Snow cover reconstruction methodology for mountainous regions based on historic in-situ observations and recent remote sensing data

A. Gafurov[1], S. Vorogushyn[1], D. Farinotti[1], D. Duethmann[1], A. Merkushkin[2], B. Merz[1]

[1] {GFZ German Research Centre for Geosciences, Section 5.4: Hydrology, Potsdam, Germany}

[2] {Uzbek Hydrometeorological Service (Uzhydromet), Tashkent, Uzbekistan}

Correcpondence to: A. Gafurov (gafurov@gfz-potsdam.de)

Abstract

Spatially distributed snow cover extent can be derived from remote sensing data with good accuracy. However, such data are available for recent decades only, after satellite missions with proper snow detection capabilities were launched. Yet, longer time series of snow cover area (SCA) are usually required e.g. for hydrological model calibration or water availability assessment in the past. We present a methodology to reconstruct historical snow coverage using recently available remote sensing data and long-term point observations of snow depth from existing meteorological stations. The methodology is mainly based on correlations between station records and spatial snow cover patterns. Additionally, topography and temporal persistence of snow patterns are taken into account. The methodology was applied to the Zerafshan River basin in Central Asia – a very data-sparse region. Reconstructed snow cover was cross-validated against independent remote sensing data and shows an accuracy of about 85%. The methodology can be used in mountainous regions to overcome the data gap for earlier decades when the availability of remote sensing snow cover data was strongly limited.

1 Introduction

Water resources from remote mountain catchments play a crucial role for the development of regions in or in the vicinity of mountain ranges (Pellicciotti et al., 2011). Seasonal snow is an important water resource in many of Earth's semiarid regions (Durand et al., 2008).

Particularly, in Central Asia, seasonal snowmelt decisively contributes to the total runoff volume (Ososkova et al., 2000, Unger-Shayesteh et al., 2013).

Information on snow cover and snow depth and ideally on snow water equivalent in Central Asian catchments is crucial for seasonal forecasts of water availability and for calibration and validation of hydrological models. However, the available sparse station-based data are insufficient to represent the snow cover variability over the large and remote mountain areas (Erickson et al., 2005). The development of remote sensing techniques during recent decades allows to derive snow cover spatially (Liu et al., 2012). Widely used remotely sensed snow cover products are from Advanced Very High Resolution Radiometer (AVHRR), Landsat and Moderate Resolution Imaging Spectroradiometer (MODIS) missions. Whereas AVHRR (launch 1978) and Landsat (launch 1972) offer remote sensing data for a longer period, MODIS is available only after 2000. However, snow cover from Landsat and AVHRR needs to be derived by end-user themselves, whereas MODIS offers already compiled snow cover product. The above mentioned snow products are extremely useful to study snow cover worldwide, however they are strongly limited by the presence of clouds. Recently, the reconstruction of snow cover time series from AVHRR data for Central Asia has been reported by Zhou et al. (2013), but they are also limited in time starting in 1986 at earliest.

Several studies used remotely sensed snow cover either as input to hydrological models (Tekeli et al., 2005, Immerzeel et al., 2008, Li et al., 2008, Wang et al., 2010) or for calibration and validation purposes (Parajka and Blöschl, 2008, Corbari et al., 2009, Liu et al., 2012, Duethmann et al., 2014). Particularly, for hydrological model calibration, spatially distributed snow cover data offer high information content required to constrain model parameters (Finger et al., 2011, Duethmann et al., 2014).

In Central Asia, continuous hydro-meteorological records are widely available from 1960s and earlier until the collapse of the Soviet Union in 1991. In contrast, continuous remote sensing snow cover data from MODIS are readily available after 2000, when station data are very scarce. We present a methodology which enables reconstructing historical snow cover pattern using long-term, point-based observations from existing meteorological stations and recent remotely sensed snow cover data. By merging high-resolution spatial satellite data with long-term station data, snow cover patterns can be reconstructed for several decades into the past.

Only a limited number of studies on snow cover reconstruction have been conducted in the past that use long term station observations and recent remote sensing data (Robinson, 1991, Brown,

2000, Frei et al., 1999, Brown and Robinson, 2011). These studies are however conducted at the continental scale under conditions of dense station network availability and neglecting the effect of topography. Robinson (1991) and Frei et al. (1999) conducted reconstruction of snow cover based on regression analysis between snow characteristics and snow cover area derived from AVHRR satellite observations. As snow characteristics both studies used snow cover duration derived from interpolated station records. Another study by Brown (2000) conducted reconstruction of snow cover for "pre-satellite era" interpolating snow depth data from station network. For grid cells of nearly 200 km, the interpolation of snow cover was done using different thresholds for snow depth and compared against NOAA SCE (Snow Cover Extent) during "satellite era". The calibration showed 2 cm to be most appropriate snow depth for accurate snow cover reconstruction based on station data. Brown and Robinson (2011) updated and extended the snow cover reconstruction of Brown (2000) to the period 1922-2010 and used this data for trend analysis of snow cover extent in the Northern Hemisphere. These studies can be helpful in assessing climate related variations of snow cover but are hardly transferable to smaller catchment scale with moderate resolution and limited station data availability.

Different to those studies mentioned above, we present a methodology for snow cover reconstruction 1) with moderate spatial resolution (500 m), 2) suitable for catchment scale hydrological studies, 3) accounting topography and 4) delivering spatially distributed snow cover maps. The methodology takes advantage of the strong control of topography on the spatial snow cover distribution. Hence, measurements from snow observation stations at different elevations can be interpreted as representative sites to predict snow cover patterns. The methodology consists of five successive steps which make use of topographic information and correlations between station records and spatial snow cover patterns. In order to test the presented methodology, snow cover reconstruction was conducted for four days (Table 1) for which independent Landsat data were available.

2 Study area

The methodology for snow cover reconstruction was developed and tested for the area containing the Upper Zerafshan River basin, Central Asia (Fig. 1).

The Upper Zerafshan River basin is located in the Gissaro-Alai Mountain Range. Elevation ranges from 658 to 5402 m.a.s.l. and basin area is about 12 000 km². The Zerafshan River basin is currently an endorheic basin in the inner Central Asia that no longer contributes to the

Amudarya River. It originates in Tadjikistan and flows towards Uzbekistan, where its water is used for agricultural production in the oases of Samarkand and Bukhara. The flow regime is strongly dominated by snow and glacier melt, as can be inferred from the temperature, precipitation, SCA and flow regimes in Fig. 2. The highest precipitation is brought by westerly flows during winter and spring, with a clear minimum during summer and early autumn (Aizen et al., 1995). The highest runoff, however, occurs during summer months and is driven by snow and glacier melt. According to MODIS Landcover product from 2009, the main land cover types in the study area are grasslands (60 %), croplands (9 %), open shrublands (7 %), woody savannahs (6 %) and permanent snow and ice covered areas (5 %).

3 Data

We used (1) daily in-situ snow depth data, (2) daily MODIS snow cover data, (3) a digital elevation model (DEM), and (4) Landsat data. The first three datasets were used for snow cover reconstruction whereas Landsat data was used as an independent dataset to validate the results.

3.1 In-situ snow depth data

Daily snow depth data in the period from 1964 to 2012 were available for seven climate stations located at different elevations (Fig. 1, Table 1). These data contain continuous snow depth measurements including records on no-snow conditions. Snow depth in these stations are recorded at 1 cm threshold depth. The data were provided by Uzbek Hydrometeorological Service (Uzhydromet).

3.2 MODIS snow cover data

MODIS daily snow cover data from the Terra satellite with 500 m spatial resolution (MOD10A, version V005) were employed for the time period of 2000 to 2012. We used MODIS Terra snow cover data due to its longer time series compared to the Aqua satellite, which delivered snow cover data only after 2002. The MODIS snow cover product is based on the Normalized Difference Snow Index (NDSI) algorithm (Hall et al., 2002a). Its accuracy was tested in different parts of the world showing good agreement with in-situ data (Klein et al., 2003, Tekeli et al., 2005, Parajka et al., 2006, Ault et al., 2006, Wang et al., 2008, Liang et al., 2008, Huang

et al., 2011, Gafurov et al., 2013, Parajka et al., 2012). The main drawback of MODIS snow cover data is the limitation due to cloud cover. There have been several studies on filtering methods for reducing cloud cover or even removing it completely (e.g., Parajka and Blöschl, 2008; Gafurov and Bárdossy, 2009; Tong et al., 2009; Hall et al., 2010; Lòpez-Burgos et al., 2013). We used original MODIS snow cover data to exclude any uncertainty that may be introduced by cloud filtering. The data were obtained from National Aeronatics and Space Administration (NASA)'s Earth Observing System Data and Information System (EOSDIS) Reverb platform. MODIS data are distributed as tiles with the size of 10° by 10°, which makes up a total of 36 horizontal (h) and 18 vertical (v) tiles covering the entire globe. In this study, the tile h23v05 was used which covers the study area completely.

3.3 Digital elevation model

The void-filled DEM with 90 m spatial resolution from NASA Shuttle Radar Topography Mission (SRTM) was used. SRTM DEM data was obtained from the CGIAR CSI (Consultative Group on International Agricultural Research, Consortium for Spatial Information) Database (www.cgiar-csi.org/data). To have the same resolution as the MODIS data (500 m), the 90 m SRTM DEM was aggregated to 500 m.

3.4 Landsat data

Optical remote sensing data from the Landsat Thematic Mapper (TM) sensor were used to validate the reconstructed snow cover maps. The Landsat data have a spatial resolution of 30 m and a temporal resolution of 16 days. Landsat data from four nearly clear-sky days in the snow season (April 10th, 1998, November 20th, 1998, April 20th, 1999 and November 15th, 1999) were used for validation purposes. Snow cover maps for the Landsat footprint (see Fig. 1) were prepared using the NDSI methodology. For a detailed description of the algorithm used for deriving snow cover maps from Landsat refer to Gafurov et al. (2013). Figure 3 shows raw and processed Landsat snow cover maps for the study area.

