Hydrology and Earth System Sciences

Cloud removal methodology from MODIS snow cover product

A. Gafurov and A. Bárdossy

Department of Hydrology and Geohydrology, Institute of Hydraulic Engineering, University of Stuttgart, Pfaffenwaldring 61, 70569 Stuttgart, Germany

Abstract. The Moderate Resolution Imaging Spectroradiometer (MODIS) employed by Terra and Aqua satellites provides spatially snow covered data with 500 m and daily temporal resolution. It delivers public domain data in raster format. The main disadvantage of the MODIS sensor is that it is unable to record observations under cloud covered regions. This is why this study focuses on estimating the pixel cover for cloud covered areas where no information is available. Our step to this product involves employing methodology based on six successive steps that estimate the pixel cover using different temporal and spatial information. The study was carried out for the Kokcha River basin located in northeastern part of Afghanistan. Snow coverage in catchments, like Kokcha, is very important where the melt-water from snow dominates the river discharge in vegetation period for irrigation purposes. Since no snow related observations were available from the region, the performance of the proposed methodology was tested using the cloud generated MODIS snow cover data as possible "ground truth" information. The results show successful performances arising from the methods applied, which resulted in all cloud coverage being removed. A validation was carried out for all subsequent steps, to be outlined below, where each step removes progressively more cloud coverage. Steps 2 to 5 (step 1 was not validated) performed very well with an average accuracy of between 90–96%, when applied one after another for the selected valid days in this study. The sixth step was the least accurate at 78%, but it led to the removal of all remaining cloud cover.

1 Introduction

One of the main sources for water availability in high altitude regions is snow. For this reason, snow data is especially important in high mountainous areas where the snow pack stays for longer periods and snow melt provides runoff and water supply for the downstream population. Many data-scarce regions worldwide, such as central Asia, greatly depend on the available snow and glacier in the mountains during spring and summer seasons for the irrigation of agricultural fields and for drinking purposes. Several researchers have previously conducted studies of the importance of mountain snow and glacier melt for the lowland regions. Aizen et al. (1995) reported 18–28% annual runoff contribution of glaciers in average for Ala-Archa basin located in central Asia. They also reported that this contribution can increase up to 40–70% in summer. The study by Singh et al. (2006) in the Himalayan basin reported an 87% glacier melt contribution to the total runoff. Kling et al. (2006) carried out a study for the whole of Austria and found the runoff coefficient for snowmelt in the alpine parts to be greater than 80%, and for the lowlands less than 50%.

Mapping the snow cover with in situ measurements is impossible for large catchments, mainly because of the extremely high cost and manpower required. Since the mid-1960s a variety of coarse and medium spatial resolution remote sensing data has been used to map several parameters of the earth. Remote sensing offers a powerful alternative for obtaining environmental data worldwide. Nowadays, many different instruments on satellites provide continuous information about the earth's actual state. The Moderate Resolution Imaging Spectroradiometer (MODIS) is one such instrument installed on the Terra and Aqua satellites. These satellites began to conduct observations in February 2000 and July 2002, respectively. The satellites pass over the equator in the morning (Terra at 10:30 a.m. in a descending mode) and afternoon (Aqua at 1:30 p.m. in an ascending mode),

Correspondence to: A. Gafurov
(abror.gafurov@iws.uni-stuttgart.de)

capturing the earth's current state and provide two snow cover images for the same location per day. The MODIS uses 36 spectral bands to estimate 44 globally available data products with a spatial resolution of 250 m, 500 m and 1 km. Among other datasets, MODIS also provides a grid of snow cover information at 500 m spatial and daily temporal resolutions. The accuracy of MODIS snow cover data has been tested by different researchers with positive results. Among them, Parajka et al. (2006) compared MODIS snow cover data with in situ information over the whole Austria and reported an agreement of about 95% between MODIS snow cover and the in situ data taken from 754 climate stations on cloud free pixels. Klein et al., (2003) reported an 88% agreement when comparing to measurements of the Upper Rio Grande Basin. Maurer et al. (2003) compared the MODIS snow covered area (SCA) product with the SCA product of National Operational Hydrologic Remote Sensing Center (NOHRSC) under cloud free conditions and concluded less cloud free misclassification (4% and 5% fewer overall for the Columbia and Missouri basins, respectively) for MODIS than with NOHRSC. The MODIS snow product accuracy study by Wang et al. (2008) reported that under clear sky conditions, the MODIS snow cover product had 94% accuracy for snow and 99% for land when compared against the in situ snow depth data taken from northern Xingjian, China. Another recent study by Wang et al. (2008) compared the standard daily, eight-day MODIS Terra and Aqua snow cover classifications and reported close to 100% agreement for land classification. For the snow-covered areas, they had high accuracy only in the winter period and decreasing accuracy for the rest of the seasons. They also had a combined daily and eight-day MODIS snow cover products and achieved better accuracy than with individual MODIS products. Elsewhere, the study by Tekeli et al. (2005) stated that "among 96 observations, 74 observations mapped as snow or land by MODIS agreed with the ground measurements with a matched ratio of 77%".

The main limitation for the direct usage of MODIS snow cover data in environmental studies is the extent of cloud covered pixels. Another limitation is that the MODIS snow cover data contains binary information where the pixel was defined as snow or no snow. No information was provided about the snow depth. This paper addresses the limitations arising from the cloud cover problem. There have been studies in the past to estimate the snow cover for cloudy pixels. For example, Parajka et al. (2008) introduced three methods to reduce the cloud covered pixels from MODIS snow products. The first method was a combination of Terra and Aqua images. The second was a cloud pixel reduction according to the information from the majority of eight neighboring pixels. The third method described the cloud pixel classification according to the most recent observation of the same pixel. They came up with a 9–21% and a 6–13% of cloud removal for the first two methods, respectively, and varying accuracies for the third method according to the temporal step.

Liang et al. (2008) developed a new daily snow cover product through combining the MODIS and AMSR-E daily snow water equivalent (SWE) data. After using in situ measurements at 20 climatic stations for validation, they came up with an overall accuracy of 76% and snow cover classification accuracy of 75% for the combined snow cover data. Additionally, their results also indicated that AMSR-E daily SWE imagery generally agreed with the MODIS daily snow cover product, with an overall agreement of 93% and a snow agreement of 97%.

In this study, six steps or methods were applied to reduce the cloud covered pixels from the MODIS snow cover data. One of the steps used a combination of the Terra and Aqua products introduced and described by Parajka et al. (2008). The goal of this study was to remove the cloud covered pixels from the snow cover data completely and to produce continuous maps of snow coverage over the study catchments.

2 Study area

The Kokcha Basin in the north-eastern part of Afghanistan was chosen as a study area for this research. The geographic location of this basin ranges in latitude from 35.40° N to 37.40° N and in longitude from 69.30° E to 71.60° E. The area of the basin is about 20 600 km^2 (Fig. 1).

The topography of this basin is very heterogeneous. The elevation ranges from 416 m above sea level (m a.s.l.) in the northwestern part to up to 6383 m a.s.l. in the southern mountains. Topographically, about 75% of the basin area corresponds to an elevation zone above 2000 m a.s.l., 50% above 3000 m a.s.l. and 30% above 4000 m a.s.l. The land cover information was also obtained from the MODIS product (MOD12Q1) and according to this map, the dominant land cover types are grasslands (43%), barren or sparsely vegetated (37%), and open shrub lands (15%). The climate is characterized by semi-arid to arid with hot summers and cold winters (Kimura et al., 2002). It is important to mention that there is almost no precipitation in the summer periods. Even so, river discharges reach their peak in the summer. Figure 2 illustrates an example of an annual precipitation and discharge cycle from this region and this clearly demonstrates that the basin discharge is highly snow driven.

