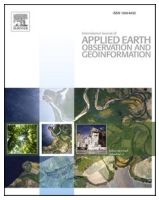
International Journal of Applied Earth Observations and Geoinformation 114 (2022) 103030

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

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



Estimating fractional snow cover in vegetated environments using MODIS surface reflectance data

Xiongxin Xiao [a], Tao He [a,*], Shunlin Liang [b], Xinyan Liu [a], Yichuan Ma [a], Shuang Liang [c], Xiaona Chen [d]

[a] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[b] Department of Geography, The University of Hong Kong, 999077, Hong Kong, China
[c] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
[d] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Keywords:
Fractional snow cover
MODIS
Forest cover
Viewing angle
North America

ABSTRACT

Advances in snow-cover mapping techniques have resulted in more accurate estimation of fractional snow cover (FSC) in areas with no vegetation; however, vegetation interference limits the accuracy of available snow cover information from satellite observations. The aim of this study was to develop a robust and enhanced FSC-retrieval algorithm using Moderate Resolution Imaging Spectroradiometer (MODIS) surface reflectance data for vegetated areas. The experiments were conducted in North America, where vegetation cover is complex and heterogeneous, using 28 Landsat-8 – MODIS image pairs acquired for the entire snow cover season (September 2015–May 2016). The FSC retrieval models were established from 20 sub-models based on the Extremely Randomized Trees method incorporating input information from multiple sources, such as commonly used variables, vegetation- and snow-related variables, location and geometry related variables, and other auxiliary variables. The FSC retrieval models were divided into forest- and non-forest types. We further investigated a canopy correction method to mitigate vegetation interference effects caused by the viewing geometry of satellite observations. The results show that the integration of 20 sub-models largely decreased model dependence on the training sample quality and improved the robustness of the model predictions. In the validation of the independent dataset, there was a noticeable improvement in FSC estimation for different land-cover and vegetation-cover types, with root-mean-square errors (RMSEs) reduced by an average of 11% compared to the Trimmed-Model. The application of canopy correction under the "Recommend" conditions (i.e., viewing zenith angle in $[45^{\circ}, 70^{\circ}]$ and fraction of forest cover in $[0, 0.3]$) improved the FSC prediction accuracy. Moreover, based on a comparison with the MOD10A1-based FSC map, our FSC estimation showed improved consistency across various vegetation coverages based on the Landsat reference FSC values, with 40% lower RMSEs and 8% increase in overall accuracy.

1. Introduction

Snow cover has critical implications for the climate and hydrology of mid- to high-latitude regions, is a major component of the water cycle in most areas of North America, which has ephemeral snow cover (Dobreva and Klein, 2011; Pulliainen et al., 2020; Xiao et al., 2022, 2020). High albedo and low thermal conductivity are distinctive features of snow cover that modulate the surface energy balance (Déry and Brown, 2007; Santolaria-Otín and Zolina, 2020). A decrease in snow cover effects the functional composition and diversity of terrestrial and maritime ecosystems in high-latitude regions (Bormann et al., 2018; Niittynen et al., 2020). Snow also affects a variety of high-latitude climate processes and feedback mechanisms (Watanabe et al., 2020; Wu and Chen, 2016), including the East Asian summer monsoon (Xiao and Duan, 2016) and global atmospheric circulation (Li et al., 2018).

Among the possible tools that provide information about snow cover, optical sensors in the visible and near-infrared spectra have been popular since the 1960s (Dietz et al., 2012; Rittger et al., 2020a). These sensors undertake continuous measurements of snow cover that can be applied in a range of rich contexts, including studies on the changes and

* Corresponding author.
E-mail address: taohers@whu.edu.cn (T. He).

https://doi.org/10.1016/j.jag.2022.103030
Received 24 April 2022; Received in revised form 14 August 2022; Accepted 19 September 2022
Available online 26 September 2022

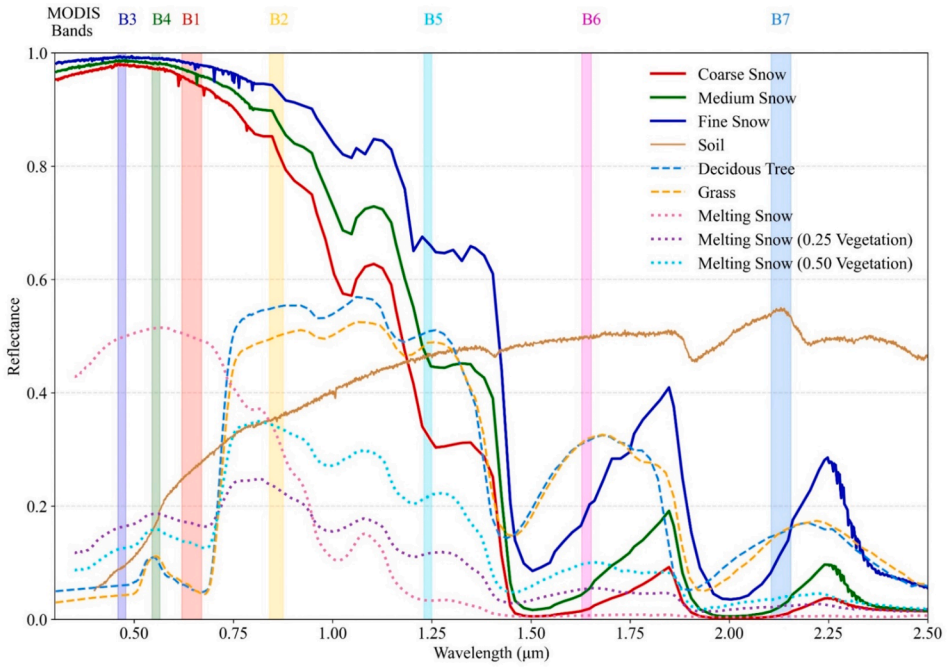


Fig. 1. Examples of spectral reflectance of the snow and non-snow endmembers from the JPL spectral library and a subset of the snow endmembers modeled through SNICAR for different effective snow grain (Baldridge et al., 2009; Warren and Wiscombe, 1980). The shadow wide indicates the spectral bandwidths of MODIS bands 1–7. The reflectance of snow within fine grain size (24 μm), medium grain size (82 μm), and coarse grain size (178 μm) generated by the SNICAR radiative transfer model are also shown.

Table 1
Summary of MODIS dataset used in this study.

Dataset Name	Variable	Notes
MOD09GA	Surface reflectance, View and solar zenith angle, view and solar azimuth angle	Possible predictor variables
MCD12Q1	Land cover types	
MOD44B	Percent Tree Cover	
MOD11A1	Land surface temperature	
MCD43A3	Black-sky albedo and White-sky albedo for visible, shortwave, NIR broadbands	
MOD10A1	Normalized Difference Snow Index (NDSI)	FSC results comparison

trends in snow cover over time and space (Bormann et al., 2018; Pulliainen et al., 2020). Moderate Resolution Imaging Spectroradiometer (MODIS) surface reflectance data is widely used for monitoring global and regional snow cover. For example, Fig. 1 displays a subset of the snow and non-snow endmember spectral libraries in the MODIS bands. Importantly, the spectral reflectivity of snow can vary depending on snowpack properties (Bohn et al., 2021; Warren, 1982), such as snow depth, grain size, liquid water content, and dust content, and thus, can vary both in time and space (Aalstad et al., 2020; Painter et al., 2003). The time-dependence of reflectance makes its variation a continuous process, which does not perfectly correlate with the visible snow area. The spatial heterogeneity of the different end-members within the mixed pixels is likely to lead to significant deviations in the observed reflectance (Fig. 1) (Hannula et al., 2020; Painter et al., 2009), which makes snow-cover detection extremely challenging. Traditionally, snow-cover mapping techniques identify pixels as binary snow cover, i.e., either snow-covered or snow-free. However, Cortés et al. (2014) pointed out that this can lead to significant bias compared to more precise fractional snow cover (FSC) metrics. Moreover, binary snow cover does not efficiently capture sub-pixels characteristics.

Traditional formula-based approaches rely strongly on a priori knowledge to obtain model parameters (Yuan et al., 2020). However, the connections and interactions within the Earth system usually exhibit multivariate and complex nonlinearities and the parameters for characterizing these processes vary over time and space (Lary et al., 2016;

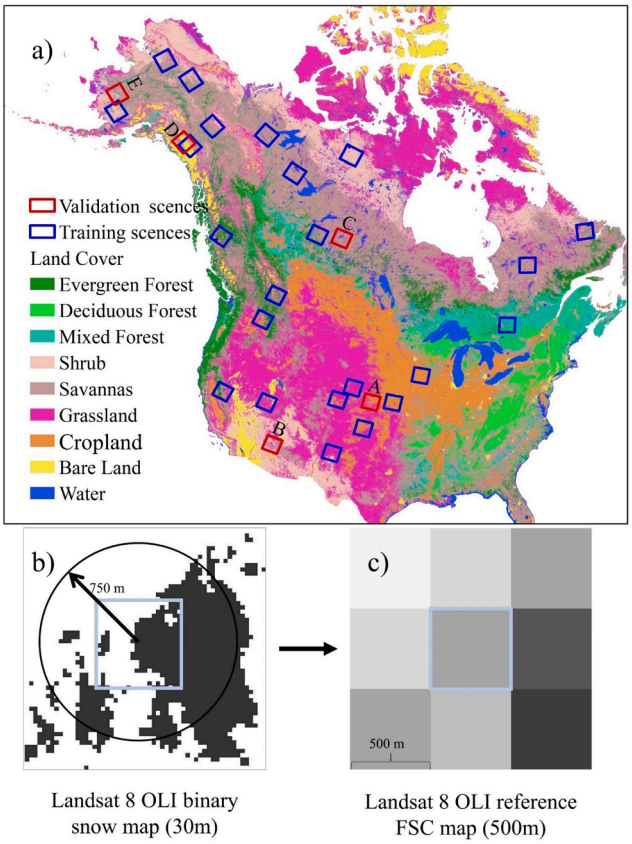


Fig. 2. a) Training and validation Landsat scenes across North America, and background is land cover information from MODIS products; b) Landsat 8 OLI binary snow map with 30 m resolution were used to calculate reference FSC map (500 m); c) within a 750 m circular radius.

Salcedo-Sanz et al., 2020). To complement traditional formula-based approaches, machine learning methods have become critical in environmental remote sensing sciences and have clear advantages for

Table 2
The description of the possible predictor variables of FSC estimation.

Category	Variable name	Description
Generality	B1	Surface reflectance in band 1 (620–670 nm)
	B2	Surface reflectance in band 2 (841–876 nm)
	B3	Surface reflectance in band 3 (459–479 nm)
	B4	Surface reflectance in band 4 (545–565 nm)
	B5	Surface reflectance in band 5 (620–670 nm)
	B6	Surface reflectance in band 6 (1628–1652 nm)
	B7	Surface reflectance in band 7 (2105–2155 nm)
	NDSI	Normalized difference snow index
	NDVI	Normalized difference vegetation index
	LC	Land cover type in regrouped classification schemes
Vegetation-Snow	NDFSI	Normalized Difference Forest Snow Index
	URSI	Universal Ratio Snow Index
	RSI	Ratio Snow Index
	ARSI	Adjust Ratio Snow Index
	RVI	Ratio Vegetation Index
	DVI	Difference Vegetation Index
	FVC	Fraction of forest cover
Location and Geometry	VZA	View zenith angle
	SZA	Solar zenith angle
	RAA	Relative azimuth angle: the absolute value of the difference between sensor and solar azimuth angle
	Latitude	The pixel geographical location information: latitude
	Longitude	The pixel geographical location information: longitude
Auxiliary	LST	Land surface temperature
	DOY	Day of year
	Ab_Vis	Blue-sky Albedo for visible broadband
	Ab_NIR	Blue-sky Albedo for near infrared red broadband
	Ab_SW	Blue-sky Albedo for shortwave broadband

monitoring change and achieving more accurate predictions (Kuter et al., 2022; Reichstein et al., 2019; Wang et al., 2022).

Briefly, there are three common approaches to estimating FSC from MODIS data: (1) the use of a linear empirical relationship between the normalized difference snow index (NDSI) and MODIS FSC based on FSC data generated from a finer-resolution reference snow cover map (Salomonson and Appel, 2004, 2006; Wang et al., 2021). Many linear regression methods have been used for local/ continental areas, and the MODIS standard method (Eq. (1)) is the most representative method for estimating FSC. (2) Linear spectral unmixing assumes that the reflectance of a given pixel is a linear combination of several surface endmembers (e.g., vegetation, snow, soil, and rock) (Metsämäki et al., 2005, 2012; Painter et al., 2003, 2009). Notably, the spectral unmixing snow cover product produced using the MODIS Snow-Covered Area and Grain-Size algorithm (Painter et al., 2009) is superior to the standard MODIS snow cover products (Rittger et al., 2013, 2020a). (3) In addition to methods (1) and (2), an increasing number of studies related to FSC estimation have used machine learning methods in recent decades (e.g., artificial neural networks and random forest methods) (Czyzowska-Wisniewski et al., 2015; Dobreva and Klein, 2011; Kuter, 2021; Kuter et al., 2018). Nevertheless, there are still challenges in the use of retrieval algorithms based on machine learning methods, which are based on single or few bands (or reflectance-based indices) that fail to predict FSC with high accuracy under dense vegetation cover (Czyzowska-Wisniewski et al., 2015; Dobreva and Klein, 2011).

The prediction of FSC with MODIS surface reflectance data has some limitations, owning to several factors such as cloud/snow misclassification (Parajka and Blöschl, 2008; Riggs et al., 2017), off-nadir viewing angles (Arsenault et al., 2014; Dozier et al., 2008), and forest cover (Heinilä et al., 2014; Rittger et al., 2020b), resulting in low accuracy. There have been many studies on FSC estimation that demonstrate the challenge of vegetation cover area. For example, in a complex alpine-forested area, Czyzowska-Wisniewski et al. (2015) developed a FSC estimation algorithm based on machine learning using Landsat images from the continental United States. Metsämäki et al. (2015) also applied the revised SCAmod method, which enables accurate FSC estimation, particularly for forests, to generate GlobSnow Snow Extent products. When there is significant forest coverage, the combined effect of viewing geometry and vegetation cover typically restricts the accuracy of FSC estimation (Rittger et al., 2020b; Xin et al., 2012). To address this, previous studies have applied different FSC adjustment methods to reduce the inaccuracies in the snow-cover area estimation from MODIS introduced by canopy obstruction (Cheng et al., 2021; Kostadinov et al., 2019; Margulis et al., 2019; Raleigh et al., 2013; Rittger et al., 2013, 2020b). For example, Margulis et al. (2019) accounted for the exposed rock fraction and forest cover to predict the surface FSC of mid-latitude montane areas using MODIS-derived FSC products, while Raleigh et al. (2013) applied the canopy correction method to adjust for the canopy occlusion in satellite-derived snow cover area products using the static FVC (fraction of vegetation cover), Rittger et al. (2020) used simultaneous FVC observations to reduce the bias in satellite-derived snow cover area products and mitigate the related issue of satellite observation angle changes in forested areas.

This study aimed to develop an enhanced FSC estimation algorithm based on a machine learning method for complex and heterogeneous vegetated cover environments while accounting for viewing geometry and forest cover. Such an algorithm is required to accurately estimate FSC and must show robust performance for various land-cover types. To address these needs, this study applied an advanced machine learning method and data mining technique to integrate multisource remote sensing information and other auxiliary data to obtain continental-scale FSC maps for complex surface environments. Section 2 describes the MODIS series products and Landsat data used as well as the predictor variables. Section 3 introduces the experimental design of the FSC retrieval algorithm along with the validation metrics. The validation results are presented in Section 4, and Section 5 discusses the influences of the predictor variables, the Landsat-based reference FSC, and snow-cover mapping in forested areas. Finally, our conclusions are provided in Section 6.

