Validation of snow extent time series derived from AVHRR GAC data at Himalaya-Hindukush

Xiaodan Wu[1,2], Kathrin Naegeli[2], Carlo Marin[3] and Stefan Wunderle[2]

[1]College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, China
5 [2]Institute of Geography and Oeschger Center for Climate Change Research, University of Bern, Hallerstrasse 12, CH-3012 Bern, Switzerland
[3]EURAC research, 39100 Bolzano - Italy

Correspondence to: Xiaodan Wu (wuxd@lzu.edu.cn)

Abstract. Long-term monitoring of snow cover is crucial for climate and hydrology studies. To meet the increasing demand
10 for a long-term, consistent snow product, an exceptional snow cover climatology was generated dating back to the 1980s using AVHRR GAC data. However, the retrieval of snow extent is not straightforward due to artifacts introduced during data processing, which are partly caused by the coarse spatial resolution of AVHRR GAC data, but also heterogeneous land cover/topography. Therefore, the accuracy and consistency of this long-term AVHRR GAC snow cover climatology needs to be carefully evaluated prior to its application. Here, we extensively validate the AVHRR GAC snow cover extent dataset for
15 the Hindu Kush Himalaya (HKH) region. The mountainous HKH region is of high importance for climate change, impact and adaptation studies. Additionally, the influences of snow depth, land cover type, elevation, slope, aspect, and topographical variability, as well as the sensor-to-sensor consistency have been explored using a snow dataset based on long-term in situ stations and high-resolution Landsat TM data. Moreover, the performance of the AVHRR GAC snow cover dataset was also compared to that of MODIS (MOD10A1 V006). Our analysis shows an overall accuracy of 94% in
20 comparison with in situ station data. Using a ±3 days temporal filter caused a slight decrease in accuracy (from 94 to 92%), which is still comparable to MOD10A1 V006 (93.6%). Validation against Landsat5 TM data over region of "P140-R40/41" indicated overall RMSEs of about 13% and 16% and overall Biases of about -1% and -2% for the AVHRR GAC raw and gap-filled snow datasets, respectively. It can be concluded that the here validated AVHRR GAC snow cover climatology is a highly valuable and powerful dataset to assess environmental changes in the HKH due to its good quality, unique temporal
25 coverage (1982-2018), and inter-sensor/satellite consistency.

1 Introduction

Snow cover is an important indicator to estimate climatic changes and a key input for climate, atmospheric, hydrological, and ecosystem models (Fletcher et al., 2009; Hüsler et al., 2012; Xiao et al., 2018). On one hand, snow cover exacerbate the effect of global warming through the positive feedback between snow and albedo (Serreze & Francis, 2006). Furthermore, it
30 affects the hydrometeorological balance through snow melt (Simpson et al., 1998). On the other hand, snow cover is severely

affected by climate change due to its high sensitivity to changes in temperature and precipitation (Brown & Mote, 2009). Therefore, accurate monitoring of its long-term behavior is a vital issue in improving weather and climate prediction, helping water management decisions, and investigating climate change impacts on environmental variables (Arsenault et al., 2014; Sun et al., 2020).

35   Optical satellite data provide important data sources for snow cover retrieval through the contrasting spectral behavior of snow relative to other natural surfaces in the visible and middle infrared region (Tedesco, 2014; Zhou et al., 2013). The global spatial coverage of satellite data makes it an efficient data source to improve our knowledge of snow cover dynamics (Siljamo & Hyvärinen, 2011; Solberg et al., 2010). Many satellites have been used to generate snow cover products at various spatial and temporal resolutions, such as AMSR-E (KELLY, 2003), MODIS (Hall et al., 2002), AVHRR (Simpson

40   et al., 1998), VIIRS (Key et al., 2013), and Landsat (Rosenthal and Dozier, 1996). In particular, new generation satellite sensors (e.g., MODIS, VIIRS) generally show an advantage over old sensors such as AVHRR and TM/ETM who suffer from significant saturation over snow in the visible channels (WMO, 2012). Nevertheless, AVHRR offers the unique opportunity to generate a consistent snow product over a 30-year normal climate period (IPCC, 2013), and thus remains vitally important. In response to the systematic observation requirements of Global Climate Observing System (GCOS), the

45   ESA Climate Change Initiative (CCI) has emphasized the necessity of generating consistent, high quality long-term datasets over the last 30 years as timely contribution to the ECV databases. Following this recommendation, the Remote Sensing Group at University of Bern, Switzerland has generated a global time series of daily fractional snow cover product from AVHRR GAC data since 1982.

However, retrieving snow cover from AVHRR GAC data is a challenging task from several observational and

50   environmental standpoints (Hall and Riggs, 2007). First, the signal observed over a certain area does not necessarily represent its actual condition, but may depend on the circumstances of observation or the artifacts introduced during data sampling (Cracknell, 1997) and processing, such as cloud contamination (Crawford, 2015; Zhou et al., 2013), atmospheric constituents (e. g., water vapor, aerosols) (Gafurov et al., 2012), topographic illumination, solar zenith angles, off-nadir viewing angles, and bidirectional reflectance distribution function (BRDF) effects (Crawford, 2015). Second, dense

55   vegetation like forest can cause problems in snow mapping because snow on ground is often not visible from above depending on the density of canopy (Gafurov et al., 2013). Last but not the least, the change between the 3B and 3A channel with the launch of AVHRR/3 instrument in 1998 may cause discontinuities in a satellite based time series (Heidinger et al., 2004). All these factors decrease reliability of snow cover retrieval. Therefore, it is necessary to comprehensively evaluate the accuracy of this AVHRR GAC snow cover extent.

