Mapping snow cover at fine resolution in complex and forested terrain

Emma Troy Boudreau

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science in Civil Engineering

University of Washington

2025

Committee:

Jessica Lundquist

Nicoleta Cristea

Program Authorized to Offer Degree:

Civil and Environmental Engineering

University of Washington

Abstract

Mapping snow cover at fine resolution in complex and forested terrain

Emma Troy Boudreau

Chair of the Supervisory Committee:

Jessica Lundquist

Nicoleta Cristea

Department of Civil and Environmental Engineering

Complex topography in mountain basins allows deep snow accumulation in depressions throughout the winter season. During the melt period, snow on exposed terrain disappears faster, leaving behind a patchy mosaic of lingering snow cover. These persistent snow patches help sustain late-season streamflow during the drier summer months. In this work, we explored methods to observe heterogeneous snow cover in a complex subbasin in the Sierra Nevada, CA, using Planet Labs, Inc. 3 m resolution commercial optical satellites and an existing random forest model (Yang et al., 2023). Here, we tested two different model training configurations and a spatio-temporal post-processing approach to improve snow mapping throughout the season for three separate years, paying particular attention to snow in forested areas. We used lidar only for evaluation so that the methods used can be applied anywhere. The new model training approach

with temporal post-processing and the new spatio-temporal post-processing approach, showcased better results than prior methods, and achieved F1 scores of 0.86 and 0.88 in the forest, respectively. We then compared basin-wide snow-covered area from the 3 m resolution snow maps with the Moderate Resolution Imaging Spectroradiometer (MODIS) Snow-Covered Area and Grain Size (STC-MODSCAG) product (Rittger et al., 2020), which is coarser resolution but more readily available. They compared well, but on average our snow maps reported snow disappeared over 3 weeks later than STC-MODSCAG, and STC-MODSCAG missed ~303,000 m^2 of snow cover that the PlanetScope snow maps identified. Our results demonstrate the capability of high-resolution imagery for detecting snow patches relevant for ecology and the potential for improved snow cover mapping in forested basins using the model training methods or the spatial post-processing methods introduced in this work.

1. Introduction

The western U.S. coastal mountain ranges (e.g., Sierra Nevada, Cascades) are wet in the winter and dry in the summer. Spring snowmelt feeds tributary streams, controlling groundwater table levels (Lowry et al., 2010), and the date of snow disappearance dictates the vitality of meadow ecosystems and phenology timing in the drier months (Breckheimer et al., 2020; Sethi et al., 2020; Theobald et al., 2017). Heterogeneous snow cover generally extends the snowmelt period compared to an equivalent uniform snowpack, playing a significant role in water retention for seasonal snow-dominated catchments (Williams et al., 2009), and modeling efforts have supported this when simulating late-season streamflow at basin scale (Brauchli et al., 2017; Luce et al., 1998; Lundquist et al., 2005; Lundquist and Dettinger, 2005; Sexstone et al., 2020; Seyfried and Wilcox, 1995). Consequently, understanding patterns of fine-scale (<10 m) snow distribution is important for studies of subbasin scale hydrology and ecology.

Snow cover extent has been observed with remote sensing since the 1970s**.** Optical snow cover mapping typically exploits the differences in a pixel's spectral reflectance in the green band, where snow is bright, and the short-wave infrared (SWIR) band, where snow is dark. Initial efforts to produce binary snow cover maps with this method calculate the normalized difference snow index (NDSI) and follow a rule based approach to label pixels 'snow' or 'no snow' by thresholding this index (Dozier and Marks, 1987; Gascoin et al., 2019; Hall et al., 1995). Spectral mixture analysis was developed as a method to obtain subpixel information by inverting multispectral data, addressing the need for more detailed information on snow cover (Adams et al., 1995; Mertes et al., 1993; Nolin et al., 1993; Okin et al., 2001; Roberts et al., 1998). Snow cover can be derived from satellite imagery surface reflectance products like NASA's Moderate Resolution Imaging Spectroradiometer (MODIS), NOAA's Visible Infrared Spectroradiometer Suite (VIIRS), and NASA's recent Harmonized Landsat Sentinel-2 (HLS), but these products are not capable of capturing snow cover at fine spatial scales (<10 m). For example a snow cover and grain size product derived from MODIS, called the Snow-Covered Area and Grain size (MODSCAG), is a spectral unmixing product that provides near-daily data, but with a spatial resolution of ~500 m (Painter et al., 2009). Conversely, airborne light detection and ranging (lidar) instruments can detect snow distribution patterns caused by fine-scale processes like wind distribution and vegetation influences (Currier et al., 2019; Mazzotti et al., 2019), but they are costly and only flown in select basins, unable to observe the temporal evolution of snow throughout the melt season.

Raw near-daily snow cover observations from satellites are influenced by sensor noise, sensor viewing angle, cloud cover, shadows, image artifacts, and canopy interference. Despite these limitations, observations over time can facilitate better estimates of snow conditions through filtering, smoothing, and interpolation (Dozier et al., 2008). These post-processing methods address sporadic fluctuations in snow cover well, but canopy cover still remains a major challenge (Raleigh et al., 2013; Rittger et al., 2013). Products like the Spatially and Temporally Complete MODSCAG fSCA (STC-MODSCAG) product (Rittger et al., 2020) apply spatial post-processing methods that factor in the viewable gap fraction (VGF) and viewing angle to adjust snow cover estimates from pixel retrievals that are affected by canopy. Despite the significant improvements with spatially and temporally adjusted products, ~500 m is likely too coarse to represent ecological scale processes.

Over the past decade there has been an emergence of fine-resolution optical satellite imagery with near-daily observations made available by commercial companies. Recent efforts

have leveraged machine learning algorithms with this imagery to create binary snow cover maps (Cannistra et al., 2021; Hu and Shean, 2022; John et al., 2022; Kiewiet et al., 2022; Pflug et al., 2024; Wang et al., 2025; Yang et al., 2023). Research has shown Planet Labs Inc., PlanetScope (PS) 4-band multispectral imagery (red [R], green [G], blue [B], and near-infrared [NIR]), can capture snow patterns at 3 m resolution, and resolve the date snow disappears on average within 3 days (Pflug et al., 2024), both scales relevant for ecology.

In this work, we explore methods to observe heterogeneous snow cover at fine scales in a basin with complex topography through 1) an evaluation of two model training methods and spatial and temporal post-processing with PS imagery and 2) a comparison with the STC-MODSCAG fSCA product, with the specific objective to quantify snow covered area (SCA) in a small basin, while paying particular attention to forest areas. We want to better understand the advantages of using PS imagery in a complex basin for ecology purposes. We hypothesize that snow detection accuracy in the forest can be improved during the model training stages by segregating forested and non-forested pixels. Given the fine resolution of PS imagery, we hypothesize that snow cover information in small forest gaps can be used to interpolate snow under canopy cover as an alternative means to address optical imagery limitations in forests. Lastly, we hypothesize that PS maps can identify late-season, heterogeneous snow that drives summer streamflow, while a coarse-resolution product cannot. Using four-band information (R, G, B, and NIR) from PS high-resolution optical satellite images, 3 m airborne lidar snow depth maps, the STC-MODSCAG product, and a random forest model, we test the following questions:

1. Can we segregate non-forested and forested pixels when training machine learning models to improve snow cover estimates in forested domains?
2. Can we apply automated spatial and temporal post-processing to PS snow cover maps to improve snow cover estimates in forested domains?
3. How well can we measure late season snow cover with PS snow maps and STC-MODSCAG?

2. Background on high-resolution snow mapping

Recent efforts to map fine-scale snow cover investigated different machine learning models, model training techniques, and model performance for different land cover conditions and domains (Cannistra et al., 2021; Hu and Shean, 2022; John et al., 2022; Pflug et al., 2024; Yang et al., 2023). Cannistra et al. (2021) successfully created 3 m resolution snow maps using a convolutional neural network (CNN) and the visible and NIR bands provided by PS. This model was trained using lidar-derived snow cover labels for 370 million pixels. They showed a slight decrease in model performance when testing model transferability across climatically different regions. Aside from minimal transferability limitations, they also determined model issues in identifying snow within canopy cover, achieving mean F1 scores of 0.76 in non-forest regions and 0.57 for forested regions in the Tuolumne basin, where the model was trained in. Building on Cannistra's work, John et al. (2022) used the same CNN model with PS imagery, but augmented the model features to include topographic and vegetation information from a 3 m Digital Elevation Model (DEM) and Canopy Height Model (CHM). They found that vegetation information like Normalized Difference Vegetation Index (NDVI) improved the model transferability more than DEM information like northness, slope, and aspect. Although these models performed well, neural networks need a significant number of training pixels, requiring efficient labeling methods. Thus, both studies used lidar-derived binary snow maps for training

labels. This approach introduces lidar and auxiliary datasets (DEM and CHM) and limits the number of PS scenes the model is trained on, consequently restricting the range of image conditions the model is exposed to during training.

To address training grievances that accompany deep learning models, Hu and Shean (2022) used a less complex Random Forest (RF) model. With this model trained on ~400,000 training pixels and Maxar WorldView-2 and WorldView-3 imagery, they tested the optimal input configuration to create land cover maps, detecting snow, shaded snow, vegetation, water, and exposed ground. When using a RF model, the information from neighboring pixels is not used in the pixel-wise classification. This feature is particularly important for identifying pixels in the open that represent late-season snow patches with high spatial variability (Lu and Weng, 2007). Hu and Shean (2022) showed that the RF model had the highest accuracy for bright snow and vegetation classes. They also suggested that feature class expansion (more and narrower classes), with distinct spectral signatures, could improve model performance.

Yang et al. (2023) also used a simple RF model but with PS imagery to map snow cover in montane meadows throughout the season. This model was trained on one PS image with manually labeled pixels for training. Although less sophisticated and computationally expensive, the RF model outperformed a CNN-based method and showed good transferability across four different study sites**.** Yang et al. (2023) found model problems with canopy interference, mixed pixels, and shadows, and showed that the PS SCA is more accurate in larger forest gaps (15 m < to < 27 m), and areas farther away from the canopy edges (> 10 m). Yang et al. (2023) suggested that including several images in the training data to represent different seasons, terrain shadows, and different sensor instrument types may improve the model. Pflug et al. (2024) addressed this suggestion and manually labelled pixels from several images throughout the season to create snow cover maps to reconstruct snow water equivalent (SWE). Pflug et al. (2024) also employed temporal post-processing using a 3-observation moving window to adjust for incorrect snow predictions related to viewing angles and image artifacts.

3. Study domain and data

3.1 Study domain

The Tuolumne watershed benefits from rich field and airborne data collections and has been studied extensively (Henn et al., 2018b, 2018a; Lundquist et al., 2005, 2003; Lundquist and Dettinger, 2005; Pflug and Lundquist, 2020). Winter precipitation falls in the form of snow and melts in the warmer months, feeding the Tuolumne River. Within the Upper Tuolumne watershed, there are 6 tributaries that feed into the Tuolumne Meadows, one of the largest sub-alpine meadows in the Sierra Nevada. We focus specifically on the Budd Creek basin, a relatively small north-facing basin ranging from 2,607 m to 3,389 m in elevation and covering 7.57 km^2. Lodgepole Pine is the dominant tree species, and 33% of the basin is covered by canopy cover and the rest of the basin is exposed granite (Cooper et al., 2006). All of the contributing tributaries dry prior to the end of the summer, but Budd Creek sustains the longest streamflow, often flowing a month or two later than the other tributaries (Lundquist and Flint, 2006).

