

## Threshold-Based Adaptive Gaussian Mixture Model Integration (TA-GMMI) Algorithm for Mapping Snow Cover in Mountainous Terrain

Yonghong Zhang<sup>1,2</sup>, Guangyi Ma<sup>1,2,\*</sup>, Wei Tian<sup>3</sup>, Jiangeng Wang<sup>4</sup> and Shiwei Chen<sup>1,2</sup>

<sup>1</sup>School of Automation, Nanjing University of Information Science & Technology, Nanjing, 210044, China

<sup>2</sup>Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Nanjing University of Information Science & Technology, Nanjing, 210044, China

<sup>3</sup>School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing, 210044, China

<sup>4</sup>School of Atmospheric Physics, Nanjing University of Information Science & Technology, Nanjing, 210044, China

\*Corresponding Author: Guangyi Ma. Email: xzmgy@nuist.edu.cn

Received: 08 April 2020; Accepted: 03 May 2020

**Abstract:** Snow cover is an important parameter in the fields of computer modeling, engineering technology and energy development. With the extensive growth of novel hardware and software compositions creating smart, cyber physical systems' (CPS) efficient end-to-end workflows. In order to provide accurate snow detection results for the CPS's terminal, this paper proposed a snow cover detection algorithm based on the unsupervised Gaussian mixture model (GMM) for the FY-4A satellite data. At present, most snow cover detection algorithms mainly utilize the characteristics of the optical spectrum, which is based on the normalized difference snow index (NDSI) with thresholds in different wavebands. These algorithms require a large amount of manually labeled data for statistical analysis to obtain the appropriate thresholds for the study area. Consideration must be given to both the high and low elevations in the study area. It is difficult to extract all snow by a fixed threshold in mountainous and rugged terrains. In this research, we avoid relying on a manual analysis for different elevations. Therefore, an algorithm based on the GMM is proposed, integrating the threshold-based algorithm and the GMM. First, the threshold-based algorithm with transferred thresholds from other satellites' analysis results are used to coarsely classify the surface objects. These results are then used to initialize the parameters of the GMM. Finally, the parameters of that model are updated by an expectation-maximum (EM) iteration algorithm, and the final results are outputted when the iterative conditions end. The results show that this algorithm can adjust itself to mountainous terrain with different elevations, and exhibits a better performance than the threshold-based algorithm. Compared with orbit satellites' snow products, the accuracy of the algorithm used for FY-4A is improved by nearly 2%, and the snow detection rate is increased by nearly 6%. Moreover, compared with microwave sensors' snow products, the accuracy is increased by nearly 3%. The validation results show that the proposed algorithm can be adapted to a complex terrain environment in mountainous areas and exhibits good performance under a transferred threshold without manually assigned labels.

**Keywords:** Cyber physical systems; FY-4A; snow cover; Gaussian mixture model



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1 Introduction

Snow cover in the Qinghai-Tibetan Plateau (TP) plays an important role in hydrological processes, surface radiation, and climate. The TP exerts profound influences not only on the local climate and environment but also on the global atmospheric circulation through its thermal and mechanical forcing [1]. According to the physical and optical characteristics of snow, it has higher reflectance in the visible and near-infrared bands than other objects, and the reflectivity of pure snow can be as high as 70% or more, which is similar to clouds but different from low-reflective water and land surfaces [2]. The reflectivity of snow and ice in visible wavebands is stronger than that in short-wave infrared bands, and most clouds have a higher reflectance in short infrared bands than snow. Therefore, the optical spectrum in 1.55  $\mu\text{m}$  and 1.75  $\mu\text{m}$  can be used to separate snow from clouds [3]. Hall et al. [4] proposed a global daily snow cover automated mapping algorithm called SNOWMAP by analyzing Landsat/TM data, this algorithm is listed by the National Snow and Ice Data Center (NSIDC) as the standard algorithm for MODIS ice and snow products. SNOWMAP uses a standard grouping technology, which mainly utilizes the normalized difference snow index (NDSI) and other spectral band thresholds to identify snow. However, this algorithm is limited by many factors. For example, the reflectivity of the same object collected by different sensors will be different. Moreover, in low-resolution images, the amount of snow contained in one pixel may impact the reflectivity greatly. According to the analysis of the spectral characteristics of snow, the optical characteristics of snow are particularly similar to those of ice/water in visible light and near-infrared wavebands, which is also the reason why it is difficult to distinguish snow directly by a fixed threshold. Furthermore, its optical reflectivity is also determined by the snow depth, snow grain size, and whether the snow is wet or dry [5]. The geographical and climatic environment of the TP is complicated, and frozen soil and snow cover are mixed. Additionally, the elevation difference between the east and the west is large, and the overlapping mountains lead to different light intensities. Therefore, the snow cover in the TP cannot be obtained by a fixed threshold [4].

As machine learning becomes a hot research area, more and more people are introducing machine learning algorithms to solve remote sensing problems. Duro et al. [6] compared the pixel-based and object-based image classification algorithms by analyzing the application of multiple supervised machine learning algorithms in feature classification; they investigated the decision tree (DT), random tree (RT) and support vector machine (SVM). The results showed that the object-based image classification algorithms exhibited better performance. He et al. [7] proposed an algorithm to map the snow cover in mountainous terrain based on the SVM, and the results showed that this algorithm obtained better performance than the threshold-based algorithm. The successful application of machine learning algorithms in solving remote sensing problems shows that this method has potential in snow identification for moderate resolution satellites' data.

China's new generation of the geostationary satellite FY-4A, has multi-temporal data that are appropriate for cloud removal and snow cover monitoring. The onboard Advanced Geostationary Radiation Imager (AGRI) has 14 channels, that can be used for snow identification. The threshold-based algorithm needs to manually label data and analyze the reflectance of snow before the threshold can be obtained. Hence, we proposed an automated snow cover mapping algorithm based on the GMM for snow cover mapping. First, the coarse classification results are given by the threshold-based algorithm. Then the coarse classification results are used to initialize the parameters of the GMM. Finally, the EM iteration algorithms are used to update the parameters of the GMM and check the iterated condition. Lastly, the classification results are outputted when the iterated conditions are satisfied. Since this algorithm is mainly based on unsupervised machine learning algorithms, it can adapt to different conditions. This algorithm does not require manual labels, and it can be used for the TP without any adjustments.

We briefly review previous work on snow cover detection in Section 2, present a detailed description of our datasets and method in Sections 3 and 4, and discuss the experimental validation of our method in Section 5. Conclusions are summarized in Section 6.

## 2 Related Work

Research on Alpine glacier variations in the TP has achieved great success during the last few decades [8–10]. Snow cover has received extensive attention in research [4,11,12]. Remote sensing techniques are useful for acquiring near-real time snow cover data. There are several remote sensing algorithms and satellite data used for monitoring snow cover [4,13–15]. For example, MODIS swath snow and ice products [4], IMS [14], GlobSnow [13], and Fengyun-2 series snow products [16]. These products are derived from optical sensors or microwave sensors. One major drawback of optical sensors such as those used by MODIS is their inability to provide surface information under cloudy conditions. This strongly limits the number of useful snow observations as abundant cloud cover is present in the Alpine region, particularly during the winter season. Meanwhile, the major drawback of microwave sensors is that snow detection is impacted by complicated environments, particularly in the Alpine region [17].

