

DERIVATION OF SUPRAGLACIAL DEBRIS COVER BY MACHINE LEARNING ALGORITHMS ON THE GEE PLATFORM: A CASE STUDY OF GLACIERS IN THE HUNZA VALLEY

Fuming Xie ^{1,2}, Shiyin Liu ^{1,2,3,*}, Yongpeng Gao ^{1,2}, Yu Zhu ^{1,2}, Kunpeng Wu ^{1,2}, Miaomiao Qi ^{1,2}, Shimei Duan ^{1,2}, Adnan M. Tahir ⁴

¹ Yunnan Key Laboratory of International Rivers and Transboundary Eco-Security, Yunnan University, Kunming 650091, China

² Institute of International Rivers and Eco-security, Yunnan University, Kunming, Yunnan 650091, China

³ State Key Laboratory of Cryospheric Sciences, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China

⁴ Department of Environmental Sciences, COMSATS University Islamabad, Abbottabad Campus, Abbottabad 22060, Pakistan

Commission III, WG III/9

KEY WORDS: Supraglacial debris cover, Hunza Valley, Machine learning, Otsu, Google Earth Engine

ABSTRACT:

Calculating the spatial-temporal distribution of supraglacial debris cover on glaciers is essential for understanding mass balance processes, glacier lake outburst floods, hydrological predictions, and glacier fluctuations that have attracted attention in recent years. However, due to the reflectance of supraglacial debris is similar to that of non-glacier slopes, mapping supraglacial debris cover based on optical remote sensing remains challenging. In this paper, we used NDSI and machine learning algorithm to extract debris cover on glaciers in Hunza Valley, Pakistan. Our result showed that the RF model has the best classification accuracy with kappa coefficient of 0.94 and overall accuracy of 96%. The debris-covered area increased by 21.31% from 1990 to 2019 (394.76 km² ~ 478.88 km²) in the study area. Results and the method are of significance in the assessment of meltwater modeling for glaciers with debris cover.

1. INTRODUCTION

Glaciers, also known as ‘alpine solid reservoirs’, are not only a promising natural freshwater resource but also a sensitive indicator of global climate change (Kaab et al., 2012; Yang, 1995; Zemp et al., 2019;). Recent studies have been conducted on mass balance (Dehecq et al., 2018; Zemp et al., 2019), glacial area change (Patel et al., 2019; Paul et al., 2013), surface velocity (Altena et al., 2019; Garg et al., 2019), glacial-hydrological modeling (Shrestha et al., 2015), and glaciers’ response to climate change (Rowan et al., 2015; Scherler et al., 2011). However, accurate glacier boundary outlines are always fundamental for change detection and model validation. At present, large-scale glacier inventory initiatives mainly include Global Land Ice Measurements from Space (GLIMS) (Raup et al., 2007), Randolph Glacier Inventory (RGI) (Pfeffer et al., 2014), Glacier Area Mapping for Discharge from the Asian Mountains (GAMDAM) (Nuimura et al., 2015), and the Second Glacier Inventory of China (Liu et al., 2015). All of these were aimed at generating a global dataset of land glaciers. However, the alpine glaciers that are widely distributed in the Pamir, Karakorum, Kunlun, Nyainqentangula, and Himalaya Mountains (Khan et al., 2015; Scherler et al., 2011; Shukla and Garg, 2019; Zhang et al., 2016) are heavily covered by debris. Due to the reflectance of supraglacial debris is similar to that of non-glacier slopes (Paul et al., 2004) and because there is a lack of continuous, large-scale, high-quality optical images that are not affected by cloud and terrain shadows, mapping the supraglacial debris cover based on remote sensing is challenging compared to mapping clean ice or snow.

In the past, attempts at boundary identification of glaciers with debris cover were made with the assistance of individual parameters such as the normalized difference vegetation index (NDVI), normalized difference snow index (NDSI), normalized difference water index (NDWI), and spectral band ratio thresholds (e.g., near-infrared/short-wave infrared (NIR/SWIR)) or their combination from optical remote sensing images (Alifu et al., 2016; Bolch et al., 2010; Mölg et al., 2018). The aim was to distinguish between clean ice and debris-covered ice. These methods can robustly delineate clean ice or snow, but they cannot accurately and automatically classify debris-covered glacier ice as distinct from clean ice and the surrounding land surface (Robson et al., 2015). This has stimulated studies on the use of other parameters such as geomorphic parameters derived from a digital elevation model (DEM) (Patel et al., 2019; Paul et al., 2004) and thermal characteristics from the infrared band (Singh and Goyal, 2018), as well as utilizing the coherence change between two successive synthetic aperture radar (SAR) images (Janke et al., 2015; Robson et al., 2015). However, complex pre-processing and severe terrain noise of SAR data make large-scale applications difficult.

In recent years, machine-learning-based classification methods have been applied to identify glacial outlines (Racoviteanu and Williams, 2012; Zhang et al., 2019). Machine learning has advantages for extracting land surface information from remote sensing images (Maxwell et al., 2018) because mining large-scale and time-series land information from massive remote sensing data is a computationally intensive task and requires powerful computing platforms for analysis. Fortunately, some geospatial cloud-computing platforms are emerging to meet this demand, such as Google Earth Engine (GEE), Amazon Web

* Corresponding author

Services (AWS), Earth Server (ES), and Earth Observation Data Centre (EOC) (Guo et al., 2020). Among these, GEE has obvious advantages because it is an open-source cloud-based platform for planetary-scale geospatial analysis that integrates mainstream free satellite data such as the Landsat archive, Sentinel series imagery, and other terrain products and climate data (Gorelick et al., 2017). It can quickly pull information from massive satellite-image data on various high-impact societal issues such as forest resources (Hazel et al., 2016), water resources (Pekel et al., 2016), land use classification (Dong et al., 2016), and other fields.

The objective of this study is to develop an automatic algorithm for identifying debris-covered ice and mapping its spatio-temporal distribution by combining glacier inventory data and remotely sensed images based on the GEE geospatial analysis platform. The study is focused on glaciers in the Hunza Valley in the Karakorum Mountains of Pakistan. Based on the traditional knowledge-based approach, Otsu's method was utilized to optimize thresholds of band-based indices and machine learning algorithms such as random forest (RF), support vector machine (SVM), and classification and regression tree (CART) that are used to classify supraglacial features, including debris-covered glacier ice, based on the spatial resolution of the images used. Raw spectral information, band ratios, and color-to-grayscale conversion from Landsat 5/8 optical satellite imagery and the topographical components derived from SRTM DEM products were extracted as feature variables in the machine learning model. The same scheme was used to generate a time-series of the debris-covered area and bare ice area in the study area. Finally, the results were comprehensively analyzed together with other data derived at the same time.