The Kokcha River is the tributary to Pandj River and Pandj River is one of the main water sources for Amudarya, the largest river in Central Asia. The historical discharge analyses indicates that about 65% of the total annual volume of water from the Kokcha Basin is generated during the summer season. The snowmelt from the high mountains is the main water source for agriculture downstream in spring and summer seasons. Thus, the snow cover information for the catchment is very important when studying the water balances of the region.
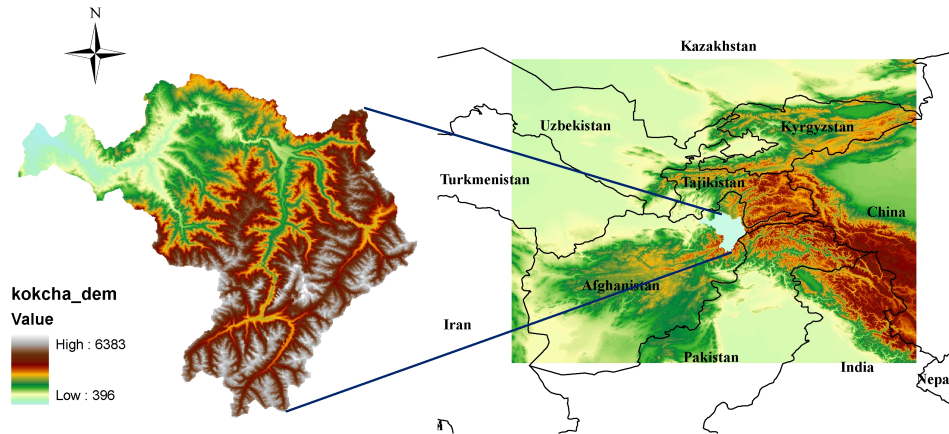
Fig. 1. Kokcha catchment.

3 Data

The MODIS instruments installed on the Terra and Aqua satellites collect data as part of NASA's Earth Observing System (EOS) program. The first MODIS instrument on the Terra satellite was launched in December 1999 and began to deliver data in February 2000. The second MODIS instrument on the Aqua satellite was launched in May 2002 and started to deliver data in July 2002. The Terra satellite crosses the equator from south to north at about 10:30 a.m. and the Aqua satellite from north to south at about 1:30 p.m. The Normalized Difference Snow Index (NDSI) method was used as the MODIS snow mapping algorithm with a set of thresholds (cf. Hall et al., 2002). The MODIS snow cover product is partly distributed as tiles of about 10° by 10° worldwide. There are a total of 36 horizontal (H) and 18 vertical (V) tiles covering the entire globe. The H23V05 tile that covers complete study area was used for this study.

The MODIS snow cover product can be ordered free of charge through the Distributed Active Archive Center (DAAC) located at the National Snow and Ice Data Center (NSIDC). The MOD10A1 (MODIS / TERRA SNOW COVER DAILY L3 GLOBAL 500 m SIN GRID V005) snow product with 500 m spatial and daily temporal resolution was obtained for the Kokcha basin for the whole year of 2003. This data was provided in an Hierarchical Data Format (HDF). The MODIS Reprojection Tool (MRT) was used in this study to convert from HDF format into an ArcGIS compatible GEOTIF format for visualization. The estimation of cloudy pixels was done in the ASCII format.

Additionally, the Hole-filled seamless Shuttle Radar Topography Mission (SRTM) digital elevation data with 90 m spatial resolution was obtained from the International Centre for Tropical Agriculture (CIAT) and this was also used as ASCII format in this methodology. The resolution of DEM data was aggregated to 500 m to match the spatial resolution of snow cover data.
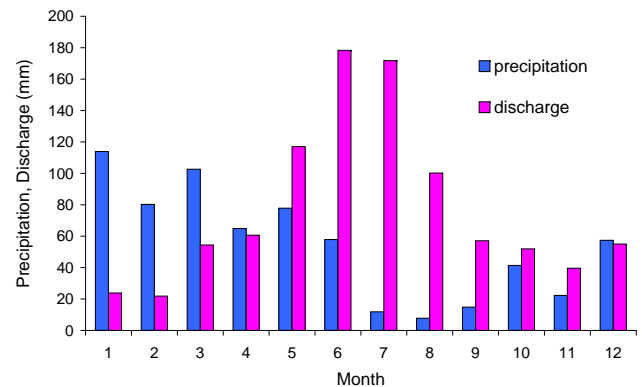

Fig. 2. Annual cycle of precipitation and discharge.

4 Methodology

As mentioned above, the methodology was based on six steps that were developed to estimate the actual pixel cover for cloud covered areas. The steps are based on spatial and temporal information of cloud covered pixels. Every step was essentially a processing step, and removed some fraction of cloud covered pixels. The resulting cloud reduced snow product generated from each step was used as input for the next step where more clouds are removed. Subsequent steps reduce the cloud progressively until the raster dataset was completely cloud free. The raster matrix analyzed for Kokcha basin consists of 459 columns and 395 rows. One compete year of 2003 was taken as the study period. The original 365 MODIS TERRA and 365 MODIS AQUA snow cover data were used and cloud covered pixels were removed. The analyses were done using the FORTRAN programming language. Below, each processing step is explained in detail.

The first step is based on the combination of Terra and Aqua satellite images. Since the cloud coverage is in continuous movement, it is possible to have different cloud

coverage during the observation from two satellites that fly through study area with a few hours time difference. The assumption for this step was that no snowmelt or snowfall occured within this time shift. If the pixel was observed as cloud covered by one satellite and cloud free by another satellite, the resulting cloud free pixel was assigned as true pixel cover. Using this step the maximum cloud free coverage was obtained from two satellites. The formula is given as follows in Eq. (1):

$$S_{(y,x,t)} = \max(S^A_{(y,x,t)}, S^T_{(y,x,t)}) \qquad (1)$$

where y is the index for row (vertical); x is the index for column (horizontal); t is the index for day (temporal) of pixel S. S^A and S^T stand for the Aqua and Terra pixels, respectively. Equation (1) was applied for both snow covered and land covered pixels. Maximum snow coverage and maximum land coverage was obtained from two satellite products.