2. Satellite datasets and their preparation

2.1. MODIS datasets

This study acquired 28 MODIS–Landsat image pairs from October 2015 to May 2016 across North America. MOD09GA V6 provides atmospherically corrected surface reflectance data (500 m) in bands 1–7 and contains sensor property information (view and solar zenith angles, view and solar azimuth angles; 1 km). As snow cover is strongly affected by land cover type (Rittger et al., 2020b), this study used MODIS land cover classification data (MCD12Q1, 500 m) with the International Geosphere-Biosphere Program (IGBP) classification scheme (Friedl et al., 2010), and the MOD44B Vegetation Continuous Fields product that provides quantitative information on vegetation cover. MODIS series data under clear sky conditions, including LST (Land surface temperature) from MOD11A1 and MCD43A3 albedo data, were also collected as auxiliary information. MCD43A3 produces daily black-sky albedo and white-sky albedo data at local solar noon for seven bands and the visible, near infrared (NIR), and shortwave bands at 500 m resolution. Blue-sky albedo was calculated as a weighted combination of black-sky albedo and white-sky albedo (He et al., 2018). MOD10A1 Normalized Difference Snow Index (NDSI) products were chosen as reference data for comparison with the derived FSC maps (Section 4.3). MOD09GA, MCD12Q1, MOD44B, MOD11A1, and MCD43A3 (Dimiceli et al., 2015; Lucht et al., 2000; Vermote and Wolfe, 2015; Wan and Dozier, 1996) are freely accessible from the Land Processes Distributed Active Archive Center (https://lpdaac.usgs.gov), and MOD10A1 is available from the National Snow and Ice Data Center website (https://nsidc.org/data/MOD10A1). The MOD10A1 V6 snow cover

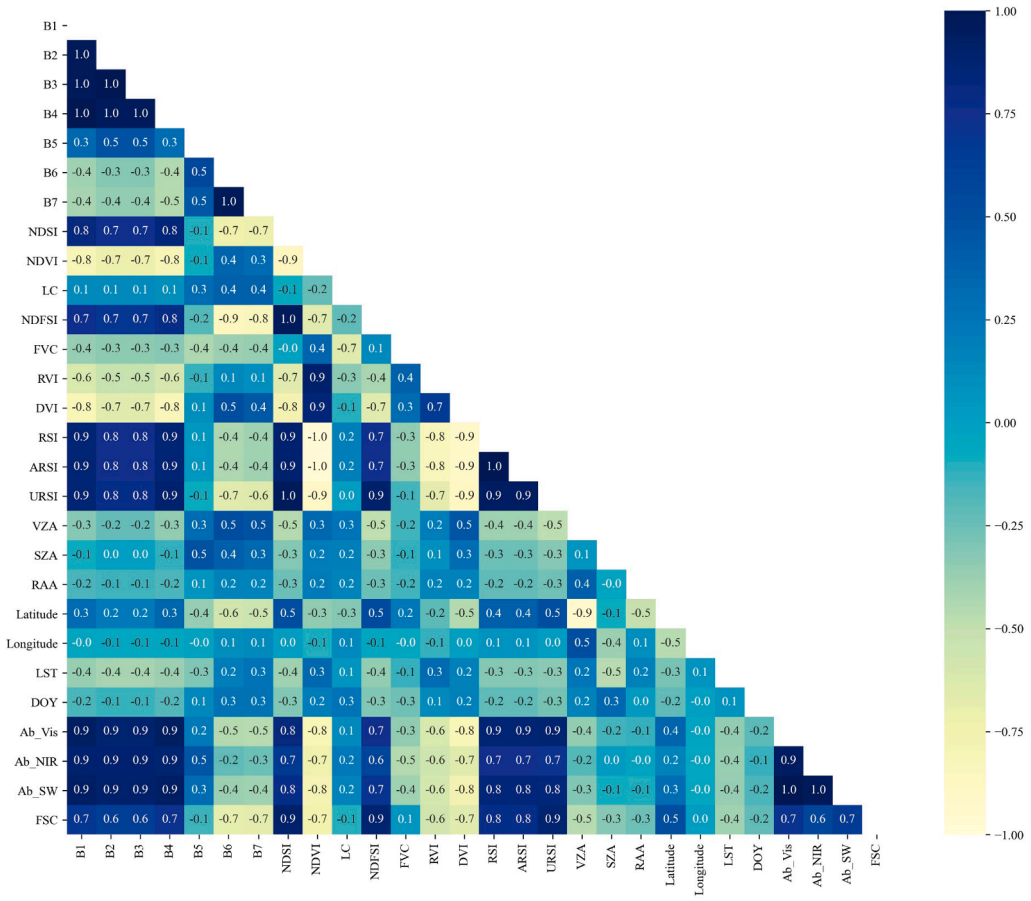


Fig. 3. The Cross-correlation coefficient between the possible predictor variables using all training scene pixels (cf. Table 2). Grids are colored as per correlation coefficients.

product only provides NDSI data with a value range of 0–100 and non-snow classes (>100) (Riggs and Hall, 2015). The MODIS snow product user guide (Riggs and Hall, 2015) suggests using Eq. (1), here referred to as MOD10A1 FSC, to calculate FSC from NDSI data (Salomonson and Appel, 2006).

$$\mathrm{FSC} = -0.01 + 1.45 * \mathrm{NDSI} \ (0 \leq \mathrm{NDSI} \leq 1) \quad (1)$$

If FSC < 0, then FSC = 0; If FSC > 1, then FSC = 1.

The series of MODIS products used (Table 1) were re-projected to the same projection, and resampled to 500 m resolution using the nearest neighbor method. The land cover classes of the MCD12Q1 product were regrouped from 17 to the following nine classes: evergreen forest, deciduous forest, mixed forest, shrub, savannas, grasslands, cropland, bare land, and water (Appendix Table A and Fig. 2). The reflectance data of a given pixel in MOD09GA were excluded if the surface reflectance provided was an invalid value or if it rendered the reflectance-based indices invalid and out-of-range. Additionally, we discarded any pixels with clouds or shadows diagnosed in Landsat or MOD35 cloud masks, as well as water bodies identified by MCD12Q1.

2.2. Landsat-8 Operational Land Imager (OLI) images and reference FSC

Landsat observations have often been used as reference baseline snow observations (Kuter, 2021; Kuter et al., 2018). Landsat satellites are nadir viewing; thus, there are fewer limitations when facing rugged terrain or densely forested regions (Cortés et al., 2014). The United States Geological Survey (USGS, https://earthexplorer.usgs.gov) delivers Landsat-8 Operational Land Imager (OLI) land surface reflectance data. Almost an entire snow season including fresh snow, aged snow, and melting snow, with different land cover types distributed across multiple latitudes, was covered by the 28 Landsat scenes (23 for training and five for validation) selected for this study (Fig. 2a and Appendix Table B). The selected Landsat scenes had minimal cloud cover (<1%) and were in the Level 1 T (L1T) format, meaning that their geo-registration precision was high (Wulder et al., 2012).

First, Landsat-8 images for snow cover were determined in binary using the SNOMAP algorithm developed by Hall et al. (1995) (Fig. 2b), which has been widely used in FSC retrieval studies (Dobreva and Klein, 2011; Kuter et al., 2018; Salomonson and Appel, 2006). To avoid geo-location uncertainties, reference 500 m resolution FSC maps (cf. Fig. 2c) were produced by aggregating the Landsat-8 binary maps in a 750 m radius circle at the center of the MODIS pixel (Fig. 2b) (Dobreva and Klein, 2011; Painter et al., 2009; Wolfe, 2006; Wu et al., 2022). During the aggregation of reference FSC maps, quality control was used to screen out the cloud pixels (including clouds and cloud shadows), which were determined using the Fmask algorithm (Qiu et al., 2019). Notably, this study converted FSC to a range of 0–1. Section 5.2 further discusses the influence of SNOMAP accuracy on the FSC estimation from the MODIS surface reflectance data.

2.3. FSC-related variables

To date, numerous reflectance-based indices have been successfully applied to monitor land surface elements such as snow cover, vegetation, and water bodies. This study selected the following reflectance-based indices as predictor variables for FSC retrieval: NDSI (Eq. (2)), NDVI (Normalized difference vegetation index, Eq. (3)), NDFSI (Normalized Difference Forest Snow Index, Eq. (4)), URSI (Universal Ratio Snow Index, Eq. (5)), RSI (Ratio Snow Index, Eq. (6)), ARSI (Adjust Ratio Snow Index, Eq. (7)), RVI (Ratio Vegetation Index, Eq. (8)), and

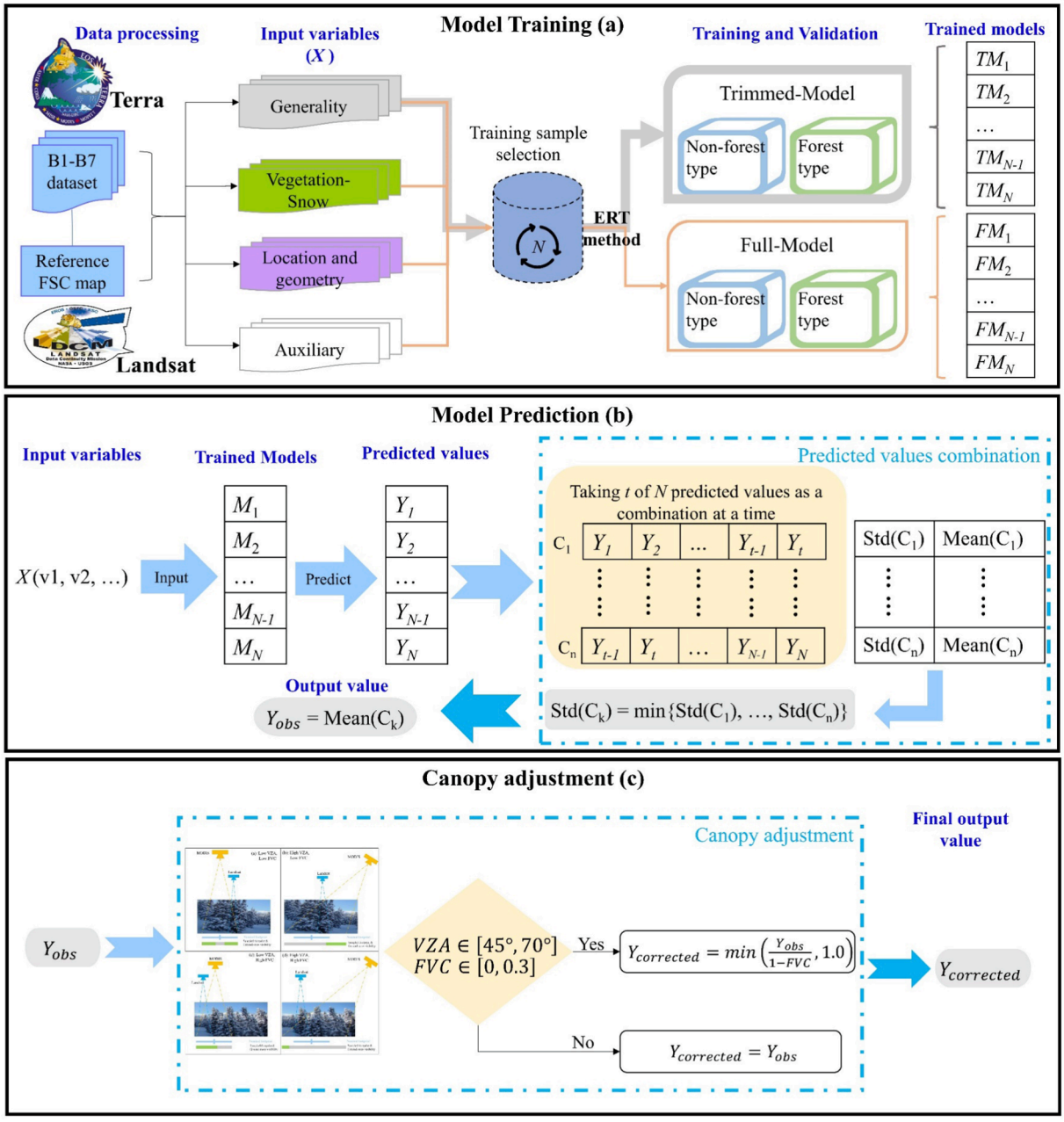


Fig. 4. The schematic diagram of model training (a), model prediction (b), and canopy adjustment (c) for the FSC retrieval. *N* and *t* was set to 20 and 11, respectively.

DVI (Difference Vegetation Index, Eq. (9)). In addition to the eight reflectance-based indices and the reflectance in seven bands, 12 other predictor variables, summarized in Table 2, were chosen for the FSC estimation. To test the hypothesis that other snow-related variables improve the accuracy of FSC retrieval in the presence of frequently used input predictor variables (i.e., the "*Generality*" category), we utilized a machine learning model to estimate FSC with multiple inputs. Fig. 3 presents the cross-correlation coefficient between the predictor variable candidates and the FSC using all the training scene pixels. The predictor variables were divided into the following four categories:

(1) Widely used variables (named as, "*Generality*"). Among the snow-related predictors used in the FSC retrieval based on machine learning methods, B1–B7, NDSI, NDVI, and land cover (LC) are the most common variables (Czyzowska-Wisniewski et al., 2015; Dobreva and Klein, 2011; Kuter, 2021; Kuter et al., 2018; Moosavi et al., 2014). The reflectance of the seven bands was used to describe the variation in snow properties (Czyzowska-Wisniewski et al., 2015; Dobreva and Klein, 2011). The NDSI, developed by Hall et al. (1995) is a crucial snow-related variable used to identify snow. The NDVI is also used to assist in the determination and/or identification of snow cover (Dobreva and Klein, 2011; Lv and Pomeroy, 2019). Additionally, Dobreva and Klein (2011) point out that LC as an input variable helps improve FSC estimation.

(2) Vegetation-snow related candidate variables (named as, "*Vegetation-Snow*"), such as NDFSI, URSI, RSI, ARSI, RVI, DVI, and FVC, were included. Previous studies have found that the NDSI have limited capability to detect snow cover in forested areas (Kostadinov et al., 2019; Lv and Pomeroy, 2019; Raleigh et al., 2013; Rittger et al., 2020b; Wang et al., 2018), while the NDFSI developed by Wang et al. (2018) has a better capability in this

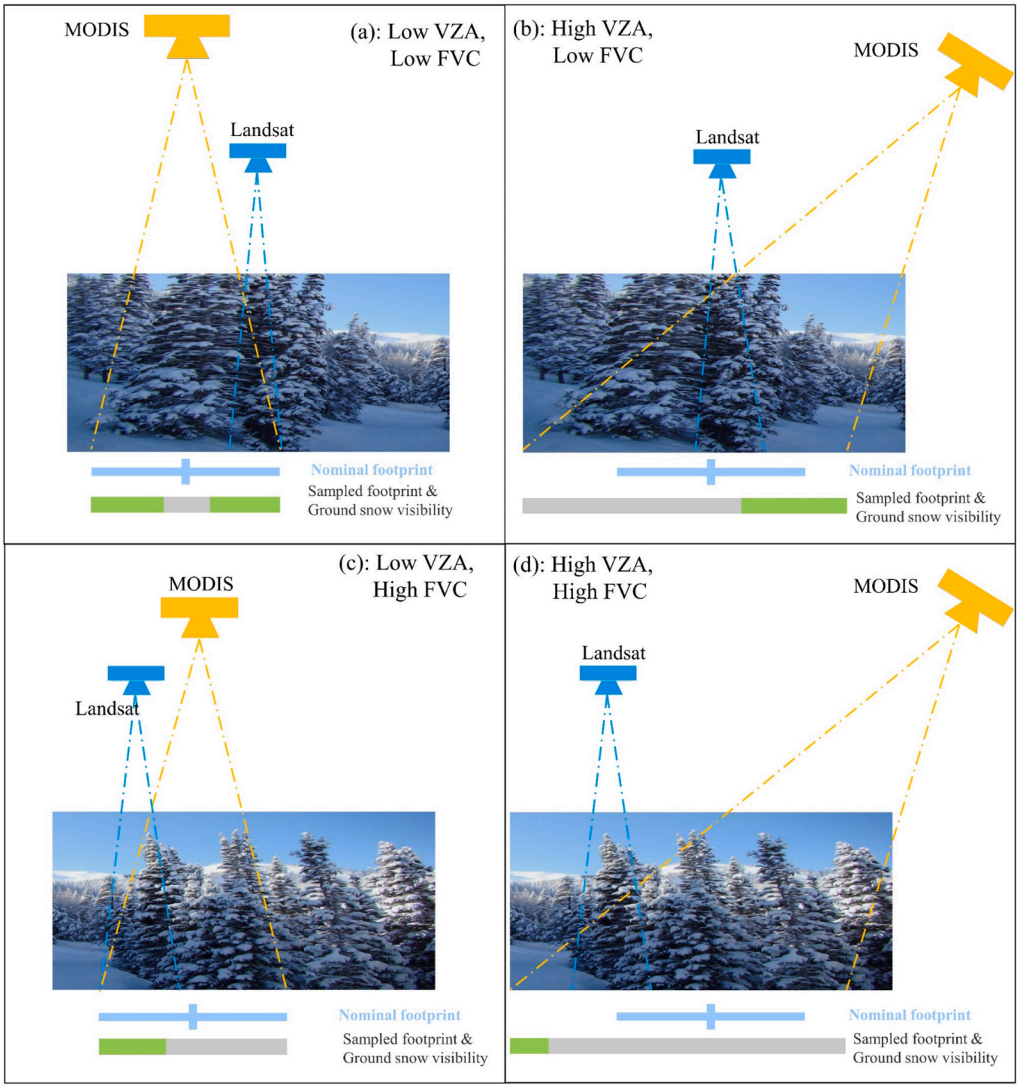


Fig. 5. The viewing geometry effects of MODIS caused by the varying of VZA and FVC on FSC estimation within MODIS sampled footprint. The green and gray rectangle represent ground snow and forest respectively observed by the MODIS sensor within its sampled footprint. The blue lines indicate MODIS pixel center and its nominal footprint (spatial resolution: ~500 m). a) vs b) and c) vs d) illustrative impacts of viewing geometry on sampled footprint and ground snow visibility at low (left) and high (right) VZA. a) vs c) and b) vs d) illustrative impacts of FVC on ground snow visibility at low (top) and high (bottom) FVC. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

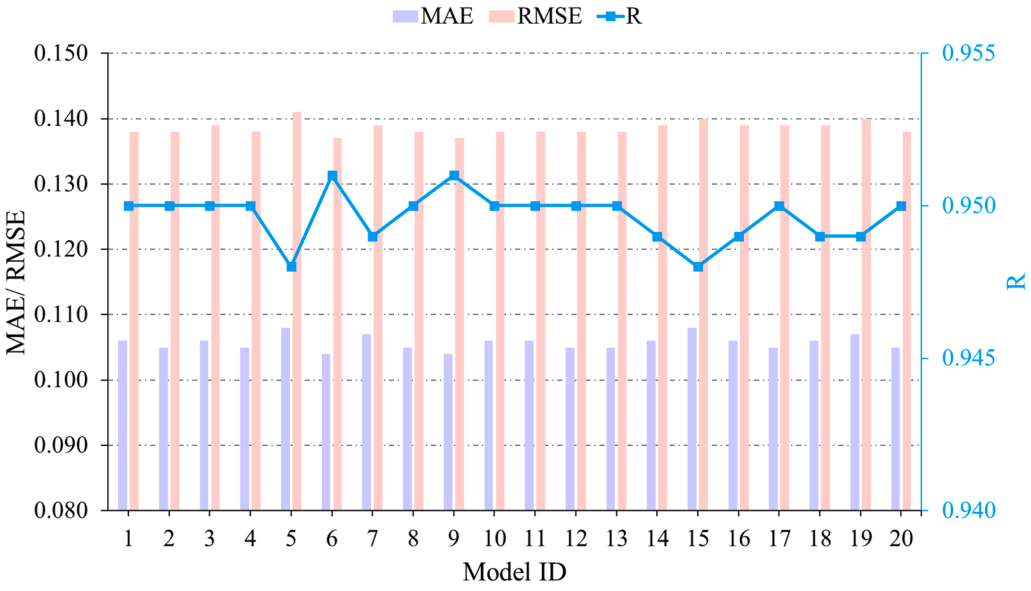


Fig. 6. Accuracy of 20 FSC retrieval sub-models (Full-Model) on the combined validation samples.