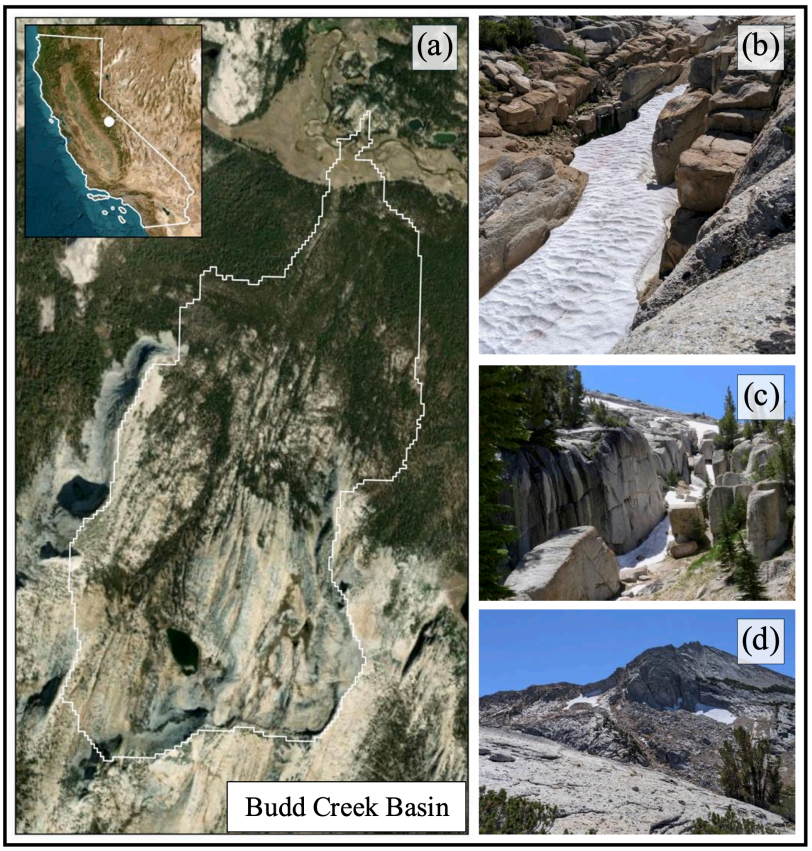

Figure 1. (a) Study domain, Budd Creek Basin (white outline). Examples of complex topography in Budd Creek Basin characterized by (b) fissures, (c) deep depressions, and (d) steep backwalls.

This late-season, sustained streamflow is the result of accumulating snow in complex topography and the reason behind choosing this study site. With deep fissures and fine-scale depressions, snow remains in these complex features longer than the surrounding exposed snowpack. These lasting snow piles impact late-season streamflow in Budd Creek by prolonging low flow conditions. Most satellites cannot capture snow in the fissures and depressions, which impact modeled snow depth, melt, and routing.

Our investigation covers 1 April to 30 September in 2020, 2022 and 2023. Both 2020 and 2022 were relatively average snow years, while 2023 was a record high snow year for the Sierra Nevada. PS imagery from 2020 were from the two earlier generations of sensors, so we included this year to understand if there are any limitations to sensor type with these methods. The latter two generations have more similar alignment in wavelength ranges for each band, but the bands in the earliest generation have less distinct separation (Frazier and Hemingway, 2021). More information on wavelength bands and band separation is described in Yang et al., 2023; Section 2.2.1.

3.2 Data products

ASO ID (Lidar)	Acquisition Date (Lidar)	Budd Basin fSCA (%) (Lidar-derived)	PlanetScope Scene ID	PlanetScope Sensor
ASO_Tuolumne_2020Apr13-14_snowdepth_3m	04-13-2020	99.0	20200413_184937_62_1058	PS2.SD
			20200413_1631141_1050	PS2
ASO_Tuolumne_Mosaic_2020May7_snowdepth_3m	05-07-2020	73.5	20200506_183733_58_105c	PS2.SD
ASO_Tuolumne_Mosaic_2020May21-23_snowdepth_3m	05-21-2020	46.0	20200523_162534_0f2a	PS2
ASO_Tuolumne_Mosaic_2022Apr5_snowdepth_3m	04-05-2022	90.3	20220403_174654_26_2451	PSB.SD
			20220407_174501_19_2445	PSB.SD
ASO_Tuolumne_Mosaic_2022Apr29_snowdepth_3m	04-29-2022	93.8	20220430_181902_39_2482	PSB.SD
ASO_Tuolumne_Mosaic_2022May18_snowdepth_3m	05-18-2022	60.3	20220518_182210_98_249b_3B	PSB.SD
ASO_Tuolumne_Mosaic_2023Apr27_snowdepth_3m	04-27-2023	96.6	20230426_174533_77_2451	PSB.SD
ASO_Tuolumne_Mosaic_2023Jun01-02_snowdepth_3m	06-01-2023	94.0	20230601_175109_84_24c8	PSB.SD
ASO_Tuolumne_Mosaic_2023Jun26-27_snowdepth_3m	06-26-2023	78.3	20230629_175450_42_24ce	PSB.SD

Table 1. ASO (lidar) ID, lidar acquisition date, fraction of snow cover in Budd Basin from lidar, PlanetScope ID, and PlanetScope sensor generation used for evaluation.

We downloaded PS imagery acquired between 1 April and 30 September for 2020, 2022, or 2023, that covered over 95% of the Budd Basin domain and had less than 5% cloud coverage. We used PS's 4-band orthorectified and radiometrically corrected Surface Reflectance product. This product has been corrected for consistency in space and time with atmospheric conditions. Most of the scenes processed were from the latest generation, PSB.SD, available from mid-March 2020 to current time. More information on PS imagery and image quality qualifications can be found in Supporting Information (Text S1).

For each year, repeat ASO lidar flights are conducted from as early as February to as late as the end of June. The dates when these flights occur vary from year to year, but within the Upper Tuolumne watershed, there are typically 2 to 3 flights after April 1st. These snow depth maps have been successfully used as validation for performance assessment of the PS-derived snow maps (Cannistra et al., 2021; Yang et al. 2023). Here, we also generated lidar-derived snow cover maps for our validation purposes consistent with the previous studies by thresholding the snow depth at 10 cm (ASO ID information in Table 1).

To ensure we only classified pixels with seasonal snow cover, we created and used a binary river network mask to remove Budd Lake using the United States Geological Survey (USGS) National Hydrography Dataset (NHD) (U.S. Geological Survey, 2024).

In Section 5.3, we compare our maps against the STC-MODSCAG fSCA product (Rittger et al., 2020). STC-MODSCAG is the latest available product from the evolution of snow cover observations from NASA's MODIS, which have increasingly improved snow estimates (Hall et al., 1995; Notarnicola et al., 2013; Painter et al., 2009; Salomonson and Appel, 2006, 2004). Canopy limitations are addressed by using the Viewable Gap Fraction (VGF), calculated with the MODSCAG fraction of vegetation (fVEG), and maintaining observation angle consistency by only using one sensor. This iteration of MODIS derived snow cover data was created by applying smoothing splines in the time dimension to the canopy adjusted fSCA that weighs nadir observations higher than off-nadir observations (Dozier et al., 2008). STC-MODSCAG is the most recent, off-the-shelf, post-processed snow cover product that uses smoothing and interpolation. We downloaded the h08v05 tile, which covers California's Sierra Nevada, for 2020, 2022, and 2023 from the NASA National Snow and Ice Data Center (NSIDC) DAAC Snow Today (see Data availability statement).

4. Methods

To create our snow cover maps, we employed three main steps. First, we trained the model to generate raw snow cover maps (Fig. 2, Step 1, see Section 4.1). Next, we performed systematic post-processing methods to the snow cover maps generated in Step 1 (Fig. 2, Step 2, see Section 4.2) to improve SCA under the canopy. Finally, we evaluated both the raw snow cover maps and the post-processed snow maps (Fig.2, Step 3, see Section 4.3).

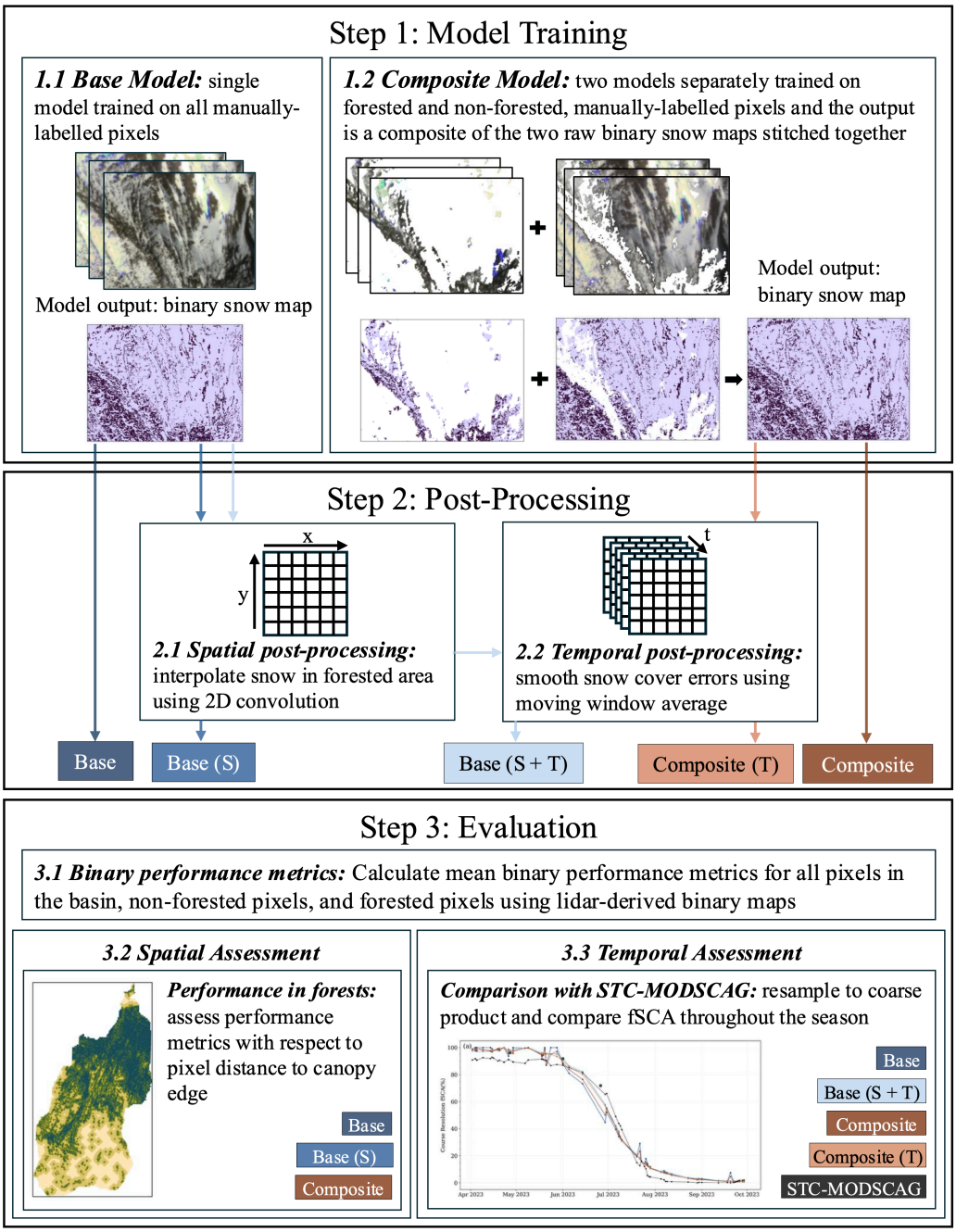

Figure 2. Conceptual figure detailing the three main steps in this work, model training, post-processing, and evaluation, corresponding to Sections 4.1.1 to 4.3.3. Imagery © 2022 Planet Labs, used under license.