60   To date, on the one hand, the performance of the AVHRR GAC snow product has not been fully explored. Of particular importance is the performance of the product over different platform operated periods, land cover types, elevations, aspect, slopes, topographies, and snow depths. On the other hand, the Hindu Kush Himalaya (HKH) area is of special interest due to large area, rich diversity of climates, hydrology, ecology, and biology (Wester et al., 2019). In this study, we therefore focus our comprehensive validation study of the ESA CCI+ snow product over HKH during snow seasons. This mountain area

65   features varying climate and snow conditions, and offers distinct characteristics of snow cover with shallowness, patchiness, and frequent short duration ephemeral snow (Qin et al., 2006). It provides an ideal area to fully investigate the performance of the AVHRR GAC snow cover product in different conditions. The validation was performed by using snow climatology derived from in situ measurements, high-resolution Landsat Thematic Mapper (TM) data, and third-party MODIS snow products following well-established snow validation and inter-comparison procedures. Long-term in situ measurements

70   enable us to assess sensor-to-sensor consistency with regard to time series application and characterize the influence of snow depth (SD) on snow cover accuracy. The comparison with Landsat TM data enables us to account for various factors affecting data quality, including land cover, elevation, topography, and orientation (aspect/slope). While the comparison with MODIS snow product is focused on their agreement with in situ measurements, as well as the comparison of their absolute values. Section 2 describes the study area and data. The validation methodology is explained in Section 3. The performance

75   of AVHRR GAC snow dataset is presented and discussed in Section 4. A brief conclusion is presented at the end.

2 Study area and data

2.1 Study area

HKH covers a mountainous region of more than 4 million km2 within the geographic area between about 16° to 40° N latitude and 60° to 105° E longitude. It extends across all or parts of eight countries, namely Afghanistan, Bangladesh,

80   Bhutan, China, India, Myanmar, Nepal, and Pakistan (You et al., 2017). Moreover, it contains the highest concentration of snow and ice outside the Polar Regions and is thus referred to as the Third Pole (Wester et al., 2019).

This region is one of the most dynamic, fragile, and complex mountain systems in the world due to the rich diversity of climatic, hydrological, and ecological characteristics. The climate conditions range from tropical (<500 masl) to high alpine and nival zones (> 6000 masl), with a principal vertical vegetation regime composed of tropical and subtropical rainforests,

85   temperate broadleaf, deciduous, or mixed forests, temperate coniferous forests, alpine moist and dry scrub, meadows, and desert steppe (Guangwei, 2002). The main land cover of this region is rangeland, which covers approximately 54% of the total area. Agriculture and forest are also present, accounting for 26% and 14% of this region, respectively. 5% of this region is permanent snow and glaciers, and 1% is water bodies (Ning et al., 2014; Wester et al., 2019).

The validation based on in situ sites covers mainly the eastern part of HKH (Fig. 1a). Therefore, the validation based on

90   Landsat TM/ETM data was divided into two parts: east and west. Due to the fact that it is very challenging to find enough cloud-free Landsat TM data over such a large area, validation against Landsat TM data was performed in detail using two tiles of Landsat data (path 140, rows 40 and 41, denoted as "P140-R40/41") (Fig.1c), covering a diverse region on the Nepal/Tibet border centered around Mount Everest. This region was chosen because it contains the greatest elevation range in the Himalayas. The northernmost part of this region are areas on the Tibetan plateau exceeding 6000 m a.s.l. where

95   vegetation change is occurring rapidly (Qiu, 2016). Furthermore, it covers a broad range of climatic conditions (Bookhagen

& Burbank, 2006). Therefore, this region is a microcosm of the range of conditions experienced across the wide HKH and thus provides a good point for investigating snow extent accuracy under different conditions (Anderson et al., 2020).
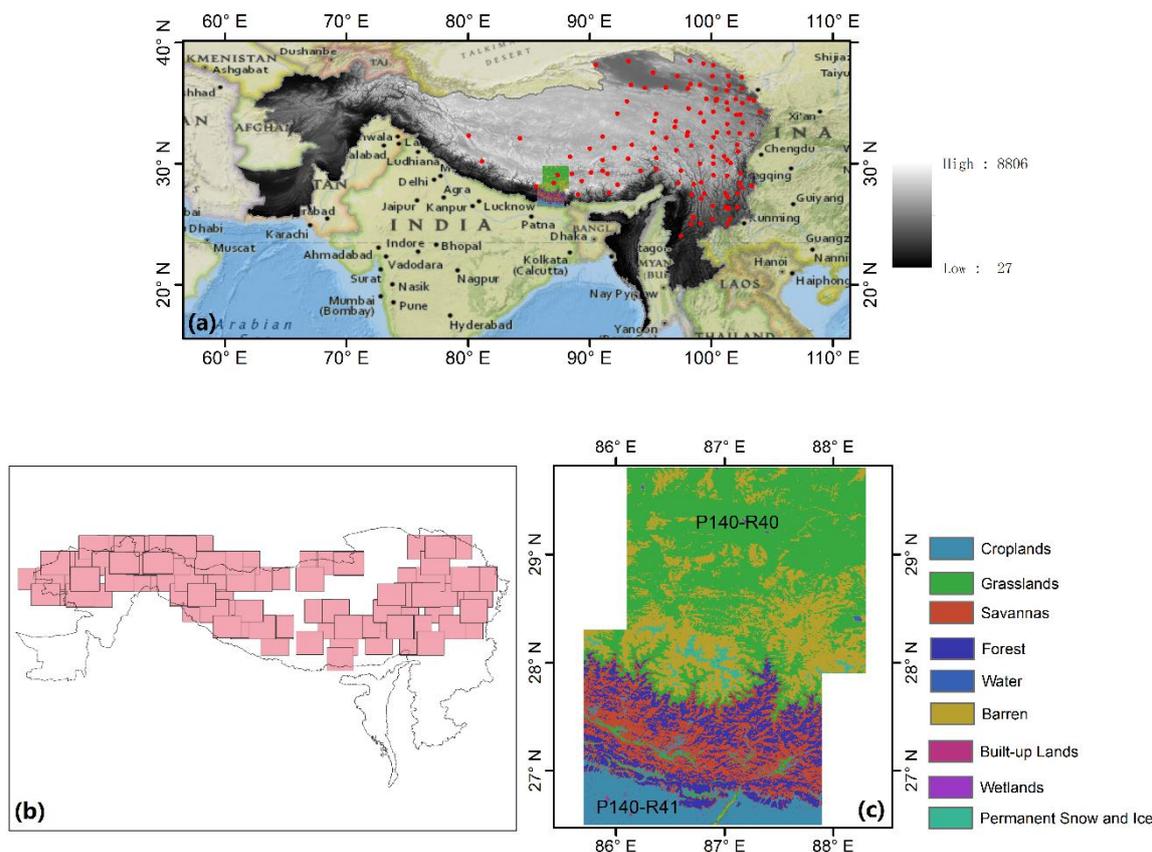