2. STUDY AREA

The Hunza Valley is an area measuring ~11000 km² located in western Karakoram, Northern Pakistan (latitude 36°00'15"~37°05'23" N, longitude 74°02'57" ~ 75°46'48" E) (Figure 1). The topography across Hunza Valley is characterized by large altitudinal variations from 1,341 m to 7,831 m above sea level (a.s.l.). The valley is home to glaciers with a total area of ~3600 km² that accounts for ~33% of the basin area based on the RGI 6.0 dataset (RGI Consortium, 2017). Most of the glaciers (e.g., Hispar, Batura, Barpu) are debris-covered and in a state of surging and advancing (Bhambri et al., 2017). Debris-covered glaciers are potential factors driving glacial lake outburst floods (GLOFs) (Bhambri et al., 2019), which represent a major threat to local people, their properties, and infrastructure such as Karakoram Highway (KKH) in the Hunza Valley. Climatologically, the study area is arid to semi-arid and lies in the subtropical climate zone, with significant variations in precipitation and temperature (Immerzeel et al., 2012). Based on the MODIS 1 km LST Daily products, the mean land surface temperature (LST) for the entire region is -12.9°C in January and 20.1°C in July. Precipitation is mainly controlled by Indian monsoons and the westerlies, and average annual precipitation is between 180 mm and 690 mm (Qureshi et al., 2017). Snow cover occupies approximately 80% of the basin's land surface in the winter and decreases to 30% in the summer (Tahir et al., 2011). Types of land cover in the basin include forest (0.36%), shrubs (16.12%), farmland (0.76%), and barren land (82.78%). The main soil types include Leptosols (LP), rock outcrop soil (RK), and glaciated soil (GG); further, the major soil component in the region is highly active clay, followed by rock outcrops and glacial soil (Ali et al., 2018; Garee et al., 2017).

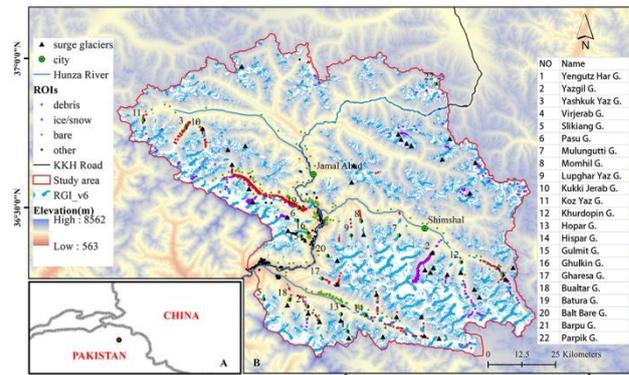


Figure 1. The geographical location of the Hunza Valley with region of interest (ROI) sample sites of various glaciers. Surging glaciers are marked with triangles (Bhambri et al., 2017).

3. DATA AND METHODS

3.1 Data and Processing

We used cloud-free imageries of Landsat 5 Thematic Mapper (TM) and Landsat 8 Operational Land Imager (OLI) acquired in each ablation season and are available on GEE. We used Sentinel-2A/B Multi-Spectral Instrument images with a spatial resolution of 10m and a recycle time of 5 days for validation purpose. The 1 arc-second SRTM DEM V3 was used to derive topographical parameters and elevation gradient, like slope, aspect, and plan curvature. RGI v6.0 glacier inventory for the region was used to constrain the spatial extend of model runs for debris cover extractions which have been modified based on Sentinel-2 images. The processing procedure can be summarised as follows: the selection of images excluding seasonal snow cover was done by confining images acquired during the ablation season (e.g., during July, or days 200 to 270 ± 10 days). Then, the pixel digital number (DN) values were converted to atmospheric top-level reflectance (TOA) (Chander et al., 2009). Third, images (pixels) with the least clouds were selected by applying a simple Landsat cloud score algorithm available on GEE, generating a composite image with no-cloud pixels of scenes in the region, and computing per-band percentile values (25%) from the accepted pixels. Whole image data pre-processing and the subsequent classification process were implemented by coding on the GEE platform.

3.2 Supraglacial Debris Extraction and Validation

3.2.1 Debris-covered ice extraction: In this work, we used three algorithms to delineate boundaries for glaciers with debris cover. The first method is a multi-criterion-based approach that uses thresholds of NDSI < 0.4 (Dozier, 1989) to distinguish clean ice/snow from debris-covered ice and NDVI < 0.1 to exclude non-glaciers. However, a fixed NDSI threshold of 0.4 may not be applicable for all regions. We then developed an optimization method based on Otsu algorithms (Otsu) to optimize the NDSI threshold to better distinguish clean ice/snow and debris-covered ice in different areas. An Otsu algorithm is an automatic non-parametric and unsupervised method for thresholding that is used to automatically detect targets in computer vision and image processing fields (Ng, 2006). Otsu is a global threshold method, and its principles are the following: Assume that the grey value of an image is 1 ~ N, divide it into two groups at k value, G₀ = [1~k] and G₁ = [k + 1~N], and calculate the probability of the two groups, ω₀ and ω₁, the average values for each group (μ₀ and μ₁), and the

entire image (μ). Then, the variance of the two groups can be calculated by the following equation:

$$\sigma^2(k) = \omega_0(\mu_0 - \mu)^2 + \omega_1(\mu_1 - \mu)^2$$

where, $\sigma^2(k)$ is a threshold selection function. By changing the k value within $1 \sim N$, the k value at which $\sigma^2(k)$ is maximized is the required threshold.

