–March), (ii) downscaling input features to finer spatial resolutions that would be more consistent with point measurements (Mital et al. 2022), and (iii) development of a new spatiotemporal cross-validation strategy.

We also acknowledge another important limitation of our study. We upscaled ASO data to a resolution of 800 m. Although this procedure results in loss of information, such a concession was necessary to incorporate gridded estimates of meteorological datasets.

6. Conclusions

In this study, we developed a machine learning model that predicts spatially distributed mountain snowpack by combining gridded estimates of meteorological and satellite data with lidar maps of SWE. High-resolution lidar maps provide information about distribution of SWE in exceptional detail, enabling us to investigate and exploit some of the factors affecting the spatial distribution of SWE. We incorporated information about the physical processes governing snow accumulation and melt as well as snow characteristics by deriving 15 different variables from gridded estimates of precipitation, temperature, surface reflectance, elevation, and canopy. Our ML framework showed that by incorporating these variables, it is possible to train a model using sparsely available lidar-derived SWE and predict SWE at unmapped time points. Within our study area, the mean R^2 value using our approach was 0.57, which was a significant improvement over SNODAS estimates (mean R^2 = 0.13). Improvements were also observed in mean values of RMSE (which was 13 cm for ML modeling and 20 cm for SNODAS).

The ML framework also helps to analyze what variables are important for modeling spatial distribution of snowpack. We observed that, at the spatial resolution of 800 m, meteorological variables that characterize atmospheric processes are more important than dynamic surface variables that characterize the properties of snow deposited on the ground. The presented framework lends itself to spatiotemporal transferability, as we combined maps across multiple basins and time points. By expanding the applicability of remotely sensed lidar data to unmapped time points, our framework can model the spatially distributed SWE that can be used to constrain models seeking to estimate snowpack and streamflow characteristics. Improved estimates of snowpack and water availability are critical for sustainable management of water resources.

Acknowledgments. This work was funded by the ExaSheds project, which was supported by the Office of Biological and Environmental Research Earth and Environmental Systems Sciences Division Data Management Program, of the U.S. Department of Energy Office of Science, under Award DE-AC02-05CH11231. The proprietary PRISM data (800-m resolution) were purchased with funding from the Watershed Function Scientific Focus Area funded by the Office of Biological and Environmental Research of the U.S. Department of Energy Office of Science, under Award DE-AC02-05CH11231.

Data availability statement. Data analyzed in this study were a reanalysis of existing data, which are available at sources described in section 2c.

TABLE A2. The R^2 values for RF model 1 using tuned hyperparameters: ntrees (number of trees), mtry (number of features to consider when looking for the best split), and nodesize (minimum number of samples in a leaf node). RMSE and bias values are omitted for brevity.

Basin	Date	ntrees	mtry	nodesize	R^2	R^2 (using default values)
BR	19 Apr 2019	750	3	1	0.61	0.62
	24 Jun 2019	500	5	1	0.69	0.66
CB	4 Apr 2016	350	3	1	0.75	0.77
CM	7 Apr 2019	350	3	1	0.08	0.13
	10 Jun 2019	350	3	1	0.83	0.83
GE	31 Mar 2018	50	5	3	0.66	0.67
	24 May 2018	500	5	1	0.17	0.28
	7 Apr 2019	350	3	1	0.18	0.16
	10 Jun 2019	350	3	1	0.87	0.88
GT	30 Mar 2018	350	3	1	0.39	0.37
	8 Apr 2019	750	3	1	0.68	0.71
	9 Jun 2019	350	3	1	0.78	0.78

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

TABLE A3. RF model performance using a parsimonious feature space: (i) accumulated snowfall, (ii) T_{mean}, (iii) longitude, (iv) elevation, (v) NDSI, (vi) NDSI7, and (viii) NDVI.

Basin	Date	R^2	RMSE (cm)	Relative bias (%)
BR	19 Apr 2019	0.69	12	7
	24 Jun 2019	0.70	11	9
CB	4 Apr 2016	0.67	9	−12
CM	7 Apr 2019	0.19	20	−29
	10 Jun 2019	0.82	17	3
GE	31 Mar 2018	0.57	10	−34
	24 May 2018	0.43	6	−92
	7 Apr 2019	0.17	29	35
	10 Jun 2019	0.85	14	10
GT	30 Mar 2018	−0.07	9	−18
	8 Apr 2019	0.59	11	−19
	9 Jun 2019	0.82	11	−27
Mean		0.54	13	25[a]

[a] Mean of absolute values of relative bias.

