

# Cloud removing method for daily snow mapping over Central Asia and Xinjiang, China

Xiaoqi Yu<sup>1,2</sup>, Yubao Qiu<sup>1</sup>, Huadong Guo<sup>1</sup>, Lijuan Chen<sup>1,3</sup>

1 Institute of Remote Sensing and Digital Earth, Laboratory of Digital Earth Science, Dengzhuang South Road, Haidian District, Beijing, China, 100094, qiuyb@radi.ac.cn

2 College of Geographical Information System, Liaoning Technical University, Fuxin, China, 123000

3 College of Geographical Science, Fujian Normal University, Fuzhou, China  
Type the author addresses here

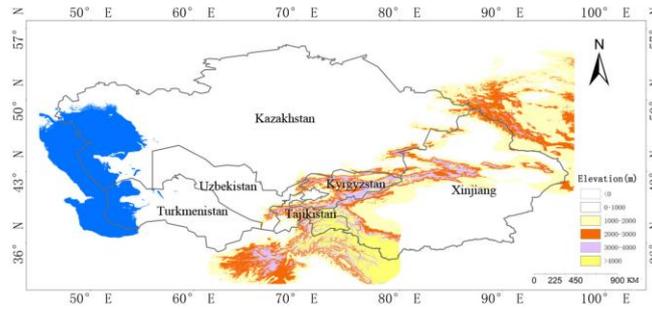
**Abstract.** Central Asia and Xinjiang, China are conjunct areas, located in the hinterland of the Eurasian continent, where the snowfall is an important water resource supplement form. The induced seasonal snow cover is vital factors to the regional energy and water balance, remote sensing plays a key role in the snow mapping field, while the daily remote sensing products are normally contaminated by the occurrence of cloud, that obviously obstacles the utility of snow cover parameters. In this paper, based on the daily snow product from Moderate Resolution Imaging Spectroradiometer (MODIS A1), a cloud removing method was developed by considering the regional snow distribution characteristics with latitude and altitude dependence respectively. In the end, the daily cloud free products was compared with the same period of eight days MODIS standard product, revealing that the cloud free snow products are reasonable, while could provide higher temporal resolution, and more details over Center Asia and Xinjiang Province.

## 1. Introduction

Central Asia and Xinjiang is one of the main snow area in the middle and low latitude. Snow and snow water resources plays an important role in the feedback regulation and supply of climate and water cycle in Central Asia and Xinjiang Province.

At present, optical remote sensing monitoring of snow cover interfered by the weather, the cloud has the greatest impact on snow monitoring in the study area<sup>[1]</sup>. A lot of research based on MODIS Snow Cover products to the cloud removing has done, integrated Central Asia with unique geographical features, it's necessary to combined with a variety of algorithms for cloud removal. Firstly, the temporal and spatial distribution characteristics of snow cover are analyzed, then using the mountain-plain partitioned algorithm, formed a set of cloud algorithm processing of daily snow cover products suitable for Central Asia and China's Xinjiang region. The elevation and study area conditions are showed by the following figure.





**Figure 1.** Distribution of Central Asia and Xinjiang region.

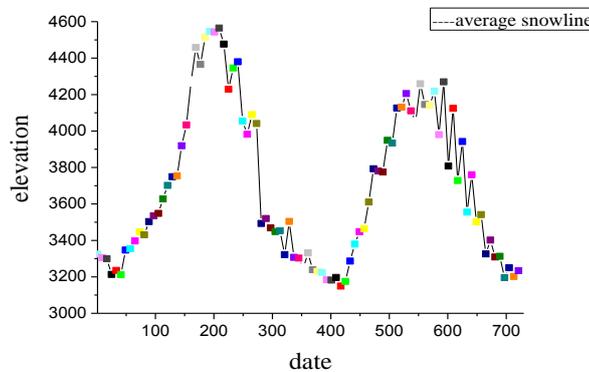
**2. Data and characteristic analysis**

MODIS snow product used in the study is based on the NDSI threshold method to get the daily global snow cover products. The V005 version of Terra MODIS and Aqua MODIS daily snow classification products MOD10A1 and MYD10A1 were download by NSIDC website. Digital elevation model DEM (Elevation Model) is selected by the USGSEROS data center where responsible for data archiving and reporting SRTM\_DEM elevation data to the public.