Since Landsat has a spatial resolution of 30 m and snow reconstruction was performed for 500 m pixels based on MODIS resolution, the processed Landsat snow cover maps were spatially aggregated to 500 m resolution. This was done by classifying each of the 500 m pixels as

snow-covered or snow-free, based on the majority of the 30 m Landsat pixels within the 500 m pixel.

4 Methodology

The methodology presented hereafter is based on similarity of different locations in terms of presence or absence of snow at a given time. The idea is to use the information about the presence of snow from one location in order to estimate the presence of snow at another location. The similarity between different locations was assessed using both observed snow cover at meteorological stations (i.e. records of snow depth > 0) and MODIS snow cover data. This similarity was quantified using the concept of conditional probability. The conditional probability gives the probability of one event (e.g. snow cover at a pixel) to occur given that another event (e.g. snow depth>0 at station) has already occurred. Additionally to the similarity between two locations, temporally persistent monthly snow cover patterns and elevation based classifications were used. The methodology consists of five successive steps where each step estimates a certain fraction of snow cover. This leads to a complete snow cover reconstruction after step 5. Once the similarities and temporally persistent snow fields are established using existing stations and remote sensing data, the methodology can be applied to other time periods and is solely based on snow records at meteorological stations. The following five steps are detailed in the next sections:

1. Pixel to station CP fields
2. Temporally persistent monthly probability fields
3. Pixel to pixel CP fields
4. Usage of elevation information
5. Pixel to station CP for CP <1

4.1 Pixel to station conditional probability

In the first step, we consider the CP of each pixel, given the observed data from a set of snow stations. We compute the CP of each pixel as follows:

$$P^s(S_{x,y}|S_n) = \frac{\sum(1-ABS(S_{x,y,t}-S_{n,t}))}{N_{x,y}} \quad \forall \quad S_{n,t} = 1 \tag{1}$$

$$P^l(S_{x,y}|S_n) = \frac{\sum(1-ABS(S_{x,y,t}-S_{n,t}))}{N_{x,y}} \quad \forall \quad S_{n,t} = 0 \tag{2}$$

where $P^s(S_{x,y}|S_n)$ ($P^l(S_{x,y}|S_n)$) is the CP of a pixel with coordinates x, y to be covered by snow (land) given that the station n also records snow depth>0 (=0) at the same time. $S_{x,y,t}$ and $S_{n,t}$ are binary variables indicating the presence ($S = 1$) or absence ($S = 0$) of snow at pixel x, y and station n for day t, respectively. $N_{x,y}$ is the number of observations simultaneously available at pixel x, y (excluding cloud covered days) and station n over the 12 years (2000-2012), for which station n showed snow ($S = 1$) or snow free ($S = 0$) conditions. The value of $P^s(S_{x,y}|S_n)$ ($P^l(S_{x,y}|S_n)$) varies from 0 to 1, with $P^s(S_{x,y}|S_n) = 1$ ($P^l(S_{x,y}|S_n) = 1$) indicating that a pixel at x, y was always observed as snow covered (snow free) in the MODIS data when station n measured snow depth > 0 (= 0), whilst $P^s(S_{x,y}|S_n) = 0$ ($P^l(S_{x,y}|S_n) = 0$) indicates an opposite relationship, i.e. that the MODIS product at x, y always showed snow free (snow covered) conditions when station n had snow depth > 0 (= 0) .

CPs were computed for each MODIS pixel in the study area (total of 169 776 pixels) using over 12 years of available MODIS data and observed snow depth measurements. Hence, the daily snow cover maps from MODIS are treated as snow observation for each 500 m grid cell, giving rise to a very dense "observation network". An example for a CP map for snow and land conditions for Chimgan station is given in Fig. 4. In total, 14 maps were derived (two maps for every of the 7 stations: one for $P^s(S_{x,y}|S_n)$ and one for $P^l(S_{x,y}|S_n)$).

The number of pixels with $P^s(S_{x,y}|S_n) = 1$ ($P^l(S_{x,y}|S_n) = 1$) varies from station to station. The higher the number of pixels with $P^s(S_{x,y}|S_n) = 1$, the higher is the predictive power of the station for snow classification. Similarly, the higher the number of pixels with $P^l(S_{x,y}|S_n) = 1$, the higher is the predictive power of the station to predict snow-free conditions. In order to quantify the predictive power of each station, we introduce two terms: Snow Predictability Index (SPI) and Land Predictability Index (LPI). These terms give the fractions of the reconstruction domain with $P^s(S_{x,y}|S_n)$ = 1 for a given station for snow and land conditions, respectively:

$$SPI_n = \frac{\sum(P^s(S_{x,y}|S_n)=1)}{N} \cdot 100 \qquad [\%] \tag{3}$$

$$LPI_n = \frac{\sum(P^l(S_{x,y}|S_n)=1)}{N} \cdot 100 \qquad [\%] \tag{4}$$

where SPI_n and LPI_n are the Snow Predictability Index and the Land Predictability Index of station n, respectively. N is the total number of pixels (169 776) in the entire study area. Pixels with $P^s(S_{x,y}|S_n) = 1$ in Fig. 4 (top) add up to 10.2 % which is the SPI value of the Chimgan station for the entire domain. This means that when the Chimgan station shows a snow depth above zero, 10.2 % of the study area can be classified as snow covered as well. Pixels with $P^l(S_{x,y}|S_n) = 1$ in Fig. 4 (bottom) add up to 9.0 %, which means that when the Chimgan station shows snow-free conditions, 9.0 % of the study area can be assigned as snow free.

Figure 4 shows that $P^s(S_{x,y}|S_n) = 1$ occurs mainly at high elevations (cf. Fig. 1), whilst $P^l(S_{x,y}|S_n) = 1$ occurs mainly at low elevations. This is not surprising since pixels with an elevation far higher than the elevation of Chimgan station tend to be snow covered if Chimgan station records positive snow depth, whilst pixels with an elevation far below the elevation of Chimgan station tend to be snow free if Chimgan station records snow depth of zero. Table 1 shows the SPI and LPI values for each station. The stations Kul' and Minchukur near the Zerafshan basin (see Fig. 1) have the highest SPI values (23.2% and 22.6%, respectively). Other stations, located further away from the catchment, have smaller SPI values. Noticeably, Oigaing station located farthest away from the Zerafshan basin has the highest LPI value (12.3 %). This can be explained with the high elevation of the station. When the station indicates snow free conditions, pixels with significantly lower elevation are likely to be snow-free as well. Assuming that the dependencies remain stable in time, the computed CPs of each pixel can be used to classify individual pixels for any arbitrary day prior to the availability of MODIS data (before 2000) for which station records are available:

$$S_{x,y,t} = 1 \qquad if \qquad \left(P^s(S_{x,y}|S_n) = 1 \; and \; S_{n,t} = 1\right) \tag{5}$$

$$S_{x,y,t} = 0 \qquad if \qquad \left(P^l(S_{x,y}|S_n) = 1 \; and \; S_{n,t} = 0\right) \tag{6}$$

This step leads to a partially reconstructed snow cover map which is further enhanced in the next steps.

4.2 Monthly probability fields

Snow cover extent is a seasonally variable parameter. Accordingly, the probability of a certain pixel to be covered by snow or land varies with time. The second step for reconstructing snow cover is based on the observation that during different months, certain pixels are snow covered or snow free with high confidence. The spatial distributions of such temporally persistent patterns can be identified using the available MODIS snow cover data in the period 2000-2012. A "*monthly probability*" (MP) of each pixel to be covered by snow or land in a certain month is computed according to:

$$MP^{s}_{x,y,m} = \frac{\sum(S_{(x,y,t),(t\in m)}=1)}{N_{x,y,m}} \quad (7)$$

$$MP^{l}_{x,y,m} = \frac{\sum(S_{(x,y,t),(t\in m)}=0)}{N_{x,y,m}} \quad (8)$$

where $MP^{s}_{x,y,m}$ and $MP^{l}_{x,y,m}$ are the probabilities of pixel x, y to be covered by snow or land in month m, respectively. $S_{(x,y,t)}$ indicates the coverage (snow or land) of pixel x, y on day t. $N_{x,y,m}$ is the total number of MODIS observations of pixel x, y and month m in the period 2000-2012. The maximum value of $MP^{s}_{x,y,m}$ ($MP^{l}_{x,y,m}$) is 1, meaning that the pixel x, y was covered always by snow (land) in month m during the cloud-free days in the period 2000-2012. Computation of MPs for every pixel in the study area leads to MP maps for all 12 months as illustrated exemplarily in Fig. 5 for April.

Pixels with $MP^{s}_{x,y,m} = 1$ in Fig. 5, i.e. pixels that were always snow covered in April, add up to 12.7 % of the whole area. This means that 12.7 % of the study area can be classified as snow covered in April. Pixels with $MP^{l}_{x,y,m} = 1$ add up to 14.9 %, i.e. 14.9 % of the domain can be classified as snow free in April. To remain consistent with the terminology used in the first step, we call the sum of pixels with $MP^{s}_{x,y,m} = 1$ ($MP^{l}_{x,y,m} = 1$) the *monthly SPI (LPI)* value for snow (land). Monthly *SPI and LPI* are defined in a similar way as in step 1 (Eq. 3 and 4).

