Remote Sensing of Environment 258 (2021) 112399

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

High-resolution CubeSat imagery and machine learning for detailed snow-covered area

Anthony F. Cannistra [a,*], David E. Shean [b], Nicoleta C. Cristea [b]

[a] Department of Biology, University of Washington, Seattle, WA, United States
[b] Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, United States

ARTICLE INFO

Editor: Dr. Menghua Wang

Keywords:
Planet
PlanetScope
Machine learning
Seasonal snow
Snow covered area
Airborne lidar
Supervised classification

ABSTRACT

Snow cover affects a diverse array of physical, ecological, and societal systems. As such, the development of optical remote sensing techniques to measure snow-covered area (SCA) has enabled progress in a wide variety of research domains. However, in many cases, the spatial and temporal resolutions of currently available remotely sensed SCA products are insufficient to capture SCA evolution at spatial and temporal resolutions relevant to the study of fine-scale spatially heterogeneous phenomena. We developed a convolutional neural network-based method to identify snow covered area using the ~3 m, 4-band PlanetScope optical satellite image dataset with ~daily, near-global coverage. By comparing our model performance to snow extent derived from high-resolution airborne lidar differential depth measurements and satellite platforms in two North American sites (Sierra Nevada, CA, USA and Rocky Mountains, CO, USA), we show that these emerging image archives have great potential to accurately observe snow-covered area at high spatial and temporal resolutions despite limited radiometric bandwidth and band placement. We achieve average snow classification F-Scores of 0.73 in our training basin and 0.67 in a climatically-distinct out-of-sample basin, suggesting opportunities for model transferability. We also evaluate the performance of these data in forested regions, suggesting avenues for further research. The unparalleled spatial and temporal coverage of CubeSat imagery offers an excellent opportunity for satellite remote sensing of snow, with real implications for ecological and water resource applications.

1. Introduction

The development of optical remote sensing techniques to measure snow-covered area (SCA) over time has enabled progress in a wide variety of research domains. Satellite observations of SCA from sensors such as Landsat Thematic Mapper (TM)/Operational Land Imager (OLI), MODIS Terra, and Sentinel-2 Multispectral Instrument (MSI) have applications in hydrology and water resource management (Andreadis and Lettenmaier, 2006; Schattan et al., 2020), atmospheric science (Fernandes et al., 2009; Painter et al., 2009), ecology (Boelman et al., 2019; Carlson et al., 2015; Dedieu et al., 2016) and in climate adaptation and mitigation (Immerzeel et al., 2009).

In many cases, the spatial and temporal resolutions of currently available remotely sensed SCA products are insufficient to capture SCA evolution at spatial and temporal resolutions relevant to the study of fine-scale spatially heterogeneous phenomena. For example, high-resolution SCA observations, when employed in a data-assimilation context, may improve runoff projections and similar hydrologic simulations by capturing fine-scale variability in snow cover (Clark et al., 2011; Luce et al., 1999; Lundquist and Dettinger, 2005). In mountain ecosystems, the composition of plant communities and associated phenological events such as flowering and growth vary as a function of snow cover extent and snow duration (Choler, 2005; Ford et al., 2013; Theobald et al., 2017; Venn et al., 2011), which exhibit significant heterogeneity over small spatial and temporal scales (1–10 m; hours to days during ablation) (Clark et al., 2011; Little et al., 1994; Rochefort et al., 1994).

Airborne lidar surveys (ALS) surveys can provide snow observations at fine spatial resolution (e.g., 3 m, Airborne Snow Observatory; Painter et al. (2016)), but are limited to a relatively small spatial extent, leading to sparse temporal coverage. Comparatively, existing satellite platforms are able to observe snow at a much broader spatial extent but at coarser spatial resolution and variable temporal resolution. Instruments carried by orbiting platforms like Sentinel-2 and Landsat 8, for example, can be used to generate snow observation products (e.g., Normalized Difference Snow Index (NDSI), Hall and Riggs (2011) or fractional snow-

* Corresponding author.
E-mail address: tonycan@uw.edu (A.F. Cannistra).

https://doi.org/10.1016/j.rse.2021.112399
Received 13 July 2020; Received in revised form 9 March 2021; Accepted 12 March 2021
Available online 22 March 2021

covered area [fSCA], Painter et al. (2009)) at 10 m (Sentinel-2) or 30 m (Landsat 8) resolution. These platforms have mid-latitude revisit times of between 5 and 16 days (though harmonized products can offer shorter revisit times). Despite these opportunities, a planetary-scale observational SCA dataset with both high spatial (< 5–10 m) resolution and frequent temporal (~daily) resolution does not yet exist.

Emerging constellations of small imaging satellites (CubeSats) can potentially bridge the temporal and spatial resolution gap between ALS surveys and satellite observations, but methods to derive SCA from these platforms have yet to be studied in detail. Planet, a CubeSat developer/vendor in California, USA, operates the PlanetScope constellation of over 150 small (3U CubeSat form factor) optical imaging satellites. This constellation offers 4-band multispectral imagery (red [R], green [G], blue [B] and near-infrared [NIR]) at ~3–5 m resolution with ~daily revisit times across most of the earth's land surface (±81.5 degrees latitude), amounting to a cumulative observational footprint of over 200 M $km^2 day^{-1}$ (Planet Labs, Inc., 2019b). However, the spectral bandwidth and available calibration options for these instruments limits the ability of traditional radiometric or spectral analyses for snow. In particular, the visible and shortwave infrared (SWIR) bands typically used for computation of NDSI (e.g. MODIS Terra bands 4 [545 nm – 565 nm] and 6 [1628 nm – 1652 nm] [Riggs and Hall (2015)]; Landsat 8 TM bands 2 [450 nm – 515 nm] and 5 [845 nm – 885 nm] (Dozier, 1989)) [or Landsat 8 OLI bands 3 (533 nm – 590 nm) and 6 (1567 nm - 1651 nm)] are not measured by satellites in the PlanetScope constellation (Fig. 1).

Despite radiometric quality limitations, the PlanetScope imagery still provides useful snow observations that can be extracted using modern computer vision techniques. Motivated by methodological developments in machine learning on images that have been successful in other domains (e.g. biomedical imaging, Ronneberger et al. (2015)), remote sensing researchers are now applying these techniques to satellite imagery (e.g. for the identification of human structures (Iglovikov et al., 2018) or glacier calving fronts (Mohajerani et al., 2019)). Here we focus on developing and evaluating a convolutional neural network-based approach to enable the delineation of snow-covered area in high-resolution PlanetScope imagery. We use high-resolution, airborne lidar-derived snow cover data to serve as "ground truth" for co-located and contemporaneous PlanetScope imagery, providing a labeled dataset for neural network model training. We train and evaluate the model for well-studied snow sites, and assess both the absolute model performance compared to the ALS ground truth and the relative model performance across particular variables of interest (e.g. F-Score, Balanced Accuracy). Here, we develop a convolutional neural network (CNN) approach to derive snow-covered area from the 4-band PlanetScope imagery. We use high-resolution airborne lidar surveys (ALS) in the Sierra Nevada (California, USA) and Rocky Mountains (Colorado, USA) (Painter et al., 2016)) to train the model (Section 2.2.3) and evaluate model performance relative to ALS and other optical satellite SCA data products.

2. Methods and data

This study has three components: paired imagery and airborne lidar acquisition within the Tuolumne Basin, CA, USA and Upper Gunnison Basin, CO, USA study sites, convolutional neural network classification model development and training, and model performance evaluation. These steps are laid out in detail in this section and schematically described in Fig. 2. We describe the convolutional neutral network methodology and corresponding data requirements, our chosen study sites, the data acquisition procedure, and our performance evaluation methodology.

2.1. Study sites

We considered two study sites for our analysis: the Upper Tuolumne Basin in the Sierra Nevada mountains of California, USA (37.89°N, −119.25°W), and the Gunnison/East River Basin in the Central Rocky Mountains of Colorado, USA (39.08°N, −107.14°W). As demonstrated in Fig. 3, our area of analysis in the Colorado zone is actually located in the upper headwaters of the Crystal River/Roaring Fork River basin. However, ALS data for this region (see Section 2.2.3) are labeled with the Gunnison/East River dataset label, so we use that reference to maintain consistency with the data library. These sites were selected based on their temporally dense archives of airborne lidar-derived snow depth data and the fact that they span multiple climatological zones (Fig. 3), which enabled more thorough snowcover model performance evaluation. Primarily within a maritime climate, the Tuolumne region exhibits higher and more variable snow water equivalent accumulation when compared to the Gunnison site. While temperature patterns are similar, precipitation patterns exhibit more variability at the Tuolumne site. Tuolumne temperature, precipitation, and SWE observations were derived from a corrected time series at a California Department of Water Resources snow pillow site at Dana Meadows (SWE: 1980–2015; Temp. + Precip.: 2002–2015). Gunnison climatological observations were obtained from the United States Department of Agriculture National Resources Conservation Service SNOTEL site at North Lost Trail, CO (1985–2019; site No. 669; https://nrcs.usda.gov). We leveraged data from the Upper Tuolumne Basin for model training and evaluation, and performed additional evaluation in the Gunnison/East River basin (see Section 2.5).

2.2. Data acquisition and processing

2.2.1. PlanetScope satellite imagery

Planet is a commercial satellite imagery company that operates Earth

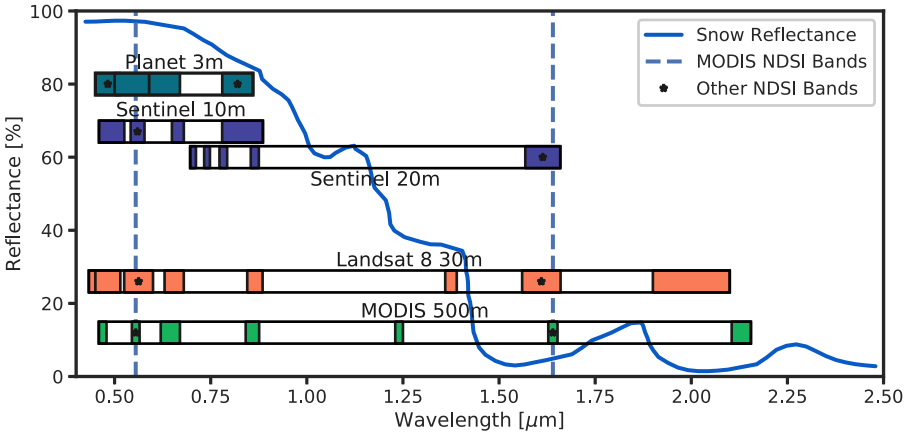

Fig. 1. Comparison of bands for Earth Observing satellites and reflectance of snow. Planet PlanetScope (cyan) observes a limited region of the snow reflectance spectrum (from Painter et al., 2009; solid blue line) compared to other optical sensors. Sentinel-2 (dark blue), Landsat 8 (orange), and MODIS (green) platforms observe a broad range of the visible, NIR and SWIR spectrum and thus are able to more readily differentiate snow using radiometric indices such as NDSI (e.g. Hall and Riggs (2007)). Bands used to compute NDSI for each observing platform are marked with asterisks, and MODIS bands are highlighted with dotted blue lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