Currently there are many cloud algorithms, Andreas et al<sup>[2]</sup> carrying out the study on the applicability of the different algorithms, including: MODIS daily combination, adjacent temporal deduction, near four pixels method, near eight pixels method, elevation filtering, long time series method in order to deal with the cloud. In this paper, based on the near eight pixels method to analyze characteristics. And try to improve the snowline method.

*2.1. Variation of the snowline characteristics with altitude dependency*

In Central Asia, there was less snow distributing in the valleys and basins, the article will divide the elevation into three segments, respectively, less than 0,0-2000meters, and 2000-8429meter.



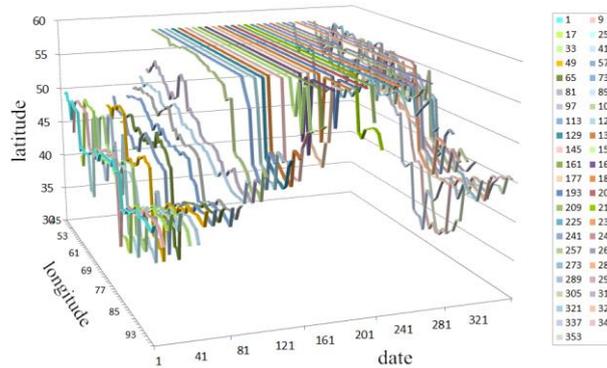
**Figure 2.** The average figure over mountain area.

Form eight days maximum snow products in 2010,2011 years,seeking average snowline of 2000-8429 meters, getting the change rule of mountain snowline for two years. (The horizontal coordinates of 100 shows 100th days in 2010, 400 shows 35th days in 2011).In 200-250, 580-650 days, elevation level fluctuation, which indicates that there is a phenomenon of sporadic snowfall and melting in the period of melting snow, but the snow accumulation season is a continuous snowfall, there is no melting phenomenon.

*2.2. Variation of the snowline characteristics with latitude dependency*

According to eight days maximum snow cover products in 2010, in Central Asia and China's Xinjiang region, when statistical longitude is 45 to 96 degrees and there are most land pixel value below snow

pixels, statistics the latitude of variation range of snow pixel values, the results as shown in figure 3, the date of 2010 first to 365th days, straight line represents no snow.

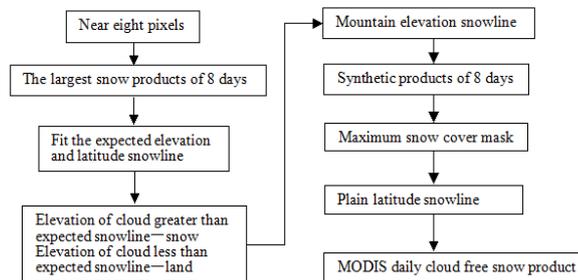


**Figure 3.** The snowline latitude curve.

Every 8 days to get a snowline of latitude, the latitude snowline range is approximately 30 to 58 degree, the range of 1-105 days and 297-365 days (i.e. integrable snowy season), longitude roughly in the range of 0 to 74, most of the terrain is plain areas, the latitude is linear, that is, there is no snow in the lower elevation areas in the summer, longitude in the range of 75 to 96, latitude curve fluctuation, contains mountains and high altitude areas where snow does not melt completely, so snowline of latitude lower than the plain areas.