Figure 1. (a) The HKH region and the distribution of *in situ* station locations (red dots); (b) The distribution of Landsat scenes available over the whole HKH; (c) Land cover type of the "P140-R40/41" region of interest which corresponds to Landsat path 140, rows 41 and 42 on the Nepal/Tibet border.

2.2 AVHRR GAC snow extent retrieval

We used the Fundamental Climate Data Record (FCDR) based on daily composites of AVHRR GAC data (doi:10.5676/DWD/ESA_Cloud_cci/AVHRR-PM/V003) produced in the ESA Cloud CCI project (Stengel et al., 2020). The data were pre-processed with an improved geocoding and an inter-channel and inter-sensor calibration using PyGAC (Devastahle et al., 2017). Alongside the daily reflectance and brightness temperature information, an excellent cloud mask including pixel-based uncertainty information is provided (Stengel et al. 2017, Stengel et al. 2020). The fractional snow retrieval method by Salomonson and Appel (2006), which is based on a statistical linear regression between fractional snow cover (FSC) and the Normalized Difference Snow Index (NDSI) (FSC = $-0.01 + 1.45 \times$ NDSI), was utilized. To reduce the effect of cloud coverage, a temporal filter was used with application of a closest ± 3 days of snow-cover observations. In this

The Cryosphere
Discussions

EGU

study, both raw daily retrieval of AVHRR GAC snow extent (denoted by "AVHRR_Raw") and cloud-gap-filled daily AVHRR GAC snow extent (denoted as "AVHRR_Gap filled") were evaluated.

2.3 Validation data sets

Reference data to be used for validation of a snow satellite dataset should cover a similar spatial or temporal extent, ideally
115 acquired at the same point in time. In situ station data provide valuable long time series measurements to check the temporal performance of the AVHRR GAC snow extent and sensor-to-sensor consistency. Considering the limitations resulted from the spatial scale mismatch between in situ and satellite observations, snow maps generated from Landsat TM/ETM were also utilized, enabling us to check the dependence of AVHRR GAC snow extent accuracy on land cover type and topography. Additionally, MODIS snow product, as the most widely used snow product, was also compared to the in situ station data to
120 evaluate the AVHRR GAC snow cover extent from a relative perspective. The data and validation purpose are briefly summarized in Table 1.

Table 1. Validation data and purpose.

Data type	Data	Method	Validation purpose
In situ-based	Snow depth	Direct comparison between *in situ* and AVHRR GAC snow dataset	Investigate the sensor-to-sensor consistency; Accuracy dependency on snow depth
High spatial resolution images (Landsat)	Landsat TM/ETM, Land cover, topography, elevation, orientation	Quantitative analysis of accuracy of AVHRR snow extent	Accuracy dependency on land cover type, topography, elevation
Satellite snow product (MODIS)	Snow cover	Comparing their performance with respect to *in situ* data as well as their absolute values for a short period to show the relation to AVHRR GAC time series	Relative validation

2.3.1 In situ snow depth measurements

In situ data were provided by China Meteorological Administration (https://data.cma.cn/en). Daily snow depth (SD)
125 measurements (118*32*365) are obtained from 118 stations located at different elevations ranging from 776 to 8530 m above sea level. SD was usually measured over a large flat area using rulers at 08:00 every day. Three measurements were made at least 10 meters away and their mathematical mean was used as the daily snow depth. In particular, if snowfall occurred after 08:00, a second measurement at 14:00 or a third measurement at 20:00 were needed depending on the time of

snowfall. The data were rounded to the nearest centimeter. Thus, SD less than 0.5 cm would be labeled as 0 cm in the record.

130   Detailed quality control was made to flag suspicious values. The period from 1982 to 2013 was used to prove the long-term consistency of the AVHRR GAC snow extent.

2.3.2 Landsat TM/ETM data and processing

Data from Landsat were explored partly due to its high spatial resolution of 30 m and partly due to its longest operation period, which almost entirely covers all NOAA sensors (NOAA-6 to -19). To mitigate the effect of cloud, the validation over

135   "P140-R40/41" was restricted to clear-sky (cloud no more than 10%) scenes of Landsat 5 TM during snow seasons (46*2 scenes from 1984 until 2013; downloaded from https://glovis.usgs.gov/). But the validation over the whole HKH was restricted to Landsat clear-sky scenes from 1999 to 2018 (around 200 scenes) (Fig. 1b). Level-1 Precision and Terrain corrected ("L1TP") data were selected since they have been radiometrically and geometrically corrected. Although many methods (Klein et al., 1998; Dozier and Painter, 2004; Salomonson and Appel, 2006) have been used to generate snow data

140   from Landsat imagery, only the fractional snow method by Salomonson and Appel (2006) was employed to derive high-resolution snow maps in this study. As the snow retrieval algorithm applied to the AVHRR GAC dataset identifies viewable snow, and Salomonson and Appel (2006) provides the same thematic information and was thus applied to the Landsat data.