The second step is based on temporal combination of cloud covered pixels. One or two days of information both forwards and backwards was considered to estimate actual pixel cover. It can occur that one day was densely cloud covered while the previous or next days were less cloud covered. This is why the previous and next days were checked as to whether they contain more accurate information. If the beginning and ending days of the observed period were covered by snow cover and the middle day was covered by cloud cover, the cloud covered day was also set as being snow covered. The same applies for land cover. The formula is given in Eq. (2):

$$S_{(y,x,t)} = 1 \quad \text{if} \quad (S_{(y,x,t-1)} = 1 \quad \text{and} \quad S_{(y,x,t+1)} = 1) \qquad (2)$$

Since the MODIS snow cover product is a binary quantity, 1 corresponds to snow cover and 0 for land cover. The second step first checks the previous day and the next day with one shift. If the analyzed cloud covered pixel has been estimated already using one temporal shift (previous and next day), then the procedure for this pixel was finished since the actual pixel cover for cloud cover was estimated. If the cloud covered pixel was not estimated using Eq. (2), then the time step was shifted back one day and forwards one day, as stated in Eqs. (3) and (4). The second step does not take into consideration a two-day backward shift and two-day forward shift. The maximum time span was two days backward and one day forward or, alternatively, one day backwards and two days forward.

$$S_{(y,x,t)} = 1 \quad \text{if} \quad (S_{(y,x,t-2)} = 1 \quad \text{and} \quad S_{(y,x,t+1)} = 1) \qquad (3)$$

$$S_{(y,x,t)} = 1 \quad \text{if} \quad (S_{(y,x,t-1)} = 1 \quad \text{and} \quad S_{(y,x,t+2)} = 1) \qquad (4)$$

Here, the assumption was made that the snow cover stayed constant if the weather was cloud covered. Possible snowmelt could contradict this assumption. However, solar radiation is the main source for snowmelt and since the pixel was covered by cloud on that day, which reflects a large proportion of this radiation, the possibility for snowmelt can be neglected. The probability that the cloud covered pixel was actually snow covered was higher (because of possible snowfall when clouds exist) than the probability that it melted away. Additional temperature information for those pixels on that day would give more information on whether the snowmelt can occur or not, but this was not considered in this study.

The third step was based on the snow transition elevation. This step was well suited for pixels at very low and very high altitudes. The concept was to find the minimum elevation where snow existed and the snow covered elevation where all pixels above this level were covered by snow. The latter is commonly called the snow line, where snow cover was continuous above this elevation. These elevation lines are then set as threshold elevations (minimum and maximum snow lines). Based on this, all cloud covered pixels below the minimum snow elevation line were assigned as land covered pixels and all pixels above maximum snow line were assigned as snow covered pixels. The formula is given in Eqs. (5) and (6):

$$S_{(y,x,t)} = 0 \quad \text{if} \quad (H_{(y,x)} < H^S_{\min}(t)) \qquad (5)$$

$$S_{(y,x,t)} = 1 \quad \text{if} \quad (H_{(y,x)} > H^S_{\max}(t)) \qquad (6)$$

where $H_{(y,x)}$ is the elevation of a pixel at (y, x) location; $H^S_{\min}(t)$ is the minimum snow covered elevation (the elevation of lowest snow covered pixel) assuming that there are no snow covered areas below this elevation on day t and $H^S_{\max}(t)$ is the maximum snow covered elevation (maximum snow line where all pixels in this line are covered by snow) on day t. This step mainly leads to the removal of cloudy pixels at very low and very high elevations. It can happen that the tops of the mountains were covered by clouds on that day while the snow line was at lower elevation. In this case, any cloud covered pixel that is located higher than the snow line was assigned as a snow covered pixel. It is the same for land covered pixels. Any cloud covered pixel that is located lower than the minimum snow covered pixel was assigned as land covered. The condition for estimating cloud covered pixels in this step was the amount of cloud free pixels resulting from the first and second steps. In order to make sure that the minimum and maximum snow covered elevations were correctly assigned, the basin should be at least 70% cloud free. If not, this step could not be applied. If there are few cloud free pixels in the basin, the maximum and the minimum snow elevations may be incorrectly assessed which may lead to enormous errors.

The fourth step was based on the spatial combination of neighboring pixels. Four direct "side-bordering" neighboring pixels of the cloudy pixel were examined. If at least three pixels are defined as snow, then the cloudy pixel was also assigned to be a snow covered pixel as well. The same applies
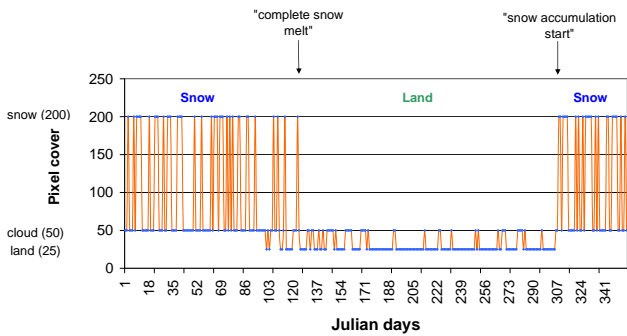
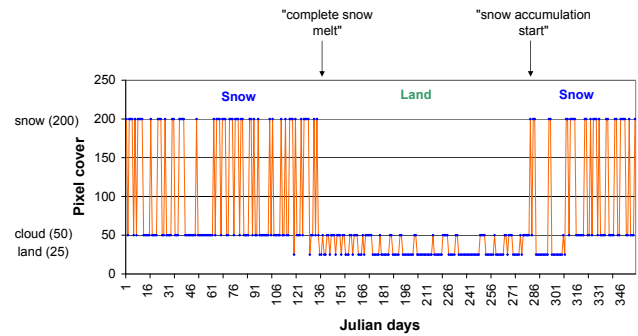**Fig. 3a.** Pixel cover time series at elevation 2901 m a.s.l.

Fig. 3b. Pixel cover time series at elevation 3294 m a.s.l.

for land covered pixels. It is possible that this assumption is not always true, but the probability of a pixel having the same cover as at least three of its direct neighboring cells was higher than the middle pixel having the opposite cover.

The fifth step was based on a combination of (eight) neighboring pixels, and particularly, the elevation between them. Here, all neighbouring pixels defined as cloud free pixel were examined for their cover and elevation. Should any of these direct lying eight neighbouring pixels be covered by snow and their elevation is lower than the cloudy pixel, then the cloudy pixel was assigned as being snow covered as well. This step is based on the fact that the temperature decreases as elevation increases, and therefore the snow at higher located pixels should melt later than the lower located pixels. If a pixel is covered by snow and any neighbouring pixel located at higher elevation was covered by cloud, then the higher elevation should also be covered by snow, theoretically. The formula describing this eventuality is given in Eq. (7):

$$S_{(y,x,t)} = 1 \quad \text{if} \quad \left(S_{(y+k,x+k,t)(k\in(-1,1))} = 1 \quad \text{and} \right.$$
$$\left. H_{(y+k,x+k)(k\in(-1,1))} < H_{(y,x)} \right) \tag{7}$$

The sixth step was based on a time series of each pixel over an entire year. For each pixel, 365 days were taken as the time series where pixel cover was given. The pixel cover starts with mainly snow and cloud cover until snowmelt on this pixel occurs and continues with land or cloud cover until snow accumulation starts. Using this information, the day when the pixel is no longer covered by snow and the day when the snow accumulations start should be able to be obtained. Accordingly, these threshold days were set as "complete snowmelt" and "snow accumulation start". Figure 3a and b show an illustration of this method.