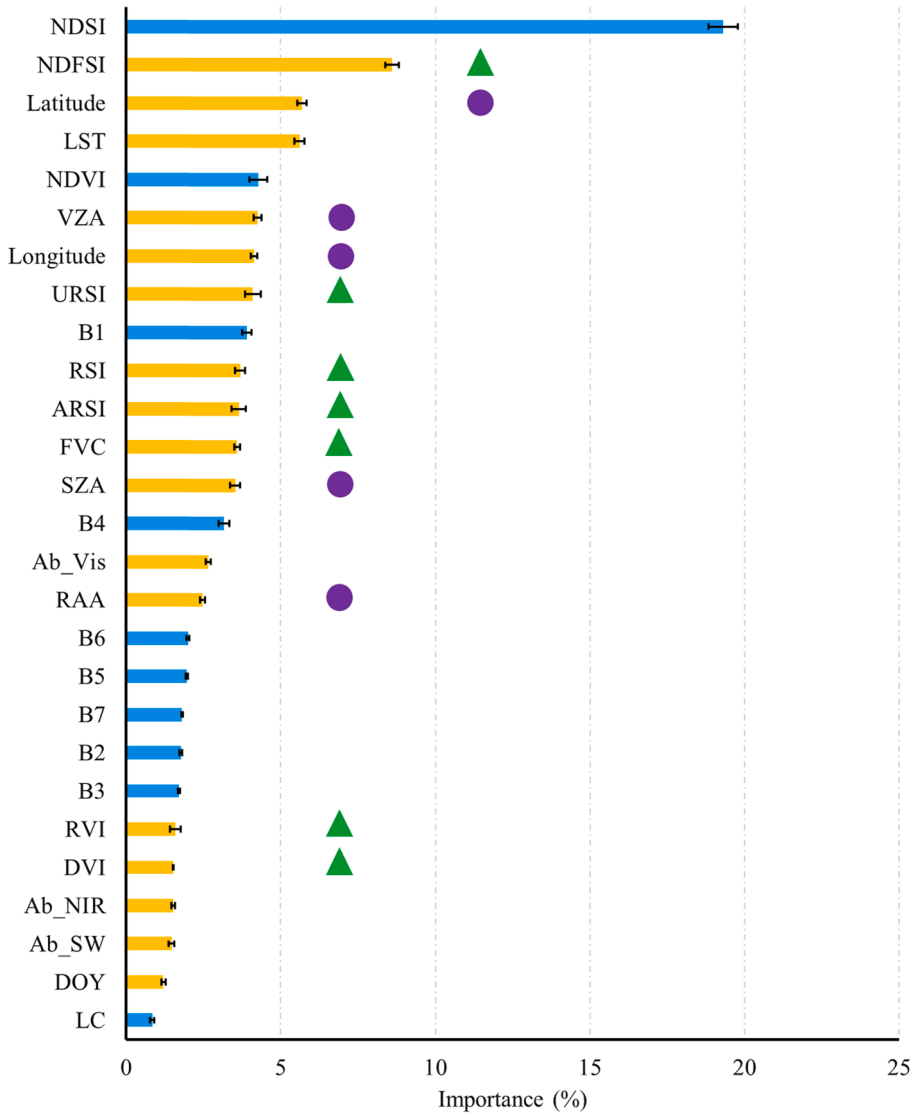


Fig. 7. The importance of predictor variables of forest-type for Full-Model. Error bars (black line) represents the standard deviation on each bar of mean. Table 2 shows the variable name abbreviations. Triangles (green) and circles (purple) indicate the predictor variable of the vegetation-snow category, the location and geometry category in Table 2, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

regard (Lv and Pomeroy, 2019). Our results using satellite observation data illustrate that NDFSI has a strong relationship with FSC (R = 0.9; Fig. 3). The URSI proposed by Wang et al. (2021) showed better performance in FSC estimation compared to the NDSI for forested areas; and the URSI and FSC are also strongly correlated (R = 0.9; Fig. 3). The FVC variable was obtained from the MOD44B tree cover percentage data. Snow has a higher reflectivity value in the red band (B1) and a low reflectivity value in the near-infrared band (B2), whereas the opposite is true for vegetation (Fig. 1), e.g., low reflectivity in the B1 and higher reflectivity in the B2 band. Thus, the RVI, DVI, and FVC were included as predictors to reduce the impacts of vegetation. Inspired by the development of the RVI, the RSI and ARSI were designed to be proportional to B1 and B2 when describing the variation of snow cover. The relationship between the RSI (or the ARSI) and FSC may be spatiotemporally variable ($R_{RSI-FSC}$: 0.8; $R_{ARSI-FSC}$: 0.8; Fig. 3).

(3) Location and geometry related variables (named as, "*Location and Geometry*") include latitude, longitude, VZA (view zenith angle), SZA (solar zenith angle), and RAA (relative azimuth angle). The distribution of snow cover depends on the location, the pixel geographical location information, latitude, and longitude, which can be used as inputs for snow cover monitoring (Yang et al., 2019). Interception or obstruction of the vegetation canopy in optical satellite observations affects FSC estimates, producing large uncertainties and errors (Liu et al., 2008; Rittger et al., 2020b; Xin et al., 2012). It is likely that additional angle-related variables (VZA, SZA, and RAA) are required to assist in FSC retrieval under vegetation cover.

(4) The auxiliary variables (named as, "*Auxiliary*") used were LST, DOY (day of year), Ab_Vis (blue-sky albedo for visible broadband), Ab_NIR (blue-sky albedo for near-infrared red broadband), and Ab_SW (blue-sky albedo for shortwave broadband). Surface temperature is critical auxiliary data for reducing snow cover commission errors in MODIS snow cover products (Riggs et al., 2017) and significantly improves FSC estimation (Liang et al., 2017). Snow age influences the spectral characteristics of the visible and infrared bands (Dietz et al., 2012; Riggs et al., 2017); thus, this study used the DOY as temporal information (Table 2). Previous studies have reported that many factors including snow pollution (e.g., dust and black carbon) (Hao et al., 2013), snow physical properties (e.g., grain size and snow water equivalent),

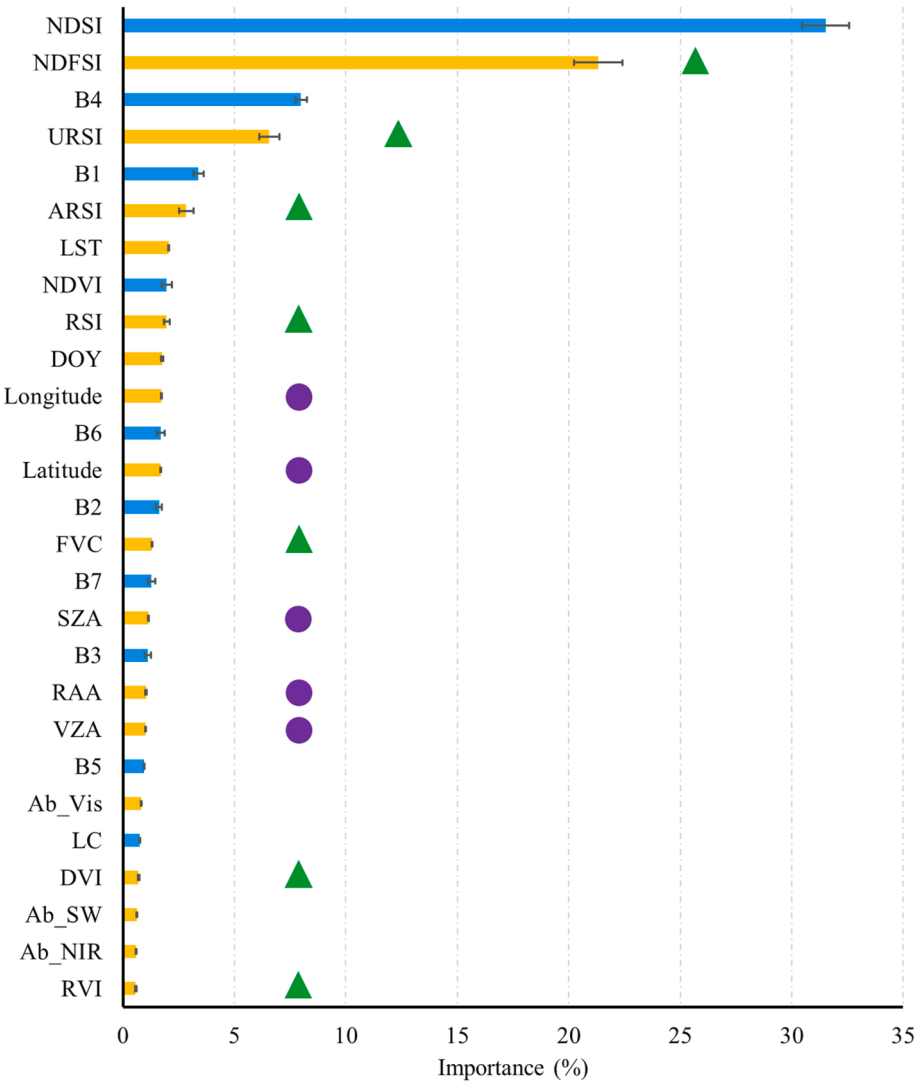


Fig. 8. The importance of predictor variables of non-forest-type for Full-Model. Error bars (black line) represents the standard deviation on each bar of mean. Table 2 shows the variable name abbreviations. Triangles (green) and circles (purple) indicate the predictor variable of the vegetation-snow category, the location and geometry category in Table 2, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Model validation on combined validation dataset: Trimmed-Model, Full-Model. Boldface lines indicate that the corresponding model has the best predictive performance in the validation scenes. The number in parentheses donates the reduction of RMSE.

Scenes ID	Mode name	*R*	MAE	RMSE
Combined	Trimmed-Model	0.937	0.114	0.154
	Full-Model	**0.951**	**0.105**	**0.137 (−10 %)**
Scene A	Trimmed-Model	0.968	0.033	0.077
	Full-Model	**0.974**	**0.039**	**0.069 (−10 %)**
Scene B	Trimmed-Model	0.930	0.087	0.144
	Full-Model	**0.950**	**0.073**	**0.122 (−15 %)**
Scene C	Trimmed-Model	0.837	0.161	0.184
	Full-Model	**0.886**	**0.137**	**0.155 (−16 %)**
Scene D	Trimmed-Model	0.911	0.131	0.169
	Full-Model	**0.915**	**0.139**	**0.165 (−2%)**
Scene E	Trimmed-Model	0.882	0.108	0.138
	Full-Model	**0.912**	**0.094**	**0.120 (−13 %)**

and bidirectional reflectance distribution function (BRDF), cause significant differences in snow reflectance. Therefore, we followed the suggestion of (Metsämäki et al., 2015) by considering three albedo variables (i.e., Ab_Vis, Ab_NIR, and Ab_SW).

$$NDSI = (B4 - B6)/(B4 + B6) \quad (2)$$

$$NDVI = (B2 - B1)/(B2 + B1) \quad (3)$$

$$NDFSI = (B2 - B6)/(B2 + B6) \quad (4)$$

$$URSI = B4/(B2 + B6) \quad (5)$$

$$RSI = B1/B2 \quad (6)$$

$$ARSI = (B1 - B2)/B2 \quad (7)$$

$$RVI = B2/B1 \quad (8)$$

$$DVI = B1 - B2 \quad (9)$$

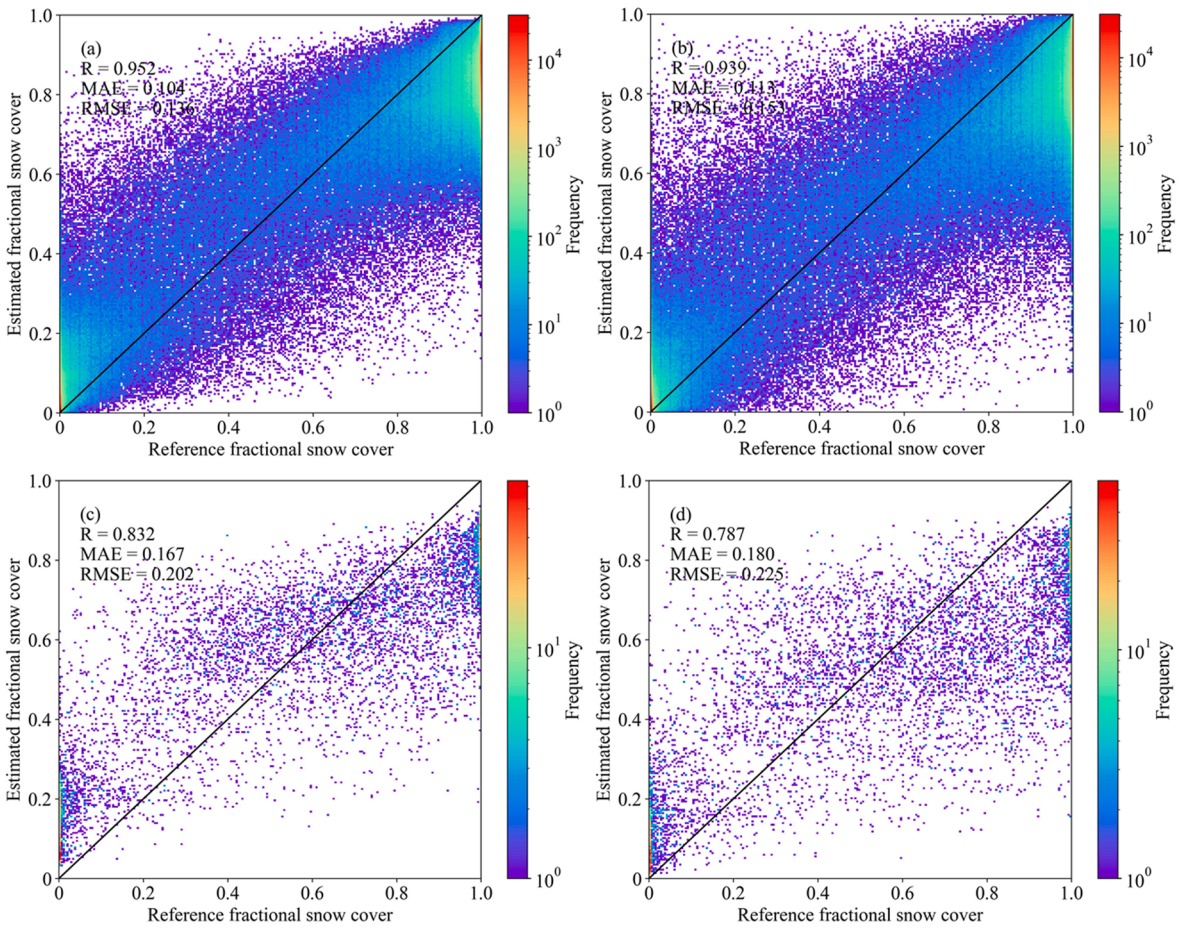


Fig. 9. The scatter density plot for two FSC retrieval models: Full-Model (left) and Trimmed-Model (right) on combined validation data. Non-forest type: (a – b); Forest type: (c – d).