Spectral bands 2 (0.53–0.61 μm) and 5 (1.55–1.75 μm) were used to provide 30m NDSI estimates (Equation (1)), and FSC was retrieved using the statistical linear regression between FSC and NDSI (Equation (2)) (Salomonson and Appel,

145   2006). Then the 30 m FSC were resampled (nearest neighbor method) and projected to a geographic projection of 0.05 ° to identify FSC within a given AVHRR GAC pixel.

$$NDSI = (B2 - B5)/(B2 + B5) \tag{1}$$
$$FSC = -0.01 + 1.45 \times NDSI \tag{2}$$

where $B2$ and $B5$ denote the spectral bands 2 and 5, respectively.

150   2.3.3 MODIS snow cover product

The Terra MODIS Level 3, collection 6, 500 m daily snow cover products (MOD10A1) (Hall et al., 2016) over HKH from 2000 to 2013 were obtained through Google Earth Engine (GEE). MODIS snow detection algorithm also uses NDSI and other criteria tests (Riggs et al., 2015). Instead of directly providing binary snow-cover area (SCA) and FSC, V006 version provides NDSI_Snow_Cover and NDSI. The former is reported in the range of 0-100 with other features identified by mask

155   values, while the latter represents the real NDSI values multiplied by 10000, which is calculated for all pixels (Riggs et al., 2016). This treatment provides more information and great flexibility to enhance the accuracy of the product, because NDSI range is not necessarily restricted to 0.4 to 1.0 for snow detection. Actually, NDSI_Snow_Cover function very similarly to FSC in V005 version since they can be linked using the equation of FSC = −0.01 + 1.45 × NDSI. Compared to the previous version, V006 version made great improvements on atmospheric correction, cloud cover, and quality index. Furthermore, the

160   algorithm takes pixel's elevation into account, which is especially important for elevated snow-covered surfaces in spring.

The Cryosphere
Discussions

2.3.4 SRTM DEM

The DEM information was obtained from the SRTM dataset, which provides a nearly global coverage with a spatial resolution of 90 m. In this study, the elevation, slope, aspect, and topographical variability were derived using this dataset in order to investigate their influences on accuracy of AVHRR GAC snow extent product. The topographical variability within

165   a certain AVHRR GAC pixel was determined by calculating the standard deviation of elevations of all sub-pixels within its spatial extent. While the elevation, slope, and aspect were resampled to match the resolution of AVHRR GAC snow dataset.

2.3.5 MODIS Land cover data

The MODIS Terra/Aqua Combined Annual Level 3 Global 500-m Version 6 land cover dataset (MCD12Q1) was generated using a supervised classification methodology (Friedl et al., 2010). In this study, the International Geosphere–Biosphere

170   Programme (IGBP) of MCD12Q1 mosaic was used to investigate the difference in accuracy over different land cover types. It includes 11 types of natural vegetation, 3 types of developed and mosaic lands, and 3 types of non-vegetated lands, which have been reclassified into nine major classes: forest, grassland, savannas, croplands, built-up lands, barren, permanent snow and ice, water body, and wetlands.

3 Methods

175   ### 3.1 Direct-comparison based on in situ data

Although the validation based on in situ sites leaves issues of scale unresolved and therefore likely accompanied by uncertainties, in situ observations are nevertheless an important data source for validation in regions lacking long-term distributed validation data (Hao et al., 2018; Yang et al., 2015; Mir et al., 2015; Hüsler et al., 2012). Since there is no reliable way to convert SD to FSC, both FSC and SD information were converted to binary information by applying appropriate

180   thresholds, respectively. Different thresholds have been suggested for in situ SD measurements to determine whether the associated pixel is covered by snow, ranging from 0 cm to 5 cm (Parajka et al., 2012; Hori et al., 2017; Hao et al., 2019; Huang et al., 2018; Liu et al., 2018; Zhang et al., 2019; Gascoin et al., 2019). Therefore, the sensitivity of thresholds was tested by computing accuracy metrics with SD increasing from 1 cm to 5 cm. The FSC maps were transferred from fractional to binary snow information by applying a threshold of FSC≥50%. The value of 50% is widely used and accepted

185   in snow cover detection (Wunderle et al., 2016; Mir et al., 2015; Crawford, 2015; Marchane et al., 2015; Hall et al., 2007).

Concerning the comparison of spatial satellite data with in situ measurements, a point-wise comparison was implemented. To relate in situ "point" measurements with AVHRR GAC "area" snow information, both the center pixel containing the in situ "point" measurement and the 3 × 3 pixels centered around this "point" were used, respectively. This treatment took into consideration the influence of data noise, geometric mismatch, and spatial heterogeneity. Furthermore,

The Cryosphere
Discussions
Open Access
EGU

190   the absence or presence of snow indicated by in situ observations is assumed to be representative of at least a 3×3 pixel area
but this depends on topography. Consequently, there are altogether 10 combination cases for accuracy assessment (Table 2).

Table 2. A short summary of all the combinations of thresholds

Cases	Combinations	Cases	Combinations
case1	SD>=1cm, center pixel	case6	SD>=1cm, 3×3 pixels
case2	SD>=2cm, center pixel	case7	SD>=2cm, 3×3 pixels
case3	SD>=3cm, center pixel	case8	SD>=3cm, 3×3 pixels
case4	SD>=4cm, center pixel	case9	SD>=4cm, 3×3 pixels
case5	SD>=5cm, center pixel	case10	SD>=5cm, 3×3 pixels

The 2×2-contingency table statistics (Table 3) was utilized to indicate the quality of snow product. If both reference
195   data and snow product identified the pixel as snow, it is labeled as a hit (a); if neither of them indicated the pixel as snow, it
is labeled as zero (d); if the snow product indicates the pixel as snow, but the reference data is not, it is marked as false (b);
and if the opposite occurs, it is indicated as a miss (c) (Hüsler et al., 2012; Siljamo et al., 2011).