The surface cover values 200, 50 and 25 in Fig. 3a and b correspond to snow, cloud and land cover, respectively. It can be seen from Fig. 3a that this pixel at elevation of 2091 m a.s.l. was covered mainly by snow (200) and cloud (50) until snow completely melts on the Julian day of 121 (1 May); the snow fall on this pixel starts on the Julian day of

311 (7 November). At higher elevation as shown in Fig. 3b, the snow completely melts later, as to be expected, at day 137 (17 May), while snow fall begins earlier at day 286 (13 October). This analysis was carried out for each pixel and all thresholds (snowmelt and snow accumulation) days were assigned according to the pixel cover time series information. All days for the same pixel after "complete snowmelt" and before "snow accumulation start" were set as land covered (25). For the pixel plotted in Fig. 3a, this translates into cloud covered days between 1 May and 7 November being set as land covered (25), and all cloud covered days from 1 January to 1 May and 7 November and 31 December as being snow covered (200). The formulas for this step are given in Eqs. (8) and (9):

$$S_{(y,x,t)} = 1 \quad \forall \quad t^A \leq t < t^M \tag{8}$$

$$S_{(y,x,t)} = 0 \quad \forall \quad t^A > t \geq t^M \tag{9}$$

where t^M and t^A are the threshold days of snowmelt and snow accumulation start, respectively and t is the day. Short term snow falls or snowmelts were not considered in this step and the first snowmelt day and the first snow covered day were taken as threshold days for eliminating clouds. A test was carried out where several snow free days after first snowmelt and several snow covered days after first snow appearance day were taken into consideration in order to assign the threshold days more precisely. The test with first snowmelt and first snow appearance day after the snow covered and land covered pixels, respectively, performed the best and were chosen for this step. Time series analyses were done from the beginning of the spring season (1 March) to the beginning of the spring season of the following year, because of possible large snow fall events happening in winter months. The Julian days of January and February were added to last year's Julian days in order to prevent error due to the lower Julian day value in those two months when applying Eqs. (8) and (9).

The sequence of steps applied in this methodology corresponded to the level of accuracy and effectiveness of each step. For example, the first step, which was based on the

Table 1. Cloud cover of original Aqua and Terra products and after implementation of each step (in%).

Day	AQUA	TERRA	Cloud cover remaining after implementation of steps (%)					
			1	2	3	4	5	6
27 Jan	98.2	98.7	97.1	50.5	23.8	22.3	17.7	0.0
12 Feb	66.9	91.8	65.2	40.1	29.0	27.7	22.9	0.0
16 Mar	76.0	90.8	70.5	27.4	26.4	24.4	18.3	0.0
14 Apr	80.4	47.5	42.1	36.0	22.4	21.5	16.5	0.0
18 Apr	89.6	99.4	89.0	34.0	6.4	5.0	3.2	0.0
1 May	90.6	91.8	83.6	79.2	31.7	31.0	28.4	0.0
22 May	79.2	62.9	55.5	13.7	13.6	12.2	9.4	0.0
22 Jun	57.1	54.2	34.6	18.1	8.7	8.0	6.5	0.0
11 Aug	25.8	21.7	10.8	8.2	4.9	4.4	4.3	0.0
28 Aug	67.9	28.2	21.1	5.3	3.7	3.1	3.1	0.0
26 Sep	89.7	91.2	83.5	16.4	3.8	3.1	2.6	0.0
6 Oct	64.7	45.0	28.7	15.8	14.5	13.2	11.4	0.0
2 Nov	91.5	30.6	30.6	7.3	1.2	1.0	0.8	0.0
21 Nov	94.2	62.7	60.8	21.2	12.8	11.3	9.0	0.0
3 Dec	97.5	91.2	89.3	52.9	41.0	39.3	29.1	0.0
29 Dec	98.2	98.9	97.1	77.8	42.4	41.6	31.1	0.0

Fig. 4. Cloud cover of original MODIS Aqua, MODIS Terra images and after implementation of five steps for 12 February 2003. **(a)** original MODIS Terra, **(b)** original MODIS Aqua, **(c)** after step 1, **(d)** after step 2, **(e)** after step 3, **(f)** after step 4, **(g)** after step 5, **(h)** after step 6. Colour legend: light blue-snow, dark brown-cloud, and green – no snow (land).

combination of the Terra and Aqua products, should theoretically be most effective and accurate since these concerns to observations with only a few hours of time shift. The probability of snow covered pixel observed by Terra satellite being observed as no snow cover by the Aqua satellite should be negligible since the time difference is small. Although the study by Wang et al. (2009) reported the Terra satellite to observe more snow cover than the Aqua satellite, but their agreements were close to 100%. The second step, where the combination of two previous and two next days were considered, is assumed to be most effective and accurate in this stage of cloud removal which generally removed most clouds with high accuracy comparing to amount of cloud removal. This was useful for the third step, where the threshold elevations for minimum and maximum snow elevation were assigned only if at least 70% of the study area was cloud free. The forth and fifth steps remove generally smaller portions of clouds, but are still considered as part of the methodology since their accuracy leads to an overall improvement of the resulting processed snow cover data quality. The sixth step where all remaining clouds were removed was applied at the end of the methodology because of its less accuracy than the other steps. Applying the step six as the last increased the overall accuracy of the methodology.

5 Results

Table 1 summarizes the cloud cover fraction of the original MODIS Aqua and MODIS Terra snow cover products and the cloud cover fraction after each step had been implemented for the Kokcha Basin for several days in the year 2003. In Table 2, the cloud fraction that was removed from each step was summarized for those days. Figures 4 and 5
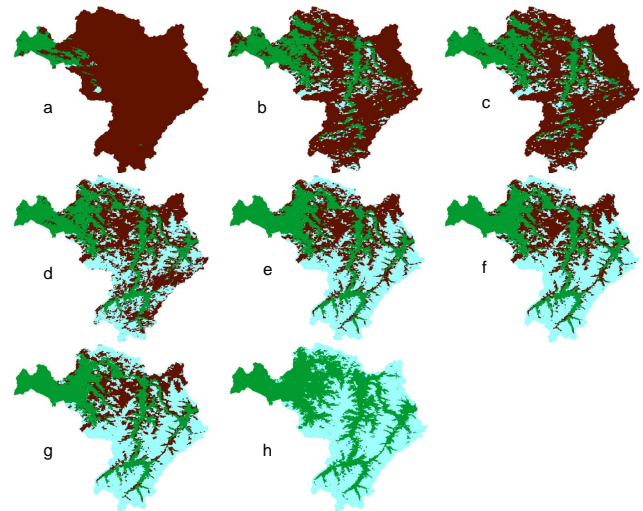
illustrate the cloud fraction removed after each step for the days of 12 February and 21 November, respectively. The results for these days are also given in Tables 1 and 2.

For 12 February 2003, the MODIS sensor the Terra satellite observed 91.8% cloud coverage (Fig. 4a) whereas the Aqua satellite after a time shift of a few hours observed 66.9% cloud cover (Fig. 4b). Obviously, this reflected the fact that the clouds were in continuous dynamic movement. The first step, which maximized cloud free coverage and was based on the combination of the Terra and Aqua product, removed 26.6% and 1.7% cloud cover compared to the Terra and Aqua products, respectively (Fig. 4c). More cloud cover was removed from the TERRA MODIS product due to less cloud coverage observed by the AQUA MODIS after few hours of time shift.

Processing using the second step, based on temporal combination of cloud covered pixels, removed 25% cloud cover, while 40.1% cloud cover remained (Fig. 4d). As mentioned before, this step performed well when either two of previous two or next two days were cloud free. The cloud coverage for the two days before the 12 February 2003 was 21% and 8%, and for the two days after the 12 February 2003, respectively was 68% and 89%. Particular to this case, more information on actual pixel cover was obtained from the previous two and next one day since they are less cloud covered than the next second day.