Various techniques have been developed to remove cloud masks from optical images. They mostly include multi-day maximum snow cover composites [18], a combination of microwave and similar data [13], and multi-temporal satellite data [19]. At present, the most used satellite orbit for snow detection is the polar orbit, which enables high spatial resolution. The National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) has been used to map the spatial extent of snow cover [20]. Hall et al. [21] proposed a method for mapping global snow cover using MODIS data. Gascoin et al. [22] introduced high-resolution operational snow cover maps from Sentinel-2 and Landsat-8 data. Pepe et al. [23] exploited the combined use of the ENVISAT Medium Resolution Imaging Spectrometer Instrument (MERIS) and Advanced Along Track Scanning Radiometer (AATSR) acquisitions and topographic data for mapping snow cover. In high latitudes, where snow is most often present, the instruments on board geostationary satellites have low viewing angles, resulting in poor spatial resolution. Although the polar orbit satellites offer a global view, they produce few daily images (1–4 images) of a certain region. However, instruments on geostationary satellites offer excellent temporal resolution data. This advantage can be used for cloud removal because the clouds move out. De Ruyter de Wildt et al. [24] proposed an algorithm for deriving snow cover maps from Spinning Enhanced Visible and Infrared Imager (SEVIRI) on board the second generation of Meteosat (MSG) satellites. Furthermore, Siljamo et al. [25] introduced a new geostationary snow cover product for SEVIRI on board the MSG satellites. Lastly, Yang et al. [19] proposed a method for snow detection using a combination of the Visible and Infrared Spin Scan-Radiometer (VISSR) on board the geostationary satellites FY-2D and FY-2E and the Microwave Radiation Imager (MWRI) on board the polar orbiting satellite FY-3B.

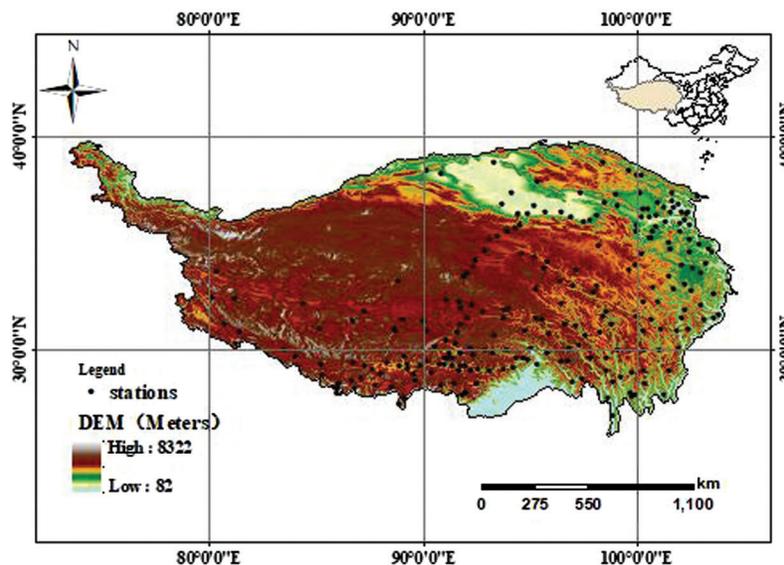
Currently, most methods for snow-cover detection based on geostationary satellite data utilize spectral-based threshold algorithms. This method depends on too many criteria and statistics of the threshold. Recent advances in machine learning have convincingly demonstrated high capability in learning remote sensing models with large datasets. Duro et al. [6] compared pixel-based and object-based image analysis approaches for classifying broad land cover classes over agricultural landscapes using three supervised machine learning algorithms: the decision tree (DT), random forest (RF), and the support vector machine (SVM). Qian et al. [26] compared the performance of four machine learning classifiers-SVM, normal Bayes (NB), classification and regression tree (CART) and K nearest neighbor (KNN) to classify very high resolution images. Zhan [27] proposed a deep learning system to classify cloud and snow with fully convolutional neural networks at the pixel level. Furthermore, Huang et al. [28] proposed a method using

the SVM to classify the glacier areas. This means the machine learning algorithm has potential in snow cover mapping using multi-spectral data with low spatial resolution.

### 3 Description of the Study Area and Data

#### 3.1 The Qinghai-Tibetan Plateau

The TP is located in southwest China between the Pamirs and Hengduan mountains. It covers an area of  $2.96 \times 10^6 \text{ km}^2$  at an average elevation of 4500 m [29]. Fig. 1 is the digital elevation model (DEM) of the TP, and the color bar indicates the altitude. It can be seen from Fig. 1 that there is a great elevation difference between the eastern and western parts of the TP. The western region of the TP is the Himalayas, with an elevation of more than 4 km. There are many mountains as high as 5 km or even 7 km above sea level, where the geographic environment is complex. The eastern part, at the end of the Himalayas, is lower than the western part, usually below 4 km [30]. The mean annual temperature decreases from the edges of the plateau towards the center with altitude, ranging from  $-15^\circ\text{C}$  to  $10^\circ\text{C}$  [31]. According to the research of Gao et al. [32], the total cloud cover in the TP is generally between 45% and 70%, reaching maximum coverage in winter.



**Figure 1:** Distribution of meteorological stations over Qinghai-Tibetan Plateau in China. The dots show the meteorological stations and the color bar indicates the altitude of the Qinghai-Tibetan Plateau

#### 3.2 The Situ Station in TP

The data used in this research were collected from February 6, 2019 to February 11, 2019 and were provided by the China Meteorological Administration (CMA). The distribution of in situ stations over TP in China are shown in Fig. 1. The black dots show the geographical position of the stations, including both manual and automatic weather stations. These data contain information of snow and precipitation. The collection time is from 8:00 BJT to 8:00 BJT the next day.

#### 3.3 Description of the FY-4A Data

FY-4A is situated at  $104.7^\circ\text{E}$  over the equator after May 25 2017. AGRI can generate an observation image of the China region within one minute sometime, and the multi-spectral data has 14 channels (Tab. 1). Channel 2 is a high-resolution visible channel that has a spatial resolution of 0.5 km at the

**Table 1:** FY-4 advanced geostationary radiation image (AGRI) bands (1 km resolution)

Number	Channel	Spatial resolution (km)	Central wavelength ( $\mu\text{m}$ )	Wavelength ( $\mu\text{m}$ )
1	VIS-blue	1	0.47	0.45–0.49
2	VIS-red	0.5–1	0.65	0.55–0.75
3	NIR	1	0.83	0.75–0.90
4	SWIR	2	1.375	1.36–1.39
5		2	1.61	1.58–1.64
6		2–4	2.25	2.1–2.35
7	MWIR	2	3.75	3.5–4.0 (high)
8		4	3.75	3.5–4.0 (low)
9	WV	4	6.25	5.8–6.7
10		4	7.1	6.9–7.3
11	LWIR	4	8.5	8.0–9.0
12		4	10.7	10.3–11.3
13		4	12.0	11.5–12.5
14		4	13.5	13.2–13.8

sub-satellite point (SSP). Channels 1 and 3 have spatial resolutions of 1 km, and channels 4, 7 have spatial resolutions of 2 km at the SSP. All of the other channels have a spatial resolution of 4 km at the SSP. The wavebands of AGRI can be grouped into visible (VIS), near infrared (NIR), short-wavelength infrared (SWIR), mid-wavelength infrared (MWIR), water vapor (WV) and long-wavelength infrared (LWIR). We used data from 2019 to develop and test the method.