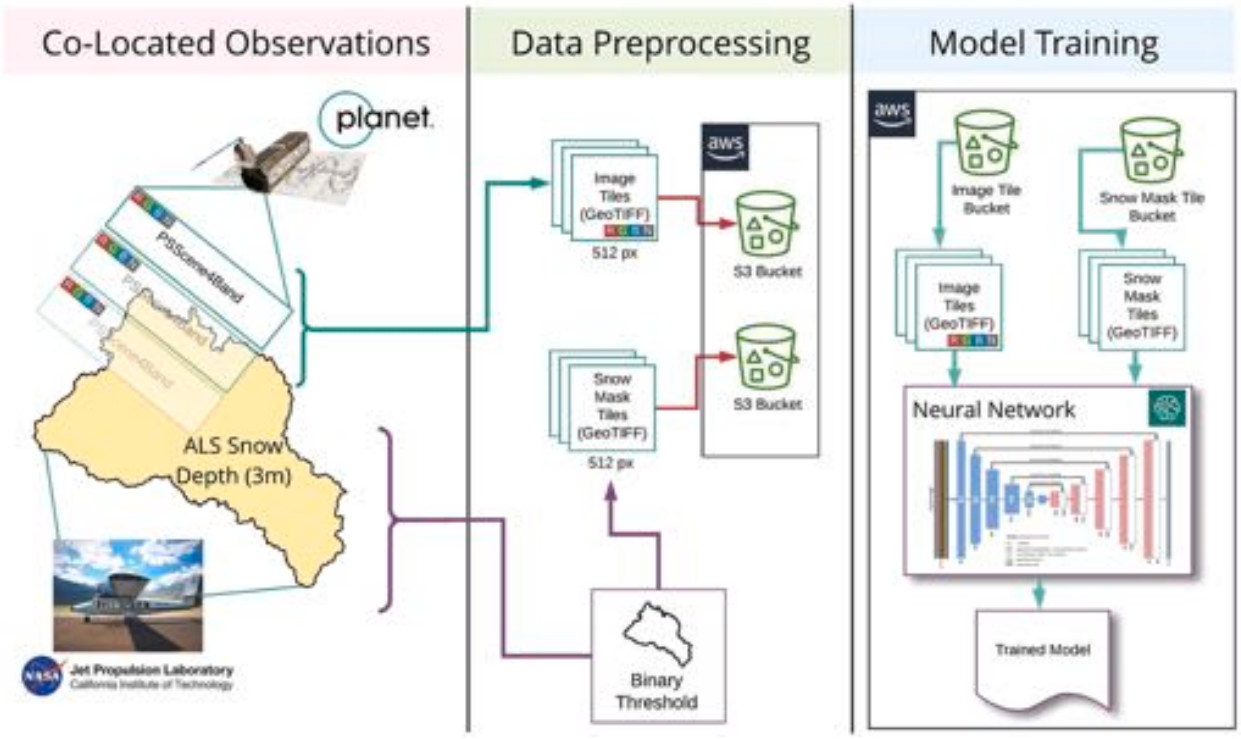

Fig. 2. Conceptual data processing and model training workflow. Our model training begins with contemporaneous, co-located PlanetScope imagery and ALS snow depth data, which are pre-processed into standardized, gridded GeoTIFFs. These GeoTIFF files are then used to train a convolutional neural network. The neural network structure is derived from Iglovikov et al. (2018), and uses 4 band PlanetScope GeoTIFF files as input.

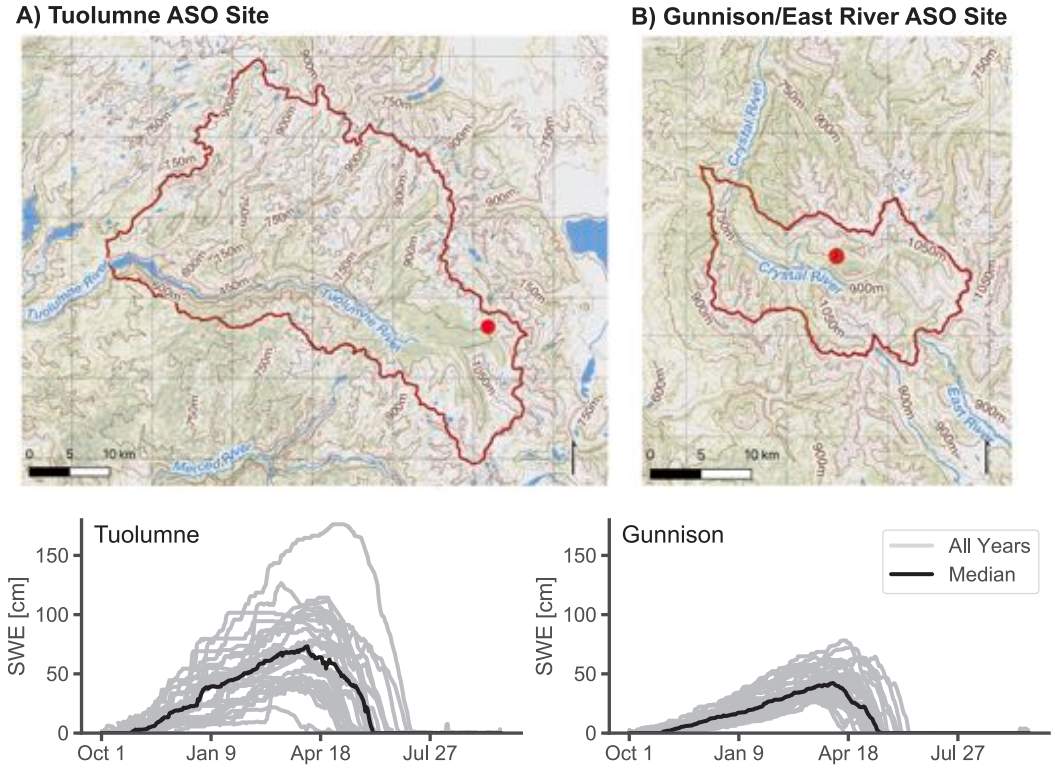

Fig. 3. Maps and plots of median daily snow water equivalent (bottom; black line; all years shown in gray) are shown for the Upper Tuolumne site in California, USA (37.89°N, −119.25°W) and the Upper Gunnison site in Colorado, USA (39.08°N, −107.14°W).. Maps show elevation, primary rivers, and tree canopy coverage (green shading). Tuolumne observations (1980–2015) were obtained from the Dana Meadows snow pillow via the California Department of Water Resources. Gunnison observations were obtained from the United States Department of Agriculture National Resources Conservation Service SNOTEL site at North Lost Trail, CO (1985–2019); site No. 669; https://nrcs.usda.gov. Red circles on maps indicate SWE measurement site locations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

observing satellite constellations, including the "PlanetScope" (PS) constellation of approximately 130 small ("3U" form factor: 10x10x30 cm) satellites in sun-synchronous orbit. This constellation collects approximately 200 M km^2 day^{-1} of optical (red, green, blue, and NIR) land-surface imagery at ~3.7 m GSD (at nadir) between ±81.5 degrees latitude, with approximately daily revisit times. With ongoing sensor development and satellite launches, the PlanetScope constellation currently includes multiple generations of instruments, including the Dove and Dove-R satellites (corresponding to "PS2" and "PS2.SD" data products, respectively). These instruments have similar design, with comparable (but not identical) spectral band centers and bandwidths (Table 1). For this study, we use the PS and PS2.SD products interchangeably, and develop a method that can leverage the full PlanetScope archive.

We used the Level-3B PlanetScope "Analytic Ortho Scene" orthorectified multispectral surface reflectance data products. Planet's internal data processing pipeline converts Level-1B top-of-atmosphere radiance (derived by applying sensor darkfield and flat field corrections to raw image data) to Level-3B surface reflectance using near-real-time MODIS data and the 6SV2.1 radiative transfer code (Kotchenova et al., 2008; Planet Labs, Inc., 2019b). These products have geo-location errors of less than 10 m RMSE (Planet Labs, Inc., 2019b). We contend that small errors in PlanetScope pixel alignment manifest relatively minor errors in our computed performance metrics, especially because we compare two binary snow maps: even a 2–3 pixel (6–9 m) misalignment would only cause misclassifications to be assigned at the edges of snow patches. Since we are working at high resolution, the number of pixels subject to these misclassifications would be lower than in a data source with coarse resolution. Only small errors in pixel alignment are likely to be present in these data, and the central results of this study are affected in only a minor way by them.

2.2.2. Scene selection and acquisition

For model training, we selected archived PlanetScope scenes with spatial and temporal overlap with our airborne lidar "ground truth" snowcover data. We selected images within a 7 day window around the lidar acquistion dates to ensure a higher probability of cloud-free imagery for similar snow conditions. PlanetScope imagery was available contemporaneously (on the same day) with airborne lidar data for both Gunnison and Tuolumne sites (see Tables A.2 and A.3). PlanetScope images were manually inspected for relative cloud fraction, and images with the best cloud-free coverage and shortest time offset were selected for inclusion into our analysis. Planet offers a cloud mask for the purpose of excluding cloudy pixels from analysis, but we found that the cloud mask often incorrectly classifies snow as cloud. Due to this limitation we chose to inspect images for clouds manually.

We used porder version 0.5.7, an open-source tool (Roy, 2019) for the Planet Orders v2 Application Programming Interface (API) (Planet Labs, Inc., 2019a), to both query the PlanetScope catalog and submit orders. Analytic Ortho Scene assets were queried via the "PSScene4-Band" identifier and the "analytic_sr" bundle identifier. We used the Planet Clips API to return only those pixels overlapping our areas of interest (e.g., lidar footprints), both to conserve our imagery quota and reduce data volume.

Table 1
Snow cover datasets used in this study for comparison to PlanetScope-derived snow cover. The "binary procedure" column describes the technique used to derive binary snow cover mask from continuous data fields. The Ref. column describes the reference used to derive the binarization threshold used.