The third step based on snow transition elevation removed 11% cloud cover, while 29% remained (Fig. 4e). This step can be applied well at very low and high altitudes. For 12 February 2003, the maximum snow elevation zone
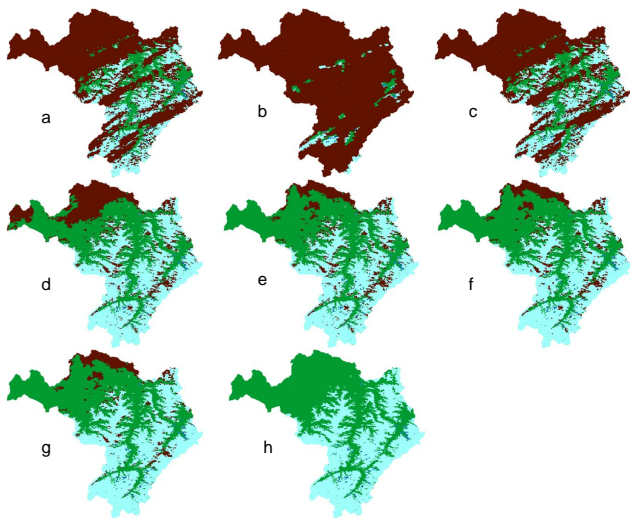
Fig. 5. Cloud cover of original MODIS Aqua, MODIS Terra images and after implementation of five steps for 21 November 2003. **(a)** original MODIS Terra, **(b)** original MODIS Aqua, **(c)** after step 1, **(d)** after step 2, **(e)** after step 3, **(f)** after step 4, **(g)** after step 5, **(h)** after step 6. Colour legend: light blue-snow, dark brown-cloud, and green – no snow (land).

Table 2. Performance of six subsequent steps when applied together (in %).

| Day | Steps | | | | | |
	1	2	3	4	5	6
27 Jan	1.6	46.7	26.6	1.6	4.5	17.7
12 Feb	26.6	25.0	11.1	1.3	4.8	22.9
16 Mar	20.3	43.1	1.0	2.0	6.2	18.3
14 Apr	5.4	6.1	13.6	1.0	5.0	16.5
18 Apr	10.4	55.0	27.7	1.3	1.8	3.2
1 May	8.2	4.4	47.5	0.7	2.5	28.4
22 May	7.4	41.9	0.1	1.4	2.8	9.4
22 Jun	19.6	16.5	9.3	0.8	1.5	6.5
11 Aug	10.8	2.6	3.3	0.5	0.2	4.3
28 Aug	7.2	15.7	1.6	0.6	0.1	3.1
26 Sep	7.7	67.1	12.5	0.7	0.5	2.6
6 Oct	16.3	12.9	1.3	1.2	1.8	11.4
2 Nov	0.1	23.3	6.2	0.1	0.2	0.8
21 Nov	33.3	39.6	8.4	1.5	2.3	9.0
3 Dec	1.9	36.4	11.9	1.7	10.2	29.1
26 Dec	1.8	19.3	35.5	0.8	10.5	31.1

Table 3. Contingency table for Terra and Aqua products in %.

March 7	AquaSnow	AquaLand
TerraSnow	84.61	0.09
TerraLand	0.97	14.33

(completely snow covered line) was 4322 m a.s.l. and the minimum elevation of last snow covered pixel observed was 1356 m a.s.l. All the pixels located above the maximum snow elevation line assigned as cloud cover in MODIS original product could be reassigned as snow cover, since these pixels should theoretically be covered by snow. The same applied to minimum snow elevation zones where all pixels located below this threshold line should be theoretically land covered pixels. As mentioned in the methodology section, the study area should be at least 70% cloud free in order to correctly assign the maximum and minimum snow lines. If not, this step can not be used. Assigning threshold snow lines when the basin was more cloud covered results in incorrect information for the threshold lines, which could enormously influence the cloud removal accuracy.

The fourth step based on neighboring pixel information removed only 1.3% cloud cover, with 27.7% cloud cover remaining (Fig. 4f). Although this step did not remove much cloud cover, it still formed part of the methodology due to its low error level. This helped to minimize the overall error.

The fifth step was based on neighboring pixel cover information, and its elevation removed 4.8% cloud cover, with 22.9% remaining (Fig. 4g). Processing with this step resulted in different cloud cover removal for different days because it was dependent upon cloud structure and was a function of time and the elevation of the neighboring pixels.

Finally, the sixth step which is based on time series information removed all of the remaining 22.9% cloud cover. The sixth step sets thresholds "complete snowmelt" and "snow accumulation start" for a single pixel over a whole year. Ac-

cording to this temporal threshold, the rest of the cloud coverage was removed (Fig. 4h). After the "complete snowmelt" threshold, all of the previous days of the same pixel are assigned as snow covered. After the "snow accumulation start" threshold, all the upcoming days of the same pixel are assigned to have snow cover as well. The days in between the two thresholds that are covered by cloud are assigned to have land cover.

For 21 November 2003, the MODIS sensor on the TERRA satellite observed 62.7% cloud cover (Fig. 5a) whereas the AQUA satellite observed 94% cloud cover (Fig. 5b). The implementation of the first step removed 33.3% cloud cover when compared to AQUA product and 1.9% cloud cover when compared to the Terra product; 60.8% cloud cover remained after the combination of AQUA and TERRA products (Fig. 5c). The second step removed 36.4% cloud cover with 21.2% remaining (Fig. 5d). The levels of cloud cover for the previous two and next two days amounted to 9%, 26%, 16% and 92%, respectively. From the third step, 8.4% cloud cover could be removed with 12.8% still remaining (Fig. 5e). The maximum snow elevation zone for this day was observed at 4319 m a.s.l.; the minimum snow elevation for the last snow covered pixel was at 1961 m a.s.l. The fourth step
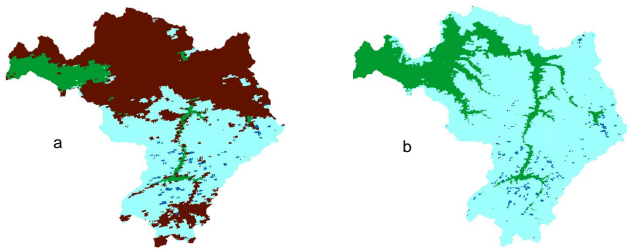
Fig. 6. MODIS Original Aqua (**a**) and processed (**b**) snow cover product for 8 March 2003.
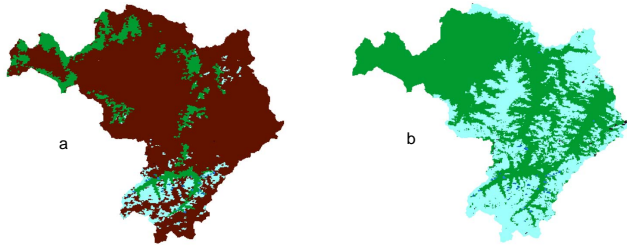

Fig. 8. MODIS Original Aqua (**a**) and processed (**b**) snow cover products for 22 May 2003.

Fig. 7. MODIS Original Terra (**a**) and processed (**b**) snow cover product for 7 April 2003.

Fig. 9. Comparison of Terra original and processed (after step 5) data for the year 2003.

removed a further 1.5% cloud cover, leaving 11.3% cloud covered pixels remaining (Fig. 5f). The fifth step led to the removal of 2.3% cloud cover with 9% left over (Fig. 5g). Finally, the last step removed all of the remaining 9% cloud cover, as exemplified in Fig. 5h.