Table 3. Contingency table used to determine probability of detection and bias.

Snow product (AVHRR GAC, MODIS)		Reference data (*In situ*)		
		yes	**no**	
	yes	a: hit	b: false	a+b
	no	c: miss	d: zero	c+d
		a+c	b+d	n=a+b+c+d

Based on these measures, indicators such as accuracy (ACC), Heidke Skill Score (HSS), and bias (Bias) were
200   determined (Equations (3-5)) (Hüsler et al., 2012). ACC denotes the percentage of correctly classified pixels divided by total
number of pixels. ACC values closer to 1 denotes a better agreement between the snow product and the reference data, while
a value of 0 corresponds to complete disagreement. However, it is strongly influenced by the most frequent category (i.e. in
summer) (Hüsler et al., 2012) and thus ideally requires an equal distribution of categories. Hence, we confine our accuracy
assessment to the snow season (from October to March) only, a limitation that was performed by other studies as well (Yang
205   et al., 2015; Gafurov et al., 2012; Hüsler et al., 2012; Huang et al., 2011). The HSS and Bias provide refined measures in
case that the frequency distribution within the validation subsets is not equal. The former describes the proportion of pixels
correctly classified over the number correct by chance in the total absence of skill. Negative values indicate that the chance
performance is better, 0 represents no skill, and a perfect performance obtains a HSS of 1 (Hüsler et al., 2012). It is generally
true that a value above 0.3 denote a relatively good score for a reasonably sized sample for binary forecast (Singh, 2015).
210   The latter, described by the ratio of the number of snow cover pixels to the number of reference data pixels, is a relative
measure to detect overestimation (value is higher than 1) or underestimation of snow (value is less than 1). Unbiased results
should have the value of 1.

The Cryosphere
Discussions

$$ACC = \frac{a+d}{a+b+c+d} \tag{3}$$

$$HSS = \frac{2(ad-bc)}{(a+c)(c+d)+(a+b)(b+d)} \tag{4}$$

215 $$Bias = \frac{a+b}{a+c} \tag{5}$$

The validation follow two types of strategies: First, the snow cover data time series of satellite and in situ station were compared, resulting in accuracy indicators over each station. This validation allow us to check the spatial divergence of accuracy within different sites as well as the effect of land cover on accuracy of satellite derived snow information. Second, the snow cover data of all in situ stations and the corresponding satellite snow pixel values were combined respectively for

220 each day in snow season. Furthermore, the comparison was implemented on daily basis, resulting in a time series of accuracy indicators for the whole period. In this way, the sensor-to-sensor consistency of accuracy can be evaluated. Additionally, in order to assess the product performance with respect to the temporal variability of snow cover, the binary metrics are summarized and analyzed for each month. An analysis of increase/decrease of accuracy with respect to FSC and SD was also included to explore the influence of smaller snow patches on the accuracy.

225 **3.2 Validation based on high-resolution snow cover maps**

In this validation framework, Landsat TM/ETM aggregated FSC were used as the reference to assess the performance of AVHRR GAC snow. Snow free values are treated as 0% snow and fully snow-covered pixel is assigned 100% snow. This validation approach provides two benefits: (i) quantitative validation results (fractional metrics) are obtained and (ii) the spatial performance of the AVHRR GAC snow cover dataset is evaluated by a pixel by pixel comparison. Statistical indexes

230 such as the root mean square error (RMSE), mean bias (mBias) and the coefficient of correlation (R) were employed to quantitatively characterize the accuracy.

The evaluation was conducted over the 'P140-R40/41' and whole HKH respectively from two aspects. First, the pixel values over the entire study period were combined, resulting maps containing RMSE and mBias for each pixel. The validation over 'P140-R40/41'offers the possibility to assess the effects of potential influence factors (land cover, elevation,

235 slope, aspect, and topography variability) on the accuracy of AVHRR snow dataset given the relatively high availability of scenes (46*2). Furthermore, the analysis was also performed for the whole HKH, eastern and western areas, forested and non-forested areas, mountain and plain areas. Second, the values of each pixel of the selected map were combined, resulting in a RMSE, mBias, and R per scene. In this way, the influence of the acquisition date on the performance of AVHRR GAC snow dataset can be investigated. This kind of validation was only performed over 'P140-R40/41' given the relatively high

240 availability of scenes.

3.3 Relative comparison with MOD10A1

In this study, the comparison with MOD10A1 was carried out from two aspects: their agreement with in situ measurements, as well as their absolute values. To achieve this, MOD10A1 V006 was compared with in situ station data as described in Section 3.1, to evaluate the overall quality of the newly derived AVHRR GAC snow cover climatology in relation to the well-used MODIS product. It is expected that the major difference in agreements between the two satellite-derived snow cover products and the in situ measurements reflect their relative performance, which is either due to the quality of the applied processing and snow cover retrieval algorithms or the general satellite data characteristics. As for the comparison of their absolute values, the 500 m FSC derived from MOD10A1 were resampled and projected to the resolution of AVHRR GAC pixel. RMSE, mBias, and R were calculated to measure their difference.
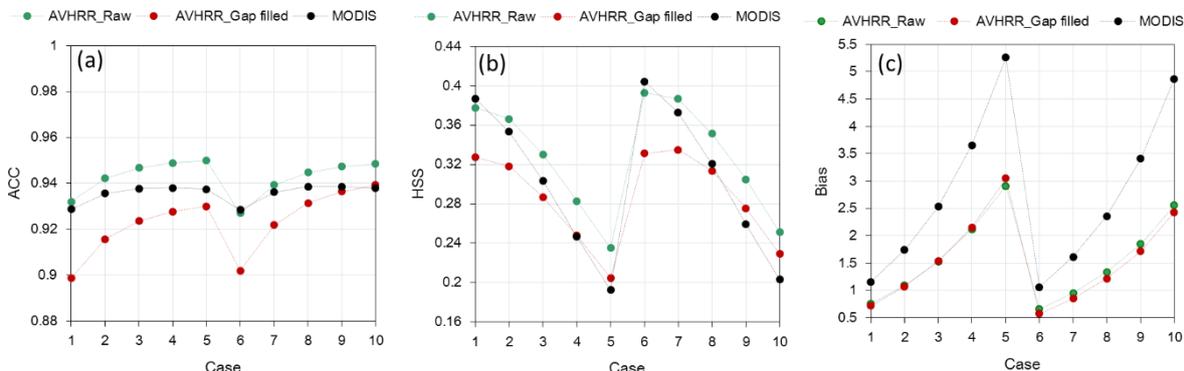
4 Results and discussion