3. Methodology

3.1. Extremely randomized trees (ERT) method

The extremely randomized trees (ERT) method, which is an ensemble learning method (Geurts et al., 2006), was selected to establish the nonlinear relationship between FSC and the selected predictor variables. Unlike the random forest method, the ERT method randomly chooses split nodes and inputs the entire training sample to build tree models (Geurts et al., 2006). All the samples are feathers randomly selected by the ERT method, which may induce the mining of more information on variables and provides better results than those obtained by the random forest method. An advantage of this approach is that the over-fitting effect is negligible. We implemented the ERT method with default parameters using the Python program editor; the number of trees was 100, the number of features was sqrt(n) (where n is the number of input variables), and the minimum number of samples per split was two (Geurts et al., 2006). The importance measures of the input predictor variables was described by the Gini index (GI) (Geurts et al., 2006; Jiang et al., 2009) to aid the interpretation of model responses to the different inputs.

3.2. Training sample selection

The performance of a machine learning method is closely related to the quality of the training sample (Millard and Richardson, 2015; Xiao et al., 2018, 2021). Importantly, the training sample should adequately represent different land cover types and different FSC values to increase the model generality and inversion accuracy. We adopted a stratified random sampling strategy to select samples based on FSC values with an interval of 0.1 and land cover type. The final 379,522 training samples (72,038 for the forest-type and 307,484 for the non-forest type) represented a randomly selected sample from each training scene (see Appendix Table B). Figure C in the Appendix presents the distribution of the target and predictor variables in the training sample (cf. Appendix Fig. D for the validation data). In addition, to test and mitigate the impact of the training sample quality, *N* paired training, and testing sets were selected using the previously described sampling strategies (Fig. 4a). Before training the Full-Model, the training sample sets were divided into forest- and non-forest types based on land cover type (Appendix Table A).

3.3. Design of the FSC retrieval model

In the model training phase, to mitigate the dependence on training samples and construct a robust FSC retrieval model, we propose an approach to obtain *N* retrieval models by repeated training with the ERT method based on *N* different training samples (Fig. 4a). The Full-Model was then established by repeating this procedure for both the forest-type and non-forest-type training samples. In the prediction phase (Fig. 4b), the predictor variables $X(v_1, v_2, v_3, \cdots\cdots)$ were input into *N* retrieval models $(M_1, M_2, \cdots\cdots, M_{N-1}, M_N)$, and the predicted values of the target variables $(Y_1, Y_2, \cdots\cdots, Y_{N-1}, Y_N)$ were generated. Then, the combinations of *t* predicted values taken from the *N* predicted values were enumerated $(C_1, C_2, \cdots\cdots, C_n)$. Correspondingly, we obtained the standard deviation and mean of the predicted value combinations $Std(C_1), Std(C_2), \cdots\cdots$, $Std(C_n)$ and $Mean(C_1), Mean(C_2), \cdots\cdots, Mean(C_n)$, respectively. Subsequently, we obtained the minimum standard deviation ($std(C_k)$ =

Fig. 10. The RMSE vary with land cover types and FVC for two FSC retrieval models: Trimmed-Model and Full-Model (cf. Table 3).

$\min\{Std(C_1), Std(C_2), \cdots\cdots, Std(C_n)\}$), and the final predicted value (Y) was obtained by averaging the corresponding predicted-value combinations (i.e., $Mean(C_k)$) (Fig. 4b). When training the N FSC retrieval models, the importance measures of each predictor variable were calculated using all the predictor variables (Figs. 7 and 8). For simplicity, the experimental tests had N and t set to 20 and 11, respectively. This meant that the minimum majority of the predicted values was selected.

3.4. Canopy adjustment for the combined effect of VZA-FVC

The impacts of the variations in FVC and VZA on ground snow visibility to the MODIS sensor are illustrated in Fig. 5. As VZA increases, the sampled footprint of each pixel elongates, and less ground snow is visible under dense vegetation cover. For a larger VZA, the pixels in the cross-track direction are stretched even worse (Dozier et al., 2008; Margulis et al., 2019), making the FSC estimation more complex, which is not considered further here. As the forest canopy can completely or partially obscure ground snow, one solution is to adjust the viewable snow cover by using the FVC to account for areas known to be covered by forests (Rittger et al., 2020b). Canopy adjustments are often based on the assumption that snow under a forest canopy has the same proportions as in the viewable gap; it is assumed that the fraction of the viewable gap is equal to 1 − FVC, such that Eq. (10) can be used for forest canopy adjustment (Rittger et al., 2020b). In Section 4.3, experimental tests were designed to analyze the error of the corrected FSC estimates that vary with the FVC and VZA values when using Eq. (10) for the forest canopy adjustment in the validation scenes. Finally, based on the results (Figs. 12 and 13), a canopy adjustment strategy was adopted, as depicted in Fig. 4c. In Eq. (10), $FSC_{corrected}$ is the satellite FSC data corrected for forest canopy occlusion, and FSC_{obs} is satellite-derived FSC.

$$FSC_{corrected} = min\left(\frac{FSC_{obs}}{1 - FVC}, 1.0\right) \quad (10)$$

3.5. Model performance validation

3.5.1. Accuracy metrics

Three common indices were used to evaluate the performance of the FSC retrieval models: Pearson's correlation coefficient (R), mean absolute error (MAE), and root mean square error (RMSE). In addition, six binary metrics (see Table C in the Appendix; Overall Accuracy (OA), Recall, Precision, Specificity, F1-score, and Cohen's kappa coefficient) were used to test the binarized FSC estimates against Landsat FSC using the binarization thresholds FSC = 0.5 (Rittger et al., 2013), as described in Section 4.4.

3.5.2. Effect of additional predictor variables (Trimmed-Model)

To investigate the contribution of the additional variables (*Vegetation-Snow*, and *Location and Geometry*, as well as the *Auxiliary* variables) to the generalization ability of the FSC retrieval model (Table 2), separate training datasets with 10 input variables (B1-B7, NDSI, NDVI, and LC) were created by excluding the additional variables from each training sample dataset obtained in Section 3.2, which were used to train with ERT method, the established model refer to as the Trimmed-Model (Fig. 4a). In addition, the validation dataset with the same input variables were used to evaluate the performance of the Trimmed-Model. The Trimmed-Model was built for both the forest and non-forest type.

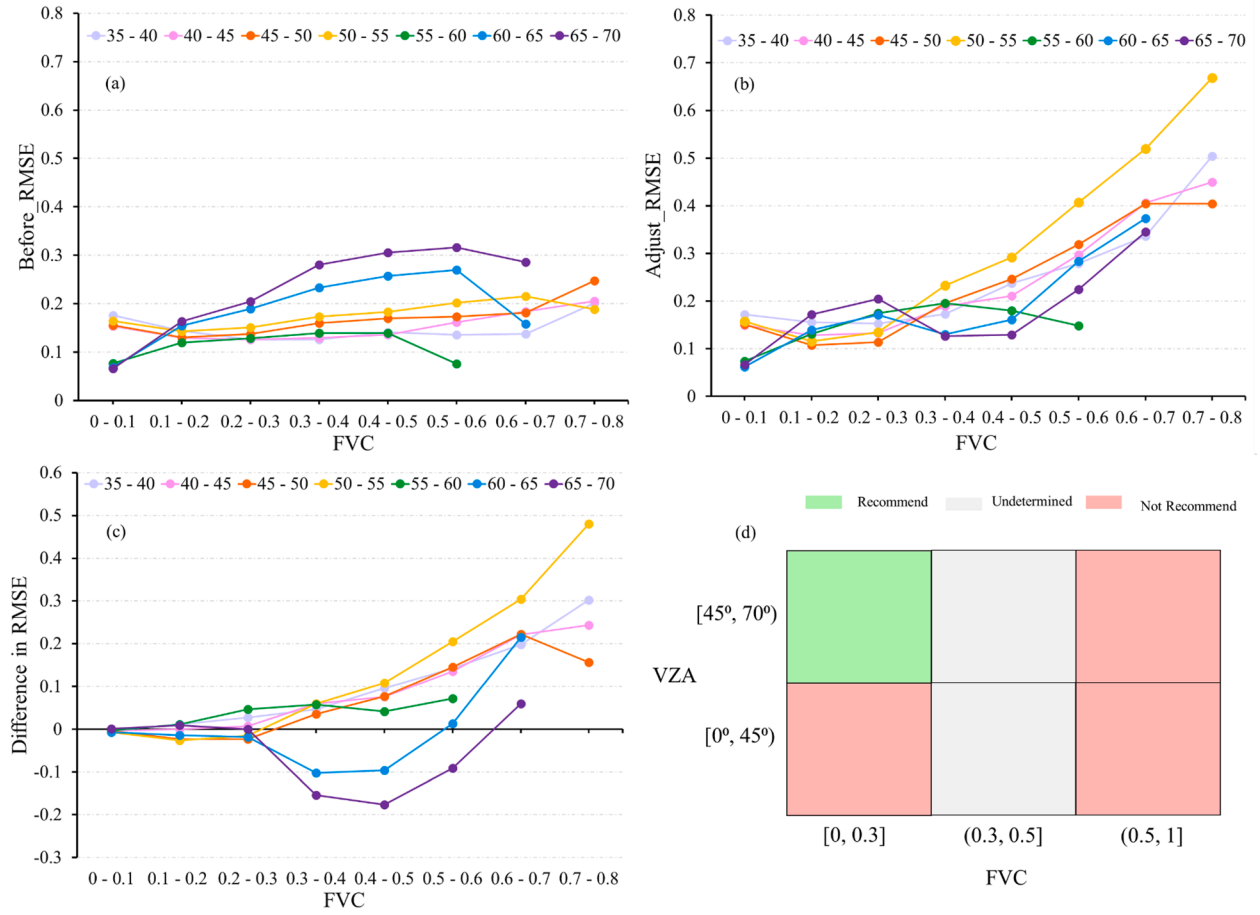


Fig. 11. The VZA and FVC coupled effect in estimating FSC. Difference equates to Adjust_RMSE subtract Before_RMSE. "*Recommend*", the adjustment is recommended in VZA ∈ [45°, 70°) and FVC ∈ [0, 0.3] conditions; "*Not recommend*", it is not recommended in denser forest cover conditions (FVC ∈ (0.5, 1]), and in VZA ∈ [0°, 45°) and FVC ∈ [0, 0.3] conditions; "Undetermined", whether the adjustment is conducted within medium forest coverage to be determined (FVC ∈ [0.3, 0.5]).

3.5.3. Comparison with the MOD10A1 FSC map

We compared our retrieval results with the MOD10A1 FSC map acquired on the same date for independent validation scenes. The MOD10A1 FSC was derived using the MOD10A1 NDSI and Eq. (1) (see section 2.1). Each metric was then computed with respect to the Landsat reference FSC and its binarized values.

4. Results

4.1. Analysis of the trained models

4.1.1. Model robustness dependence on the training sample

With regard to the dependence on the training sample, this section assesses whether the 20 trained models have significant statistical differences in FSC retrieval on five validation scenes and the combined validation data using *repeated measures analysis of variance* (RANOVA) (Huson, 2003; Kuter, 2021). The results show that the predictor values of the 20 FSC retrieval models did not differ statistically based on the F-test ($p < 0.01$). This indicated that the training sample selection technique did not lead to a significant difference in the training models. However, the accuracy metrics of the FSC estimates of the 20 retrieval models on the combined validation data (Fig. 6) show that compared to the Landsat reference FSC values, the predicted values did have slight differences, with R in the range 0.948–0.951, MAE in the range 0.104–0.108, and RMSE in the range 0.137–0.141. The accuracy metrics of the 20 FSC retrieval models for the five individual validation datasets are also provided in Appendix Table D. Compared with the 20 models, the Full-model minimizes the uncertainty caused by the training sample features with different noise levels and improves the prediction robustness of the FSC retrievals. This emphasizes the necessity of integrating all of the retrieval models.

4.1.2. Contributions of the predictor variables

Importance measures help understand a model's decision-making process between the input variables and the associated target. Figs. 7 and 8 show the importance of the predictor variables of the forest- and non-forest types models, respectively. The forest-type model identified the NDSI and the NDFSI as the two largest contributors to the FSC estimates; the mean importance of the NDSI was 19.3 % compared to 8.6 % for the NDFSI. The NDFSI has rarely been used for FSC inversion, and, as such, we did not expect this to be the second largest contributor. Similarly, the two most important contributors in the non-forest-type model were the NDSI (31.5 %) and the NDFSI (21.3 %) (Fig. 8). Across 40 runs of the retrieval models (both non-forest and forest types), the NDSI and the NDFSI ranked first and second in terms of importance in all cases. These two predictor variables also showed a strong correlation with FSC (~0.9, Fig. 3). As seen in Fig. 7, the top 10 contributors have only three predictor variables that belong to the '*Generality*' category (i.e., NDSI, NDVI, B1). The newly introduced variables (orange bars) have a higher ranking, while the contribution of the spectral indices (green triangles) was greater than that of the seven spectral bands in both the forest-type and non-forest-type models. This confirms that the spectral indices may enhance the transferability of the established retrieval models (Belgiu and Drăgu, 2016). Compared with the non-forest-type model, the forest-type model is more sensitive to variations in the *Location and Geometry* variables (purple circles) (Heinilä

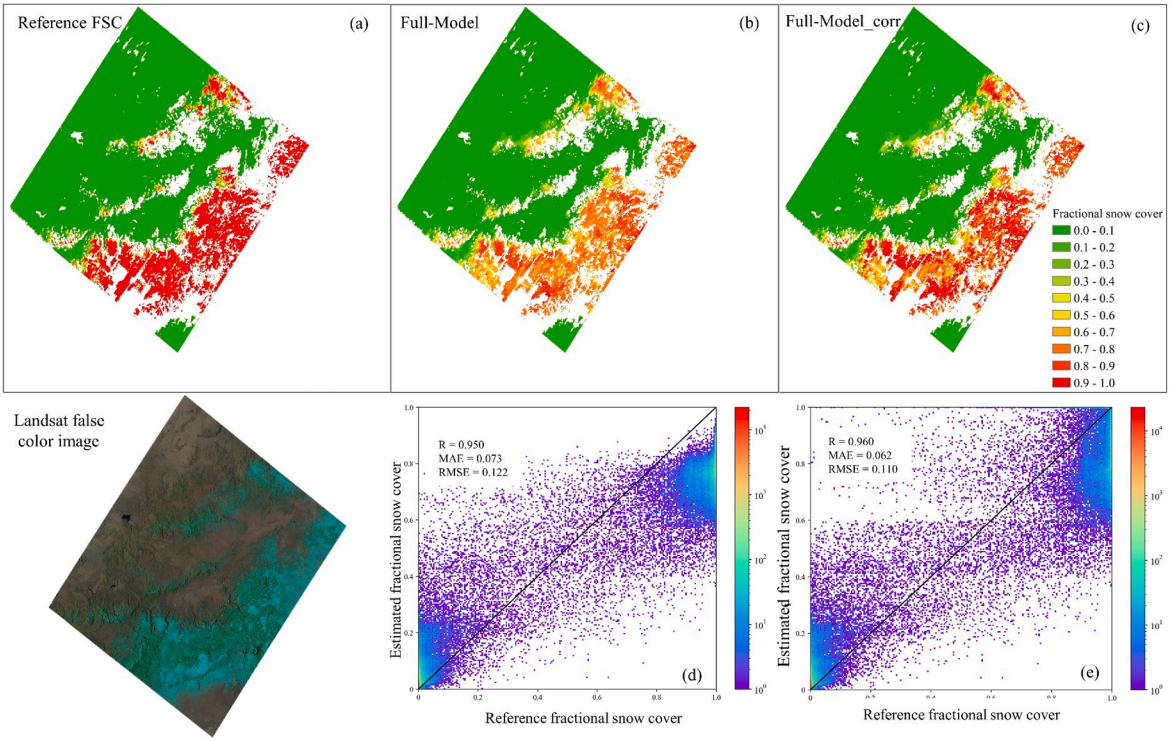


Fig. 12. Landsat false color images, Landsat-derived reference FSC maps and the example of VZA-FVC adjustment in Scene B that represents snow cover in forest regions, where VZA is in the range [59.86°, 65.31°] and FVC is in the range [0, 0.66]. Whiter background in (a) – (c) indicates invalid or no data. (d) and (e) are the scatter density figure for the corresponding (b) and (c) images, respectively. 4 spatial images are in the same projection coordinate system (Lambert azimuthal equal area).

