



## Research papers

# Enhanced scaling effects significantly lower the ability of MODIS normalized difference snow index to estimate fractional and binary snow cover on the Tibetan Plateau

Hongbo Zhang<sup>a,b,c,\*</sup>, Fan Zhang<sup>d,e,f</sup>, Guoqing Zhang<sup>d,e</sup>, Wei Yan<sup>g</sup>, Sien Li<sup>a</sup>

<sup>a</sup> College of Water Resources & Civil Engineering, China Agricultural University, Beijing, China

<sup>b</sup> State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing, China

<sup>c</sup> State Key Laboratory of Cryospheric Science, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou, China

<sup>d</sup> Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences (CAS), Beijing, China

<sup>e</sup> CAS Center for Excellence in Tibetan Plateau Earth Sciences, Beijing, China

<sup>f</sup> University of Chinese Academy of Sciences, Beijing, China

<sup>g</sup> School of Geographic Sciences, Xinyang Normal University, Xinyang, China



## ARTICLE INFO

This manuscript was handled by Dr Emmanouil Anagnostou Editor-in-Chief

## Keywords:

Scaling effects  
MODIS fractional snow cover  
Binary snow cover  
NDSI threshold  
Tibetan Plateau

## ABSTRACT

MODIS fractional (FSC) and binary (BSC) snow-cover data are important for obtaining accurate spatiotemporal snow-cover information for the Tibetan Plateau (TP) where rapid warming is closely related to snow-cover changes. However, FSC and BSC data are no longer provided in the newly released version (v6) of the MODIS snow cover product, having been replaced with normalized difference snow index (NDSI) data. Recent studies have observed clearly lower accuracy of MODIS snow cover data on the TP than in other areas, possibly implying that there are strong scaling effects due to the complex terrain and land cover which are not well understood. A total of 353 Landsat-8 scenes covering most parts of the TP are used to establish a new empirical relationship between FSC and MODIS NDSI for FSC estimation and a new NDSI threshold for BSC estimation. The results indicate that the new regression model (mean root-mean-squared-deviation (RMSD): 0.22) has a better FSC estimation accuracy than the previously used global reference equation (mean RMSD: 0.24) and that the new NDSI threshold of 0.29 (mean Cohen's Kappa (CK): 0.49) outperforms the global reference NDSI threshold of 0.4 (mean CK: 0.40) in BSC estimation. The relatively low accuracy could be due to an enhanced scaling effect. The 30-m Landsat-8 NDSI data are upscaled to 500-m (MODIS) to analyze the scaling effects on FSC and BSC estimates made using MODIS NDSI. We find that the methods using MODIS NDSI have much lower estimation accuracy, for both FSC and BSC, compared with those using upscaled Landsat-8 NDSI. An analysis of variance (ANOVA) test which considers 512 combinations of aspect, slope and normalized difference vegetation index (NDVI) further demonstrates that the enhanced scaling effects are mainly caused by terrain factors (i.e. aspect and slope). The optimal NDSI threshold for estimating BSC generally increases with slope and decreases as the aspect varies from the southeast to northwest. This study has important implications for the optimal use of MODIS NDSI snow cover data on the TP and highlights the importance of developing more advanced methods which take more factors into account.

## 1. Introduction

Snow is a key component of the cryosphere and has important effects on climate and water resources. Snow cover, with its high albedo, is closely linked to regional climate change via the “snow albedo feedback” mechanism (Ma et al., 2020; Thackeray et al., 2019). In addition, in high-elevation areas such as the Himalayas, seasonal snow melt is an

important source of water (Armstrong et al., 2019). Therefore, accurate snow cover information is crucial for research on the Tibetan Plateau (TP), where snow cover changes play a key role in the “Elevation Dependent Warming” (Pepin et al., 2019; You et al., 2019), and where snow melt makes relatively large contributions (>20%) to total runoff in the headwater catchments of several major Asian rivers (Zhang et al., 2013b).

In mountainous areas, remote sensing is often the most efficient way

\* Corresponding author at: College of Water Resources & Civil Engineering, China Agricultural University, Beijing, China.

E-mail address: [zhanghongbo@cau.edu.cn](mailto:zhanghongbo@cau.edu.cn) (H. Zhang).

## Nomenclature

BSC	Binary snow cover		
B3 <sub>L8_30m</sub>	Landsat-8 band 3 at the original resolution of 30 m	LBMU	An LBM model of the average (or “universal”) linear relationships that are established using samples from all the 201 training scenes
B3 <sub>L8_500m</sub>	Landsat-8 band 3 aggregated to MODIS scale (i.e. ~500 m)	MA	A type A model that takes MODIS normalized difference snow index as the independent variable and snow-cover fraction as the dependent variable
B4 <sub>L8_30m</sub>	Landsat-8 band 4 at the original resolution of 30 m	MAU	An MA model of the average (or “universal”) linear relationships that are established using samples from all the 201 training scenes
B5 <sub>L8_30m</sub>	Landsat-8 band 5 at the original resolution of 30 m	MB	A type B model that takes snow-cover fraction as the independent variable and MODIS normalized difference snow index as the dependent variable
B6 <sub>L8_30m</sub>	Landsat-8 band 6 at the original resolution of 30 m	MBU	An MB model of the average (or “universal”) linear relationships that are established using samples from all the 201 training scenes
B6 <sub>L8_500m</sub>	Landsat-8 band 6 aggregated to MODIS scale (i.e. ~500 m)	MR	The global reference model that was used for producing global MODIS snow-cover fraction
FSC	Fractional snow cover	NDSI <sub>L8_30m</sub>	Landsat-8 normalized difference snow index at the original resolution of 30 m
LAI	A type A model that takes NDSI <sub>L8I</sub> as the independent variable and snow cover fraction as the dependent variable	NDSI <sub>L8I</sub>	The normalized difference snow index calculated with pixel-aggregated Landsat-8 bands (as Input)
LAIU	An LAI model of the average (or “universal”) linear relationships that are established using samples from all the 201 training scenes	NDSI <sub>L8M</sub>	The Mean normalized difference snow index that is calculated by aggregating all the corresponding 30-m Landsat-8 normalized difference snow index values within a MODIS pixel
LAM	A type A model that takes NDSI <sub>L8M</sub> as the independent variable and snow cover fraction as the dependent variable	NDSI <sub>MOD</sub>	MODIS normalized difference snow index
LAMU	An LAM model of the average (or “universal”) linear relationships that are established using samples from all the 201 training scenes		
LBI	A type B model that takes snow cover fraction as the independent variable and NDSI <sub>L8I</sub> as the dependent variable		
LBIU	An LBI model of the average (or “universal”) linear relationships that are established using samples from all the 201 training scenes		
LBM	A type B model that takes snow-cover fraction as the		

to carry out large-scale snow mapping, due to the sparse coverage of ground-based observations (Zhang and Ma, 2018). This is especially the case for the TP, which has an average elevation greater than 4000 m a.s.l. and no regular weather stations above 5000 m. As snow has a relatively high reflectance in the visible wavelengths and a relatively low reflectance in the shortwave infrared band, the normalized difference snow index (NDSI) is commonly used for snow detection for a wide range of sensors, including MODIS (Moderate Resolution Imaging Spectroradiometer), Landsat TM (Thematic Mapper), ETM+ (Enhanced Thematic Mapper Plus), OLI (Operational Land Imager) and Sentinel-2 MSI (MultiSpectral Instrument) (Crawford, 2015; Gascoïn et al., 2019; Hall et al., 1995; Salomonson and Appel, 2004; Yin et al., 2013). Out of these products, the MODIS daily snow cover, with its relatively high temporal frequency (twice a day), is probably the most popular, although it does have a somewhat coarse spatial resolution of ~500 m. Many studies have used MODIS fractional or binary snow-cover data for the TP for a variety of applications, such as analyzing the changes in snow-cover area (Pu et al., 2007; Zhang et al., 2013a), snow line (Choubin et al., 2019; Tang et al., 2014), snow-covered days (Tang et al., 2013a; Zhang et al., 2012) and snow-cover phenology (Wang et al., 2017), and for providing important snow cover information for driving, calibrating or verifying hydrological models (Immerzeel et al., 2009; Shrestha et al., 2011; Zhang et al., 2015). In addition, most of the composite snow-cover products use MODIS snow cover data as an important input (Chen et al., 2018; Huang et al., 2017; Yu et al., 2016).

However, the MODIS fractional and binary snow-cover data are no longer provided in the newest version (v6) of the MODIS snow-cover product (Riggs et al., 2016b). A good linear relationship between NDSI and snow-cover fraction was the theoretical basis of MODIS fractional snow-cover (FSC) data (Riggs et al., 2006). The original regression equation was developed from 14 Landsat scenes and subsequently implemented as a global reference equation for producing MODIS Terra FSC data around the world, with the validation root-mean-squared-

deviation (RMSD) and correlation coefficient (R) reported to be about 0.1 and 0.9, respectively (Salomonson and Appel, 2006). The MODIS binary (i.e., snow or non-snow) snow cover data (BSC) uses the assumption that when the NDSI threshold >0.4 at least 50% of the pixel is covered by snow (Crawford, 2015). However, increasingly, evidence suggests that either the globally applied equation, or the spatially fixed NDSI threshold, might not be optimal for local applications with variations in landscape and satellite viewing conditions (Härer et al., 2018; Mishra et al., 2009; Riggs et al., 2017). Thus, both FSC and BSC have now been replaced by NDSI snow-cover data containing only NDSI values for each pixel. These data are expected to be suitable for a wider range of applications. This change poses important questions about what regression equation between FSC and MODIS NDSI should be applied, and which NDSI threshold should be used, when using the new MODIS snow cover product on the TP. Some recent studies, based on daily snow depth observations (Zhang et al., 2020, 2019b), suggest that 0.1 is the optimal NDSI threshold for BSC estimation in China (including the TP). However, their sparse snow-depth observations may not represent actual FSC. The value of FSC which corresponds to an NDSI threshold of 0.1 also remains unclear.

The scale issue is an important factor affecting the estimation accuracy of remotely sensed variables (McCabe and Wood, 2006). Land-surface heterogeneity including different topography and land-cover conditions has been demonstrated to greatly impact the prediction of various environmental parameters, such as leaf area index (Garrigues et al., 2006; Liang, 2000), surface energy flux (Kustas et al., 2004) and evapotranspiration (Long et al., 2011; Sharma et al., 2016). In these studies, the aggregation of higher-resolution satellite images (e.g. 30 m Landsat missions) to larger pixel sizes, such as 500 m and 1 km, is a common way to mimic lower-resolution sensors. Recent studies have also found that, while the accuracy of MODIS snow cover data is generally good for stations outside the TP, it is much worse for those within the TP (Zhang et al., 2019b). Thus, there could be strong scaling

effects on FSC and BSC estimates for the TP. However, to-date, few studies have investigated the scaling effects on FSC or BSC estimation using NDSI. Härer et al. (2018) calibrate the NDSI threshold for a range of spatial resolutions, from 30 m to 990 m, using a camera-based snow-cover map, but they do not test for MODIS snow-cover data. The effects of elevation or land-cover types on the accuracy of MODIS snow cover data have been analyzed (Arsenault et al., 2014; Huang et al., 2011; Rittger et al., 2013; Zhang et al., 2019b), but the scale influence on FSC or BSC estimation from MODIS NDSI is poorly understood, especially for the TP.

Moderate resolution Landsat images have been used for validating MODIS fractional snow-cover data or developing new local relationships (Cyzowska-Wisniewski et al., 2015; Dobрева and Klein, 2011; Hou et al., 2020; Kuter et al., 2018; Wang et al., 2018). For example, Rittger et al. (2013) assessed MODIS snow-cover data based on 172 Landsat images, but only 25 of these were of the TP, and these were clustered in the Nepal Himalayas. The limited number and spatial coverage of reference images restricts our knowledge about snow mapping accuracy using MODIS NDSI and the factors influencing it. Obviously, more reference scenes should be used to increase the spatial representativeness and to establish a more reliable NDSI threshold for BSC estimation and a more reliable FSC relationship for FSC estimation.