### 3. Daily cloud-free algorithm

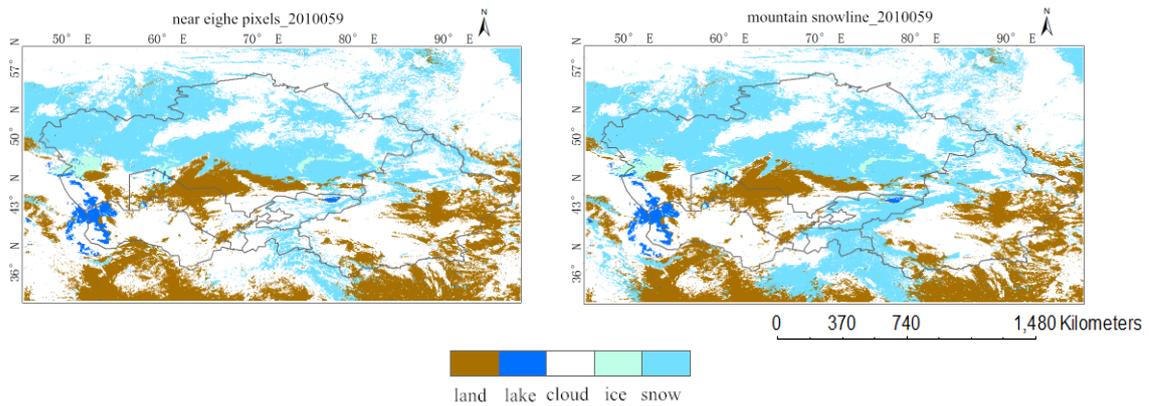
According to the snowline changes of latitude and elevation mentioned above, we draw a feasible scheme as figure 4 shows.



**Figure 4.** Cloud removal algorithm flow chart.

#### 3.1. Snowline algorithm based on elevation over mountain area

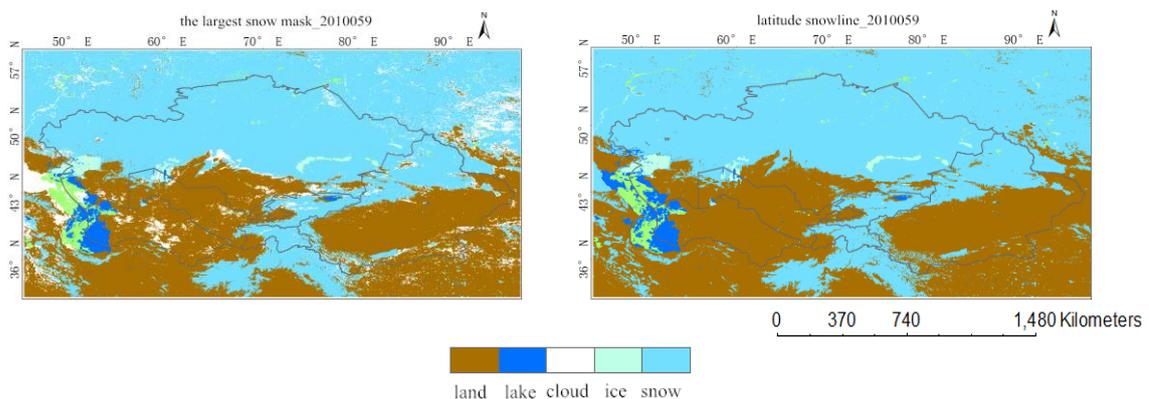
Extraction of 2000 meters above the area, to deal with the cloud by the average snowline. Snowl method based on: 1) the same scene images,the cloud height is greater than or equal to all the snow pixel average elevation, cloud pixels are classified as snow; 2) the cloud pixel height less than or equal to all land pixel average elevation, cloud pixels are classified into land; 3) the cloud height between all the snow pixel average elevation and land elevation, cloud pixels are classified into a piece of snow. We using elevation snowline method to remove some cloud based on the near eight pixels data,and figure 5 is the comparison chart of near eight pixels and snowline processing.



**Figure 5.** Comparison between near eight pixels and elevation snowline method.

### 3.2. Snowline algorithm based on latitude over plain area

Based on the products which are obtained by the mountain elevation snowline algorithm, 8 days products synthesized of maximum snow and land, get 8 days maximum difference of snow and land. According to the known latitude snowline on plain area, dealing with the maximum snow cover film products by latitude snowline method. Judge: above the latitude snowline and altitude is greater than 0, the cloud pixel assigned to snow. Under the latitude snowline, altitude area greater than 0, the cloud pixel assigned to land. Then the cloud is completely removed, and getting MODIS daily cloud-free snow products. Figure 6 is the comparison chart of maximum snow mask and latitude snowline method.



