Remote Sensing of Environment 264 (2021) 112608

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Multi-sensor fusion using random forests for daily fractional snow cover at 30 m

Karl Rittger [a,b,*], Mitchell Krock [c], William Kleiber [c], Edward H. Bair [b], Mary J. Brodzik [d], Thomas R. Stephenson [e], Balaji Rajagopalan [f], Kat J. Bormann [g,h], Thomas H. Painter [g]

[a] Institute of Arctic and Alpine Research, University of Colorado, 4001 Discovery Dr, Boulder, CO 80303, USA
[b] Earth Research Institute, University of California, Santa Barbara, CA 93106, USA
[c] Department of Applied Mathematics, University of Colorado, Boulder, CO 80309, USA
[d] National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, USA
[e] Sierra Nevada Bighorn Sheep Recovery Program, California Department of Fish and Wildlife, 787 North Main St., Suite 220, Bishop, CA 93514, USA
[f] Department of Civil, Environmental and Architectural Engineering & Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, USA
[g] Airborne Snow Observatories Inc., Mammoth Lakes, CA 93546, USA
[h] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

ARTICLE INFO

Editor: Menghua Wang

Keywords:
MODIS
Landsat
Fractional snow cover
Fusion
Downscaling
Spectral mixture analysis
Random forest

ABSTRACT

In addition to providing water for nearly 2 billion people, snow drives resource selection by wildlife and influences the behavior and demography of many species. Because snow cover is highly spatially and temporally variable, mapping its extent using currently available satellite data remains a challenge. At present, there are no sensors acquiring daily data of Earth's entire surface at fine spatial resolutions (< 30 m) in wavelengths required for snow cover retrieval, namely: visible, near-infrared, and shortwave infrared. Fine scale observations at 30 m from Landsat are available at 16-day intervals since 1982 and at 8-day intervals since 1999. However, over this duration, snow can accumulate, ablate, or both, making the Landsat data ineffective for many applications. Conversely, the Moderate Resolution Imaging Spectroradiometer (MODIS) atmospherically corrected daily reflectance data, have a coarse spatial resolution of 463 m and thus, are not ideal for snow cover mapping either. This spatial and temporal resolution tradeoff limits the use of these data for a wide range of snow cover applications and indicates a pressing need for data fusion. To address this need, we use a physically-based, spectral-mixture-analysis approach for mapping fractional snow cover (fSCA) and a two-stage random forest algorithm to produce daily 30 m fSCA. We test our algorithm in the US Sierra Nevada and find MODIS fSCA is the most important predictor. We cross validate using 170 Landsat scenes and while snow cover varies immensely in time we find little variation in errors between seasons, a small bias of 0.01, and an overall accuracy of 0.97 with slightly higher precision than recall. This technique for accurate, daily, high-resolution snow cover retrievals could be applied more broadly for analyses of regional energy budget, validating snow cover in global and regional models, and for quantifying changes in the availability of biotic resources in ecosystems.

1. Introduction

One quarter to one sixth of the global population depends on water from snow and ice melt (Barnett et al., 2005; Mankin et al., 2015). Snow is a natural reservoir of water, but it has a complex spatial and temporal distribution that make accurate creation of snow cover maps and the estimation of snow water equivalent (SWE) difficult. These difficulties in snow estimation manifest in runoff forecast errors and can have significant impacts on water managers (Stillinger et al., 2021), agricultural users, and others who depend on accurate water allocation estimates.

In addition to its impact on water resources, snow influences the survival of mountain ungulates and profoundly affects the energetic costs of living in colder environments (Conner et al., 2018; Stephenson et al., 2020). In mountainous or northern climates, snow determines the availability of forage and the energy that must be expended to acquire

* Corresponding author.
E-mail address: karl.rittger@colorado.edu (K. Rittger).

https://doi.org/10.1016/j.rse.2021.112608
Received 1 August 2020; Received in revised form 6 July 2021; Accepted 10 July 2021

food for much of the year (Parker et al., 1984; Schwab and Pitt, 1987). For example, bighorn sheep in alpine environments exhibit multiple migratory behaviors including alpine residency during winter (Spitz et al., 2018) and select habitat in response to snowscapes. Finer scale snow metrics improve predictive models of the movement and resource selection of these animals (Mahoney et al., 2018). Consequently, snow cover is an essential parameter in habitat models designed to predict resource selection (Spitz et al., 2020).

To better understand these impacts of snow cover on both humans and wildlife, we need to look at the challenges surrounding the satellite retrieval of fractional snow cover (fSCA) and related SWE estimates. To calculate SWE, accurate estimates of fSCA are needed, e.g. in energy balance models (Bair et al., 2016; Rittger et al., 2016). In a machine learning approach to predicting SWE, fSCA was the most important predictor (Bair et al., 2018a). However, accurate estimates of fSCA are difficult to calculate at both a high spatial and temporal resolution. This is especially true in mountainous terrain where the spatial variability of snow caused by wind redistribution (Liston and Sturm, 1998) and other factors can be substantial. For example, Airborne Snow Observatory (ASO, Painter et al., 2016) retrievals that use a combination of lidar and hyperspectral sensors to estimate SWE at 50 m resolution routinely produce adjacent pixels with an order of magnitude difference in SWE. Likewise, bare and snow-covered areas are often found within a few meters of each other because of differences in topography or vegetation (Rosenthal and Dozier, 1996; Selkowitz et al., 2014). To further complicate matters, areal snow cover can change on the order of hours, especially during melt periods where snow can be present in the morning and gone by afternoon. Rapid temporal changes influence accurate modeling of snowmelt, requiring imagery with a temporal resolution of at least one day (Slater et al., 2013). SWE models forced using 30 m fSCA data have shown more accurate results compared to those models forced with 500 m fSCA (Molotch and Margulis, 2008; Margulis et al., 2019).

Thus, for accurate spatial models of snow distribution and snowmelt, two competing requirements are needed: decameter scale spatial resolution and daily temporal resolution. No satellite or constellation of satellites currently satisfies these requirements. Three widely used sensors for remote sensing of snow are MODIS (daily, 500 m resolution), Landsat 5, and Landsat 7 (16-day for each or 8-day combined, 30 m resolution). More recently, data are available from Landsat 8 (16-day, 30 m resolution) and Sentinel 2a and 2b (10-day for each or 5-day combined, 10, 20, or 60 m resolution depending on band). Individually, none of these satellites satisfy our requirements, nor do they when combined together, though the scheduled launch of Landsat 9 in September 2021 will bring us one step closer.

The need for high-temporal and spatial-resolution snow cover products motivates sensor fusion, and there have been a few studies where fusion techniques have been applied to take advantage of both the spatial and temporal resolution, though many use binary snow cover for coarse resolution data, fine resolution data, or both. Baumgartner et al. (1987) compared Landsat Multispectral Scanner System with Advanced Very High Resolution Radiometer (AVHRR) imagery and suggested the two could be fused, but stopped short of creating a fused product. Durand et al. (2008) created a fused MODIS and Landsat product. The researchers combined fSCA derived from coarsening the MOD10A v.4 binary snow cover data (Hall et al., 2002) to 1 km with 30 m fractional snow cover estimated from Landsat 7 Enhanced Thematic Mapper+ (ETM+) reflectance data (Painter et al., 2003; Painter et al., 2009). The novel linear technique they used to combine the two data streams, when applied to the upper Rio Grande in Colorado (US) to constrain a SWE reconstruction model, resulted in a 51% reduction in mean absolute error (MAE) and a 49% reduction in bias for SWE. Likewise, (Durand et al., 2008) found that using ETM+ data upscaled to 100 m provided substantial improvement in SWE with 23% MAE compared to 50% MAE for the coarser MODIS data.

Like Durand et al. (2008), Berman et al. (2018) used MODIS and Landsat snow maps but took an alternate approach with a dynamic time warping technique to create daily 30 m snow cover estimates. Instead of binary snow cover, they used fractional snow cover from MOD10A1 (Salomonson and Appel, 2004). Berman et al. (2018) report a Root Mean Squared Error (RMSE) of 31% to 68% for fSCA, validated with ground measurements (snow pillows and time-lapse cameras). It is notable that the regions selected in that study are particularly challenging forest locations for optical remote sensing, with 60% to >80% canopy cover. High MODIS sensor view zenith angles that can completely obstruct snow cover in forests (Dozier et al., 2008; Rittger et al., 2020) were not accounted for, and only cloud-free (not gap-filled or smoothed) imagery (Hall et al., 2010) was used, thereby greatly reducing the effective temporal resolution for the inputs. Several studies have downscaled and compared snow cover between the products e.g., Landsat with MODIS snow cover data (Walters et al., 2014; Li et al., 2015) and find reasonable accuracy, but the maps are binary, which overestimate snow at high fractions and underestimate snow at low fractions (Rittger et al., 2013). A number of previous studies focusing on snow cover but not SWE have combined MODIS data and binary snow maps from Landsat at 30 m or unmanned aerial vehicles at similar or higher spatial resolution (Dobreva and Klein, 2011; Moosavi et al., 2014; Liang et al., 2017; Kuter et al., 2018; Liu et al., 2020; Kuter, 2021). The studies used varying methods such as linear regression, multivariate adaptive regression splines, neural networks, support vector machines and random forests, but focused on improving the accuracy of daily maps at 500 m rather than providing a higher spatial resolution daily product.

A goal of data fusion is that the fused product is more consistent or accurate than the individual parts. These earlier fusion methods used either the Normalized Difference Snow Index (NDSI) or binary versions of the MOD10A snow cover product. With improved fSCA retrieval algorithms for both MODIS and Landsat input data streams, along with appropriate treatment of data gaps, clouds, high sensor view angles, and saturation, an improved daily moderate-resolution snow cover fusion product may be realized. While the combination of newest generation satellites Landsat 8, Sentinel 2a, and Sentinel 2b provide observations of Earth's surface approximately every three days (Claverie et al., 2018), these new data do not address snow mapping in the historical context. In addition, as previously noted, snow can still accumulate, melt, or both within three days' period supporting the need for fusion in both the historical and future contexts. Analysis comparing the 30 m snow maps at the three-day interval to MODIS snow maps will inform our understanding of the various methods later described in estimating daily snow cover accurately before the application of interpolation or fusion methods.

Compared to previous methods, our efforts described in this paper represent a step forward through the use of spectral mixture analysis for snow cover mapping at both the Landsat and MODIS resolution together with a new machine learning approach to combine them. We describe the study area in the Sierra Nevada USA, the satellite data that include observations from Landsat and MODIS, and algorithms that include spectral mixture analysis used to estimate fSCA in Section 2. Section 2 also describes the fusion methods. In particular, we use random forests, but introduce a novel two-stage conditional approach that generalizes the original random forest concept. Validation metrics follow, along with sampling in Section 3. Section 4 shows the results and discusses the training sample size analysis, variable importance, seasonal performance, and limitations of the approach. Section 5 concludes with a summary and future directions.