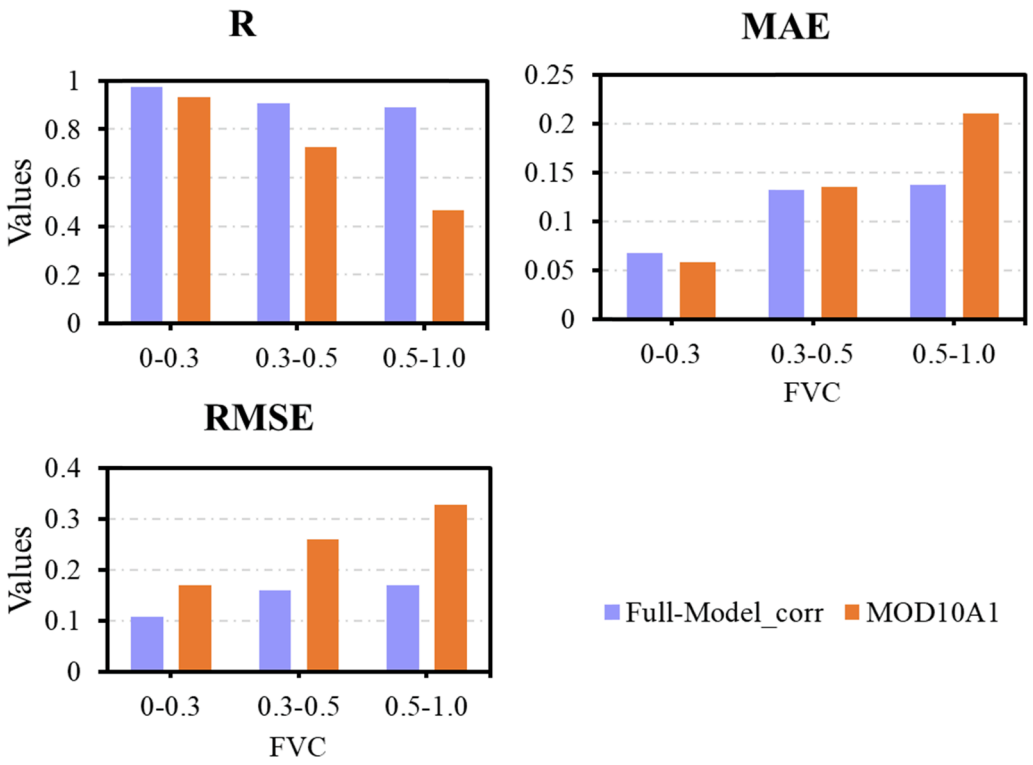


Fig. 13. Comparison of the variation of three accuracy metrics for the Full-Model_corr algorithm and MOD10A1 with the combined validation scene FSC dataset divided based on FVC.

Table 4
Accuracy metrics for the predicted FSC (Before) and the adjusted FSC (After) by the recommend adjustment rules in five independent validation scenes (A–E) (cf. Fig. 12).

		Before (Full-Model)	After (Full-Model_corr)
Scene A	*R*	0.974	0.975
	MAE	0.039	0.037
	RMSE	0.069	0.067
Scene B	*R*	0.950	0.960
	MAE	0.073	0.062
	RMSE	0.122	0.110
Scene C	*R*	0.886	0.911
	MAE	0.137	0.096
	RMSE	0.155	0.138
Scene D	*R*	0.915	0.917
	MAE	0.139	0.136
	RMSE	0.165	0.164
Scene E	*R*	0.912	0.912
	MAE	0.094	0.091
	RMSE	0.120	0.120

Table 5
The comparison of FSC map generated from Full-Model_corr and MOD10A1 in five independent validation scenes (A-E).

	FSC map name	*R*	MAE	RMSE
Scene A	Full-Model_corr	0.971	0.030	0.056
	MOD10A1	0.956	0.017	0.068
Scene B	Full-Model_corr	0.961	0.060	0.107
	MOD10A1	0.879	0.056	0.186
Scene C	Full-Model_corr	0.915	0.092	0.134
	MOD10A1	0.690	0.106	0.241
Scene D	Full-Model_corr	0.928	0.129	0.156
	MOD10A1	0.856	0.087	0.217
Scene E	Full-Model_corr	0.912	0.088	0.115
	MOD10A1	0.714	0.105	0.197

et al., 2014; Rittger et al., 2020b; Xin et al., 2012), revealing that viewing geometry partially accounted for the variation in surface reflectance.

4.2. Model prediction performance

The independent validation dataset was used to analyze the performance of the Trimmed-Model and Full-Model. Table 3 summarizes the accuracy metrics for the five individual scenes and the combined validation data. In general, compared to the Trimmed-Model, the Full-Model seemed to perform better in estimating FSC with an approximate 10 % improvement based on the *R*, MAE, and RMSE metrics (boldface in Table 3). This means that the FSC values predicted by the Full-Model are more consistent with the Landsat reference FSC values. It is notable that the retrieval models tended to have larger errors for areas of forest cover than for those with non-forest cover; the RMSEs of the Full-Model were 0.202 and 0.136 for these two cover types (Fig. 9a vs 9c; left panel), and 0.225 and 0.153 (Fig. 9b vs 9d; right panel) for Trimmed-Model respectively. Among the five validation scenes, the retrieval models performed poorly in Scene C (lowest *R*), which had the highest vegetation coverage. This indicates that forest cover was a large source of error that impeded the accuracy of the FSC estimates. Moreover, the effects of vegetation obscuration and viewing angle, especially in forested areas, may not have been completely removed in these two models.

Compared with the Trimmed-Model, the Full-Model provided the lowest RMSE values (~0.08–0.25) for all the land cover types (Fig. 10). The non-forest type yields a relatively good retrieval accuracy, with an RMSE < 0.16, with the lowest RMSE (~0.08) obtained for cropland. In contrast, for forest cover types (evergreen, deciduous, and mixed forest), both models yielded poor FSC retrievals, with RMSE > 0.18, although the performance of the Full-Model was relatively better. A similar improvement was seen in the sensitivity test (Fig. 10); in regions with less forest cover (FVC < 0.1), the Trimmed-Model and Full-Model had RMSEs of 0.13 and 0.12, respectively. The RMSE values increased with an increase in FVC ($0.1 \leq$ FVC ≤ 0.6), although the performance of the Full-Model was comparatively better. In contrast, the Full-Model did not show a significant improvement in regions with dense forest cover (FVC > 0.6).

4.3. Canopy-adjusted FSC

As illustrated in Figs. 10 and 11a, the proposed FSC retrieval model (Full-Model) did not completely remove the coupled impact of VZA and FVC in the forest cover types. Higher RMSE values are often associated with denser forest cover and larger VZA values. Fig. 11a depicts the variation in RMSE values for the validation data under various VZA and FVC conditions, and Appendix Figs. A and B display the distributions of VZA and FVC in the validation data. Limited the number of samples, results for a VZA $> 60°$ and FVC > 0.5 give slightly different RMSE trends. Subsequently, we applied Eq. (10) to adjust the VZA-FVC effect, and the results (Fig. 11b and 11c) suggest that the errors of all the adjusted FSC values were not obviously reduced.

The application of Eq. (10) may introduce significant uncertainties or errors in the predicted FSC when the land surface has a denser forest cover (FVC > 0.5; Fig. 11b). In contrast, a slight improvement in the estimated FSC was achieved when the FVC was < 0.3 and VZA was $> 45°$. This demonstrates that Eq. (10) alone may not fully eliminate the influence of the VZA-FVC effect when estimating FSC. In the case of medium forest cover ($0.3 <$ FVC < 0.5), there is no clear conclusion regarding the VZA-FVC effect. According to the adjustment results, the operation of Eq. (10) under different VZA and FVC conditions have been summarized in Fig. 11d, where "*Recommend*" indicates the recommended adjustment in the conditions using Eq. (10) (VZA $\in [45^{\circ}, 70^{\circ})$ and FVC $\in [0, 0.3]$]; "*Not recommend*" indicates the adjustment conditions not recommended for dense forest cover conditions (FVC $\in (0.5, 1]$), or in VZA $\in [0^{\circ}, 45^{\circ})$ and FVC $\in [0, 0.3]$; "*Undetermined*" indicates that it remains undetermined whether Eq. (10) should be utilized for medium forest coverage (FVC $\in [0.3, 0.5]$).

After performing the "*Recommend*" adjustments, the error of the FSC estimates was reduced in the five validation scenes (A–E), with higher *R* and lower MAE and RMSE values (Table 4). A visual example of the VZA-FVC adjustment for FSC retrieval is shown in Fig. 12. In this case, the Full-Model performs quite well in predicting the FSC of Scene B ($R = 0.950$, MAE $= 0.073$, RMSE $= 0.122$). Fig. 12b and 12d show large underestimations in reference high FSC values (~1.0), where is vegetation cover region. Even though there was underestimation and overestimation in the adjusted FSC map (Fig. 12c and 12e), the accuracy metrics were significantly improved, with higher *R* (0.960 vs 0.950) and lower MAE (0.062 vs 0.073) and RMSE (0.110 vs 0.122) values, respectively. In most cases, the FSC values were improved at higher values (Fig. 12c and 12e).

4.4. Validation analysis and FSC map comparison

The MOD10A1 V6 FSC maps were compared to those generated by the Full-Model_corr algorithm. Table 5 summarizes the accuracy metrics for the five validation scenes. Generally, the Full-Model_corr showed a better capability for predicting the spatial variability of snow cover than MOD10A1 with reference to the accuracy metrics, with higher *R* and lower MAE and RMSE values (Table 5 and Fig. 13). Overall, there were significant improvements with the Full-Model_corr algorithm compared to MOD10A1 in the combined validation data, with *R* values of 0.961 vs 0.893 and RMSE values of 0.124 vs 0.202, respectively. According to the FVC values, three forest coverage classes (0–0.3, 0.3–0.5, and 0.5–1.0) were used to create three independent validation subsets to examine the

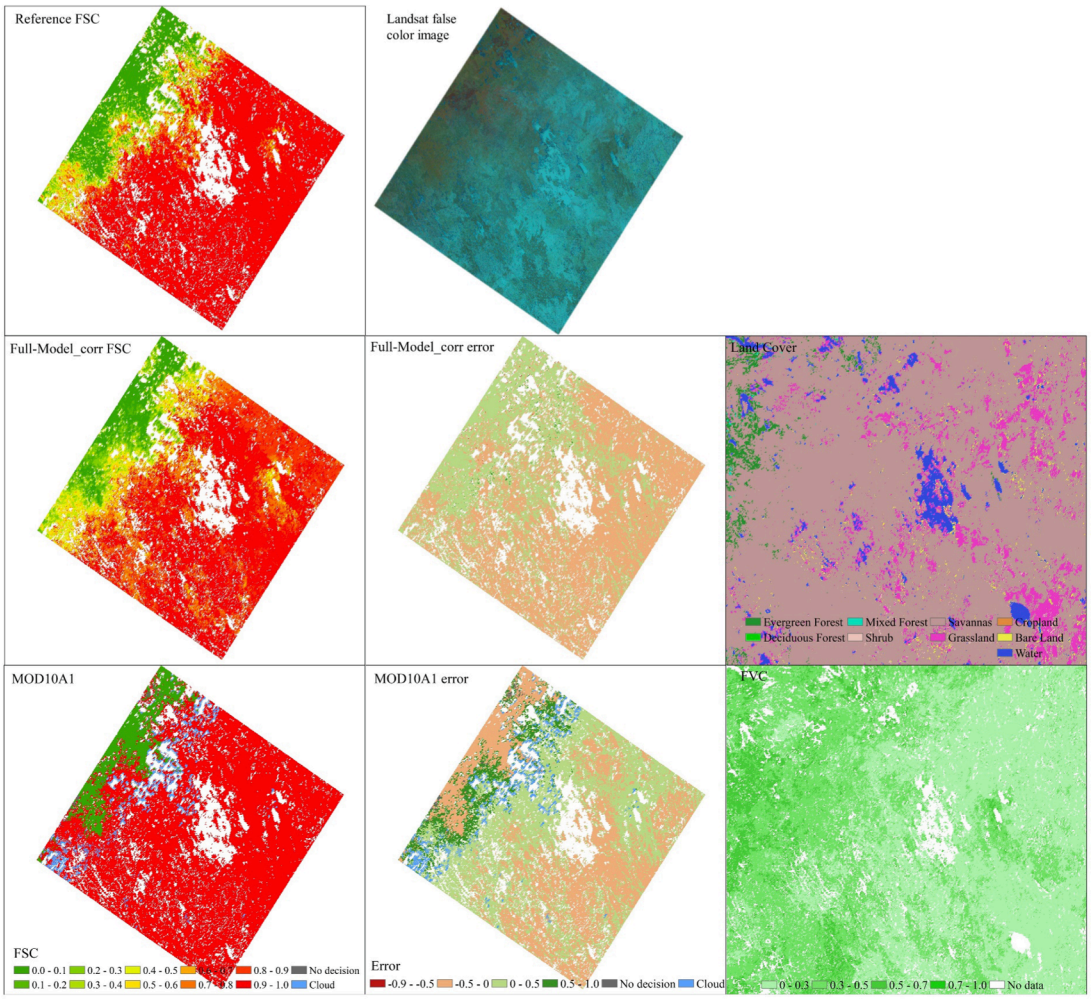


Fig. 14. Landsat false color images, Landsat-derived reference FSC maps, the predicted FSC map comparison in Scene C by Full-Model_corr and MOD10A1 algorithm, and the corresponding error maps (i.e., Error = $FSC_{generated} - FSC_{reference}$). The error values are in the ranges [-1, 1], where the negative and positive error values represent underestimation and overestimation, respectively. White background in five images indicate the invalid or no data. 8 images are in the same projection coordinate system (Lambert azimuthal equal area).

prediction performance of two methods further (Fig. 13). The FSC prediction of the Full-Model_corr algorithm in low FVC (0–0.3) performs best in comparison to the reference FSC, with a RMSE of approximately 0.1. Compared to the MOD10A1 FSC data, the reduction in the RMSE of the FSC estimations for vegetation areas was approximately 40 %.

Further analysis of the two scenes (Scenes B and D) shows that the Full-Model_corr and the MOD10A1 results are quite successful in deriving the spatial distribution of FSC. However, the Full-Model_corr FSC map captures snow cover variability better than the MOD10A1 FSC map, with a better agreement with the Landsat reference FSC data (Figs. 14 and 15, Table 5). Notably, the Full-Model_corr underestimates at the edge of the snow-covered and snow-free areas (i.e., the snowline) in mountainous regions (Fig. 15), while the MOD10A1 FSC map suffers from overestimation at the snowline (error > 0.5; Figs. 14 and 15). The MOD10A1 data also tended to significantly overestimate lower reference FSC values (error > 0.5, corresponding to the edges of more densely vegetated areas) compared to the Full-Model_corr FSC map (Figs. 14 and 15). There is also evidence of the misclassification of snow cover as cloud in the MOD10A1 FSC map due to the inherent cloud-cover algorithm error in the MOD10A1 product (Riggs and Hall, 2015); as observed in previous studies (Kuter, 2021), snow/cloud classification occurs at the edges of the snow cover area in mountainous regions and forest regions with sporadic snow cover.

To further evaluate the improvements achieved by our proposed algorithm, we binarized the FSC data derived from the Full-Model_corr and the MOD10A1 product (500 m resolution) using a threshold of 0.5. Based on the binarization results, the Full-Model_corr FSC data performed better than MOD10A1 FSC data on the five validation scenes (Fig. 16) with higher OA (0.962 vs 0.896) and Kappa coefficients (0.921 vs 0.783). Specifically, there is an almost 8 % improvement in the OA of snow-cover detection using the Full-Model_corr algorithm. When the FVC increases in areas with snow and vegetation, the weaker spectra signal of snow (Fig. 1) and the effects of viewing geometry led to inaccurate FSC predictions and greater commission errors (i.e., lower *Precision;* Fig. 17). According to the results shown in Figs. 13 and 17, FSC prediction in densely vegetated areas was significantly improved using the Full-Model_corr.