**Figure 6.** Comparison between the largest snow mask and latitude snowline.

## 4. Comparative analysis and verification

In order to assess and validate the snow classification accuracy, the snow cover area precision indexes are used to evaluate the snow classification for each step, indicators include the total classification accuracy (Accuracy), Recall, snow classification accuracy (Precision) and F value. The total accuracy represent the probability that a pixel is classified correctly, but when the image has large tracts of no snow regions, this index may be misleading, if the image is only 10% snow pixels, then the total accuracy will reach 90% because most pixels are divided into land. Recall is the probability of detection of snow pixels, i.e. a ratio between correct classification of snow pixel number and the actual number of all snow sample pixels. Precision is the ratio between correct classification of snow pixel number and classification of all the snow pixel. Recall and Precision are correlative interplay, high recall rates show that the correct classification of snow pixel number has high accuracy compared with the ground stations, and the high precision indicate that correct classification of snow pixels has high accuracy relative to the total snow pixels number of MODIS. Recall is high, does not mean that the Precision is higher, so using F value, perhaps the most useful indicators, balances Recall and Precision.

There are 363 temporal images from October 1,2012 to February 28,2014 in Central Asian region, and daily snow depth data provided by 103 ground stations, through the confusion matrix (Table 1)and formula (1), (2), (3), (4) to evaluate the snow classification image, analysis the precision of MODIS (MOD10A1, MYD10A1) data under the clear weather and the synthetic products in the process of cloud removing, considering the following samples: 1) the sample number a shows snow both in MODIS and ground station records (snow depth >0); 2) b shows that snow in ground station records but not in MODIS; 3) c shows that snow in MODIS but not in ground station records ; 4) d shows no snow in either the MODIS or ground station records.

**Table 1.** Confusion matrix of accuracy verification.

|          |       |      |
|----------|-------|------|
|          | MODIS |      |
| stations | snow  | land |
| snow     | a     | b    |
| land     | c     | d    |

The index formula of snow accuracy is as follows:

$$Accuracy = \frac{a + d}{a + b + c + d} \times 100\% \quad (1)$$

$$Recall = \frac{a}{a + b} \times 100\% \quad (2)$$

$$Precision = \frac{a}{a + c} \times 100\% \quad (3)$$

$$f = 2 \frac{Precision \times Recall}{Precision + Recall} \times 100\% = \frac{2a}{2a + b + c} \times 100\% \quad (4)$$

The number of samples of raw data MOD10A1 ,MYD10A1 and each process product in the confusion matrix is shown in table 2.