4.1 Model training

We tested two different random forest model training approaches, referred to as the Base model (Yang et al., 2023; Pflug et al., 2024) (Fig. 2, Step 1.1) and the Composite model (Fig. 2, Step 1.2). We used the training methods from Yang et al. (2023) and Pflug et al. (2024) for the Base model and trained a single model for the entire basin. We then developed a new model training approach, referred to as the Composite model, that applied two separately trained models for forested and non-forested pixels. For both approaches, we used the R, G, B, and NIR bands, following the methods presented in Yang et al. (2023) and Pflug et al. (2024) to map snow cover with 4-band PS imagery. Like Yang et al. (2023), we used the sklearn.ensemble RandomForestClassifier and sklearn.model_selection to create the model with 10 estimators, a max depth of 10, 3 maximum features, and 1 random state for both the Base model and the Composite model. Each of these training techniques are described in the following sections.

4.1.1 Base training

For the Base training method, we manually selected a training dataset by individually labelling 13,000 pixels as snow, no snow, and unobserved using an interactive Matplotlib Python tool. Pixels were sampled from 14 PS images from 2022 that exhibited a range of image quality conditions (Table 1; Section 3.2). Snow and no snow pixels were labeled from the same 10 PS images, and unobserved pixels were selected from the remaining 4 images used in training that exhibited high saturation, glare, or other image artifacts. Different from Yang et al. (2023), this three-category labeling approach follows the methods of Pflug et al. (2024).

4.1.2 Composite training

Snow in the forest can often have a lower reflectance than snow in the open across the red, green, and blue bands (discussed in depth in Section 6.2). To evaluate if training these distinctly different categories will lead to better predictions in both classes, we trained two different models: one model on forested pixels and one model on non-forested pixels. While we do not introduce new classes to a singular model, a technique known as feature class expansion, we apply the concept that treating these pixels separately will result in better predictions due to different spectral properties between the two groups. Both the forested-trained model and the open-trained model were trained using the same 14 PS images from 2022 (Table 1; Section 3.2), with the same label categories (snow, no snow, and unobserved), and sample sizes (a total 13,000 pixels). This method will be referred to as the Composite method throughout.

Forested pixels were defined by a PS-derived canopy mask. To identify forested pixels, we generated an NDVI map in the basin for 15 snow free images and compiled them into one by taking the median NDVI for each pixel. We chose images from peak growing season, to best represent where canopy cover can exist, but the NDVI values may be impacted by when in the growing season these images are collected. We compared median NDVI to canopy height, from the lidar CHM (3 m resolution) to evaluate the range of NDVI values to include in the mask. Previous studies mapping snow cover have defined forested pixels as any pixel with canopy height over 2 m (Mazzotti et al., 2019; Yang et al., 2023). Based on NDVI to canopy height distribution, we found an NDVI of 0.35 roughly corresponded to 2 m (Fig. S1 in Supporting Information). Therefore, we threshold the median NDVI, defining any pixel with a

median NDVI greater than 0.35 as forested and anything under 0.35 canopy-free. Using an NDVI-generated canopy mask can reduce introducing an auxiliary dataset (the CHM) and be useful for domains without available lidar.

4.2 Post-processing

4.2.1 Spatial post-processing

Yang et al. (2023) showed that the RF model had a robust performance for pixels 10 m (> 3 pixels) away from the forest edge and for pixels in relatively large forest gaps, 15 m to 17 m (5 to 6 pixels) away from the forest edge. To test our hypothesis that snow cover information in forest gaps, made available by high-resolution imagery, can be used to interpolate snow cover in forested areas, we employed a hybrid post-processing approach by introducing contextual information to a pixel-wise classification (Lu and Weng, 2007) (Fig. 2, Step 2.1). In doing this, we applied a 2D convolution to pixels within the canopy mask, testing various neighborhood sizes and snow threshold requirements necessary to update a pixel label. The neighborhood size dictates the number of surrounding pixels that are considered, and the snow threshold determines how many of the pixels in the surrounding neighborhood need to be labeled snow to update the target pixel to snow. We used a uniformly distributed smoothing circular kernel. We tested a range of radius values for the kernel: 1 pixel (3 m), 3 pixels (9 m), 7 pixels (21 m), 10 pixels (30 m) and a range of threshold percents between 10% and 60%. Within the scipy.signal Python package, we use the convolve2d() function to implement this approach. We ultimately moved forward using a radius of 3 pixels and a 30% threshold. This method will be referred to as the Base (S) model. The convolution was applied to pixels within the canopy mask, preserving the original labels for open pixels. We tested this post-processing technique to the Base model, as an alternative method to the Composite model training approach for mapping snow cover in forested domains. We made this decision to determine if forest pixels are best addressed in model training stages (Composite) or post-processing stages (Base (S)).

4.2.2 Temporal post-processing

To address fluctuations in snow cover related to sensor noise, image artifacts, shadows, canopy effects, and cloud interference, we employ further temporal post-processing to the raw model outputs. Examples of these image conditions that impact accurate snow maps can be found in Supporting Information (Fig. S2). Snow maps can be treated like three dimensional cubes, in which the x and y of each individual pixel in a map can be represented as f(x, y), and when maps throughout the season are stacked together, a third temporal dimension is introduced, and each pixel can be represented as f(x, y, t).

We applied automatic temporal post-processing to the Base (S) model and Composite model under the assumption that the domain has seasonal snow coverage and minimal spring snow accumulation (Fig. 2, Step 2.2). We used a temporal median filter with a 5 observation moving window. We then assigned all unobserved pixels in the first time step to be snow and any unobserved pixels in following timesteps were set to the value of the previous timestep, an approach used by Pflug et al. (2024). Snow cover fluctuations in the forest were likely due to varying viewing angles from scene-to-scene while fluctuations of snow cover in the open were often related to saturated images (typically earlier in the season), and both were susceptible to

cloud coverage throughout the season. We refer to the temporally post-processed maps as Base (S+T) (both spatially and temporally adjusted) and Composite (T).

4.3 Evaluation

We evaluated our maps by testing the overall performance using binary performance metrics (Fig. 2, Step 3.1, Section 4.3.1). We did this for each of the five versions of maps generated (Base, Base (S), Base (S+T), Composite, and Composite (T)). We then performed a spatial assessment using distance to canopy edge (DCE) metrics (Mazzotti et al., 2019) to test model performance for forest clusters, forest edge, and gaps (Fig. 2, Step 3.2, Section 4.3.2). To better understand only spatially related contributions from the methods presented, we tested the Base maps, the spatially post-processed Base (Base (S)) maps, and the Composite maps, excluding temporally adjusted maps for the DCE evaluation. Lastly, we assessed the raw snow maps (Base and Composite) and the temporally adjusted maps (Base (S+T) and Composite (T)) and compared them with STC-MODSCAG (Fig. 2, Step 3.3, Section 4.3.3).

4.3.1 Performance metrics

We calculated binary performance metrics for every PS-derived snow cover map that had a corresponding lidar flight within 3 days of the retrieval date. Each pixel in the map was classified as a true positive (TP), true negative (TN), false positive (FP), or a false negative (FN). Pixels classified as snow in both the PS-derived map and the validation ASO flight are TPs. A TN is when both the PS-derived pixel and the validation pixel are classified as no snow. Pixels classified as snow in the PS derived map but not the ASO flight are FPs (commission), and FNs occur when the RF classified pixel is no snow when the validation pixel is classified as snow (omission). Using the TPs, TNs, FPs, and FNs we calculated the precision, recall, and F1 score.

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

Precision is used to determine how well the model accurately predicts snow. This is the number of true snow labels (both the PS and lidar derived snow maps have snow in a pixel) divided by all snow pixels (both correctly and incorrectly labeled snow pixel in the PS map). Lower precision values are typically a result of higher FP's or commission errors, meaning misclassification labeling snow when the pixel is not snow covered.

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

Recall is the number of TPs (correct snow labels) divided by the sum of the TPs and FNs (incorrectly labeled no snow). While precision quantifies how many of the snow predictions we made were true, recall quantifies how many of the actual snow pixels we labelled correctly. Lower recall values are typically a result of higher FNs or omission, meaning the model is underpredicting snow cover.

$$F1\ score = 2\ x\ \frac{Precision\ x\ Recall}{Precision + Recall}$$

The F1 score synthesizes both precision and recall and is used to evaluate the models overall balance of commission errors and omission errors. A higher F1 score implies that the model is maximizing precision and recall. In the case that there is no snow on the ground and the model reports no snow on the ground (TP=0), this metric is no longer valuable in assessing the model. Since all validation maps we used were above 46% snow covered, and all the evaluation

snow maps reported some snow cover, we did not encounter this situation. We provided the mean binary performance metrics of the Base, Base (S), Base (S+T), Composite, and Composite (T) maps including data from all three years for three general pixel categories; open pixels, canopy pixels, and all pixels (Fig. 2, Step 3.1). We used the NDVI-derived canopy mask (Section 4.1.2) to determine pixel categories.

4.3.2 Performance in forests

We used the PS NDVI-derived canopy mask (Section 4.1.2) to separate pixels in the basin into 8 categories based on DCE (Mazzotti et al., 2019) metrics. We used the same six categories as the evaluation methods in Yang et al. (2023), but with two additional categories; 'Open' and 'Dense Canopy'. The six categories from Yang et al. (2023) were forest clusters (-9 to -3 m), forest edges (-3 to 3 m), small gaps (3 to 9 m), medium gaps (9 to 15 m), large gaps (15 to 27 m), and outside forest (27 to 100 m). We considered any pixels in the 'Open' category (DCE greater than 100 m) to be unaffected by the canopy. While Yang et al. (2023) excludes any pixels with DCE < -9 m, on the basis that optical sensors are not suitable for mapping snow blocked by dense canopy, we chose to include these pixels to test the ability to interpolate snow with dense canopy with the spatial post-processing method.

Here, we present F1 scores for each DCE category (Fig. 2, Sep 3.2). We did not apply the temporal post-processing in the forest evaluation using DCE metrics. This decision was made in efforts to understand how the spatial post-processing of the Base model (Base (S)) and the segregation of forested and non-forested pixels in training (Composite) affected performance compared to the base model.

4.3.3 Comparison with STC-MODSCAG

We then assessed the basin-wide fraction of snow covered area from our post-processed maps with the STC-MODSAG fSCA to understand the benefits and drawbacks of using a high-resolution commercial product to a coarser resolution (~500 m) publicly available, long-term product (Fig. 2, Step 3.3). We downloaded the STC-MODSCAG from 1 April to 30 September for each year to calculate the basin-wide fraction of snow cover throughout the season (https://nsidc.org/snow-today; Rittger et al., 2020). Here, we resampled our PS-derived snow maps to match the resolution of STC-MODSCAG, dropping any pixels that partially lie within the basin at the coarse resolution. We then used the resampled PS maps to mask the STC-MODSCAG maps so that both products only included pixels that were entirely within the domain bounds. In doing this, we acknowledge that the coarser resolution basin-wide fSCA for both products may not cover the entire basin domain and areas along the edge of the basin may not be represented. Due to the basin's relatively small size and unique shape, only 63 coarse-resolution pixels represented the basin. We used the snow fraction variable in the STC-MODSCAG dataset for each date we had a corresponding PS scene.