Here, our objectives are: (1) re-evaluate the ability of the global linear regression equation to estimate FSC from MODIS NDSI and establish a new, more accurate equation for the TP; (2) determine the optimal NDSI thresholds for detecting snow pixels with FSC > 50% for the TP; (3) investigate scaling effects on the accuracy of both MODIS BSC and FSC estimation. For the first time, as many as 353 Landsat OLI scenes from across the TP are used to examine FSC and BSC estimates based on MODIS NDSI. To the best of our knowledge, this is also the first study to investigate scaling effects on FSC and BSC estimates from MODIS NDSI by making a comparison with upscaled Landsat-8 NDSI data. The paper is organized as follows: the data used, including MODIS NDSI snow-cover data, Landsat-8 OLI reference data and other auxiliary data are described in Section 2; The methods used for FSC and BSC estimation based on MODIS NDSI and upscaled Landsat-8 NDSI, and scale influence are presented in Section 3; Results concerning the accuracy of the new regression model and also new optimal NDSI thresholds for FSC and BSC estimates made using MODIS NDSI are presented in Section 4, along with a discussion of the influence of scaling factors, uncertainties and limitations.

## 2. Data

### 2.1. MODIS NDSI snow-cover data

We use the “MODIS/Terra Snow Cover Daily L3 Global 500 m SIN Grid, Version 6” data sets because Terra and Landsat-8 have similar overpass times. The data used are for the period 2013–2015. The MODIS NDSI data contains two layers: “NDSI\_Snow\_Cover” and raw NDSI. MODIS NDSI ( $NDSI_{MOD}$ ) is calculated following equation (1).

$$NDSI_{MOD} = (M4 - M6)/(M4 + M6) \quad (1)$$

where M4 is the MODIS visible light band 4 (wavelength: 0.545–0.565  $\mu\text{m}$ ) and M6 is the shortwave infrared band 6 (1.628–1.652  $\mu\text{m}$ ). NDSI\_Snow\_Cover only contains NDSI values in the range 0.1–1, since values less than 0.1 are converted to 0, indicating non-snow. The raw NDSI data layer includes the entire range of NDSI values, including negative values. Since the original empirical equation, used in previous MODIS snow products, was developed based on the entire range of NDSI (Salomonson and Appel, 2004, 2006), the raw NDSI values are used here. It should be noted that, to reduce uncertainties, we have excluded all NDSI pixels that did not pass the screening tests of the product due to low visible reflectance, high surface temperature, high shortwave infrared reflectance or low illumination conditions (Riggs et al., 2016a).

### 2.2. Landsat 8 OLI reference data

Landsat 8 OLI data are used as the FSC “truth” following Kuter et al. (2018) and Hou et al. (2020). These data have a relatively high spatial resolution of about 30 m but also a relatively long revisit time of about 16 days. Initially, all the available Landsat 8 Level 1 T scenes with cloud cover <10% during 2013–2015 were downloaded from the USGS Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA). These are provided as top-of-atmosphere (TOA) reflectance through the online service of ESPA. They are then classified as snow or non-snow pixels based on the SNOMAP algorithm (Hall et al., 1995). The NDSI for each 30-m Landsat-8 pixel ( $NDSI_{L8\_30m}$ ) is initially calculated according to Equation (2):

$$NDSI_{L8\_30m} = (B3_{L8\_30m} - B6_{L8\_30m}) / (B3_{L8\_30m} + B6_{L8\_30m}) \quad (2)$$

where  $B3_{L8\_30m}$  is Landsat-8 OLI Band 3 (green, wavelength: 0.525–0.600  $\mu\text{m}$ ) and  $B6_{L8\_30m}$  is Band 6 (shortwave near-infrared, wavelength: 1.560–1.660  $\mu\text{m}$ ). It should be noted that the subscript “\_30m” is used to distinguish the band or NDSI at the original resolution of 30 m from that at the upscaled resolution of 500 m. The pixels with  $NDSI > 0.4$ ,  $Band\ 3 > 0.1$  and  $Band\ 5 > 0.11$  are classified as 30-m snow pixels. The FSC “truth” is calculated for each 500-m MODIS NDSI pixel, by counting the number of 30 m snow pixels contained in the corresponding MODIS pixel: the method used in the development of the MODIS FSC algorithm (Salomonson and Appel, 2004, 2006). Due to the difficulty inherent in distinguishing snow from clouds in Landsat 8 OLI data (Stillinger et al., 2019), all the “cloudy” 500 m pixels (those with >5% of the 30 m pixels flagged as “cloud” or “cloud shadow”) are eliminated, to reduce potential errors. In total 353 scenes were selected: 201 scenes from 2013 to 2014, used for training the empirical relationship, and 152 scenes from 2015, used for validation (Fig. 1).

### 2.3. Auxiliary data

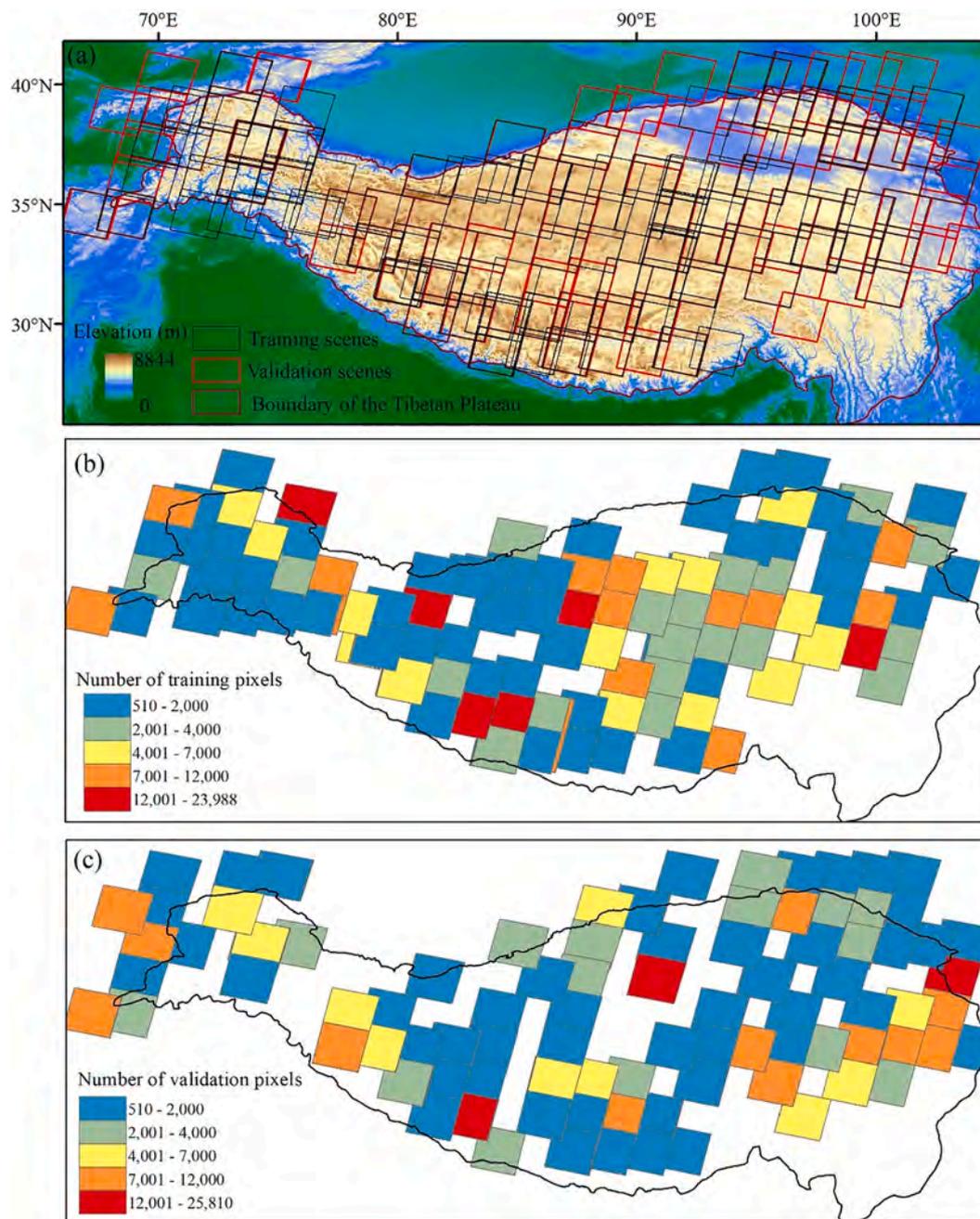
The land cover type data were obtained from the MODIS Land Cover Product (MCD12Q1) Collection 6. We use the International Geosphere-Biosphere Programme (IGBP) classification schemes. The spatial distribution of the major land-cover types of the TP is shown in Fig. 2. At the pixel scale, the spectral characteristics of snow may be weakened due to being mixed with the characteristics of other land-cover types. In particular, the dense canopy of forests may obstruct snow signals, and errors in FSC estimation have been found to be significantly correlated with vegetation fraction (Rittger et al., 2013). Since the normalized difference vegetation index (NDVI) is closely related to vegetation fraction (Baret et al., 1995), NDVI is also derived from the Landsat-8 OLI data, to evaluate the effects of land-cover types. For each 30-m Landsat-8 pixel, NDVI is initially calculated from equation (3):

$$NDVI_{L8\_30m} = (B5_{L8\_30m} - B4_{L8\_30m}) / (B5_{L8\_30m} + B4_{L8\_30m}) \quad (3)$$

where  $B5_{L8\_30m}$  is Landsat-8 OLI Band 5 (near-infrared) and  $B4_{L8\_30m}$  is the Band 4 (red). The NDVI at MODIS scale is then calculated by averaging all the  $NDVI_{L8\_30m}$  pixels contained in the corresponding MODIS pixel. Elevation information is obtained from Version 3 of the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM) data, which has a similar spatial resolution ( $\sim 30$  m) to Landsat 8 OLI. The elevation at the MODIS scale is then calculated by aggregating the original 30-m DEM pixels contained in the corresponding MODIS pixel. Finally, slope and aspect are calculated based on the aggregated elevation data at the MODIS scale.

## 3. Methodology

Fig. 3 presents an overview of the input data, the processes for using MODIS NDSI and upscaled Landsat-8 NDSI for establishing different FSC



**Fig. 1.** Map of the study area (a), and the location and sample number of MODIS-scale pixels with “observed” snow cover fraction between 0.1 and 0.95 for the 201 training Landsat-8 scenes (b) and the 152 validation Landsat-8 scenes (c). The elevation data displayed in (a) are from the Shuttle Radar Topography Mission (SRTM) dataset.

and BSC estimation methods, and how the scale-influence analysis is conducted. Firstly, a total of 353 Landsat-8 scenes with 30 m resolution are processed to 500-m FSC pixels and subsequently used as “truth” for training and validating the BSC and FSC estimates derived from MODIS NDSI. The Landsat-8 images are also converted to 500-m NDSI pixels for constructing the Landsat-8 based BSC and FSC schemes which are then compared with the MODIS-based schemes for analyzing the scaling effects. The following sections describe these processes in detail.