2. Data and methods

Our study region in the Sierra Nevada USA spans from the North Fork of the American River basin in the north to the South Fork of the Kern River basin in the south (Fig. 1). This area of the Sierra Nevada is characterized by high vertical relief (up to 3000 m), with areas on the west side of the crest receiving far more precipitation than those to the east. For example, SWE on the ground measured at 2940 m on Mammoth Mountain near the crest reaches an annual peak of 128 cm (Bair et al.,

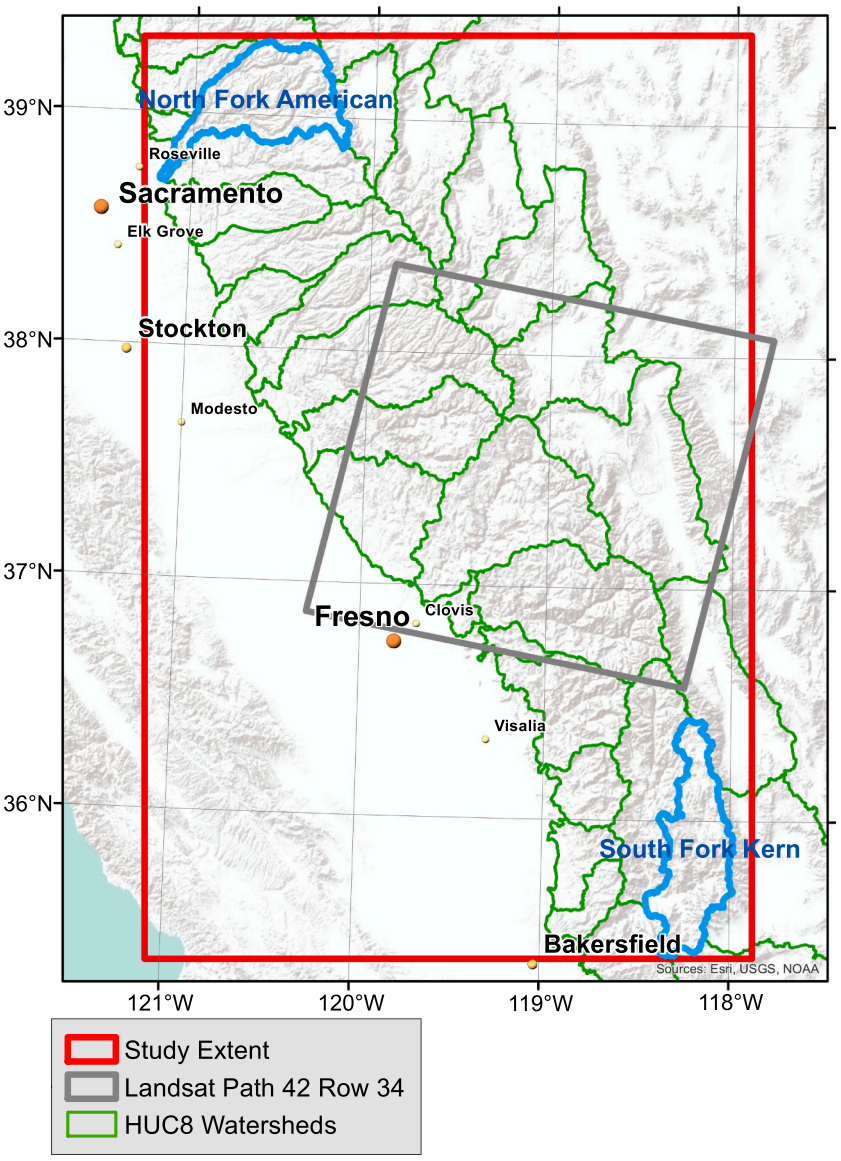

Fig. 1. Study extent in the California Sierra Nevada. HUC8 basins from the North Fork of the American River basin to the South Fork of the Kern River basin are shown in green.

2018b), while Bishop CA (1260 m), in the Owens Valley to the east, receives a scant 13 cm of annual precipitation and rarely has snow on the ground. Vegetation varies from chaparral to dense evergreen forests on the west side, to alpine tundra at the crest, to desert plants in the eastern valleys. The reliance of California on snow melt, the number of reservoirs, and both state and federal water projects in this region make the region of particular interest.

The MODIS and Landsat data used to generate fSCA to analyze our study region are described in Section 2.1. The snow cover fusion method we use to blend the input data is described in Section 2.2.

2.1. Fractional snow cover data

We used daily 500 m surface reflectance data from MODIS Terra, and 16-day 30 m surface reflectance data from Landsat 5 and Landsat 7 to generate fSCA maps for each sensor. We used available Landsat scenes in path 42 row 34 with up to 35% cloud cover and manually removed scenes with significant cloud over snow. The resulting number of images totaled 170 scenes in the eleven-year period from 2000 to 2011. The corresponding MODIS scene was processed for each date with a Landsat image. The 11-year Landsat record captures the majority of spatial variability in wet and dry years (Rittger et al., 2016), though it is worth noting that 2015, the driest year on record, and 2017, the wettest year on record, occurred outside this period. Periodically updating input data is likely to improve the results. We show examples for wet and dry years in the results and discussion in Section 4. Annual and monthly distributions of images are shown in Fig. 2. Landsat 7 suffered from an issue with the scan line corrector in 2003, so we selected only images before May 31, 2003, resulting in more images from 2000, 2001, and 2002, as seen in Fig. 2a. Cloud cover is more prevalent in the winter, resulting in a reduced number of images in October through February as seen in Fig. 2b. Sections 2.1.1 and 2.1.2 describe the process to create the fSCA maps from MODIS and Landsat data in more detail, respectively. These fSCA maps are then used as inputs to the data fusion method described in Section 2.2.

2.1.1. MODIS

The process to create fSCA maps from MODIS begins with MOD09GA Collection 6 (Vermote et al., 2015) atmospherically corrected surface reflectance available from the Land Processes Distributed Archive

Center (http://lpdaac.usgs.gov) in sinusoidal projection at 463 m spatial resolution. MOD09GA is used as input to the MODIS Snow Cover and Grain Size algorithm (MODSCAG), a spectral mixing algorithm that produces snow cover maps (Painter et al., 2009). However, the snow cover maps from MODSCAG cannot be utilized directly because the input MOD09GA surface reflectances suffer from data gaps and errors. Likewise, the ground is often obscured by cloud cover (Dozier et al., 2008). In addition, MODIS also sees only the viewable snow cover fraction. These errors in the input data are propagated through to the fSCA maps; thus, we must make corrections to the MODSCAG fSCA maps before they can be applied to the data fusion technique. First, we correct for tree projections and off-nadir viewing that masks snow to retrieve on-ground snow cover fraction (Raleigh et al., 2013; Rittger et al., 2020). We do this by using the time series of fractional snow cover along with MOD09GA quality flags and weight observations by the satellite view angle (Dozier et al., 2008), adjusting viewable snow cover to estimate snow cover on the ground using the vegetation endmember (Rittger et al., 2020). We reproject these corrected, gap-filled estimates of fractional snow cover on the ground to 500 m resolution and the Universal Transverse Mercator (UTM) coordinate system to match Landsat data. These data can now be used in training our fSCA fusion model as described in Section 2.2.2 and for prediction in Section 2.2.3.

2.1.2. Landsat

To create fSCA maps from Landsat, a similar spectral mixture algorithm that was used for MODIS was also applied to Landsat surface reflectance, as well as the same adjustment for viewable snow cover to estimate snow on the ground. As Landsat provides nadir views, no weighting is needed as with MODIS for off-nadir sensor views. Cloud cover was not present in many of the Landsat reflectance images and were manually removed when present. Previous work found little bias between the Landsat and MODIS estimates (Rittger et al., 2013; Rittger et al., 2021). However, Landsat 5 and Landsat 7 both frequently saturate over snow in bands 1, 2, and 3. For pixels that are not saturated in any band, we ran spectral mixture analysis with six bands, 1 to 5 and 7. If pixels have saturation in some bands, we used a reduced set of bands, choosing those that are not saturated. Note that band 1 typically saturates first, followed by band 3, then band 2. If bands 1, 2, and 3 are all saturated, we assume full snow cover. These methods remove clear biases introduced in other studies that assume binary snow cover at 30 m or do not address saturation at all. We used the Landsat fractional snow cover on the ground for training in the snow cover fusion described in 2.2.2.

The method described above differs from a similar implementation by the United States Geological Survey (USGS) (Department of the Interior, 2021). The USGS approach uses the same underlying spectral mixture algorithm but differs in both cloud masking and adjustment from viewable snow cover fraction to snow cover fraction on the ground. The USGS cloud masking employs an additional spectral test; however, mixed pixels can appear spectrally similar to clouds (Stillinger et al., 2019). The resulting snow maps may include both errors of omission and commission. The USGS approach also uses a static tree canopy percent to adjust viewable snow cover (Raleigh et al., 2013). As compared to our approach that uses the vegetation fraction estimated simultaneously, our approach is shown to produce improved results (Rittger et al., 2020). Additionally, the USGS neighborhood technique used in dense forest would increase biases between MODIS and Landsat snow mapping, possibly decreasing the accuracy of our fusion technique. A final reason for using our own procedure is that we hope to expand this work to snow albedo (Bair et al., 2019) based on snow grain size (Painter et al., 2009) and the reduction of snow albedo from light absorbing impurities (Painter et al., 2012). The USGS produces neither of these.

2.2. Snow cover fusion methods

In this section we briefly introduce random forests and discuss how they are adapted for training and predicting daily snow cover fraction at 30 m.

2.2.1. Random forests

We use a two-stage random forest algorithm trained on 16-day Landsat data, daily MODIS data, and physiographic information to predict 30 m fSCA. The first stage represents a classification step that predicts a class of zero, one, or non-boundary values (i.e., fSCA in the open interval (0,1)). For the non-boundary classification (not zero or one), a second random forest predicts values between zero and one fSCA. Before addressing the fusion technique further, we provide a brief background of random forests and discuss their utility as predictive models.