5. Results

5.1 Overall performance

Group	Model	F1 Score	Precision	Recall
Canopy Pixels	Base	0.77 ± 0.02	0.92 ± 0.08	0.70 ± 0.00
	Base (S)	0.82 ± 0.03	0.90 ± 0.10	0.78 ± 0.02
	Base (S+T)	0.88 ± 0.03	0.94 ± 0.05	0.84 ± 0.02
	Composite	0.84 ± 0.04	0.88 ± 0.11	0.82 ± 0.02
	Composite (T)	0.86 ± 0.05	0.88 ± 0.11	0.86 ± 0.01
Open Pixels	Base	0.93 ± 0.03	0.92 ± 0.03	0.95 ± 0.04
	Base (S)	0.93 ± 0.03	0.92 ± 0.03	0.95 ± 0.04
	Base (S+T)	0.94 ± 0.03	0.90 ± 0.05	0.98 ± 0.01
	Composite	0.93 ± 0.03	0.92 ± 0.05	0.94 ± 0.05
	Composite (T)	0.94 ± 0.02	0.92 ± 0.04	0.96 ± 0.03
All Pixels	Base	0.86 ± 0.02	0.92 ± 0.04	0.82 ± 0.01
	Base (S)	0.88 ± 0.02	0.91 ± 0.05	0.86 ± 0.01
	Base (S+T)	0.91 ± 0.03	0.92 ± 0.05	0.90 ± 0.01
	Composite	0.89 ± 0.03	0.91 ± 0.06	0.87 ± 0.02
	Composite (T)	0.90 ± 0.03	0.90 ± 0.05	0.91 ± 0.00

Table 2. Mean F1 score, precision, and recall calculated for dates with available PS imagery and corresponding lidar observations in 2020, 2022, and 2023 for each snow mapping method. Each mean metric includes 11 observation dates with PS imagery across the three years. (S) refers to spatial and (T) refers to temporal post-processing.

We evaluated each model's overall performance in open pixel, canopy pixels, and both pixels combined using 11 PS derived snow maps not used in model training and across the 3 years in this study. In our overall performance evaluation against the available lidar-derived snow cover maps, we saw robust and similar performance among all models for the open pixels and an increased performance for canopy pixels with post-processing for the Base model and Composite model, resulting in an overall increase in performance for the entire basin (Table 2).

As expected, high omission errors in forested pixels were the sources of uncertainty among all methods. This is evident from low recall, and values in the forest averaged about ~0.16 lower than those in the open. Between the five methods analyzed, the spatially and temporally adjusted Base model (Base (S+T)) and the temporal adjusted Composite model (Composite (T)) had the highest overall F1 scores for canopy pixels and open pixels, achieving mean F1 scores of 0.91 and 0.90, respectively, for all pixels. The Base (S+T) model and the Composite (T) model improved mean F1 scores for forest pixels the most by 0.11 and 0.09, respectively.

The Base (S) model also performed better than the Base model in the forest, and improved mean F1 scores by 0.05. Additionally, the performance of the Base and Base (S) models is identical for open pixels because the spatial post-processing is only applied to pixels within the canopy mask. Lastly, we emphasize binary metrics do not entirely capture the utility of temporal post-processing, since they are derived from a handful of snapshots throughout the season, but they provide a general understanding of how the model performs (more on this in Section 6.4).

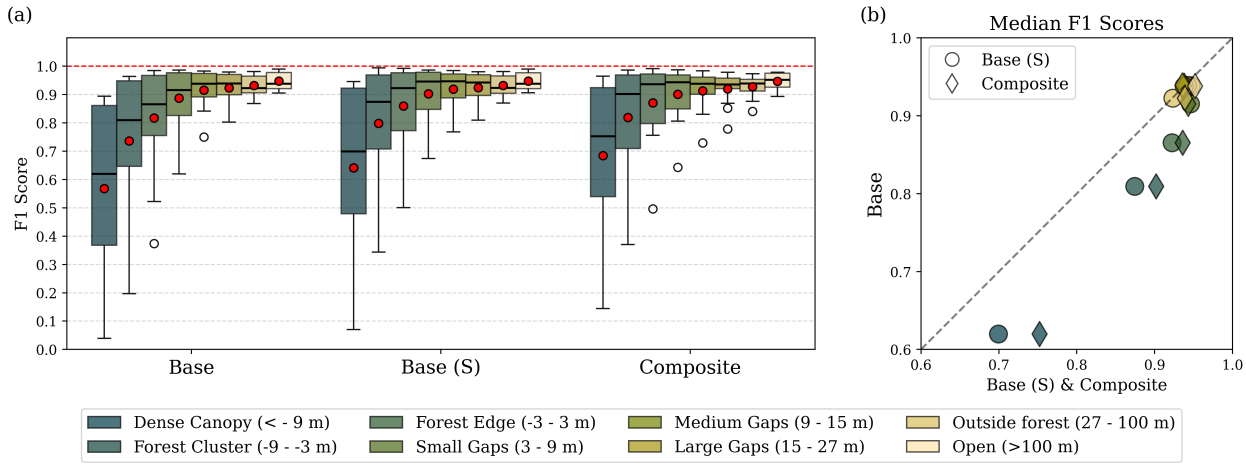

Figure 3. (a) F1 scores for pixel classes derived from distance to canopy edge (DCE) metrics for the Base model, the spatially post-processed Base model (Base (S)), and the Composite model. These metrics include maps generated from PS scenes with corresponding lidar flights for 2020, 2022, and 2023. The red dots represent the mean F1 score for each class. The boxplot shows the median (horizontal black line; median values listed in Supporting Information, Table S1) and range of F1 scores for each class. Concept from Yang et al. (2023). (b) scatter plot of Base median F1 scores to Base (S) (circle) and Composite (diamond) median F1 scores for each DCE class. Colors on (a) and (b) correspond to DCE category.

5.2 Evaluation of SCA over open and forested areas

To better understand model performance across space and with respect to canopy cover, we grouped pixels according to the DCE and calculated the F1 score (Fig. 3). Both the Base (S) model and the Composite model showed increased model performance for dense canopy, forest cluster, and forest edge. The most notable improvements with the Base (S) maps and the Composite maps are in the forest cluster category, achieving median F1 scores of 0.87 and 0.90, respectively, while the Base model performed relatively poor with a median F1 score of 0.81 (Fig. 3b and Table S1). Similarly, both the Base (S) model and the Composite model performed well in the forest edge category, achieving median F1 scores of 0.92 and 0.94, respectively. While both also performed better in the dense canopy, they had poor performance and achieved median F1 scores of 0.70 and 0.75 for Base (S) and Composite, respectively.

All models performed well for all pixels greater than 3 m away from the canopy. For these categories, the models were more consistent, showing a smaller spread of F1 scores (Fig. 3a). Median F1 scores for any pixels outside of the forest (the outside forest and open categories) were the same for the Base and the Base (S) model because the spatial component of the post-processing was only applied to forested pixels. While all models had robust performance for the two open classes, the Composite model did show slightly higher performance, with a median F1 score that was 0.02 higher, but this was within the noise levels.

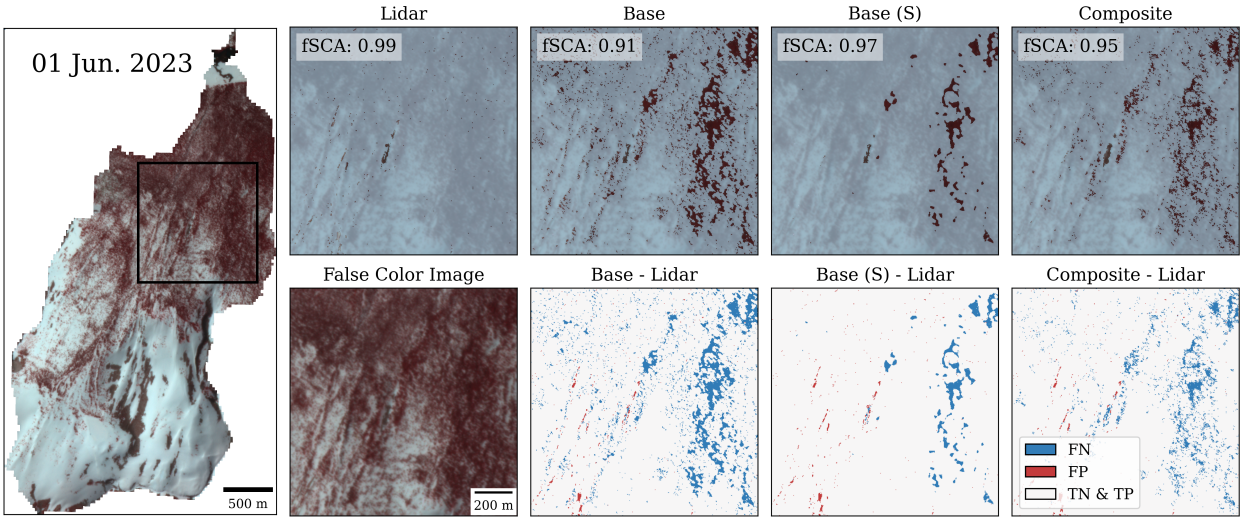

Figure 4. Examples of snow cover distribution for PlanetScope-derived snow maps on 01 June 2023 with insets zoomed in on a forested area in the basin. PlanetScope false color image of Budd Basin on 01 June 2023 (left column) (Imagery © 2023 Planet Labs) with inset bounding box (back outline). Inset with snow cover (light blue) overlaid on PlanetScope false color image (top row) with fSCA labelled (left to right): lidar-derived, Base model, Base (S) model, Composite model. Bottom row insets (left to right): PlanetScope false color image, Base and lidar-derived difference map, Base (S) and lidar-derived difference map, composite and lidar-derived difference map. For the difference maps, blue represents false negatives (FN), red represents false positives (FP), and cream represents true negative and true positives (both PlanetScope snow maps and lidar-derived snow maps).

The Base (S) model addressed incorrectly labeled no snow pixels that were spatially distributed amongst correctly labeled pixels as well as mislabeled pixels that were located on the edge of false negative clusters well (Fig. 4). Clusters of false negative pixels were less dense for the Composite model than both the Base and the Base (S) model. False negatives, omission errors, were the dominant source of uncertainty in snow maps from 01 June 2023, and this was similar to the other maps used in the evaluation process. We found commission errors for all models were minimal in the forest (Fig. 4) but occasionally occurred in the open where there were steep slopes and bare rocks that have high reflectance (Supporting Information, Fig. S3). This could be a result of band alignment-related image quality conditions due to abrupt changes in elevation, which is a known issue with PS imagery (Planet Team, 2022).