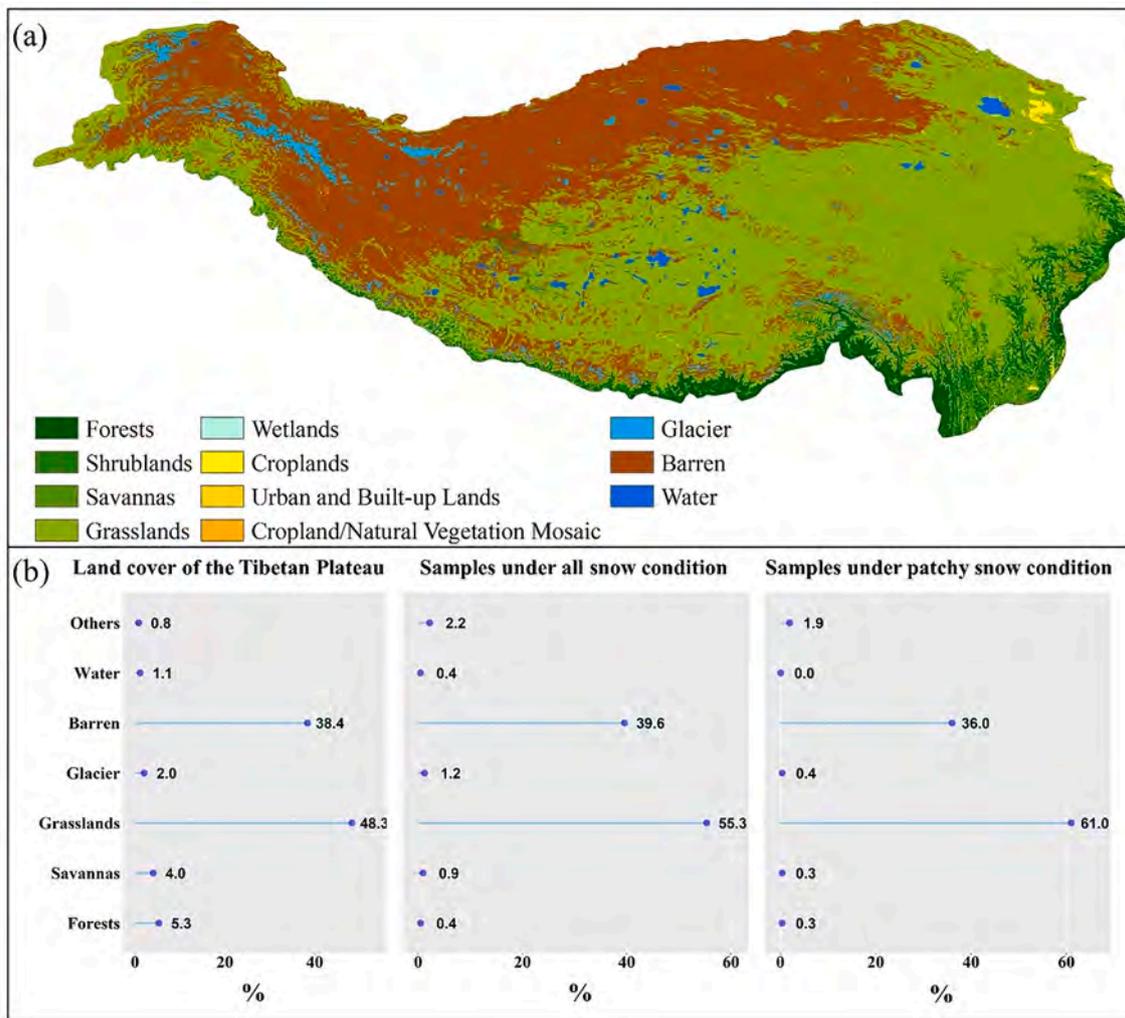
### 3.1. FSC estimation

#### 3.1.1. Using MODIS NDSI for FSC estimation

Constructing a linear regression between the 500-m  $NDSI_{MOD}$  and the FSC derived from 30-m Landsat pixels within the corresponding

MODIS pixel is a popular way to estimate FSC, which was also used for the MODIS snow-cover product Version 5 (Riggs et al., 2006). According to Salomonson and Appel (2006), only MODIS pixels with a true FSC in the range of 10 – 95% are considered for building the regression relationship to reduce the negative effects of multiple samples clustering at very low or very high FSC. All the 353 selected Landsat 8 scenes have at least 500 corresponding 500-m MODIS pixels meeting the criteria. More detailed information is shown in Fig. 1.

According to Salomonson and Appel (2004), there are generally two ways of building such an empirical relationship: “model MA”, in which  $NDSI_{MOD}$  is taken as the independent variable and FSC as the dependent variable; and “model MB” where FSC is the independent variable and  $NDSI_{MOD}$  is the response. It is clear that the equation for MB has to be inverted algebraically to estimate FSC from NDSI, but this approach is



**Fig. 2.** The spatial distribution (a) and percentage (b) of major land cover types on the Tibetan Plateau. The percentage of major land-cover types for the samples used in this study including those under all-snow and patchy-snow conditions, are shown in panel (b). Note: some types have been combined for brevity, e.g. “Woody Savannas” and “Savannas” are combined as “Savannas”. The land-cover types are derived from MODIS Land Cover Type Product (MCD12Q1) Collection 6.

considered to be better than MA because of the larger variance and errors in NDSI compared to FSC (Salomonson and Appel, 2004, 2006). For each Landsat-8 training scene, both models, MA and MB, are built. Both models have 201 sets of coefficients derived from the 201 training scenes. Using samples from all the 201 training scenes, two average (or “universal”) linear relationships are established, called “model MAU” and “model MBU”. The reference linear regression model, which was used for producing the global MODIS FSC developed by Salomonson and Appel (2006), is also tested here for comparison, and named “model MR”. Considering the meaningful ranges of NDSI and FSC, all estimated FSC < 0 or those with NDSI ≤ 0 are set to 0% FSC, and those with values > 1 are set to 100% FSC.

In total, five types of model (MA, MB, MAU, MBU and MR) are built based on NDSI<sub>MOD</sub>. Of these models, the one with the highest validation accuracy is considered to be the best model that can be achieved through a linear regression between NDSI<sub>MOD</sub> and FSC.

### 3.1.2. Upscaling Landsat 8 NDSI to MODIS scale for FSC estimation

To further explore the effects of scaling on the accuracy of the linear regression models based on NDSI, two types of upscaled Landsat-8 NDSI are constructed: NDSI<sub>L8M</sub>, a mean NDSI, calculated for each MODIS pixel by aggregating all the corresponding Landsat-8 NDSI (i.e., NDSI<sub>L8,30m</sub>) values; and NDSI<sub>L8I</sub>, an NDSI calculation made using equation (4),

$$NDSI_{L8I} = (B3_{L8,500m} - B6_{L8,500m}) / (B3_{L8,500m} + B6_{L8,500m}) \quad (4)$$

where the MODIS-scale B3 (B3<sub>L8,500m</sub>) and B6 (B6<sub>L8,500m</sub>) are initially calculated by averaging the TOA reflectance of the 30 m Landsat B3<sub>L8,30m</sub> and B6<sub>L8,30m</sub> pixels contained in the corresponding MODIS pixel, respectively. For NDSI<sub>L8M</sub>, four types of model “model LAM”, “model LBM”, “model LAMU” and “model LBMU” are constructed by replacing NDSI<sub>MOD</sub> with NDSI<sub>L8M</sub> in models MA, MB, MAU and MBU. Similarly, the other four types of model: “model LAI”, “model LBI”, “model LAIU” and “model LBIU” are constructed with NDSI<sub>L8I</sub> as the input. A comparison of the NDSI<sub>L8M</sub>-based and NDSI<sub>L8I</sub>-based models illustrates the effects of the different methods of NDSI calculation on the upscaling.

### 3.1.3. Evaluation of the accuracy of FSC estimation

Previous studies have shown that the unbalanced distribution of samples (e.g. the huge number of 0% FSC pixels) may inflate the evaluated accuracy (Zhang et al., 2019b). Two validation conditions are thus considered in this study: 1) the “all-snow condition”, in which all the samples covering the entire range of FSC between 0 and 1 are used for validation; 2) the “patchy-snow condition”, in which only FSC samples between 0.1 and 0.95 are used for validation. To evaluate the accuracy of the FSC estimates, both RMSD and R are calculated by comparing the estimated and “true” FSC values. In accordance with previous FSC

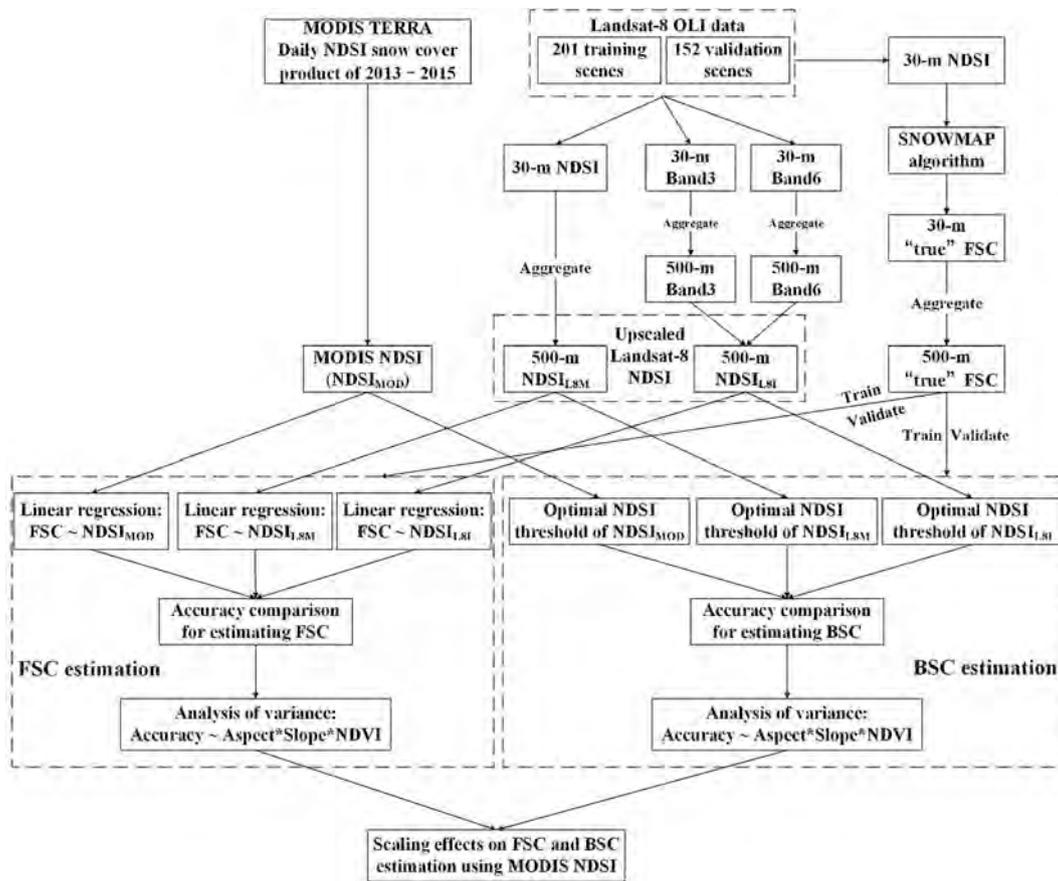


Fig. 3. Flow chart illustrating the processes for establishing new fractional (FSC) and binary (BSC) snow-cover estimation methods based on MODIS NDSI and upscaled Landsat-8 NDSI, and how the scaling effects are analyzed.

estimates (Salomonson and Appel, 2004; Shreve et al., 2009), RMSD is selected as the primary evaluation metric for the deterministic comparison of the different models. The model with the highest accuracy, out of models MA, MB, MAU, MBU and MR, is selected as the best linear regression model for estimating FSC based on MODIS NDSI. The same comparison is conducted for the models based on upscaled Landsat NDSI, i.e. models LA, LB, LAIU, LBIU, LAMU and LBMU.

### 3.1.4. Analysis of scaling effects on FSC estimation

NDSI<sub>L8M</sub> and NDSI<sub>L8I</sub> cannot be directly compared with NDSI<sub>MOD</sub> because they involve different band locations and sensors, but their performances in estimating FSC are comparable, because they both make the same assumption of a strong linear relationship between FSC and NDSI. The spectral radiance is generally linear when aggregated from relatively high (30 m) to relatively low (e.g. 1 km) resolutions (Liang, 2000) and NDSI is calculated using simple band mathematics. Therefore, we can assume that, if the scaling effects on the FSC estimation using MODIS NDSI are small, the models using NDSI<sub>MOD</sub> should have a similar accuracy to those using NDSI<sub>L8I</sub> or NDSI<sub>L8M</sub>, although, as discussed below, the upscaling processes for constructing NDSI<sub>L8M</sub> or NDSI<sub>L8I</sub> may also be affected to some degree by pixel heterogeneity.

In mountainous areas, land-surface radiance is dependent on topography and land cover (Meyer et al., 1993), and so three factors, aspect, slope and NDVI, have been selected to describe the varying topography and land-cover conditions (Czyzowska-Wisniewski et al., 2015; Hou et al., 2020; Liang et al., 2017). If there is a significant difference in accuracy between the models based on NDSI<sub>MOD</sub> and those based on NDSI<sub>L8I</sub> or NDSI<sub>L8M</sub>, there may be strong scaling effects present. Such scale influence can be analyzed by investigating how the accuracy of the NDSI<sub>MOD</sub>-based models, and its differences compared to

models using NDSI<sub>L8I</sub> or NDSI<sub>L8M</sub>, are affected by these three factors.

The analysis of variance (ANOVA) is a common technique used for calculating the contributions of different factors (Zhang et al., 2019a), and is used here to quantify the effects of the three scaling factors. Since aspect generally has eight classifications, slopes and NDVI have also been classified into eight groups according to their respective histogram distributions. The eight slope groups are 0–2, 2–4, 4–6, 6–8, 8–10, 10–12, 12–14 and >14 degrees. The eight NDVI groups are <0, 0–0.015, 0.015–0.03, 0.03–0.045, 0.045–0.06, 0.06–0.075, 0.075–0.09 and >0.09. Thus, the samples are divided into a total of 512 (=8 × 8 × 8) groups representing different aspect, slope and NDVI conditions. For each group, the R between the estimated and “true” FSC is calculated. The variance of the estimation accuracy (i.e. R) can thus be partitioned into different components related to different factors. The interactions between different factors are also considered in ANOVA. The contributions of individual factors, or their interaction terms, are calculated based on their sum of squares. It should be noted that the use of R rather than RMSD as the accuracy measurement for ANOVA, is mainly because RMSD is very susceptible to the magnitude of FSC and, as the differences in the FSC of different groups can be very large, use of RMSD could lead to unreliable results.