Data	Obs. Type	Spatial Res.	Temporal Res.	Binary Procedure	Ref.
ASO Snow Depth	Airborne lidar	3 m	Weekly-Monthly, during ablation season	Threshold: Depth > 10 cm	Painter et al. (2016)
Sentinel 2 NDSI	Satellite	10 m	5 days	Threshold: NDSI >0.42	Drusch et al. (2012)
Landsat 8 fSCA	Satellite	30 m	16 Days	Threshold: fSCA >0.5	U.S. Geological Survey (2018)

2.2.3. Airborne lidar

Training and evaluating a neural network model requires high-resolution snowcover classification products to serve as "ground truth" per-pixel labels of snow presence/absence. For this study, we use snowcover masks derived from NASA/JPL Airborne Snow Observatory (ASO) data products (Painter et al., 2016). ASO releases gridded 3-m snow depth rasters for select watersheds across the western United States with monthly-weekly repeat intervals from mid-winter through late spring (March – June). These data are available from the National Snow and Ice Data Center's Distributed Data Access Center (`https://nsidc.org/data/ASO_3M_SD`). We defined a snow depth threshold to produce binary snow classification masks (snow, not snow) from the lidar-derived snow depth products, assuming the ASO pixels with snow depth above this threshold were completely snow covered and ASO pixels below this threshold were completely snow free. Our chosen threshold value was intended to reflect a conservative estimate of the vertical accuracy of lidar measurement of snow. Painter et al. (2016) observed vertical accuracy of 8 cm based on an assessment of ASO data in unforested terrain without topographic complexity, and Currier et al. (2019) observed a range of ALS vertical accuracy between 8 and 16 cm in mixed terrain. Our assessment of the literature combined with the heterogeneous nature of the Tuolumne watershed (both in vegetation cover and topographic complexity, which can reduce ALS accuracy) led us to conservatively choose a 10 cm threshold to generate a binary snow mask from the ASO observations. Using an ASO-derived snow depth collection in the Tuolumne basin from May 2018 (ID # `ASO_3M_SD_USCATE_20180528`), we observed a 0.97% difference in SCA between the 8 cm and 10 cm thresholds, which represents a discrepancy of only 4.43 km^2. We believe that our chosen threshold represents a conservative estimate of the uncertainty present in ALS measurements of the Tuolumne watershed. In addition, we derived a tree canopy mask from these data by applying a threshold of 1 m to a canopy height model derived from the ASO ALS survey data (see Ferraz et al. (2018) for methodology). These canopy height model data were obtained via personal correspondence with Jeff Deems (Upper Gunnison Basin) and Kat Bormann (Tuolumne basin). The threshold of 1 m was chosen following from previous efforts demonstrating differential ALS performance in shrub-covered vs. canopy-covered regions when compared to terrestrial surveys (Currier et al., 2019).

2.3. Convolutional neural network model

A convolutional neural network is a type of machine learning model, or algorithm. "Machine learning" is a term that is used to describe a set of statistical techniques to build predictive models of an outcome variable from data. Models are "trained" or "fit" to data by selecting a "training" subset of examples from the population of data. These examples are used to derive a predictive relationship with the response variable, and the methodology used to derive this relationship varies for different machine learning approaches. Once fit, models are assessed for their ability to accurately predict response variables given "unseen" samples of data (the "test" subset) which is disjoint from the training set.

In this study we employed a "supervised learning" approach, wherein the presence of the response variable in the data (known as a data "label") guides an algorithmic search for a statistical relationship between the input data ("features") and the response ("label"). Once a supervised learning model is fit using data that contains the response variable, the resulting statistical relationship can be employed to predict the response variable from unlabeled data.

Identifying the spatial extent and categorical classification of regions within images is known as "image segmentation" or "instance segmentation." Identification of snow-covered regions in satellite imagery fits well within this task definition. In our version of the task, the four bands of PlanetScope imagery at each pixel represent the input data to our model (the "features", see Section 2.2.1), and airborne lidar-derived binary snow presence represent the response variable ("labels", see Section 2.2.3).

We employ a convolutional neural network to accomplish this image segmentation task. Neural networks are specific types of machine learning methods designed to extract statistically meaningful linear combinations of input features from data (PlanetScope bands) and model a dependent variable (snow presence/absence) as a nonlinear function of these derived linear combinations (Hastie et al., 2009, Section 11.1). The particular network used in this study is based in the "U-Net" network architecture, a network previously shown to perform well in biomedical image segmentation (Ronneberger et al., 2015), modified by Iglovikov et al. (2018) to perform building detection in satellite remote sensing imagery. We use the resulting network, known as "TernausNetV2", for our satellite image segmentation task. The input to this model is a 512 × 512 pixel segment of a 4-band PlanetScope image, paired with it's corresponding ASO snow mask for training, and the output is a binary (0/1) snow mask. The TernausNetV2 network employs an "encoder-decoder" architecture with skip connections—this type of network allows for the identification and combination of both high-level (e.g. coarse) and low-level (e.g. fine) feature maps, or spatial representations of desired features (in our case: snow-covered regions). A desirable characteristic of this type of network is the potential for "transferability" across a wide variety of input types: for example, a single trained model could potentially perform well across different snow types or geography, despite not being "trained" on all input types. For further details regarding the precise layering architecture of the neural network, refer to Iglovikov et al. (2018). To our knowledge this method has not been applied to the segmentation of snow in satellite imagery. It is worth noting that many other machine learning methodologies exist, most of which would perform suitably given the snow classification task presented above, given well-formed training data. However, we have chosen this method due to its demonstrated successes with other types of classification in satellite imagery.

2.4. Model training

The neural network was trained using data from the Upper Tuolumne Basin site. For training, we paired a given set of binary snow mask tiles from a single ASO collection with the corresponding set of co-located, contemporaneous PS imagery tiles acquired within 7 days of the ASO acquisition date (Fig. 2), and divided this set of image-mask pairs into training and testing subsets via a 70/30% split. Our neural network and model training routines were modified from the implementation known as "Robosat.pink" (Courtin and Hofmann, 2019). Each training effort involved 50 epochs with a batch size of 7 and a learning rate of 2.5×10^{-5} (these values were chosen based on the sensible defaults presented in the Robosat.pink source neural network implementation, Courtin and Hofmann (2019)). We leveraged two ASO collection dates (Table A.2) in the Tuolumne region by repeating the above procedure for each ASO collection, initializing each subsequent model training procedure with the weights derived from the previous model training run. This allowed the training process to build upon previous training runs. The open-source implementation of this procedure and all associated data processing workflows are available on GitHub at `https://github.com/acannistra/planet-snowcover`.

2.5. Model performance evaluation

We use the ASO-derived snow cover mask as "ground truth" for evaluation of our PS model SCA, and compute pixel-based metrics of performance of our PS-derived snow mask with reference to the ASO snow mask. Additionally, to compare against existing SCA datasets, we acquired SCA products from Sentinel-2 and Landsat-8 that were both contemporaneous and co-located with the PlanetScope and ASO lidar snow masks described above for each study site. We computed normalized difference snow index (NDSI) from Sentinel-2 bands 3 (10 m) and 11 (nearest-neighbor resampled to 10 m) as:

$$\mathrm{NDSI} = \frac{\mathrm{B3} - \mathrm{B11}}{\mathrm{B3} + \mathrm{B11}}$$

and generated snowcover masks using an NDSI threshold of 0.42 (Drusch et al., 2012). We used 30-m Landsat 8 fSCA products available from the United States Geological Survey (U.S. Geological Survey, 2018), and generated snowcover masks using an fSCA threshold of 0.5. A summary of all SCA products used in this study is presented in Table 1.

For each snowcover dataset (e.g. CubeSat SCA, Landsat-8 fSCA, and Sentinel-2 NDSI) we computed precision, recall, F-score and balanced accuracy with reference to contemporaneous airborne lidar-derived snow cover (see Table 1).

Precision is the percentage of snow classifications predicted by our model that are also snow classifications in the compared dataset:

$$\mathrm{Precision} = \frac{\mathrm{True\ Positives}}{\mathrm{True\ Positives} + \mathrm{False\ Positives}} \tag{1}$$

Recall is the percentage of true snow classifications predicted by our model that are also true snow classifications in the compared dataset

$$\mathrm{Recall} = \frac{\mathrm{True\ Positives}}{\mathrm{True\ Positives} + \mathrm{False\ Negatives}} \tag{2}$$

F-Score or F1 score is the harmonic mean of precision and recall:

$$\mathrm{FScore} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{3}$$

Balanced accuracy normalizes the true positive and true negative predictions by the number of true positive and true negative samples to allow for a less biased assessment of accuracy given the relative accuracy of each prediction type. For the binary (two-class) case (e.g. snow/no-snow), it is computed as the arithmetic mean of sensitivity (true positive rate) and specificity (true negative rate) (see Mosley (2013) and Kelleher et al. (2015)):

$$\mathrm{Balanced\ Accuracy} = \frac{\mathrm{True\ Positive\ Rate} + \mathrm{True\ Negative\ Rate}}{2} \tag{4}$$

2.5.1. Within-region assessment and out-of-region transferability

We created an assessment regime using the above metrics to assess the model performance within the training region (Upper Tuolumne, CA) and compared it to model performance outside of the training region (Gunnison, CO). This strategy to allowed us to determine the transferability of a given trained model to a geographically/climatically distinct region. To do this we identified PS imagery that overlapped with an ASO "ground truth" snow cover mask in each basin, excluding any individual image with an overlap footprint of less than 2 M pixels (roughly 18 km2). In assessing Tuolumne (within-region) performance, we selected a subset of PS imagery via the 30% "test set" identified during the model training procedure which overlapped with a single ASO collection (`ASO_3M_SD_USCATE_20180528`, see Table A.2 and Fig. 4). This ensured that the chosen imagery was not part of the model training procedure. For the Gunnison (out-of-region) assessment, we selected a set of PS images that overlapped a single ASO collection (`ASO_3M_SD_USCOGE_20180524`, see Table A.3). We then used the trained model from the Tuolumne basin to predict snowcover for these Gunnison images, and computed the mean and standard deviation of each metric outlined above.