5.3 Comparison with coarser resolution STC-MODSCAG

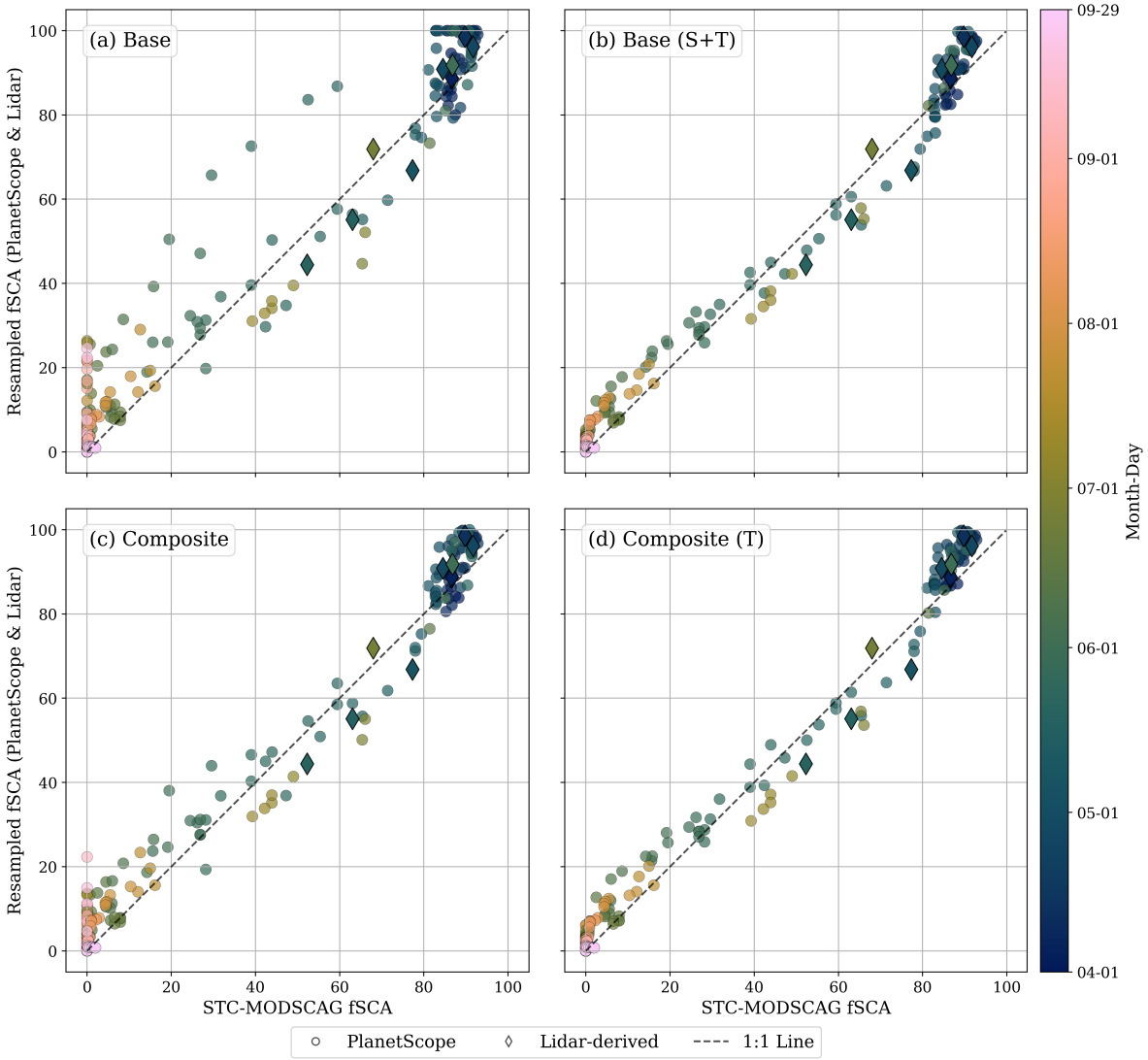

Figure 5. Comparison of resampled PlanetScope-derived Budd Basin fSCA (circle marker) and resampled lidar-derived Budd Basin fSCA (diamond marker) to the STC-MODSCAG fSCA. The four snow mapping methods using PlanetScope imagery compared are the Base model (top left), the Base (S+T) model (top right), the Composite model (bottom left), and the Composite (T) model (bottom right). Marker color corresponds to the day and month of the water year. All PlanetScope derived maps and lidar-derived snow maps from 2020, 2022, and 2023 with the corresponding STC-MODSCAG map are shown. The last acquisition date in the PS data is 29 September.

We saw general agreement between PS-derived fSCA values and corresponding STC-MODSCAG fSCA values from all images evaluated in 2020, 2022, and 2023 (Fig. 5). Among the aggregated 264 available dates with PS and STC-MODSCAG, we have 9 days with lidar observations and STC-MODSCAG, spanning a basin-wide fSCA of 98% to 44%.

The raw Base snow maps and the STC-MODSCAG fSCA had the largest deviation from the 1:1 line, with several PS fSCA reporting comparatively high values (Fig. 5, top left panel). This was then corrected in the Base (S+T) maps (Fig. 5, top right panel). Early in the season,

when the basin was greater than ~90% snow-covered, there were several dates when PS maps identified nearly 100% snow cover while STC-MODSCAG did not (Fig. 5, top left panel). This was likely the result of over saturated pixels and image conditions like blooming, which was also corrected for with the spatial and temporal post-processing. The temporal adjustments on the Composite maps were less extreme, but the raw composite maps experienced fewer extreme fSCA values (Fig. 5, bottom row). The Base model had a mean 4.2% +/- 8.1% fSCA difference and the Composite model had a 2.6% +/- 5.0% mean fSCA difference. The Base (S+T) model and the Composite (T) model agreed most with the STC-MODSCAG fSCA and had mean differences of 1.6% +/- 3.8% and 2.0% +/- 3.7%, respectively.

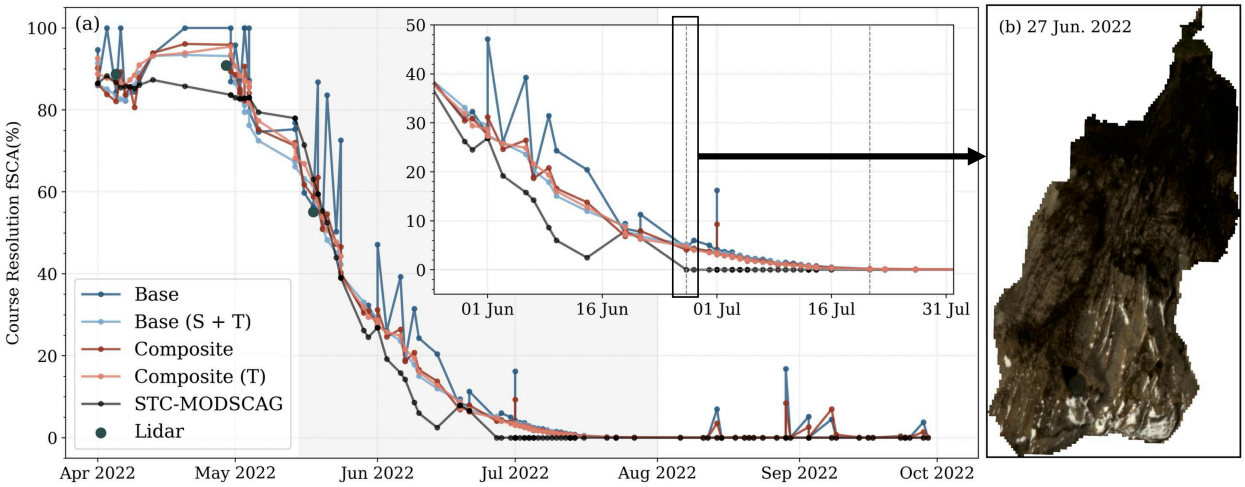

Figure 6. (a) 2022 basin-wide fSCA timeseries from the Base model (dark blue), Base (S+T) model (light blue), the Composite model (red), the Composite (T) model (pink), and the STC-MODSCAG (black). Markers indicate fSCA retrievals. Dark grey marker represents ASO (lidar-derived) fSCA retrievals. Inset shows snow disappearance timeframe within the basin. Vertical dashed lines on 06-27 and 07-21 indicate the date of snow disappearance for STC-MODSCAG maps and PS maps, respectively. (b) PlanetScope red, green, and blue image on 27 June 2022 (PS ID: 20220627_182341_54_2495) (Imagery © 2022 Planet Labs), the melt out date for STC-MODSCAG in Budd Basin.

Year	STC-MODSCAG DSD	PlanetScope DSD	PlanetScope $fSCA_{t=\text{STC-MODSCAG DSD}}$
2020	06-23	07-30	4%
2022	06-27	07-21	5%
2023	08-29	NA	3%

Table 3. The spatially and temporally complete Moderate Resolution Imaging Spectroradiometer Snow Covered-Area and Grain Size (STC-MODSCAG) Date of Snow Disappearance (DSD) and the Median DSD for the Base snow maps, the spatially and temporally post-processed Base (Base (S+T)) snow maps, the Composite snow maps, and the temporally post-processed Composite (Composite (T)) snow maps as well as the corresponding PS-derived Budd Basin fSCA when STC-MODSCAG reports no snow in the basin for 2020, 2022, and 2023.

The PS derived maps and the STC-MODSCAG maps showed general agreement between snow cover, except for the tail ends of the season, when PS identified more snow in both cases (Fig. 6). The STC-MODSCAG time series had a dramatic SCA decrease from about 30% to 0% while the slope of the fSCA line for all PS methods was more gradual in that range. When fSCA values were above 80% in the basin, STC-MODSCAG fSCA values were generally lower than both the ASO (lidar) fSCA and all PS methods (Fig. 6a) in 2022. A similar trend was observed for both 2020 (Supporting information Fig. S4) and 2023 (Supporting information Fig. S5), all except for two to three observations skewed from image artifacts. Fig. 6a shows several of these large spikes in the Base model (black), which are from image quality issues (early season spikes) and clouds (September 2022 spikes). The Base (S+T) model and the Composite (T) models did not show late season spikes from cloud coverage, showcasing the effectiveness of temporal smoothing.

The PS-derived melt out date in the basin was 24 days later than the STC-MODSCAG melt out date in 2022 (Fig. 6a). When the coarse product observed no snow in the basin, PS maps identified on average ~303,000 m^2 of snow cover in 2022. This remaining snow covered 4% of the basin's total area and was tucked away in deep fissures and depressions at high elevations. These lasting snow piles that PS can see are the reason Budd Creek has the longest sustained streamflow of the tributaries in the Upper Tuolumne River Basin. While we do not have lidar data to compare to the PS fSCA on the date STC-MODSCAG identifies no snow in the basin, the snow is visible to the eye from PS RGB imagery (Fig. 6b) and is regularly documented by hikers in the area. The remaining snow that is missed by the coarser sensors and visually identified in PS imagery evident in all three years in our analysis (Supporting information Fig. S4 and Fig. S5). Similarly to 2022, STC-MODSCAG melt out date was 37 days earlier in 2020 and over a month earlier in 2023, although PS still reported snow cover in the last scene on 29th September in 2023 (Table 3). STC-MODSCAG missed ~378,800 m^2 (5% of the basin) and ~227,000 m^2 (3% of the basin) of snow cover in the basin for 2020 and 2023, respectively. Our comparison with STC-MODSCAG demonstrated the value PS derived snow maps have for the application of basin hydrology and ecology when identifying the date snow disappears in a basin.

6. Discussion

6.1 Overall performance

While all methods had robust performance in our evaluation with lidar derived snow maps, it is evident that improvements in snow mapping can be made during the model training stages and post-processing stages of mapping snow cover. All methods performed well across the three distinct water years, showcasing versatility across snow conditions.