## 3.2. BSC estimation

### 3.2.1. Using MODIS NDSI and upscaled Landsat-8 NDSI for BSC estimation

MODIS binary snow-cover data were previously produced using the NDSI threshold of 0.4. All the 500 m “true” FSC pixels are reclassified as either snow (FSC > 50%) or non-snow (FSC ≤ 50%) pixels. Previous studies have shown that 0.4 may not be the optimal NDSI threshold for

many areas, including the TP (Zhang et al., 2020, 2019b). Two types of NDSI threshold are considered here: 1) a locally optimal NDSI threshold, for which a sensitivity test of different NDSI thresholds (ranging from 0.1 to 0.7) on BSC estimation accuracy is implemented to find an optimal NDSI threshold for each training scene, following Zhang et al. (2019b); 2) a region-wide optimal NDSI threshold, for which a similar sensitivity test, but based on samples from all 201 training scenes, is conducted, with the NDSI threshold that achieves the highest average accuracy selected. Both types of NDSI threshold are calculated for  $NDSI_{MOD}$ ,  $NDSI_{L8I}$  and  $NDSI_{L8M}$ , respectively.

### 3.2.2. Evaluation of BSC estimation accuracy

To evaluate BSC estimation accuracy, firstly a confusion matrix is defined: 1) true FSC > 0.5 and satellite predicts snow-covered (i.e.  $NDSI > NDSI$  threshold), defined as “TP” (true positive); 2) true FSC > 0.5 while satellite predicts snow-free (i.e.  $NDSI \leq NDSI$  threshold), defined as “FN” (false negative); 3) true FSC  $\leq$  0.5 while satellite predicts snow-covered, defined as “FP” (false positive); 4) true FSC  $\leq$  0.5 and satellite predicts snow-free, defined as “TN” (true negative). Four types of evaluation metrics consisting of Probability of Detection (POD), Precision (PC), F-score (FS) and Cohen’s Kappa (CK) are then calculated using equations (5) – (8), respectively.

$$POD = \frac{TP}{TP + FN} \quad (5)$$

$$PC = \frac{TP}{TP + FP} \quad (6)$$

$$FS = \frac{2 \times POD \times PC}{POD + PC} \quad (7)$$

$$CK = \frac{GA - Pr(e)}{1 - Pr(e)} \quad (8)$$

where GA and Pr(e) are calculated from Equations (9) – (11).

$$Pr(e) = \left( \frac{TP + FP}{Total} \times \frac{TP + FN}{Total} \right) + \left( \frac{TN + FP}{Total} \times \frac{TN + FN}{Total} \right) \quad (9)$$

$$GA = \frac{TP + TN}{Total} \quad (10)$$

$$Total = TP + FN + FP + TN \quad (11)$$

CK is selected as the primary evaluation metric in deterministic comparisons such as selecting the optimal NDSI thresholds, following Zhang et al. (2019b). FS, POD and PC are calculated to enable better understanding of the error sources.

### 3.2.3. Analysis of scaling effects on BSC estimation

The analysis of scaling effects on BSC estimation is similar to the procedure described in Section 3.1.4. Firstly, the accuracy of the methods using MODIS NDSI (i.e.  $NDSI_{MOD}$ ) and upscaled Landsat-8 NDSI (i.e.  $NDSI_{L8I}$  and  $NDSI_{L8M}$ ) is compared. If, for example, using  $NDSI_{MOD}$  shows significantly lower accuracy than using  $NDSI_{L8I}$  or  $NDSI_{L8M}$ , then there could be large scaling effects on the BSC estimates based on  $NDSI_{MOD}$ . The effects of the three scaling factors, aspect, slope and NDVI are then analyzed as follows. The same 512 sample groups of different aspect, slope and NDVI conditions as used for FSC estimation are used. For each group, the estimation accuracy (CK) is calculated, based on the estimated and “true” BSC. The effects of different scaling factors are further quantified based on ANOVA of the estimation accuracy using MODIS NDSI and upscaled Landsat-8 NDSI or their accuracy differences.

## 4. Results and discussion

### 4.1. The performance of FSC estimation models based on MODIS NDSI

For all the 201 training scenes, the RMSD of model MA is always smaller than that of model MR, as shown in Fig. 4a. The average RMSDs of the model MA is 0.18 compared to 0.24 for model MR. This result confirms that a local empirical relationship can be more suitable for a specific region than the global reference equation. However, an individual scene-based model, which has relatively good accuracy with the training data, may not be suitable for the whole TP. Therefore, the best performing of the 201 MA models, referred to as “model MAT”, is determined by comparing their validation accuracy in conditions of both all-snow and patchy-snow. Similarly, the best model MB is also selected, and referred to as “model MBT”. Although model MB was expected to be better than MA, both models MA and MAT show generally better performance than models MB and MBT in the training and validation scenes, respectively (Fig. 4). Models MAT and MR show almost equal accuracy in the validation under all-snow condition. However, the validation accuracy of MAT is clearly better than that of MR under patchy-snow conditions. Although models MAU and MBU both consider all the training samples, possibly making them more “universal” (Salomonson and Appel, 2004), their validation accuracy is significantly lower than MAT under either all-snow or patchy-snow conditions (Fig. 4b). Thus, model MAT is considered to be the best out of all the FSC estimation models based on  $NDSI_{MOD}$  used here.

The validation accuracy of MAT under all-snow conditions is relatively good, with an RMSD as low as  $\sim$ 0.09, even lower than the values of  $\sim$ 0.1 reported by Salomonson and Appel (2006) and Salomonson and Appel (2004). However, this high validation accuracy may be mainly because of the very unbalanced distribution of validation data with snow-free pixels (i.e. FSC = 0%) accounting for  $\sim$ 73% of the data. MAT shows much lower accuracy (RMSD: 0.22) under patchy-snow conditions. This finding may also explain why some local studies found MODIS fractional snow-cover data was accurate in the Heihe river basin of the northeastern TP, with relatively low RMSD values of 0.09 – 0.11 (Hou and Huang, 2014), while others, such as Hou et al. (2020) and Tang et al. (2013b), reported large RMSD values greater than 0.2. As the validation results under patchy-snow conditions are clearly more reliable, the ANOVA is conducted under these conditions.

### 4.2. Scaling effects on FSC estimation

The comparison between models based on upscaled Landsat-8 NDSI shows that the “universal” models that use all the training samples (i.e. models LAIU, LBIU, LAMU and LBMU) have slightly better accuracy than the models which use samples from individual scenes (i.e. models LAI, LBI, LAM and LBM), although the results are not shown here. Fig. 5a shows that for all the training Landsat-8 scenes, the RMSDs of models LAI (mean RMSD: 0.1) and LAM (mean RMSD: 0.09) are clearly lower than that of model MA (mean RMSD: 0.18). The validation results for models LAIU, LBIU, LAMU and LBMU also show significantly better accuracy than model MAT under both all-snow and patchy-snow conditions (Fig. 5b). The large difference in accuracy between models based on  $NDSI_{MOD}$  and those based on  $NDSI_{L8I}$  and  $NDSI_{L8M}$  indicates there may be strong scaling effects on FSC estimation using MODIS NDSI.

Although  $NDSI_{L8I}$  and  $NDSI_{L8M}$  are calculated from  $NDSI_{L8,30m}$  in two different ways, their performances in the FSC estimation are quite similar, with little difference in accuracy, as shown in Fig. 5. As  $NDSI_{L8I}$  is constructed using a method closer to that used for  $NDSI_{MOD}$ , the ANOVA is conducted based on results from MAT and LBIU. The ANOVA test shows that, although the single contribution of aspect, slope or NDVI to the variances of the accuracy of model MAT (i.e.  $R_{MAT}$ ) is not very large, their combined contributions (including the interaction terms) can be as large as 61%, indicating a dominant scaling effect (Fig. 6a). Although it seems that the scale issue also has effects on FSC estimation