Decision trees are a simple type of statistical model. A single tree tends to produce poor predictions, but when combined in series can provide accurate predictions. The premise is that a decision tree splits into two branches based on features or predictors (e.g., MODIS fSCA $\geq$

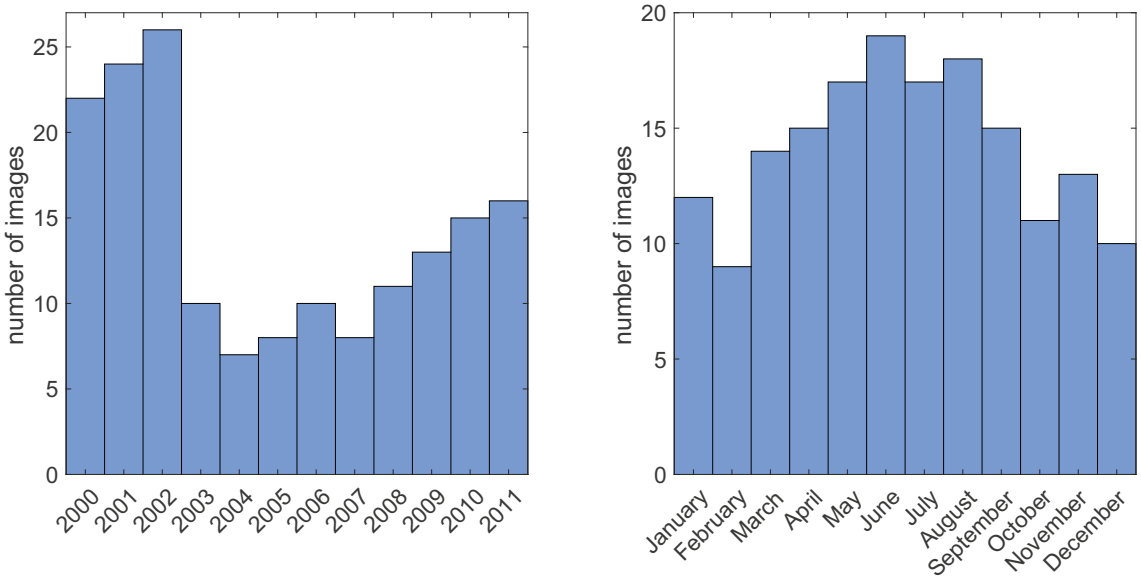

Fig. 2. Number of images used (a) by year (b) by month.

0.5 versus <0.5) and produces a separate fitted value on each side of the split. Predictions simply follow the tree branches according to the splits, using the final terminal value as the prediction. The main difficulties are in selecting the locations of the splits and associated prediction values for each subregion in the predictor space. Different goodness-of-fit criteria are used in the classification versus regression trees; but largely, the ideas remain the same.

Random forests (Breiman, 2001) are useful algorithms for both continuous data (a regression forest) and categorical data (a classification forest). In both cases, the forest is a collection of decision trees where the results from the decision trees are pooled together to create the final prediction. For classification, the final prediction is the most popular classification value (i.e., the mode) among the classification trees; while, for regression, the final prediction is the average value of the component regression tree predictions. The benefit of using multiple decision trees is a reduction in variance. That is, individual decision trees are highly variable models in that addition/removal of a few data points can drastically change individual tree predictions. The combination of trees in the forest reduces variability and provides a robust prediction framework (Hastie et al., 2009).

An important property of a regression tree is that it constrains predictions to lie within the range of the observed values. Therefore, using a single random forest would be a viable alternative to our two-stage forest approach, as it would yield predictions from 0 to 1. However, in initial experiments, we found the single-step forest tended to underestimate areas with snow cover of 1, and thus produce biased estimates of fSCA. When compared to a simpler approach, such as logistic regression for classification followed by linear regression, the random forest model was able to capture more complex features in the estimated snow map in early exploratory model fits.

We use the *ranger* R package for all random forest computations in this work (Wright and Ziegler, 2017). The bulk of *ranger* code is written in C++, but it is easily accessed through a user-friendly R package where the main commands are "ranger" for growing forests and "predict.ranger" for prediction. An important property of the random forest algorithm is that computations involving the individual decision trees can be run in parallel. In the *ranger* package, the default setting for these computations is the number of available cores.

2.2.2. Predictor variables

The two-stage random forest model at any given 30 m pixel uses the corresponding value from the 500 m MODIS fSCA, as well as several physiographic predictor variables at a 30 m resolution shown in Table 1, which correspond to geophysical features in our study area in the US Sierra Nevada (Fig. 1). These variables were used because of previous positive results (Liston et al., 2007; Fassnacht et al., 2012; Bair et al., 2018a) showing the predictive capability for snow. We also include the day of year as a predictor variable to account for potential seasonal interactions of the predictor variables. However, we note the discontinuity in the use of day of year and suggest future work use either a sine curve to approximate persistent year-to-year patterns or solar illumination to capture the inherent physical processes.

2.2.3. Two-stage random forests: classification and regression

Our approach separately fits a classification and regression random forest on training data and combines them during the prediction step. The total data size is approximately 1.172 billion pixels. In both cases, we use a completely randomized design for selecting training data points, resulting in a small subset of the available Landsat data and providing the opportunity for validation with >95% of existing observations from Landsat. We randomly choose a fixed number of pixels (see Section 4.1) on each day with available pairs of MODIS and Landsat imagery and given the accuracy of the predictions and robustness for increasing sample sizes, a stratified selection did not seem necessary. For the classification case, these are chosen from all available pixels. For the regression case, they are chosen from all available pixels with non-boundary Landsat values. The random forests are separately estimated but combined in the prediction step.

Table 1
Predictors variables (assigned at 30 m spatial resolution).

Name	Description
DOY	Day of year
Elevation	Height above sea level (SRTM, Farr et al., 2007)
Slope	Topographic slope calculated using elevation
Aspect	Topographic aspect calculated using elevation
Land cover type	National Land Cover Dataset (NLCD, Homer et al., 2007)
Latitude	Pixel center latitude
Longitude	Pixel center longitude
Forest height	LANDFIRE forest height (LANDFIRE, Rollins, 2009)
North/west/southwest barrier difference	Elevation difference between pixel and highest pixel in each direction also called shield height
West/southwest distance to ocean	Pixel distance to ocean in each direction
Windspeed	Average windspeed (Liston and Sturm, 1998; Cosgrove et al., 2003)

In Section 4.1, we display the effect of changing the number of randomly selected training points. It is worth noting that, at most, fewer than 5% of the available pixels are selected as training data (even at the largest number of points per day); but, we are still able to produce robust downscaled predictions as shown in Section 4 with our error metrics presented in Section 3. In our study, we use random forests with 100 trees and allow for splits over three variables at each node. In additional experiments (not shown), we found that using more trees (e.g., the default value 500) produced slightly better prediction scores at the cost of substantially increased computation time and memory. Since the total number of training pixels is much smaller than the total number of pixels to predict, we configured computational resources to favor a larger number of training points per day rather than a larger number of trees. The classification trees use the Gini impurity index (defined in Section 4.2) to select splits, while the regression trees use variance reduction (Hastie et al., 2009).

With the fitted forests, prediction is straightforward: at any pixel, the classification tree produces three fitted probabilities of fSCA falling into either class 0, the open interval (0,1) or 1. Note, for notational ease for the rest of the paper, we will call these three classes Z for zero, B for between, and H for 100%, respectively. The highest probability class is chosen. If this class is the non-boundary case B, the predicted value at that pixel is the associated prediction from the regression random forest. Such an approach will speed up predictions, since the regression random forest predictions are only necessary in pixels that are classified as non-boundary.

Random forests have several tuning parameters, including the number of features which are randomly sampled at each node, the minimum number of observations per tree leaf, as well as the number of trees. In our experiments, we followed the number of node split variables as recommended in Hastie et al. (2009), using the square root of the number of features. The minimum number of observations per tree leaf were set to the default values of the *ranger* package, namely 10 for classification and 5 for regression. We performed an experiment to determine the number of trees. For computational feasibility, we selected a subregion of the US Sierra Nevada containing 2,560,000 Landsat pixels and fitted random forest classification models with 100 to 500 trees in increments of 100. The results suggested that using 100 trees gave predictive fits with negligible degradation in accuracy compared to 500 trees (a typical default value used in practice); so, for the ensuing analysis, we use 100 trees.

A fundamental benefit of taking a statistical approach is the explicit availability of uncertainty measures. For example, Fig. 3a-c shows MODIS and Landsat inputs and the output from the two-stage random forest model for February 16, 2008. Fig. 3d-f shows the probabilities of falling in the three classes of Z, B, and H. Recall that the downscaling algorithm selects the class with highest probability in producing the

final estimate, but the associated probabilities can give a measure of confidence in the final estimate. For example, a close look at the probability maps in Fig. 3e reveals that there are nontrivial probabilities in the B class surrounding Lake Tahoe near the northern end of the study area, suggesting that there is reduced confidence of the final classification for this area. However, the probability in the H class is significantly less; and the probability in the Z class is significantly more than both; so, Z is chosen as the predicted value. We also note some apparent artifacts in the strictly nonzero (Fig. 3e and f) classification probability maps aligning with the footprint of the Landsat data shown in Fig. 1 and Fig. 3c. The square region artifact appears where Landsat training data are available, and the model tends to be more confident in classifications in this region because of such availability of training data; however, the approach is able to extrapolate to regions without direct training imagery, at a cost of lower class confidences.

3. Validation metrics and sampling

There are several predictor variables with missing (not-applicable) or null values and also Landsat images with null values over the study region, for example, in cloudy or deeply shaded areas and large water bodies. These pixels are removed from validation and sampling along with pixels saturated in all three bands as described in Section 2.1.2. Pixels saturated in all three bands would lead to an overestimate in snow cover (Rittger et al., 2021). To grow the random forests, we select a subset of the remaining data in the study region and create the forest using only these training data. As previously described, the sampling for training is done randomly by drawing a randomly selected set of the viable data over each of the 170 days. We considered daily sample sizes of 5×10^4, 1×10^5, 2×10^5, and 3×10^5 pixels. In computational terms, the largest sample size corresponds to growing a forest with about 51 M training data points.