6.2 The Composite and the Base (S) model

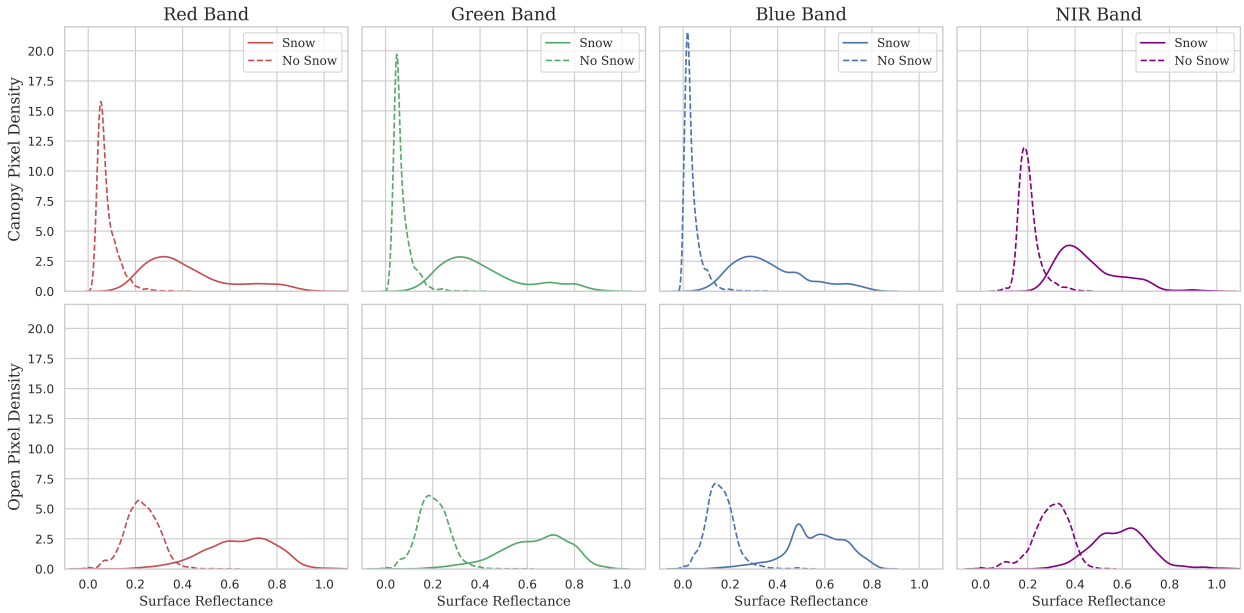

Figure 7. PlanetScope training data reflectance value distribution for Red (red), Green (green), Blue (blue), and Near infrared (purple) bands for canopy-covered pixels (top row) and open pixels (bottom row) separated by manual snow (solid) and no snow (dashed) labels. Both Open and Forested graphs show snow and no snow distributions that contain 5,000 data points each.

The Composite model was tested with the assumption that forested and non-forested pixels have distinctly different reflectance values among the four bands provided by PS imagery. Fig. 7 shows the distributions of reflectance values for forested and non-forested pixels in the red, green, blue, and near-infrared bands. Pixels without snow have lower reflectance values, and as seen across all four bands, no snow pixels particularly in the canopy had lower reflectance values and had a narrower spread of reflectance values. Canopy pixel density curves for snow and no snow generally intersect where the peak of the no snow, open pixel density curve is. Segregating forested and non-forested pixels in the training process mitigates 'confusion' the model might experience within this range of values. As for snow in the forest, which was our initial target when testing the Composite model, the distribution among all bands showed a right skew, with lower reflectance values. Lower reflectance values in the forest is a known phenomenon, and a result of mixed pixels, canopy coverage, or shadows, which is why detecting snow in the forest from optical sensors is difficult. A model can gain information from this general trend to inform better predictions.

Ultimately, treating forested and non-forested pixels separately increased model performance. Mean F1 scores in the forest improved 0.05 from the Base to the Composite model (Table 2). An unanticipated result of segregating the two models in the beginning was fewer extreme fluctuations in snow cover from image quality issues that the Base model was subject to (Fig. 6a, Supporting information Fig. S4 and Fig. S5).

The spatial post-processing also improved snow cover estimates. This method improved overall forest mean F1 scores from 0.77 to 0.82. The spatial post-processing corrected for snow cover best when the omission errors were distributed well among correct snow labels and pixels

at the boundary of clusters of omission errors. This method lacked effectiveness in dense clusters of false negatives (Fig. 4).

Our DCE analysis (Section 5.2) showed that both the Base (S) model and the Composite model are promising methods for mapping snow in the forest. For the forest edge pixels, the Base (S) model achieved a median F1 score of 0.92 and the Composite model achieved a median F1 score of 0.94, while the Base model achieved a median F1 score of 0.87. The Base (S) and Composite model also performed better in small forest gaps, and had a median F1 score of 0.95 and 0.94, respectively, while the Base had a F1 score of 0.92. The Base model and Base (S) model had median F1 scores lower than 0.90 for forest clusters (Supporting Information, Table S1. Yang et al. (2023) similarly reported poor performance in forest clusters, but it is important to note this is not a direct comparison and the results reported by Yang et al. (2023) were not from snow maps in Budd Basin. The composite model did achieve a median F1 score of 0.90 in forest clusters.

6.3 Temporal post-processing

Temporal smoothing addressed sensor noise, image artefacts, canopy shading, and cloud interference, which can cause sporadic changes in snow cover from one timestep to the next (Dozier et al., 2008; Pflug et al., 2024). Relevant to PS derived snow maps specifically, Pflug et al. (2024) found that automated temporal post-processing most improves snow cover estimates by correcting forest-edge grid cells that were blocked by canopy at off-nadir angles. We evaluated the need for temporal post-processing with the snow mapping methods introduced in this work and found increased F1 scores, captured by select observation dates, as well as more realistic fSCA time series throughout the melt season. The study reaffirms the necessity of augmenting temporal post-processing to raw snow maps from optical satellite imagery.

6.4 Limitations and future considerations

Within the scope of this study, there are some limitations to note. We test this work on one basin with a forest composition that does not reflect the forest composition of all other basins. Relevant to the Composite method, we do not comment on the transferability of this model to basins in other regions, and this may require additional training of a new model for each location. As for the spatial post-processing of the Base (Base (S)) model, the radius of 3 pixels and threshold of 30% snow cover for the neighborhood statistics worked for the canopy density in Budd Creek Basin but might not be appropriate for another forest. A basin with more sparse canopy cover may require a lower threshold. We suggest testing the appropriate combination of radius and thresholds for the basin of interest.

While our ‘truth’ datasets provide adequate spatial information, they lack sufficient temporal information. Within the three years of this study, the lidar acquisitions show 46% or more snow cover within the basin and only three days per season. Consequently, we cannot say for certain how these methods do in the forest when the basin is predominantly not covered in snow. We do feel confident in how our model performs in the open based on visual assessment of PS RGB images, but quantitatively validating this is difficult. Additionally, John et al. (2022) tested several lidar-derived SCA thresholds ranging from 3 cm and 20 cm against a PS-derived SCA map and found the F1 score decreased with a threshold under 8 cm and above 10 cm (John

et al., 2022), but this can be a source for errors. While the community has adopted lidar as a 'truth' dataset, we acknowledge no data is free of uncertainty.

Beyond PS imagery, we believe the Composite method might be applicable to other datasets. With that said, the utility of this approach most likely would be limited to an upper bound of sensor resolution, requiring pixel size to be representative of a canopy covered area. Similarly, the spatial post-processing is feasible due to the very detailed nature of PS that enables robust snow detection rates in forest gaps and is necessary for interpolating snow based on neighborhood statistics; therefore, we are skeptical this approach would work with a coarser product.

7. Conclusion

Detailed and frequent snow cover data is essential for hydrology and ecology in seasonally snow-covered mountain subbasins. While relatively accurate snow maps can be derived from PS imagery, data are limited by observation frequency, cloud coverage, sensor noise, image artefacts, and canopy interference. Snow maps from optical sensors are subject to these obscuring factors, and thus require filtering, smoothing, and interpolation. We investigated best practices for mapping snow in complex and forested domains. We developed a new model training technique, called the Composite model, and a spatial post-processing technique to improve snow detection rates for pixels obstructed by canopy cover and canopy shadows. Our methods improved mean F1 scores in the forest from 0.77, with our Base model, to 0.82 and 0.84, for the spatially post-processed Base model and the Composite model, respectively. We then showed the importance of temporal smoothing for snow cover mapping and further improved snow detection rates, achieving mean F1 scores of 0.88 and 0.86 for forested pixels with the spatially and temporally post-processed Base model and the temporally post-processed Composite model, respectively. We compared the PS-derived snow maps to the long term, moderate resolution, STC-MODSCAG fSCA product and found our maps can identify the last ~303,000 m^2 of snow cover that the coarse product could not. As a result, our maps reported the snow disappeared in the basin over 3 weeks later than the date STC-MODSCAG reported no snow. The work presented here shows promising efforts to better map snow for small basin ecology and hydrology.

Acknowledgements:

The author gratefully acknowledges this work was partially funded by the Future Rivers program at the University of Washington as part of a NSF National Research Traineeship award (DGE 1922004) and the Valle Scholarship & Scandinavian Exchange Program at the University of Washington. The author also thanks the Mountain Hydrology Research Group team at the University of Washington for support and feedback, the Remote Sensing group at the Norwegian Geotechnical Institute for hosting the author during this work, Eric Gagliano in the Terrain Analysis and Cryosphere Observation Lab at the University of Washington for programming knowledge and support in solving technical challenges, and the Planet Labs Support Team for support with satellite imagery questions. The author would like to give a special thanks to Justin Pflug, Kehan Yang, and Steven Pestana for their guidance and expertise with this project and Jessica Lundquist and Nicoleta Cristea for their mentorship throughout this program.

Supporting Information:

Text S1. Planet Data API categorizes each published image as 'test' or 'standard' based on image quality qualifications. Most images pass the established criteria to meet the 'standard' requirements, but some do not based on lack of ground lock or band alignment issues. More information on PS publishing can be found in Planet's technical specification (Planet Team, 2022). In this work we created snow maps from 279 PS scenes. Within the three focus years of this study, 55 scenes were considered 'test' quality, and 51 of the 'test' quality images occurred in April and May when the basin was predominantly snow covered, and image saturation levels were high. Only 18 images in April and May among the three years were in the 'standard' category. We included these images to test the functionality of temporal smoothing for better snow estimates and avoid long periods between observations.

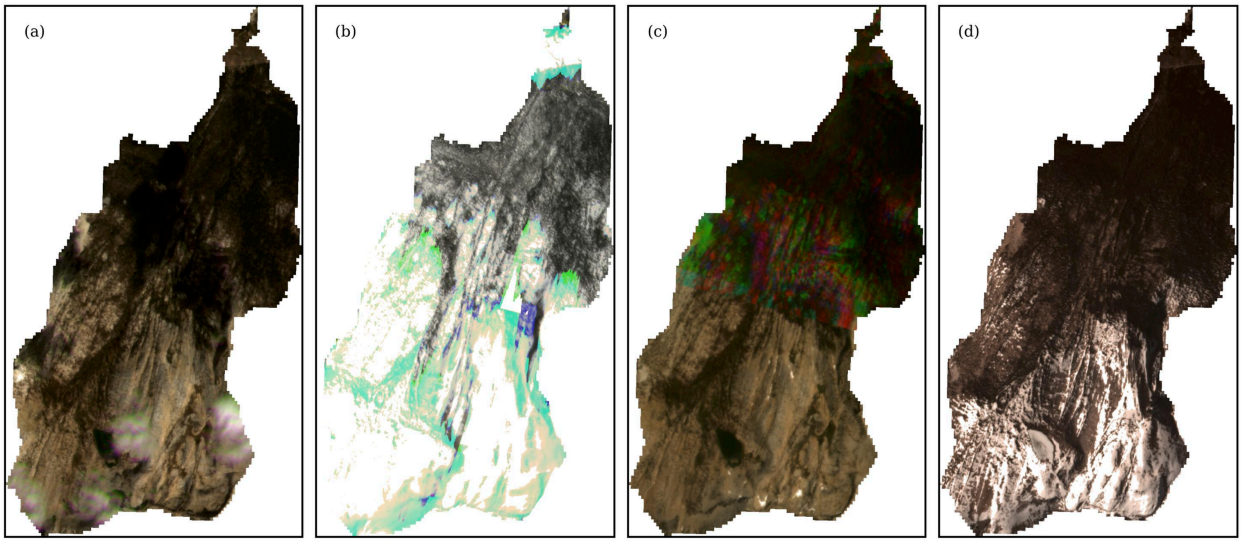

Figure S1. Examples of image characteristics in PlanetScope (PS) 4-band imagery that cause issues with mapping snow. Examples of (a) cloud coverage in PS image (PS ID: 20220928_183320_79_2426) (Imagery © 2022 Planet Labs), (b) saturation and blooming (PS ID: 20220420_182139_58_2490) (Imagery © 2022 Planet Labs), (c) degraded image quality (PS ID: 20220723_182239_82_2492) (Imagery © 2022 Planet Labs), and shadows (PS ID: 20200528_162440_0f2a) (Imagery © 2020 Planet Labs).