### 3.4 Comparison Data

A series of Moderate-Resolution Imaging Spectroradiometer (MODIS) snow products at various spatial and temporal resolutions from the Terra satellite has been available since February 2000 [18]. In this research the MODIS snow products MOD10A1 and MYD10A1 V6 from February 6, 2019 to February 11, 2019 are used for comparison with FY-4A results. The MOD10A1 and MYD10A1 are derived from Terra and Aqua satellites' data, which transit TP in the morning and afternoon, respectively. Additionally, the snow products derived from microwave sensors are used for comparison. The GlobSnow SE V2 database with 1 km resolution is used, which was established by the European Space Agency's Data User Element (DUE) project GlobSnow to create a snow extent (SE) and snow water equivalent (SWE) database [13]. IMS snow products are produced and provided by the National Oceanic and Atmospheric Administration (NOAA). This product is derived from different data, including that from the NOAA polar orbit satellite, GOES, GMS, the MT-SAT multi-functional transport satellite, METEO-SAT, US Department of Defense (US DOD) polar orbit satellites, American National Defense Meteorological Satellite Program (DMSP), and US National Ice Center (NIC) Sea Ice products. In this paper, the IMS v3 product is used. In order to compare the results of this article with the snow products of geostationary satellites, this article uses the Fengyun-2 snow product for comparison. The FY-2 swath snow products are derived from FY-2F, FY-2G and FY-2H.

## 4 Mathematical Technology

An automated snow cover mapping algorithm is proposed in this research. This algorithm can be used for rugged terrain without analyzing a fixed threshold for different object surfaces. This algorithm is divided

into three steps, a threshold-based algorithm is improved according to the terrain. The mask obtained from the first step can be used as a pseudo label to reclassify the snow by the GMM. Finally, the parameter of the GMM is iterated with the EM algorithm and the classification is outputted when the iteration conditions are satisfied.

#### 4.1 Description of the Threshold-Based Algorithm for Mapping Snow Cover

##### 4.1.1 Snow Cover Consideration

Dozier et al. [3] found that planetary reflectance of snow  $R_p$  in Landsat TM 0.45  $\mu\text{m}$  is greater than about 0.16;  $R_p(1.57)$  is less than about 0.2; and the NDSI is greater than about 0.4. Since the AGRI on board FY-4A contains a 1.6  $\mu\text{m}$  (FY Band 5) channel, the value of the NDSI can be calculated by FY-4A/AGRI data. FY is the FY-4A's abbreviation. The NDSI may be calculated using FY-4A/AGRI data as follows:

$$\text{NDSI} = \frac{\text{FY Band 2} - \text{FY Band 5}}{\text{FY Band 2} + \text{FY Band 5}} \quad (1)$$

Considering the complex environment of TP, some other wavebands will be used to assist the identification. Snow and ice are considerably more reflective in the visible spectrum than in the short-wave IR part of the spectrum, and the reflectance of most clouds remains high in the short-wave IR, while the reflectance of snow is low [34]. These bands mainly correct the pixels that are misidentified as snow. The primary change in reflectance occurs in the visible wavelengths as snow has a much higher visible reflectance than soil, leaves or trees [18]. This behavior can be captured by the normalized difference vegetation index (NDVI) [34]. If the NDVI =  $\sim 0.1$ , the pixel may be mapped as snow even if the NDSI is  $< 0.4$  [37]. The NDVI using FY-4A/AGRI can be formulated as follows:

$$\text{NDVI} = \frac{\text{FY Band 3} - \text{FY Band 2}}{\text{FY Band 3} + \text{FY Band 2}} \quad (2)$$

##### 4.1.2 Cloud Cover Consideration

The NDSI is capable of separating most snow and clouds [19]. The combination of  $r_{0.64} > 0.25$  and  $r_{1.6} > 0.30$  can be used to distinguish many clouds from snow, where  $r$  represent the reflectivity of the objects [3]. Cumulus clouds are generally readily distinguished from snow because the reflectance of cumulus clouds remains high in the region of the spectrum from 1.58 to 1.64  $\mu\text{m}$  (FY Band 5), whereas the reflectance of snow drops. In the same way, the cirrus clouds remain high in the region of the spectrum from 1.36 to 1.39  $\mu\text{m}$  (FY Band 4). Since the elevation of the TP is high, the clouds may be higher than that in the plain. The range of 1.36–1.39  $\mu\text{m}$  is very suitable for distinguishing between high clouds and surface snow, because it mainly detects radiation from levels above 500 hPa and hardly any surface radiation [24]. Thus, the clouds can be distinguished by the combination of 0.65  $\mu\text{m}$  (FY Band 2), 1.61  $\mu\text{m}$  (FY Band 5), and 1.375  $\mu\text{m}$  (FY Band 4). If  $r_{0.65} > 0.25$  and  $r_{1.61} > 0.30$  or  $r_{1.375} > 0.058$ , the pixel will be regarded as cloud.

##### 4.1.3 Ice Cover Consideration

According to the research of Paul et al. [38], channels 3, 4, and 5 of Landsat Thematic Mapper (TM) can be used for glacier debris recognition. FY bands 2, 3 and 4 are used to separate glacier debris from snow. The formulation is noted as formulation (3), where  $a$  and  $b$  denote constants.

$$\frac{\text{FY Band 3}}{\text{FY Band 4}} < a * \frac{\text{FY Band 2}}{\text{FY Band 4}} + b \quad (3)$$

##### 4.1.4 Water Mask Consideration

Xu [39] proposed a modified normalized difference water index (MNWDI) to detect the water body. This algorithm was based on the NDWI [40]. Xu found that the NDWI could not completely suppress the

information irrelevant to the water body, and many non-water body information, especially the water body information within the city, were still included in the extraction of the water body information. Hence, he proposed the MNDWI to identify water bodies. This index uses the green and moderate infrared wavebands, which are denoted as  $R_{Green}$  and  $R_{MIR}$ .  $R_{Green}$  and  $R_{MIR}$  are the radiances in the green visible and moderate infrared wavebands. The MNDWI can be formulated as formulation (4) when using FY-4A/AGRI data, and the MNDWI > 0.1 can be thought of as a body of water.

$$\text{MNDWI} = \frac{\text{FY Band 2} - \text{FY Band 5}}{\text{FY Band 2} + \text{FY Band 5}} \quad (4)$$

## 4.2 Use of Gaussian Mixture Model for Filtering

### 4.2.1 Gaussian Mixture Model

The gaussian mixture model (GMM) is used to estimate the probability density of samples [35]. We assume that different object surfaces satisfy a normal distribution with different parameters, and the GMM is seen as a linear combination of probability density function of normal distribution. As seen in formulation (5), the distribution is determined by the mean  $\mu$  and variance  $\Sigma$  of the object. The linear combination of (5) is formulated in (6), where  $\alpha_i$  is the mixed coefficient. The sum of  $\alpha_i$  is 1, and is denoted as  $\sum_{i=1}^k \alpha_i = 1$ .  $k$  is the total number of class. The input of (6) is FY-4A/AGRI 2 km imagery with 7 channels, which is denoted as  $x_j = \{x_j^1, x_j^2, \dots, x_j^7\}$ , where  $x_j \in R^{H \times W \times 7}$ .  $H$  and  $W$  means height and width of the image, respectively.  $z_j$  is the prediction class of  $x_j$ . The posterior probability distribution of different objects can be denoted as  $P_M(z_j = i | x_j)$ , and is calculated by formulation (7).

$$p(x|\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)} \quad (5)$$

$$p_M(x) = \sum_{i=1}^k \alpha_i \cdot p(x|\mu_i, \Sigma_i) \quad (6)$$

$$p_M(z_j = i | x_j) = \frac{p(z_j = i) \cdot p_M(x_j | z_j = i)}{p_M(x_j)} \quad (7)$$