The third algorithm is a machine-learning algorithm that includes SVM (Suykens and Vandewalle, 1999), RF (Liaw and Wiener, 2002), and CART (Breiman et al., 1984). In this study, default parameters were used, but 500 trees were set for the RF classification, and the kernel function of the radial basis function (RBF) was applied in the SVM model. It is obvious that a single spectrum cannot fully solve the problem of the similarity of ice covered with debris to the surrounding terrain. We generated 14 feature variables: original spectrum (b1, b2, b3, b4, b5, b6, b7), band ratios (nir/swir1), NDVI [(nir-red)/(nir+red)], NDSI [(green-swir1)/(green+swir1)], NDWI [(green-nir)/(green+nir)], and luminance and geomorphic parameters (slope and aspect). Training data were visually sampled based on high-spatial-resolution Sentinel-2 and Google Earth images. Samples in the regions of interest (ROIs) were divided into clean ice or snow (ice/snow), debris-covered ice (debris), bare land, and others (e.g., vegetation, villages, rivers and lakes, and shadows) according to land cover types in the Hunza Valley (Ali et al., 2018). For example, for 2019, 1,024 sample points were selected, including 373 debris, 356 ice/snow, 270 bare land, and 205 other points. The spatial distribution of ROIs for 2019 is shown in Figure 1B.

3.2.2 Post-classification processing: The classified results were processed by applying slope threshold and RGI6.0 + outlines and removing the ‘salt-and-pepper effect’ and glacier ice and debris-covered areas measuring less than 0.05 km².

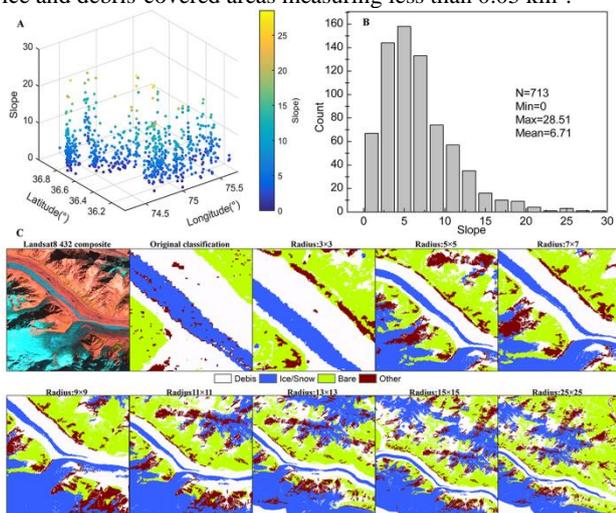


Figure 2. Post-classification processing: A) 3D distribution of slope in the debris-covered area of glaciers; B) histogram of slope statistics; C) comparison of removal effects under different kernel radius and kernel types, 5x5 square type, and 50 connected pixels were selected.

(i) *Slope threshold* This is a key parameter for delineating glacier areas with debris cover. Some early studies proposed various thresholds for slope. For example, a slope < 24° can be used to distinguish debris-covered glaciers and the surrounding terrain (Paul et al., 2004). Some proposed smaller values such as a slope < 12° (Alifu et al., 2015) or a slope < 14~16° (Robson et al., 2015). We hypothesize that the slope threshold shows spatial heterogeneity for various glacierised mountains.

To obtain a typical value for this region, we selected 713 sampling points uniformly for glacier and non-glacier debris areas of the Hunza Valley. Our statistical results show that slopes of less than 25° dominate the glacier area (99% of the glacier debris area) with a mean slope of 6.71° (Figure 2A, B). Therefore, a slope threshold of 25°, which is consistent with Paul’s suggestion, was used to confine the glacier debris area from the surrounding debris area (Paul et al., 2004).

(ii) *Removal of the ‘salt-and-pepper effect’* This effect is common in pixel-based classification of land surfaces. To remove island pixels, we found that a 5x5 square kernel with 50 connected pixels was the best choice to remove this effect after comparison of different kernel sizes (kernel radius of 3x3, 5x5, 7x7, 9x9, 11x11, 13x13, 15x15, and 25x25) and types (square, circle, octagon, diamond, cross, and plus) and connected pixel counts (5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 1000) (Figure 2C).

3.2.3 Evaluation of classification accuracy: To evaluate the accuracy of classification models and the authenticity of the estimation results, we used two methods. For machine learning, a cross-validation method was used by dividing the total sample into two parts, with 70% of the sample points from each class randomly selected to train the model and 30% of sample points withheld as a validation dataset. Using the validation data, a confusion matrix was generated to assess the accuracy of predictions across class and overall accuracy through the Kappa coefficient and overall accuracy (OA). Another accuracy evaluation method using the ‘round robin’ method is based on multiple manual digitizations, as proposed by Paul and colleagues (Paul et al., 2017). Based on high-resolution Sentinel-2 and Google Earth images, we selected three glaciers (Kukki Jerab, Virjerab, and Yashkuk Yaz) and performed manual digital digitization five times and used the mean value for evaluating automatically derived extent and standard deviation for digitization accuracy.

4. RESULTS

4.1 Accuracy Analysis

Figure 3A shows that results from the machine learning algorithms generally have high accuracy, with Kappa coefficients ranging from 0.82 to 0.95. In comparison, classification accuracy for the SVM model is slightly lower than for the RF model, whose Kappa coefficient is 0.9 and overall accuracy is 96.02%. Table 1 shows the debris-covered area on three debris-covered glaciers derived by machine learning algorithms and by manual digitization. Manual digitization for each glacier by five professionals shows decreased standard deviations for the largest to the small glaciers, specifically Yashkuk Yaz (26.77 ± 2.19 km²), Kukki Jerab (15.19 ± 0.80 km²), and Virjerab (21.56 ± 0.35 km²) glaciers. The linear fit between the mean of the digitizations and automatic estimates indicates that all the machine learning algorithms can produce results close to reality, but the RF model is slightly better than the other two with reference to the determination R² coefficient of 0.98 for RF, 0.97 for the multi-index method, and 0.93 for the Otsu-based model (Figure 3B). Table 2 shows the confusion matrix of model prediction versus validation data using the RF model and 2019 remote sensing data. The accuracy of RF model prediction for different land types is more than 90%, especially for clean ice/snow with user and producer accuracy of 100%. Considering the accuracy of the RF model for satellite data acquired at different times, the Kappa coefficient of the RF model was within a range of 0.92–0.95 (Figure 4).

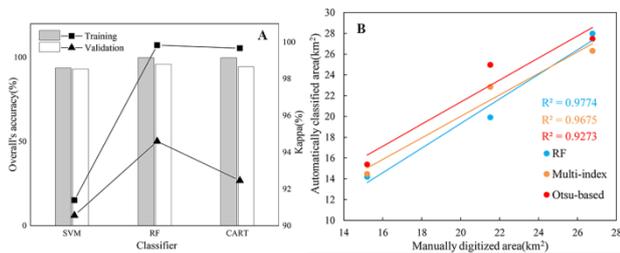


Figure 3. A) Model accuracy of three machine learning algorithms (bar plot is overall's accuracy and line chart are Kappa coefficient); B) linear fit of estimated values to manually digitized values.