Validation is performed using all data not used in training. Summary statistics are calculated on a day-to-day basis by comparing an estimated image (C) to the true image (T). In our case, the estimated image C is snow cover fraction from the two-stage random forest model and T is the Landsat snow cover. We use a set of binary metrics that rely on common classification of identifying a pixel in C as containing snow or not. True positive (TP) indicates that the estimated and true fSCA values are both larger than zero ($C > 0$ & $T > 0$); true negative (TN) indicates that the estimated and true fSCA values are both zero ($C = 0$ & $T = 0$); false positive (FP) indicates that the estimated fSCA value is larger than zero, but the true fSCA value is zero ($C > 0$ & $T = 0$); and false negative (FN) indicates that the estimated fSCA value is zero, but the true fSCA value is larger than zero ($C = 0$ & $T > 0$). Based on these values, we report precision, recall, specificity, the F-score, and accuracy.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

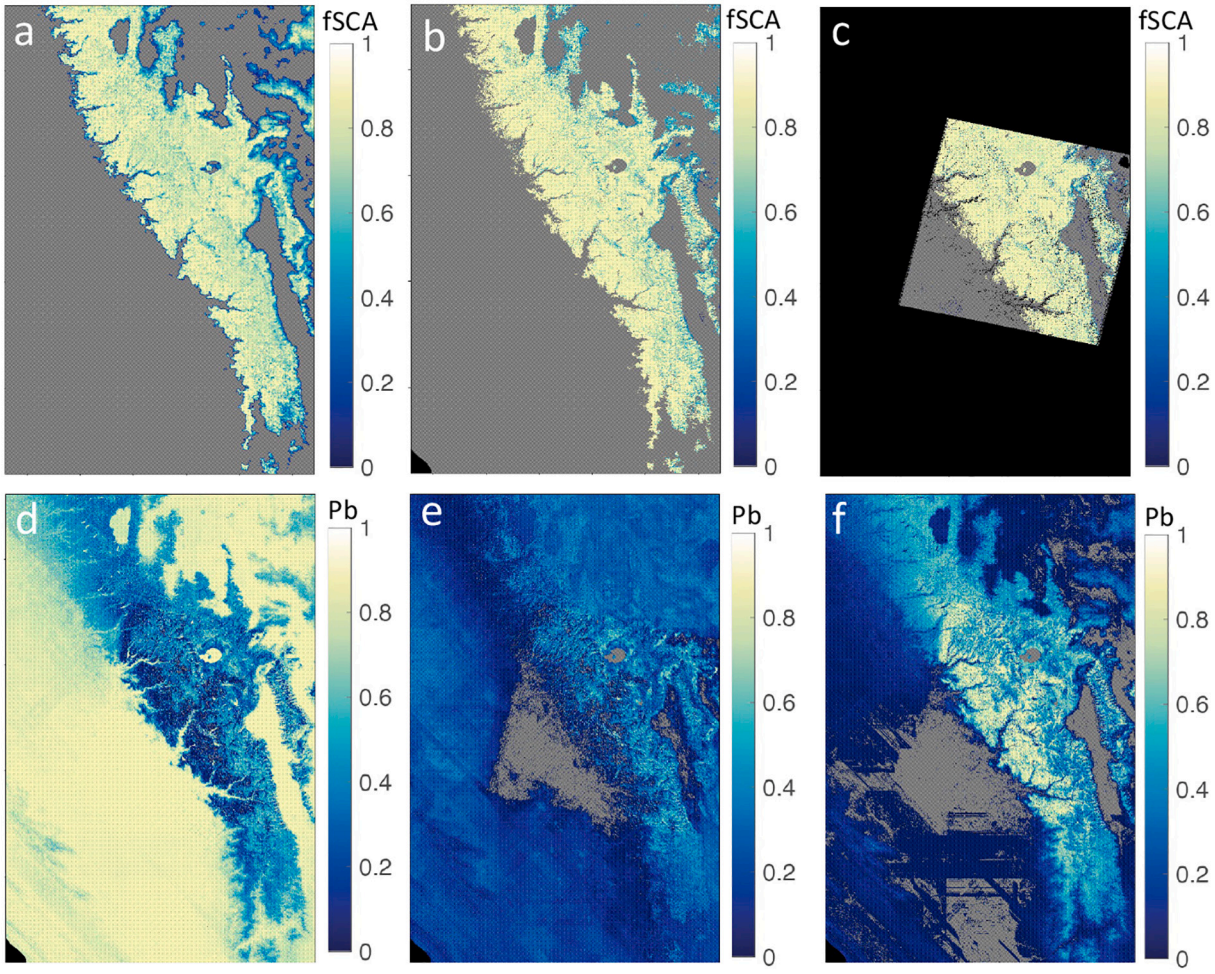

Fig. 3. Snow cover inputs and downscaling uncertainty covering the region in the Study Extent in Fig. 1 for February 16, 2008. The top row contains (a) MODIS fSCA at 500 m, (b) two-stage random forest fSCA at 30 m and (c) Landsat fSCA for this day. The bottom row contains probabilities (Pb) of (d) class Z, 0 fSCA, (e) class B, a value between 0 and 1 fSCA and (f) class H, 1 fSCA, respectively.

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

$$F = 2\,\frac{Precision\ x\ Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN} \tag{4}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

A small precision (Eq. 1, positive predictive value) indicates many false positives relative to true positives. A small recall (Eq. 2, sensitivity or true predictive rate) indicates many false negatives compared to true positives. A small specificity (Eq. 3, true negative rate) indicates many false positives relative to true negatives. The F-score (Eq. 4) incorporates both recall and precision. Accuracy (Eq. 5) quantifies positive classifications as a ratio with the total number of pixels.

We also consider the case of fractional snow cover where snow is greater than zero in either or both C or T. We ignore the zero values so as not to bolster the error estimates by the numerous snow free areas. We report the median difference of C-T, mean difference of C-T, and the RMSE (Eq. 6) of C-T.

$$RMSE = \sqrt{\frac{1}{N}\sum_{N}\left(f_{SCA}^{C} - f_{SCA}^{T}\right)^{2}} \tag{6}$$

4. Results and discussion

To illustrate the fSCA based on our approach, we consider two example days with distinct climatological characteristics. For the ensuing results, we fit the two-stage model and downscale on a cluster with 193 GB of RAM and 2× Intel Xeon Gold 6130, 2.1GHz 16-core Skylake processors. Fig. 4 shows data for January 25, 2006, and March 17, 2007, which were unusually dry and wet years respectively (Rittger et al., 2016). The MODIS fSCA used as a predictor is shown in Fig. 4a and d, downscaled predictions in Fig. 4b and e, and Landsat validating values Fig. 4c and f. The two-stage downscaling model successfully identifies the areas in both climatological extremes of where and how much snow is present. The high values of precision, recall, specificity, F-score, and accuracy for January 25, 2006, are shown in Fig. 4b and for March 17, 2007, are shown in Fig. 4e. The predictive RMSEs over true or predicted positive fSCA (i.e. excluding snow-free areas) are 0.25 and 0.34 for the two days.

Generally, the downscaling approach produces the finer resolution that is a physically reasonable portrayal of the distribution of fSCA over the mountainous terrain. For example, the California Central Valley in the southwest is snow free, and the blocking effect of the Sierra Nevada leaves higher elevations snow-free to the northeast. Another effect of downscaling that is evident on both dates is the clear sharpening of the edges and gradient of fSCA between areas of no snow and snow. This is visible in Fig. 4b and e near the snowline where snow transitions quickly

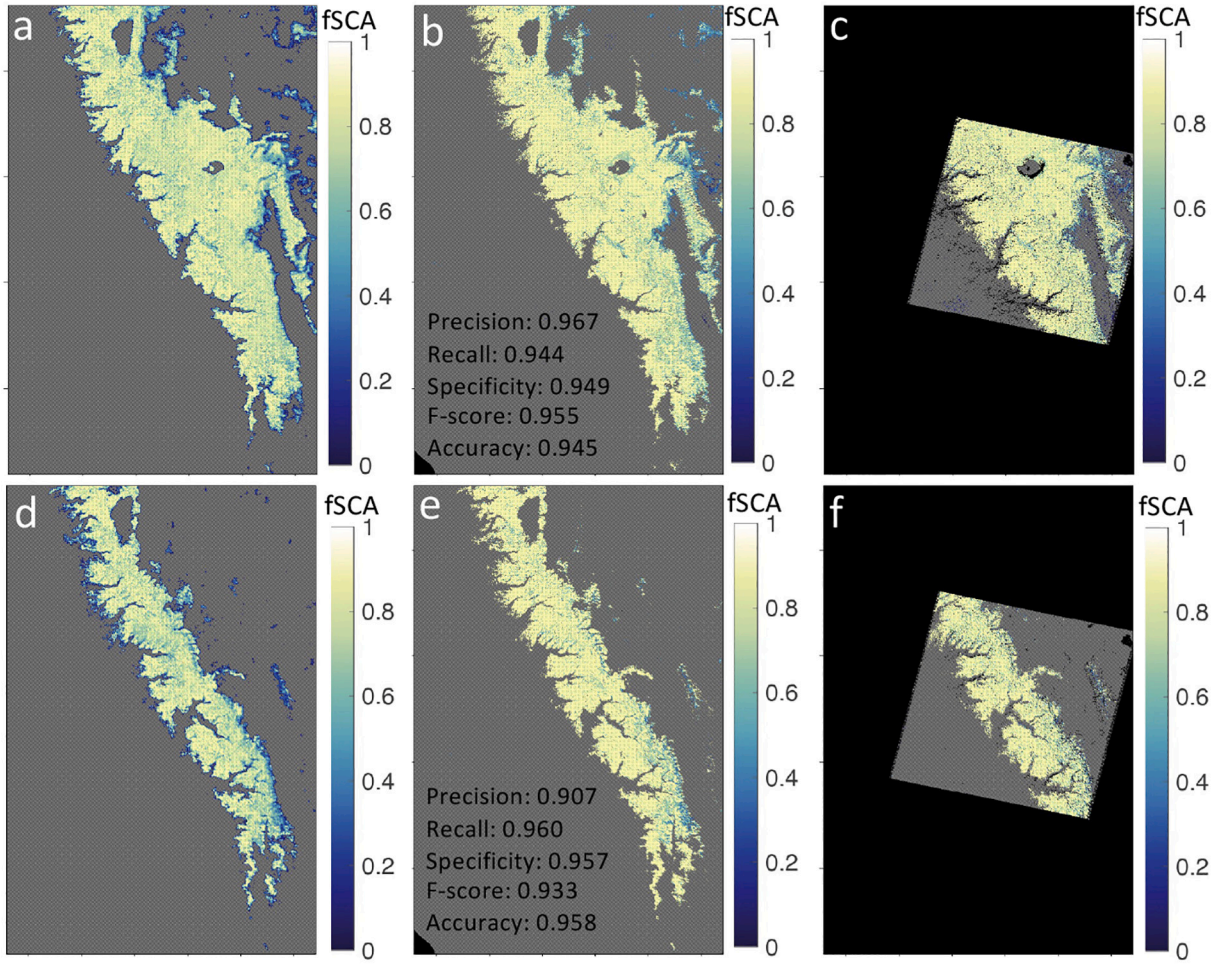

Fig. 4. Example of spatially downscaled predictions covering the Study Extent shown in Fig. 1. Images are fSCA for MODIS (a and d), downscaled prediction (b and e) and validating Landsat (c and f) for January 25, 2006 (top row), and March 17, 2007 (bottom row). Black indicates missing values.

from a high percent to zero. The coarser resolution MODIS snow cover (shown in Fig. 4a and d) tends to smear or smooth out these gradients, while the modeled snow cover (Fig. 4b and e) produces a more detailed spatial distribution like that of Landsat (Fig. 4c and f).