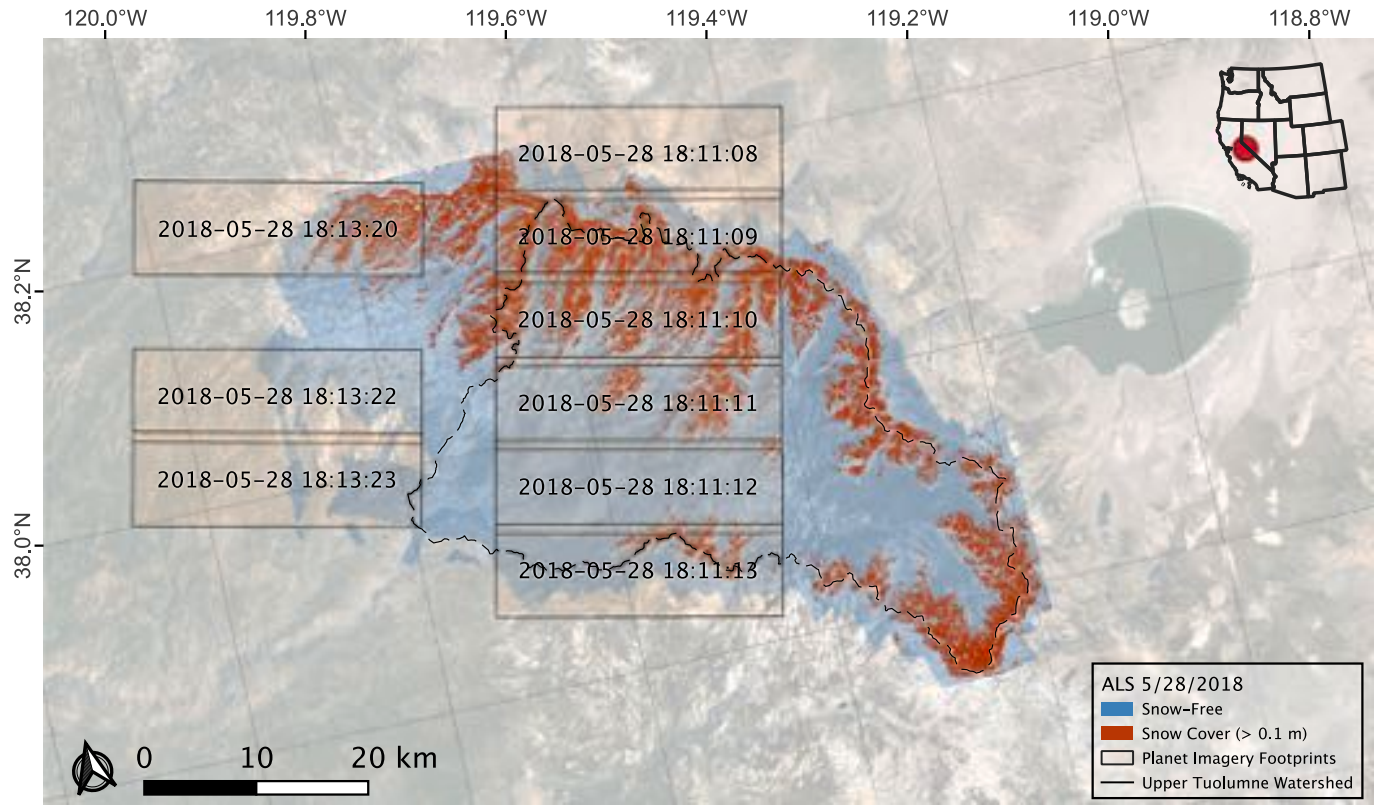

Fig. 4. A subset of the cloud-free May 28, 2018 PlanetScope scenes (boxes) and the contemporaneous May 28, 2018 ALS data (`ASO_3M_SD_USCATE_20180528`) for Tuolumne site. Red pixels were snow-covered in the ALS "ground truth" data, while blue pixels were identified as snow-free. Both datasets were used for model training and within-region evaluation. See Table A.2 for a full list of the PS scenes used for training. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.5.2. Comparison with other satellite datasets

We compared our PS model results with co-located snowcover products generated from other satellite datasets (Table 1) acquired within 5–15 days of the ASO observation acquisition date. We applied nearest-neighbor resampling to the ASO snow cover rasters to upsample ASO to match the native spatial resolution of each satellite snow cover dataset. Means and standard deviations of each metric described above were then computed with reference to the contemporaneous binarized ASO collection.

2.5.3. Performance under tree canopy

We also examined the effect of tree canopy cover on our PS model performance. In particular, we evaluated the ability of high-resolution PlanetScope imagery to identify snow in canopy-covered regions versus uncovered regions, and compared this performance to the snowcover products from other satellite sources. We applied the tree canopy mask described above to PS-based model predictions and each of the alternative SCA datasets and computed snow identification performance as above. We performed this analysis for all images included in earlier performance analyses at both the Tuolumne and Gunnison sites. Percentage of tree canopy coverage in these images was computed to be between 22 and 55%.

3. Results

We divide the evaluation of our snow classification model into three assessments. We first show that the PS snow classification model trained using data from the Upper Tuolumne Basin, California, USA produces improved snow classification results when compared with other satellite sources. Second, we find comparable performance when assessing the same model (fitted with Tuolumne data) using out-of-region data from the Gunnison River/East River Basin, Colorado, USA. Finally, we demonstrate mixed model performance under tree canopies in both the Tuolumne and the Upper Gunnison basins when compared to other SCA products.

3.1. Within-region model performance (Tuolumne, CA)

We used out-of-sample data from Tuolumne to compare our PS model output to the reference ASO snow cover, and snowcover products derived from Sentinel-2 NDSI and Landsat 8 fSCA across four metrics of performance (Fig. 5C). The PS model consistently outperforms the other, lower-resolution satellite products, though the standard deviation computed across test images do overlap between the PS model statistics and those of other products. For example, the PS model produces results with a mean F-score of 0.73 ($S.D. = 0.12$), compared to 0.64 (Landsat 8, $S.D. = 0.15$) and 0.63 (Sentinel-2, $S.D. = 0.14$). We also find differences in balanced accuracy, with a mean score of 0.82 ($S.D. = 0.07$) for our PS model, compared to 0.75 (Landsat 8, $S.D. = 0.07$) and 0.75 (Sentinel-2, $S.D. = 0.13$). See Fig. 5 and Table A.4.

3.2. Out of region model performance (Gunnison/East River basin, CO)

When evaluating performance using out-of-region PS imagery from the Gunnison/East River Basin, CO, the Tuolumne-trained model exhibits comparable performance to Landsat 8 fSCA- and Sentinel-2 NDSI-derived snow masks across three of four classification metrics (precision differs; Fig. 6C). For example, balanced accuracy for our PS Model was 0.75 ($S.D. = 0.08$), comparable to 0.77 (Landsat 8, S.D. = 0.10) and 0.76 (Sentinel-2, S.D. = 0.10); similarly, F-score for the PS Model was 0.67 (S.D. = 0.18), comparable to 0.73 (Landsat 8, S.D. = 0.19) and 0.68 (Sentinel-2, S.D. = 0.23). Fig. 7 provides an illustration of a 2 km by 2

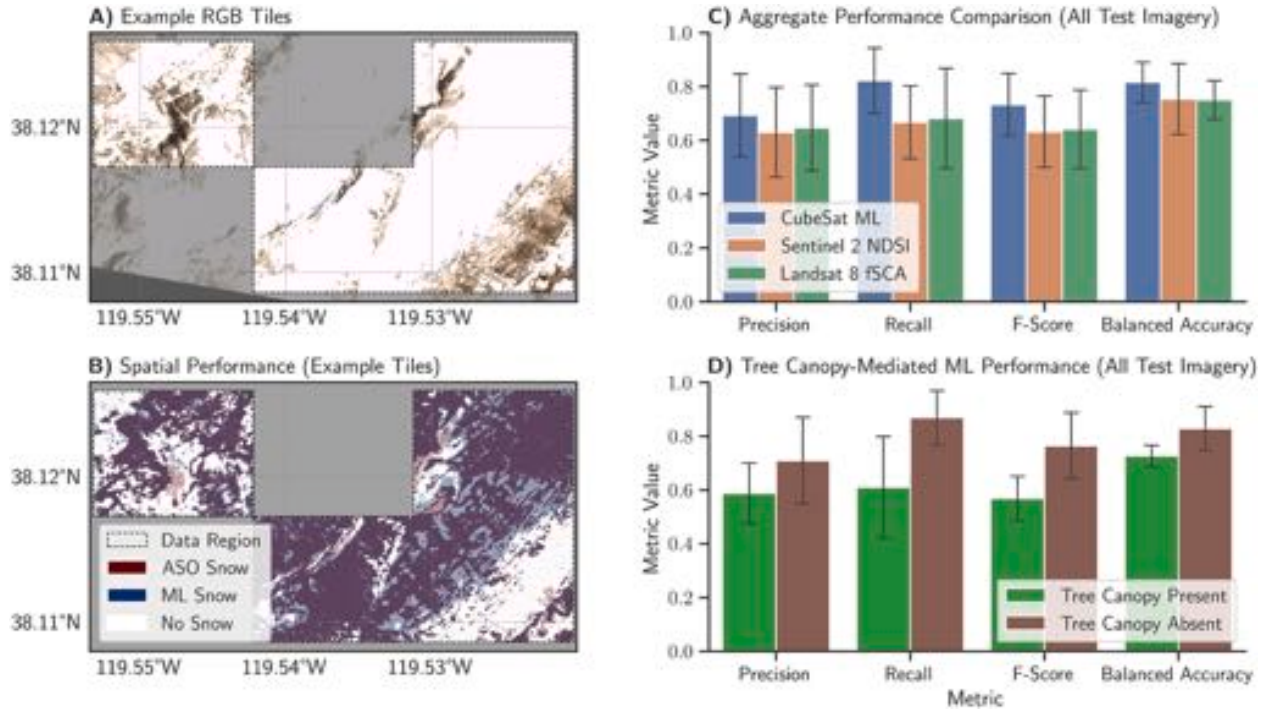

Fig. 5. Metrics of SCA segmentation performance, Tuolumne, CA. A) Example out-of-sample ("test") RGB PS tiles, gray shading indicates regions randomly withheld for training and therefore not suitable for evaluation; B) binary snow classification from PS model (blue) and ASO (red). Overlap is shown in darker hue. Gaps in ASO snow-covered area are due to the threshold value of 0.1 m during the binarization of ASO snow depth. C) Aggregate performance comparison across all Tuolumne out-of-sample imagery, compared to Sentinel-2 and Landsat 8 snow classifications. Bars indicate standard deviation across imagery. PS model snow masks perform as well or better than other classifications. D) Mean tree canopy-mediated model performance of PS model-derived snow cover, aggregated across all imagery. Bars indicate standard deviation. Canopy cover decreases model performance across all four metrics. Full statistics available in Tables A.4 and A.6. Includes copyrighted material of Planet Labs, Inc. All Rights Reserved. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

km segment of the data used to compute these metrics, highlighting the improved performance of PS-based snow masks over other methods. In addition, we observe that the presence of tree canopy (e.g. vegetation with height greater than 1 m from a lidar-derived canopy height model) causes all four performance metrics to decline in this single scene (Fig. 6, panel D). Complete metrics are available in Tables A.5 and A.7.

3.3. Model performance beneath tree canopy

For the Gunnison site, the PS model performance across three of four metrics (Recall, F-Score, and Balanced Accuracy) was lower for canopy-covered areas compared to both canopy-free areas and mean performance across entire images (Fig. 8A). Similar reduced accuracy under canopy was also observed in the 10 m Sentinel 2 SCA data product (Fig. 8B), but was considerably less prominent in the 30 m Landsat 8 fSCA data product (Fig. 8C). Full metrics are available in Table A.7.

4. Discussion

We demonstrated, through evaluation across two climatically distinct sites, that PlanetScope imagery can offer reliable discrimination of snow-covered regions at high spatial (~3 m) and temporal (~1–2 d) resolution (Figs. 5; 6C). Our convolutional neural network methodology enables SCA identification from PlanetScope data through deep learning and spectral-spatial pattern recognition. This approach leverages all available PlanetScope bands and proved successful to exploit the RGB and NIR band placements that posed challenges in deriving SCA via traditional radiometric indices (Fig. 1). This contribution represents both a novel application of an existing method (convolutional neural networks) to the detection of snow in remotely sensed optical imagery and a detailed assessment of PlanetScope imagery for snow remote sensing applications.