APPENDIX

Additional Figures and Tables

Figure A1 shows accumulated snowfall (averaged over the spatial extent of an ASO snapshot), as a function of backward accumulated days, for early-season snapshots not shown in Fig. 2. Figure A2 shows the PDD sum (averaged over the spatial extent of an ASO snapshot), as a function of backward accumulated days, for early-season snapshots not shown in Fig. 3. Table A1 lists the precise order of predictor variables for each RF model, which was maintained for repeatability. Table A2 presents R^2 values for an alternative RF model 1, which was trained using tuned hyperparameter values (section 3d) but yielded results that are similar to those shown in Table 1. Table A3 shows the RF model performance using a parsimonious set of features, which was determined using the sequential feature selection algorithm (section 3h).

REFERENCES

Aguilar, C., J. Herrero, and M. J. Polo, 2010: Topographic effects on solar radiation distribution in mountainous watersheds and their influence on reference evapotranspiration estimates at watershed scale. *Hydrol. Earth Syst. Sci.*, **14**, 2479–2494, https://doi.org/10.5194/hess-14-2479-2010.

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

Bair, E. H., A. Abreu Calfa, K. Rittger, and J. Dozier, 2018: Using machine learning for real-time estimates of snow water equivalent in the watersheds of Afghanistan. *Cryosphere*, **12**, 1579–1594, https://doi.org/10.5194/tc-12-1579-2018.

Bales, R. C., N. P. Molotch, T. H. Painter, M. D. Dettinger, R. Rice, and J. Dozier, 2006: Mountain hydrology of the western United States. *Water Resour. Res.*, **42**, W08432, https://doi.org/10.1029/2005WR004387.

Balk, B., and K. Elder, 2000: Combining binary decision tree and geostatistical methods to estimate snow distribution in a mountain watershed. *Water Resour. Res.*, **36**, 13–26, https://doi.org/10.1029/1999WR900251.

Barrett, A. P., 2003: National Operational Hydrologic Remote Sensing Center Snow Data Assimilation System (SNODAS) products at NSIDC. NSIDC Special Rep. 11, 19 pp., https://nsidc.org/sites/nsidc.org/files/files/nsidc_special_report_11.pdf.

Behrangi, A., K. J. Bormann, and T. H. Painter, 2018: Using the airborne snow observatory to assess remotely sensed snowfall products in the California Sierra Nevada. *Water Resour. Res.*, **54**, 7331–7346, https://doi.org/10.1029/2018WR023108.

Bergeron, T., 1965: On the low-level redistribution of atmospheric water caused by orography. *Suppl. Proc. Int. Conf. Cloud Phys.*, Tokyo and Sapporo, Japan, World Meteorological Organization, 96–100.

Bergstra, J., and Y. Bengio, 2012: Random search for hyper-parameter optimization. *J. Mach. Learn. Res.*, **13**, 281–305.

Biau, G., and E. Scornet, 2016: A random forest guided tour. *TEST*, **25**, 197–227, https://doi.org/10.1007/s11749-016-0481-7.

Brasnett, B., 1999: A global analysis of snow depth for numerical weather prediction. *J. Appl. Meteor. Climatol.*, **38**, 726–740, https://doi.org/10.1175/1520-0450(1999)038<0726:AGAOSD>2.0.CO;2.

Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, https://doi.org/10.1023/A:1010933404324.

Brown, R. D., B. Brasnett, and D. Robinson, 2003: Gridded North American monthly snow depth and snow water equivalent for GCM evaluation. *Atmos.–Ocean*, **41**, 1–14, https://doi.org/10.3137/ao.410101.

Broxton, P. D., N. Dawson, and X. Zeng, 2016: Linking snowfall and snow accumulation to generate spatial maps of SWE and snow depth. *Earth Space Sci.*, **3**, 246–256, https://doi.org/10.1002/2016EA000174.