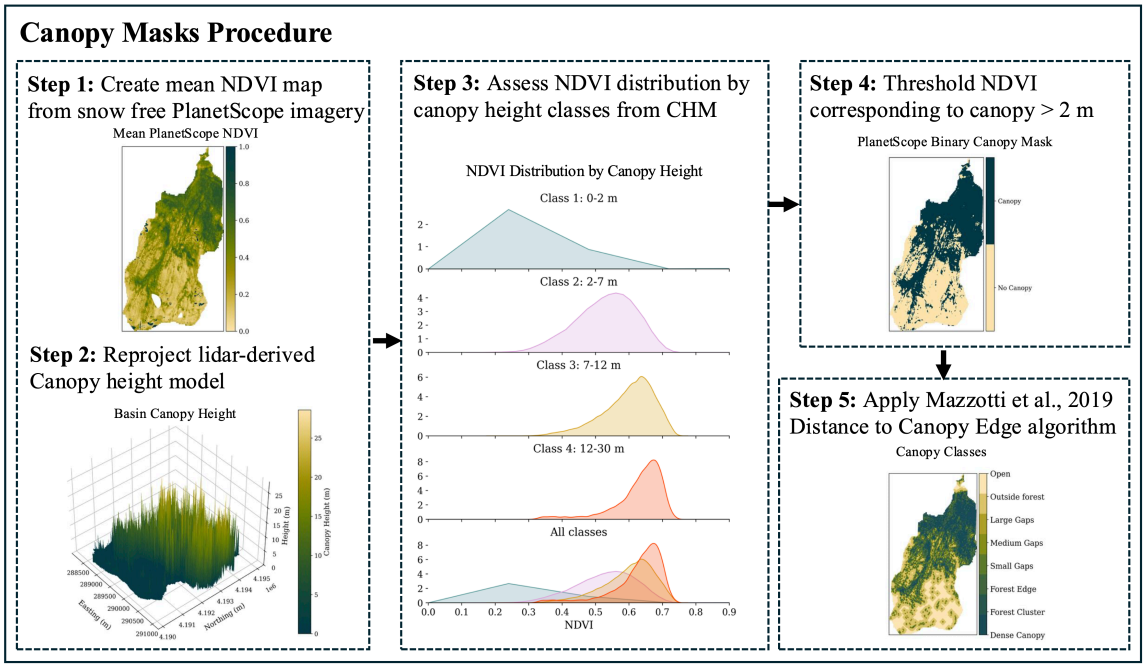

Figure S2. Conceptual diagram showing steps to create binary canopy mask used for training the Composite model and distance to canopy edge classes (Mazzotti et al., 2019) for spatial assessment.

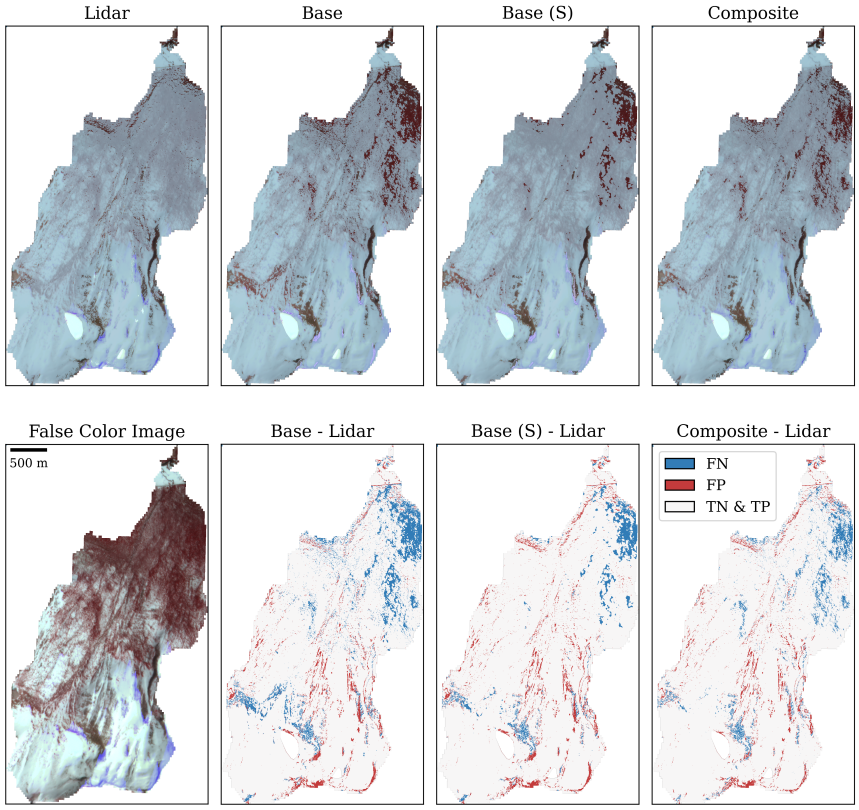

Figure S3. (Top row) Snow cover maps for lidar, Base, spatially post-processed Base, and Composite methods on 30 April 2022 (PS ID: 20220430_181902_39_2482) (Imagery © 2022 Planet Labs). False color image and difference maps for following methods in top row (bottom row).

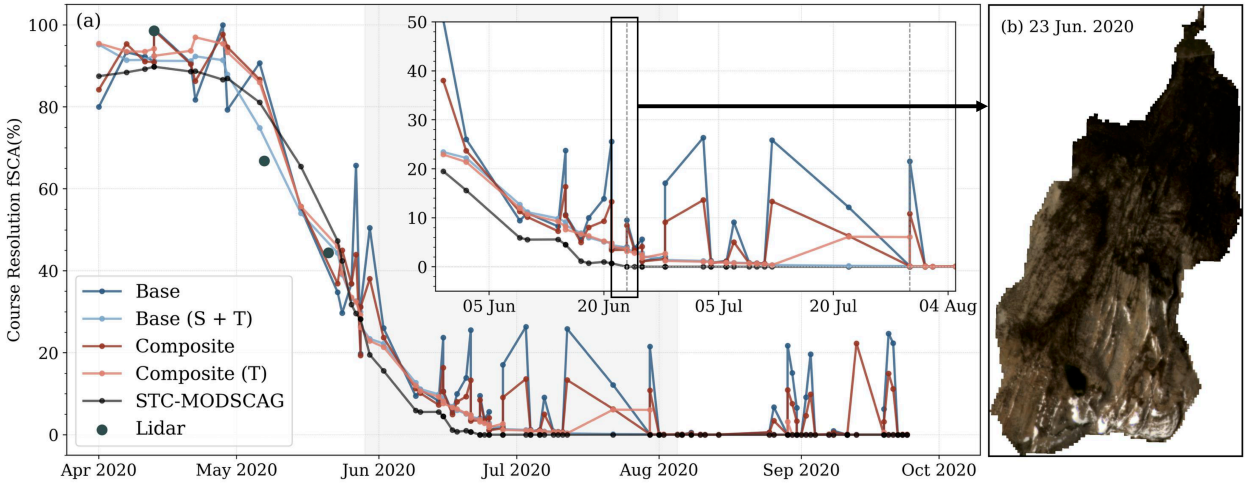

Figure S4. (a) 2020 fSCA timeseries from the Base model (dark blue), Base (S+T) model (light blue), the Composite model (purple), the Composite (T) model (pink), and the STC-MODSCAG (black). Markers indicate fSCA retrievals. Dark grey marker represents ASO (lidar-derived) fSCA retrieval. Inset showing end of season. Vertical dashed lines on 06-23 and 07-30 indicate the date of snow disappearance for STC-MODSCAG maps and PS maps, respectively. (b) PlanetScope red, green, and blue image on 23 June 2020 (PS ID: 20200623_175932_77_2277) (Imagery © 2020 Planet Labs), the melt out date for STC-MODSCAG in Budd Basin.

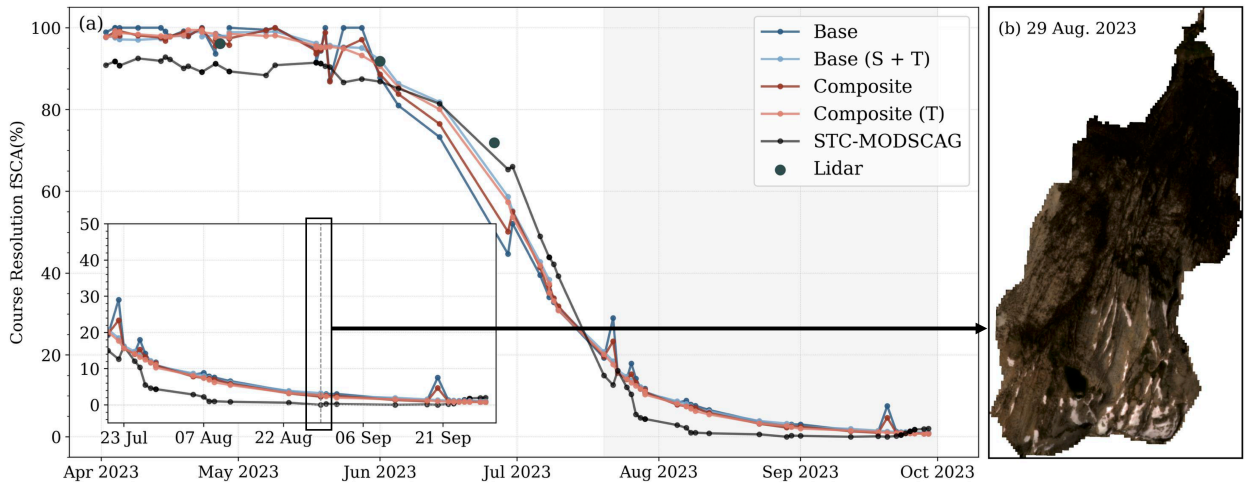

Figure S5. (a) 2023 fSCA timeseries from the Base model (dark blue), Base (S+T) model (light blue), the Composite model (purple), the Composite (T) model (pink), and the STC-MODSCAG (black). Markers indicate fSCA retrievals. Dark grey marker represents ASO (lidar-derived) fSCA retrieval. Inset showing end of season. The vertical dashed line on 08-29 indicates the date of snow disappearance for STC-MODSCAG maps. (b) PlanetScope red, green, and blue image on 08 Aug. 2023 (PS ID: 20230829_183119_13_2488) (Imagery © 2023 Planet Labs), the melt out date for STC-MODSCAG in Budd Basin.

Model	DCE Class	F1	Precision	Recall
Base	Dense Canopy	0.62	0.98	0.45
	Forest Cluster	0.81	0.96	0.79
	Forest Edge	0.87	0.96	0.87
	Small Gaps	0.92	0.95	0.90
	Medium Gaps	0.94	0.94	0.95
	Large Gaps	0.94	0.94	0.97
	Outside Forest	0.92	0.93	0.98
	Open	0.94	0.93	0.99
Base (S)	Dense Canopy	0.70	0.98	0.54
	Forest Cluster	0.87	0.96	0.87
	Forest Edge	0.92	0.96	0.91
	Small Gaps	0.95	0.95	0.95
	Medium Gaps	0.95	0.94	0.97
	Large Gaps	0.94	0.94	0.98
	Outside Forest	0.92	0.93	0.99
	Open	0.94	0.93	0.99
Composite	Dense Canopy	0.75	0.98	0.63
	Forest Cluster	0.90	0.96	0.89
	Forest Edge	0.94	0.95	0.92
	Small Gaps	0.94	0.94	0.95
	Medium Gaps	0.94	0.93	0.95
	Large Gaps	0.94	0.92	0.95
	Outside Forest	0.94	0.91	0.97
	Open	0.95	0.95	0.98

Table S1. Median F1 score, Precision, Recall, and Accuracy for the Base model, Base (S) model, and Composite model in the 8 DCE classes. Values correspond to Figure 3 in Section 5.1.