Figs. 5 and 6 show error characteristics for the days corresponding to Fig. 4. Fig. 5 contains the average of the downscaled estimates along with one standard deviation of the downscaled estimates for Landsat values in bins of size 0.1. Note that the x-axis values correspond to the average Landsat fSCA in each bin (e.g., the bin [0,0.1] has an average Landsat fSCA of close to zero in both days). We see a slight tendency for the downscaled predictions to overestimate fSCA at varying levels, which is more extreme in the latter image. However, error bars contain the identity line (indicating perfect downscaling). Fig. 6 shows RMSE as functions of elevation, slope, and aspect for both days, in 20 equally spaced bins. Errors are elevated at higher elevations and higher slopes, but generally around 0.25 for the worst cases, while there is not as much apparent structure over varying aspects.

At the scale of the entire Sierra Nevada, it is difficult to discern some of the detail in the fSCA. In Fig. 7a and e, false color images are shown where snow is cyan, soil/rock is red, forests are green, and lakes are black for December 8, 2002 (top), and April 23, 2003 (bottom). The fractional snow cover is shown in Fig. 7b and f from Landsat and Fig. 7c and g from MODIS. Fig. 7d and h show the fused snow cover at 30 m. Deep shadows and areas masked by clouds in the Landsat data for both dates are shown as black, but stand out in the December image (Fig. 7b). The modeled fSCA shown in Fig. 7d appears to correctly produce zero values in the valleys and lower elevations in the east while also filling deeply shadowed area in the west with positive values. In both Fig. 7c and g, MODIS fSCA, the mixed pixels at the snow line smear fully or mostly covered pixels with no-snow pixels, while the higher spatial resolution of the fused version (Fig. 7d and h), better spatially distributes the snow cover.

Higher spatial resolution is important in accurately representing the influence of snow cover in winter landscapes. In mountainous terrain at the latitudes of this study, aspect strongly affects the persistence of snow cover. This effect is pronounced at fine spatial scales, particularly in deeply dissected mountains such as the Sierra Nevada. Snow ablates on south facing aspects due to increased solar radiation and is redistributed from west facing aspects by wind. Coarser scale images often fail to characterize the variation in snow cover by aspect in steep mountainous terrain.

From both visual inspection and validation metrics (Table 2 and Table 3), the initial results demonstrate skill in improving the spatial resolution of daily snow cover from MODIS. Next, we discuss the details of model training and consider quantifying the quality of downscaled predictions.

4.1. Training sample size analysis

Here we present the validation metrics for each of the forests grown with different daily training sample sizes. The data are initially split into training and testing subsets. We vary the classification forest training sample size by varying the number of randomly chosen pixels per training day (of 170) between 5×10^4 and 3×10^5. The second-stage regression forest is always trained on the subset of the classification training data that are non-boundary, and so it may vary based on the day of year and component samples chosen for the classification training set. We consider total training samples of size 8.5×10^6, 17×10^6, 34×10^6, and 51×10^6; these values correspond to random sampling from each of 170 available days of data 5×10^4, 1×10^5, 2×10^5, and 3×10^5 pixels. Hold-out testing/validation sets are the total sample size (approximately 1.172 billion) minus the training size, resulting in testing/validation data sizes of more than 1.1 billion.

Validation statistics are provided in Table 2 (see Section 3 for details) using validation set Landsat data, ignoring saturated pixels. The results are averaged temporally across the 170 available days and weighted by the number of pixels in each Landsat scene. In the final row of Table 2, we report the computation time for growing the classification forest with the training data. The classification step in the process of fitting our model dominates the total computing time and the total memory usage.

Based on the results in Table 2, we see a slight improvement in predictive performance by increasing sample size. However, the improvements are relatively small compared to the substantial increase in computation time required to fit the larger sample size datasets. For example, the largest forest (3×10^5) required over 150 GB of RAM and two days of computation time on a node with 32 multithreaded cores, while the smallest forest (5×10^4) finished growing in about an hour on the same hardware. A forest with 400,000 training points per day was also tested, but it exhausted the available memory limit of 180 GB (13 GB allocated for operating system) before growth was complete. A two-step random forest approach for estimating fSCA seems to provide robust downscaled predictions even when using a very small portion of data to fit the model. Given the accurate results of training on such a small portion of the data, we could use Landsat 7 after May 31, 2003, with the scan line corrector off, sampling pixels with available data in future modeling efforts.

For the smallest number of samples (in Table 2), the specificity statistic is the highest, indicating that in regions with no snow the predicted fSCA rarely indicates that there is snow. The accuracy metric, which uses all cases, is the next highest statistic and reports a value of 0.965. This refers to the ability of the predicted products to correctly specify whether there is snow or not, which is attained for all but 3.5% of cases. Recall is lower than precision, indicating that the predicted snow cover

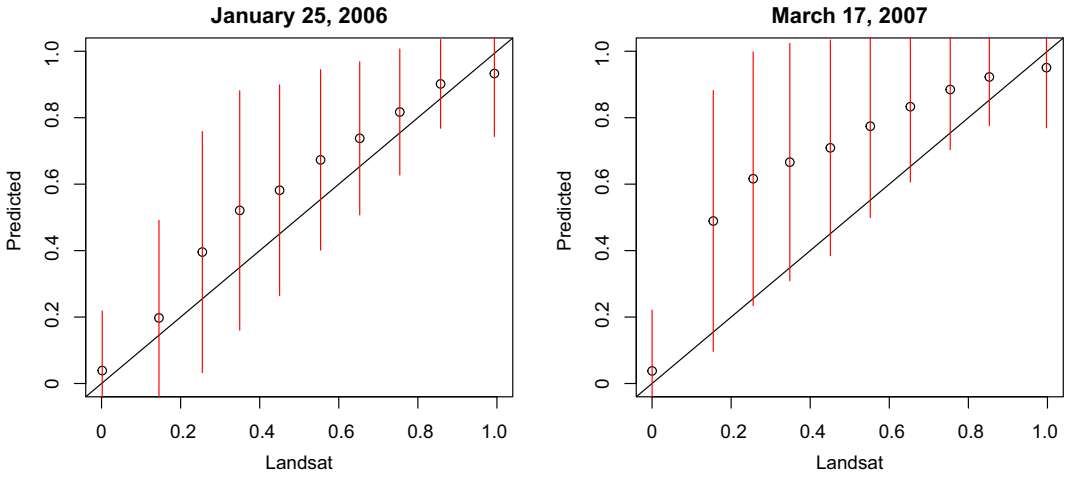

Fig. 5. Downscaled fSCA versus true Landsat corresponding to the days in Fig. 4. Vertical bars indicate one standard deviation.

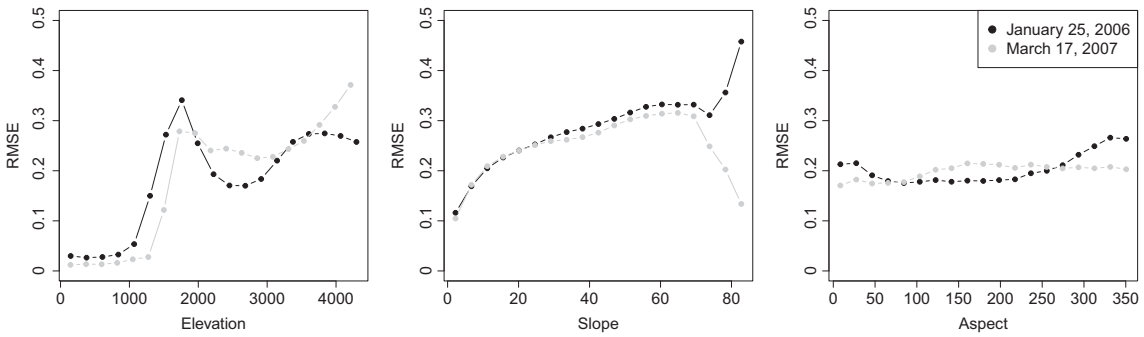

Fig. 6. Root mean squared error fSCA as a function of elevation (m), slope (degrees) and aspect (degrees) for the two example days of Fig. 4.

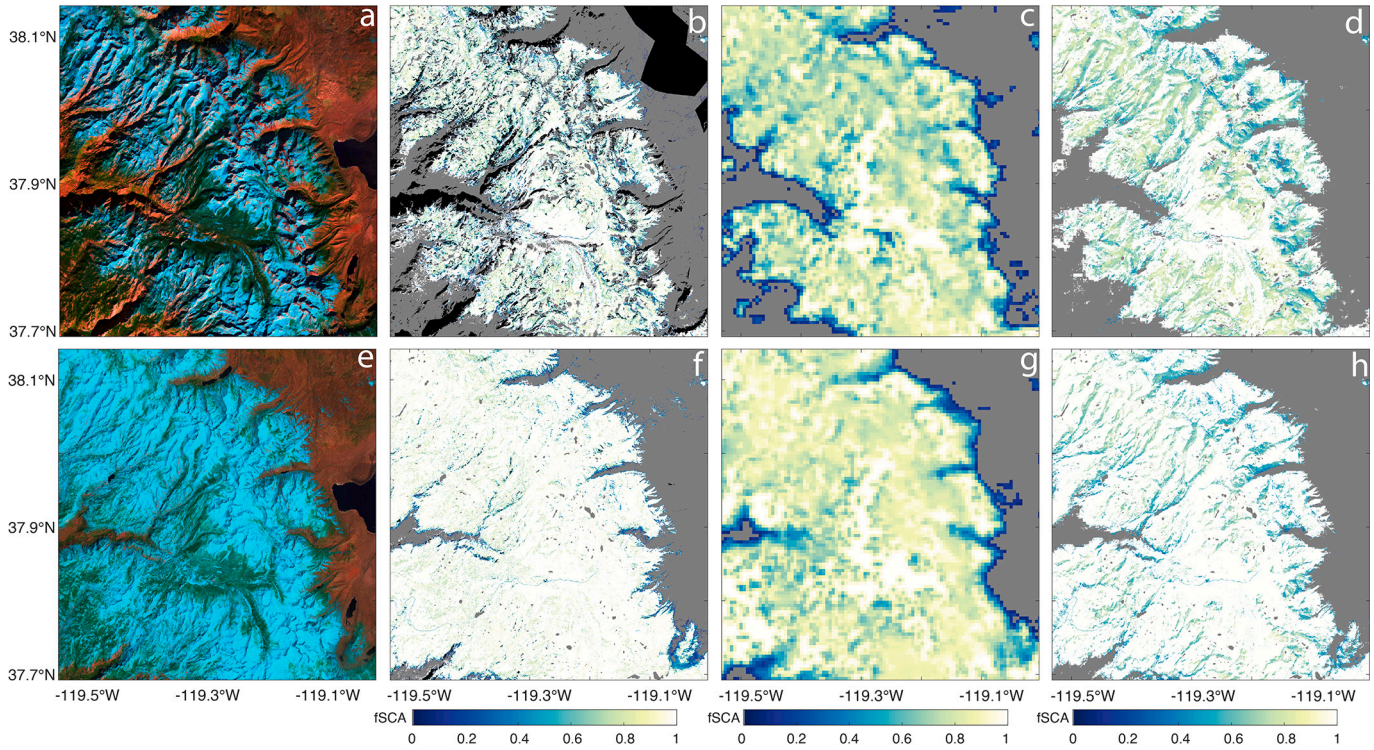

Fig. 7. For December 8, 2002 (top) and April 23, 2003 (bottom), false color using band 6, 4, and 2 which shows snow as cyan, rock and soil as red, forest as green, and water as black (a & e), fSCA from Landsat (b & f), fSCA from MODIS (c and g), and fused snow cover at 30 m (d and h).