Bühler, Y., M. Marty, L. Egli, J. Veitinger, T. Jonas, P. Thee, and C. Ginzler, 2015: Snow depth mapping in high-alpine catchments using digital photogrammetry. *Cryosphere*, **9**, 229–243, https://doi.org/10.5194/tc-9-229-2015.

Carroll, T., D. Cline, G. Fall, A. Nilsson, L. Li, and A. Rost, 2001: NOHRSC operations and the simulation of snow cover properties for the coterminous U.S. *Proc. 69th Annual Meeting of the Western Snow Conf.*, Sun Valley, ID, Western Snow Conference, 1–14, https://westernsnowconference.org/sites/westernsnowconference.org/PDFs/2001Carroll.pdf.

Cartwright, K., C. Mahoney, and C. Hopkinson, 2022: Machine learning based imputation of mountain snowpack depth within an operational LiDAR sampling framework in southwest Alberta. *Can. J. Remote Sens.*, **48**, 107–125, https://doi.org/10.1080/07038992.2021.1988540.

Chandrashekar, G., and F. Sahin, 2014: A survey on feature selection methods. *Comput. Electr. Eng.*, **40**, 16–28, https://doi.org/10.1016/j.compeleceng.2013.11.024.

Chen, X., and Coauthors, 2020: Integrating field observations and process-based modeling to predict watershed water quality under environmental perturbations. *J. Hydrol.*, **602**, 125762, https://doi.org/10.1016/j.jhydrol.2020.125762.

Chicco, D., M. J. Warrens, and G. Jurman, 2021: The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.*, **7**, e623, https://doi.org/10.7717/peerj-cs.623.

Clark, M. P., and Coauthors, 2011: Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review. *Water Resour. Res.*, **47**, W07539, https://doi.org/10.1029/2011WR010745.

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

Clow, D. W., L. Nanus, K. L. Verdin, and J. Schmidt, 2012: Evaluation of SNODAS snow depth and snow water equivalent estimates for the Colorado Rocky Mountains, USA. *Hydrol. Processes*, **26**, 2583–2591, https://doi.org/10.1002/hyp.9385.

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.*, **28**, 2031–2064, https://doi.org/10.1002/joc.1688.

Deems, J. S., S. R. Fassnacht, and K. J. Elder, 2008: Interannual consistency in fractal snow depth patterns at two Colorado mountain sites. *J. Hydrometeor.*, **9**, 977–988, https://doi.org/10.1175/2008JHM901.1.

de Rosnay, P., G. Balsamo, C. Albergel, J. Muñoz-Sabater, and L. Isaksen, 2014: Initialisation of land surface variables for numerical weather prediction. *Surv. Geophys.*, **35**, 607–621, https://doi.org/10.1007/s10712-012-9207-x.

Dozier, J., and J. Frew, 1990: Rapid calculation of terrain parameters for radiation modeling from digital elevation data. *IEEE Trans. Geosci. Remote Sens.*, **28**, 963–969, https://doi.org/10.1109/36.58986.

——, E. H. Bair, and R. E. Davis, 2016: Estimating the spatial distribution of snow water equivalent in the world's mountains. *WIREs Water*, **3**, 461–474, https://doi.org/10.1002/wat2.1140.

Dwivedi, D., and Coauthors, 2022: Imputation of contiguous gaps and extremes of subhourly groundwater time series using random forests. *J. Mach. Learn. Model. Comput.*, **3** (2), 1–22, https://doi.org/10.1615/JMachLearnModelComput.2021038774.

Dye, D. G., and C. J. Tucker, 2003: Seasonality and trends of snow-cover, vegetation index, and temperature in northern Eurasia. *Geophys. Res. Lett.*, **30**, 1405–1408, https://doi.org/10.1029/2002GL016384.

Erickson, T. A., M. W. Williams, and A. Winstral, 2005: Persistence of topographic controls on the spatial distribution of snow in rugged mountain terrain, Colorado, United States. *Water Resour. Res.*, **41**, W04014, https://doi.org/10.1029/2003WR002973.

Fassnacht, S. R., K. A. Dressler, and R. C. Bales, 2003: Snow water equivalent interpolation for the Colorado River basin from snow telemetry (SNOTEL) data. *Water Resour. Res.*, **39**, 1208, https://doi.org/10.1029/2002WR001512.