4.1. Comparison to current SCA approaches

Existing efforts to measure SCA at similarly high resolutions often rely upon layered methodologies, where each additional processing step (e.g. statistical or physically-based downscaling, fractional SCA algorithms) has the potential to add complexity, external data dependence, and error. While effective, these approaches can be limited by the methodological and technical assumptions inherent in downscaling coarse observational data. For example, while downscaling MODIS observations is a reliable approach in most cases, observational factors such as sensor viewing angle and forest presence have been shown to significantly degrade fSCA quality and ultimately to affect the downscaled SCA product (Cristea and Lundquist, 2016; Rittger et al., 2019). Fractional SCA derived from sensors with large field of view such as MODIS can be affected by viewing angle and resampling strategy, with errors as large as 50% over forested areas at the end of MODIS scan line (Xin et al., 2012). These errors will be smaller in PlanetScope data. PlanetScope images are released with nadir view angles of less than 5 degrees and fewer than 20% saturated pixels. Here we have demonstrated the effectiveness of a SCA retrieval method at native ~3 m GSD from daily, near-nadir PlanetScope observations. We note that the choice of 0.5 as our Landsat 8 fSCA binarization threshold is one that could introduce bias into our comparative results: 30 m pixels that have fSCA values of <0.5 certainly are not necessarily "snow-free," though we take them as such in our comparative analysis. However, we examined both 0.3 and 0.75 fSCA binarization thresholds, and discovered relatively minor differences in F-Score and Balanced Accuracy (with respect to ASO coarsened to 30 m, as above) when compared to a binarization threshold of 0.5. These differences were a result of compensatory offsets

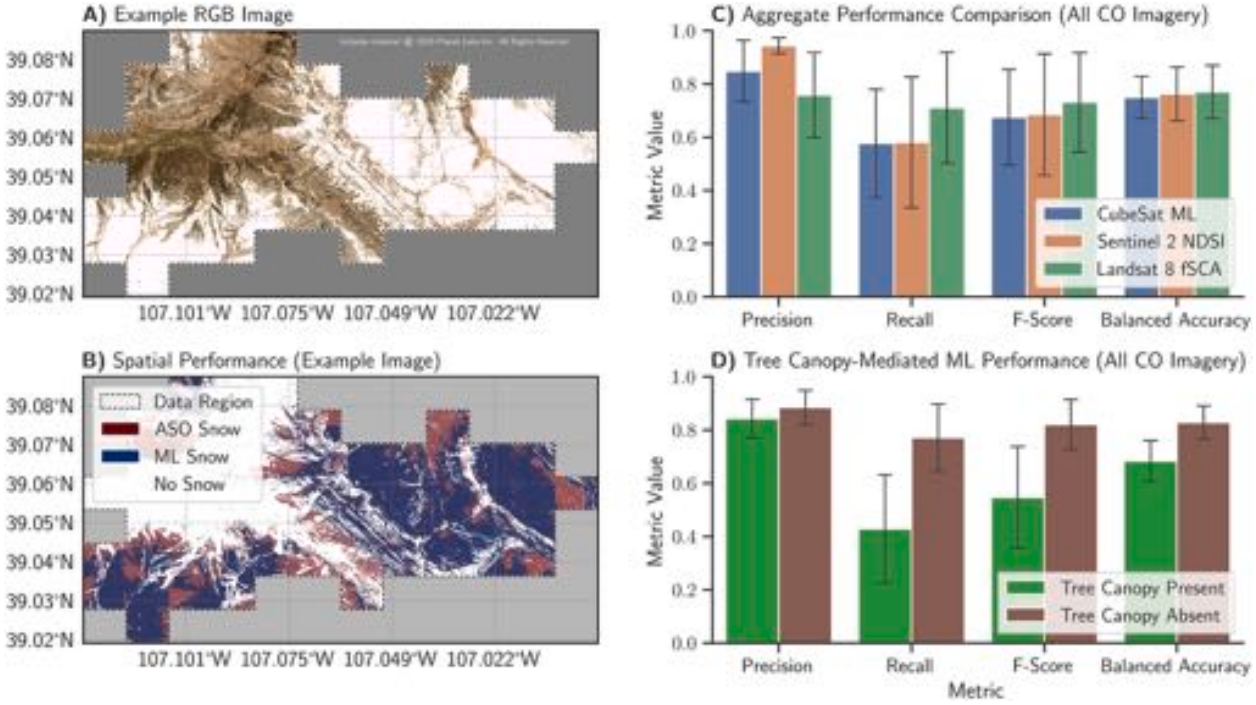

Fig. 6. Out-of-region performance of Tuolumne-trained snow-covered area identification model in the Gunnison River Basin, Colorado, USA. A) Sample PS image. B) Classification results. Our model exhibits systematic misclassification behavior, demonstrated here in a selected image (Planet Scene ID#: `20180524_172637_0f2d`, Panel A). Regions of the selected image that contain snow (Panel B; red coloring) are mis-classified as being snow-free by our model (Panel B; absence of blue coloring). C) Comparison with other satellite data. The PS model shows comparable performance to SCA derived from Sentinel-2 and Landsat 8 platforms across four metrics with reference to ASO ground truth (Panel C; Bars represent mean metric values across multiple distinct images covering a single ASO collection, error bars show standard deviation). D) Performance under canopy. Model performance metrics decline in the presence of tree canopy cover (canopy height > 1 m, Panel D) as determined by an ASO-derived canopy height model. Bars show means; error bars indicate standard deviation across all Gunnison imagery. Includes copyrighted material of Planet Labs, Inc. All Rights Reserved. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

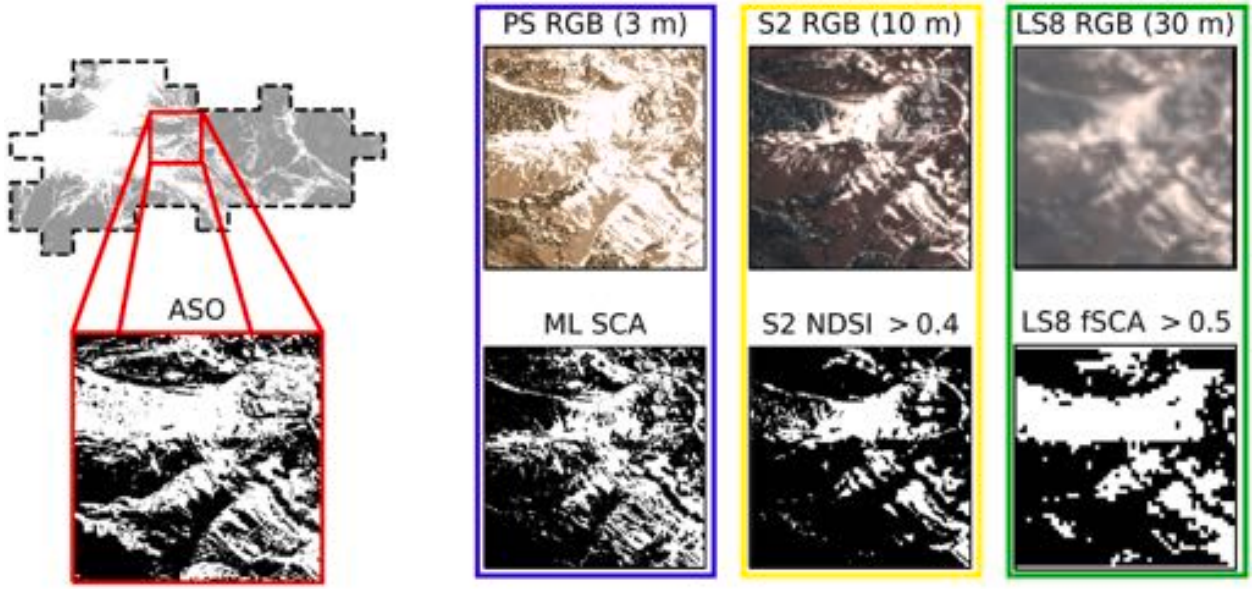

Fig. 7. Fine-scale patterns of snow distribution can be observed in 3 m PlanetScope derived SCA in contrast to coarser observations. This figure compares contemporaneous RGB observations and derived SCA for three optical satellite platforms (PlanetScope, Sentinel 2, and Landsat 8, see Table 1) to lidar-derived SCA within a 2 km by 2 km region selected from a single PS Scene (Planet Scene ID# `20180524_172637_0f2d`) within the Upper Gunnison River basin domain, depicted by dotted-line region and in Fig. 6. Includes copyrighted material of Planet Labs, Inc. All Rights Reserved.

of precision and recall: lower fSCA thresholds, for example, caused an increase in Recall (the fraction of True Positives), but a decrease in Precision, compared to the 0.5 threshold. Because both precision and recall are factors in computing F Score and Balanced Accuracy, these compensatory differences led to relatively small changes in those metrics (see Fig. A.1).

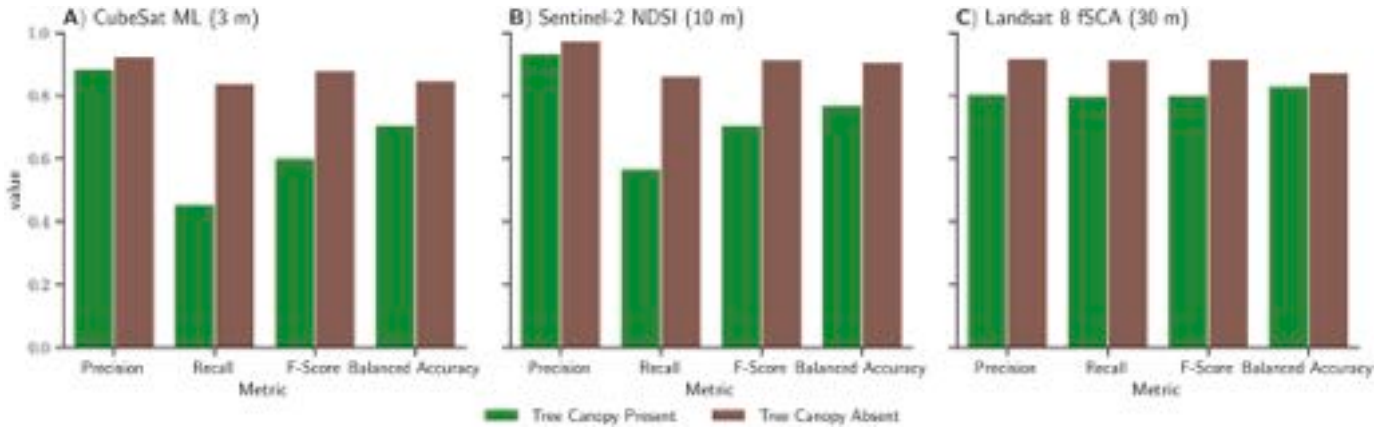

Fig. 8. Effects of tree canopy on SCA identification performance for a single Gunnison, CO PS scene (Percent Tree Canopy: 55%). Panels show impact of tree canopy on snow classification performance with A) the CubeSat (PS) model B) Sentinel-2 NDSI, and C) Landsat 8 fSCA for a single PlanetScope image (ID#: `20180524_172637_0f2d`, Fig. 6A) in the Upper Gunnison Basin, CO. Finer resolution SCA products (3 m CubeSat ML and 10 m Sentinel 2 NDSI, panels A and B) exhibit more pronounced performance differences between canopy-covered areas (green bars) and canopy-free areas (brown bars) than a coarser observational product (Landsat 8), where the discrepancies are less pronounced. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Comparisons of PS-derived SCA with other high-resolution optical satellite SCA data products (e.g. Figs. 5 and 6C) show that our method offers an accurate observational SCA technique. This finding is also promising for emerging multi-sensor fusion methodologies. Combining multisource and multitemporal remote sensing data with machine learning models has been suggested as a suitable method to overcome limitations of individual data sources (e.g. resolution, measurement type) (Ghamisi et al., 2019). For example, the global PlanetScope archive of daily imagery is an avenue for sourcing high-resolution imagery which could be further refined by other sensing methodologies with improved spectral bandwidth and band placement. Our methodology, when applied to this fused data product, may demonstrate further improved SCA segmentation performance while maintaining desirable spatial and temporal resolution characteristics. The advantages of using high-resolution snow data are once again shown in Fig. 7, where finer-scale patterns of snow distribution can be observed in PlanetScope-derived SCA relative to coarser products. The ability to observe these fine-scale patterns is a principal advantage of PS-derived data for scientific applications.