### 4.2.2 Updating the Parameters

The expectation-maximization (EM) algorithm allows us to estimate the value of potential parameters. The EM algorithm can be divided into two steps. First, the values of  $\gamma_{ji}$  are calculated according to the initial parameters (E step). Then, the values of parameters  $\mu_i$ ,  $\Sigma_i$  and  $\alpha_i$  are updated during the iteration using the following formulations (M step). In the following formulations, the  $P_M(z_j = i | x_j)$  is denoted as  $\gamma_{ji}$  for simplicity.

$$\mu_i = \frac{\sum_{j=1}^m \gamma_{ji} x_j}{\sum_{j=1}^m \gamma_{ji}} \quad (8)$$

$$\Sigma_i = \frac{\sum_{j=1}^m \gamma_{ji}(x_j - \mu_i)(x_j - \mu_i)^T}{\sum_{j=1}^m \gamma_{ji}} \tag{9}$$

$$\alpha_i = \frac{1}{m} \sum_{j=1}^m \gamma_{ji} \tag{10}$$

### 4.3 Adaptive Classification Model

We proposed an adaptive classification model called the threshold-based adaptive Gaussian mixture model integration (TA-GMMI) algorithm. In this method, the masks of different objects are marked using the threshold-based algorithm. These coarse classification results are used in the initial process of the GMM. A flow chart of different objects is shown in Fig. 2. The parameters in Algorithm 1 lines 2–4 are initialized using the masks of cloud, snow, ice, water, and soil. The loop started in the first line of the algorithm acts to superimpose all of the data in a day to find the largest union set that belongs to the snow category. The EM algorithm is used to update the parameters of the GMM. In lines 6 and 7, the posterior probability is first calculated using the initialized parameters of means and covariance (E step). The purpose of initializing the parameters is to obtain the specific classification object. Then the new values of these parameters are updated by the formulation in lines 9–13 (M step). In lines 17–20, the different objects are classified into different classes. Line 21 shows the maximum union of all recognition results in a day.

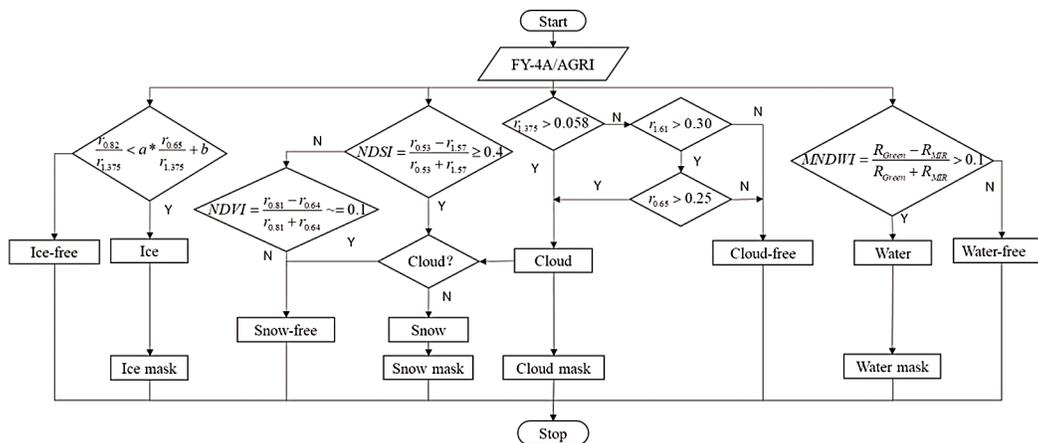


Figure 2: The flow chart of different objects

### 4.4 Assessment Criterion

The results of TA-GMMI will be compared with in situ station data. The confusion matrix comparing the daily snow cover product data against in situ observations is shown in Tab. 2. *TP* and *TN* represent the total number of correct snow and no snow detections, respectively. *FP* and *FN* represent the total number of incorrect snow and no snow detections, respectively. We use overall accuracy, snow detection rate, omission error and commission error as evaluation indicators.

---

**Algorithm 1:** TA-GMMI algorithm
 

---

Input: Dataset  $D = \{x_1, x_2, \dots, x_m\}$ , where  $x_d = \{x_1^d, x_2^d, \dots, x_p^d \mid d = 1, 2, \dots, m\}$ .

Process:

- 1: for  $n = 1, 2, \dots, m$  do:
- 2: use threshold to achieve the coarse classification results. The pixels of the figure can be classified into different classes  $\{C_1, C_2, \dots, C_k\}$ , and  $C_{\lambda_j} = C_{\lambda_j} \cup \{x_j^n\}, 1 \leq j \leq p$ .
- 3: calculate the means of different classes using the coarse classification results.  $\mu = \{\mu_1, \mu_2, \dots, \mu_k\}$ ,  $\Sigma = \{\Sigma_1, \Sigma_2, \dots, \Sigma_k\}$ .
- 4: initialize the parameters of the Gaussian mixture model  $\{\alpha_i, \mu_i, \Sigma_i \mid 1 \leq i \leq k\}$
- 5: repeat:
- 6: for  $j = 1, 2, \dots, p$  do:
- 7:  $\gamma_{ji} = p_M(z_j = i \mid x_j) (1 \leq i \leq k)$
- 8: end for
- 9: for  $i = 1, 2, \dots, k$  do:
- 10: update  $\mu_i'$  by using formulation (8)
- 11: update  $\Sigma_i'$  by using formulation (9)
- 12: update  $\alpha_i$  by using formulation (10)
- 13: end for
- 14: update  $\{\alpha_i, \mu_i, \Sigma_i \mid 1 \leq i \leq k\}$  with  $\{\alpha_i', \mu_i', \Sigma_i' \mid 1 \leq i \leq k\}$
- 15: until that satisfy the condition that
- 16:  $C_i = \emptyset (1 \leq i \leq k)$

Continuation of Algorithm 1

- 17: for  $j = 1, 2, \dots, p$  do:
- 18:  $\gamma_j = \arg \max_{i \in \{1, 2, \dots, k\}} \gamma_{ji}$
- 19:  $C_{\lambda_j} = C_{\lambda_j} \cup \{x_j\}$
- 20: end for
- 21:  $C_{snow} = C_{\lambda_j}^d \cup \{x_j^{d-1}\}$
- 22: end for

Output: the results of the classification  $C_{snow}$

---

The overall accuracy is defined as

$$\text{Overall accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (11)$$

The snow detection rate is defined as the ratio of snow cover pixels detected by snow products to the total number of snow cover events:

$$\text{Snow detection rate} = \frac{TP}{TP + FN} \quad (12)$$

$FN$  represents the total number of missing snow pixels, which indicates that the snow product identified uncovered land (no snow) at a certain location, but snow was present in the in situ observation. This event was described as an omission error:

$$\text{Omission error} = \frac{FN}{TP + FN + FP + TN} \quad (13)$$

When snow product data indicated snow cover, but the in situ observation was snow-free, the event was labeled as a commission error. The commission error is formulated as follows:

$$\text{Commission error} = \frac{FP}{TP + FN + FP + TN} \quad (14)$$

**Table 2:** Confusion matrix comparing daily snow cover product against in situ observations

Ground Observation	Daily Snow-Cover Product	
	Snow	No Snow
Snow	$TP$	$FN$
No snow	$FP$	$TN$

Continuous indices are employed to evaluate the skill of each algorithm in snow detection. Statistics in this category include the root-mean-square error (RMSE), correlation coefficient (CC), and mean absolute error (MAE), which are calculated by the following equations:

$$RMSE = \frac{1}{n} \sqrt{\sum_{i=1}^n (Sim_i - Obs_i)^2}, \quad (15)$$

$$CC = \frac{\frac{1}{n} \sum_{i=1}^n (Sim_i - \overline{Sim_i})(Obs_i - \overline{Obs_i})}{\sigma_{Sim} \sigma_{Obs}}, \quad (16)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Sim_i - Obs_i|, \quad (17)$$

where “Sim” stands for simulation and “Obs” stands for truth labels.