Ferraz, A., S. Saatchi, K. Bormann, and T. Painter, 2018: Fusion of NASA airborne snow observatory (ASO) lidar time series over mountain forest landscapes. *Remote Sens.*, **10**, 164, https://doi.org/10.3390/rs10020164.

Gan, Y., Y. Zhang, C. Kongoli, C. Grassotti, Y. Liu, Y.-K. Lee, and D.-J. Seo, 2021: Evaluation and blending of ATMS and AMSR2 snow water equivalent retrievals over the conterminous United States. *Remote Sens. Environ.*, **254**, 112280, https://doi.org/10.1016/j.rse.2020.112280.

Grünewald, T., Y. Bühler, and M. Lehning, 2014: Elevation dependency of mountain snow depth. *Cryosphere*, **8**, 2381–2394, https://doi.org/10.5194/tc-8-2381-2014.

Hall, D. K., G. A. Riggs, and V. V. Salomonson, 1995: Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote Sens. Environ.*, **54**, 127–140, https://doi.org/10.1016/0034-4257(95)00137-P.

Hamlet, A. F., P. W. Mote, M. P. Clark, and D. P. Lettenmaier, 2005: Effects of temperature and precipitation variability on snowpack trends in the western United States. *J. Climate*, **18**, 4545–4561, https://doi.org/10.1175/JCLI3538.1.

Harpold, A. A., S. Rajagopal, J. B. Crews, T. Winchell, and R. Schumer, 2017: Relative humidity has uneven effects on shifts from snow to rain over the western U.S. *Geophys. Res. Lett.*, **44**, 9742–9750, https://doi.org/10.1002/2017GL075046.

Hedrick, A. R., and Coauthors, 2018: Direct insertion of NASA airborne snow observatory-derived snow depth time series into the *iSnobal* energy balance snow model. *Water Resour. Res.*, **54**, 8045–8063, https://doi.org/10.1029/2018WR023190.

Hock, R., 2003: Temperature index melt modelling in mountain areas. *J. Hydrol.*, **282**, 104–115, https://doi.org/10.1016/S0022-1694(03)00257-9.

James, T., A. Evans, E. Madly, and C. Kelly, 2014: The economic importance of the Colorado River to the basin region. L. William Seidman Research Institute Rep., 54 pp., https://businessforwater.org/wp-content/uploads/2016/12/PTF-Final-121814.pdf.

Jennings, K. S., and N. P. Molotch, 2019: The sensitivity of modeled snow accumulation and melt to precipitation phase methods across a climatic gradient. *Hydrol. Earth Syst. Sci.*, **23**, 3765–3786, https://doi.org/10.5194/hess-23-3765-2019.

——, T. S. Winchell, B. Livneh, and N. P. Molotch, 2018: Spatial variation of the rain–snow temperature threshold across the Northern Hemisphere. *Nat. Commun.*, **9**, 1148, https://doi.org/10.1038/s41467-018-03629-7.

Jost, G., M. Weiler, D. R. Gluns, and Y. Alila, 2007: The influence of forest and topography on snow accumulation and melt at the watershed-scale. *J. Hydrol.*, **347**, 101–115, https://doi.org/10.1016/j.jhydrol.2007.09.006.

Karasiak, N., J.-F. Dejoux, C. Monteil, and D. Sheeren, 2022: Spatial dependence between training and test sets: Another pitfall of classification accuracy assessment in remote sensing. *Mach. Learn.*, **111**, 2715–2740, https://doi.org/10.1007/s10994-021-05972-1.

Kim, R. S., M. Durand, D. Li, E. Baldo, S. A. Margulis, M. Dumont, and S. Morin, 2019: Estimating alpine snow depth by combining multifrequency passive radiance observations with ensemble snowpack modeling. *Remote Sens. Environ.*, **226** (8), 1–15, https://doi.org/10.1016/j.rse.2019.03.016.

Kirchner, P. B., R. C. Bales, N. P. Molotch, J. Flanagan, and Q. Guo, 2014: LiDAR measurement of seasonal snow accumulation along an elevation gradient in the southern Sierra Nevada, California. *Hydrol. Earth Syst. Sci.*, **18**, 4261–4275, https://doi.org/10.5194/hess-18-4261-2014.