Table 2
Validation metrics for random forest predictions as functions of training sample size. Columns represent different numbers of training sample sizes per day. Rows represent validation statistics (see text for details) averaged over all validation data.

Training samples per day	5×10^4	1×10^5	2×10^5	3×10^5
Precision	0.932	0.935	0.939	0.941
Recall	0.887	0.892	0.898	0.901
Specificity	0.984	0.985	0.985	0.986
F score	0.909	0.913	0.918	0.921
Accuracy	0.965	0.966	0.968	0.969
Mean difference	0.003	0.003	0.004	0.004
Root mean squared error	0.340	0.330	0.320	0.314
Computation time	1.2 h	6 h	22.8 h	40.5 h

maps miss snow more frequently than falsely prescribe snow. While the RMSE is 0.34 for pixels with snow (i.e. excluding zeros) and errors are balanced around zero with a mean fSCA difference of 0.003 as we would expect with a random forest model.

Table 3
Same cross-validation results as Table 1 for the final model split by season. Acronyms are December, January, February (DJF), March, April, May (MAM), June, July, August (JJA) and September, October, November (SON).

	DJF	MAM	JJA	SON
Precision	0.94	0.93	0.94	0.95
Recall	0.90	0.88	0.91	0.91
Specificity	0.99	0.99	0.98	0.99
F score	0.92	0.91	0.92	0.93
Accuracy	0.97	0.97	0.96	0.98
Mean difference	0.01	0.00	0.01	0.00
Root mean squared error	0.32	0.33	0.32	0.29

4.2. *Variable importance*

Individual decision trees are useful for model interpretation and determining relative influence of particular features and predictors on the final model prediction. Determining the relative influence of features in random forests is less straightforward due to multiple trees but is

often quantified using variable importance measures. As we have separate classification and regression forests, we use two standard measures of variable importance: the Gini index for classification (Hastie et al., 2008), and variance reduction for regression. The *ranger* package provides an option for growing a random forest with the "importance" option, providing both indices as byproducts of the estimation framework.

The Gini index can be interpreted as quantifying the decrease in node impurity associated with a particular predictor which is calculated from the Gini node impurity or the power of discrimination across the classes. For a feature that is split over a given node t with J classes (in our case $J = 3$), the Gini node impurity is

$$\sum_{j=1}^{J} \widehat{p_j}(t)\left(1-\widehat{p_j}(t)\right) \tag{7}$$

where $\widehat{p_j}(t)$is the estimated probability of class j at node t. Lower values of Gini impurity indicate a purer node. This can be intuitively seen: if the jth feature is discriminative at node t, then $\widehat{p_j}(t)$ will be close to zero or one, and the corresponding Gini impurity will be smaller. By contrast, if the feature is not discriminative then $\widehat{p_j}(t) \approx 0.5$ and the product $\widehat{p_j}(t)\left(1-\widehat{p_j}(t)\right)$ will be maximized, leading to a larger Gini impurity. The Gini index is the decrease in impurity between a given node and weighted impurities at the two child nodes; larger values of the Gini index indicate more important features.

The variable importance measure for regression is variance reduction where the importance for a feature that is split at a given node is calculated as the improvement in predictive squared error compared to no split at that node. Large variance reduction indicates more important features that have good predictive capability. For random forests, in both cases, the variable importance measures are summed over all nodes and all trees on which a particular feature is split.

The scaled variable importance metrics for the classification and regression forests show the relative importance of each predictor (Fig. 8), with higher values indicating higher importance. Note that importance is based on different statistics for classification and regression, so the values are not directly comparable. However, in both forests the most important predictor variables are MODIS fSCA, day of year, and elevation.

4.3. Seasonal performance

We now consider the behavior of the two-stage model over different seasons. A seasonal breakdown of the overall statistics in Table 2 shows little impact of seasonality on the model performance (Table 3). Results are reported only for the largest training size (3×10^5 samples per day). This lack of seasonality in the statistics indicates that the model can capture many characteristics of snow cover that may differ across time and space. For example, spatial distribution during accumulation and melt periods can differ. Snow is lighter and may be redistributed by wind more easily during accumulation; while during snowmelt, snow is denser and less likely to be transported. In addition, topographic shadowing may result in different patterns during melt and accumulation because of seasonal changes in solar geometry.

Fig. 9 illustrates the seasonality of each statistic, showing averages for each day of the year. A large seasonal variability is seen in recall and F-score which is not obvious in Table 3 statistics. This is because those scenes with low recall and F-score have very few pixels and as such do not heavily weigh the seasonal statistics. Fig. 9 shows that the model performs well, except for the summer where recall is low. Starting around June, the recall (and corresponding F-score) falls from near one to near zero, which suggests the classification procedure produced many false negatives in this period. In particular, the model failed to identify snowy regions and instead estimated no snow in the area, which is not surprising because MODIS does not easily detect the small and heterogeneous patches of snow in the summer.

Fig. 10 and Fig. 11 illustrate conditional performance of the model for non-boundary values of fSCA. Specifically, Fig. 10 shows mean/median difference and RMSE over days of the year for only those values of Landsat snow cover in class B.

Fig. 11 shows the same statistics, but for the cases where the downscaling model also predicts values in class B. It is interesting to note that when the model correctly predicts nontrivial fSCA, there is essentially zero bias during the summer. This suggests that the negative bias seen in Fig. 10 is primarily caused by the misclassification error in the first stage forest's misclassification error, possibly from patchy snow, rather than the second stage forest error.

4.4. Limitations of the approach and future improvements

This approach is limited by using MODIS imagery at 500 m. That is, the coarse resolution of MODIS implies that 16×16 nested pixels in the target Landsat image share the same MODIS prediction value. This results in a byproduct of apparent artifacts in the fused images that sometimes show clear boundaries along MODIS pixel edges. We are currently implementing the same framework with pre-smoothed MODIS imagery, which is bilinearly interpolated during reprojection from sinusoidal projection at 463 m to match the scale of 30 m Landsat in UTM projection to produce a 20-year time series based on the methods presented in this paper. A second drawback is that the statistical model is fit globally—that is, the two-stage model is fit on, and predicts, values over the whole region. However, we expect the statistical relationship to vary between the physiographic features, MODIS, and Landsat with geographic location. Future efforts will be devoted to developing locally

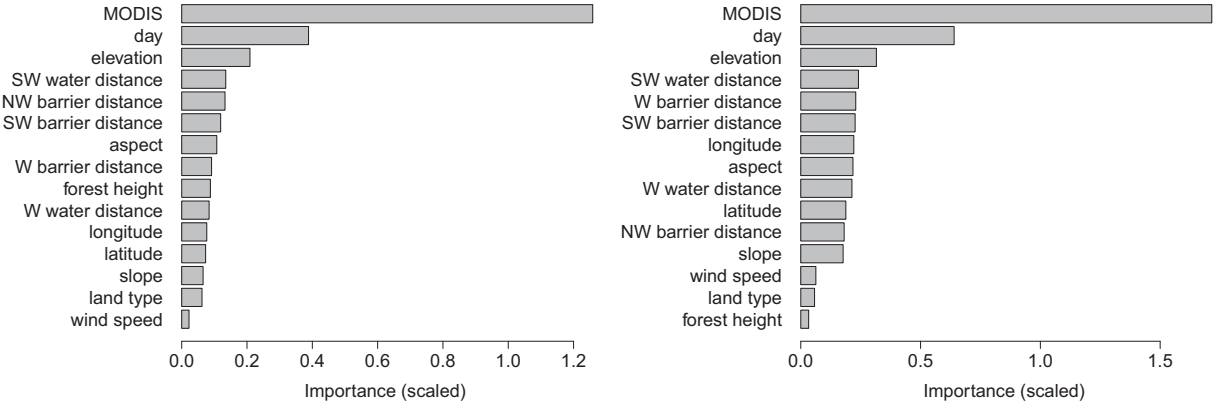

Fig. 8. Variable importance for the classification forest based on the Gini index (left) and for the regression forest based on variance reduction (right). Values on the left plot are scaled by $1 \times \mathbf{10^6}$and values on the right plot are scaled by $1 \times \mathbf{10^8}$.

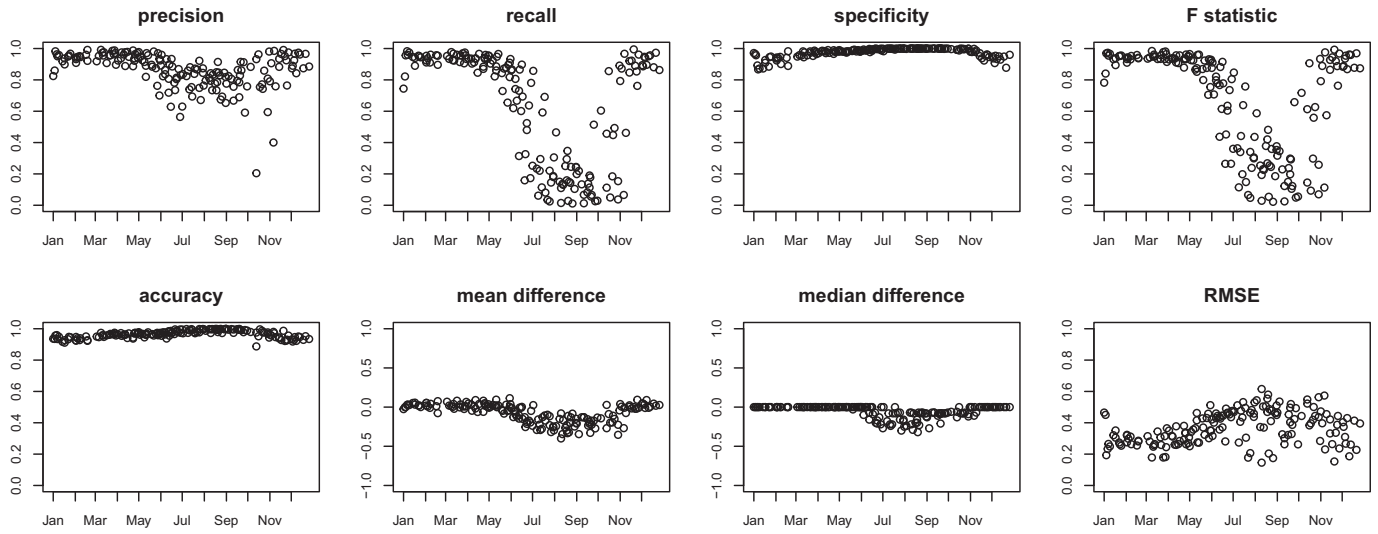

Fig. 9. Validation metrics for the downscaled prediction versus validating Landsat image, plotted by the corresponding day over all 170 scenes used.