In addition, to verify the accuracy and reliability of RF model estimation, we also compared the results with those of others (Figure 4). For example, the manual digitization results of Mölg and colleagues (Mölg et al., 2018) are an area of 583.56 km², which is much larger than the debris cover area of this study in 2010 (440.10 km²). This is consistent with the overestimation described by the authors. Another result automatically extracted by Scherler and colleagues (Scherler et al., 2018) is roughly consistent with the results from this study, with a residual error of 25~33 km².

Table 1 Comparison of debris cover area values derived from automatic extraction algorithms and manual digitisation

Glacier	Multiple digitisations (km ²)					Std.	Mean	RF	Multi-index	Otsu-based
	1	2	3	4	5					
Kukki Jerab	15.67	16.13	15.04	15.12	14.01	0.80	15.19	14.19	14.47	15.38
Virjerab	21.67	21.94	21.46	20.98	21.56	0.35	21.52	19.92	22.85	24.98
Yashkuk Yaz	25.90	27.26	26.20	30.21	24.30	2.19	26.77	27.98	26.32	27.48

Table 2 Confusion matrix for RF model classification in 2019 (model prediction (top) and validation data (left))

	Debris	Ice/snow	Bare	Other	Total	Producer accuracy
Debris	108	0	4	2	114	94.74%
Ice/snow	0	115	0	0	115	100%
Bare	4	0	79	0	83	95.18%
Other	3	0	2	60	65	92.31%
Total	115	115	85	62	377	OA: 96.02%
User accuracy	93.91%	100%	92.94%	96.77%	-	Kappa: 94.60%

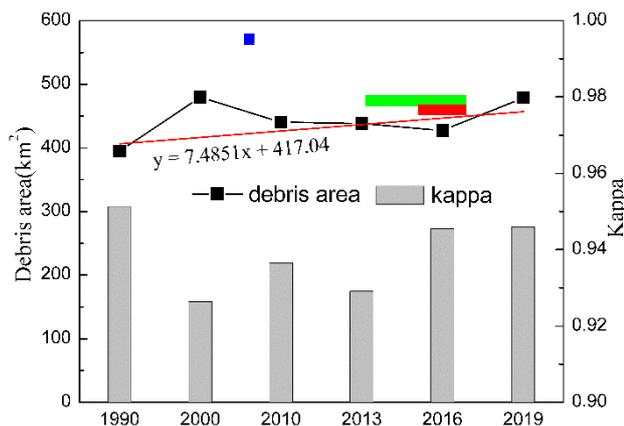


Figure 4. Debris cover area of the study area in different periods. Black represents estimates by the RF classifier in this paper. The grey histogram represents the estimation accuracy (Kappa coefficient) of the RF model for the corresponding period. Blue represents the manual digitization results of Mölg and colleagues (Mölg et al., 2018) based on Google Earth images and 2007 ALOS-1 PALSAR-1 images. Yellow and red represent those automatically extracted by Scherler and colleagues (Scherler et al., 2018) using Landsat 8 composite images from 2013–2017 and Sentinel-2 composite images from 2015–2017, respectively.

4.2 Mapping Supraglacial Debris Cover

The glacial debris cover area for six periods (1990, 2000, 2010, 2013, 2016, and 2019) in the Hunza Valley was delineated based on the RF model. The total area of glaciers with debris cover in the study area showed a clear increase from 1990 (394.76 km²) to 2019 (478.88 km²) (Figure 4). This increasing trend has been seen for other major glaciers during the same

period as shown in Figure 5. This means glaciers in the region have been becoming dirty since 1990.

The results for bare ice and supraglacial debris cover area are based on the RF model (shown in Figure 6A) and on images acquired in 2019 for all glaciers in the Hunza Valley. More than 78% of the clean ice/snow area lies higher than 5,500 m a.s.l., while about 80% of the debris-covered ice is distributed between 4,000 and 5,000 m a.s.l. The median elevations were approximately 5,365 m for clean ice/snow and 4,075 m for debris-covered ice. The median elevation of glaciers in Hunza is 5,230 m a.s.l., which is sometimes referred to as the equilibrium-line altitude of glaciers in the Hunza basin (Qureshi et al., 2017). This means that supraglacial debris is extensively distributed in the lower part of the ablation area of glaciers. Most glaciers have a North (N) and Northeast (NE) aspect, accounting for 38.7% of the glacier area, and only a few glaciers have a West (W) aspect (Figure 6B).

Overall, the debris-covered area of each glacier has also increased over time but varies by altitude gradient. In the low-altitude area (i.e., < 3,200 m a.s.l.), the debris-covered area steadily changes without a significant increase, with an average variance of about 0.1 km² but increases significantly in the mid-altitude range (3,200 ~ 4,200 m a.s.l.). When the elevation is higher than 4,200 m a.s.l., the debris-covered area increases with large variance, especially in areas above 5,000 m. Seasonal snow is a key factor that causes large variance in the debris cover of these areas. The essence of debris expansion is the conversion of clean ice into impure ice, which changes its reflectance on remote sensing images, causing it to be identified as the debris class. The main material source of these moraines is impure ice formed by dust accumulation. Moreover, due to deglaciation, rocks in the ice body are exposed and accumulate, the bedrock on both sides of a glacier disintegrates under the

influence of external forces, or debris such as rolled clastic rock accumulates on the ice surface because of climate change.