5. Discussion

5.1. Predictor variables

The MODIS snow cover product V6 provides NDSI data with a range of 0 to 100 (Riggs and Hall, 2015) and neglects the cases where the NDSI < 0 (Huang et al., 2018), whereas Xin et al. (2012) found that the NDSI of snow cover pixels was negative in forested areas. This implies that using NDSI > 0 to determine snow tends to considerably underestimate the snow cover area because it excludes those areas corresponding to

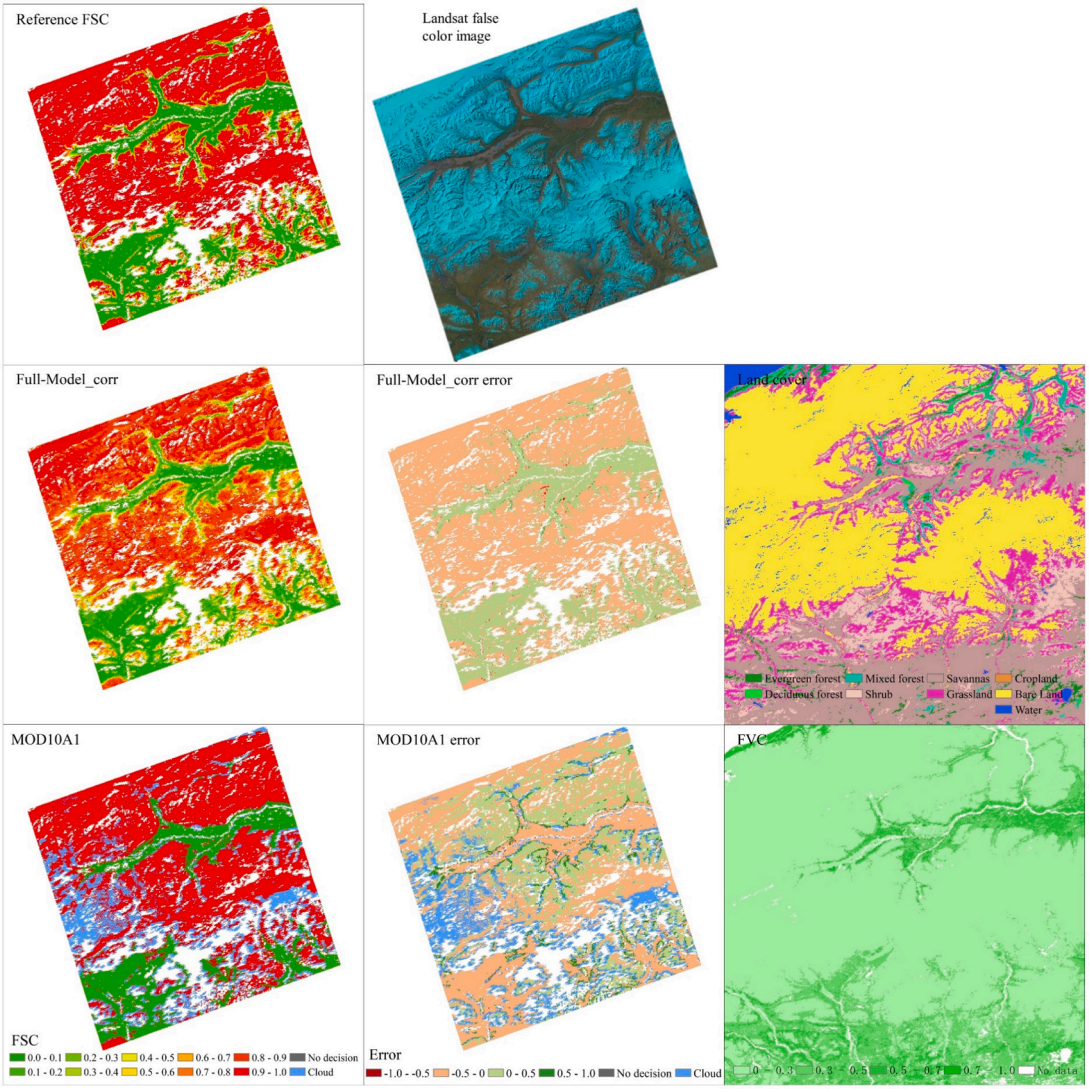


Fig. 15. Landsat false color images, Landsat-derived reference FSC maps, the predicted FSC map comparison in Scene D by Full-Model_corr and MOD10A1 algorithm, and the corresponding error maps (i.e., Error = $FSC_{generated} - FSC_{reference}$). The error values are in the ranges [-1, 1], where the negative and positive error values represent underestimation and overestimation, respectively. White background in five images indicate the invalid or no data. 8 images are in the same projection coordinate system (Lambert azimuthal equal area).

NDSI < 0. Therefore, this study extended the NDSI range for FSC estimation to less than 0. As shown in Fig. 1, the RSI and ARSI have the potential to distinguish snow-related signals in non- or low-vegetation cover areas because their spectral responses distinguish snow from vegetation (i.e., trees and grass). However, the RSI (and the ARSI) may have no distinctly unique advantages for densely vegetated areas (Fig. 1). For the majority of mixed pixels, the diversity of variable properties relating to surface-covered end-members makes it challenging to apply the snow cover identification strategy extensively. Additionally, the contributions of the 27 predictor variables in the non-forest-type model show completely different rank orders to those of the forest-type model (Figs. 7 and 8). This is probably because the spectral features of snowpack vary with land cover type and spatiotemporal conditions.

For model input variable selection, several studies have used feature selection methods based on importance measurements to avoid time-intensive procedures (Belgiu and Drăgu, 2016; Xiao et al., 2021; Yang et al., 2021). The ranking of variable importance of machine learning methods is highly dependent on the training sample as well as the method used. For example, using multivariate adaptive regression splines (MARS) and random forest, Kuter (2021) showed that completely different importance ranks when inputting the same variables. Therefore, model variable feature selection methods may be arbitrary, and the transferability of the optimized variable features and the associated models may be limited when applying to a new study area. The predictor variables that are most suitable for FSC estimation need to be mor thoroughly investigated, and the importance ranking of variables in one study area should not be directly adopted in another area without prior checking. While our study focused specifically on vegetation, it should be note that terrain-related variables also have significant impacts on snow distribution (Czyzowska-Wisniewski et al., 2015).

5.2. Limitations of the Landsat-based reference FSC

It is well-established that Landsat snow cover observations are relatively more accurate than those derived from MODIS data (Salomonson and Appel, 2004). Consequently, Landsat-derived reference FSC

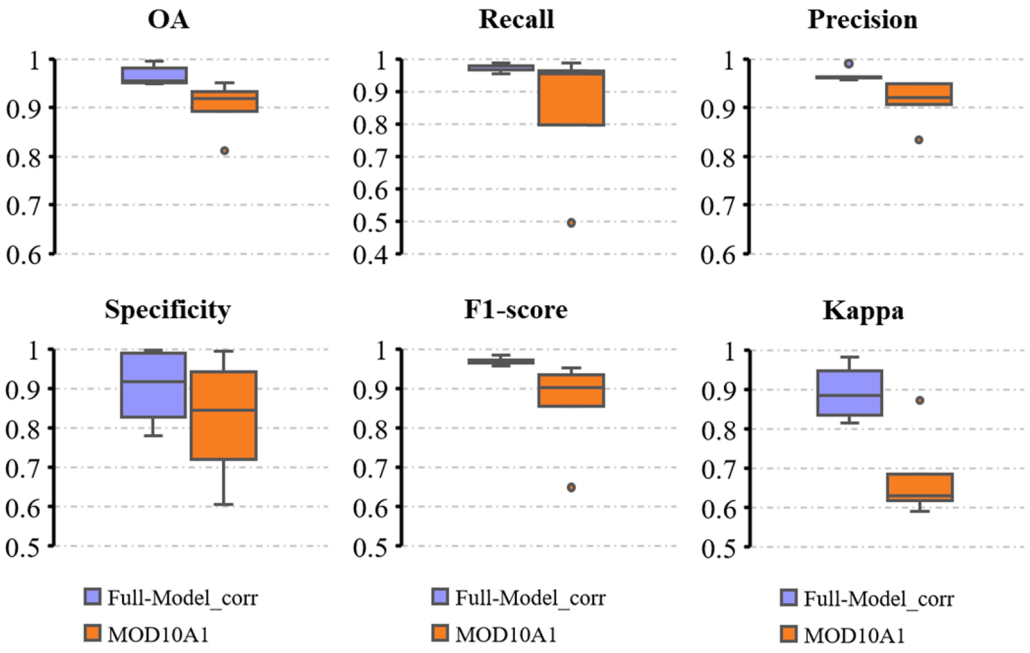


Fig. 16. Comparison of the variation of binary statistics from Full-Model_corr algorithm and MOD10A1 employed FSC binarization thresholds 0.5 in the combined validation scene.

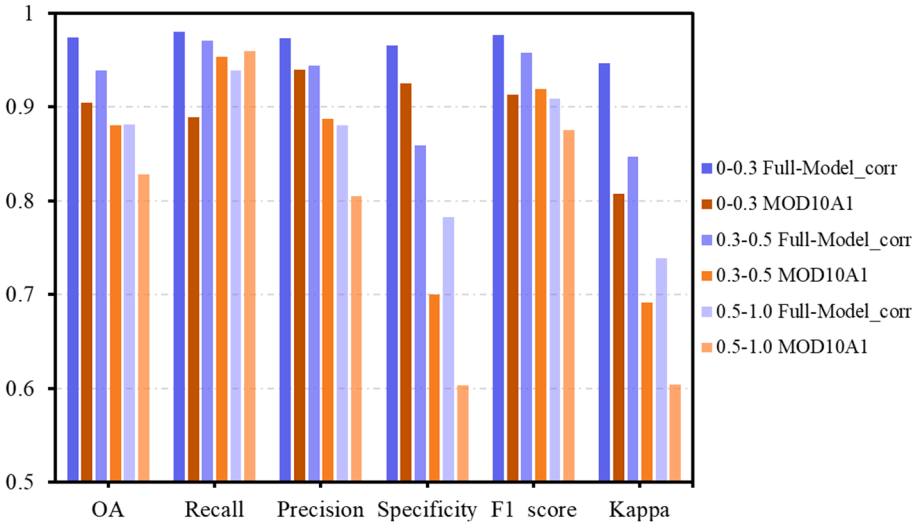


Fig. 17. Summary of binary statistics from Full-Model_corr algorithm and MOD10A1 with FSC dataset divided based on FVC. The FSC data in five scenes (A-E) was converted into binary values using FSC binarization thresholds 0.5.

datasets have been widely employed as the "true" observations in a large number of studies related to FSC retrieval using medium resolution remote sensing reflectance data (Dobreva and Klein, 2011; Kuter, 2021; Kuter et al., 2018; Metsämäki et al., 2015). Nevertheless, Metsämäki et al. (2015) showed that the validation results for the GlobSnow Snow Extent product were strongly dependent on the quality of the Landsat reference data, which to a certain extent, could not provide representative "true" observations for dense forest. In the case of the SNOMAP algorithm, highly patchy snow cover (i.e., low FSC) cannot be easily detected, while greater snow coverage (FSC $>$ 0.6) can be detected with high accuracy (0.98) (Hall et al., 1995). Because of forest cover, patchy snow, and sensor observation conditions, as pointed out by Kuter et al. (2018), the binary snow cover detected by the SNOMAP algorithm (Hall et al., 1995) do have intrinsic errors, resulting in an inevitable bias in the generated Landsat reference FSC (Berman et al., 2018; Rittger et al., 2013). Therefore, in some special cases, Landsat-based FSC reference maps may not represent the actual ground snow, especially in forested areas (Margulis et al., 2019; Metsämäki et al., 2015).

5.3. *Snow mapping in forest covered areas*

There are two main reasons why the observed FSC misrepresents the actual snow cover in forested areas (Rittger et al., 2020b). First, tree canopies and tall vegetation obscure satellite observations, making ground snow invisible in the satellite nadir view (Kostadinov et al., 2019; Liu et al., 2008; Raleigh et al., 2013). Second, an increase in the sensor VZA decreases the visible proportion of forest gaps and stretches the pixels (Rittger et al., 2020b; Xin et al., 2012). To improve FSC estimation from surface reflectance in vegetated areas, it is necessary to establish a robust retrieval model and adopt a canopy adjustment

approach. However, without considering the reference FSC error, the direct adoption of a canopy adjustment approach may not yield the expected better results under all VZA-FVC conditions (Rittger et al., 2020b). This may be because under the assumptions in Eq. (10), the visible FSC in open areas is applied in the FSC estimation in the canopy-covered areas but is not universally applicable across all land cover types (Kostadinov et al., 2019; Raleigh et al., 2013).

There are many studies related to canopy snow interception and the validation of sub-canopy snow. For example, Lv and Pomeroy (2019) assessed snow cover interception in large forests using satellite observations, while others have tried to adjust for canopy effect in FSC mapping using LIDAR data (Kostadinov et al., 2019), and, at a local scale, ground measurements and MODIS data (Raleigh et al., 2013). Additionally, differential remote sensing observation times increase the complexity of applying canopy adjustment when estimating FSC. Raleigh et al. (2013) reported that multiple variables and factors influence snow cover accumulation and ablation rates in forested regions, resulting in different spatial patterns beneath the forest canopy and within forests viewable gap by satellite (Varhola et al., 2010).

6. Conclusions

Vegetation cover is a challenging factor in FSC estimation using satellite-based remote sensing data. To enhance FSC retrieval accuracy from MODIS surface reflectance data for complex and heterogeneous vegetation cover environments, we applied various types of variables derived from surface reflectance data and multiple MODIS products (*Generality* variables, *Vegetation-Snow* related variables, *Location and Geometry* related variables, and *Auxiliary* variables). Twenty sub-models trained using the ERT method were integrated to alleviate dependence on the quality of the training sample. Furthermore, the canopy adjustment method was implemented to reduce the influence of canopy obscuration on FSC estimation in forested regions. Our validation results based on the independent Landsat dataset indicate that the proposed algorithm offers better performance for FSC prediction, by an average of 11 %, compared to the Trimmed-Model under different vegetation covers. The application of the "*Recommend*" canopy adjustment method with the conditions of VZA $\in [45^{\circ}, 70^{\circ}]$ and FVC $\in [0, 0.3]$ can obtain more accurate FSC values relative to the Landsat reference FSC data. When compared with the MOD10A1 FSC datasets, the FSC prediction and snow cover detection using our algorithm were significantly enhanced under dense vegetation covers (RMSE = 0.124 vs 0.202 and OA = 0.921 vs 0.784, respectively; Figs. 13 and 16). However, because of some overestimation and underestimation of FSC still occur, future efforts should be made to separate the snow reflectance signal from that of other surface materials and evaluate outputs against more Landsat images encompassing a range of vegetation types. In particular, for 500 m resolutions, a larger number of samples with different VZA and FVC values are needed to further verify the proposed canopy adjustment method.

CRediT authorship contribution statement

Xiongxin Xiao: Conceptualization, Methodology, Writing – original draft, Investigation, Visualization. **Tao He:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition. **Shunlin Liang:** Conceptualization, Writing – review & editing. **Xinyan Liu:** Resources, Validation, Writing – review & editing. **Yichuan Ma:** Resources, Validation, Writing – review & editing. **Shuang Liang:** Writing – review & editing. **Xiaona Chen:** Writing – review & editing.

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.

Data availability

Data will be made available on request.

Acknowledgements

This study was funded by the National Natural Science Foundation of China Grant (42090012), the Hubei Province Natural Science Foundation Grant (2021CFA082), and the National Key Research and Development Program of China Grant (2020YFA0608704). We appreciate the valuable suggestions of Dr. Feng Tian from School of Remote Sensing and Information Engineering of Wuhan University. We also thank the anonymous reviewers for their valuable comments and suggestions.

Appendix

See the Table A1, Table B1, Table C1, Table D1.

Table A
The regrouped classification of MCD12Q1 IGBP Land cover classification scheme.

Original classes	Original ID	Reclassified classes	Reclassified ID	Reclass labels
Evergreen needleleaf forest	1	Evergreen forest	1	Forest
Evergreen broadleaf forest	2			
Deciduous needleleaf forest	3	Deciduous forest	2	
Deciduous broadleaf forest	4			
Mixed forests	5	Mixed forest	3	
Closed shrublands	6	Shrub	4	Non-Forest
Open shrublands	7			
Woody savannas	8	Savannas	5	
Savannas	9			
Grasslands	10	Grasslands	6	
Permanent wetlands	11			
Croplands	12	Croplands	7	
Cropland/natural vegetation mosaic	14			
Urban and built-up lands	13	Bare land	8	
Permanent snow and ice	15			
Barren	16			
Water Bodies	17	Water	9	Water

Table B
The information of selected Landsat 8 OLI training and validation scenes. Total training sample is 379,522 (cf. Fig. C in the Appendix), and 72,038 for forest-type and 307,484 for non-forest type.

Scene ID	WRS-2 path/ row	Date acquired	Numbers of available pixels	
			Training sample	Testing sample
Training scenes				
1	009/022	5 May 2016	10,273	119,575
2	016/023	6 May 2016	19,090	121,133
3	018/027	29 Jan. 2016	16,930	64,279
4	026/030	4 Dec. 2015	18,535	129,923
5	028/032	3 Jan. 2016	10,504	93,243
6	030/034	16 Dec. 2015	2,095	81,668
7	032/031	31 Jan. 2016	2,214	86,341
8	032/036	30 Dec. 2015	7,278	56,259
9	033/032	22 Jan. 2016	16,101	64,352
10	039/033	13 Nov. 2015	20,092	103,777
11	042/020	10 April 2016	26,125	91,135
12	043/014	19 May 2016	16,666	99,358
13	043/025	1 April 2016	30,200	94,932
14	043/027	25 Nov. 2015	21,709	65,250
15	043/033	16 March 2016	15,693	123,495
16	049/016	27 April 2016	9,330	126,742
17	052/022	31 March 2016	30,125	92,315
18	056/014	28 April 2016	8,864	139,849
19	063/015	12 March 2016	19,265	102,614
20	063/017	4 Oct. 2015	13,506	71,183
21	071/013	12 Oct. 2015	22,796	79,581
22	073/018	18 March 2016	19,836	96,593
23	076/013	24 April 2016	22,295	105,984
Validation scenes				
A	030/032	1 Jan. 2016	58,564	
B	037/036	17 Dec. 2015	121,376	
C	039/020	21 April 2016	139,057	
D	064/017	20 April 2016	144,688	
E	075/017	16 March 2016	135,525	
Combined	–	–	599,210	

Table C
Binary accuracy metrics description.