As also seen from Tables 1 and 2, different steps perform differently and this depended mainly on the structure of cloud cover.

Using above mentioned methodology where six subsequent steps are used, all pixels with cloud cover from original MODIS snow cover products were reassigned as either snow or land cover. The images in Figs. 6, 7, and 8 show the original MODIS snow cover products compared to the MODIS snow cover product after applying six subsequent steps that removed cloud cover for different days.

The overall performance of the methodology for the year 2003 is illustrated in Fig. 9 where the original Terra satellite cloud cover fraction is compared with the results of cloud cover fraction after applying steps 1 to 5. Seeing as the sixth step has removed the rest of clouds and the fraction of cloud after this is therefore zero, it has not been plotted in Fig. 9. A visible improvement in the snow product can be observed after the processing with the fifth step.

6 Validation

Since there is no information available from the study area to validate the methodology, the original MODIS snow cover products with least cloud cover were used for validation purposes. Several days with least cloud covered MODIS snow
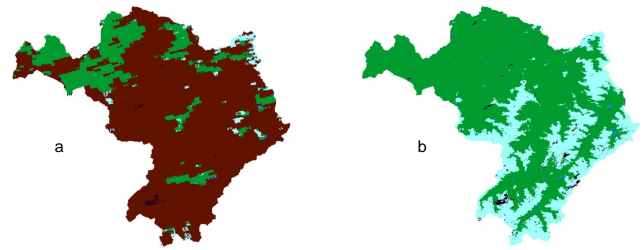
product were filled by clouds of another densely cloud covered snow product. In this way we cloud generate "observed" snow cover product where the performance of the six previously mentioned steps can be validated. In the validation section we neglected the first step which is based on combination of Terra and Aqua snow cover products. Since the time shift between two satellites was a few hours, it was accepted that the coverage remained more or less unchanged. The contingency table for Aqua and Terra snow products is illustrated in Table 1. The table shows good accuracy for 7 March and little disagreement for 8 April. More than 4% of pixel cover was assigned as land from TERRA satellite and snow from AQUA satellite and this can be due to the snowfall during this shifting time on that day.

For validation purposes, 26 March 2003 (validation day 1) and 17 April 2003 (validation day 2) were selected because of least cloud coverage observed on these days. For validation day 1, Terra satellite observed less cloud cover (6.6%) than the Aqua satellite (14.3%) and for this reason the original Terra snow cover product from this day was filled by the cloud cover values of 22 March 2003 Terra snow cover product. The cloud cover fraction of the study area on 22 March 2003 was 98%. For validation day 2, the Aqua satellite observed less cloud cover with only 3% over the entire Kokcha Basin. The original Aqua cloud cover pixels from 14 April 2003 with 80% cloud fraction were assigned to this day for validation purposes. The generated snow cover maps with assigned cloud cover pixels were used as an input for a proposed methodology (excluding step 1) and the results were compared with the original snow cover products without cloud filling. Tables 4 and 5 show the validation

Table 4. Validation results for each step in% for 26 March 2003.

Steps	Total Cloud Elimination	TRUE	FALSE	Snow to Snow	Land to Land	Snow to Land	Land to Snow
2	0.50	0.46	0.02	0.35	0.11	0.01	0.02
3	35.52	34.61	0.84	0.00	34.61	0.84	0.00
4	0.41	0.30	0.11	0.01	0.29	0.11	0.00
5	1.25	1.17	0.04	1.11	0.07	0.03	0.01
6	62.32	56.07	4.98	52.68	3.38	0.32	4.65
Total	100.00	92.61	5.99	54.15	38.46	1.31	4.68

Table 5. Validation results for each step in% for 17 April 2003.

Steps	Total Cloud Elimination	TRUE	FALSE	Snow to Snow	Land to Land	Snow to Land	Land to Snow
2	10.34	9.42	0.54	3.70	5.72	0.36	0.17
3	38.17	38.17	0.00	0.06	38.11	0.00	0.00
4	1.02	1.00	0.02	0.14	0.86	0.02	0.01
5	12.09	11.37	0.50	7.11	4.26	0.26	0.25
6	38.38	31.53	6.11	21.74	9.78	0.38	5.73
Total	100.00	91.49	7.17	32.75	58.73	1.02	6.16

results for validation day 1 and validation day 2.

As it can be seen from the table, the validation results were reasonably good. As mentioned already, different steps perform differently depending on the structure of the cloud coverage. For example, on validation day 1 (Table 4), steps 3 and 6 performed best whereas the other steps performed less well. Step 2, which was based on temporal information, did not perform well, because the previous two and next two days of 26 March had cloud coverage of 59%, 100%, 93% and 97%, respectively

Step 3, which was based on snow transition elevation, performed very well with a disagreement of only 0.84% for the entire catchment. Although step 4 usually returned the lowest amount of cloud removal, it was still included in the methodology. Step 5 was based on the neighboring pixel cover information and the elevation between them. This step also removed a small percentage of clouds, though with high accuracy.

As should be clear, the remaining cloud cover after step 5 was removed after processing with step 6. The processing technique of step 6 was applied at the end because its results were usually at the highest disagreement compared to other steps. Usually, the technique of step 6 tends to assign more snow for land covered pixels. This can occur due to incorrect assignment of "complete snowmelt" and "snow accumulation start" threshold days in step 6. The difficulty in assigning precisely the threshold days comes from the pos-

sibility for short spells of snow fall and snowmelt during the snow accumulation and snowmelt transition period.

On validation day 2 (Table 5), steps 2 and 5 also performed reasonably well. Step 2 removed 10.3% where the cloud coverage of the previous two and next two days was 100%, 55%, 89% and 75%, respectively. Step 3 removed 38.2% cloud cover with 100% accuracy. Step 4 on the other hand, performed poorly but, as already stated, with high accuracy. The performance of step 5 varied, because it was dependent on the cloud coverage structure as seen from both validation days. The rest of the cloud cover (38.4%) was removed after processing with step 6.

As is illustrated in the contingency tables, the proposed methodology showed high accuracy for removing clouds from original cloud containing MODIS snow cover product. Processing with step 6 resulted in the lowest accuracy among the steps tested, though, nevertheless, it was more acceptable to have about 6% cloud coverage error than having almost 62% of the entire basin being covered by clouds (validation day 1). This performance also changed from day to day as can be seen from validation results. Some more validation results for 2003 are given in Table 6 and Figs. 10 and 11.

7 Validation of individual steps

In addition to the validation of the processing where each step removed progressively more and more cloud, the

Table 6. Validation result for different days in 2003.