4.1 The comparison with in situ data

4.1.1 The threshold sensitivity analysis

To test the sensitivity of the in situ SD threshold for the snow cover detection, the overall accuracy metrics were computed by combining data of all in situ sites throughout the study period (from 1982 to 2013 for the AVHRR GAC-derived snow and from 2000 to 2013 for MOD10A1). The variations of ACC, HSS, and Bias with all the threshold combinations (Table 2) are shown in Fig. 2. We can notice that the ACC for all three datasets (AVHRR raw, AVHRR gap-filled, and MODIS snow data) are generally high, with values larger than 0.9 for all combination cases. Nevertheless, ACC changes considerably among the 10 cases. An increasing SD threshold positively affect ACC (Fig. 2a). Generally, the rate of change for ACC is highest between the 1 and 2 cm SD threshold and flattens for greater SD thresholds. This is valid for all three satellite datasets and independent of the pixel(s) considered (center pixel versus 3 ×3 pixels, Table2).

It is noticeable that ACC increases at the expense of the other metrics of HSS and Bias. In fact, both HSS and Bias feature the opposite behavior than ACC, i.e. an increasing SD threshold, negatively affects HSS and bias (Fig. 2(b and c)). Again, these results are generally valid for all three satellite datasets and independent of the pixel(s) considered. However, interestingly for both AVHRR datasets the bias changes from underestimation for a SD threshold of 1 cm to overestimation for SD thresholds ≥ 2 cm. Unlike AVHRR, MODIS snow product exhibits a consistent overestimation with increasing SD threshold. Generally, the rate of change for HSS and Bias is greatest between the 5 and 4 cm SD threshold and flattens for smaller SD thresholds. Considering Bias, a SD threshold of 2 cm (case2 and 7) and of 1 cm (case1 and 6) maximizes the overall accuracy of the AVHRR GAC snow cover dataset and the MODIS snow product, respectively. In other words, the presence of snow can be best detected by the AVHRR GAC dataset for in situ snow depth measurement of 2 cm.

Figure 2. Sensitivity analysis of product accuracy to the choice of the snow depth threshold in the *in situ* station data. The overall error is a spatio-temporally integrated statistical measure.

When it comes to the influences of geometric mismatch or spatial heterogeneity (center pixel versus 3 × 3 pixels, Table2), they are negligible for the variation trend of accuracy indicators with SD threshold but cannot be ignored for the absolute value of accuracy indicators. However, the latter effect is not fixed and varies with satellite datasets and accuracy indicators. For the AVHRR raw snow dataset, the ACC based on center pixel (cases 1-5) are consistently larger than those based on the average of 3 × 3 pixels (cases 6-10) at each specific threshold on SD (Fig. 2a). While a contrary phenomenon can be observed for the AVHRR gap-filled snow cover dataset, which exhibits smaller ACC based on center pixel than based on the average of 3 × 3 pixels. The magnitude of ACC for MODIS seems to be unaffected by the choice of pixel(s) considered. Unlike the results for ACC, the HSS of all three satellite datasets and all five SD thresholds based on the center pixel are consistently smaller than those based on the average of 3 × 3 pixels, indicating that the effect of geometric mismatch and spatial heterogeneity on HSS is systematic and basically independent of the satellite datasets and SD thresholds. A similar result can be found for the Bias, which is consistently larger based on the center pixel than those based on the average of 3 × 3 pixels for each SD threshold and all three satellite datasets.

Since the performance of the satellite snow datasets is represented by three indicators, a compromise is needed in the choice of SD threshold to both account for increasing ACC and decreasing HSS and Bias with increasing SD threshold. Here, we adopted 2 cm as the optimum threshold to transform in situ SD measurements to snow-covered or snow-free information as the threshold of 2 cm is always the visible break for ACC, HSS, and Bias (Fig. 2). Furthermore, both the center pixel and the average of 3 × 3 pixels will be considered since their effects on magnitude of ACC are not consistent among these three satellite datasets. Consequently, the validation results based on case2 and case7 will be focused in the following analyses.