Klein, A. G., D. K. Hall, and G. A. Riggs, 1998: Improving snow cover mapping in forests through the use of a canopy reflectance model. *Hydrol. Processes*, **12**, 1723–1744, https://doi.org/10.1002/(SICI)1099-1085(199808/09)12:10/11<1723::AID-HYP691>3.0.CO;2-2.

Kongoli, C., J. Key, and T. Smith, 2019: Mapping of snow depth by blending satellite and in-situ data using two-dimensional optimal interpolation—Application to AMSR2. *Remote Sens.*, **11**, 3049, https://doi.org/10.3390/rs11243049.

Largeron, C., and Coauthors, 2020: Toward snow cover estimation in mountainous areas using modern data assimilation methods: A review. *Front. Earth Sci.*, **8**, 325, https://doi.org/10.3389/feart.2020.00325.

Lehning, M., T. Grünewald, and M. Schirmer, 2011: Mountain snow distribution governed by an altitudinal gradient and terrain roughness. *Geophys. Res. Lett.*, **38**, L19504, https://doi.org/10.1029/2011GL048927.

Liu, Y., C. D. Peters-Lidard, S. V. Kumar, K. R. Arsenault, and D. M. Mocko, 2015: Blending satellite-based snow depth products with in situ observations for streamflow predictions

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

in the Upper Colorado River basin. *Water Resour. Res.*, **51**, 1182–1202, https://doi.org/10.1002/2014WR016606.

López, G., and F. J. Batlles, 2014: Estimating solar radiation from MODIS data. *Energy Procedia*, **49**, 2362–2369, https://doi.org/10.1016/j.egypro.2014.03.250.

Luce, C. H., D. G. Tarboton, and K. R. Cooley, 1998: The influence of the spatial distribution of snow on basin-averaged snowmelt. *Hydrol. Processes*, **12**, 1671–1683, https://doi.org/10.1002/(SICI)1099-1085(199808/09)12:10/11<1671::AID-HYP688>3.0.CO;2-N.

Lundquist, J. D., M. Hughes, B. Henn, E. D. Gutmann, B. Livneh, J. Dozier, and P. Neiman, 2015: High-elevation precipitation patterns: Using snow measurements to assess daily gridded datasets across the Sierra Nevada, California. *J. Hydrometeor.*, **16**, 1773–1792, https://doi.org/10.1175/JHM-D-15-0019.1.

Lyons, M. B., D. A. Keith, S. R. Phinn, T. J. Mason, and J. Elith, 2018: A comparison of resampling methods for remote sensing classification and accuracy assessment. *Remote Sens. Environ.*, **208**, 145–153, https://doi.org/10.1016/j.rse.2018.02.026.

Malek, S. A., R. C. Bales, and S. D. Glaser, 2020: Estimation of daily spatial snow water equivalent from historical snow maps and limited in-situ measurements. *Hydrology*, **7**, 46, https://doi.org/10.3390/hydrology7030046.

Mankin, J. S., D. Viviroli, D. Singh, A. Y. Hoekstra, and N. S. Diffenbaugh, 2015: The potential for snow to supply human water demand in the present and future. *Environ. Res. Lett.*, **10**, 114016, https://doi.org/10.1088/1748-9326/10/11/114016.

Margulis, S. A., Y. Fang, D. Li, D. P. Lettenmaier, and K. Andreadis, 2019: The utility of infrequent snow depth images for deriving continuous space-time estimates of seasonal snow water equivalent. *Geophys. Res. Lett.*, **46**, 5331–5340, https://doi.org/10.1029/2019GL082507.

Marks, D., and J. Dozier, 1992: Climate and energy exchange at the snow surface in the alpine region of the Sierra Nevada: 2. Snow cover energy balance. *Water Resour. Res.*, **28**, 3043–3054, https://doi.org/10.1029/92WR01483.

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

Mital, U., D. Dwivedi, J. B. Brown, B. Faybishenko, S. L. Painter, and C. I. Steefel, 2020: Sequential imputation of missing spatio-temporal precipitation data using random forests. *Front. Water*, **2**, 20, https://doi.org/10.3389/frwa.2020.00020.