The main idea in this step is to transfer these temporally persistent monthly spatial snow / land patterns (SPI_m / LPI_m) to the past to reconstruct historical snow cover. However, the validity of these temporally persistent spatial snow / land patterns over longer time in the past is not assured, due to e.g. potential warmer / cooler or wetter / dryer climate conditions. In order to account for possible climatic variability, we introduce a buffer as vertical elevation shift from month-specific minimum snow and maximum land lines. We define a month-specific minimum snow line ($H^{s}_{min,m}$) and a month-specific maximum land line ($H^{l}_{max,m}$) as:

$$H^s_{min,m} = \min\left(H_{x,y}\right) \quad \forall \quad MP^s_{x,y,m} = 1 \tag{9}$$

$$H^l_{max,m} = \max\left(H_{x,y}\right) \quad \forall \quad MP^l_{x,y,m} = 1 \tag{10}$$

where $H_{x,y}$ is the elevation of the pixel x, y. H^l_{max} is thus the maximum elevation of all pixels with $MP^l_{x,y,m} = 1$ in month m. Note that below the altitude H^l_{max} not all pixels necessarily have $MP^l_{x,y,m} = 1$. Similarly, H^s_{min} is the minimum elevation of all pixels with $MP^s_{x,y,m} = 1$ in month m. Again, H^s_{min} does not necessarily represent the elevation above which all pixels have $MP^s_{x,y,m} = 1$. Table 3 lists monthly SPI and LPI values, as well as $H^s_{min,m}$ and $H^l_{max,m}$ values for the selected area.

These $H^s_{min,m}$, $H^l_{max,m}$, and monthly SPI / LPI maps were used to further reconstruct the snow cover resulting from step 1:

$$S_{x,y,t} = 1 \; if \; \left(H_{x,y} > H^s_{min,m} + buffer \;\; and \;\; MP^s_{x,y,m} = 1\right) \qquad t \in m \tag{11}$$

$$S_{x,y,t} = 0 \; if \; \left(H_{x,y} < H^l_{max,m} - buffer \; and \;\; MP^l_{x,y,m} = 1\right) \qquad t \in m \tag{12}$$

where $buffer$ is a parameter accounting for the possible vertical shift in snow line. In order to account for climate variability not represented by the period for which satellite observations are available, "buffer" was set to 500 m. Due to absence of historical data on snow line variations in the region, the buffer was estimated corresponding to the maximum observed variation in the Equilibrium Line Altitude (ELA) of Abramov Glacier (Table 3, see Fig. 1for location) for the period 1972-1998 (WGMS 2001, Pertziger 1996), and is, thus, a conservative estimate for the variations in snow line for the study area.

4.3 Pixel to pixel conditional probability

In step 1, CPs of each pixel in accordance to station records were computed, and any pixel that had $P^s(S_{x,y}|S_n) = 1$ was classified according to the station record. The idea behind the third step is similar, but CPs of each pixel in accordance to other pixels are computed this time. In such a way, the state of different pixels is used as a predictor for snow cover elsewhere. We define the CP of any pixel with coordinates x, y to be covered by snow (land) given that another pixel i, j is covered by snow (land) as follows:

$$P^s(S_{i,j}|S_{x,y}) = \frac{\sum(1 - ABS(S_{i,j,t} - S_{x,y,t}))}{N_{i,j}} \quad \forall \quad S_{x,y,t} = 1 \tag{13}$$

$$P^l(S_{i,j}|S_{x,y}) = \frac{\sum(1-ABS(S_{i,j,t}-S_{x,y,t}))}{N_{i,j}} \quad \forall \quad S_{x,y,t} = 0 \qquad (14)$$

where $S_{i,j,t}$ indicates whether the pixel with coordinates i,j is snow covered ($S = 1$) or snow free ($S = 0$) for a given day t, and $N_{i,j}$ is the total number of valid observations (clear sky, no cloud) at pixel i,j simultaneously available for a given condition ($S = 1\ or\ S = 0$) in the period 2000-2012.

The computation of $P^s(S_{i,j}|S_{x,y})$ and $P^l(S_{i,j}|S_{x,y})$ according to Eq. (13) and (14) is repeated in an "all-versus-all" procedure, which means that all possible combinations of (x,y) and (i,j) are considered. For the region of interest, this yields at maximum 339 552 (2 x total number of pixels) CP maps (i.e. two maps for every pixel: for snow and land condition). However, not all of these maps were used for snow reconstruction since some pixels may have no perfect dependence (no pixels with CP=1) to any other pixel in the study area. An example of the CP maps for snow and snow free conditions for the pixel located at x =100 and y =100 is given in Fig. 6.

The pixel with coordinates x =100, y =100 has $P^s(S_{i,j}|S_{100,100}) = 1$ for 30 684 (18 %) other pixels in the study area, which means that when the particular pixel was snow covered during 2000-2012, 30 684 other pixels were always snow covered as well. Thus, is the SPI value of that pixel is 18%, and this can be interpreted as the predictive power for snow of the pixel for the entire study area. Analogously, 23.770 pixels (14 %) have $P^l(S_{i,j}|S_{100,100}) = 1$ with this pixel and the LPI value of this pixel is 14%. The SPI and LPI values of each pixel are derived through Eq. (3) and (4), and illustrated in Fig. 7 for all pixels in the study area.

The maximum value of Fig. 7 (top) is 46%, meaning that, according to the observations of the period 2000-2012, 46% (78 189 pixels) of the study area was always snow covered when that particular pixel was snow covered. The maximum LPI value (Fig. 7 bottom) is 88 %, meaning that this particular pixel is able to predict snow free conditions for 88% (149 685 pixels) of the basin. These two pixels with maximum SPI and LPI values are located within an area which has high predictive power for snow and land, respectively. When interpreting Fig. 7, three further features are worth notice: (1) pixels with SPI=0 or LPI=0 exist as well; these pixels have no predictive power and are therefore not used in the snow cover reconstruction; (2) SPI and LPI maps generally reflect the topography of the catchment: lower elevation pixels have higher SPI values and pixels at higher elevations have higher LPI values; (3) snow free pixels are easier to predict than snow covered ones.

The SPI and LPI maps were used for classifying pixels that are still undefined after the previous steps:

$$S_{i,j,t} = 1 \quad if \quad \left(P^s(S_{i,j}|S_{x,y}) = 1 \; and \; S_{x,y,t} = 1\right) \tag{15}$$

$$S_{i,j,t} = 0 \quad if \quad \left(P^s(S_{i,j}|S_{x,y}) = 1 \; and \; S_{x,y,t} = 0\right) \tag{16}$$

Since in this step SPI and LPI maps were generated for every pixel in the basin, this step tends to classify a significantly larger area than the first step where only 7 stations were used for constructing CPs.

4.4 Snow cover estimation using elevation information from neighbouring pixels

This step is adapted from Gafurov et al. (2009) and is based on the information of neighbouring pixels. Let us consider a pixel that has not been classified as snow covered or snow free in any of the previous steps. If any of the adjacent eight pixels is covered by snow and the elevation of that snow-covered pixel is lower than the pixel that is still undefined, then the undefined pixel is classified as snow covered. The same idea is applied for snow-free pixels. Hence, this step can be formalized as follows:

$$S_{x,y,t} = 1 \quad if \quad S_{x+k,y+k,t} = 1 \quad and \;\; H_{x+k,y+k} < H_{x,y} \qquad k \in (-1,1) \tag{17}$$

$$S_{x,y,t} = 0 \quad if \quad S_{x+k,y+k.t} = 0 \quad and \;\; H_{x+k,y+k} > H_{x,y} \qquad k \in (-1,1) \tag{18}$$

This step takes only elevation of neighbouring pixel into account. However, in areas where factors others than elevation have an influence on neighbouring pixel condition (e.g. pixels located near to water surfaces or forests), additional information such as land cover map could be introduced into this step.

4.5 Snow cover estimation with CP<1

In the last step, the $P^s(S_{x,y}|S_n)$ and $P^l(S_{x,y}|S_n)$ values calculated in step 1 are used again. Whereas in step 1 only CP$= 1$ conditions were considered, this step considers $CP < 1$ relations (cf. Fig 4) to classify still undefined pixels. Still undefined pixels are classified according to the

highest CP value found amongst all 14 CP values available (7 CPs for snow and 7 CPs for land). However, since the stations are located at different elevations with varying snow covered (snow free) days, their computed CPs can have varying levels of confidence depending on the number of observations used to construct probabilities. To account for this in the reconstruction, we compute lower bound confidence intervals (CI) of each CP. As the CP estimates follow a binomial distribution, we compute lower bound CI of CP according to:

$$CI^{low}_{(P^s(S_{x,y}|S_n))} = P^s(S_{x,y}|S_n) - z\sqrt{\frac{1}{N_{x,y}}P^s(S_{x,y}|S_n)(1-P^s(S_{x,y}|S_n))} \quad (19)$$

$$CI^{low}_{(P^l(S_{x,y}|S_n))} = P^l(S_{x,y}|S_n) - z\sqrt{\frac{1}{N_{x,y}}P^l(S_{x,y}|S_n)(1-P^l(S_{x,y}|S_n))} \quad (20)$$

where $CI^{low}_{(P^s(S_{x,y}|S_n))}$ and $CI^{low}_{(P^l(S_{x,y}|S_n))}$ are lower bound of 95 % CI of CP for snow and land conditions, respectively. z is the constant for 95 % confidence level (1.96). The computed lower bound CI for each CP will help to classify still undefined pixel coverage based on highest confidence level for $P(S_{x,y}|S_n) < 1$ case among 14 stations. Thus, we use highest lower bound CI among all CPs of this particular pixel to be decisive for reconstruction.

$$S_{x,y,t} = 1 \quad if \quad \max(CI^{low}_{(P^s(S_{x,y}|S_n))}) > \max(CI^{low}_{(P^l(S_{x,y}|S_n))}) \qquad n \in 1{:}7 \quad (21)$$

$$S_{x,y,t} = 0 \quad if \quad \max(CI^{low}_{(P^l(S_{x,y}|S_n))}) > \max(CI^{low}_{(P^s(S_{x,y}|S_n))}) \qquad n \in 1{:}7 \quad (22)$$

Taking maximum lower bound CI values for still undefined pixels in the last step allows to complete the classification for all pixels. However, since in this step $P(S_{x,y}|S_n)$<1 was considered, the reconstruction is subject to uncertainty that stems from non-perfect agreement between station records and a pixel in the period 2000-2012.