## 5 Experiment

### 5.1 Comparison with Other Snow Products

In order to verify the feasibility of the algorithm, we use a variety of international mainstream snow products. In addition to snow products derived from optical sensor data, snow products derived from microwave sensors’ data are also used for comparison.

As can be seen from [Tab. 3](#), the TA-GMMI achieves the highest overall accuracy at 77.9% in this research, demonstrating an increase of by nearly 2% compared with MOD10A1, and nearly 1% compared with MYD10A1; moreover, the snow detection rate increases by nearly 6%. This means that

**Table 3:** Results of different snow products

Snow products	Assessment Index			
	OA	SD	OE	CE
MOD10A1	75.79	7.42	21.20	3.01
MYD10A1	76.44	39.13	21.72	1.83
GlobSnow SE	76.70	2.28	22.38	0.910
IMS	73.69	28.57	16.36	9.94
FY-2F	74.86	57.95	9.63	15.49
FY-2G	72.82	63.42	8.32	18.85
FY-2H	71.87	57.38	9.76	18.35
FY-4A (ours)	77.09	45.14	12.56	10.34

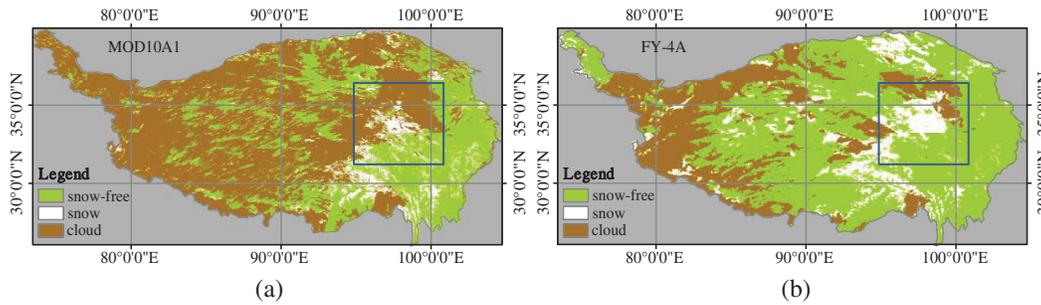
TA-GMMI used for FY-4A/AGRI multi-temporal data is not only feasible but can get more accurate snow detections than polar orbit satellite's snow products under clouds. Compared with the snow products derived from the FY-2 swath satellite, the snow detection rate is low and the omission error is relatively high; however, its commission error is the lowest of several geostationary satellite snow products (FY-2 swath snow products). This indicates that FY-2 series snow cover products are more inclined to recognize snow pixels, resulting in a higher commission error. It can be seen that the performance of the proposed algorithm based on Fengyun-4 satellite data is between that of ice and snow products made from polar orbit satellite and geostationary orbit satellite data, and can guarantee high identification accuracy and high snow detection rate. By comparing the results of the proposed algorithm to those of the microwave snow cover products GlobSnow SE and IMS, their accuracy is improved by 0.39% and 3.4%, respectively, and the snow detection rate is increased by 42.86% and 16.57%, respectively. Thus, our algorithm can detect more snow cover than microwave snow cover products. The microwave snow products are not disturbed by clouds, which means that the multi-temporal data of FY-4A can detect snow cover under clouds. Furthermore, the snow detection rate is also reduced: 3.8% lower than that of IMS and 9.82% lower than that of GlobSnow SE. This decrease indicates that there are more pixels missed by microwave snow products in cloud-free pixels, and some of these pixels could be detected by TA-GMMI algorithm. However, the reason why the accuracy of microwave data in Qinghai-Tibetan Plateau is not very high may be that the accuracy is affected by terrain, snow depth and so on.

### 5.2 Comparison of FY-4A and MOD10A1

MODIS snow cover products have good performance under clear skies [18]. Cloud obscuration is the main limit of the MODIS snow cover products. The daily snow cover product MOD10A1 is compared with FY-4A images Fig. 3a shows the result of MOD10A1 while Fig. 3b shows the result of TA-GMMI for FY-4A/AGRI. In order to verify the capability of cloud removal using FY-4A, we chose a cloudy day. It can be seen that the TA-GMMI detected more snow in the middle of the TP (blue box). In Fig. 3a, the snow in the middle of the TP cannot be identified because of cloud obscuration. In Fig. 3b, we can see that the cloud content is very low. In the middle of the TP, the cloud can be removed by the multi-temporal data of FY-4A/AGRI.

### 5.3 Ablation Experiment

In order to verify that the TA-GMMI algorithm will improve the recognition accuracy, we compare this algorithm to the threshold-based algorithm and MODIS daily snow products. We compare the results with



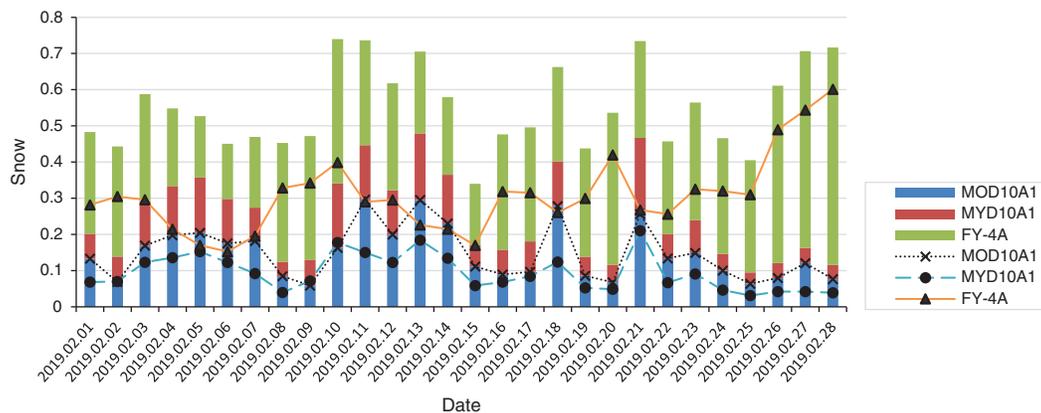
**Figure 3:** Results of (a) MOD10A1 and (b) FY-4A at 04:00UTC in February 1, 2019

**Table 4:** Results of ablation experiment

Snow products	Algorithm	MAE (%)	RMSE (%)	CC (%)
MOD10A1	SNOWMAP	34.04	71.35	92.09
MYD10A1	SNOWMAP	43.39	83.51	91.20
FY-4A	Threshold-based algorithm	12.64	47.39	91.49
FY-4A	TA-GMMI	6.77	23.67	92.19

MODIS snow products to show that the use of FY-4A/AGRI data can reduce identification errors. We use the mean absolute error (MAE), root mean square error (RMSE) and correlation coefficient (CC) to measure the results, which are shown in Tab. 4.