References:

Adams, J.B., Sabol, D.E., Kapos, V., Almeida Filho, R., Roberts, D.A., Smith, M.O., Gillespie, A.R., 1995. Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon. Remote Sens. Environ. 52, 137–154. https://doi.org/10.1016/0034-4257(94)00098-8

Brauchli, T., Trujillo, E., Huwald, H., Lehning, M., 2017. Influence of Slope-Scale Snowmelt on Catchment Response Simulated With the Alpine3D Model. Water Resour. Res. 53, 10723–10739. https://doi.org/10.1002/2017WR021278

Breckheimer, I.K., Theobald, E.J., Cristea, N.C., Wilson, A.K., Lundquist, J.D., Rochefort, R.M., HilleRisLambers, J., 2020. Crowd-sourced data reveal social–ecological mismatches in phenology driven by climate. Front. Ecol. Environ. 18, 76–82. https://doi.org/10.1002/fee.2142

Cannistra, A.F., Shean, D.E., Cristea, N.C., 2021. High-resolution CubeSat imagery and machine learning for detailed snow-covered area. Remote Sens. Environ. 258, 112399. https://doi.org/10.1016/j.rse.2021.112399

Cooper, D.J., Lundquist, J.D., Flint, L.E., Lott, F., Flint, A.L., Wolf, E., King, J., 2006. Effects of the Tioga Road on Hydrologic Processes and Lodgepole Pine Invasion into Tuolumne Meadows, Yosemite National Park. Yosemite National Park.

Currier, W.R., Pflug, J., Mazzotti, G., Jonas, T., Deems, J.S., Bormann, K.J., Painter, T.H., Hiemstra, C.A., Gelvin, A., Uhlmann, Z., Spaete, L., Glenn, N.F., Lundquist, J.D., 2019. Comparing Aerial Lidar Observations With Terrestrial Lidar and Snow-Probe Transects From NASA's 2017 SnowEx Campaign. Water Resour. Res. 55, 6285–6294. https://doi.org/10.1029/2018WR024533

Dozier, J., Marks, D., 1987. Snow Mapping and Classification from Landsat Thematic Mapper Data. Ann. Glaciol. 9, 97–103. https://doi.org/10.3189/S026030550000046X

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., Hydrologic Remote Sensing 31, 1515–1526. https://doi.org/10.1016/j.advwatres.2008.08.011

Frazier, A.E., Hemingway, B.L., 2021. A Technical Review of Planet Smallsat Data: Practical Considerations for Processing and Using PlanetScope Imagery. Remote Sens. 13, 3930. https://doi.org/10.3390/rs13193930

Gascoin, S., Grizonnet, M., Bouchet, M., Salgues, G., Hagolle, O., 2019. Theia Snow collection: high-resolution operational snow cover maps from Sentinel-2 and Landsat-8 data.

Hall, D.K., Riggs, G.A., Salomonson, V.V., 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

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

Henn, B., Clark, M.P., Kavetski, D., Newman, A.J., Hughes, M., McGurk, B., Lundquist, J.D., 2018a. Spatiotemporal patterns of precipitation inferred from streamflow observations across the Sierra Nevada mountain range. J. Hydrol. 556, 993–1012. https://doi.org/10.1016/j.jhydrol.2016.08.009

Henn, B., Painter, T.H., Bormann, K.J., McGurk, B., Flint, A.L., Flint, L.E., White, V., Lundquist, J.D., 2018b. High-Elevation Evapotranspiration Estimates During Drought: Using Streamflow and NASA Airborne Snow Observatory SWE Observations to Close the Upper Tuolumne River Basin Water Balance. Water Resour. Res. 54, 746–766. https://doi.org/10.1002/2017WR020473

Hu, J.M., Shean, D., 2022. Improving Mountain Snow and Land Cover Mapping Using Very-High-Resolution (VHR) Optical Satellite Images and Random Forest Machine Learning Models. Remote Sens. 14, 4227. https://doi.org/10.3390/rs14174227

John, A., Cannistra, A.F., Yang, K., Tan, A., Shean, D., Hille Ris Lambers, J., Cristea, N., 2022. High-Resolution Snow-Covered Area Mapping in Forested Mountain Ecosystems Using PlanetScope Imagery. Remote Sens. 14, 3409. https://doi.org/10.3390/rs14143409

Kiewiet, L., Trujillo, E., Hedrick, A., Havens, S., Hale, K., Seyfried, M., Kampf, S., Godsey, S.E., 2022. Effects of spatial and temporal variability in surface water inputs on streamflow generation and cessation in the rain–snow transition zone. Hydrol. Earth Syst. Sci. 26, 2779–2796. https://doi.org/10.5194/hess-26-2779-2022

Lowry, C.S., Deems, J.S., Loheide Ii, S.P., Lundquist, J.D., 2010. Linking snowmelt-derived fluxes and groundwater flow in a high elevation meadow system, Sierra Nevada Mountains, California. Hydrol. Process. 24, 2821–2833. https://doi.org/10.1002/hyp.7714

Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. Int. J. Remote Sens. 28, 823–870. https://doi.org/10.1080/01431160600746456

Luce, C.H., Tarboton, D.G., Cooley, K.R., 1998. The influence of the spatial distribution of snow on basin-averaged snowmelt. Hydrol. Process. 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., Cayan, D.R., Dettinger, M.D., 2003. Meteorology and Hydrology in Yosemite National Park: A Sensor Network Application, in: Zhao, F., Guibas, L. (Eds.), Information Processing in Sensor Networks. Springer, Berlin, Heidelberg, pp. 518–528. https://doi.org/10.1007/3-540-36978-3_35

Lundquist, J.D., Dettinger, M.D., 2005. How snowpack heterogeneity affects diurnal streamflow timing. Water Resour. Res. 41. https://doi.org/10.1029/2004WR003649

Lundquist, J.D., Dettinger, M.D., Cayan, D.R., 2005. Snow-fed streamflow timing at different basin scales: Case study of the Tuolumne River above Hetch Hetchy, Yosemite, California. Water Resour. Res. 41. https://doi.org/10.1029/2004WR003933

Lundquist, J.D., Flint, A.L., 2006. Onset of Snowmelt and Streamflow in 2004 in the Western United States: How Shading May Affect Spring Streamflow Timing in a Warmer World. https://doi.org/10.1175/JHM539.1

Mazzotti, G., Currier, W.R., Deems, J.S., Pflug, J.M., Lundquist, J.D., Jonas, T., 2019. Revisiting Snow Cover Variability and Canopy Structure Within Forest Stands: Insights From Airborne Lidar Data. Water Resour. Res. 55, 6198–6216. https://doi.org/10.1029/2019WR024898

Mertes, L.A.K., Smith, M.O., Adams, J.B., 1993. Estimating suspended sediment concentrations in surface waters of the Amazon River wetlands from Landsat images. Remote Sens. Environ. 43, 281–301. https://doi.org/10.1016/0034-4257(93)90071-5

Nolin, A.W., Dozier, J., Mertes, L.A.K., 1993. Mapping alpine snow using a spectral mixture modeling technique. Ann. Glaciol. 17, 121–124. https://doi.org/10.3189/S0260305500012702

Notarnicola, C., Duguay, M., Moelg, N., Schellenberger, T., Tetzlaff, A., Monsorno, R., Costa, A., Steurer, C., Zebisch, M., 2013. Snow Cover Maps from MODIS Images at 250 m Resolution, Part 1: Algorithm Description. Remote Sens. 5, 110–126. https://doi.org/10.3390/rs5010110

Okin, G.S., Murray, B., Schlesinger, W.H., 2001. Degradation of sandy arid shrubland environments: observations, process modelling, and management implications. J. Arid Environ. 47, 123–144. https://doi.org/10.1006/jare.2000.0711

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

Pflug, J.M., Lundquist, J.D., 2020. Inferring Distributed Snow Depth by Leveraging Snow Pattern Repeatability: Investigation Using 47 Lidar Observations in the Tuolumne Watershed, Sierra Nevada, California. Water Resour. Res. 56, e2020WR027243. https://doi.org/10.1029/2020WR027243

Pflug, J.M., Yang, K., Cristea, N., Boudreau, E.T., Vuyovich, C.M., Kumar, S.V., 2024. Using Commercial Satellite Imagery to Reconstruct 3 m and Daily Spring Snow Water Equivalent. Water Resour. Res. 60, e2024WR037983. https://doi.org/10.1029/2024WR037983

Planet Labs Inc., 2022. Imagery Product Specification: Reprocessing SuperDove/Dove-R archive with band alignment improvements; Planet Labs, Inc.: San Francisco, CA, USA.

Planet Team, 2022. Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA, USA. https://docs.planet.com/.

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., 35th Year Anniversary Issue 51, 367–380. https://doi.org/10.1016/j.advwatres.2012.03.002

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, e2019WR024914. https://doi.org/10.1029/2019WR024914

Roberts, D.A., Gardner, M., Church, R., Ustin, S., Scheer, G., Green, R.O., 1998. Mapping Chaparral in the Santa Monica Mountains Using Multiple Endmember Spectral Mixture Models. Remote Sens. Environ. 65, 267–279. https://doi.org/10.1016/S0034-4257(98)00037-6

Salomonson, V.V., Appel, I., 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

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

Sethi, M.L., Theobald, E.J., Breckheimer, I., Hille Ris Lambers, J., 2020. Early snowmelt and warmer, drier summers shrink postflowering transition times in subalpine wildflowers. Ecology 101, e03171. https://doi.org/10.1002/ecy.3171

Sexstone, G.A., Driscoll, J.M., Hay, L.E., Hammond, J.C., Barnhart, T.B., 2020. Runoff sensitivity to snow depletion curve representation within a continental scale hydrologic model. Hydrol. Process. 34, 2365–2380. https://doi.org/10.1002/hyp.13735

Seyfried, M.S., Wilcox, B.P., 1995. Scale and the Nature of Spatial Variability: Field Examples Having Implications for Hydrologic Modeling. Water Resour. Res. 31, 173–184. https://doi.org/10.1029/94WR02025

Theobald, E., Breckheimer, I., HilleRisLambers, J., 2017. Climate drives phenological reassembly of a mountain wildflower meadow community. Ecology 98. https://doi.org/10.1002/ecy.1996

U.S. Geological Survey, National Geospatial Program, 20240105, USGS National Hydrography Dataset Best Resolution (NHD) for Hydrological Unit (HU) 8 - 18040008 (published 20240105) FileGDB: U.S. Geological Survey.

Wang, Z., Jaya Baskar, J.V., Sistla Naga Sai, M.S., Svoma, B., Vivoni, E.R., 2025. Spatiotemporal Patterns of Intermittent Snow Cover From PlanetScope Imagery Using Deep Learning. Geophys. Res. Lett. 52, e2025GL116582. https://doi.org/10.1029/2025GL116582

Williams, C.J., McNamara, J.P., Chandler, D.G., 2009. Controls on the temporal and spatial variability of soil moisture in a mountainous landscape: the signature of snow and complex terrain. Hydrol. Earth Syst. Sci. 13, 1325–1336. https://doi.org/10.5194/hess-13-1325-2009

Yang, K., John, A., Shean, D., Lundquist, J.D., Sun, Z., Yao, F., Todoran, S., Cristea, N., 2023. High-resolution mapping of snow cover in montane meadows and forests using Planet imagery and machine learning. Front. Water 5.