5 Results and discussion

Applying the five steps described above, snow cover maps for the area containing Zerafshan basin were reconstructed for four days in 1998 and 1999. The maps contain binary information showing whether a given pixel was covered by snow or not. The accuracy of the reconstructed snow cover maps was assessed by comparing against independent snow maps derived from four Landsat images from the same days. The validation could be done only for recent years (1998 and 1999) due to availability of cloud free Landsat images during the snow period only for

those days. The comparison was performed on a pixel-to-pixel basis, and the accuracy was assessed in a contingency table (Table 4). In Table 4, the sum of percentages of "SS" and "LL" columns represent the degree of accuracy after each reconstruction step, related to the total share of reconstructed pixels. Accordingly, the sum of "SL" and "LS" indicate the error in relation to the total percentage.

As an example, Fig. 8 shows the reconstructed and Landsat-derived snow cover maps for April 10, 1998. The comparison of these maps results in 85.7 % of correct reconstruction (cases SS+LL in Table 4) and 14.3 % of erroneous reconstruction (SL+LS). Steps 1-4 show high accuracy with only little erroneous reconstruction (ER) whereas step 5 has the lowest accuracy in all validation days. However, the reconstruction fraction (RF) is very high in step 5 compared to previous steps. Note that ER may also be enhanced by erroneous snow cover estimation from raw Landsat data and due to the spatial aggregation of Landsat 30 m original resolution to 500 m. Another potential bias may come from similar approaches (NDSI) used to map snow cover both for MODIS snow cover maps which are used to assess CPs between station and pixels and Landsat snow cover maps which are used to validate reconstructed snow cover maps. However, different threshold values than MODIS were used to map snow cover from Landsat assuring best visual validation of snow and snow free surface cover.

In order to better illustrate snow reconstruction in step 5, Fig. 9 shows the areal fraction for which the reconstruction was performed in steps 1-4 and maximum lower bound CI obtained in step 5 under CP<1 condition for the validation day of April, 10. Most of the still unclassified pixels after steps1-4 have CI values close to 1 and only few pixels have a lower CI value (reddish and light blue colours in Fig. 9). Figure 10 illustrates the trade-off between RF and ER as a function of lower bound CI in step 5. For example, for the validation day of April 10, ER from steps 1-4 adds up to 1.5 % (Table 4) and RF to 51.4 % (snow and land classes in Fig. 9). With decreasing CI, RF increases, but at the cost of an increased ER. However, Fig. 10 also shows that RF is relatively high until about CI=0.9 with increasing ER. In all four days used for validation, an almost complete reconstruction is achieved with CI>0.9.

Figures 8, 9 and 10 also demonstrate that the methodology provides two types of results for snow cover reconstruction: deterministic and probabilistic snow cover maps. Deterministic maps result from the complete classification of pixels (Fig. 8) with binary information (snow / snow free) taking CI<1 in step 5 into account at the expense of the overall accuracy. However, the accuracy is still quite high with a range of 83.3-85.7 % for the four validation days and is only slightly less than the accuracy of MODIS snow cover product in Central Asia (ca. 92%)

when compared to Landsat snow information (Gafurov et al., 2013). Alternatively, probabilistic snow cover maps (Fig. 9) deliver a partial snow cover reconstruction with high accuracy resulting from steps 1-4, and, as result from step 5, a probability statement for snow cover for the remaining pixels.

The validation days for this study were chosen deliberately from snow melt and snow accumulation season (transition period) where snow cover estimation is particularly challenging. For the time outside the snow melt or snow accumulation period, higher accuracies can be expected since a higher fraction can be reconstructed in the first 4 steps already. During snow transition period, snow cover conditions such as ephemeral snow cover can occur which exacerbates snow cover estimation. However, in the reconstruction process using steps 1, 2 and 3 the conditions with ephemeral snow cover (in the period 2000-2012) are accounted as well. Under such conditions, station or MODIS data may see different snow cover than the reality. In such cases (e.g. MODIS sees "land" although there is ephemeral snow whilst the station sees "snow" since it is a manual point recording with a certain threshold), CP and MP of pixel gets the value of < 1 as they do not show the same event and is not used in the first three steps for reconstruction. Only distinct snow cover records from both station and MODIS are used to identify snow covered areas in these step with CP=1 (MP=1). Reduced CP values that may partly be due to ephemeral snow cover are, however, used in step 5 in order to classify areas still undefined in the steps 1-4 and may thus contribute to the accuracy loss in step 5.

The validation of reconstructed snow cover maps were done using independent Landsat data in this study. Alternatively, the AVHRR snow cover data, which is also available beyond the MODIS data availability in the past, can be used for validation purposes. However, AVHRR snow cover data has a coarser spatial resolution (~ 1.1 km) than the resolution (500 m) used in this study. Unfortunately, processed AVHRR snow cover data was not available at the time of manuscript writing and remains alternative data to be used for validation.

6 Limitations of the methodology

The predictive power of the observations at meteorological stations for snow cover reconstruction is limited by the elevation range of the stations. If all meteorological stations are located at high elevations, they will be good predictors during summer for snow-free conditions, but will perform poorly when predicting snow-covered areas during winter due to their elevation and correspondingly lower SPI values. Conversely, low-elevation stations are better

indicators for snow-covered pixels at higher altitudes than they are for snow-free ones. Hence, a wide spread in station elevation is optimal for accurate snow cover reconstruction. In our case study, the application of the presented methodology suffered from the small number of station data (only 7 stations). A higher number of stations would lead to a higher number of SPI and LPI maps and would allow to reconstruct a larger areal fraction of snow cover in the first four steps with high accuracy. Noticeably, the stations do not need to be located inside the area of interest.

Reconstruction of the snow cover for the past is based on the assumptions that (1) the calibration period, i.e. the MODIS data period, is representative for the past period, and (2) the relationship between station records and spatial snow patterns derived from MODIS data is stationary, i.e does not significantly change in time. A calibration period which lacks extreme conditions, e.g. snow-rich or snow-scarce years, might lead to larger errors in the reconstruction. A longer calibration period is expected to lead to more robust relationships for reconstructing snow cover.

The problem of representativity of the MODIS period in the reconstruction step 2 is tackled by the introduction of the elevation buffer to capture the effect of inter-annual temporal variability of snow line elevation. For this the temporal variability of the recorded ELA from the neighbouring Abramov glacier was used as a proxy. Through changes in climatic conditions of the calibration period going beyond temporal variability of the snow line elevation in the reconstruction period, the relationships between station records and some pixels (step 1) and between pixels (step 3) may become non-representative. This occurs if snow line in the future/calibration period more often separates the station of the pixels compared to the reconstruction period. Hence, an analysis of temperature and precipitation trends and comparison of climatology between calibration and reconstruction periods may provide some confidence on representativeness of the relationships used.

The statistical relationship (CP) between point measurements and aereal patterns computed in this study highly depend on topography. Since Zerafshan basin has a very heterogeneous topography with high elevation range, good predicitve power (SPI and LPI) of individial stations could be obtained. This is important to estimate initial snow cover in the first step which is a base input for next steps (except step 2). Thus, we can conclude that the methodology is well applicable for mountainous areas where high SPI and LPI values can be obtained. However, it might be difficult to exploit statistical relationship between point measurements and aereal pattern in lowland areas and is a subject to be tested.

7 Conclusion

In this study, a methodology for reconstructing past snow cover using historical in-situ snow depth data, recent remote sensing snow cover data and topographic data was presented. The methodology is based on (1) constructing relationships between station observations and remote sensing data, (2) estimating the monthly variation of snow cover from remote sensing data, (3) deriving pixel-to-pixel relationships using remote sensing data, and (4) using neighbourhood relations. Once the dependence between individual pixels and station records is derived, this dependence is used to reconstruct past snow cover based solely on station records.

The methodology was applied to a study area containing the Zerafshan River basin – a basin with high topographic gradients – in Central Asia and showed correct classification in the range of 83.3 to 85.7 % when compared to four Landsat snow cover scenes. This high agreement is noteworthy, given that only 7 stations and 12 years of remote sensing data were available. Moreover, snow cover reconstruction was done for snowmelt and onset season when snow classification is challenging compared to outside snowmelt and onset seasons where large areas are easy to classify as snow or land. The agreement is only slightly less than that of original MODIS snow cover product with accuracy of about 92% for Central Asia when compared to Landsat (Gafurov et al., 2013). Just 12 years of MODIS data were sufficient to extract stable patterns of snow cover and relate them to station records in the Zerafshan basin with heterogeneous topography. Hence, we conclude that the developed methodology is suitable to derive past snow cover in remote mountainous regions such as the Zerafshan basin with very limited data availability. Reconstructed snow cover patterns can be used for hydrological model calibration / validation and for understanding snow cover dynamics over large areas prior the age of satellite observations. The performance of here presented methodology for non-mountainous areas remains open question.

Acknowledgement

This work was carried out in the frame of the CAWa (Water in Central Asia) project (http://www.cawa-project.net, Contract No. AA7090002), funded by the German Federal Foreign Office as part of the "Berlin Process". Doris Duethmann was supported by the SuMaRiO project (Sustainable Management of River Oases along the Tarim River/China), funded by the BMBF (German Ministry for Education and Research) — Funding Measure ‘‘Sustainable Land Management’’, reference number: LLA2-02.