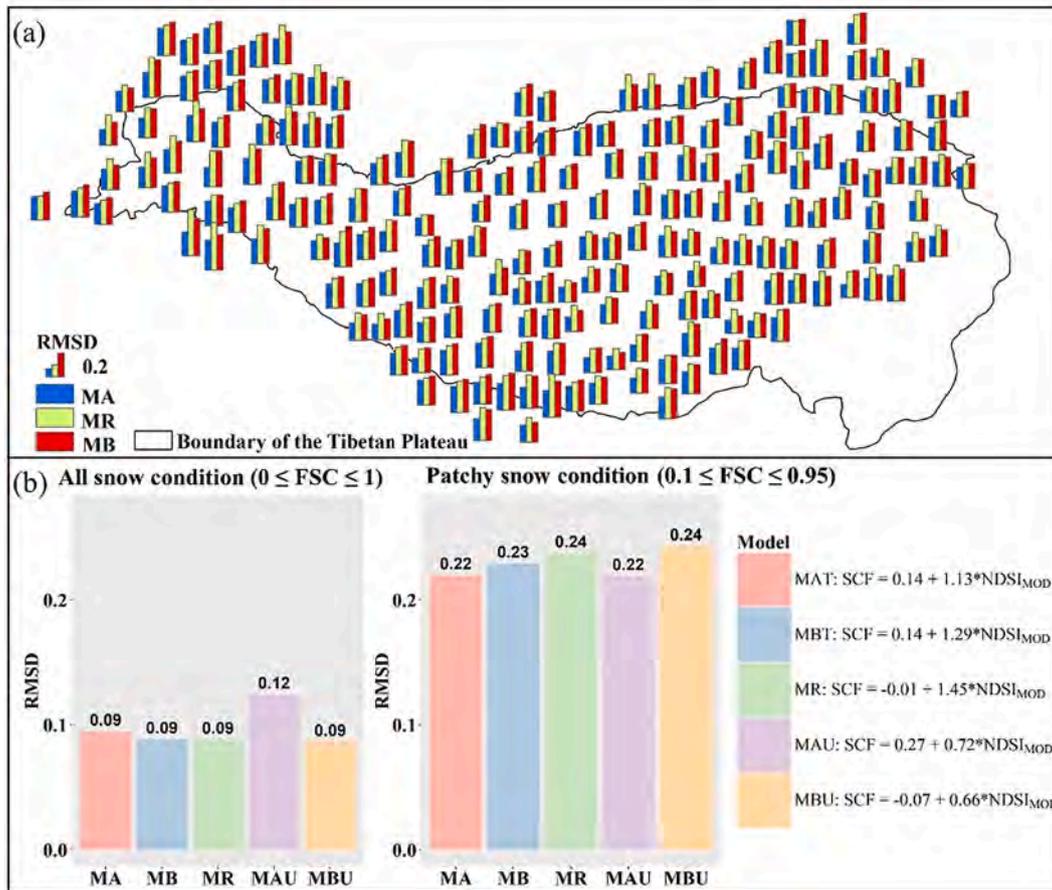


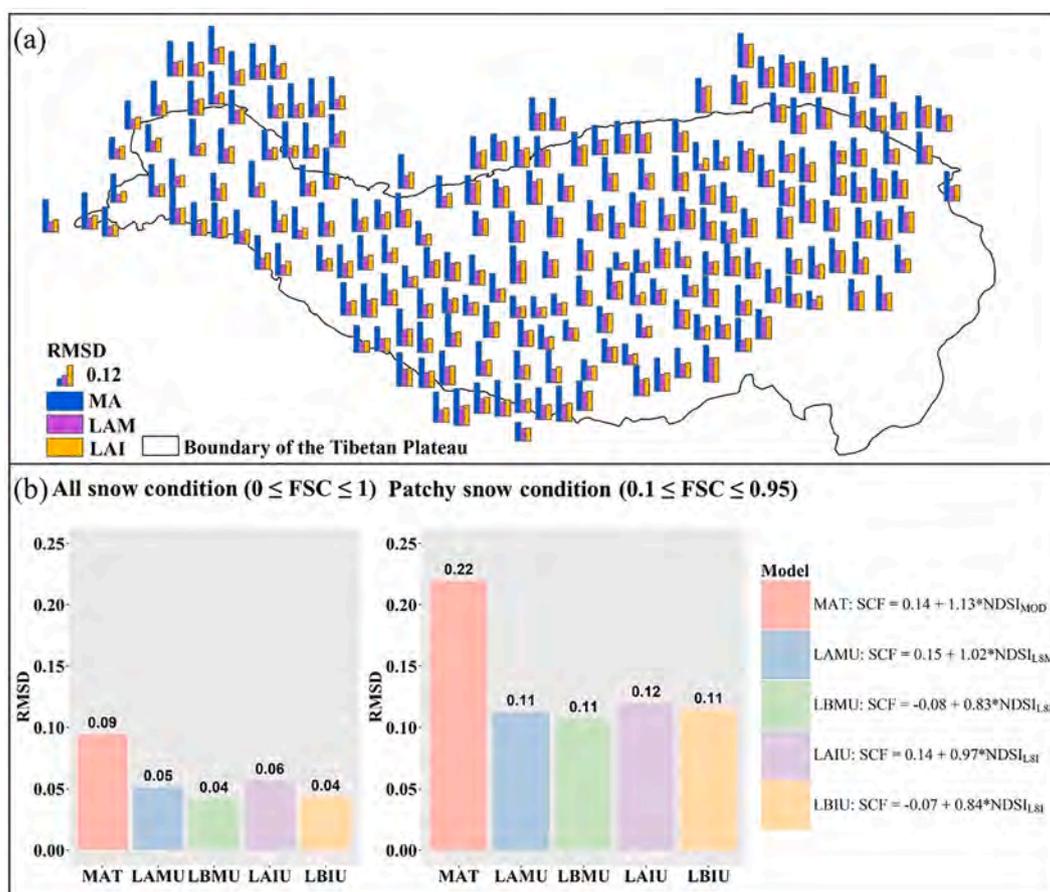
Fig. 4. Accuracy comparison of models based on MODIS NDSI (i.e.  $NDSI_{MOD}$ ). The comparisons of RMSD for models MA, MR, and MB are shown for each of the 201 training Landsat scenes separately (a). The accuracy comparison between models of the best MA, the best MB, MR, MAU and MBU is plotted using RMSD calculated using samples from the 152 validation scenes (b). Note: the locations of some scenes have been adjusted slightly to prevent overlapping in panel (a). MAT: the best model among the 201 MA models; MBT: the best model among the 201 MB models.

using  $NDSI_{L8I}$ , the combined effects of aspect, slope, NDVI and their interactions on the accuracy of model LBIU (i.e.  $R_{LBIU}$ ) are much smaller, making a total contribution of 41%. Thus, a comparison of Fig. 6a and 6b shows that there are enhanced scaling effects on FSC estimation based on  $NDSI_{MOD}$ . It is also interesting that NDVI is a major scaling factor of the variance of  $R_{LBIU}$  (Fig. 6b), whereas aspect and slope are obviously more important factors of the variance of  $R_{MAT}$ . Thus, the enhanced scaling effects on MODIS NDSI may be mainly caused by the complex topography rather than land cover heterogeneity. The large difference in FSC estimations between  $NDSI_{MOD}$  and  $NDSI_{L8I}$  also indicates that upscaling Landsat-8 NDSI to 500-m cannot mimic the true MODIS NDSI well. The linear aggregation process of Landsat-8 NDSI may not be able to reflect the complex effects of terrains on the possibly non-linearly scaling behavior of NDSI and the linear relationship between NDSI and FSC may be degraded for MODIS due to such scaling effects. For example, the illumination conditions within a MODIS pixel may be greatly affected by aspect and slope which could further influence the spectral radiance (Crawford et al., 2013; Mishra et al., 2009). We further conduct an ANOVA of the accuracy difference between the two models (i.e.  $R_{LBIU}$  minus  $R_{MAT}$ ). As mentioned before,  $R_{LBIU}$  is always better than  $R_{MAT}$ , and the ANOVA result shows that the superiority of  $R_{LBIU}$  is strongly controlled by the two terrain factors of aspect and slope (Fig. 6c). This result further demonstrates that complex terrain is the main factor lowering the skill of MODIS NDSI for snow-cover estimation on the TP.

#### 4.3. Optimal NDSI threshold for binary snow cover (FSC ≥ 50%) estimation

Using locally optimal NDSI thresholds and the same training samples as those for FSC estimation, both  $NDSI_{L8I}$  (mean CK: 0.89) and  $NDSI_{L8M}$  (mean CK: 0.91) show much higher accuracy than  $NDSI_{MOD}$  (mean CK: 0.55) (Fig. 7a). The accuracy of BSC estimation is also strongly affected by the unbalanced distribution of samples. When all the samples ( $0 \leq FSC \leq 1$ ) of the 201 training scenes are used, the spatially averaged values of CK for  $NDSI_{MOD}$ ,  $NDSI_{L8M}$  and  $NDSI_{L8I}$  are greatly increased to 0.88, 0.98 and 0.98, respectively. Fig. 7b illustrates the sensitivity tests of the averaged BSC estimation accuracy on a region-wide NDSI threshold ranging from 0.1 to 0.7. The maximum accuracy corresponds to NDSI thresholds (i.e. region-wide optimal NDSI thresholds) of 0.29, 0.38 and 0.39 for  $NDSI_{MOD}$ ,  $NDSI_{L8M}$  and  $NDSI_{L8I}$ , respectively. The region-wide optimal thresholds are further used in validation and the results shown in Fig. 7c show that the spatially averaged values of CK for  $NDSI_{MOD}$ ,  $NDSI_{L8M}$  and  $NDSI_{L8I}$  are 0.49, 0.90 and 0.88 under patchy-snow conditions, and they are greatly increased to 0.85, 0.98, 0.98, under all-snow conditions. The spatially averaged values of FS for  $NDSI_{MOD}$ ,  $NDSI_{L8M}$  and  $NDSI_{L8I}$  vary similarly, with values of 0.63, 0.93 and 0.91, respectively, under patchy-snow conditions, and 0.87, 0.98, 0.98 under all-snow conditions. The much lower BSC estimation accuracy of  $NDSI_{MOD}$ , compared with that of  $NDSI_{L8I}$  or  $NDSI_{L8M}$ , in both training and validation situations, indicates that there are also possibly large scaling effects on the BSC estimates based on MODIS NDSI.

Similarly to the case for FSC estimation, the accuracy of BSC estimation using  $NDSI_{L8I}$  is very close to that achieved when using  $NDSI_{L8M}$ .



**Fig. 5.** Accuracy comparison between MODIS-based and upscaled Landsat-based FSC estimation models. The comparisons of RMSD for models MA, LAM, and LAI are shown for each of the 201 training Landsat scenes separately (a). The accuracy comparison between the MAT, LAMU, LBMU, LAIU and LBIU models is plotted using RMSD calculated using samples from the 152 validation scenes (b). Note: the locations of some scenes have been adjusted slightly to prevent overlapping in panel (a).

The results of  $NDSI_{MOD}$  and  $NDSI_{L8I}$  are thus used in the ANOVA for BSC estimation. The ANOVA of scaling effects is conducted for patchy-snow conditions because of the negative effects of the unbalanced samples distribution under all-snow conditions. The results are highly consistent with those of the FSC estimation. The total contribution of the three scaling factors, aspect, slope, NDVI and their interaction terms, to the variances of BSC estimation accuracy using  $NDSI_{MOD}$  is  $\sim 61\%$  (Fig. 8a), very close to the values for FSC estimation (Fig. 6a), although the effect of slope seems to be larger for BSC estimation. The scaling effects on  $NDSI_{L8I}$ -based BSC estimation are smaller than those on  $NDSI_{L8I}$ -based FSC estimation, with the total contribution of the three scaling factors being only  $\sim 33\%$  (Fig. 8b). Aspect and slope are the two major factors of the enhanced scaling effects on BSC estimation using  $NDSI_{MOD}$ , with their combined contribution to the variances of the accuracy differences between the two methods (i.e.  $CK_{L8I}$  minus  $CK_{MOD}$ ) being  $\sim 50\%$  (Fig. 8c).

#### 4.4. Implications of the use of the MODIS NDSI snow cover product for FSC and BSC estimation on the Tibetan Plateau

The unbalanced distribution of samples (i.e. most samples having 0% FSC) under all-snow conditions means that the accuracy of FSC estimation is overestimated by using a linear regression model based on MODIS NDSI. Under patchy-snow conditions (i.e.  $0.1 \leq FSC \leq 0.95$ ) the accuracy is much lower. Even using local regression models, the averaged RMSD across the 201 training scenes is still as high as 0.18. It should be noted that model MR, which was used in the previous version of the MODIS snow cover product, shows even lower accuracy under patchy-snow conditions with an averaged RMSD of 0.24. Our study

proposes a new regression model (i.e. model MAT) based on  $NDSI_{MOD}$  selected from the candidate models derived from 201 Landsat-8 scenes. However, the accuracy improvement when using this new model, is very limited, with the validation RMSD still as high as 0.22. In contrast, the models using upscaled Landsat-8 NDSI perform much better, with an averaged validation RMSD of 0.11–0.12. Thus, the ability of MODIS NDSI to estimate FSC seems to be weakened and there could be larger uncertainties than expected when using FSC values estimated from MODIS NDSI on the TP. The ANOVA results indicate that more importance should be attached to terrain factors, including aspect and slope, in future FSC estimations. Considering that the linearity between FSC and MODIS NDSI may be greatly degraded by scaling effects, more complicated methods that can incorporate more spectral bands and factors such as MODSCAG (MODIS Snow-Covered Area and Grain size) (Painter et al., 2009), artificial neural networks (Hou and Huang, 2014; Hou et al., 2020), multivariate adaptive regression splines (Kuter et al., 2018) and random forests (Liu et al., 2020), should be used to create more accurate FSC products for the TP.

For BSC estimation, we find that 0.29 could be the optimal NDSI threshold for BSC mapping on the TP using MODIS NDSI data, a value which seems to be in conflict with some previous studies, which suggest, based on daily snow depth observations, a NDSI threshold of 0.1 for use in China, including the TP (Zhang et al., 2020, 2019b). It should be noted that point observations of snow depth may not mean that the corresponding MODIS pixel has a value of  $FSC > 50\%$ . In addition to the obviously different type of data used as “observations”, another important reason for the discrepancy could be that the NDSI threshold of 0.1 may be more accurate for describing whether snow is present within a MODIS pixel, as indicated by (Riggs et al., 2017), rather than whether