The comparison of results show that the threshold-based algorithm and TA-GMMI algorithm used for FY-4A/AGRI data exhibit better performance than MODIS snow products. From Tab. 4, we can see that the MAE and RMSE of snow cover products derived from FY-4A/AGRI data is much lower than that of MODIS. The main reason for this is that the objects can be detected when the cloud moves away. Since MOD10A1 and MYD10A1 obtain one image each day, it is difficult for them to identify the objects clearly under the cloud. Compared with the threshold-based algorithm, the MAE and RMSE of TA-GMMI are decreased, and the CC is improved. This means that the TA-GMMI algorithm can achieve more accurate results than the threshold-based algorithm. This also indicates that the TA-GMMI can be adaptive without adjusting the threshold.



**Figure 4:** Statistical results of the MOD10A1, MYD10A1 and FY-4A in the TP

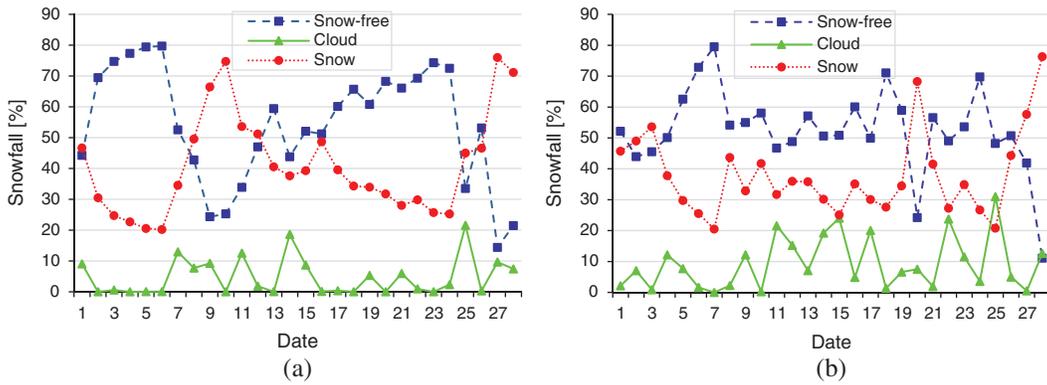


Figure 5: Statistical results in snow, and cloud cover in (a) Shigaze and (b) Yushu

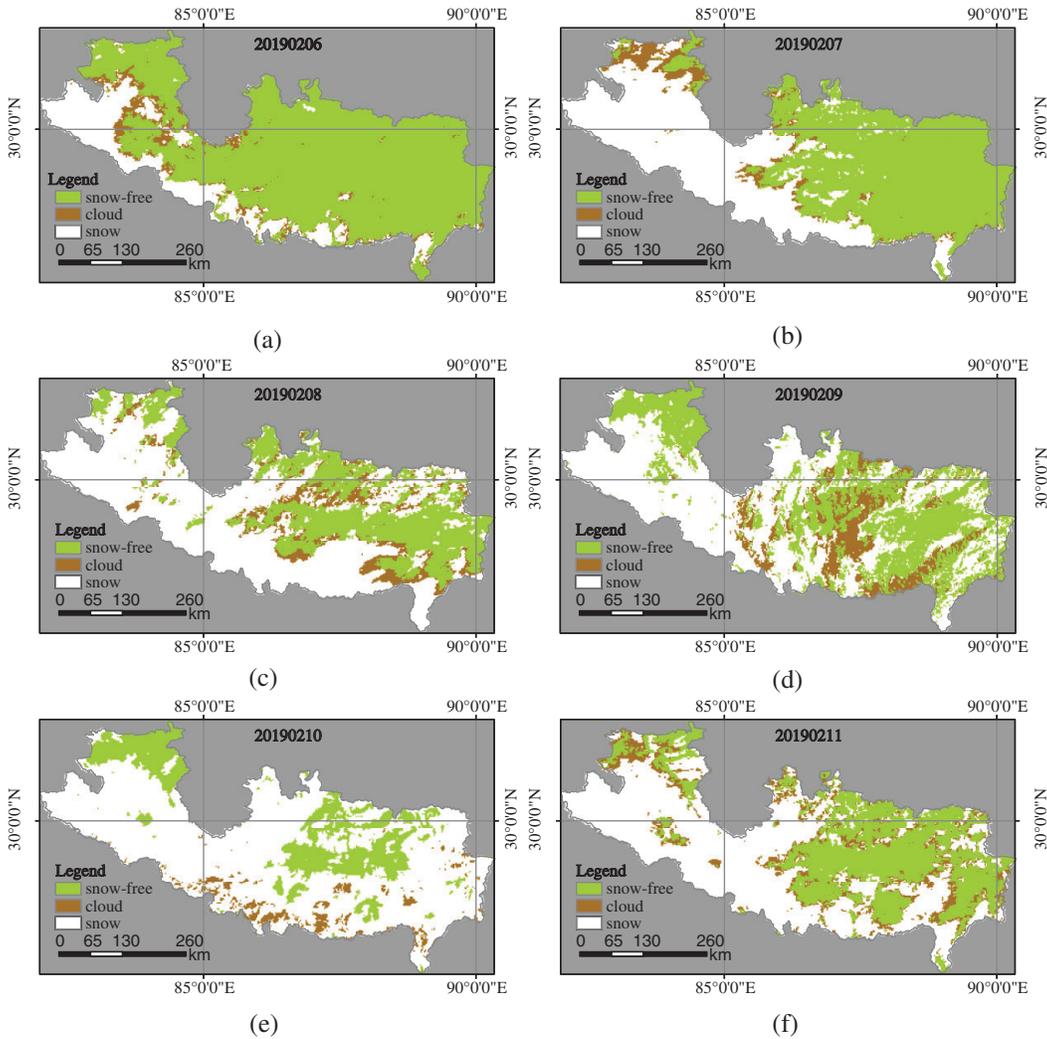
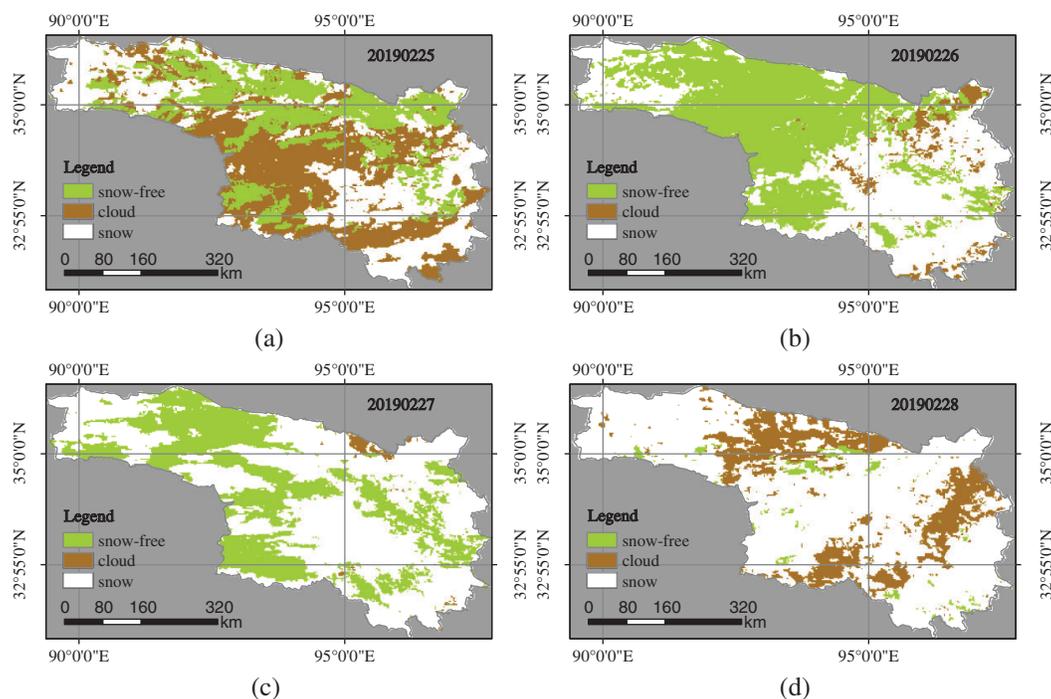