References

Aizen, V., Aizen, E., and Melack, J.: Climate snow cover, glaciers and runoff in the Tien Shan, Central Asia, Water Resour. Bull., 31(6), 1113-1129, 1995

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

Brown, R.: Northern Hemisphere Snow Cover Variability and Change, 1915-97, J. Climate, 13, 2339-2355, 2000

Brown, R. and Robinson D.: Northern Hemisphere spring snow cover variability and change over 1922–2010 including an assessment of uncertainty. The Cryosphere, 5, 219–229, doi: 10.5194/tc-5-219-2011, 2011

Corbari, C., Ravazzani, G., Martinelli, J. and Mancini, M.: Elevation based correction of snow coverage retrieved from satellite images to improve model calibration, Hydrol. Earth Syst. Sci., 13, 639-649, 2009

Duethmann, D., Peters, J., Blume, T., Vorogushyn, S. and Güntner, A.: The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia, Water Resour. Res., 50, 2002-2021, doi: 10.1002/2013WR014382, 2014

Durand, M., Molotch, N. and Margulis, A.: A Bayesian approach to snow water equivalent reconstruction, J. Geoohys. Res., 113, D20117, doi: 10.1029/2008JD009894, 2008

Erickson, T., Williams, M. and Winstral, A.: Persistence of topographic controls on the spatial distribution of snow in rugged mountain terrain, Colorado, United States, Water Resour. Res., 41, W04014, doi: 10.1029/2003WR002973, 2005

Finger, D., Pellicciotti, F., Konz, M., Rimkus, S. and Burlando, P.: The value of glacier mass balance, satellite snow cover images and hourly discharge for improving the performance of a physically based distributed hydrological model, Water Resour. Res., 47, W07519, doi: 10.1029/2010WR009824, 2011

Frei, A., Robinson, D. and Hughes, M.: North American Snow Extent: 1900-1994, Int. J. Climatol., 19, 1517-1534, 1999

Gafurov, A. and Bárdossy, A.: Cloud removal methodology from MODIS snow cover products, Hydrol. Earth Syst. Sc., 13, 1-13, 2009

Gafurov, A., Kriegel, D., Vorogushyn, S. and Merz, B.: Evaluation of remotely sensed snow cover product in Central Asia, Hydrol. Res., 44.3, 506-522, doi: 10.2166/nh.2012.094, 2013

Hall, D., Riggs, G. and Salomonson, V.: Development of methods for mapping global snow cover using Moderate Resolution Imaging Spectroradiometer (MODIS) data, Remote Sens. Environ., 83, 181-194, 2002a

Hall, D., Riggs, G., Foster, J. and Kumar, S.: Development and evaluation of a cloud-gap-filled MODIS daily snow-cover product, Remote Sens. Environ., 114, 496-503, 2010

Huang, X., Liang, T., Zhang, X. and Guo, Z.: Validation of MODIS snow cover products using Landsat and ground measurements during the 2001-2005 snow seasons over northern Xinjiang, China, Int. J. Remote Sens., 32:1, 133-52, 2011

Immerzeel, W., Droogers, P., Jong, S. and Bierkens, M.: Large-scale monitoring of snow cover and runoff simulation in Himalayan river basins using remote sensing, Remote Sens. Environ., 113, 40-49, 2008

Jarvis, A., Reuter, H., Nelson, A. and Guevara, E.: Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org), 2008

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

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 Sens. Environ., 112, 3750-3761, 2008

Li, X. and Williams, M.: Snowmelt runoff modeling in an arid mountain watershed, Tarim Basin, China, Hydrol. Process., 22, 3931-3940, 2008

Liu, T., Willems, P., Feng, X., Li, Q., Huang, Y., Bao, A., Chen, X., Veroustraete, F. and Dong, Q.: On the usefulness of remote sensing input data for spatially distributed hydrological modeling: case study of the Tarim River basin in China, Hydrol. Process., 26, 335-344, 2012

López-Burgos, V., Gupta, H. V. and Clark, M.: Reducing cloud obscuration of MODIS snow cover area products by combining spatio-temporal techniques with a probability of snow approach, Hydrol. Earth Syst. Sci., 17, 1809-1823, doi: 10.5194/hess-17-1809-2013, 2013

NASA Landsat Program, Landsat TM and ETM+, USGS, Sioux Falls, 2003

NASA Land Processes Distributed Active Archive Center (LP DAAC). Land Cover Type Yearly L3 Global 500 m SIN Grid (MCD12Q1). USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, 2001.

Ososkova, T., Gorelkin, N. and Chub, V.: Water resources of Central Asia and adaptation measures for climate change, Environ. Monit. Assess., 61, 161-166, 2000

Parajka, J. and Blöschl, G.: Validation of MODIS snow cover images over Austria. Hydrol. Earth Syst. Sc., 10 (5), 679-689, 2006

Parajka, J. and Blöschl, G.: The value of MODIS snow cover data in validating and calibrating conceptual hydrological models, J. Hydrol., 358, 240-258, 2008

Parajka, J. and Blöschl, G.: Spatio-temporal combination of MODIS images – potential for snow cover mapping, Water Resour. Res., 44, 1-13, 2008

Parajka, J., Holko, L., Kostka, Z. and Blöschl, G.: MODIS snow cover mapping accuracy in a small mountain catchment – comparison between open and forest sites, Hydrol. Earth Syst. Sc., 16, 2365-2377, 2012

Pellicciotti, F., Buegrl, C., Immerzeel, W., Konz, M. and Shrestha, A.: Challenges and uncertainties in hydrological modeling of remote Hindu-Kush-Karakoram-Himalayan (HKH) Basins: Suggestions for calibration strategies, Mt. Res. Dev., 39, 39-50, 2012

Pertziger, F.: Abramov Glacier Data Reference Book: Climate, Runoff, Mass Balance, Technical University Munich, Tashkent/Munich, 1996.

Robinson, D.: Merging operational satellite and historical station snow cover data to monitor climate change, Palaeogeogr. Palaeocl. (Global Planet. Change Sect.), 90, 235-240, 1991

Tekeli, A., Akyurek, Z., Sorman, A., Sensoy, A. and Sorman, A.: Using MODIS snow cover maps in modeling snowmelt runoff process in the eastern part of Turkey, Remote Sens. Environ., 97 (2), 216-230, 2005

Unger-Shayesteh, K., Vorogushyn, S., Farinotti, D., Gafurov, A., Duethmann, D., Mandychev, A. and Merz, B.: What do we know about past changes in the water cycle of Central Asian headwaters? A review, Global Planet. Change, 110, 4–25, doi: 10.1016/j.gloplacha.2013.02.004, 2013

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 (4), 1497-1513, 2008

Tong, J., Dèry, S. and Jackson, P.: Interrelationships between MODIS/Terra remotely sensed snow cover and the hydrometeorology of the Quesnel River Basin, British Columbia, Canada, Hydrol. Earth Syst. Sci., 13, 1439-1452, 2009

Wang, J., Li, H. and Hao, X.: Responses of snowmelt runoff to climate change in an inland river basin, Norwestern China, over the past 50 years, Hydrol. Earth Syst. Sci., 14, 1979-1987, 2010

WGMS, Glacier Mass Balance Bulletin No. 6 (1998-1999). Haeberli, W., Frauenfelder, R., and Hoelzle, M. (eds.), IAHS (ICSI)/ UNEP/UNESCO, World Glacier Monitoring Service, Zurich. Available at: http://www.wgms.ch/mbb/mbb6/wgms_2001_gmbb6.pdf, last access: 2013-01-04, 2001

Zhou, H., Aizen, E. and Aizen, V.: Deriving long term snow cover extent dataset from AVHRR and MODIS data: Central Asia case study, Remote Sens. Environ., 136, 146-162, doi:10.1016/j.rse.2013.04.015, 2013

Table 1. Uzhydromet snow observation stations with indication of elevation (Elev), and SPI (Snow Predictability Index) and LPI (Land Predictability Index) values for the study area (see section methodology for definitions). The entries “records on snow reconstucted days” indicate whether a station was snow covered (0) or snow free (1) during a day for which snow cover reconstruction was conducted and Landsat scenes were available for validation.

Name	Elev (m.a.s.l)	SPI (%)	LPI (%)	Records on snow reconstructed days			
				Apr 10, 1998	Nov 20, 1998	Apr 29, 1999	Nov 15, 1999
Dukant	1984	12.4	3.7	1	0	0	1
Kamchik	2145	10.0	9.1	1	0	0	1
Kul'	2161	23.2	6.5	0	0	0	1
Minchukur	2136	22.6	4.2	0	0	0	0
Oigaing	2187	2.2	12.3	1	1	1	1
Pskem	1256	2.9	0.8	1	0	0	1
Chimgan	1677	10.2	9.0	1	0	0	1

Table 2. Monthly SPI, LPI, $H^s_{min,m}$, and $H^l_{max,m}$ values for the study region.

Month	*SPI*		*LPI*	
	Fraction (%)	H^s_{min}	Fraction (%)	H^l_{max}
1	10.1	1544	0.0	658
2	11.4	1164	0.0	2233
3	24.0	1658	0.4	2923
4	12.7	2354	14.9	3323
5	1.1	2900	32.7	3707
6	0.0	3568	49.5	4387
7	0.0	5402	66.3	4434
8	0.0	5402	78.3	4520
9	0.0	4285	51.2	4330
10	0.2	3405	25.3	4089
11	0.9	2750	0.4	2581
12	8.5	2373	0.0	1707

Table 3. ELA records of Abramov Glacier (WGMS 2001, Pertziger 1996)

Year	1972	1977	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
ELA (m.a.s.l)	4020	4393	4130	4170	4200	4220	4242	4110	4120	4250	4240	4163	4440	4130

Table 4. Contingency table (in %) for the reconstructed snow cover maps validated against four aggregated Landsat snow cover images. Four cases are distinguished: SS, LL, SL, and LS. The first (second) letter indicates the classification according to the presented algorithm (Landsat). "S" stands for "snow", "L" for "land". "Total" indicates the percentage of pixels classified after each step. Results refer to the Landsat domain (dashed line) shown in Fig. 1b.