	Ground: Snow	Ground: Snow-free
Satellite: Snow	True positive (TP)	False positive (FP)
Satellite: Snow-free	False negative (FN)	True negative (TN)
$OA = \mathrm{TP} + \mathrm{TN}/(TP+FN+FP+TN)$		
$Precision = TP/(TP+FP) = 1-CE$		
$Recall = TP/(TP+FN) = 1-OE$		
$F1\_score = (2*Precision*Recall)/(Precision+Recall)$		
$Specificity = TN/(TN+FP)$		

Table D
Performance of 20 FSC retrieval sub-models in 5 independent validation scenes (Appendix Table B). Bold face on each scene indicates the FSC results of the corresponding model with the best FSC retrieval performance.

	Scene A			Scene B			Scene C			Scene D			Scene E		
Model ID	*R*	MAE	RMSE	*R*	MAE	RMSE	*R*	MAE	RMSE	*R*	MAE	RMSE	*R*	MAE	RMSE
Model_1	0.975	0.039	0.068	0.951	0.074	0.121	0.884	0.137	0.157	0.913	0.142	0.168	0.909	0.096	0.122
Model_2	0.973	0.040	0.071	0.951	0.072	0.121	0.884	0.138	0.157	0.916	0.137	0.164	0.910	0.095	0.121
Model_3	0.974	0.039	0.069	0.946	0.078	0.127	0.887	0.135	0.155	0.914	0.140	0.167	0.910	0.095	0.121
Model_4	0.974	0.039	0.069	0.951	0.074	0.121	0.886	0.135	0.155	0.914	0.140	0.167	0.910	0.095	0.121
Model_5	0.973	0.040	0.071	0.948	0.075	0.125	0.879	0.141	0.160	0.911	0.143	0.169	0.909	0.096	0.122
Model_6	0.973	0.039	0.071	0.951	0.073	0.120	0.887	0.136	0.155	0.917	0.137	0.163	0.911	0.095	0.120
Model_7	0.972	0.040	0.072	0.949	0.075	0.124	0.881	0.140	0.159	0.913	0.142	0.167	0.910	0.095	0.121
Model_8	0.973	0.040	0.070	0.949	0.072	0.124	0.884	0.139	0.157	0.917	0.136	0.163	0.908	0.096	0.122
Model_9	0.973	0.039	0.071	0.950	0.074	0.123	**0.891**	**0.132**	**0.152**	0.914	0.140	0.167	**0.913**	**0.093**	**0.119**
Model_10	0.973	0.039	0.071	0.950	0.075	0.123	0.888	0.134	0.154	0.913	0.140	0.167	0.907	0.097	0.123
Model_11	0.975	0.037	0.068	0.951	0.072	0.121	0.880	0.141	0.159	0.915	0.140	0.166	0.910	0.096	0.121
Model_12	0.974	0.039	0.070	0.948	0.076	0.124	0.882	0.138	0.158	**0.919**	**0.135**	**0.162**	0.911	0.094	0.120
Model_13	0.971	0.041	0.073	0.951	0.072	0.121	0.889	0.134	0.154	0.910	0.144	0.170	0.912	0.094	0.120
Model_14	0.972	0.040	0.072	**0.952**	**0.070**	**0.119**	0.881	0.140	0.159	0.911	0.143	0.169	0.910	0.095	0.121
Model_15	0.974	0.039	0.070	0.942	0.082	0.131	0.884	0.138	0.157	0.912	0.142	0.168	0.910	0.095	0.121
Model_16	0.972	0.040	0.072	0.948	0.075	0.125	0.882	0.141	0.158	0.915	0.139	0.166	0.911	0.094	0.121
Model_17	0.972	0.040	0.071	0.945	0.077	0.128	0.888	0.134	0.154	0.915	0.139	0.166	0.910	0.095	0.121
Model_18	0.971	0.041	0.073	0.951	0.072	0.121	0.883	0.138	0.157	0.912	0.141	0.168	0.909	0.096	0.122
Model_19	**0.975**	**0.038**	**0.068**	0.950	0.073	0.122	0.882	0.140	0.158	0.910	0.144	0.170	0.908	0.096	0.122
Model_20	0.972	0.040	0.071	0.951	0.072	0.121	0.886	0.136	0.156	0.914	0.140	0.167	0.907	0.097	0.123
Full-Model	0.974	0.039	0.069	0.950	0.073	0.122	0.886	0.137	0.155	0.915	0.139	0.165	0.912	0.094	0.120

See the Fig. A1, Fig. B1, Fig. C1, Fig. D1.

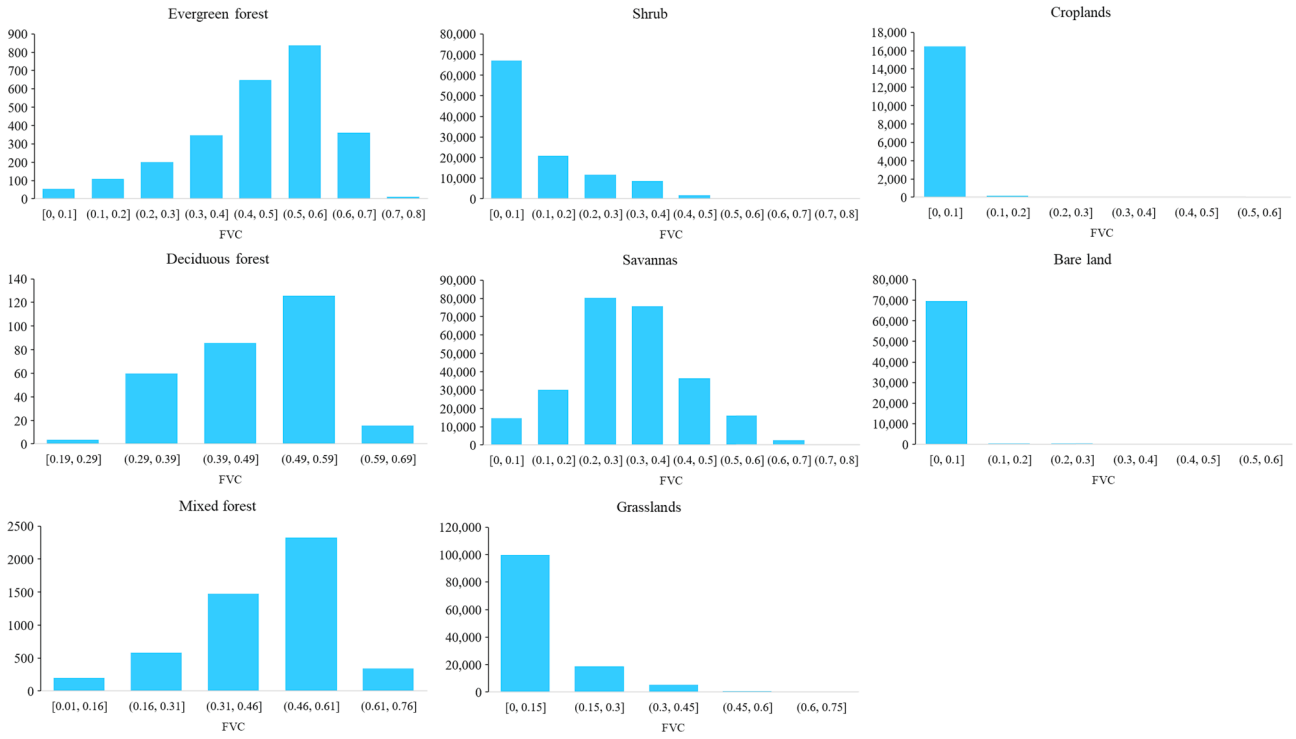


Fig. A. The FVC distribution of the combined validation sample about 599,210 pixels with respect to the land cover types (cf. Tables 2 and 3).

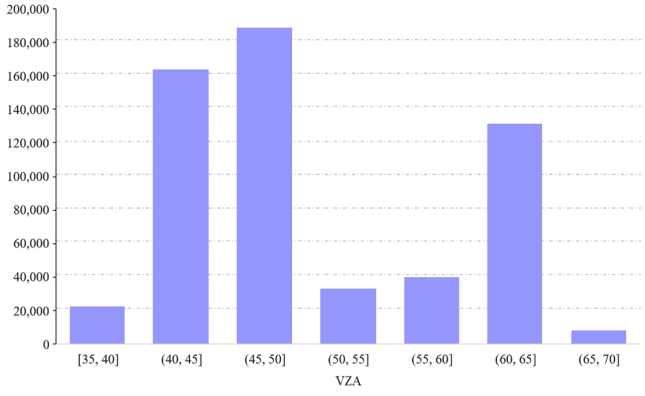


Fig. B. The VZA distribution of the combined validation sample about 599,210 pixels.

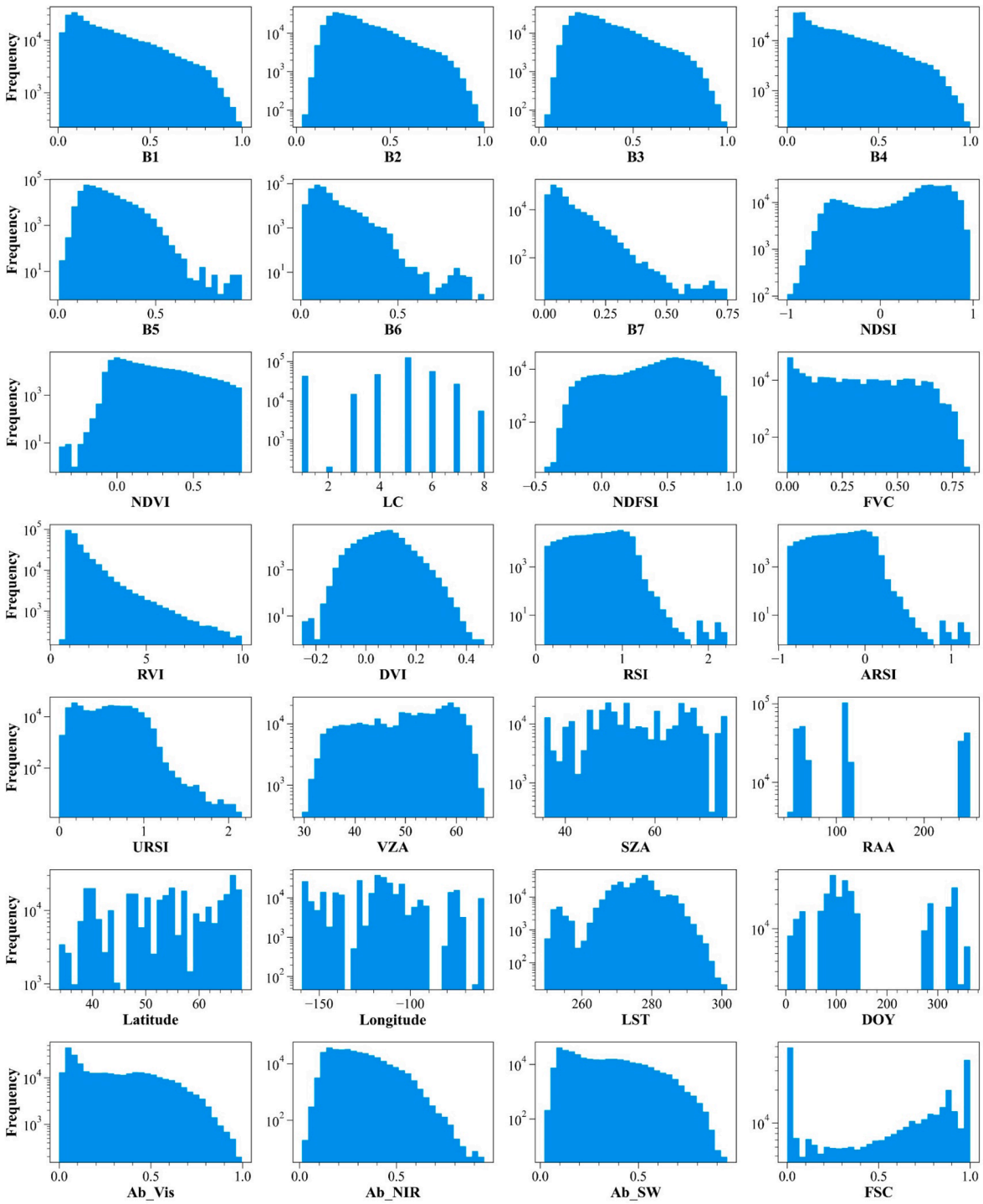


Fig. C. Histograms of 27 predictor variables and target variable (FSC) on training samples.

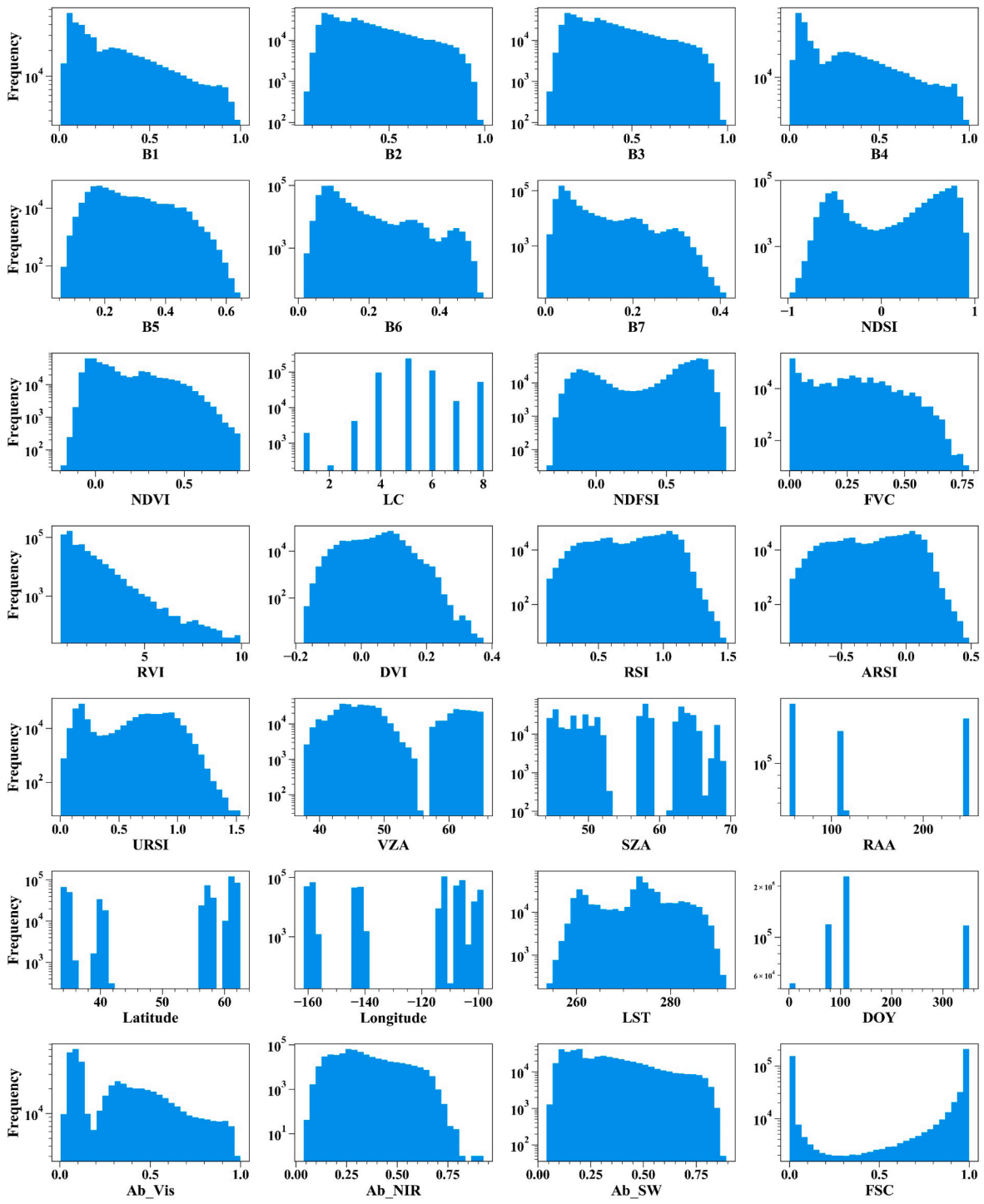


Fig. D. Histograms of 27 predictor variables and target variable (FSC) on validation data sets.

References

Aalstad, K., Westermann, S., Bertino, L., 2020. Evaluating satellite retrieved fractional snow-covered area at a high-Arctic site using terrestrial photography. Remote Sens. Environ. 239, 111618 https://doi.org/10.1016/j.rse.2019.111618.

Arsenault, K.R., Houser, P.R., De Lannoy, G.J.M., 2014. Evaluation of the MODIS snow cover fraction product. Hydrol. Process. 28, 980–998. https://doi.org/10.1002/hyp.9636.