——, ——, ——, and C. I. Steefel, 2022: Downscaled hyper-resolution (400 m) gridded datasets of daily precipitation and temperature (2008–2019) for East Taylor subbasin (western United States). *Earth Syst. Sci. Data*, **14**, 4949–4966, https://doi.org/10.5194/essd-14-4949-2022.

Mott, R., V. Vionnet, and T. Grünewald, 2018: The seasonal snow cover dynamics: Review on wind-driven coupling processes. *Front. Earth Sci.*, **6**, 197, https://doi.org/10.3389/feart.2018.00197.

National Operational Hydrologic Remote Sensing Center, 2004: Snow data assimilation system (SNODAS) data products at NSIDC, version 1. National Snow and Ice Data Center, accessed 15 October 2021, https://doi.org/10.7265/N5TB14TC.

Oaida, C. M., and Coauthors, 2019: A high-resolution data assimilation framework for snow water equivalent estimation across the western United States and validation with the airborne snow observatory. *J. Hydrometeor.*, **20**, 357–378, https://doi.org/10.1175/JHM-D-18-0009.1.

Ohmura, A., 2001: Physical basis for the temperature-based melt-index method. *J. Appl. Meteor. Climatol.*, **40**, 753–761, https://doi.org/10.1175/1520-0450(2001)040<0753:PBFTTB>2.0.CO;2.

Painter, T. H., 2018: ASO L4 lidar snow water equivalent 50m UTM Grid, version 1. National Snow and Ice Data Center, accessed 17 June 2020, https://doi.org/10.5067/M4TUH28NHL4Z.

——, A. P. Barrett, C. C. Landry, J. C. Neff, M. P. Cassidy, C. R. Lawrence, K. E. McBride, and G. L. Farmer, 2007: Impact of disturbed desert soils on duration of mountain snow cover. *Geophys. Res. Lett.*, **34**, L12502, https://doi.org/10.1029/2007GL030284.

——, A. C. Bryant, and S. M. Skiles, 2012: Radiative forcing by light absorbing impurities in snow from MODIS surface reflectance data. *Geophys. Res. Lett.*, **39**, L17502, https://doi.org/10.1029/2012GL052457.

——, and Coauthors, 2016: The airborne snow observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. *Remote Sens. Environ.*, **184**, 139–152, https://doi.org/10.1016/j.rse.2016.06.018.

Peck, E. L., T. R. Carroll, and S. C. Vandemark, 1980: Operational aerial snow surveying in the United States/Etude de neige aérienne effectuée aux Etats Unis. *Hydrol. Sci. Bull.*, **25**, 51–62, https://doi.org/10.1080/02626668009491904.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830, https://doi.org/10.48550/ARXIV.1201.0490.

Probst, P., M. N. Wright, and A.-L. Boulesteix, 2019: Hyperparameters and tuning strategies for random forest. *WIREs Data Mining Knowl. Discovery*, **9**, e1301, https://doi.org/10.1002/widm.1301.

Pudil, P., J. Novovičová, and J. Kittler, 1994: Floating search methods in feature selection. *Pattern Recognit. Lett.*, **15**, 1119–1125, https://doi.org/10.1016/0167-8655(94)90127-9.

Qin, J., Z. Chen, K. Yang, S. Liang, and W. Tang, 2011: Estimation of monthly-mean daily global solar radiation based on MODIS and TRMM products. *Appl. Energy*, **88**, 2480–2489, https://doi.org/10.1016/j.apenergy.2011.01.018.

Rittger, K., T. H. Painter, and J. Dozier, 2013: Assessment of methods for mapping snow cover from MODIS. *Adv. Water Resour.*, **51**, 367–380, https://doi.org/10.1016/j.advwatres.2012.03.002.

Roberts, D. R., and Coauthors, 2017: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, **40**, 913–929, https://doi.org/10.1111/ecog.02881.

Salomonson, V. V., and I. Appel, 2004: Estimating fractional snow cover from MODIS using the normalized difference snow index. *Remote Sens. Environ.*, **89**, 351–360, https://doi.org/10.1016/j.rse.2003.10.016.

——, and ——, 2006: Development of the Aqua MODIS NDSI fractional snow cover algorithm and validation results. *IEEE Trans. Geosci. Remote Sens.*, **44**, 1747–1756, https://doi.org/10.1109/TGRS.2006.876029.