Day	Step	SS	LL	SL	LS	Total
Apr 10, 1998	1	8.5	7.2	0.0	0.0	15.7
	2	13.0	14.9	0.1	0.1	28.1
	3	17.5	27.1	0.8	0.1	45.5
	4	20.6	29.3	1.3	0.2	51.4
	5	43.4	42.3	12.5	1.8	100.0
Nov 20, 1998	1	0.1	11.2	0.0	0.0	11.3
	2	0.2	11.3	0.0	0.0	11.5
	3	0.5	30.5	0.0	0.0	31.0
	4	1.4	33.4	0.0	0.1	34.9
	5	18.1	67.0	2.2	12.7	100.0
Apr 29, 1999	1	0.1	13.9	0.0	0.0	14.0
	2	7.6	21.5	0.0	0.5	29.6
	3	9.0	35.2	0.1	0.7	45.0
	4	11.0	38.1	0.3	0.9	50.3
	5	24.5	58.8	14.3	2.3	100.0
Nov 15, 1999	1	18.4	4.1	0.0	0.4	22.9
	2	18.4	4.2	0.0	0.4	23.0
	3	24.8	15.7	0.4	0.8	41.7
	4	28.5	18.3	0.7	1.3	48.9
	5	41.7	42.0	11.2	5.1	100.0

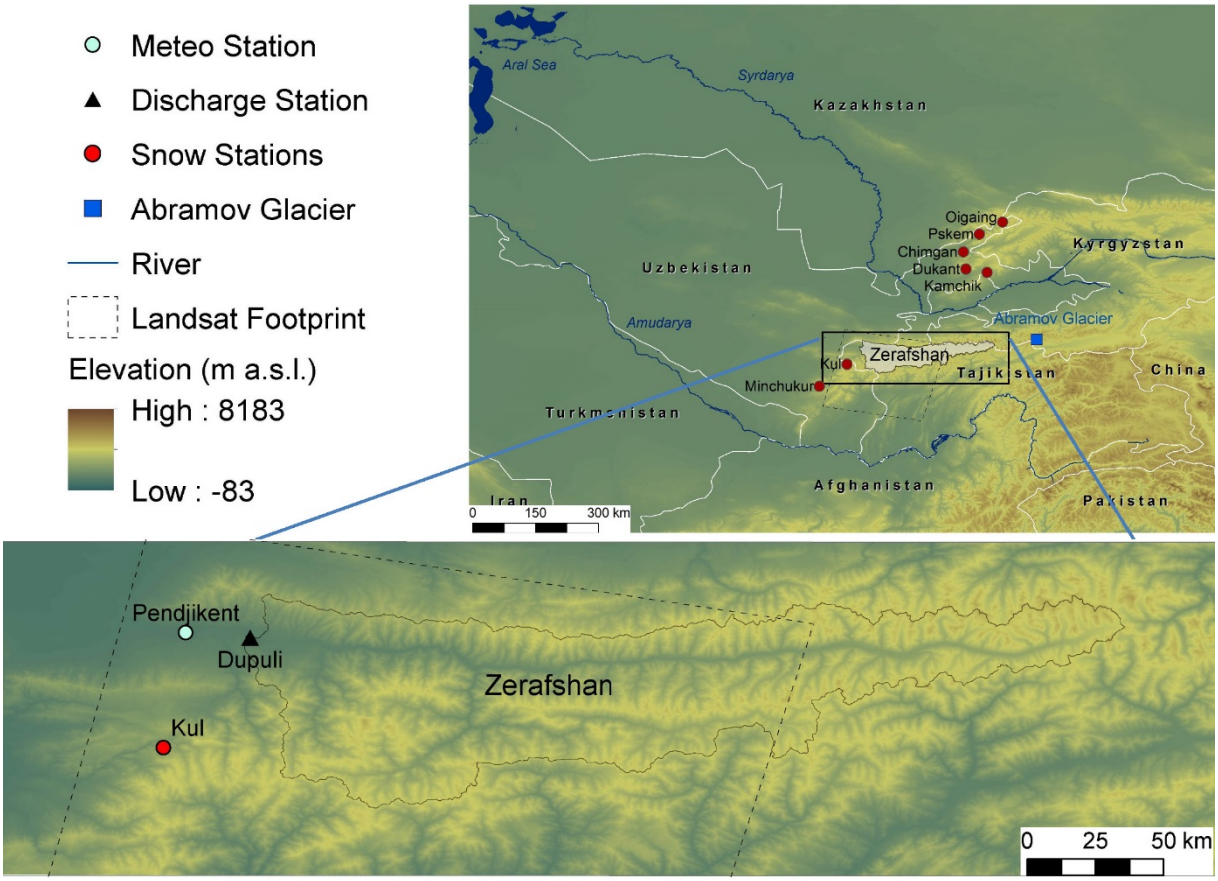


Figure 1. Location of the Upper Zerafshan River basin in the Gissaro-Alai Mountain Range, Central Asia. Snow cover reconstruction was conducted for the entire area of Fig. 1b and validated for the area with Landsat footprint.

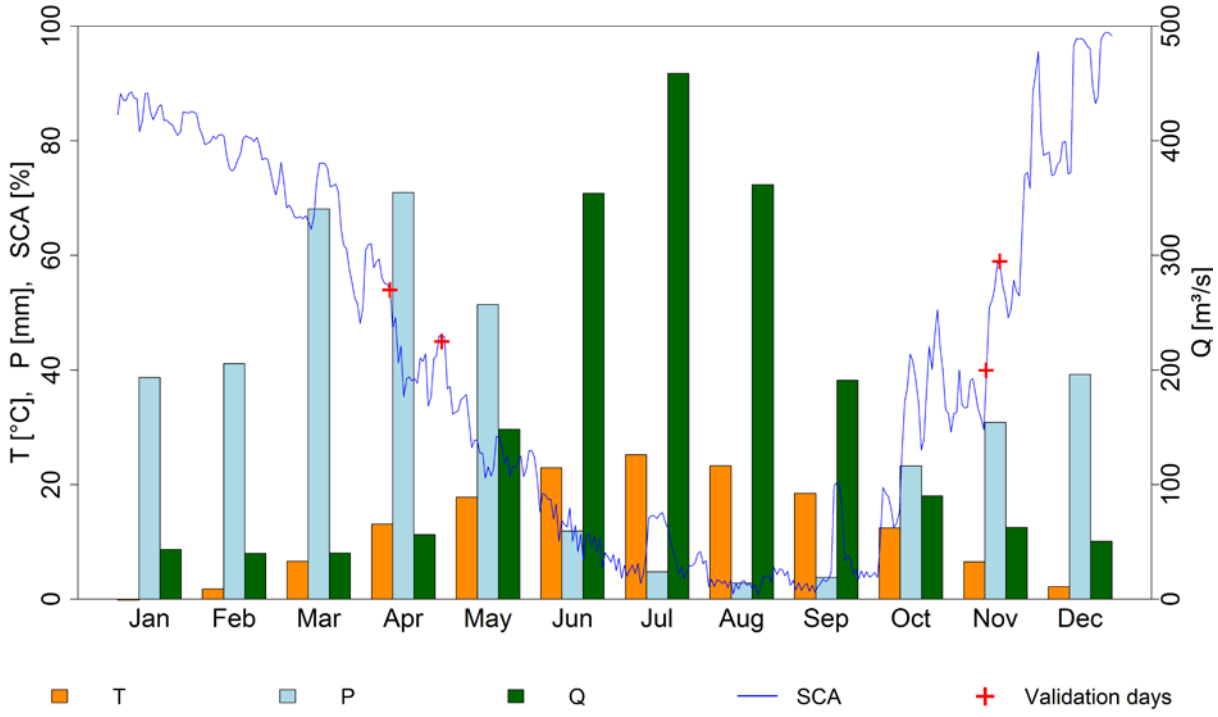


Figure 2. Monthly average air temperature (T), cumulative precipitation (P), discharge (Q) and SCA dynamics in the Zerafshan basin. T, P and Q means are based on data for the period from 1930 till 2008. Daily SCA is for 2004 obtained from MODIS and cloud eliminated using Gafurov and Bárdossy (2009). Temperature (zero in January) and precipitation data are from the Pendjikent station (1016 m a.s.l.), discharge is measured at the Dupuli gauge (see Fig. 1).