Baldridge, A.M., Hook, S.J., Grove, C.I., Rivera, G., 2009. The ASTER spectral library version 2.0. Remote Sens. Environ. 113, 711–715. https://doi.org/10.1016/j.rse.2008.11.007.

Belgiu, M., Drăgu, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

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

Bohn, N., Painter, T.H., Thompson, D.R., Carmon, N., Susiluoto, J., Turmon, M.J., Helmlinger, M.C., Green, R.O., Cook, J.M., Guanter, L., 2021. Optimal estimation of snow and ice surface parameters from imaging spectroscopy measurements. Remote Sens. Environ. 264, 112613 https://doi.org/10.1016/j.rse.2021.112613.

Bormann, K.J., Brown, R.D., Derksen, C., Painter, T.H., 2018. Estimating snow-cover trends from space. Nat. Clim. Chang. 8, 924–928. https://doi.org/10.1038/s41558-018-0318-3.

Cheng, Z., Guo, Z., Fu, P., Yang, J., Wang, Q., 2021. New insights into the effects of methane and oxygen on heat/mass transfer in reactive porous media. Int. Commun. Heat Mass Transf. 129, 105652 https://doi.org/10.1016/j.icheatmasstransfer.2021.105652.

Cortés, G., Girotto, M., Margulis, S.A., 2014. Analysis of sub-pixel snow and ice extent over the extratropical Andes using spectral unmixing of historical Landsat imagery. Remote Sens. Environ. 141, 64–78. https://doi.org/10.1016/j.rse.2013.10.023.

Czyzowska-Wisniewski, E.H., van Leeuwen, W.J.D., Hirschboeck, K.K., Marsh, S.E., Wisniewski, W.T., 2015. Fractional snow cover estimation in complex alpine-forested environments using an artificial neural network. Remote Sens. Environ. 156, 403–417. https://doi.org/10.1016/j.rse.2014.09.026.

Déry, S.J., Brown, R.D., 2007. Recent Northern Hemisphere snow cover extent trends and implications for the snow-albedo feedback. Geophys. Res. Lett. 34, L22504. https://doi.org/10.1029/2007GL031474.

Dietz, A.J., Kuenzer, C., Gessner, U., Dech, S., 2012. Remote sensing of snow - a review of available methods. Int. J. Remote Sens. 33, 4094–4134. https://doi.org/10.1080/01431161.2011.640964.

Dimiceli, C., Carroll, M., Sohlberg, R., Kim, D.H., Kelly, M., Townshend, J.R.G., 2015. MOD44B MODIS/Terra vegetation continuous fields yearly L3 global 250 m SIN grid V006. NASA EOSDIS L. Process. Distrib. Act. Arch. Cent.

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

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

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X., 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. Remote Sens. Environ. 114, 168–182. https://doi.org/10.1016/j.rse.2009.08.016.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach. Learn. 63, 3–42. https://doi.org/10.1007/s10994-006-6226-1.

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 (2), 127–140.

Hannula, H.R., Heinilä, K., Böttcher, K., Mattila, O.P., Salminen, M., Pulliainen, J., 2020. Laboratory, field, mast-borne and airborne spectral reflectance measurements of boreal landscape during spring. Earth Syst. Sci. Data 12, 719–740. https://doi.org/10.5194/essd-12-719-2020.

Hao, X.H., Wang, Jie, Wang, Jian, Zhang, P., Huang, C.L., 2013. The measurement and retrieval of the spectral reflectance of different snow grain size on Northern Xinjiang, China. Spectrosc. Spectr. Anal. https://doi.org/10.3964/j.issn.1000-0593(2013)01-0190-06.

He, T., Liang, S., Wang, D., Cao, Y., Gao, F., Yu, Y., Feng, M., 2018. Evaluating land surface albedo estimation from Landsat MSS, TM, ETM +, and OLI data based on the unified direct estimation approach. Remote Sens. Environ. 204, 181–196. https://doi.org/10.1016/j.rse.2017.10.031.

Heinilä, K., Salminen, M., Pulliainen, J., Cohen, J., Metsämäki, S., Pellikka, P., 2014. The effect of boreal forest canopy to reflectance of snow covered terrain based on airborne imaging spectrometer observations. Int. J. Appl. Earth Obs. Geoinf. 27, 31–41. https://doi.org/10.1016/j.jag.2013.06.004.

Huang, Y., Liu, H., Yu, B., Wu, J., Kang, E.L., Xu, M., Wang, S., Klein, A., Chen, Y., 2018. Improving MODIS snow products with a HMRF-based spatio-temporal modeling technique in the Upper Rio Grande Basin. Remote Sens. Environ. 204, 568–582. https://doi.org/10.1016/j.rse.2017.10.001.

Huson, L., 2003. Statistical Methods for the Analysis of Repeated Measurements. J. R. Stat. Soc. Ser. D (The Stat.) 52, 691–692. https://doi.org/10.1046/j.1467-9884.2003.t01-3-00383_5.x.

Jiang, R., Tang, W., Wu, X., Fu, W., 2009. A random forest approach to the detection of epistatic interactions in case-control studies. BMC Bioinf. 10, S65. https://doi.org/10.1186/1471-2105-10-S1-S65.

Kostadinov, T.S., Schumer, R., Hausner, M., Bormann, K.J., Gaffney, R., McGwire, K., Painter, T.H., Tyler, S., Harpold, A.A., 2019. Watershed-scale mapping of fractional snow cover under conifer forest canopy using lidar. Remote Sens. Environ. 222, 34–49. https://doi.org/10.1016/j.rse.2018.11.037.

Kuter, S., 2021. Completing the machine learning saga in fractional snow cover estimation from MODIS Terra reflectance data: Random forests versus support vector regression. Remote Sens. Environ. 255, 112294 https://doi.org/10.1016/j.rse.2021.112294.

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

Kuter, S., Bolat, K., Akyurek, Z., 2022. Remote Sensing of Environment A machine learning-based accuracy enhancement on EUMETSAT H-SAF H35 effective snow-covered area product. Remote Sens. Environ. 272, 112947 https://doi.org/10.1016/j.rse.2022.112947.

Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. Geosci. Front. 7, 3–10. https://doi.org/10.1016/j.gsf.2015.07.003.

Li, W., Guo, W., Qiu, B., Xue, Y., Hsu, P.C., Wei, J., 2018. Influence of Tibetan Plateau snow cover on East Asian atmospheric circulation at medium-range time scales. Nat. Commun. 9, 4243. https://doi.org/10.1038/s41467-018-06762-5.

Liang, H., Huang, X., Sun, Y., Wang, Y., Liang, T., 2017. Fractional Snow-Cover Mapping Based on MODIS and UAV Data over the Tibetan Plateau. Remote Sens. 9, 1332. https://doi.org/10.3390/rs9121332.

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

Lucht, W., Schaaf, C.B., Strahler, A.H., 2000. An algorithm for the retrieval of albedo from space using semiempirical BRDF models. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/36.841980.

Lv, Z., Pomeroy, J.W., 2019. Detecting intercepted snow on mountain needleleaf forest canopies using satellite remote sensing. Remote Sens. Environ. 231, 111222 https://doi.org/10.1016/j.rse.2019.111222.

Margulis, S.A., Liu, Y., Baldo, E., 2019. A Joint Landsat- and MODIS-Based Reanalysis Approach for Midlatitude Montane Seasonal Snow Characterization. Front. Earth Sci. 7, 1–23. https://doi.org/10.3389/feart.2019.00272.

Metsämäki, S.J., Anttila, S.T., Markus, H.J., Vepsäläinen, J.M., 2005. A feasible method for fractional snow cover mapping in boreal zone based on a reflectance model. Remote Sens. Environ. 95, 77–95. https://doi.org/10.1016/j.rse.2004.11.013.

Metsämäki, S., Mattila, O.-P., Pulliainen, J., Niemi, K., Luojus, K., Böttcher, K., 2012. An optical reflectance model-based method for fractional snow cover mapping applicable to continental scale. Remote Sens. Environ. 123, 508–521.

Metsämäki, S., Pulliainen, J., Salminen, M., Luojus, K., Wiesmann, A., Solberg, R., Böttcher, K., Hiltunen, M., Ripper, E., 2015. Introduction to GlobSnow Snow Extent products with considerations for accuracy assessment. Remote Sens. Environ. 156, 96–108. https://doi.org/10.1016/j.rse.2014.09.018.

Millard, K., Richardson, M., 2015. On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping. Remote Sens. 7, 8489–8515. https://doi.org/10.3390/rs70708489.

Moosavi, V., Malekinezhad, H., Shirmohammadi, B., 2014. Fractional snow cover mapping from MODIS data using wavelet-artificial intelligence hybrid models. J. Hydrol. 511, 160–170. https://doi.org/10.1016/j.jhydrol.2014.01.015.

Niittynen, P., Heikkinen, R.K., Luoto, M., 2020. Decreasing snow cover alters functional composition and diversity of Arctic tundra. Proc. Natl. Acad. Sci. USA 117, 21480–21487. https://doi.org/10.1073/pnas.2001254117.

Painter, T.H., Dozier, J., Roberts, D.A., Davis, R.E., Green, R.O., 2003. Retrieval of subpixel snow-covered area and grain size from imaging spectrometer data. Remote Sens. Environ. 85, 64–77. https://doi.org/10.1016/S0034-4257(02)00187-6.

Painter, T.H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R.E., Dozier, J., 2009. Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. Remote Sens. Environ. 113, 868–879. https://doi.org/10.1016/j.rse.2009.01.001.

Parajka, J., Blöschl, G., 2008. Spatio-temporal combination of MODIS images - potential for snow cover mapping. Water Resour. Res. 44, 1–13. https://doi.org/10.1029/2007WR006204.

Pulliainen, J., Luojus, K., Derksen, C., Mudryk, L., Lemmetyinen, J., Salminen, M., Ikonen, J., Takala, M., Cohen, J., Smolander, T., Norberg, J., 2020. Patterns and trends of Northern Hemisphere snow mass from 1980 to 2018. Nature 581, 294–298. https://doi.org/10.1038/s41586-020-2258-0.

Qiu, S., Zhu, Z., He, B., 2019. Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery. Remote Sens. Environ. 231, 111205 https://doi.org/10.1016/j.rse.2019.05.024.

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.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Riggs, G.A., Hall, D.K., Román, M.O., 2017. Overview of NASA's MODIS and Visible Infrared Imaging Radiometer Suite (VIIRS) snow-cover Earth System Data Records. Earth Syst. Sci. Data. https://doi.org/10.5194/essd-9-765-2017.

Riggs, G., Hall, D., 2015. MODIS Snow Products Collection 6 User Guide. NSIDC User Guid. Ser.

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

Rittger, K., Bormann, K.J., Bair, E.H., Dozier, J., Painter, T.H., 2020a. Evaluation of VIIRS and MODIS snow covered fraction in High Mountain Asia using Landsat 8. Front. Remote Sens. 2, 1–15. https://doi.org/10.3389/frsen.2021.647154.

Rittger, K., Raleigh, M.S., Dozier, J., Hill, A.F., Lutz, J.A., Painter, T.H., 2020b. Canopy Adjustment and Improved Cloud Detection for Remotely Sensed Snow Cover Mapping. Water Resour. Res. 56, 1–20. https://doi.org/10.1029/2019WR024914.

Salcedo-Sanz, S., Ghamisi, P., Piles, M., Werner, M., Cuadra, L., Moreno-Martínez, A., Izquierdo-Verdiguier, E., Muñoz-Marí, J., Mosavi, A., Camps-Valls, G., 2020. Machine learning information fusion in Earth observation: A comprehensive review of methods, applications and data sources. Inf. Fusion 63, 256–272. https://doi.org/10.1016/j.inffus.2020.07.004.

Salomonson, V.V., Appel, I., 2004. Estimating fractional snow cover from MODIS using the normalized difference snow index. Remote Sens. Environ. 89 (3), 351–360.

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 (7), 1747–1756.

Santolaria-Otín, M., Zolina, O., 2020. Evaluation of snow cover and snow water equivalent in the continental Arctic in CMIP5 models. Clim. Dyn. 55, 2993–3016. https://doi.org/10.1007/s00382-020-05434-9.

Varhola, A., Coops, N.C., Weiler, M., Moore, R.D., 2010. Forest canopy effects on snow accumulation and ablation: An integrative review of empirical results. J. Hydrol. 392, 219–233. https://doi.org/10.1016/j.jhydrol.2010.08.009.

Vermote, E., Wolfe, R., 2015. MOD09GA MODIS/terra surface reflectance daily L2G global 1 km and 500 m SIN grid V006. NASA EOSDIS L. Process. DAAC 10.

Wan, Z., Dozier, J., 1996. A generalized split-window algorithm for retrieving land-surface temperature from space. IEEE Trans. Geosci. Remote Sens. 34, 892–905. https://doi.org/10.1109/36.508406.

Wang, G., Jiang, L., Shi, J., Su, X., 2021. A universal ratio snow index for fractional snow cover estimation. IEEE Geosci. Remote Sens. Lett. 18, 721–725. https://doi.org/10.1109/LGRS.2020.2982053.

Wang, X., Wang, J., Che, T., Huang, X., Hao, X., Li, H., 2018. Snow cover mapping for complex mountainous forested environments based on a multi-index technique. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11, 1433–1441. https://doi.org/10.1109/JSTARS.2018.2810094.

Wang, J., Wu, X., Wen, J., Xiao, Q., Gong, B., Ma, D., Cui, Y., Lin, X., Bao, Y., 2022. Upscaling in situ site-based albedo using machine learning models: main controlling factors on results. IEEE Trans. Geosci. Remote Sens. 60, 1–16.

Warren, S.G., 1982. Optical properties of snow. Rev. Geophys. 20, 67. https://doi.org/10.1029/RG020i001p00067.

Warren, S.G., Wiscombe, W.J., 1980. A model for the Spectral albedo of snow. II: Snow containing Atmospheric Aerosols. J. Atmos. Sci. 37 (12), 2734–2745.

Watanabe, S., Kotsuki, S., Kanae, S., Tanaka, K., Higuchi, A., 2020. Snow water scarcity induced by record-breaking warm winter in 2020 in Japan. Sci. Rep. 10, 1–7. https://doi.org/10.1038/s41598-020-75440-8.

Wolfe, R.E., 2006. MODIS Geolocation. In: Qu, J.J., Gao, W., Kafatos, M., Murphy, R.E., Salomonson, V.V. (Eds.), Earth Science Satellite Remote Sensing. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 50–73.

Wu, R., Chen, S., 2016. Regional change in snow water equivalent–surface air temperature relationship over Eurasia during boreal spring. Clim. Dyn. 47, 2425–2442. https://doi.org/10.1007/s00382-015-2972-8.

Wu, X., Wen, J., Xiao, Q., Bao, Y., You, D., Wang, J., Ma, D., Lin, X., Gong, B., 2022. Quantification of the uncertainty caused by geometric registration errors in multiscale validation of satellite products. IEEE Geosci. Remote Sens. Lett. 19, 1–5.

Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., Woodcock, C.E., 2012. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. Remote Sens. Environ. 122, 2–10. https://doi.org/10.1016/j.rse.2012.01.010.

Xiao, Z., Duan, A., 2016. Impacts of Tibetan Plateau snow cover on the interannual variability of the East Asian Summer Monsoon. J. Clim. 29, 8495–8514. https://doi.org/10.1175/JCLI-D-16-0029.1.

Xiao, X., Liang, S., He, T., Wu, D., Pei, C., Gong, J., 2021. Estimating fractional snow cover from passive microwave brightness temperature data using MODIS snow cover product over North America. Cryosph. 15, 835–861. https://doi.org/10.5194/tc-15-835-2021.

Xiao, X., Zhang, T., Zhong, X., Shao, W., Li, X., 2018. Support vector regression snow-depth retrieval algorithm using passive microwave remote sensing data. Remote Sens. Environ. 210, 48–64. https://doi.org/10.1016/j.rse.2018.03.008.

Xiao, X., Zhang, T., Zhong, X., Li, X., 2020. Spatiotemporal variation of snow depth in the Northern Hemisphere from 1992 to 2016. Remote Sens. 12, 2728. https://doi.org/10.3390/rs12172728.

Xiao, X., He, T., Liang, S., Zhao, T., 2022. Improving fractional snow cover retrieval from passive microwave data using a radiative transfer model and machine learning method. IEEE Trans. Geosci. Remote Sens. 60, 1–15. https://doi.org/10.1109/TGRS.2021.3128524.

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.

Yang, J., Jiang, L., Dai, L., Pan, J., Wu, S., Wang, G., 2019. The consistency of SSM/I vs. SSMIS and the influence on snow cover detection and snow depth estimation over China. Remote Sens. 11, 1879. https://doi.org/10.3390/rs11161879.

Yang, J.W., Jiang, L.M., Lemmetyinen, J., Pan, J.M., Luojus, K., Takala, M., 2021. Improving snow depth estimation by coupling HUT-optimized effective snow grain size parameters with the random forest approach. Remote Sens. Environ. 264, 112630 https://doi.org/10.1016/j.rse.2021.112630.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sens. Environ. 241, 111716 https://doi.org/10.1016/j.rse.2020.111716.