In general, we can observe that the ACC for AVHRR raw snow cover datasets are 0.942 and 0.939 in case2 and case7, respectively. Using a temporal filter to mitigate the impact of clouds caused a slight reduction in ACC of the AVHRR gap-filled snow cover dataset to 0.916 and 0.922 in case2 and case7, respectively. The ACC of MODIS is situated between AVHRR raw and gap-filled snow datasets, and equals to 0.936 in both cases. A similar phenomenon can be found in HSS, where the AVHRR raw snow cover dataset shows the largest HSS of 0.366 and 0.387, followed by the MODIS snow
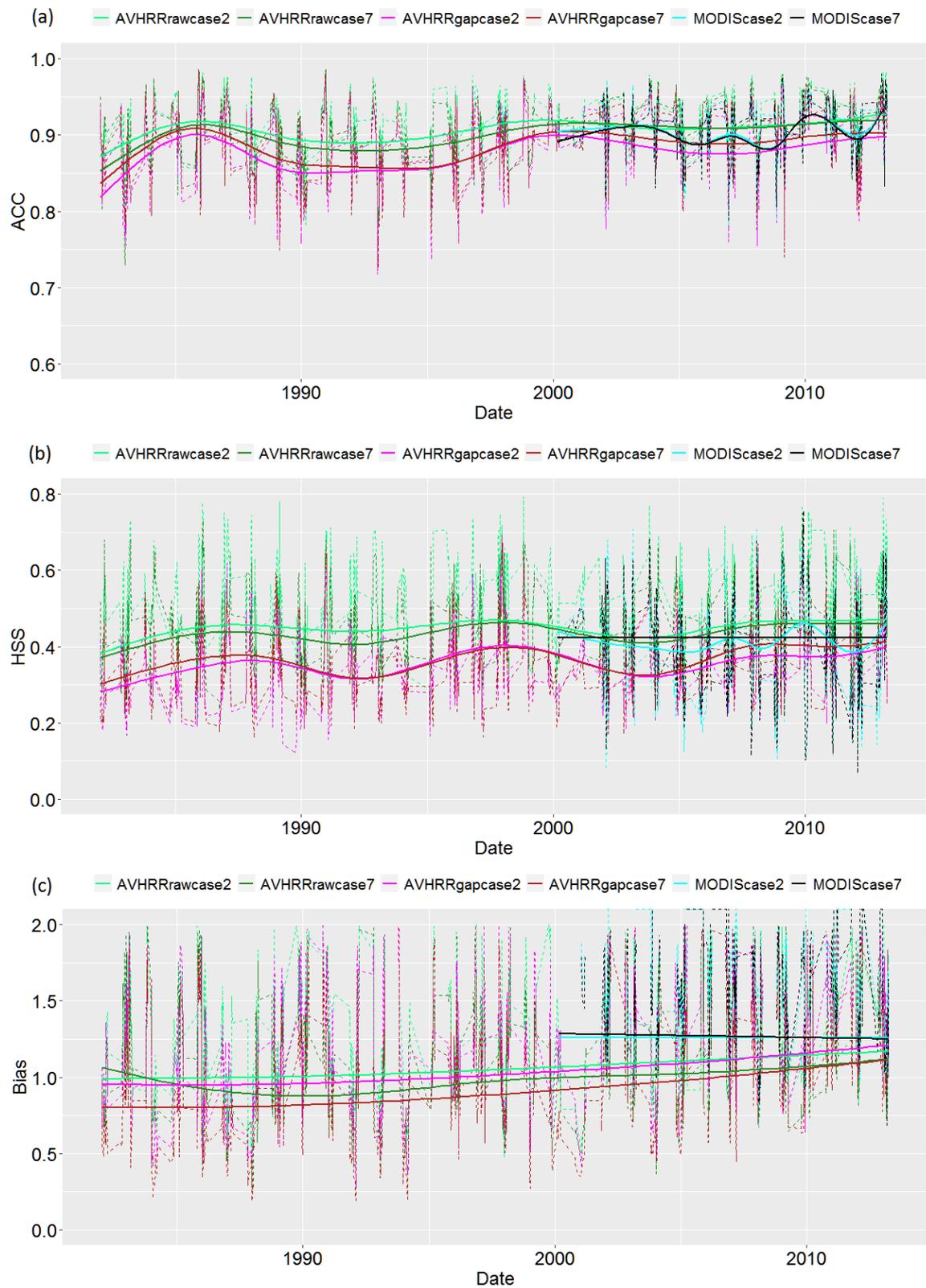
product, with 0.353 and 0.373 in case2 and case7, respectively. The AVHRR gap-filled snow cover dataset ranks last, with smallest HSS of 0.318 and 0.335 in case2 and case7, respectively. In contrast, the Bias shows a different behavior compared to ACC and HSS. While the AVHRR raw and gap-filled snow cover datasets are exhibiting very similar values (1.091 and 0.955, and 1.076 and 0.855 in case2 and case7, respectively), the MODIS snow products shows a much larger Bias of 1.740

300   and 1.613 in case2 and case7, respectively. Combining the results of ACC, HSS, and Bias, it can be concluded that the AVHRR raw snow cover dataset present a good to very good performance when compared with in situ data. The AVHRR gap-filled snow cover dataset shows a slightly worse performance, but it still maintains a high overall accuracy, which is comparable to the MODIS snow product.

4.1.2 The temporal behavior of quality indicators

305   The temporal consistency provides information on potential sensor artifacts in the time series. It was assessed by computing the validation measures for each day in snow season (Fig. 3). Both 1*1 (case2) and 3*3 (case7) pixel area were considered for three satellite snow datasets. It can be seen that the accuracy metrics show a certain inter-annual variability, especially for ACC and HSS. Furthermore, the temporal variation characteristics of these metrics derived from AVHRR raw and gap-filled datasets are basically consistent throughout the time series in the two cases.