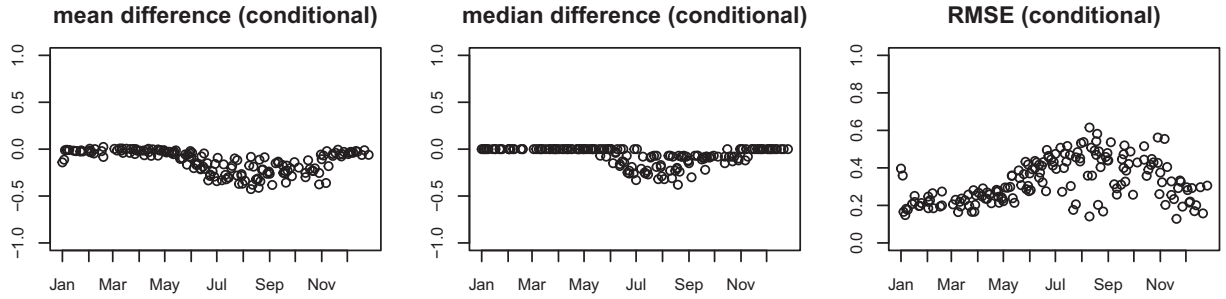

Fig. 10. Conditional statistics for all pixels where the true Landsat values are in class B for all 170 scenes used.

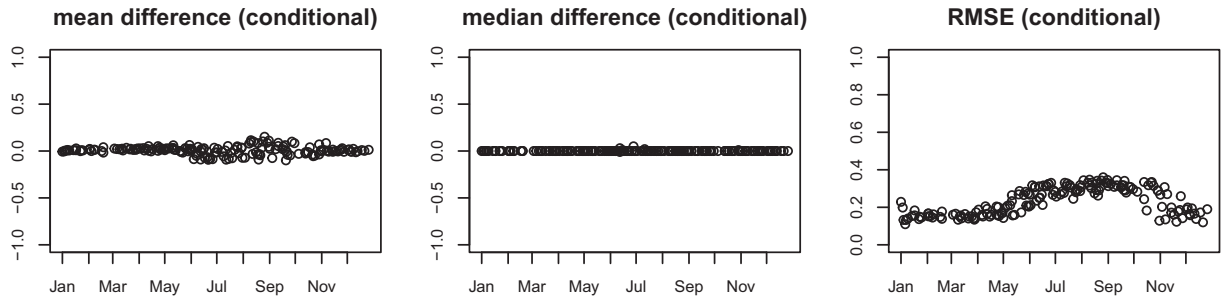

Fig. 11. Conditional statistics for all pixels where the true Landsat values and the predicted values are in class B over all 170 scenes used.

sliding models that can better adapted to local characteristics. Finally, excluding cloudy pixels is necessary for the best training data in each month. Using manually cloud masked scenes for Landsat is time intensive, but clouds are spectrally similar to snow in certain conditions (Stillinger et al., 2019); so, snow cannot always be separated from cloud automatically for each pixel. Given the small number of Landsat pixels needed, future work should focus on an automated process to select pixels over a range of fSCA that we are confident are snow.

5. Conclusion

We used a scale-invariant, two-stage random forest model along with the best available fractional snow cover estimates from spectral mixture analysis to create daily snow cover maps at 30 m. Validation metrics show that, with adequate sampling over time and space, we can combine infrequent snow mapping from 30 m data with daily data at 500 m to produce accurate daily maps of snow at 30 m resolution. The model performed similarly in all seasons of the year and at different aspects with larger errors for steeper slopes and higher elevations. Daily MODIS snow cover, day of year, and elevation were the three most important predictor variables.

Alternative validation data such as airborne data or high-resolution commercial observations will be useful for continuing validation in a range of topography including shadowed and forested areas. Snow cover under the tree canopy continues to challenge both estimatation and validation of snow on the ground, but lidar and buried temperature sensor networks can improve our understanding of snow in forests. Moving forward, observations from the Visible Infrared Imaging Radiometer Suite (VIIRS) instrument on Suomi National Polar-orbiting Partnership (NPP) and Joint Polar Satellite System (JPSS-1 and JPSS-

2) will be critical to extend the current record of fSCA from the MODIS instrument. Landsat 8 and Sentinel 2a and 2b can provide a more complete set of training data without suffering from the saturation problems of previous sensors. Spectral mixture analysis provides snow surface properties such as: snow grain size, snow albedo, and the impact of light absorbing particles. Application of this fusion method to these snow surface properties is planned and will further extend our understanding of the temporal and spatial distribution of snow. Ultimately, these fused products will be used to improve our SWE and streamflow forecasting efforts, as well as to improve Sierra Nevada bighorn sheep recovery efforts.

Funding

California Department of Fish and Wildlife, NASA award 80NSSC18K1489; NASA award 80NSSC18K0427; NOAA award NA18OAR4590380, and University of California award LFR-18-54831.

Data statement

To satisfy NASA open data policies, all data are available in online repositories in GEOTIFF format from Landsat and MODIS snow cover, predictors from Table 1, and snow cover from the 2-stage random forest model. (Rittger, 2021)

The 2-stage random forest model is currently on a public repository in GitHub (Krock et al., 2019).

CRediT authorship contribution statement

Karl Rittger: Conceptualization, Methodology, Formal analysis, Data curation, Visualization, Writing – original draft, Writing – review & editing, Project administration, Funding acquisition. **Mitchell Krock:** Methodology, Software, Validation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. **William Kleiber:** Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Supervision. **Edward H. Bair:** Conceptualization, Software, Formal analysis, Writing – original draft, Writing – review & editing, Funding acquisition. **Mary J. Brodzik:** Software, Formal analysis, Data curation, Writing – review & editing, Funding acquisition. **Thomas R. Stephenson:** Conceptualization, Writing – review & editing. **Balaji Rajagopalan:** Conceptualization, Methodology, Writing – review & editing. **Kat J. Bormann:** Writing – review & editing, Funding acquisition. **Thomas H. Painter:** Methodology, Software, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bair, E.H., Rittger, K., Davis, R.E., Painter, T.H., Dozier, J., 2016. Validating reconstruction of snow water equivalent in California's Sierra Nevada using measurements from the NASA airborne snow observatory. Water Resour. Res. 52, 8437–8460. https://doi.org/10.1002/2016WR018704.

Bair, E.H., Calfa, A.A., Rittger, K., Dozier, J., 2018a. 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.

Bair, E.H., Davis, R.E., Dozier, J., 2018b. Hourly mass and snow energy balance measurements from Mammoth Mountain, CA USA, 2011–2017. Earth System Science Data 10, 549–563. https://doi.org/10.5194/essd-10-549-2018.

Bair, E.H., Rittger, K., Skiles, S.M., Dozier, J., 2019. An examination of snow albedo estimates from MODIS and their impact on snow water equivalent reconstruction. Water Resour. Res. 55, 7826–7842. https://doi.org/10.1029/2019wr024810.

Barnett, T.P., Adam, J.C., Lettenmaier, D.P., 2005. Potential impacts of a warming climate on water availability in snow-dominated regions. Nature 438, 303–309. https://doi.org/10.1038/nature04141.

Baumgartner, M.F., Seidel, K., Martinec, J., 1987. Toward snowmelt runoff forecast based on multisensor remote-sensing Informnation. IEEE Trans. Geosci. Remote Sens. GE-25, 746–750. https://doi.org/10.1109/TGRS.1987.289744.

Berman, E.E., Bolton, D.K., Coops, N.C., Mityok, Z.K., Stenhouse, G.B., Moore, R.D., 2018. Daily estimates of Landsat fractional snow cover driven by MODIS and dynamic time-warping. Remote Sens. Environ. 216, 635–646. https://doi.org/10.1016/j.rse.2018.07.029.

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

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.C., Skakun, S.V., Justice, C., 2018. The harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sens. Environ. 219, 145–161. https://doi.org/10.1016/j.rse.2018.09.002.

Conner, M.M., Stephenson, T.R., German, D.W., Monteith, K.L., Few, A.P., Bair, E.H., 2018. Survival analysis: informing recovery of Sierra Nevada bighorn sheep. J. Wildl. Manag. 82, 1442–1458. https://doi.org/10.1002/jwmg.21490.

Cosgrove, B.A., Lohmann, D., Mitchell, K.E., Houser, P.R., Wood, E.F., Schaake, J.C., Robock, A., Marshall, C., Sheffield, J., Duan, Q., Luo, L., Higgins, R.W., Pinker, R.T., Tarpley, J.D., Meng, J., 2003. Real-time and retrospective forcing in the north American land data assimilation system (NLDAS) project. J. Geophys. Res.-Atmos. 108, 8842. https://doi.org/10.1029/2002jd003118.

Department of the Interior, U.S.G.S, 2021. Landsat Fractional Snow Covered Area (fSCA) Algorithm Description Document (ADD). https://www.usgs.gov/media/files/landsat-fractional-snow-covered-area-add.

Dobreva, I.D., Klein, A.G., 2011. Fractional snow cover mapping through artificial neural network analysis of MODIS surface reflectance. Remote Sens. Environ. 115, 3355–3366. https://doi.org/10.1016/j.rse.2011.07.018.

Dozier, J., Painter, T.H., Rittger, K., Frew, J.E., 2008. Time-space continuity of daily maps of fractional snow cover and albedo from MODIS. Adv. Water Resour. 31, 1515–1526. https://doi.org/10.1016/j.advwatres.2008.08.011.

Durand, M., Molotch, N.P., Margulis, S.A., 2008. Merging complementary remote sensing datasets in the context of snow water equivalent reconstruction. Remote Sens. Environ. 112, 1212–1225. https://doi.org/10.1016/j.rse.2007.08.010.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The shuttle radar topography Mission. Rev. Geophys. 45, RG2004 https://doi.org/10.1029/2005rg000183.