4.2. Utility of high-resolution snow cover

With the potential to be applied to the global catalog of PlanetScope imagery, our method is well-poised to serve as a useful tool across myriad scientific domains. For example, in studies of alpine ecology, snow disappearance date (SDD) has been shown to be a strong control on the regimes of ecological disturbance that mediate plant diversity and species persistence in heterogeneous alpine environments (Choler, 2005). Changing SDD has also been shown to cause reassembly in mountain wildflower communities (Theobald et al., 2017). However, due to the intensive nature of observing snow cover in the field, such ecological insights are necessarily accompanied with caveats regarding the accuracy of observed SDD or geographic scope. As such, high resolution SCA, both in space and time, could serve to both improve the accuracy of ecological studies reliant on SDD observations and allow for existing geographically distributed ecological datasets to be used in understanding the relationship of SDD to other ecological variables. Hydrologic modeling is another domain with a potential need for this approach; the spatially heterogeneous distribution of snow can be more accurately observed with higher-resolution SCA, with applications in predicting streamflow timing (Lundquist and Dettinger, 2005) and validating hydrologic model parameterizations (Clark et al., 2011; Luce et al., 1999).

4.3. Observed limitations

Similar to other studies focused on optical remote sensing of snow, tree canopy cover presents a challenge for our methodology. This is also illustrated in Fig. 7, where PlanetScope-derived SCA fails to identify snow cover in canopy-covered regions containing snow (as seen from ASO and PS RGB), but other radiometric approaches (e.g. Sentinel-2 NDSI and Landsat 8 fSCA) are more successful. More generally, the performance difference between canopy-present and canopy-absent regions observed in PS-derived snow masks is substantial (Figs. 5D; 6D; 8A), but compares to the difference observed in the Sentinel-2 NDSI-derived SCA product when examined at the Gunnison site (Fig. 8B). Though Landsat 8 fSCA performance exhibits a more muted decline in response to canopy presence (again at Gunnison site; Fig. 8C), previous studies have shown that both Landsat and MODIS experience degraded snow observation performance in tree-covered and tree-shadowed areas (Kane et al., 2008; Raleigh et al., 2013). It is also worth noting that both data source and snow-identification methodology play a role in snow-identification performance, thus it is prudent to consider the role of both variables when comparing across datasets. In addition, though we do not formally investigate the impact on cloud cover on snow identification performance, clouds are likely to present a challenge to SCA identification for our PS model, either via obfuscation or as a source for false-positives. This is especially true in comparison to other optical platforms due to the limited 4-band nature of PS imagery. However, future work in this area is warranted to examine whether the neural network model architecture can leverage the available PlanetScope bands and the unique spatial characteristics of cloud (e.g. texture, orientation) to better differentiate snow from cloud. In addition, Planet and other satellite imagery vendors continue to improve upon methodologies for discriminating clouds in their imagery products, which may improve the general applicability of our method in future iterations. The high temporal frequency of PlanetScope imagery acquisitions is another potential avenue for improving SCA detections over cloudy areas, as frequent collections could better enable cloud-free scenes to be captured. Finally, we have observed that topographical shadowing of snow is a factor that influences the quality of model-based identification of snow. Shadows from topography often “confuse” the neural network method, causing shadowed regions to be labeled snow-free. This effect is likely to be strongest at mid- to high-latitudes mid-winter, when sun illumination angles are lowest. In a similar fashion to cloud discrimination, future work should investigate the ability of a neural network methodology to discriminate snow from shadowed snow to improve the

utility of this methodology.

It is worth noting that we believe the F-score performance achieved here (0.73, compared to 0.64 and 0.63 for Landsat 8 fSCA and Sentinel-2 NDSI; Fig. 5 and Table A.4) is well within the range necessary for the operational use of this approach. The high-resolution nature of both the observational data and our "ground truth" lidar data make F-score at these resolutions a particularly conservative measure of performance. The F-score metric is the weighted harmonic mean of precision (*"How well does the model avoid labeling a snow pixel as a non-snow pixel?"*) and recall (*"How well does the model correctly identify all snow pixels?"*). Both of these metrics are challenging to achieve when comparing across observational modes (e.g. optical and lidar), but recall is particularly susceptible to the spatial heterogeneity present at high resolutions in our study sites when lidar is used as a ground-truth. Airborne lidar's ability to "see through" features such as forest canopies and be unaffected by shadows enables a "ground truth" dataset which is biased toward identifying *all* snow, compared to satellite-based observational methodologies. Any optical obfuscation (trees, rock shadow) dramatically reduces PlanetScope (and other optical observation platforms') capability when compared to airborne lidar. This obfuscation places a technical upper-bound on recall, thus affecting maximum F-score performance. We claim that the performance achieved here, especially when compared with that of similarly high-resolution SCA products (Sentinel-2, Landsat 8), represents a state-of-the-art observational contribution given the technical constraints discussed here.

Importantly, we do not consider snow identification beneath tree canopies to be a useful metric of the utility of this method, as canopy obfuscation puts a hard technical upper-bound on possible performance in these regions, as above. However, the ability of our methodology to identify snow in forested regions and in forest gaps is an important issue for further inquiry. Several techniques have been proposed to improve snow mapping in vegetated areas and among forest gaps. These include corrections using canopy reflectance models (Klein et al., 1998), observation-based approaches (Rittger et al., 2019), new radiometric indices (Wang et al., 2015), as well as combination of data from multiple sensors (Dressler et al., 2006; Durand et al., 2008; Raleigh et al., 2013). Although the current iteration of our neural network-based method does not yet achieve desirable performance in mapping snow in forested or cloudy areas, this methodology holds promise. A major contributor to the desirable performance of the convolutional "U-Net" architecture employed here is an underlying ability to "learn" meaningful spatial patterns from training data (Ronneberger et al., 2015), a quality that will perhaps improve canopy/cloud differentiation and improve SCA performance for these cases. A more robust approach for differentiating clouds from snow is also key to taking advantage of the global PlanetScope archive, as our current approach to training the model requires cloud-free scenes and thus considerable manual intervention. To leverage this potential opportunity, future work could incorporate forest canopy cover or cloud detection algorithms during the model training process, potentially allowing the model to differentiate canopy/cloud from non-obstructed snow, improving performance; however, we hypothesize that utilizing this methodology to enhance SCA segmentation performance among and between tree canopies may limit the geographic extensibility of a given model due to regional differences in forest structure.

PlanetScope imagery offers a novel lens into snow dynamics via a unique combination of spatial resolution and high revisit frequency. However, there are important technical considerations to acknowledge when utilizing these relatively new observational data. First, when compared to rigorously calibrated platforms such as MODIS or Landsat 8 OLI, imagery from Dove satellites (that comprise the PlanetScope fleet) exhibits radiometric calibration errors that can impact any analyses reliant upon them. Recent research has investigated methods to recalibrate PlanetScope imagery using coarser Landsat observations (Houborg and McCabe, 2016). Furthermore, two sensor types ("instruments") are present in the PlanetScope constellation ("PS2" and "PS2.SD") resulting from ongoing sensor development and satellite launches. These instruments have comparable (but not identical) spectral band centers and bandwidths (Table A.1). The inconsistency of the bandwidth and band centers in these PlanetScope data may affect the ability of our modeling methodology to identify snow. In addition, preliminary work in generating digital elevation models (DEMs) from PlanetScope imagery has revealed limitations in georeferencing accuracy, especially in remote or mountainous regions, potentially leading to offsets and RGB vs NIR band misalignment (Bhushan et al., 2021). These shortcomings are unlikely to be significant in applications such as snow cover identification, but future work to assess the magnitudes of error in these data is warranted.

4.4. Future work

This study demonstrates the continued utility of publicly available snow depth measurements derived from airborne lidar. Without the accurate snow depth products from the Airborne Snow Observatory, we would lack high-resolution "ground truth" data necessary for model training. Some opportunities for future work exist regarding the optimal approach to creating a binary snow mask from these snow depth data, and to integrate ASO spectrometer data for improved classification. We chose to use a threshold of 10 cm to convert the ASO snow depth field into a binary snow mask based on currently available field-validated assessments of ASO accuracy (see Section 2.2.3). However, it is likely that this choice had a significant impact on our results (see Section 4.1); qualitative examination of PlanetScope-derived SCA compared to actual contemporaneous imagery reveals that our method identifies snow-covered regions that in many cases are "no-snow" regions in the ASO-derived snow masks. This suggests that our model's identification of shallow/seasonal snowpacks may in some cases be incorrectly counted as a misclassification, lowering performance metrics. Future work should investigate the choice of threshold to determine the impacts of this choice on model performance.

The potential for our method to extract high-resolution snow cover over time across regions without ground-truth data is an important facet of our work. Given the global observational footprint of the PlanetScope imagery and the relative sparsity of repeat lidar snow depth observations, determining the skill of a model trained in a lidar-rich region and in other regions without lidar snow depth data is an important metric. In our assessment of model performance in the Upper Gunnison River Basin, CO (Fig. 6C) we observed good model performance compared to contemporaneous ALS observations when using the Tuolumne-based model to identify SCA. This result is perhaps not surprising because we chose to evaluate model performance during the ablation season (e.g. when lidar ground-truth is available), during which time the snowpack is more easily comparable between the Sierra Nevada and the Rockies. While it is necessary to take into account the above caveats regarding tree canopy structure differences across sites, this result suggests transferability potential—models trained in ALS-rich regions can be used in other regions to identify snow covered area without re-training the model in those new regions. This finding also suggests that the large but spatially constrained archive of existing ALS SCA data can be leveraged to create a model, or set of models, which is able to provide 3 m SCA observations across a much larger spatial footprint. Furthermore, the rich temporal archive of PlanetScope imagery, paired with a well-trained model as presented here, could mitigate some of the challenges of less-frequent observational platforms (such as cloudiness), and should be investigated as an avenue to supplement the temporal catalog of coarser SCA products. In summary, future work is warranted to better examine the role of the extensive PlanetScope catalog in improving training volume, investigating performance differences across geographic and topographic gradients, and improving remotely sensed SCA identification performance in cloudy areas.