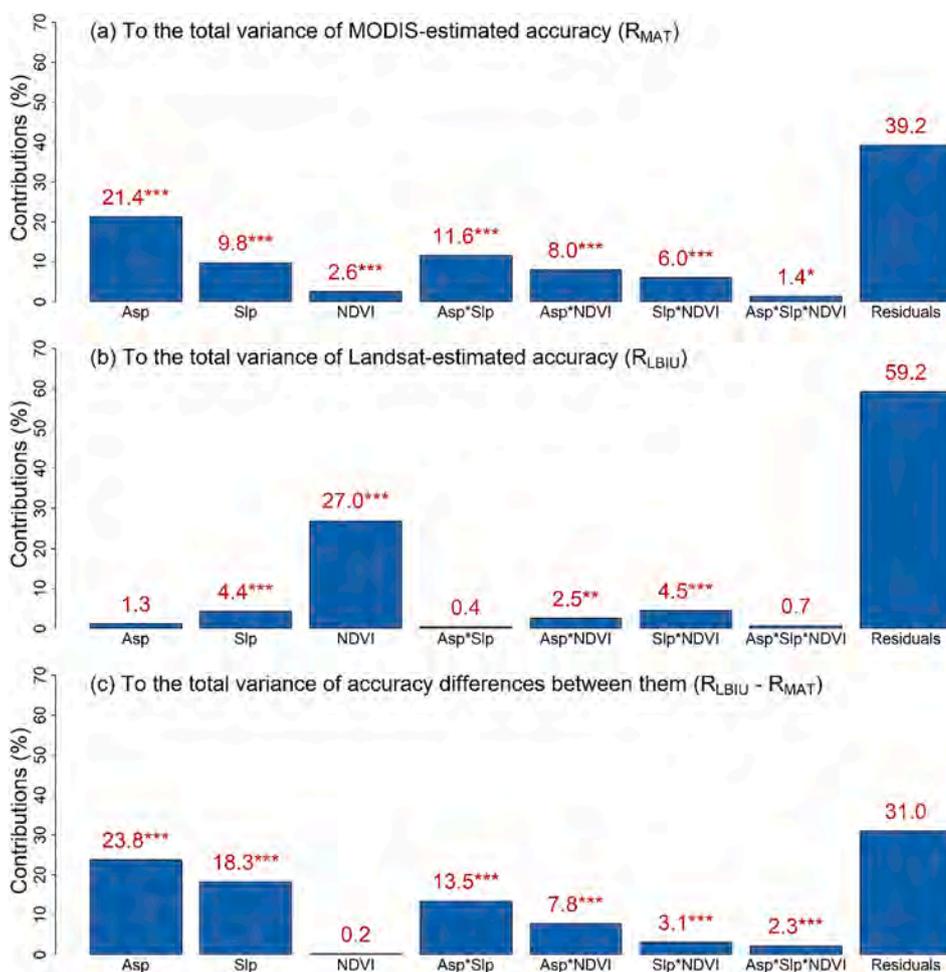


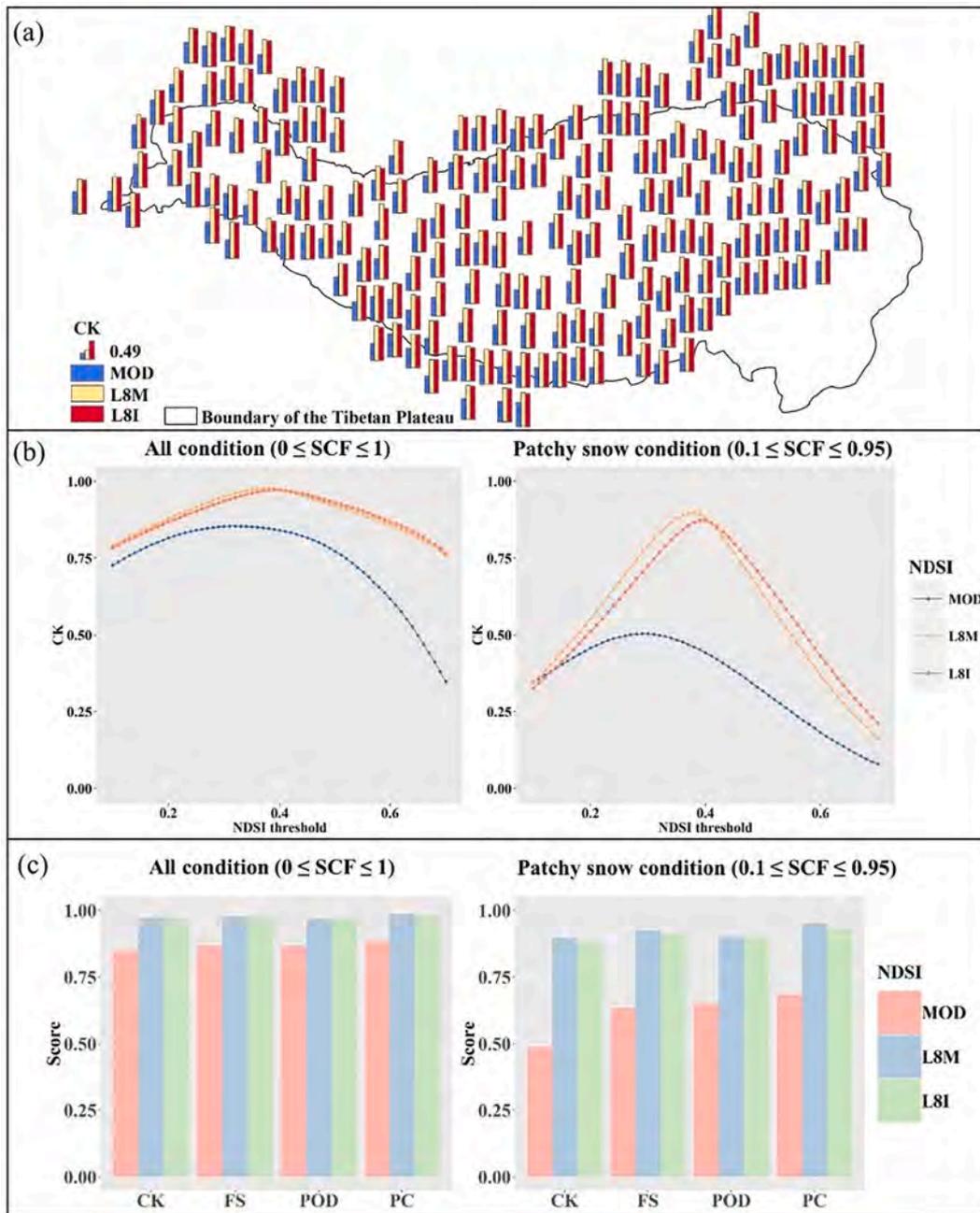
Fig. 6. The ANOVA results showing the contributions of the three scaling factors (i.e. aspect, slope and NDVI) and their interaction terms to total variances of the accuracy for the model using MODIS NDSI (i.e. model MAT) (a), the model using upscaled Landsat-8 NDSI (i.e. model LBIU) (b) and the accuracy differences between the two models (c).  $R_{MAT}$  is the correlation coefficient between the true FSC and the value estimated from model MAT,  $R_{LBIU}$  is the correlation coefficient between the true FSC and the value estimated from model LBIU; Asp is aspect; Slp is slope; \*\*\* indicates  $P$ -value < 0.001; \* indicates  $P$ -value < 0.05.

snow covers more than 50% of the pixel. Therefore, we further tested more FSC situations of snow detection with the lower limit for the FSC of a snow-covered pixel varying from 0.1 to 0.5. For each FSC situation, the optimal NDSI threshold is selected as the one with the highest averaged CK. To ensure that the results of these tests are comparable with previous studies, all-snow conditions are used. The results, shown in Fig. 9, show that the optimal NDSI threshold decreases from 0.29 to 0.1, as the lower limit of FSC decreases from 50% to 5%. It is interesting to note that the BSC estimation accuracy also decreases as the lower limit of FSC decreases. This result indicates that MODIS NDSI snow-cover data are more accurate for detecting snow pixels with  $FSC > 50\%$ , possibly because larger scaling effects may exist in pixels with relatively low FSC. Such enhanced scaling effects, mainly caused by terrain factors, are also observed when estimating BSC using MODIS NDSI. Under fragmented or patchy-snow conditions, both omission and commission errors of BSC detection using MODIS NDSI are large, with the averaged POD and PC as low as 0.65 and 0.68, respectively, resulting in relatively low values of CK (0.49) and FS (0.63). Thus, special attention is required when extracting the boundaries of snow-covered areas, such as the snow line or glacier extent, based on MODIS snow-cover data, and more advanced methods, that can better consider scaling factors, should be developed for BSC estimation on the TP. In addition, previous studies which use 0.4 as the NDSI threshold for snow mapping, may benefit from our finding about the new optimal NDSI threshold of 0.29, considering that, under patchy-snow conditions, the accuracy of using 0.29 as the NDSI threshold (mean CK: 0.49) is clearly better than that of using 0.4 (mean CK: 0.40) when the POD as low as 0.46 indicates a lot of snow omission errors. It should also be noted that one can easily calculate the new FSC and BSC data according to the new regression equation and new

threshold using the original MODIS NDSI data. This may be one reason why MODIS now only provides NDSI data (Riggs et al., 2017).

#### 4.5. Uncertainties and limitations

The “true” FSC data are derived from Landsat-8 OLI scenes following the SNOMAP algorithm (Hall et al., 1995) that uses 0.4 as the NDSI threshold. However, Härer et al. (2018) found that, in some cases, a fixed NDSI threshold of 0.4 could result in relative errors of about 24% in FSC estimation. They suggest considering seasonal variation of NDSI thresholds, but the camera-based calibration method they use will be hard to apply on the TP, where regular observation stations are very sparse. Even though MODIS TERRA and Landsat-8 have similar overpass times, the differences in acquisition time between them could also affect the results because snow depths on the TP are relatively small and snow may melt or sublimate during the time interval (Zhang et al., 2020). To solve this problem, Hou et al. (2020) suggests using image pairs with the same acquisition time for training the FSC estimation model. It is assumed that in a reliable pair of Landsat-MODIS scenes, there should be a high degree of consistency between  $NDSI_{L8I_{500m}}$  and  $NDSI_{MOD}$ . We thus implemented relatively strict criteria for increasing the reliability of the reference scenes: the only scenes selected for training or validation in this study are those with a correlation coefficient between  $NDSI_{L8I_{500m}}$  and  $NDSI_{MOD} > 0.7$ . Due to the very unbalanced composition of land cover types (Fig. 2), the land cover effects considered in ANOVA are based on NDVI rather than land cover types. It should be noted that NDVI may be more accurate in representing the effects of vegetation rather than those of land-cover types. This problem could be partially solved by collecting more samples from different land-cover types in



**Fig. 7.** Accuracy comparison of binary snow cover ( $FSC \geq 50\%$ ) detection between  $NDSI_{MOD}$ ,  $NDSI_{L8M}$  and  $NDSI_{L8I}$ . (a) Accuracy comparison of locally optimal NDSI thresholds based on samples from each training Landsat-8 scene. (b) Sensitivity tests of binary snow-cover estimation accuracy on a region-wide NDSI threshold ranging from 0.1 to 0.7 under all-snow (left) and patchy-snow (right) conditions. (c) Accuracy comparison between  $NDSI_{MOD}$ ,  $NDSI_{L8M}$  and  $NDSI_{L8I}$  using the NDSI thresholds of 0.29, 0.38 and 0.39, respectively, based on samples from the 152 validation Landsat-8 scenes. MOD:  $NDSI_{MOD}$ ; L8M:  $NDSI_{L8M}$ ; L8I:  $NDSI_{L8I}$ .

future. The results of ANOVA show that NDVI has limited effects on the accuracy of FSC and BSC estimation using MODIS NDSI. However, as shown in Fig. 10, there are truly significant accuracy differences between different land cover groups. For both FSC and BSC estimation, the accuracy for forest and savanna is significantly lower than for barren land and grassland. It should be noted that forest and savanna account for only about 5% and 4% of the area of the whole plateau, respectively, and the samples we used generally have a composition of major land cover types which is consistent with the TP as a whole (Fig. 2b). It should also be noted that, according to Sulla-Menashe and Friedl (2018), the pixels labeled here as savanna, as determined from the MODIS land-cover data, are more likely to be “sparse forest” or unidentified grasslands, as savanna is usually only found in tropical regions. The relatively

weak effects of NDVI are thus attributed to the much smaller portion of densely vegetated areas.

Though a fixed NDSI threshold of 0.29 is recommended as the optimal one in the present study, the optimal NDSI threshold may, in fact, vary with slope and aspect. All the samples are thus divided into 64 ( $=8 \times 8$ ) groups by considering the eight slope conditions and eight types of aspect. For each group, a sensitivity test of different NDSI thresholds (ranging from 0.1 to 0.7) on BSC estimation accuracy, which is the same as that in Section 3.2.1, is conducted to determine an optimal NDSI threshold. In total, 64 slope/aspect specific optimal NDSI thresholds are obtained, as shown in Fig. 10c. NDVI is not considered here, mainly because the ANOVA tests show that NDVI is not a key factor for BSC estimation on the TP. It is clear that the optimal NDSI threshold