**Table 2.** Number of samples in raw data and each process.

| No. | product                     | a    | b   | c    | d     | a3   | b3  | Accuracy(%) |                | Recall(%)   |                | Precision(%) |                | F value(%)  |                |
|-----|-----------------------------|------|-----|------|-------|------|-----|-------------|----------------|-------------|----------------|--------------|----------------|-------------|----------------|
|     |                             |      |     |      |       |      |     | All station | Snow depth>3cm | All station | Snow depth>3cm | All station  | Snow depth>3cm | All station | Snow depth>3cm |
| 1   | MOD10A1                     | 1525 | 286 | 1404 | 14307 | 1364 | 125 | 90.35       | 91.11          | 84.21       | 91.61          | 79.72        | 77.85          | 81.90       | 84.17          |
| 2   | MYD10A1                     | 1237 | 193 | 1276 | 13761 | 1171 | 67  | 91.08       | 91.75          | 86.5        | 94.59          | 79.09        | 78.17          | 82.63       | 85.60          |
| 3   | MOYD10A1                    | 2065 | 340 | 1921 | 16549 | 1874 | 161 | 89.17       | 89.85          | 85.86       | 92.09          | 81.17        | 79.64          | 83.45       | 85.41          |
| 4   | adjacent temporal deduction | 2495 | 399 | 2253 | 19092 | 2253 | 186 | 89.06       | 89.75          | 86.21       | 92.37          | 82.13        | 80.58          | 84.12       | 86.07          |
| 5   | near four pixels method     | 2560 | 404 | 2285 | 19229 | 2305 | 186 | 89.01       | 89.71          | 86.37       | 92.53          | 82.26        | 80.68          | 84.27       | 86.20          |
| 6   | near eight pixels method    | 2975 | 404 | 2587 | 19229 | 2656 | 186 | 88.13       | 88.75          | 88.04       | 93.46          | 83.57        | 81.95          | 85.75       | 87.33          |
| 7   | elevation snowline          | 3219 | 416 | 3053 | 19468 | 2768 | 186 | 86.74       | 87.29          | 88.56       | 93.7           | 83.35        | 81.15          | 85.87       | 86.98          |
| 8   | maximum snow and land mask  | 5414 | 819 | 4784 | 24276 | 4708 | 385 | 84.12       | 84.87          | 86.86       | 92.44          | 81.04        | 78.79          | 83.85       | 85.07          |
| 9   | latitude snowline           | 6091 | 895 | 5560 | 24615 | 5201 | 409 | 82.63       | 83.32          | 87.19       | 92.71          | 81.18        | 78.65          | 84.07       | 85.10          |

The cloud removal algorithm finally removed all cloud pixels by gradually controlling the cloud cover amount and the loss of precision, considering the factors of terrain features in the area and the correlation of snow in time and space, and satisfactory results were obtained. The ground stations observation data is used to validate the accuracy of MODIS daily cloudless snow products. Among them, Accuracy was 82.63%, Precision was 81.18%, Recall rate was 87.19%, F value was 84.07%. When the snow depth is small, MODIS Snow classification accuracy is not high, if only retain snow

samples that snow depth greater than 3cm , the classification accuracy is greatly improved and can be used for the dynamic monitoring of snow cover area in Central Asia.

There are large area and complex terrain over Central Asia and Xinjiang,China, and the plateau snow distribution has a certain randomness and uncertainty. The snowline algorithm based on elevation in the mountain and latitude in the plain used to cloud removing, its uncertainty may be affected by the terrain, temperature, precipitation and other factors. Gafurov<sup>[3]</sup> considered the slope factors when using snowline method, there are more sunlight in sunny slope, because of the differences of local water sources and climate in different regions, the high altitude pixel located in sunny slope is land, adjacent low altitude pixel located in shady slope is snow. These conditions without considering may cause the cloud removal algorithm in the study area have error. 8 days maximum snow or land area mask discrimination method using snow cover information correlation in short time, could remove most of the cloud and the precision is higher, but still can not accurately return to the original situation of the day's snow cover, and a little miscarriage of justice arose. Therefore, developing more perfect cloud removal algorithm and the method to verify the precision is of great significance for the real-time dynamic monitoring of the snow and analyzing the effects of snow cover on climate.

### **5. Comparative analysis and verification**

This work is supported in part by National Natural Science Foundation of China (Grant No. 41371351 and 41120114001) and Chinese Special Fund for Meteorology-Scientific Research in the Public (GYHY201206040).

### **References**

- [1] Xiaodong H and Xiaohua H 2012 *Cloud removal methodology from MODIS daily snow product* vol 34(Beijing : Journal of glaciology and geocryology ) No.5
- [2] Andreas J D and Claudia K 2012 *Remote sensing of snow –a review of available methods* ed U Gessner and S Dech ( International Journal of Remote Sensing ) p 4094-4134
- [3] Gafurov A and Bardoyy A 2009 *Cloud removal methodology from MODIS snow cover product* (Hydrology and Earth System Sciences) p 1361-1373
- [4] Parajka J and Pepe M 2010 *A regional snow-line method for estimating snow cover from MODIS during cloud cover* ( Journal of Hydrology ) p 203-212