5. Conclusions

We developed a method to identify snow covered area using PlanetScope CubeSat imagery, a ~3 m, ~daily, 4-band optical image dataset with ~global extent. Our method performs comparatively to other state-of-the-art remotely-sensed measures of snow cover (F-Score: 0.73, compared to 0.64 and 0.63 for Landsat 8 fSCA and Sentinel-2 NDSI). We show that these emerging image archives have great potential to accurately observe snow-covered area at high spatial and temporal resolutions despite limited radiometric bandwidth and band placement. We used a machine-learning based image segmentation approach, powered by a convolutional neural network modeling framework, to meet the challenge of producing a snow-covered area product from these data. Metrics of snow classification performance at our primary training and assessment site (the Upper Tuolumne Basin, CA, USA) show levels of classification accuracy comparable both to high-resolution airborne lidar surveys (ASO) and snow classifications derived from other platforms (Sentinel-2 and Landsat 8). The transferability potential of our model was examined via validation in a climatically and geographically distinct basin (Upper Gunnison, CO, USA), where we found slightly lower performance levels when compared to the model training basin. Finally, we demonstrate our model's diminished performance as a result of tree canopy cover in both study sites, suggesting the potential for future work in assessing the utility of the PS archive in identifying snow covered area in vegetated regions. The unparalleled spatial and temporal coverage of CubeSat imagery offers an excellent opportunity for satellite remote sensing of snow, with real implications for ecological and water resource applications.

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.

Acknowledgements

This work is supported by an NSF IGERT grant [DGE-1258485] and a NSF Graduate Research Fellowship [DGE-1762114] to A.F.C., a NASA THP award [80NSSC18K1405] to D.S., and an ESIP Incubator award and NSF EAR Award [EAR-1947875] to N.C. Data access and supplemental support was provided by the NASA Commercial Smallsat Data Acquisition Program 2018 Pilot Study. We also gratefully acknowledge support from Planet in solving technical challenges related to the use of their data, and from Shashank Bhushan for his productive conversations and expertise in using Planet data.

Appendix A. Appendix

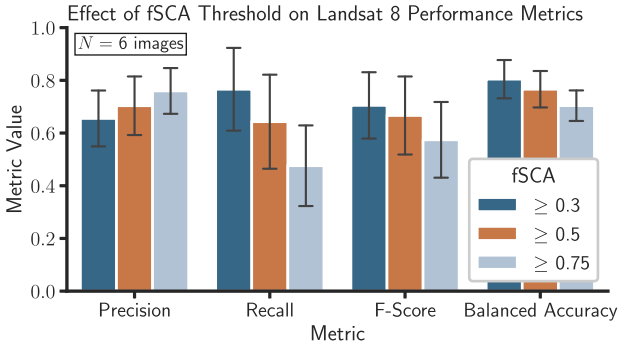

Fig. A.1: Comparison of Landsat 8-derived binary snow classification performance metrics (with reference to ASO) for different fSCA binarization thresholds in the Tuolumne basin. These values are generated from Landsat 8 pixels that overlap PlanetScope imagery and ASO snow depth observations, used elsewhere (e.g. Results, Discussion, Tables A.4–A.7) for comparision to ML-derived and Sentinel-2-based snow classification. Bars represent mean performance metric value and error bars represent standard deviation across $N = 6$ PlanetScope image footprints. Changes in recall and precision can be attributed to the changing specificity of pixel classification as fSCA threshold values change, but the opposing effects of fSCA on these two metrics leads to relatively consistent F-Score and Balanced Accuracy values with different fSCA thresholds.

Table A.1
Spectral bandwidth of PS2 and PS2.SD instruments within the PlanetScope constellation, (Planet Labs, Inc., 2019a).

	PS2	PS2.SD
Blue:	455–515 nm	464–517 nm
Green:	500–590 nm	547–585 nm
Red:	590–670 nm	650–682 nm
NIR:	780–860 nm	846–888 nm

Table A.2
Data used for training the Tuolumne regional model. Two ASO observations in the Greater Tuolumne region (`ASO_3M_SD_USCATE_20180528` and `ASO_3M_SD_USCASJ_20180601`) are overlapped with contemporaneous Planet imagery for training. These data together represent over 368 M labeled 3.7 m pixels with a footprint of approximately 5000 km^2.

ASO/Ground Truth ID	ASO Basin ID	ASO Basin Name	Planet Scene ID
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_180846_1002
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_180847_1002
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181108_1025
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181109_1025
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181110_1025
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181111_1025
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181112_1025
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181113_1025
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181319_1005
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181320_1005
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181322_1005
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181323_1005
ASO_3M_SD_USCASJ_20180601	USCASJ	SAN JOAQUIN MAIN FORK	20180601_181447_0f32
ASO_3M_SD_USCASJ_20180601	USCASJ	SAN JOAQUIN MAIN FORK	20180601_181448_0f32
ASO_3M_SD_USCASJ_20180601	USCASJ	SAN JOAQUIN MAIN FORK	20180601_181450_0f32
ASO_3M_SD_USCASJ_20180601	USCASJ	SAN JOAQUIN MAIN FORK	20180601_181451_0f32

Table A.3
Data used for model evaluation in Upper Gunnison, CO basin. One ASO observation (ASO_3M_SD_USCOGE_20180524) was used to identify contemporaneous PlanetScope imagery.

ASO/Ground Truth ID	ASO Basin ID	ASO Basin Name	Planet Scene ID
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172142_103d
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172143_103d
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172144_103d
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172145_103d
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172326_0f51
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172327_0f51
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172329_0f51
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172330_0f51
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172331_0f51
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172634_0f2d
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172635_0f2d
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172637_0f2d

Table A.4
Metrics of Snow Covered Area products in the **Tuolumne basin** as compared to ASO-derived "ground truth." Mean and Standard deviation are computed across 6 regions defined by PlanetScope scene boundaries. Only data not included in model training procedure are used in evaluation.

SCA	Metric	Mean	Std. Dev.
CubeSat ML	Balanced Accuracy	0.82	0.07
	F-Score	0.73	0.12
	Precision	0.69	0.15
	Recall	0.82	0.12
Landsat 8 fSCA	Balanced Accuracy	0.75	0.07
	F-Score	0.64	0.15
	Precision	0.65	0.16
	Recall	0.68	0.19
Sentinel 2 NDSI	Balanced Accuracy	0.75	0.13
	F-Score	0.63	0.13
	Precision	0.63	0.17
	Recall	0.67	0.14

Table A.5
Metrics of Snow Covered Area products in the **Upper Gunnison River basin** as compared to ASO-derived "ground truth." Mean and Standard deviation are computed across 12 regions defined by PlanetScope scene boundaries.

SCA	Metric	Mean	Std. Dev.
CubeSat ML	Balanced Accuracy	0.75	0.08

(*continued on next page*)

Table A.5 (*continued*)

SCA	Metric	Mean	Std. Dev.
	F-Score	0.67	0.18
	Precision	0.85	0.12
	Recall	0.58	0.20
Landsat 8 fSCA	Balanced Accuracy	0.77	0.10
	F-Score	0.73	0.19
	Precision	0.76	0.16
	Recall	0.71	0.21
Sentinel 2 NDSI	Balanced Accuracy	0.76	0.10
	F-Score	0.68	0.23
	Precision	0.94	0.03
	Recall	0.58	0.25

Table A.6
Metrics of classification performance of snow covered area products, including our PS Model, in vegetated and unvegetated regions of the **Tuolumne basin**. "Vegetation Only" corresponds to pixels containing vegetation >1 m as derived from the ASO Canopy Height Model. Mean and Standard deviation are computed across 6 regions defined by PlanetScope scene boundaries. Only data not included in model training procedure are used in evaluation.

SCA	Veg. Type	Metric	Mean	Std. Dev.
CubeSat ML	No Vegetation	Balanced Accuracy	0.83	0.08
		F-Score	0.76	0.12
		Precision	0.71	0.16
		Recall	0.87	0.10
	Vegetation Only	Balanced Accuracy	0.73	0.04
		F-Score	0.57	0.08
		Precision	0.59	0.11
		Recall	0.61	0.19
Landsat 8 fSCA	No Vegetation	Balanced Accuracy	0.75	0.08
		F-Score	0.65	0.17
		Precision	0.69	0.18
		Recall	0.67	0.20
	Vegetation Only	Balanced Accuracy	0.76	0.06
		F-Score	0.59	0.06
		Precision	0.52	0.08
		Recall	0.72	0.11
Sentinel-2 NDSI	No Vegetation	Balanced Accuracy	0.75	0.13
		F-Score	0.66	0.15
		Precision	0.68	0.19
		Recall	0.67	0.14
	Vegetation Only	Balanced Accuracy	0.73	0.13
		F-Score	0.50	0.07
		Precision	0.45	0.10
		Recall	0.61	0.14

Table A.7
Metrics of classification performance of snow covered area products, including our PS Model, in vegetated and unvegetated regions of the **Upper Gunnison River basin**. "Vegetation Only" corresponds to pixels containing vegetation > 1 m as derived from the ASO Canopy Height Model. Mean and Standard deviation are computed across 12 regions defined by PlanetScope scene boundaries.

SCA	Veg. Type	Metric	Mean	Std. Dev.
CubeSat ML	No Vegetation	Balanced Accuracy	0.83	0.06
		F-Score	0.82	0.09
		Precision	0.88	0.06
		Recall	0.77	0.13
	Vegetation Only	Balanced Accuracy	0.68	0.08
		F-Score	0.55	0.19
		Precision	0.84	0.07
		Recall	0.43	0.20
Landsat 8 fSCA	No Vegetation	Balanced Accuracy	0.78	0.12
		F-Score	0.75	0.23
		Precision	0.78	0.20
		Recall	0.73	0.25
	Vegetation Only	Balanced Accuracy	0.75	0.09
		F-Score	0.70	0.17
		Precision	0.72	0.15
		Recall	0.68	0.19
Sentinel-2 NDSI	No Vegetation	Balanced Accuracy	0.81	0.10
		F-Score	0.77	0.21

(*continued on next page*)

Table A.7 (*continued*)

SCA	Veg. Type	Metric	Mean	Std. Dev.
		Precision	0.94	0.06
		Recall	0.69	0.24
	Vegetation Only	Balanced Accuracy	0.72	0.09
		F-Score	0.60	0.23
		Precision	0.94	0.02
		Recall	0.48	0.24

References

Andreadis, K.M., Lettenmaier, D.P., 2006. Assimilating remotely sensed snow observations into a macroscale hydrology model. Adv. Water Resour. 29, 872–886. URL.