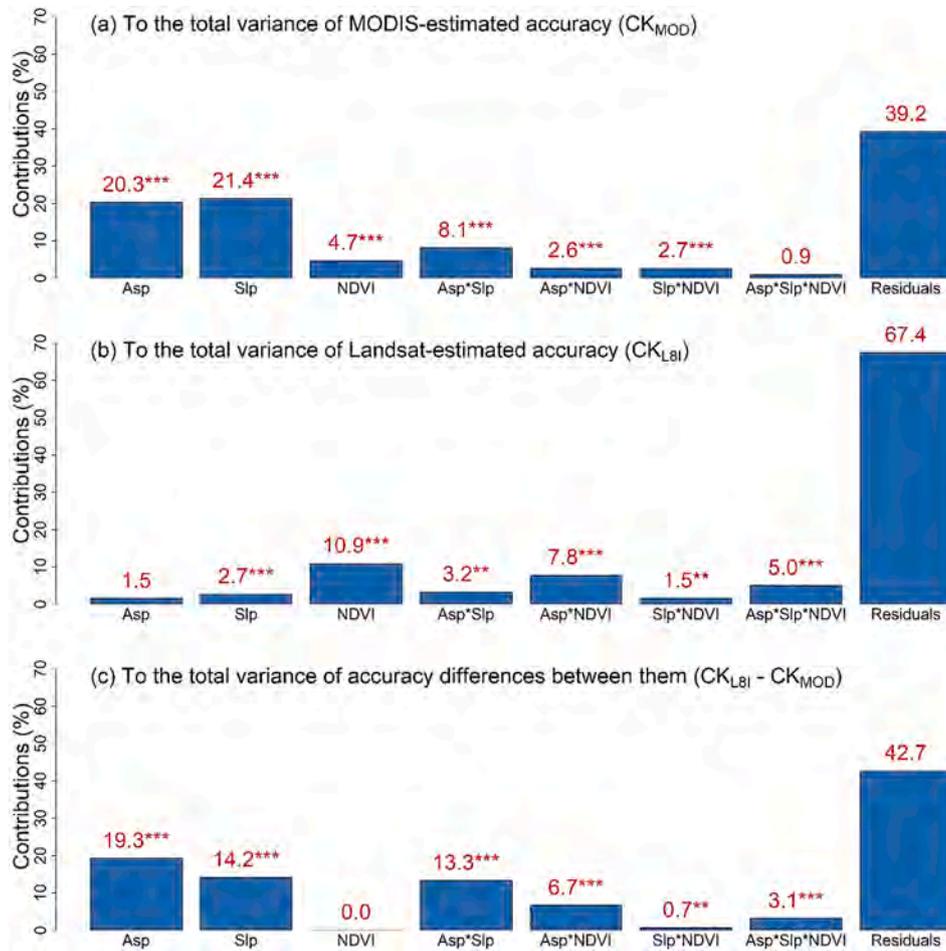


Fig. 8. The ANOVA results showing the contributions of the three scaling factors (i.e. aspect, slope and NDVI) and their interaction terms to total variances of the binary snow cover estimation accuracy under patchy-snow conditions using  $NDSI_{MOD}$  (a),  $NDSI_{L8I}$  (b) and the accuracy differences between them (c).  $CK_{MOD}$  is the CK for  $NDSI_{MOD}$ ,  $CK_{L8I}$  is the CK for  $NDSI_{L8I}$ ; Asp is aspect; Slp is slope; \*\*\* indicates  $P$ -value < 0.001; \*\* indicates  $P$ -value < 0.01; \* indicates  $P$ -value < 0.05.

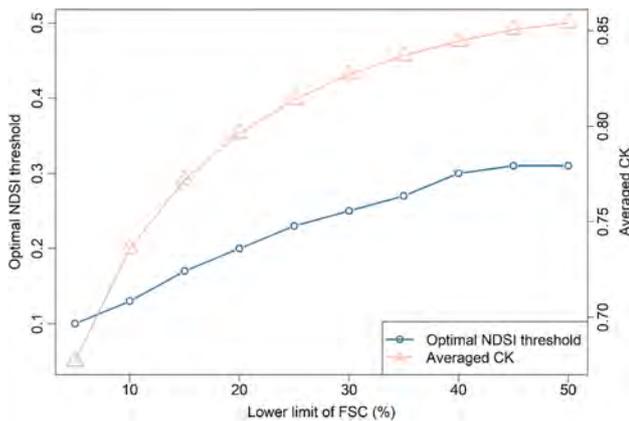
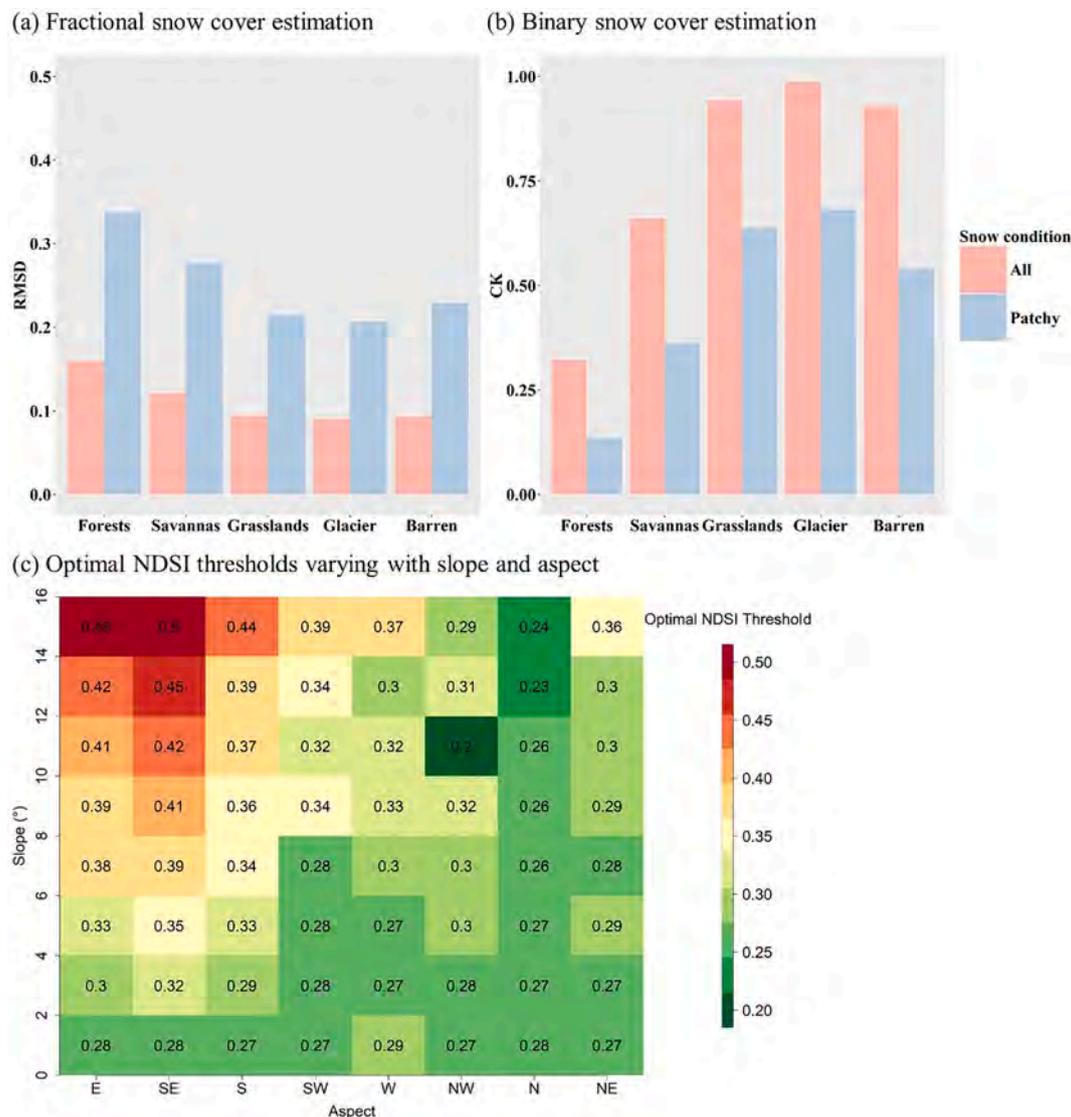


Fig. 9. The optimal NDSI thresholds for detecting snow pixels with different lower limits of FSC (ranging from 5% to 50%) and the corresponding accuracy (averaged CK).

generally increases with slope and decreases as the aspect varies from southeast to northwest. Steeper slopes tend to produce larger commission errors of snow detection. Thus, for pixels with a steep slope, the NDSI threshold tends to be high for reducing the snow commission errors. In general, we also find fewer snow omission errors for the more southeastward aspects than for the more northwestward aspects, possibly due to differences in illumination conditions. Thus, a relatively

low NDSI threshold is required to reduce the omission errors in the aspects that are more northwestward.

There are some limitations in this study. First, due to the high proportion of cloud cover, there are no training or validation Landsat-8 scenes in the southeastern part of the TP, although the scenes used covered most parts of the TP (Fig. 1). Camera-based snow-cover maps (Härer et al., 2018) or images from unmanned aerial vehicles (Liang et al., 2017; Liu et al., 2020) are possible solutions for relieving this problem in future studies. Second, the effects of forest canopy may not be sufficiently considered. Many studies have found that MODIS snow-cover data may largely underestimate FSC for heavily forested areas (Klein et al., 1998; Liu et al., 2020; Masson et al., 2018; Rittger et al., 2013; Wang et al., 2018). Such a problem can be alleviated by spectral mixture analysis (Rittger et al., 2013), considering view angle effects (Xin et al., 2012), the multi-index technique (Wang et al., 2018) or by incorporating more bands and factors using machine learning (Liu et al., 2020). Third, in some cases, it can be difficult to distinguish cloud from snow, when only using Landsat-8 bands, because of their similar multi-spectral signatures (Stillinger et al., 2019). Even if almost all the “cloud” and “cloud shadow” pixels have been removed, there could still be errors existing in the Landsat-8 cloud mask that can be corrected in the future. In addition, elevation is not included as an individual factor in ANOVA, but we find that elevation has relatively small effects on the accuracy of both FSC and BSC estimation using MODIS NDSI, compared with aspect and slope, because TP elevations are clustered in the range 3500–5500 m. According to the histogram of distribution of elevations, eight elevation groups can be defined as <4000, 4000–4200,



**Fig. 10.** Uncertainty due to land cover, slope and aspect. Panels (a) and (b) show accuracy comparisons, based on validation samples, for five major land-cover types of fractional and binary snow cover estimation, respectively. Panel (c) shows the optimal NDSI thresholds for different slopes and aspects.

4200–4400, 4400–4600, 4600–4800, 4800–5000, 5000–5200 and >5200 m. The correlation coefficients (absolute value) between the accuracy of the eight elevation groups and their group-average elevations are 0.24 and 0.49 for FSC and BSC estimations, respectively, and both are insignificant. In contrast, the correlation coefficients between the accuracy of the eight slope groups and their group-average slopes are >0.75 for both FSC and BSC estimates.

**5. Conclusions**

This study investigates the ability of the recently released MODIS NDSI snow-cover data to estimate FSC and BSC on the Tibetan Plateau, by utilizing 353 Landsat-8 images from 2013 to 2015. The images cover most parts of the plateau, where snow cover plays an important role in both regional climate change and water security. We conduct a more detailed analysis of the important scaling effects on FSC and BSC estimation using MODIS NDSI on the TP, than any previous studies, by using a much larger number of reference scenes and by including a comparison with upscaled Landsat-8 NDSI (from 30 m to 500 m).

The FSC ~ NDSI empirical relationship and the NDSI threshold are re-established for Tibetan Plateau FSC and BSC estimates, respectively, based on the MODIS NDSI snow cover product V6. A new regression

equation between FSC and MODIS NDSI (i.e. model MAT) is established from 201 training Landsat-8 scenes and shows better accuracy in estimating FSC for the Tibetan Plateau than the previously used global reference equation (i.e. model MR). A region-wide optimal NDSI threshold for BSC estimation on the Tibetan Plateau of 0.29 is found, and this has better accuracy than the global reference NDSI threshold of 0.4.

We find that there are enhanced scaling effects on both FSC and BSC estimates made using MODIS NDSI, as the accuracy of methods using MODIS NDSI are much lower than those using upscaled Landsat-8 NDSI in both FSC and BSC estimates. The ANOVA results, considering 512 different aspect, slope and NDVI combinations, show that terrain factors (aspect and slope) cause the enhanced scaling effects. Although the accuracy over forest and savanna is clearly lower than that over barren land and grassland in both FSC and BSC estimates, aspect and slope are found to be the major scaling factors for the Tibetan Plateau where 87% of the total area is barren or grassland.

Our study shows that the ability to estimate FSC and BSC from MODIS NDSI data is significantly weakened by the complex topography of the Tibetan Plateau, and thus highlights the importance of developing more advanced models that can incorporate additional factors, including, at least, aspect and slope in future studies. Given the limited improvement in the estimation accuracy of the newly established FSC

relationship, there is a particular requirement for a more efficient method for future FSC estimates in practical research related to hydrology or climate. We also describe a framework for analyzing scaling effects on MODIS-based FSC or BSC estimates, which involves upscaling Landsat-8 NDSI data for comparison and the use of ANOVA to consider different topography and land cover conditions, which is not limited to the Tibetan Plateau and may potentially be useful in other mountainous regions.

## 6. Data availability

All the data used are publicly available. The MODIS NDSI snow cover data can be downloaded from [search.earthdata.nasa.gov](https://search.earthdata.nasa.gov). The Landsat-8 OLI TOA reflectance data can be obtained from USGS Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA) ([espa.cr.usgs.gov](https://espa.cr.usgs.gov)).