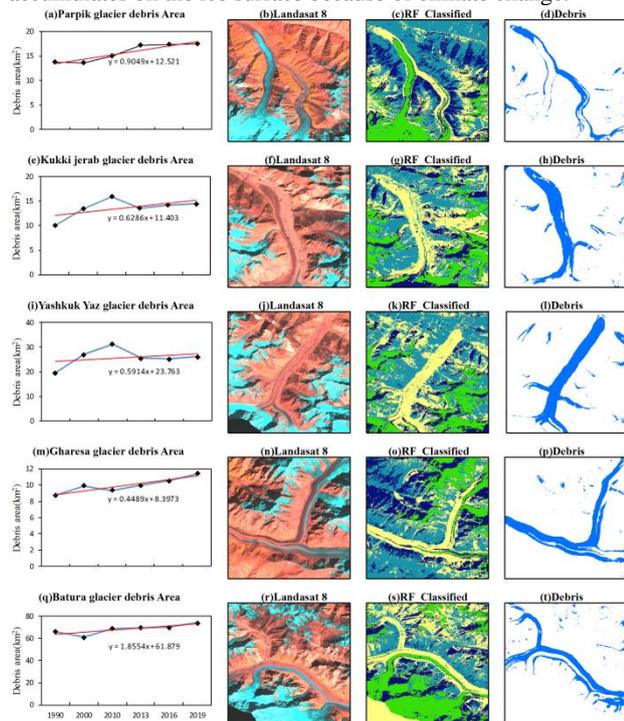


Figure 5. Random forest classifier-based area estimates for five glaciers. Column 1 is estimated debris area in different years and the trend over time, followed by an example of supraglacial debris extraction for each glacier, including the Landsat 8 satellite image in 2009 (Column 2), RF classified results (Column 3), and debris area (Column 4).

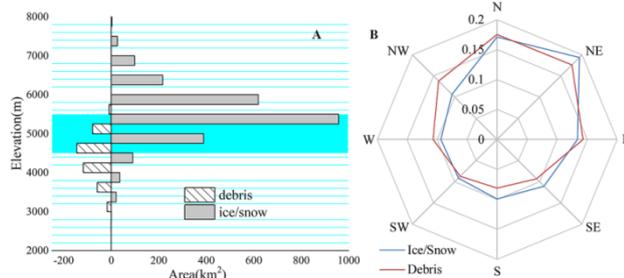


Figure 6. A) Distribution of debris-covered and clean ice/snow area by altitude based on the 2019 RF model classified result. The light blue area represents the equilibrium-line altitude (ELA) of glaciers in the Hunza Valley (4,500 ~ 5,500 m a.s.l.); B) Percentage of debris cover and ice/snow distribution by the aspect.

5. DISCUSSION

5.1 Uncertainties in Supraglacial Debris Mapping

Uncertainty analysis, including uncertainty from measurement errors, models, and scale effects, is an important step in validating the authenticity of remotely sensed products (Wu et al., 2014). Aiming to determine the extent of debris cover using remote sensing data, we analyzed sources of uncertainty in this paper. This mainly includes the following aspects: i) Errors from remote sensing data, including distortions generated by the sensor during image acquisition, and image quality and mixed pixel effects caused by spatial resolution. Satellite data directly determine the quality of the classification results, especially in high mountain areas, which are affected by clouds and steep

terrain. In this paper, to exclude the impact of fresh snow, we selected images from the melting season with heavy cloud cover. This resulted in a lack of high-quality images for some years, requiring us to use images from previous and subsequent years. ii) Errors from ground observation data, such as the counts and spatial distribution of the sample points in this paper for machine learning classification. iii) Errors from classification methods, mainly including model types, selected feature variables, value selection for model parameters or thresholds, and accuracy evaluation method. iv) The complex surface types of debris-covered glaciers, such as stagnant ice in glacier tongue areas, cliffs and ponds on debris-covered area, large moraines, and dirty ice.

Generally, the methods for delineating debris include manual digitization and multi-rule-based and supervised classification. From the nature of different algorithms, manual digitization is time-consuming, and the interpretation accuracy varies depending on the experience of the interpreter. The multi-rule-based method can quickly produce binarised results, with simple post-classification processing. However, the rule-based method generally determines the threshold based on experience, and this threshold may not be suitable for all glacier areas. Fortunately, the Otsu-based method developed in this paper addresses this problem and automatically calculates the threshold, which is more flexible than the empirical threshold. Supervised classification is complicated and requires a huge amount of calculations, but when it is combined with geospatial big data analysis platforms such as GEE, it has great potential for classification. From the perspective of the classification effect, all methods are highly robust for identifying clean ice or snow, but there is a slight difference in the identification of the debris-covered layer, which is mainly reflected in the glacier tongue and the interface area between clean ice and debris-covered ice. All methods have inaccurate identification at the glacier tongue with thicker debris or vegetation (Vezzola et al., 2016). At the interface of clean ice and debris-covered area, the effects of machine learning classification are the best, followed by the Otsu-based method. On the whole, the Otsu-based method is fast and can combine RGI data with appropriate manual corrections, making it highly suitable for large-scale glacier inventory. Further, because the pixel-based classification considers only the spectral characteristics of a single pixel, it is easy to identify the ponds and cliffs on debris-covered glaciers as water or clean ice. In addition, the results with salt-and-pepper effects often require post-classification processing, especially for machine learning algorithms. Therefore, object-based classification based on image features such as shape or topological features has been developed with higher accuracy than pixel-based methods (Kraaijenbrink et al., 2016; Robson et al., 2015). For example, Robson and colleagues (Robson et al., 2015) showed that compared with the pixel-based method, an object-based method can improve the classification of debris-covered glaciers by 10%.

5.2 New Approaches for Future Work

The biggest challenges in debris-covered ice delineation are the complex surface features of glacial tongue areas (i.e., stagnant ice) and mixed pixels caused by the coexistence of debris and clean ice on upper glacier areas. Due to high-resolution optical remote sensing, images have more detailed characteristics, making the first approach conducive to high-precision detection of targets. Extraction of multiple feature variables, such as texture features, combined with the powerful capabilities of deep learning, is a method capable of extracting these complex surface features.

Another approach makes full use of the motion characteristics of glaciers to identify debris-covered ice. Ice flows, which are a major feature distinguishing glaciers from other natural bodies, control the basic process of glacial change. Observations of glacier ice flow are of great significance to reveal glaciers' changing rules and predict their future changes. Ice flow can be monitored by remote sensing methods such as visual tracking, cross-correlation (Fahnestock et al., 2016), and interferometric radar (Paul et al., 2013). Previous studies have used glacier velocity data to analyze glacial motion mechanisms, test glacial flow theories, and identify surging or advancing glaciers (Dehecq et al., 2018; Quincey et al., 2011), but no study has used glacier mapping. Therefore, to verify the possibility of using glacial surface motion characteristics to distinguish glaciers and non-glacier areas, we conducted a test using ITS_LIVE surface velocity products, which are derived by cross-correlation feature tracking (Gardner et al., 2019). The surface velocity is generated from Landsat images, and the resampling spatial resolution is 240 m. Figure 7A shows the mean surface velocity from 2011 to 2018. The surface velocity in non-glacial areas is between 0 ~ 1 m/yr. As shown in Figure 7B and 7C, we extracted areas with a surface velocity greater than 5 m/yr and 10 m/yr and overlaid the debris cover estimated in this paper. This showed good edge matches at coarse resolution. After improving the spatial and temporal resolution of surface velocity, this motion feature can solve the problem of mixed pixels between glaciers and lateral moraines, especially for large debris-covered glaciers.