Days	Steps	Total Cloud Elimination	TRUE	FALSE	Snow to Snow	Land to Land	Snow to Land	Land to Snow
9 Feb	2	55.91	55.03	0.56	38.59	16.44	0.29	0.27
	3	11.35	10.81	0.17	9.95	0.85	0.01	0.16
	4	1.66	1.57	0.06	1.23	0.34	0.01	0.05
	5	6.96	6.37	0.42	4.14	2.23	0.14	0.28
	6	24.12	19.33	4.29	7.48	11.85	3.57	0.72
6 Mar	2	64.03	63.7	0.02	56.02	7.68	0.01	0.01
	3	13.83	12.28	0.67	6.75	5.53	0.64	0.02
	4	1.22	1.14	0.06	0.93	0.22	0.05	0
	5	2.41	2.36	0.04	1.88	0.47	0.01	0.03
	6	18.52	13.22	5.19	13.18	0.04	0	5.18
6 May	2	78.13	73.94	3.51	26.55	47.39	2.7	0.81
	3	7.12	6.58	0.35	3.87	2.72	0.02	0.32
	4	1.38	1.24	0.1	0.55	0.68	0.08	0.03
	5	4.02	3.22	0.69	1.42	1.79	0.47	0.22
	6	9.36	6.69	2.56	3.48	3.21	1.02	1.54
28 Nov	2	5.94	5.4	0.49	1.28	4.12	0.21	0.28
	3	41.71	39.93	1.38	14.64	25.29	0.22	1.16
	4	1.18	1.04	0.12	0.46	0.59	0.04	0.07
	5	5.24	4.39	0.76	1.72	2.66	0.37	0.39
	6	45.93	38.69	6.9	16.57	22.12	3.72	3.18

Table 7. Validation results for individual steps of 1 and 2.

Steps	Total Cloud Elimination	TRUE	FALSE	Snow to Snow	Land to Land	Snow to Land	Land to Snow
1	24.73	22.82	1.04	13.37	9.45	0.52	0.52
2	75.27	73.47	1.42	48.98	24.48	0.43	0.99
Total	100.00	96.29	2.46	62.35	33.93	0.95	1.51

performance of the individual steps was also tested. As a benchmark for the validation, the combination of Terra and Aqua products, i.e. which forms the basis of step 1 and gave the best results, was set against the individual validations of the other steps. Tables 7 to 11 show the validation of other steps (2 to 6) in relation to step 1. This validation was carried out for 2 February 2003.

For the individual validation, all possible cloud elimination from two steps was normalized to 100%. In Table 7, 24.7% of the cloud was eliminated using the step 1 while 75.3% of the cloud was eliminated using the step 2. As can be seen, both steps performed with little error (1.04% and 1.42%).

In the case of steps 1 and 3, most of the cloud cover was eliminated using step 1 (88%). Only around 12% was eliminated by step 3. This result shows that very high pixels in the catchment area were not covered by cloud on that day since

no value for the snow to snow result was returned. It can be seen that all cloud cover removal took place at areas of very low elevation.

As with step 3, step 4 showed little improvement in eliminating cloud cover but the performance was acceptable enough to include it in this methodology.

Step 5 performed well. More than 46% of the cloud cover was removed with 44.5% of this fraction being estimated correctly.

Step 6 eliminates all clouds according to the time series method, although results showed the highest error. Nevertheless, the error was acceptable when studying high mountainous and ungauged basins where no alternative information was available. For example, around 13% of the total 88.6 % cloud elimination was FALSE. This was perhaps due to the fact that only 2 steps (1 and 6) were included in this validation and 88.6% cloud was removed applying step 6. This

Table 8. Validation results for individual steps of 1 and 3.

Steps	Total Cloud Elimination	TRUE	FALSE	Snow to Snow	Land to Land	Snow to Land	Land to Snow
1	88.38	81.53	3.71	47.78	33.76	1.86	1.85
3	11.62	11.62	0.00	0.00	11.62	0.00	0.00
Total	100.00	93.15	3.71	47.78	45.38	1.86	1.85

Table 9. Validation results for individual steps of 1 and 4.

Steps	Total Cloud Elimination	TRUE	FALSE	Snow to Snow	Land to Land	Snow to Land	Land to Snow
1	97.73	90.16	4.1	52.83	37.33	2.06	2.04
4	2.27	2.22	0.04	1.37	0.85	0.01	0.03
Total	100	92.38	4.14	54.2	38.18	2.07	2.07
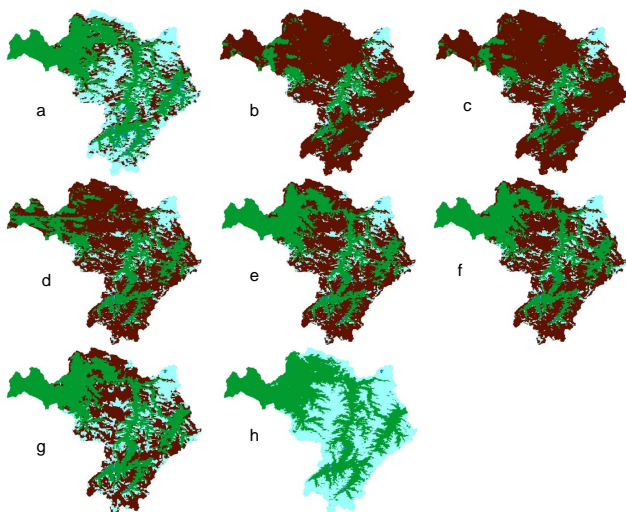
Fig. 10. Validation results for 3 January 2003. **(a)** original MODIS Terra snow cover for 3 January, **(b)** original MODIS Aqua snow cover for 6 January, **(c)** snow cover data from 3 January filled with cloud values of 6 January snow cover data, **(d)** after step 2, **(e)** after step 3, **(f)** after step 4, **(g)** after step 5, **(h)** after step 6. Colour legend: light blue-snow, dark brown-cloud, and green – no snow (land).
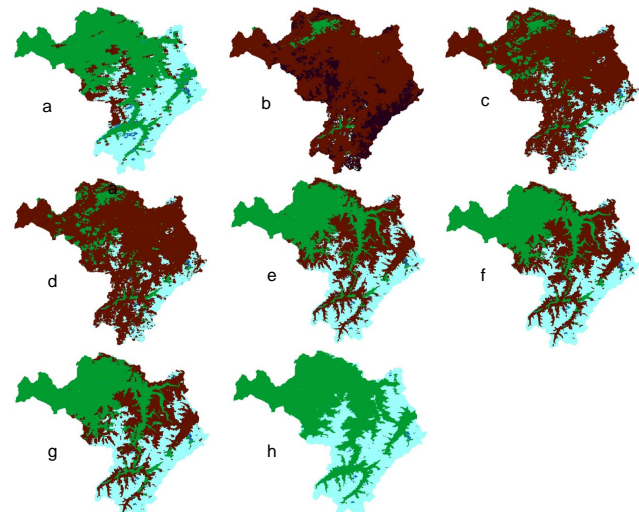
Fig. 11. Validation results for 13 May 2003. **(a)** original MODIS Terra for May13, **(b)** original MODIS Aqua for 12 May, **(c)** snow cover data from 13 May filled with cloud values of 12 May snow cover data, **(d)** after step 2, **(e)** after step 3, **(f)** after step 4, **(g)** after step 5, **(h)** after step 6. Colour legend: light blue-snow, dark brown-cloud, and green – no snow (land).

means that the total cloud removal after step 1 was carried out using step 6. This is why this step is used at the end, because it produces the largest error.

However, as mentioned already, by applying all 6 of the steps in succession, the error can be evened out due to the results of higher accuracy from the processing by the other steps. The fraction of cloud cover that remained after five

processing steps was quite low, so that the error after the sixth and final processing step would affect the overall accuracy significantly.

Table 10. Validation results for individual steps of 1 and 5.