Scalzitti, J., C. Strong, and A. Kochanski, 2016: Climate change impact on the roles of temperature and precipitation in western U.S. snowpack variability. *Geophys. Res. Lett.*, **43**, 5361–5369, https://doi.org/10.1002/2016GL068798.

Schirmer, M., V. Wirz, A. Clifton, and M. Lehning, 2011: Persistence in intra-annual snow depth distribution: 1. Measurements and topographic control. *Water Resour. Res.*, **47**, W09516, https://doi.org/10.1029/2010WR009426.

Schneider, D., and N. P. Molotch, 2016: Real-time estimation of snow water equivalent in the Upper Colorado River basin using MODIS-based SWE reconstructions and SNOTEL data.

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC

Water Resour. Res., **52**, 7892–7910, https://doi.org/10.1002/2016WR019067.

Schratz, P., J. Muenchow, E. Iturritxa, J. Richter, and A. Brenning, 2019: Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecol. Modell.*, **406**, 109–120, https://doi.org/10.1016/j.ecolmodel.2019.06.002.

Shuai, P., X. Chen, U. Mital, E. T. Coon, and D. Dwivedi, 2022: The effects of spatial and temporal resolution of gridded meteorological forcing on watershed hydrological responses. *Hydrol. Earth Syst. Sci.*, **26**, 2245–2276, https://doi.org/10.5194/hess-26-2245-2022.

Strachan, S., and C. Daly, 2017: Testing the daily PRISM air temperature model on semiarid mountain slopes. *J. Geophys. Res. Atmos.*, **122**, 5697–5715, https://doi.org/10.1002/2016JD025920.

Strobl, C., A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis, 2008: Conditional variable importance for random forests. *BMC Bioinf.*, **9**, 307, https://doi.org/10.1186/1471-2105-9-307.

Sturm, M., M. A. Goldstein, and C. Parr, 2017: Water and life from snow: A trillion dollar science question. *Water Resour. Res.*, **53**, 3534–3544, https://doi.org/10.1002/2017WR020840.

U.S. Geological Survey, 2019: The National Map—New data delivery homepage, advanced viewer, lidar visualization. USGS Doc., 2 pp., https://pubs.usgs.gov/fs/2019/3032/fs20193032.pdf.

Vermote, E., and R. Wolfe, 2015: MOD09GA MODIS/*Terra* surface reflectance daily L2G Global 1 km and 500 m SIN Grid, version 6. USGS, accessed 21 July 2021, https://doi.org/10.5067/MODIS/MOD09GA.006.

Wainwright, H. M., and Coauthors, 2020: Satellite-derived foresummer drought sensitivity of plant productivity in Rocky Mountain headwater catchments: Spatial heterogeneity and geological-geomorphological control. *Environ. Res. Lett.*, **15**, 084018, https://doi.org/10.1088/1748-9326/ab8fd0.

Wang, X.-Y., J. Wang, Z.-Y. Jiang, H.-Y. Li, and X.-H. Hao, 2015: An effective method for snow-cover mapping of dense coniferous forests in the upper Heihe River basin using Landsat operational land imager data. *Remote Sens.*, **7**, 17 246–17 257, https://doi.org/10.3390/rs71215882.

——, S. Chen, and J. Wang, 2020: An adaptive snow identification algorithm in the forests of northeast China. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **13**, 5211–5222, https://doi.org/10.1109/JSTARS.2020.3020168.

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

Zheng, Z., P. B. Kirchner, and R. C. Bales, 2016: Topographic and vegetation effects on snow accumulation in the southern Sierra Nevada: A statistical summary from lidar data. *Cryosphere*, **10**, 257–269, https://doi.org/10.5194/tc-10-257-2016.

——, N. P. Molotch, C. A. Oroza, M. H. Conklin, and R. C. Bales, 2018: Spatial snow water equivalent estimation for mountainous areas using wireless-sensor networks and remote-sensing products. *Remote Sens. Environ.*, **215**, 44–56, https://doi.org/10.1016/j.rse.2018.05.029.

Unauthenticated | Downloaded 12/02/22 12:00 AM UTC