Fassnacht, S.R., Dressler, K.A., Hultstrand, D.M., Bales, R.C., Patterson, G., 2012. Temporal inconsistencies in coarse-scale snow water equivalent patterns: Colorado River basin snow telemetry-topography regressions. Pirineos: Revista de Ecología de Montaña 167, 165–185. https://doi.org/10.3989/Pirineos.2012.167008.

Hall, D.K., Riggs, G.A., Salomonson, V.V., DiGirolamo, N.E., Bayr, K.J., 2002. MODIS snow-cover products. Remote Sens. Environ. 83, 181–194. https://doi.org/10.1016/S0034-4257(02)00095-0.

Hall, D.K., Riggs, G.A., Foster, J.L., Kumar, S.V., 2010. Development and evaluation of a cloud-gap-filled MODIS daily snow-cover product. Remote Sens. Environ. 114, 496–503. https://doi.org/10.1016/j.rse.2009.10.007.

Hastie, T., Tibshariani, R., Friedman, J., 2008. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, Second ed.

Hastie, T., Tibshariani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag.

Homer, C.G., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A.J., VanDriel, J.N., Wickham, J., 2007. Completion of the 2001 National Land Cover Database for the conterminous United States. Photogramm. Eng. Remote. Sens. 73, 337–341. http://pubs.er.usgs.gov/publication/70029996.

Krock, ML, Kleiber, W, Brodzik, MJ, Rajagopalan, B, Rittger, K, 2019. ESPFusion: Data fusion for downscaling Earth Surface Properties [Source code]. github.com/mjbrodzik/ESPFusion.

Kuter, S., 2021. Completing the machine learning saga in fractional snow cover estimation from MODIS Terra reflectance data: random forests versus support vector regression. Remote Sens. Environ. 255, 30. https://doi.org/10.1016/j.rse.2021.112294.

Kuter, S., Akyurek, Z., Weber, G.W., 2018. Retrieval of fractional snow covered area from MODIS data by multivariate adaptive regression splines. Remote Sens. Environ. 205, 236–252. https://doi.org/10.1016/j.rse.2017.11.021.

Li, H., He, Y., Hao, X., Che, T., Wang, J., Huang, X., 2015. Downscaling snow cover fraction data in mountainous regions based on simulated inhomogeneous snow ablation. Remote Sens. 7, 8995. http://www.mdpi.com/2072-4292/7/7/8995.

Liang, H., Huang, X.D., Sun, Y.H., Wang, Y.L., Liang, T.G., 2017. Fractional snow-cover mapping based on MODIS and UAV data over the Tibetan plateau. Remote Sens. 9, 19. https://doi.org/10.3390/rs9121332.

Liston, G.E., Sturm, M., 1998. A snow-transport model for complex terrain. J. Glaciol. 44, 498–516.

Liston, G.E., Haehnel, R.B., Sturm, M., Hiemstra, C.A., Berezovskaya, S., Tabler, R.D., 2007. Instruments and methods: simulating complex snow distributions in windy environments using SnowTran-3D. J. Glaciol. 53, 241–256. https://doi.org/10.3189/172756507782202865.

Liu, C.Y., Huang, X.D., Li, X.B., Liang, T.G., 2020. MODIS fractional snow cover mapping using machine learning Technology in a Mountainous Area. Remote Sens. 12, 16. https://doi.org/10.3390/rs12060962.

Mahoney, P., Liston, G., LaPoint, S., Gurarie, E., Mangipane, B., Wells, A., Brinkman, T., Eitel, J., Hebblewhite, M., Nolin, A., Boelman, N., Prugh, L., 2018. Navigating

snowscapes: scale-dependent responses of mountain sheep to snowpack properties. Ecol. Applicat. 28 https://doi.org/10.1002/eap.1773.
Mankin, J.S., Viviroli, D., Singh, D., Hoekstra, A.Y., Diffenbaugh, N.S., 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., Liu, Y., Baldo, E., 2019. A joint landsat- and MODIS-based reanalysis approach for midlatitude montane seasonal snow characterization. Front. Earth Sci. 7. https://doi.org/10.3389/feart.2019.00272.
Molotch, N.P., Margulis, S.A., 2008. Estimating the distribution of snow water equivalent using remotely sensed snow cover data and a spatially distributed snowmelt model: a multi-resolution, multi-sensor comparison. Adv. Water Resour. 31, 1503–1514. https://doi.org/10.1016/j.advwatres.2008.07.017.
Moosavi, V., Malekinezhad, H., Shirmohammadi, B., 2014. Fractional snow cover mapping from MODIS data using wavelet-artificial intelligence hybrid models. J. Hydrol. 511, 160–170. https://doi.org/10.1016/j.jhydrol.2014.01.015.
Painter, T.H., Dozier, J., Roberts, D.A., Davis, R.E., Green, R.O., 2003. Retrieval of subpixel snow-covered area and grain size from imaging spectrometer data. Remote Sens. Environ. 85, 64–77. https://doi.org/10.1016/s0034-4257(02)00187-6.
Painter, T.H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R.E., Dozier, J., 2009. Retrieval of subpixel snow-covered area, grain size, and albedo from MODIS. Remote Sens. Environ. 113, 868–879. https://doi.org/10.1016/j.rse.2009.01.001.
Painter, T.H., Bryant, A.C., Skiles, S.M., 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.
Painter, T.H., Berisford, D.F., Boardman, J.W., Bormann, K.J., Deems, J.S., Gehrke, F., Hedrick, A., Joyce, M., Laidlaw, R., Marks, D., Mattmann, C., McGurk, B., Ramirez, P., Richardson, M., Skiles, S.M., Seidel, F.C., Winstral, A., 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.
Parker, K.L., Robbins, C.T., Hanley, T.A., 1984. Energy expenditures for locomotion by mule deer and elk. J. Wildl. Manag. 48, 474–488. https://doi.org/10.2307/3801180.
Raleigh, M.S., Rittger, K., Moore, C.E., Henn, B., Lutz, J.A., Lundquist, J.D., 2013. Ground-based testing of MODIS fractional snow cover in subalpine meadows and forests of the Sierra Nevada. Remote Sens. Environ. 128, 44–57. https://doi.org/10.1016/j.rse.2012.09.016.
Rittger, K., Painter, T.H., Dozier, J., 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.
Rittger, K., 2021. Random Forest fused MODIS and Landsat snow cover from spectral mixture analysis in the Sierra Nevada, USA. Zenodo. https://doi.org/10.5281/zenodo.5144494.
Rittger, K., Bair, E.H., Kahl, A., Dozier, J., 2016. Spatial estimates of snow water equivalent from reconstruction. Adv. Water Resour. 94, 345–363. https://doi.org/10.1016/j.advwatres.2016.05.015.
Rittger, K., Raleigh, M.S., Dozier, J., Hill, A.F., Lutz, J.A., Painter, T.H., 2020. Canopy adjustment and improved cloud detection for remotely sensed snow cover mapping. Water Resour. Res. 56. https://doi.org/10.1029/2019WR024914.
Rittger, K., Bormann, K.J., Bair, E.H., Dozier, J., Painter, T.H., 2021. Evaluation of VIIRS and MODIS snow cover fraction in High-Mountain Asia using Landsat 8 OLI. Front. Remote Sens. 2 https://doi.org/10.3389/frsen.2021.647154.
Rollins, M.G., 2009. LANDFIRE: a nationally consistent vegetation, wildland fire, and fuel assessment. Int. J. Wildland Fire 18, 235–249. https://doi.org/10.1071/WF08088.
Rosenthal, W., Dozier, J., 1996. Automated mapping of montane snow cover at subpixel resolution from the Landsat thematic mapper. Water Resour. Res. 32, 115–130. https://doi.org/10.1029/95WR02718.
Salomonson, V.V., Appel, I., 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.
Schwab, F.E., Pitt, M.D., 1987. Comparison of a direct and an indirect method for estimating available winter browse. Wildl. Soc. Bull. 15, 544–548. http://www.jstor.org/stable/3782579.
Selkowitz, D., Forster, R., Caldwell, M., 2014. Prevalence of pure versus mixed snow cover pixels across spatial resolutions in alpine environments. Remote Sens. 6, 12478. http://www.mdpi.com/2072-4292/6/12/12478.
Slater, A.G., Barrett, A.P., Clark, M.P., Lundquist, J.D., Raleigh, M.S., 2013. Uncertainty in seasonal snow reconstruction: relative impacts of model forcing and image availability. Adv. Water Resour. 55, 165–177. https://doi.org/10.1016/j.advwatres.2012.07.006.
Spitz, D., Hebblewhite, M., Stephenson, T., German, D., 2018. How plastic is migratory behavior? Quantifying elevational movement in a partially-migratory alpine ungulate. Can. J. Zool. 96 https://doi.org/10.1139/cjz-2017-0367.
Spitz, D.B., Hebblewhite, M., Stephenson, T.R., 2020. Habitat predicts local prevalence of migratory behaviour in an alpine ungulate. J. Anim. Ecol. 89, 1032–1044. https://doi.org/10.1111/1365-2656.13167.
Stephenson, T.R., German, D.W., Cassirer, E.F., Walsh, D.P., Blum, M.E., Cox, M., Stewart, K.M., Monteith, K.L., 2020. Linking population performance to nutritional condition in an alpine ungulate. J. Mammal. 101, 1244–1256. https://doi.org/10.1093/jmammal/gyaa091.
Stillinger, T., Roberts, D.A., Collar, N.M., Dozier, J., 2019. Cloud masking for Landsat 8 and MODIS Terra over snow-covered terrain: error analysis and spectral similarity between snow and cloud. Water Resour. Res. 55, 6169–6184. https://doi.org/10.1029/2019wr024932.
Stillinger, T, Costello, C, Bales, RC, Dozier, J, 2021. Reservoir Operators React to Uncertainty in Snowmelt Runoff Forecasts. J. water Resour. Plann. Manage. 147 (10) https://doi.org/10.1061/(ASCE)WR.1943-5452.0001437.
Vermote, E., Roger, J.C., Ray, J.P., 2015. MODIS surface reflectance User's guide collection 6 version 1.4.
Walters, R.D., Watson, K.A., Marshall, H.-P., McNamara, J.P., Flores, A.N., 2014. A physiographic approach to downscaling fractional snow cover data in mountainous regions. Remote Sens. Environ. 152, 413–425. https://doi.org/10.1016/j.rse.2014.07.001.
Wright, M.N., Ziegler, A., 2017. *Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R*, Journal of Statistical Software. (Foundation for Open Access Statistics).