Bhushan, S., Shean, D., Alexandrov, O., & Henderson, S. (2021). Automated digital elevation model (dem) generation from very-high-resolution planet skysat triplet stereo and video imagery. ISPRS J. Photogramm. Remote Sens., 173, 151–165. Doi: https://doi.org/10.1016/j.isprsjprs.2020.12.012.

Boelman, N.T., Liston, G.E., Gurarie, E., Meddens, A.J.H., Mahoney, P.J., Kirchner, P.B., Bohrer, G., Brinkman, T.J., Cosgrove, C.L., Eitel, J.U.H., Hebblewhite, M., Kimball, J. S., LaPoint, S., Nolin, A.W., Pedersen, S.H., Prugh, L.R., Reinking, A.K., Vierling, L. A., 2019. Integrating snow science and wildlife ecology in Arctic-boreal North America. Environ. Res. Lett. 14, 010401 https://doi.org/10.1088/1748-9326/aaeec1.

Carlson, B.Z., Choler, P., Renaud, J., Dedieu, J.-P., Thuiller, W., 2015. Modelling snow cover duration improves predictions of functional and taxonomic diversity for alpine plant communities. Ann. Bot. 116, 1023–1034. https://doi.org/10.1093/aob/mcv041.

Choler, P., 2005. Consistent shifts in alpine plant traits along a mesotopographical gradient. Arct. Antarct. Alp. Res. 37, 444–453. https://doi.org/10.1657/1523-0430(2005)037[0444:CSIAPT]2.0.CO;2.

Clark, M.P., Hendrikx, J., Slater, A.G., Kavetski, D., Anderson, B., Cullen, N.J., Kerr, T., Hreinsson, E.O., Woods, R.A., 2011. Representing spatial variability of snow water equivalent in hydrologic and land-surface models: a review. Water Resour. Res. 47, W07539 https://doi.org/10.1029/2011WR010745.

Courtin, O., Hofmann, D.J., 2019. RoboSat.pink Computer Vision framework for GeoSpatial Imagery. DataPink. URL. http://RoboSat.pink.

Cristea, N.C., Lundquist, J.D., 2016. An Evaluation of Terrain-Based Downscaling of MODIS-Based Fractional Snow Covered Area Datasets over the Tuolumne River, CA Based on Lidar-Derived Snow Data, p. 4. Seattle, WA.

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.

Dedieu, J.-P., Carlson, B.Z., Bigot, S., Sirguey, P., Vionnet, V., Choler, P., 2016. On the importance of high-resolution time series of optical imagery for quantifying the effects of snow cover duration on alpine plant habitat. Remote Sens. 8, 481. URL.

Dozier, J., 1989. Spectral signature of alpine snow cover from the landsat thematic mapper. Remote Sens. Environ. 28, 9–22. URL. http://www.sciencedirect.com/science/article/pii/0034425789901016. https://doi.org/10.1016/0034-4257(89)90101-6.

Dressler, K.A., Leavesley, G.H., Bales, R.C., Fassnacht, S.R., 2006. Evaluation of gridded snow water equivalent and satellite snow cover products for mountain basins in a hydrologic model. Hydrol. Process. 20, 673–688. https://doi.org/10.1002/hyp.6130.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution Mission for GMES operational services. Remote Sens. Environ. 120, 25–36. URL:

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. URL.

Fernandes, R., Zhao, H., Wang, X., Key, J., Qu, X., Hall, A., 2009. Controls on northern hemisphere snow albedo feedback quantified using satellite earth observations. Geophys. Res. Lett. 36 https://doi.org/10.1029/2009GL040057.

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

Ford, K.R., Ettinger, A.K., Lundquist, J.D., Raleigh, M.S., Hille Ris Lambers, J., 2013. Spatial heterogeneity in ecologically important climate variables at coarse and fine scales in a high-Snow Mountain landscape. PLoS One 8, e65008. https://doi.org/10.1371/journal.pone.0065008.

Geological Survey, U.S., . Collection-1 Landsat Level-3 Fractional Snow Covered Area (FSCA) Science Product. URL. https://www.usgs.gov/centers/eros/science/usgs-eros-archive-landsat-landsat-level-3-fractional-snow-covered-area-fsca. https://doi.org/10.5066/F7XK8DS5.

Ghamisi, P., Rasti, B., Yokoya, N., Wang, Q., Hofle, B., Bruzzone, L., Bovolo, F., Chi, M., Anders, K., Gloaguen, R., Atkinson, P.M., Benediktsson, J.A., 2019. Multisource and multitemporal data fusion in remote sensing: a comprehensive review of the state of the art. IEEE Geosci. Remote Sens. Magaz. 7, 6–39. https://doi.org/10.1109/MGRS.2018.2890023.

Hall, D.K., Riggs, G.A., 2007. Accuracy assessment of the MODIS snow products. Hydrol. Process. 21, 1534–1547. https://doi.org/10.1002/hyp.6715.

Hall, D.K., Riggs, G.A., 2011. Normalized-difference snow index (NDSI). In: Singh, V.P., Singh, P., Haritashya, U.K. (Eds.), Encyclopedia of Snow, Ice and Glaciers. Springer Netherlands, Dordrecht, pp. 779–780. https://doi.org/10.1007/978-90-481-2642-2_376.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning volume 27 of Springer Series in Statistics. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-84858-7. arXiv:1010.3003.

Houborg, R., McCabe, M.F., 2016. High-resolution NDVI from Planet's constellation of earth observing Nano-satellites: a new data source for precision agriculture. Remote Sens. 8, 768. https://doi.org/10.3390/rs8090768.

Iglovikov, V, Seferbekov, S., Buslaev, A., Shvets, A., 2018. TernausNetV2: Fully Convolutional Network for Instance Segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 233–237. URL: http://arxiv.org/abs/1806.00844.

Immerzeel, W.W., Droogers, P., de Jong, S.M., Bierkens, M.F.P., 2009. Large-scale monitoring of snow cover and runoff simulation in Himalayan river basins using remote sensing. Remote Sens. Environ. 113, 40–49. https://doi.org/10.1016/j.rse.2008.08.010.

Kane, V.R., Gillespie, A.R., McGaughey, R., Lutz, J.A., Ceder, K., Franklin, J.F., 2008. Interpretation and topographic compensation of conifer canopy self-shadowing. Remote Sens. Environ. 112, 3820–3832. https://doi.org/10.1016/j.rse.2008.06.001.

Kelleher, J.D., Mac Namee, B., D'Arcy, A., 2015. Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. The MIT Press, Cambridge, Massachusetts.

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

Kotchenova, S.Y., Vermote, E.F., Levy, R., Lyapustin, A., 2008. Radiative transfer codes for atmospheric correction and aerosol retrieval: intercomparison study. Appl. Opt. 47, 2215. https://doi.org/10.1364/AO.47.002215.

Little, R., Peterson, D., Conquest, L., 1994. Regeneration of sub-alpine fir (Abies lasiocarpa) following fire - effects of climate and other factors. Can. J. Forest Res.-Revue Can. Rec. Forest. 24, 934–944. https://doi.org/10.1139/x94-123.

Luce, C.H., Tarboton, D.G., Cooley, K.R., 1999. Sub-grid parameterization of snow distribution for an energy and mass balance snow cover model. Hydrol. Process. 13, 1921–1933. https://doi.org/10.1002/(SICI)1099-1085(199909)13:12/13<1921::AID-HYP867>3.3.CO;2-J.

Lundquist, J.D., Dettinger, M.D., 2005. How snowpack heterogeneity affects diurnal streamflow timing. Water Resour. Res. 41, W05007 https://doi.org/10.1029/2004WR003649.

Mohajerani, Y., Wood, M., Velicogna, I., Rignot, E., 2019. Detection of glacier calving margins with convolutional neural networks: a case study. Remote Sens. 11, 74. https://doi.org/10.3390/rs11010074.

Mosley, L., 2013. A Balanced Approach to the Multi-Class Imbalance Problem. Graduate Theses and Dissertations. https://doi.org/10.31274/etd-180810-3375.

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. URL:

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. URL:

Planet Labs, Inc, 2019a. Planet Developer Resource Center. URL. https://developers.planet.com/docs/orders/.

Planet Labs, Inc, 2019b. Planet Imagery Product Specifications. URL. https://assets.planet.com/docs/Planet_Combined_Imagery_Product_Specs_letter_screen.pdf.

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. URL:

Riggs, G.A., Hall, D.K., 2015. MODIS Snow Products Collection 6 User Guide. Technical Report. URL: https://nsidc.org/sites/nsidc.org/files/files/MODIS-snow-user-guide-C6.pdf.

Rittger, K., Raleigh, M.S., Dozier, J., Hill, A.F., Lutz, J.A., Painter, T.H., 2019. Canopy adjustment and improved cloud detection for remotely sensed snow cover mapping. Water Resour. Res. 55 https://doi.org/10.1029/2019WR024914.

Rochefort, R.M., Little, R.L., Woodward, A., Peterson, D.L., 1994. Changes in sub-alpine tree distribution in western North America: a review of climatic and other causal factors. The Holocene 4, 89–100. https://doi.org/10.1177/095968369400400112.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. URL: http://arxiv.org/abs/1505.04597.

Roy, S., . samapriya/porder: Porder: Simple CLI for Planet ordersV2 API. URL. https://zenodo.org/record/3547667#.XhkH9FNKjOY. https://doi.org/10.5281/zenodo.3547667.

Schattan, P., Schwaizer, G., Schöber, J., Achleitner, S., 2020. The complementary value of cosmic-ray neutron sensing and snow covered area products for snow hydrological modelling. Remote Sens. Environ. 239, 111603. URL: http://www.sciencedirect.com/science/article/pii/S0034425719306236 https://doi.org/10.1016/j.rse.2019.111603.

Theobald, E.J., Breckheimer, I., HilleRisLambers, J., 2017. Climate drives phenological reassembly of a mountain wildflower meadow community. Ecology 98, 2799–2812. https://doi.org/10.1002/ecy.1996.

Venn, S.E., Green, K., Pickering, C.M., Morgan, J.W., 2011. Using plant functional traits to explain community composition across a strong environmental filter in Australian alpine snowpatches. Plant Ecol. 212, 1491–1499. https://doi.org/10.1007/s11258-011-9923-1.

Wang, X.-Y., Wang, J., Jiang, Z.-Y., Li, H.-Y., Hao, X.-H., 2015. An effective method for snow-cover mapping of dense coniferous forests in the upper Heihe River basin using Landsat operational land imager data. Remote Sens. 7, 17246–17257. https://doi.org/10.3390/rs71215882.

Xin, Q., Woodcock, C.E., Liu, J., Tan, B., Melloh, R.A., Davis, R.E., 2012. View angle effects on MODIS snow mapping in forests. Remote Sens. Environ. 118, 50–59. https://doi.org/10.1016/j.rse.2011.10.029.