Figure 6: Changes in snow cover in Shigatze from February 6, 2019 to February 11, 2019

### 5.4 Evolution Analysis

The snowfall statistics of the TP, Shigatze and Yushu regions are used for analysis. The statistics are presented in Figs. 4 and 5 to visually show the changes in snowfall. Figs. 4, 5a and 5b show the changes in snow and clouds in the TP, Shigatze and Yushu regions, respectively. The chart for Shigatze shows that the snow cover lasts from February 6, 2019 to February 11, 2019. The daily snow cover mapping results in Shigatze (Fig. 6) shows that after February 6, 2019, the snow cover expanded from the southwest to the northeast and reached the widest range on February 10, 2019. Fig. 7 contains the snow cover map in Yushu from February 25, 2019 to February 28, 2019. The chart of snow cover in Yushu shows that the snow levels were high in February and extended largely after February 25. According to the daily snow maps, we can see that the snow coverage increased from February 25 to February 28, and the coverage was the widest on February 28, covering almost the entire Yushu area. Past news and weather forecast data show that heavy snowfall occurred in Shigatze from the 6th to 10th of February, especially in Keelung, Nyalam County and some townships. Heavy snowfall also occurred in Yushu from the 25th to 28th in February. For precipitation and surface temperature data, please refer to the website <http://data.cma.cn/data/online/t/1>. According to the past weather forecast, precipitation data and news, the time of the two massive snowfall is basically the same as the time we monitored the snow cover change. Furthermore, the snowfall occurred in the same location. Accurate monitoring of the two massive snowfalls shows that by using FY-4A data and the TA-GMMI algorithm we can accurately detect snow changes and greatly reduce cloud interference. It also shows that the algorithm proposed in this paper is feasible and reliable and can be used for daily snow monitoring tasks.



**Figure 7:** Changes in snow cover in Yushu from February 25, 2019 to February 28, 2019

Comparing the data of MOD10A1 and MYD10A1 at the same period (Fig. 4), the snow cover of MOD10A1 and MYD10A1 fluctuated greatly from February 6 to February 10. There is a decrease in the TP between February 6 and February 9, because the optical sensor is obscured by clouds during snowfall. The MOD10A1 and MYD10A1 data were inconsistent with the observations of previous news and

weather forecast data. This indicates that the change has a great lag, and it is impossible to accurately and timely express the changes in snow cover in the TP. This finding also proves that the results of this paper are very useful for snow cover monitoring of the TP. From February 25 to February 28, heavy snowfall occurred in the TP, but MOD10A1 and MYD10A1 did not detect snow fluctuations. This is because ground objects cannot be detected due to cloud occlusion, therefore, the changes in snow cover on the surface cannot be captured. Ultimately, the results of TA-GMMI can be used for snow cover monitoring in the TP. Moreover, these results can provide stable and accurate monitoring results and reflect the real change in snow cover.

## 6 Conclusion

In this study, an algorithm based on the Gaussian mixture model is proposed for FY-4A multi-temporal images. It integrates the threshold-based algorithm and the Gaussian mixture model algorithms. In order to verify the feasibility of this hybrid algorithm, we compared its results with those of mainstream snow products. Compared with snow products with only one image per day, nearly 90% of cloud cover can be reduced. Compared with the snow cover products, our algorithm can obtain a lower cloud cover and a higher snow detection rate. Contrasting geostationary satellites' snow cover products, this algorithm can obtain a lower commission error. Moreover, compared to polar orbit satellite's snow cover products, our algorithm can obtain a lower omission error. The results show that FY-4A/AGRI's multi-temporal image can effectively reduce the interference caused by clouds. The proposed algorithm also exhibits a better performance than the threshold-based algorithm. This means that our algorithm can adjust to complex mountainous terrain under a fixed threshold. This will greatly reduce the difficulty of manual labeling. Finally, we take the two snow disasters that happened on the Qinghai-Tibetan Plateau as examples. The snow content in the TP is counted, and a line chart shows that the snow cover underwent two huge changes in February 2019. A comparison to meteorological data and news reports shows that the two changes are due to the time and place of the snowfall. The successful monitoring process shows that the algorithm is feasible and can conduct snow monitoring.

**Acknowledgement:** Thanks to the National Meteorological Information Center of China Meteorological Administration for offering the meteorological data.

**Funding Statement:** This study was jointly supported by National Science Foundation of China (41661144039, 41875027 and 41871238).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

1. Li, C., Su, F., Yang, D., Tong, K., Kan, B. et al. (2018). Spatiotemporal variation of snow cover over the Tibetan Plateau based on MODIS snow product, 2001–2014. *International Journal of Climatology*, 38(2), 708–728. DOI 10.1002/joc.5204.
2. Hou, H. S., Yang, H. Y. (2009). A general introduction to MODIS snow products and its researching application. *Remote Sensing Technology and Application*, 24(02), 252–256.
3. Dozier, J. (1989). Spectral signature of alpine snow cover from the Landsat Thematic Mapper. *Remote Sensing of Environment*, 28(1), 9–22. DOI 10.1016/0034-4257(89)90101-6.
4. Hall, D. K., Riggs, G. A., Salomonson, V. V., DiGirolamo, V. V., Bayr, K. J. (2002). MODIS snow-cover products. *Remote Sensing of Environment*, 83(1–2), 181–194. DOI 10.1016/S0034-4257(02)00095-0.
5. Xu, W., Ma, H., Wu, D., Yuan, W. (2017). Assessment of the daily cloud-free MODIS snow-cover product for monitoring the snow-cover phenology over the Qinghai-Tibetan plateau. *Remote Sensing*, 9(6), 585. DOI 10.3390/rs9060585.