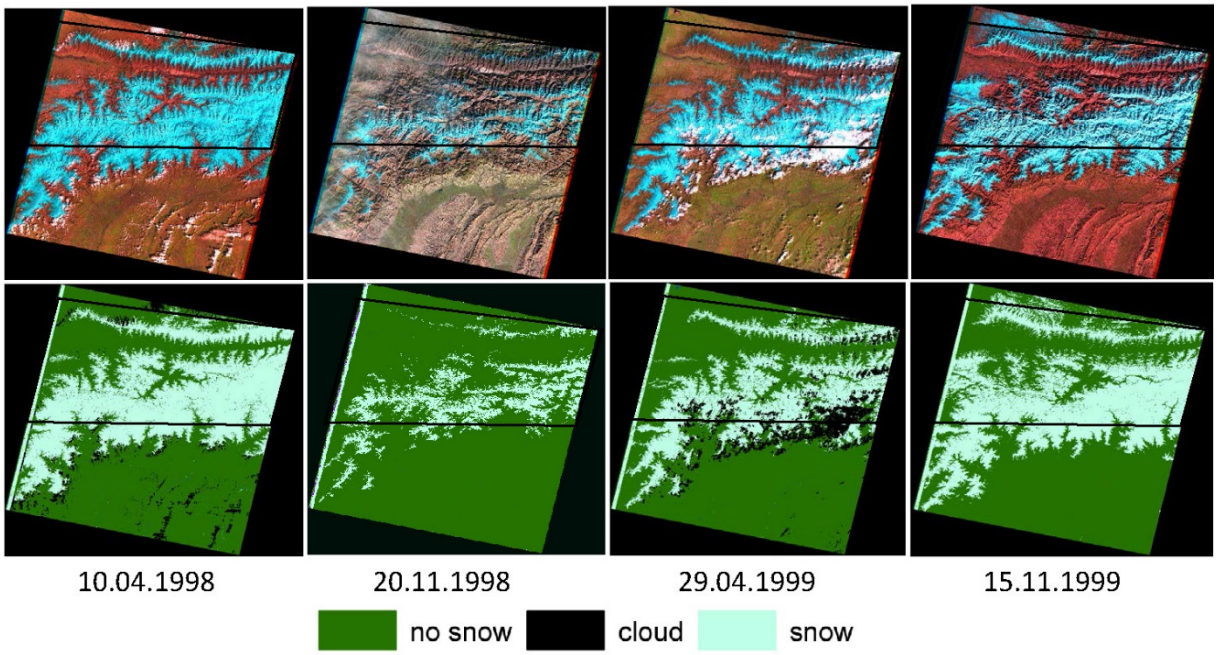


Figure 3. Original Landsat scenes (top row) and derived snow cover maps used for validation (bottom row). Black outlines show the validation domain for Zerafshan basin.

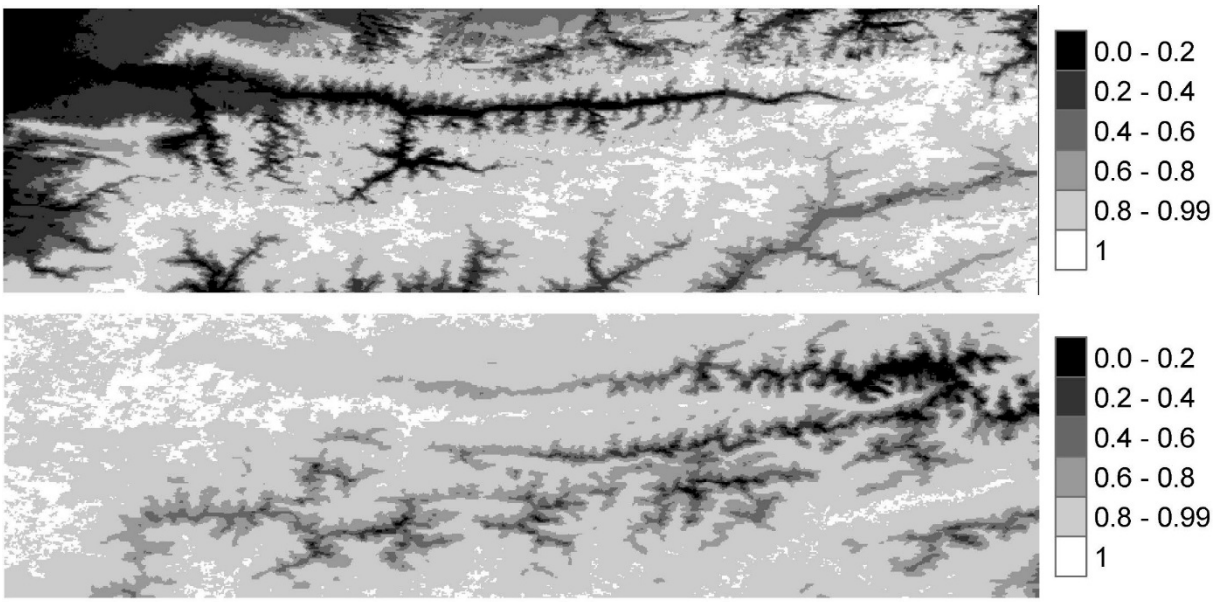


Figure 4. CP maps for snow (top) and land (bottom) conditions of Chimgan station (see Fig. 1) for the study area. The figure shows the same domain as Figure 1b

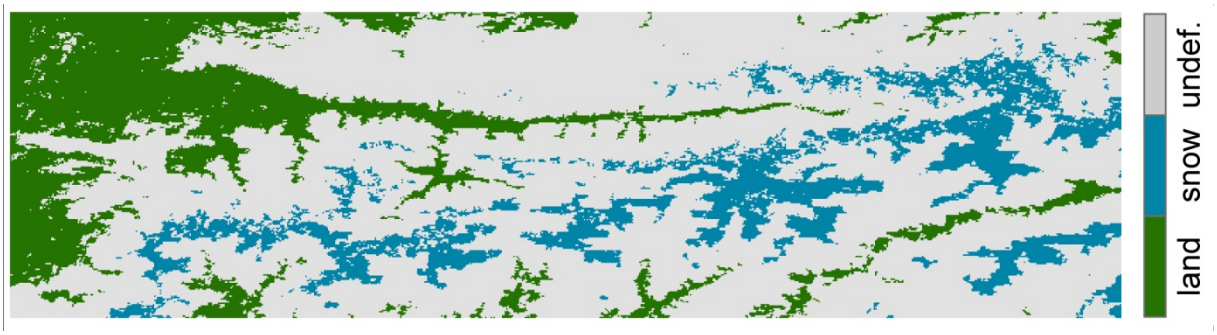


Figure 5. Temporally persistent spatial snow ($MP^s = 1$) and land ($MP^l = 1$) patterns for April in the study area shown in Fig. 1b. Pixels with "undef" indicate $MP^s < 1$ or $MP^l < 1$, for which classification is not possible in this step.

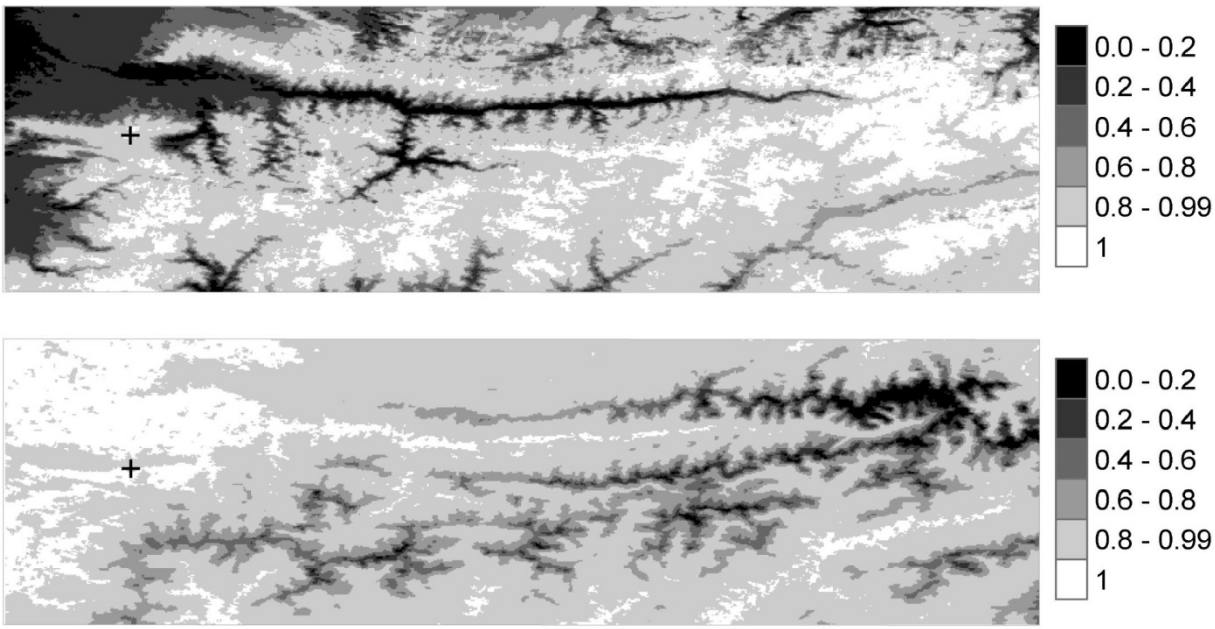


Figure 6. CP maps for snow (top) and land (bottom) conditions for the pixel $x = 100$, $y = 100$ (black cross) with elevation 2206 m.a.s.l.