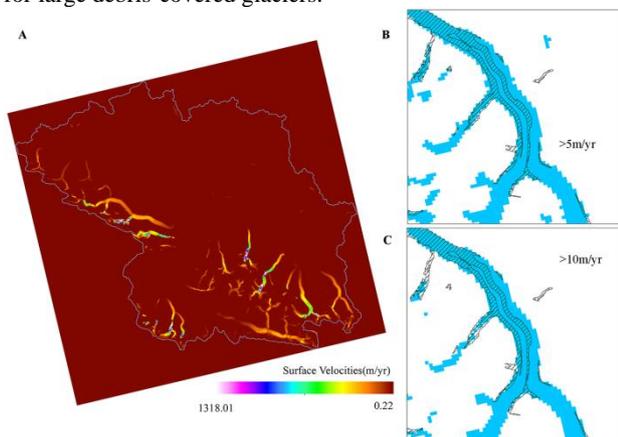


Figure 7. A) Mean surface velocity from 2011 to 2018. Data are from ITS_LIVE annual surface velocity products from the National Snow and Ice Data Centre. B) and C) are overlays of debris cover extent and area surface velocity > 5 m/yr and > 10 m/yr, respectively.

6. CONCLUSIONS AND OUTLOOK

The purpose of this study was to develop an automatic method for debris-covered ice mapping and further explore the change in the area of debris cover of glaciers over time. We coded three algorithms for debris cover extent mapping based on the GEE geospatial big data analysis platform, and we tested them with data on glaciers in the Hunza Valley using Landsat satellite imagery. The main conclusions are i) The RF has the best accuracy among the classification algorithms, with a Kappa coefficient of 0.94, an overall accuracy of 96.02%, and determination coefficient (R^2) of the linear fit between the manual digitisation of 0.98. However, after analysing the performance and practicability of the methods, we found that the Otsu-based method has great potential for large-scale glacier inventory when combined with manual corrections. ii) The supraglacial debris cover of the Hunza Valley in 1990, 2000,

2010, 2013, 2016, and 2019 was estimated based on RF classifier and Landsat images. The results show that the total debris cover area was 394.76 km² in 1990 and 478.88 km² in 2019, indicating that the debris cover on glaciers has increased over the past 30 years.

Identification of debris is a fundamental but challenging field for researching glacier change and water resources. Debris affects the melting rate of glaciers, with thin debris accelerating melting and thick debris slowing melting. Due to the heterogeneity of debris thickness, debris-covered glaciers experience differential ablation, which makes the surface moraine gather or disperse in space. Such gathering or dispersing, in turn, accelerates differential ablation. This interaction results in the formation of supraglacial ponds and ice cliffs, and these affect glacier-derived runoff to a great extent. Moreover, these cliffs and ponds are not only factors that influence the hydrological process but also sources of glacial disasters such as GLOFs, which affect production and living safety in the downstream areas. With the effects of global warming, clean glaciers in the high mountains of Asia are increasingly changing into debris-covered glaciers, which leads to a change of discharge in rivers mainly supplied by glacial meltwater. Consequently, the water resources in basins and future trends will be greatly affected. Glacier meltwater runoff is an important source of freshwater for human production and life. Research by Immerzeel et al. shows that owing to the retreat of mountain glaciers, reduction of snow cover, and widespread shrinkage and disappearance of alpine lakes, a quarter of the global population is at risk of more severe water supply problems (Immerzeel et al., 2019). Overall, studies on the identification of surface features of debris-covered glaciers (e.g., debris thickness, glacial ponds, and cliffs) and their spatial distribution and seasonal variation are meant to indicate the response of debris-covered glaciers to climate change and the effect on hydrology and water resources.

ACKNOWLEDGMENTS

This project has received funding from the National Natural Science Foundation of China (Grant No.41761144075), the Second Tibetan Plateau Scientific Expedition and Research Program (STEP, grant No. 2019QZKK0208), the Research Fund for Introducing Talents of Yunnan University (No. YJRC3201702) and the initiative for the promotion of Yunnan University in the development of a world-class university. Xie Fuming is also supported by the Innovation Fund Designated for Graduate Students of Yunnan University (Grant No. 2019226).

REFERENCES

- Ali, A.F., Xiao, C., Zhang, X., Adnan, M., Iqbal, M. and Khan, G. (2018) Projection of future streamflow of the Hunza River Basin, Karakoram Range (Pakistan) using HBV hydrological model. *J MT SCI-ENGL* 15(10): 2218-2235. doi: 10.1007/s11629-018-4907-4
- Alifu, H., Johnson, B.A. and Tateishi, R. (2016) Delineation of Debris-Covered Glaciers Based on a Combination of Geomorphometric Parameters and a TIR/NIR/SWIR Band Ratio. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 9(2): 781-792. doi: 10.1109/JSTARS.2015.2500906
- Alifu, H., Tateishi, R. and Johnson, B. (2015) A new band ratio technique for mapping debris-covered glaciers using Landsat