6. Duro, D. C., Franklin, S. E., Dubé, M. G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment*, 118, 259–272. DOI 10.1016/j.rse.2011.11.020.
7. He, G., Xiao, P., Feng, X., Zhang, X., Wang, Z. et al. (2015). Extracting snow cover in mountain areas based on SAR and optical data. *IEEE Geoscience and Remote Sensing Letters*, 12(5), 1136–1140. DOI 10.1109/LGRS.2014.2386275.
8. Mayewski, P. A., Jeschke, P. A. (1979). Himalayan and Trans-Himalayan glacier fluctuations since AD 1812. *Arctic and Alpine Research*, 11(3), 267–287. DOI 10.2307/1550417.
9. Bishop, M. P., Olsenholler, J. A., Shroder, J. F., Barry, R. G., Raup, B. H. et al. (2004). Global land ice measurements from space (GLIMS): remote sensing and GIS investigations of the Earth's cryosphere. *Geocarto International*, 19(2), 57–84. DOI 10.1080/10106040408542307.
10. Ye, Q., Kang, S., Chen, F., Wang, J. (2006). Monitoring glacier variations on Geladandong mountain, central Tibetan Plateau, from 1969 to 2002 using remote-sensing and GIS technologies. *Journal of Glaciology*, 52 (179), 537–545. DOI 10.3189/172756506781828359.
11. Pu, Z., Xu, L., Salomonson, V. V. (2007). MODIS/Terra observed seasonal variations of snow cover over the Tibetan Plateau. *Geophysical Research Letters*, 34(6), L06706.
12. Notarnicola, C., Duguay, M., Moelg, N., Schellenberger, T., Tetzlaff, A. et al. (2013). Snow cover maps from MODIS images at 250 m resolution, part 1: algorithm description. *Remote Sensing*, 5(1), 110–126. DOI 10.3390/rs5010110.
13. Metsämäki, S., Pulliainen, J., Salminen, M., Luojus, K., Wiesmann, A. et al. (2015). Introduction to GlobSnow Snow Extent products with considerations for accuracy assessment. *Remote Sensing of Environment*, 156, 96–108. DOI 10.1016/j.rse.2014.09.018.
14. Mazari, N., Tekeli, A. E., Xie, H., Sharif, H. I., Hassan, A. A. E. (2013). Assessment of ice mapping system and moderate resolution imaging spectroradiometer snow cover maps over Colorado Plateau. *Journal of Applied Remote Sensing*, 7(1), 073540. DOI 10.1117/1.JRS.7.073540.
15. Zhu, Z., Wang, S., Woodcock, C. E. (2015). Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sensing of Environment*, 159, 269–277. DOI 10.1016/j.rse.2014.12.014.
16. Li, S., Yan, H., Liu, C. (2007). Study of snow detection using FY-2C satellite data. *Journal of Remote Sensing*, 11, 406–413.
17. Liu, X., Jin, X., Ke, C. Q. (2014). Accuracy evaluation of the IMS snow and ice products in stable snow covers regions in China. *Journal of Glaciology and Geocryology*, 36, 500–507.
18. Hall, D. K., Riggs, G. A. (2007). Accuracy assessment of the MODIS snow products. *Hydrological Processes: An International Journal*, 21(12), 1534–1547. DOI 10.1002/hyp.6715.
19. Yang, J., Jiang, L., Shi, J., Wu, S., Sun, R. et al. (2014). Monitoring snow cover using Chinese meteorological satellite data over China. *Remote Sensing of Environment*, 143, 192–203. DOI 10.1016/j.rse.2013.12.022.
20. Harrison, A. R., Lucas, R. M. (1989). Multi-spectral classification of snow using NOAA AVHRR imagery. *International Journal of Remote Sensing*, 10(4–5), 907–916. DOI 10.1080/01431168908903930.
21. 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 Sensing of Environment*, 54(2), 127–140. DOI 10.1016/0034-4257(95)00137-P.
22. Gascoïn, S., Grizonnet, M., Bouchet, M., Salgues, G., Hagolle, O. (2019). Theia Snow collection: high-resolution operational snow cover maps from Sentinel-2 and Landsat-8 data. *Earth System Science Data*, 11(2), 493–514. DOI 10.5194/essd-11-493-2019.
23. Pepe, M., Brivio, P. A., Rampini, A., Nodari, F. R., Boschetti, M. (2005). Snow cover monitoring in Alpine regions using ENVISAT optical data. *International Journal of Remote Sensing*, 26(21), 4661–4667. DOI 10.1080/01431160500206635.
24. De Ruyter de Wildt, M., Seiz, G., Gruen, A. (2007). Operational snow mapping using multitemporal Meteosat SEVIRI imagery. *Remote Sensing of Environment*, 109(1), 29–41. DOI 10.1016/j.rse.2006.12.008.

25. Siljamo, N., Otto, H. (2011). New geostationary satellite-based snow-cover algorithm. *Journal of Applied Meteorology and Climatology*, 50(6), 1275–1290. DOI 10.1175/2010JAMC2568.1.
26. Qian, Y., Zhou, W., Yan, J., Li, W., Han, L. (2015). Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. *Remote Sensing*, 7(1), 153–168. DOI 10.3390/rs70100153.
27. Zhan, Y., Wang, J., Shi, J., Cheng, G., Yao, L. et al. (2017). Distinguishing cloud and snow in satellite images via deep convolutional network. *IEEE Geoscience and Remote Sensing Letters*, 14(10), 1785–1789. DOI 10.1109/LGRS.2017.2735801.
28. Huang, L., Li, Z., Tian, B. S., Chen, Q., Liu, J. L. et al. (2011). Classification and snow line detection for glacial areas using the polarimetric SAR image. *Remote Sensing of Environment*, 115(7), 1721–1732. DOI 10.1016/j.rse.2011.03.004.
29. Zhang, Y. L., Li, B. Y., Zheng, D. (2002). A discussion on the boundary and area of the Tibetan Plateau in China. *Geographical Research*, 21, 1–8.
30. Liu, R. G., Liu, Y., Xu, X. L., Ge, Q. S. (2017). Change of geographical environment in southern margin of Tibetan Plateau since 1980s. *Bulletin of Chinese Academy of Sciences*, 32(09), 1003–1013.
31. You, Q., Fraedrich, K., Ren, G., Pepin, N., Kang, S. (2013). Variability of temperature in the Tibetan Plateau based on homogenized surface stations and reanalysis data. *International Journal of Climatology*, 33(6), 1337–1347. DOI 10.1002/joc.3512.
32. Gao, X. X., Chen, Y., Zhang, W., Lu, P. X. (2017). Climate characteristics of clouds and their influence on the land-atmosphere system in the Qinghai-Tibetan Plateau. *Journal of Lanzhou University (Natural Sciences)*, 53(04), 459–466+480.
33. Riggs, G. A., Hall, D. K., Salomonson, V. V. (1994). A snow index for the Landsat thematic mapper and moderate resolution imaging spectroradiometer. *Proceedings of IGARSS'94—1994 IEEE International Geoscience and Remote Sensing Symposium*, 4, 1942–1944.
34. 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 Sensing of Environment*, 54(2), 127–140. DOI 10.1016/0034-4257(95)00137-P.
35. Burton-Johnson, A., Black, M., Fretwell, P., Joseph, K. G. (2016). An automated methodology for differentiating rock from snow, clouds and sea in Antarctica from Landsat 8 imagery: A new rock outcrop map and area estimation for the entire Antarctic continent. *Cryosphere*, 10(4), 1665–1677. DOI 10.5194/tc-10-1665-2016.
36. Verbeek, J. J., Vlassis, N., Kröse, B. (2003). Efficient greedy learning of Gaussian mixture models. *Neural computation*, 15(2), 469–485.
37. Klein, A. G., Hall, D. K., Riggs, G. A. (1998). Improving snow cover mapping in forests through the use of a canopy reflectance model. *Hydrological Processes*, 12(10–11), 1723–1744. DOI 10.1002/(SICI)1099-1085(199808/09)12:10/11<1723::AID-HYP691>3.0.CO;2-2.
38. Paul, F., Kääb, A., Maisch, M., Kellenberger, T. (2002). The new remote-sensing-derived Swiss glacier inventory: I. Methods. *Annals of Glaciology*, 34, 355–361. DOI 10.3189/172756402781817941.
39. Xu, H. Q. (2005). A study information extraction of water body with the modified normalized difference water index (MNDWI). *Journal of Remote Sensing*, 9(5), 589–595.
40. McFeeters, S. K. (1996). The use of the Normalized Difference Water Index in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425–1432. DOI 10.1080/01431169608948714.