## CRedit authorship contribution statement

**Hongbo Zhang:** Writing - original draft, Conceptualization, Methodology, Software. **Fan Zhang:** Writing - review & editing. **Guoqing Zhang:** Writing - review & editing. **Wei Yan:** Writing - review & editing. **Sien Li:** 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.

## Acknowledgements

This study is supported by the Talent Cultivation and Development Support Program by China Agricultural University (Li Sien), State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute (Grant Number: 2019nkms02), State Key Laboratory of Cryospheric Science, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences (Grant Number: SKLCS-OP-2020-13), the National Natural Science Foundation of China (Grant Number: 41701079) and the Natural Science Foundation of Jiangsu Province (Grant Number: BK20181059).

## References

- Armstrong, R.L., et al., 2019. Runoff from glacier ice and seasonal snow in High Asia: separating melt water sources in river flow. *Reg. Environ. Change* 19 (5), 1249–1261.
- Arsenault, K.R., Houser, P.R., De Lannoy, G.J., 2014. Evaluation of the MODIS snow cover fraction product. *Hydrol. Process.* 28 (3), 980–998.
- Baret, F., Clevers, J., Steven, M., 1995. The robustness of canopy gap fraction estimates from red and near-infrared reflectances: a comparison of approaches. *Remote Sens. Environ.* 54 (2), 141–151.
- Chen, X., et al., 2018. Developing a composite daily snow cover extent record over the Tibetan Plateau from 1981 to 2016 using multisource data. *Remote Sens. Environ.* 215, 284–299.
- Choubin, B., et al., 2019. Spatiotemporal dynamics assessment of snow cover to infer snowline elevation mobility in the mountainous regions. *Cold Reg. Sci. Technol.* 167, 102870.
- Crawford, C.J., 2015. MODIS Terra Collection 6 fractional snow cover validation in mountainous terrain during spring snowmelt using Landsat TM and ETM+. *Hydrol. Process.* 29 (1), 128–138.
- Crawford, C.J., Manson, S.M., Bauer, M.E., Hall, D.K., 2013. Multitemporal snow cover mapping in mountainous terrain for Landsat climate data record development. *Remote Sens. Environ.* 135, 224–233.
- Czyzowska-Wisniewski, E.H., van Leeuwen, W.J., 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.
- Dobrev, I.D., Klein, A.G., 2011. Fractional snow cover mapping through artificial neural network analysis of MODIS surface reflectance. *Remote Sens. Environ.* 115 (12), 3355–3366.

- Garrigues, S., Allard, D., Baret, F., Weiss, M., 2006. Influence of landscape spatial heterogeneity on the non-linear estimation of leaf area index from moderate spatial resolution remote sensing data. *Remote Sens. Environ.* 105 (4), 286–298.
- Gascoin, S., Grizonnet, M., Bouchet, M., Salgues, G., Hagolle, O., 2019. Thea Snow collection: high-resolution operational snow cover maps from Sentinel-2 and Landsat-8 data. *Earth Syst. Sci. Data* 11 (2).
- 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.
- Härer, S., Bernhardt, M., Siebers, M., Schulz, K., 2018. On the need for a time-and location-dependent estimation of the NDSI threshold value for reducing existing uncertainties in snow cover maps at different scales. *The Cryosphere* 12 (5), 1629.
- Hou, J., Huang, C., 2014. Improving mountainous snow cover fraction mapping via artificial neural networks combined with MODIS and ancillary topographic data. *IEEE Trans. Geosci. Remote Sensing* 52 (9), 5601–5611.
- Hou, J., Huang, C., Zhang, Y., Guo, J., 2020. On the Value of Available MODIS and Landsat8 OLI Image Pairs for MODIS Fractional Snow Cover Mapping Based on an Artificial Neural Network. *IEEE Trans. Geosci. Remote Sensing*.
- Huang, X., Deng, J., Wang, W., Feng, Q., Liang, T., 2017. Impact of climate and elevation on snow cover using integrated remote sensing snow products in Tibetan Plateau. *Remote Sens. Environ.* 190, 274–288.
- Huang, X., Liang, T., Zhang, X., Guo, Z., 2011. Validation of MODIS snow cover products using Landsat and ground measurements during the 2001–2005 snow seasons over northern Xinjiang, China. *Int. J. Remote Sensing* 32 (1), 133–152.
- 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 (1), 40–49.
- 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 (10–11), 1723–1744.
- Kustas, W.P., et al., 2004. Effects of remote sensing pixel resolution on modeled energy flux variability of croplands in Iowa. *Remote Sens. Environ.* 92 (4), 535–547.
- 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.
- 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 (12), 1332.
- Liang, S., 2000. Numerical experiments on the spatial scaling of land surface albedo and leaf area index. *Remote Sensing Rev.* 19 (1–4), 225–242.
- Liu, C., Huang, X., Li, X., Liang, T., 2020. MODIS fractional snow cover mapping using machine learning technology in a mountainous area. *Remote Sens.* 12 (6), 962.
- Long, D., Singh, V.P., Li, Z.L., 2011. How sensitive is SEBAL to changes in input variables, domain size and satellite sensor? *J. Geophys. Res.: Atmos.* 116 (D21).
- Ma, N., Yu, K., Zhang, Y., Zhai, J., Zhang, H., 2020. Ground observed climatology and trend in snow cover phenology across China with consideration of snow-free breaks. *Clim. Dyn.* 55, 2867–2887.
- Masson, T., et al., 2018. An assessment of existing methodologies to retrieve snow cover fraction from MODIS data. *Remote Sens.* 10 (4), 619.
- McCabe, M.F., Wood, E.F., 2006. Scale influences on the remote estimation of evapotranspiration using multiple satellite sensors. *Remote Sens. Environ.* 105 (4), 271–285.
- Meyer, P., Itten, K.I., Kellenberger, T., Sandmeier, S., Sandmeier, R., 1993. Radiometric corrections of topographically induced effects on Landsat TM data in an alpine environment. *ISPRS J. Photogramm. Remote Sens.* 48 (4), 17–28.
- Mishra, V., Negi, H., Rawat, A., Chaturvedi, A., Singh, R., 2009. Retrieval of sub-pixel snow cover information in the Himalayan region using medium and coarse resolution remote sensing data. *Int. J. Remote Sens.* 30 (18), 4707–4731.
- Painter, T.H., et al., 2009. Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. *Remote Sens. Environ.* 113 (4), 868–879.
- Pepin, N., et al., 2019. An examination of temperature trends at high elevations across the Tibetan Plateau: the use of MODIS LST to understand patterns of elevation-dependent warming. *J. Geophys. Res.: Atmos.*
- Pu, Z., Xu, L., Salomonson, V.V., 2007. MODIS/Terra observed seasonal variations of snow cover over the Tibetan Plateau. *Geophys. Res. Lett.* 34 (6).
- Riggs, G.A., Hall, D.K., Román, M.O., 2016a. MODIS snow products user guide for Collection 6.
- Riggs, G.A., Hall, D.K., Román, M.O., 2016b. MODIS Snow Products User Guide for Collection 6. Digital Media.
- 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* 9 (2), 765.
- Riggs, G.A., Hall, D.K., Salomonson, V.V., 2006. MODIS snow products user guide to collection 5. Digital Media 80 (6), 1–80.
- Rittger, K., Painter, T.H., Dozier, J., 2013. Assessment of methods for mapping snow cover from MODIS. *Adv. Water Resour.* 51, 367–380.
- Salomonson, 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 Sensing* 44 (7), 1747–1756.
- Sharma, V., Kilic, A., Irmak, S., 2016. Impact of scale/resolution on evapotranspiration from Landsat and MODIS images. *Water Resour. Res.* 52 (3), 1800–1819.
- Shrestha, M., Wang, L., Koike, T., Xue, Y., Hirabayashi, Y., 2011. Modeling the spatial distribution of snow cover in the Dudhkoshi Region of the Nepal Himalayas. *J. Hydrometeorol.* 13 (1), 204–222.

- Shreve, C.M., Okin, G.S., Painter, T.H., 2009. Indices for estimating fractional snow cover in the western Tibetan Plateau. *J. Glaciol.* 55 (192), 737–745.
- Stillinger, T., Roberts, D.A., Collar, N.M., Dozier, J., 2019. Cloud masking for Landsat 8 and MODIS Terra over snow-covered terrain: Error analysis and spectral similarity between snow and cloud. *Water Resour. Res.* 55 (7), 6169–6184.
- Sulla-Menashe, D., Friedl, M.A., 2018. User guide to collection 6 MODIS land cover (MCD12Q1 and MCD12C1) product. USGS: Reston, VA, USA: 1–18.
- Tang, Z., et al., 2014. Extraction and assessment of snowline altitude over the Tibetan plateau using MODIS fractional snow cover data (2001 to 2013). *J. Appl. Remote Sens.* 8 (1), 084689.
- Tang, Z.G., Wang, J., Li, H.Y., Yan, L.L., 2013a. Spatiotemporal changes of snow cover over the Tibetan plateau based on cloud-removed moderate resolution imaging spectroradiometer fractional snow cover product from 2001 to 2011. *J. Appl. Remote Sens.* 7, 14.
- Tang, Z.G., Wang, J., Yan, L.L., Li, H., Liang, J., 2013b. Estimating sub-pixel snow cover from MODIS in Qinghai-Tibet Plateau. *J. Arid Land Resour. Environ.* 27 (11), 33–38.
- Thackeray, C.W., Derksen, C., Fletcher, C.G., Hall, A., 2019. Snow and climate: feedbacks, drivers, and indices of change. *Current Climate Change Reports* 5 (4), 322–333.
- Wang, X., Jian, W., Tao, C., Huang, 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.* PP(99), 1–9.
- Wang, X., Wu, C., Wang, H., Gonsamo, A., Liu, Z., 2017. No evidence of widespread decline of snow cover on the Tibetan Plateau over 2000–2015. *Sci. Rep.* 7 (1), 14645.
- Xin, Q., et al., 2012. View angle effects on MODIS snow mapping in forests. *Remote Sens. Environ.* 118, 50–59.
- Yin, D., Cao, X., Chen, X., Shao, Y., Chen, J., 2013. Comparison of automatic thresholding methods for snow-cover mapping using Landsat TM imagery. *Int. J. Remote Sens.* 34 (19), 6529–6538.
- You, Q., et al., 2019. Review of snow cover variation over the Tibetan Plateau and its influence on the broad climate system. *Earth Sci. Rev.* 103043.
- Yu, J., et al., 2016. Developing daily cloud-free snow composite products from MODIS terra-aqua and IMS for the Tibetan Plateau. *IEEE Trans. Geosci. Remote Sensing* 54 (4), 2171–2180.
- Zhang, B., et al., 2013a. Monitoring changes of snow cover, lake and vegetation phenology in Nam Co Lake Basin (Tibetan Plateau) using remote SENSING (2000–2009). *J. Gt. Lakes Res.* 39 (2), 224–233.
- Zhang, F., et al., 2015. Snow cover and runoff modelling in a high mountain catchment with scarce data: effects of temperature and precipitation parameters. *Hydrol. Process.* 29 (1), 52–65.
- Zhang, G., et al., 2019a. Regional differences of lake evolution across China during 1960s–2015 and its natural and anthropogenic causes. *Remote Sens. Environ.* 221, 386–404.
- Zhang, G.Q., Xie, H.J., Yao, T.D., Liang, T.G., Kang, S.C., 2012. Snow cover dynamics of four lake basins over Tibetan Plateau using time series MODIS data (2001–2010). *Water Resour. Res.* 48.
- Zhang, H., Zhang, F., Che, T., Wang, S., 2020. Comparative evaluation of VIIRS daily snow cover product with MODIS for snow detection in China based on ground observations. *Sci. Total Environ.* 724, 138156.
- Zhang, H., et al., 2019b. Ground-based evaluation of MODIS snow cover product V6 across China: Implications for the selection of NDSI threshold. *Sci. Total Environ.* 651, 2712–2726.
- Zhang, Y., Ma, N., 2018. Spatiotemporal variability of snow cover and snow water equivalent in the last three decades over eurasia. *J. Hydrol.* 559.
- Zhang, L., Su, F., Yang, D., Hao, Z., Tong, K., 2013b. Discharge regime and simulation for the upstream of major rivers over Tibetan Plateau. *J. Geophys. Res.: Atmos.* 118 (15), 8500–8518.