Steps	Total Cloud Elimination	TRUE	FALSE	Snow to Snow	Land to Land	Snow to Land	Land to Snow
1	53.65	49.5	2.25	29	20.49	1.13	1.12
5	46.35	44.5	1.25	32.83	11.68	0.54	0.71
Total	100	94	3.5	61.83	32.17	1.67	1.83

Table 11. Validation results for individual steps of 1 and 6.

Steps	Total Cloud Elimination	TRUE	FALSE	Snow to Snow	Land to Land	Snow to Land	Land to Snow
1	11.41	10.52	0.48	6.17	4.36	0.24	0.24
6	88.59	74.26	13.43	52.49	21.77	11.13	2.3
Total	100	84.78	13.91	58.66	26.13	11.37	2.54

8 Discussion and conclusion

The objective of this study was to remove cloud covered areas from the original MODIS Terra and Aqua satellite snow cover products to obtain snow cover information for data limited regions such as the Kokcha Basin where no snow cover data is available locally. The methodology proposed to remove clouds from original MODIS snow cover products used a combination of six steps which were applied successively one after another. At each processing stage, each step estimated the pixel cover that was covered by cloud. The results of each step were used as an input for next round of processing where more and more cloud covered pixels was estimated and eliminated.

The methodology behind the proposed six steps was based on a combination of temporal and spatial information from original MODIS snow cover products from the TERRA and AQUA satellites and an elevation map. The validation results showed a high degree of accuracy when comparing original "cloud free" snow cover filled with cloud cover from a different day MODIS snow cover data. Steps 1 to 5 had a performance accuracy of more than 95%. After final processing was done using step 6, where all remaining cloud covered pixels were removed according to the time series information of each pixel, the accuracy still remained at above 90% overall. As was shown in validation section, different days gave different performances, and this was obviously linked to the dynamic structure of cloud cover.

Very high accuracies with some steps were observed, although it was often usual in those cases that only a small fraction of the clouds were removed. The validation of individual steps was also carried out where each step was tested separately, in relation to the combination of the Terra and Aqua products as a benchmark. The results of this validation indicated that the methodology was reasonably robust.

Only the last processing, step 6, based on the time series information, produced errors of more than 10%. The initial processing with step 1 gave the most accurate results, since it was based on a true estimation of pixel cover by the same sensor a few hours before or after a time shift. Processing with step 6 gave the highest cloud removal amounts. On the other hand, as just mentioned, the accuracy of this processing step was very low due to the false estimation of the complete snowmelt and snow accumulation date thresholds.

For the proposed methodology the FORTRAN program was developed and this can be tested in any region where snow information plays an important role for hydrological cycle. As snow plays key role in many mountainous regions, the daily snow cover data may enormously improve the modeling experience in hydrology.

Using proposed methodology it was possible to estimate the snow cover dynamics of mountainous regions. The original MODIS snow cover product that contains cloud covered pixels can be processed using this methodology, meaning that snow cover data with 500 m spatial and daily resolution can be prepared. MODIS also offers eight day composite snow cover information with little or no cloud cover, but this is the maximum snow cover extent for eight composite days. Yet, a considerable fraction of snow could melt or fall within eight days time period, which may not be enough information for higher temporal resolution modelling purposes. This is why the daily snow cover information is very valuable also for model calibration and for validation purposes. Such information can be extremely helpful when modeling available water resources in mountainous areas where snowmelt in spring or summer becomes a valuable raw material for energy production, agriculture and for drinking purposes in lowland areas.

References

Aizen, V., Aizen, E., Glazirin, G., and Loaiciga, H. A.: Simulation of daily runoff in Central Asian alpine watersheds, J. Hydrol., 238, 15–34, 2000.

Ault, T., Czajkovski, K., Benko, T., Coss, J., Struble, J., Spongberg, A., Templin, M., and Gross, C.: Validation of the MODIS snow product and cloud mask using the student and NWS cooperative station observations in the Lower Great Lakes Region, Remote Sens. Environ., 105, 341–353, 2006.

Hall, D. K., Riggs, G. A., Salomonson, V. V., DiGirolamo E. N., and Bayr J. K.: MODIS snow-cover products, Remote Sens. Environ., 83, 181–194, 2002.

Hall, D. K., Riggs, G. A., and Salomonson, V. V.: 2007, updated daily. MODIS/Aqua Snow Cover Daily L3 Global 500 m Grid V005, [01.01.2003–31.12.2003]. Boulder, Colorado USA: National Snow and Ice Data Center. Digital media, 2007.

Jarvis, A., Reuter, H.I., Nelson, A., and Guevara, E.: Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT), available from http://srtm.csi.cgiar.org, 2008.

Kimura, F. and Sato, T.: The roles of orography on Northeastern Asia dry climate, International Workshop on Terrestrial Change in Mongolia, 2002.

Klein, A., Barnett, A., and Lee, S.: Evaluation of MODIS snow-cover products in the Upper Rio Grande River Basin, Geophys. Res. Abstr., 5, 12420, 2003

Kling, H., Rürst, J., and Nachtnebel, H. P.: Seasonal, spatially distributed modelling of accumulation and melting of snow for computing runoff in a long-term, large-basin water balance model, J. Hydrol. Process., 20, 2141–2156, 2006.

Liang, T., Zhang, X., Xie, H., Wu, C., Feng, Q., Huang, X., and Chen, Q.: Toward improved daily snow cover mapping with advanced combination of MODIS and AMSR-E measurements, Remote Sene. Environ., 112, 3750–3761, 2008.

Maurer, E., Rhoads, J., Dubayah, R., and Lattenmeier, D.: Evaluation of the snow covered area data product from MODIS, J. Hydrol. Process., 17, 59–71, 2003.

Parajka, J. and Blöschl, G.: Validation of MODIS snow cover images over Austria, Hydrol. Earth Syst. Sci., 10, 679–689, 2006, http://www.hydrol-earth-syst-sci.net/10/679/2006/.

Parajka J. and Blöschl, G.: Spatio-temporal combination of MODIS images – potential for snow cover mapping, Water Resour. Res. 44, W03406, doi:10.1029/2007WR006204, 2008.

Pepe, M., Boschetti, L., Brivio, P., and Rampini, A.: Accuracy benefits of fuzzy classifier in remote sensing data classification of snow, Fuzzy Systems Conference, Milan, Italy, 23–26 July 2007, 9796063, 2007.

Rodell, M. and Houser, P. R.: Updating a Land Surface Model with MODIS – Derived Snow Cover, J. Hydrometeorol., 5, 1064–1075, 2004.

Tekeli, A. E., Akyürek, Z., Sorman, A. A., Sensoy, A., and Sorman A. Ü.: Using MODIS snow cover maps in modeling snowmelt runoff process in the eastern part of Turkey, Remote Sens. Environ., 97. 216–230, 2005.

Singh, P., Arora, M., and Goel, N.: Effect of climate change on runoff of glacierized Himalayan basin, J. Hydrol. Process., 20, 1979–1992, 2006.

Wang, X., Hie, H., Liang, T., and Huang, X.: Comparison of validation of MODIS standard and new combination of Terra and Aqua snow cover products in northern Xinjiang, China, J. Hydrol. Process., 23, 419–429, 2009.

Wang, X., Xie, H., and Liang, T.: Evaluation of MODIS snow cover and cloud mask and its application in Northern Xinjiang, China, Remote Sens. Environ., 112, 1497–1513, 2008.