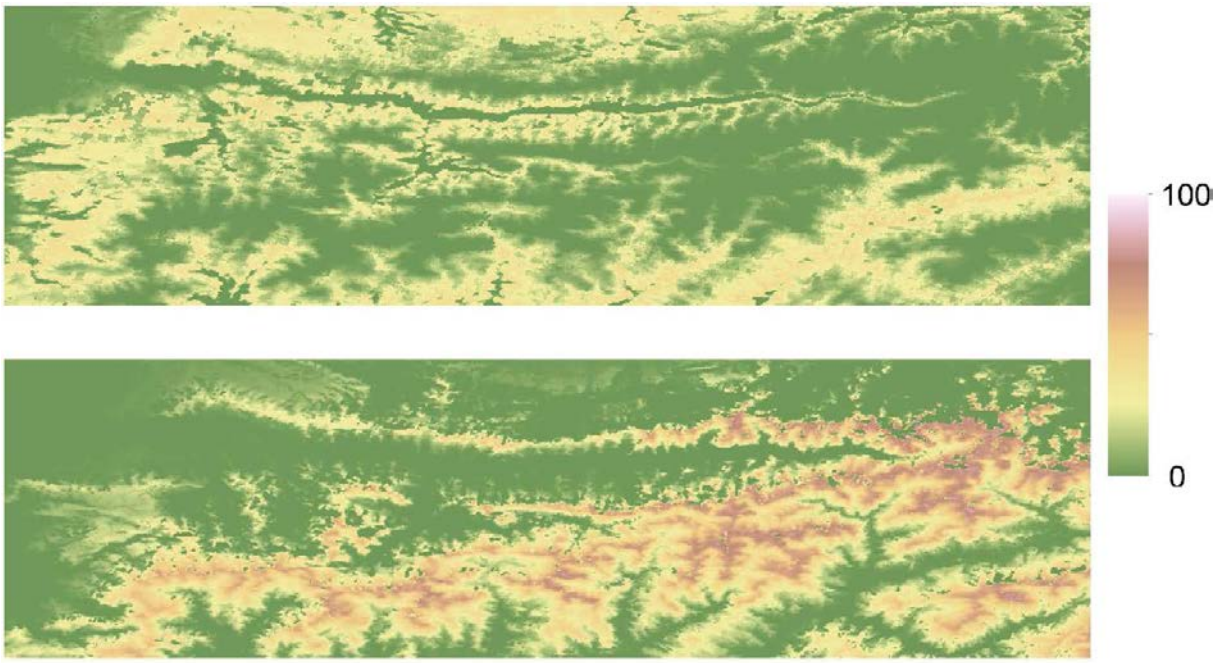


Figure 7. *SPI* (top) and *LPI* (bottom) values of each pixel (in %) in the study area defined in Fig. 1b.

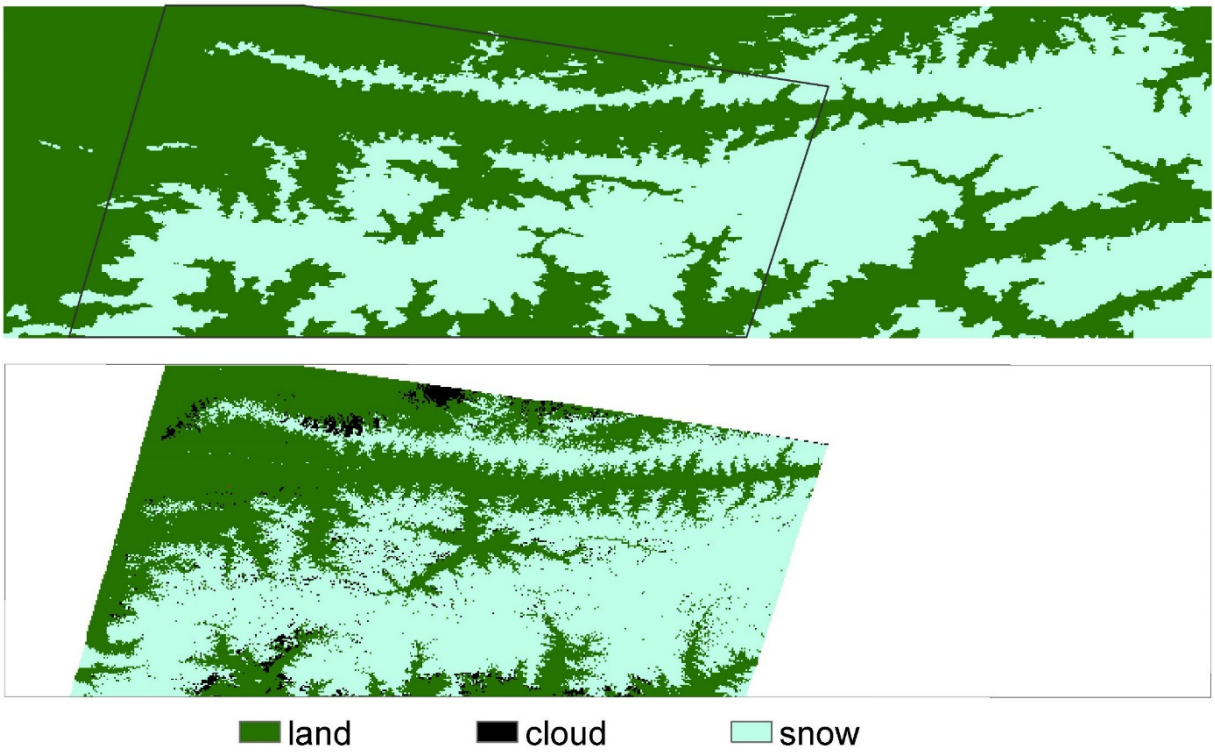


Figure 8. Reconstructed (top) and Landsat (bottom) snow cover maps for April 10, 1998.

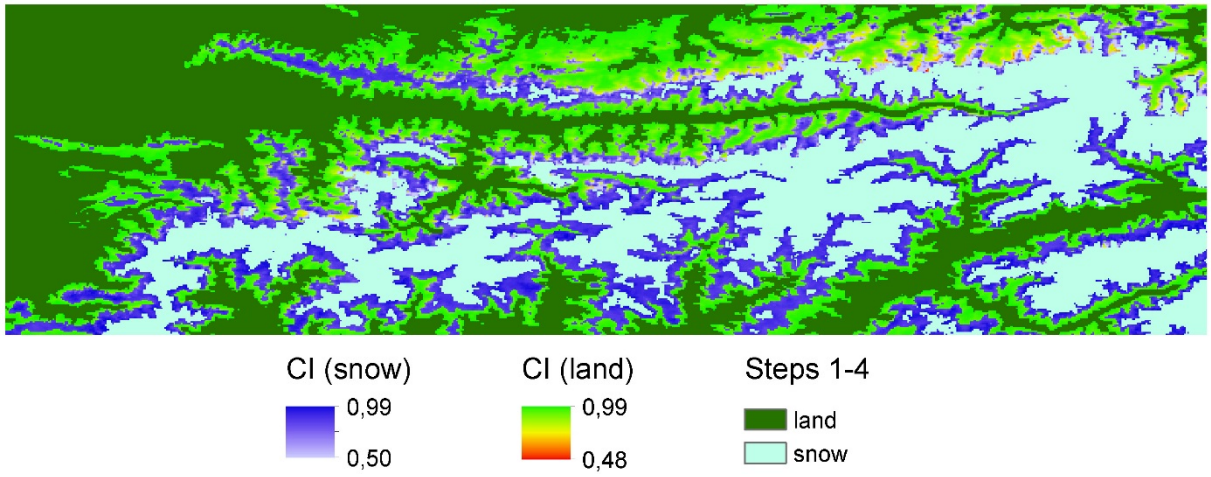
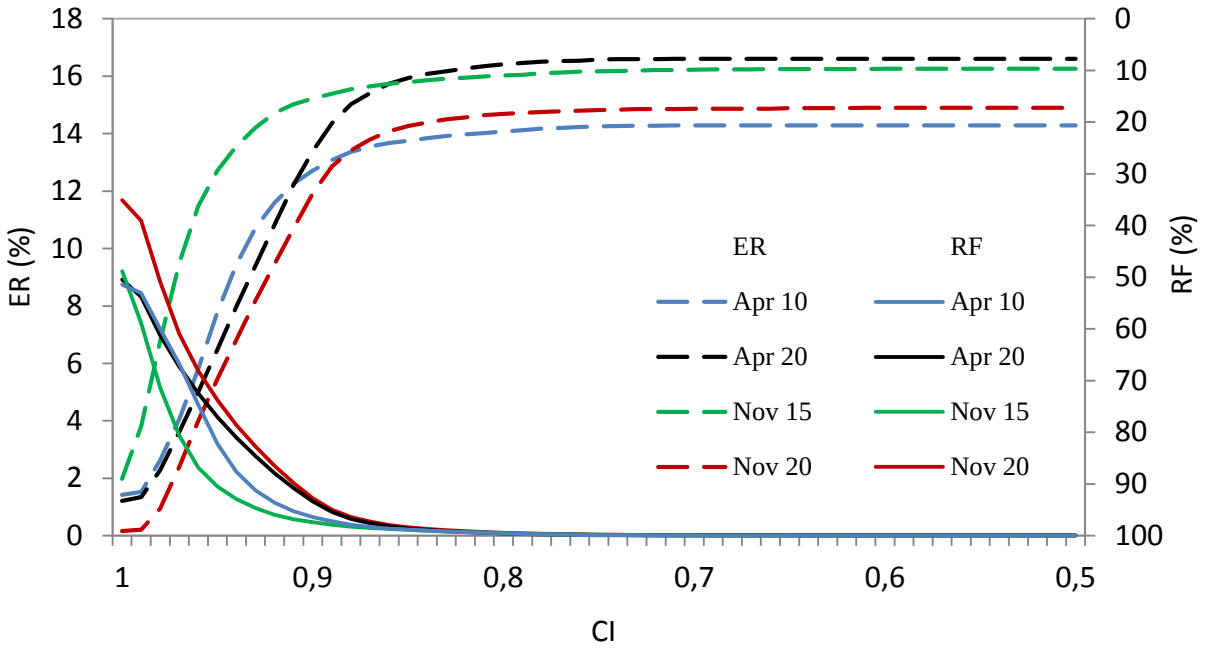


Figure 9. Fraction of reconstruction in steps 1-4 and maximum CI values for snow and land in step 5 for the study area illustrated in Fig. 1b.



1

2 Figure 10. Trade-off between erroneous reconstruction (ER, dashed lines) and reconstruction
3 fraction (RF, solid lines) after step 4 as a function of CI.

4