- imagery and a digital elevation model. *Int. J. Remote Sens.* 36(8): 2063-2075. doi: 10.1080/2150704x.2015.1034886
- Altena, B., Scambos, T., Fahnestock, M. and Kääh, A. (2019) Extracting recent short-term glacier velocity evolution over southern Alaska and the Yukon from a large collection of Landsat data. *The Cryosphere* 13(3): 795-814. doi: 10.5194/tc-13-795-2019
- Bhambri, R., Hewitt, K., Kawishwar, P., Kumar, A., Verma, A., Snehmami et al. (2019) Ice-dams, outburst floods, and movement heterogeneity of glaciers, Karakoram. *Glob. Planet. Change* 180: 100-116. doi: 10.1016/j.gloplacha.2019.05.004
- Bhambri, R., Hewitt, K., Kawishwar, P. and Pratap, B. (2017) Surge-type and surge-modified glaciers in the Karakoram. *Sci Rep* 7(1): 15391. doi: 10.1038/s41598-017-15473-8
- Bolch, T., Menounos, B. and Wheate, R. (2010) Landsat-based inventory of glaciers in western Canada, 1985–2005. *Remote Sens. Environ.* 114(1): 127-137. doi: 10.1016/j.rse.2009.08.015
- Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984) Classification and Regression Trees. doi: 10.1201/9781315139470
- Chander, G., Markham, B.L. and Helder, D.L. (2009) Summary of Current Radiometric Calibration Coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI Sensors. *Remote Sens. Environ.* 113(5): 893-903. doi: 10.1016/j.rse.2009.01.007
- Dehecq, A., Gourmelen, N., Gardner, A.S., Brun, F., Goldberg, D., Nienow, P.W. et al. (2018) Twenty-first century glacier slowdown driven by mass loss in High Mountain Asia. *Nat. Geosci.* 12(1): 22-27. doi: 10.1038/s41561-018-0271-9
- Dong, J., Xiao, X., Menarguez, M.A., Zhang, G., Qin, Y., Thau, D. et al. (2016) Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. *Remote Sens Environ* 185: 142-154. doi: 10.1016/j.rse.2016.02.016
- Fahnestock, M., Scambos, T., Moon, T., Gardner, A., Haran, T. and Klinger, M. (2016) Rapid large-area mapping of ice flow using Landsat 8. *Remote Sens. Environ.* 185: 84-94. doi: 10.1016/j.rse.2015.11.023
- Garee, K., Chen, X., Bao, A., Wang, Y. and Meng, F. (2017) Hydrological Modeling of the Upper Indus Basin: A Case Study from a High-Altitude Glacierized Catchment Hunza. *Water* 9(1)doi: 10.3390/w9010017
- Garg, P.K., Shukla, A. and Jasrotia, A.S. (2019) On the strongly imbalanced state of glaciers in the Sikkim, eastern Himalaya, India. *Sci Total Environ* 691: 16-35. doi: 10.1016/j.scitotenv.2019.07.086
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. and Moore, R. (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202: 18-27. doi: 10.1016/j.rse.2017.06.031
- Guo, H., Goodchild, M.F. and Annoni, A., (2020). Manual of Digital Earth. *Springer Singapore*doi: 10.1007/978-981-32-9915-3
- Hazel, J., Krutkramelis, K., Mooney, P., Tomschik, M., Gerow, K., Oakey, J. et al. (2016) High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* 342(6160): 850-853. doi: 10.1126/science.1244693
- Immerzeel, W.W., Pellicciotti, F. and Shrestha, A.B. (2012) Glaciers as a Proxy to Quantify the Spatial Distribution of Precipitation in the Hunza Basin. *Mt. Res. Dev.* 32(1): 30-38. doi: 10.1659/MRD-JOURNAL-D-11-00097.1
- Janke, J.R., Bellisario, A.C. and Ferrando, F.A. (2015) Classification of debris-covered glaciers and rock glaciers in the Andes of central Chile. *Geomorphology* 241: 98-121. doi: 10.1016/j.geomorph.2015.03.034
- Kaab, A., Berthier, E., Nuth, C., Gardelle, J. and Arnaud, Y. (2012) Contrasting patterns of early twenty-first-century glacier mass change in the Himalayas. *Nature* 488(7412): 495-8. doi: 10.1038/nature11324
- Khan, A., Naz, B.S. and Bowling, L.C. (2015) Separating snow, clean and debris covered ice in the Upper Indus Basin, Hindukush-Karakoram-Himalayas, using Landsat images between 1998 and 2002. *J. Hydrol.* 521: 46-64. doi: 10.1016/j.jhydrol.2014.11.048
- Kraaijenbrink, P.D.A., Shea, J.M., Pellicciotti, F., Jong, S.M.d. and Immerzeel, W.W. (2016) Object-based analysis of unmanned aerial vehicle imagery to map and characterise surface features on a debris-covered glacier. *Remote Sens. Environ.* 186: 581-595. doi: 10.1016/j.rse.2016.09.013
- Liaw, A. and Wiener, M. (2002) Classification and Regression by RandomForest. *R News* 2: 18-22.
- Liu, S., Yao, X., Shangguan, D., Guo, W., Wei, J., Xu, J. et al. (2015) The contemporary glaciers in China based on the Second Chinese Glacier Inventory. *ACTA GEOGRAPHICA SINICA* 70(1): 3-16. doi: 10.11821/dlxb201501001
- Maxwell, A.E., Warner, T.A. and Fang, F. (2018) Implementation of machine-learning classification in remote sensing: an applied review. *Int. J. Remote Sens.* 39(9): 2784-2817. doi: 10.1080/01431161.2018.1433343
- Mölg, N., Bolch, T., Rastner, P., Strozzi, T. and Paul, F. (2018) A consistent glacier inventory for Karakoram and Pamir derived from Landsat data: distribution of debris cover and mapping challenges. *Earth Syst. Sci. Data* 10: 1807-1827. doi: 10.5194/essd-10-1807-2018
- Ng, H.-F. (2006) Automatic thresholding for defect detection. *Pattern Recognit. Lett.* 27(14): 1644-1649. doi: 10.1016/j.patrec.2006.03.009
- Nuimura, T., Sakai, A., Taniguchi, K., Nagai, H., Lamsal, D., Tsutaki, S. et al. (2015) The GAMDAM glacier inventory: a quality-controlled inventory of Asian glaciers. *The Cryosphere* 9(3): 849-864. doi: 10.5194/tc-9-849-2015
- Patel, A., Prajapati, R., Dharpure, J.K., Mani, S. and Chauhan, D. (2019) Mapping and monitoring of glacier areal changes using multispectral and elevation data: A case study over Chhota-Shigri glacier. *Earth Science Informatics*doi: 10.1007/s12145-019-00388-x

- Paul, F., Barrand, N.E., Baumann, S., Berthier, E., Bolch, T., Casey, K. et al. (2017) On the accuracy of glacier outlines derived from remote-sensing data. *Ann. Glaciol.* 54(63): 171-182. doi: 10.3189/2013AoG63A296
- Paul, F., Bolch, T., Kääb, A., Nagler, T., Nuth, C., Scharrer, K. et al. (2013) The glaciers climate change initiative: Methods for creating glacier area, elevation change and velocity products. *Remote Sens. Environ.* 162: 408-426. doi: 10.1016/j.rse.2013.07.043
- Paul, F., Huggel, C. and Kääb, A. (2004) Combining satellite multispectral image data and a digital elevation model for mapping debris-covered glaciers. *Remote Sens. Environ.* 89(4): 510-518. doi: 10.1016/j.rse.2003.11.007
- Pekel, J.F., Cottam, A., Gorelick, N. and Belward, A.S. (2016) High-resolution mapping of global surface water and its long-term changes. *Nature* 540(7633): 418-422. doi: 10.1038/nature20584
- Pfeffer, W.T., Arendt, A.A., Bliss, A., Bolch, T., Cogley, J.G., Gardner, A.S. et al. (2014) The Randolph Glacier Inventory: a globally complete inventory of glaciers. *J. Glaciol.* 60(221): 537-552. doi: 10.3189/2014JoG13J176
- Quincey, D.J., Braun, M., Glasser, N.F., Bishop, M.P., Hewitt, K. and Luckman, A. (2011) Karakoram glacier surge dynamics. *Geophys. Res. Lett.* 38(18)doi: 10.1029/2011gl049004
- Qureshi, M.A., Yi, C., Xu, X. and Li, Y. (2017) Glacier status during the period 1973-2014 in the Hunza Basin, Western Karakoram. *Quat. Int.*: 1-12. doi: 10.1016/j.quaint.2016.08.029
- Racoviteanu, A. and Williams, M.W. (2012) Decision Tree and Texture Analysis for Mapping Debris-Covered Glaciers in the Kangchenjunga Area, Eastern Himalaya. *Remote Sens.* 4(10): 3078-3109. doi: 10.3390/rs4103078
- Raup, B., Racoviteanu, A., Khalsa, S.J.S., Helm, C., Armstrong, R. and Arnaud, Y. (2007) The GLIMS geospatial glacier database: A new tool for studying glacier change. *Glob. Planet. Change* 56(1-2): 101-110. doi: 10.1016/j.gloplacha.2006.07.018
- Robson, B.A., Nuth, C., Dahl, S.O., Hölbling, D., Strozzi, T. and Nielsen, P.R. (2015) Automated classification of debris-covered glaciers combining optical, SAR and topographic data in an object-based environment. *Remote Sens. Environ.* 170: 372-387. doi: 10.1016/j.rse.2015.10.001
- Rowan, A.V., Egholm, D.L., Quincey, D.J. and Glasser, N.F. (2015) Modelling the feedbacks between mass balance, ice flow and debris transport to predict the response to climate change of debris-covered glaciers in the Himalaya. *Earth Planet. Sci. Lett.* 430: 427-438. doi: 10.1016/j.epsl.2015.09.004
- Scherler, D., Bookhagen, B. and Strecker, M.R. (2011) Spatially variable response of Himalayan glaciers to climate change affected by debris cover. *Nat. Geosci.* 4(3): 156-159. doi: 10.1038/ngeo1068
- Scherler, D., Wulf, H. and Gorelick, N. (2018) Global Assessment of Supraglacial Debris-Cover Extents. *Geophys. Res. Lett.* 45(21): 11,798-11,805. doi: 10.1029/2018gl080158
- Shrestha, M., Koike, T., Hirabayashi, Y., Xue, Y., Wang, L., Rasul, G. et al. (2015) Integrated simulation of snow and glacier melt in water and energy balance-based, distributed hydrological modeling framework at Hunza River Basin of Pakistan Karakoram region. *J. Geophys. Res.-Atmos.* 120(10): 4889-4919. doi: 10.1002/2014jd022666
- Shukla, A. and Garg, P.K. (2019) Evolution of a debris-covered glacier in the western Himalaya during the last four decades (1971–2016): A multiparametric assessment using remote sensing and field observations. *Geomorphology* 341: 1-14. doi: 10.1016/j.geomorph.2019.05.009
- Singh, V. and Goyal, M.K. (2018) An improved coupled framework for Glacier classification: an integration of optical and thermal infrared remote-sensing bands. *Int. J. Remote Sens.* 39(20): 6864-6892. doi: 10.1080/01431161.2018.1468104
- Suykens, J.A.K. and Vandewalle, J. (1999) Least Squares Support Vector Machine Classifiers. *Neural Processing Letters* 9: 293-300. doi: 10.1.1.43.6438
- Tahir, A.A., Chevallier, P., Arnaud, Y., Neppel, L. and Ahmad, B. (2011) Modeling snowmelt-runoff under climate scenarios in the Hunza River basin, Karakoram Range, Northern Pakistan. *J. Hydrol.* 409(1-2): 104-117. doi: 10.1016/j.jhydrol.2011.08.035
- Vezzola, L.C., Diolaiuti, G.A., D'Agata, C., Smiraglia, C. and Pelfini, M. (2016) Assessing glacier features supporting supraglacial trees: A case study of the Miage debris-covered Glacier (Italian Alps). *The Holocene* 26(7): 1138-1148. doi: 10.1177/0959683616632883
- Wu, X., Xiao, Q., Wen, J., Liu, Q., Peng, J. and Li, X. (2014) Advances in uncertainty analysis for the validation of remote sensing products: Take leaf area index for example. *J. Remote Sens.* 18(5): 1011-1023. doi: 10.11834/jrs.20143332
- Yang, Z. (1995) Glacier meltwater runoff in China and its nourishment to river. *Chinese Geographical Science* 5(1): 66-76.
- Zemp, M., Huss, M., Thibert, E., Eckert, N., McNabb, R., Huber, J. et al. (2019) Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016. *Nature* 568(7752): 382-386. doi: 10.1038/s41586-019-1071-0
- Zhang, J., Jia, L., Menenti, M. and Hu, G. (2019) Glacier Facies Mapping Using a Machine-Learning Algorithm: The Parlung Zangbo Basin Case Study. *Remote Sens.* 11(4): 452. doi: 10.3390/rs11040452
- Zhang, Z., Liu, S., Wei, J., Xu, J., Guo, W., Bao, W. et al. (2016) Mass Change of Glaciers in Muztag Ata-Kongur Tagh, Eastern Pamir, China from 1971/76 to 2013/14 as Derived from Remote Sensing Data. *PLoS One* 11(1): e0147327. doi: 10.1371/journal.pone.0147327