310   In the time series derived by AVHRR GAC data, ACC is basically distributed between about 85% and 90% (Fig. 3a). An obvious increase of ACC can be observed from 1982 to 1985, followed by a decrease in ACC from 1985 to 1990. From 1990 to 1995, ACC is relatively stable, followed by an increase from 1995 to 2000. From 2000, ACC shows a very slightly decreasing trend until 2007 from when it was substantially retained. Differently from the previous assessments, the ACC of AVHRR snow datasets at the beginning of the time series (1982-1995) is slightly worse than the end of the time series

315   (1996-2013). Similar to ACC, the HSS also shows clearly visible fluctuations in the time series (Fig. 3b). A slight increase of HSS can be found from 1982 until 1988, followed by a decrease from 1988 to 1992, and then an increase appear again until 1998 when HSS reach the maximum value (approximately 0.47 and 0.4 for AVHRR raw and gap-filled snow cover datasets, respectively). From 1998, the HSS of the two datasets show a decreasing trend until 2004, from when it continues to increase year by year and reaches the maximum in the final year. The Bias shows a different behavior compared to ACC and HSS

320   (Fig. 3c), which shows a slightly increasing trend throughout the time series except for the AVHRR raw snow datasets in case7. When the case2 was focused, the Bias of AVHRR raw snow dataset basically equals to 1 at the beginning of the time series and reach the maximum value of 1.15 in the end. By contrast, the AVHRR gap-filled snow dataset slightly underestimate snow cover at the beginning with Bias very close to 1 and reach the maximum value of 1.2. Nevertheless, a considerable underestimation can be observed in AVHRR raw snow dataset before 2000 and gap-filled dataset before 2005

325   when the case7 was considered.

Figure 3. The time series (denoted by the dashed lines) of ACC, HSS, and Bias for AVHRR raw, AVHRR gap-filled, and MODIS snow data during the experimental period. A simple moving average with box dimension n=7 was applied to the time series in order to reduce the noise and uncover patterns in the data. The solid lines represent the smoothing line with the method of "auto".
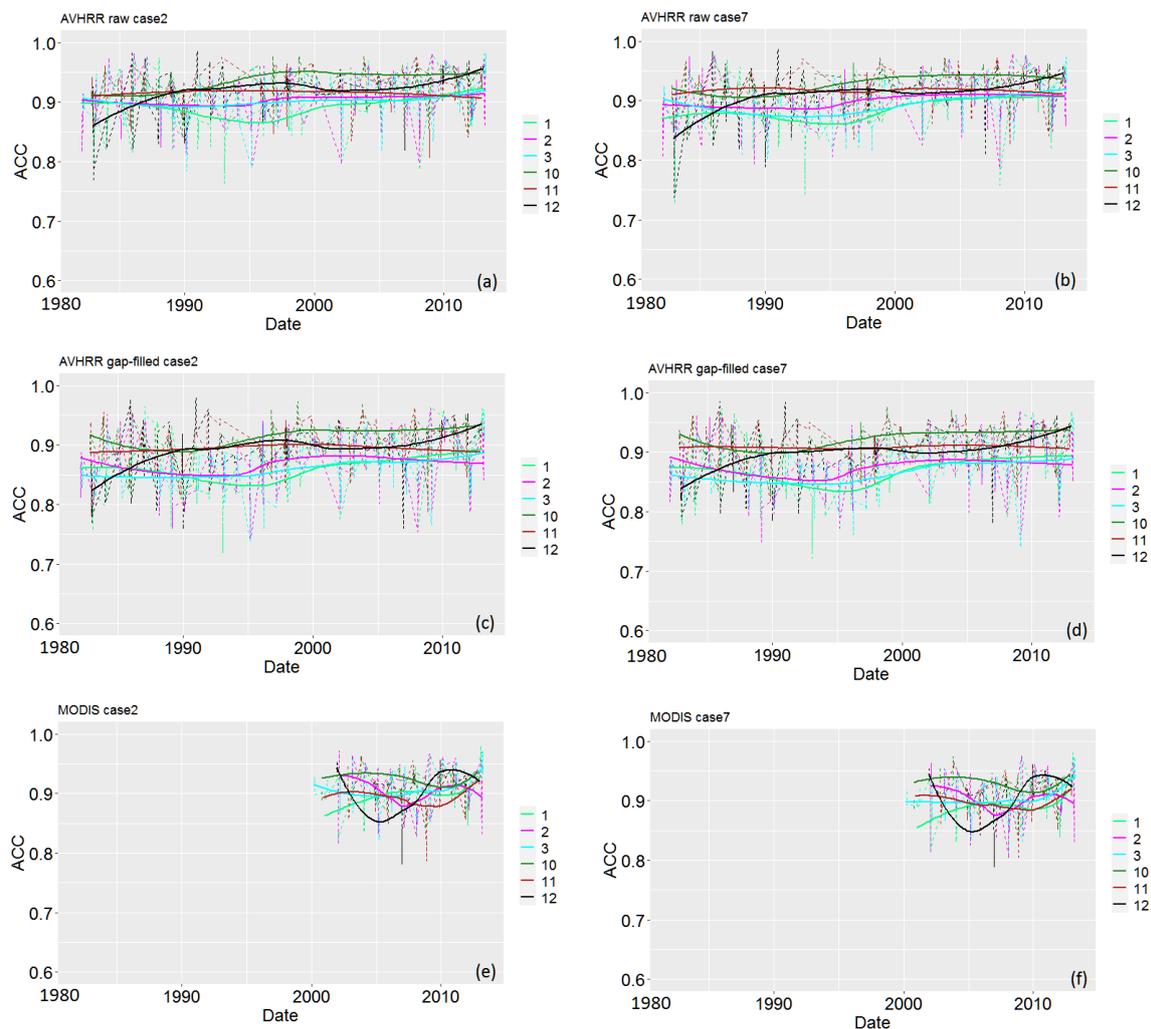
From Fig. 3, it is obvious that AVHRR raw snow dataset performs better than the gap-filled snow dataset. The ACC and HSS of the former is consistently larger than that of the latter in two cases throughout the time series (Fig. 3(a-b)). As for the Bias, a slight difference can be observed between them in case2, and their difference narrow down with the time. Nevertheless, the Bias of AVHRR raw snow dataset is closer to 1 than the gap-filled snow dataset before 2005 in case7 (Fig. 3c). In comparison to the 1-pixel case, AVHRR raw snow dataset shows decrease in ACC and HSS with aggregation but the AVHRR gap-filled snow dataset shows increase (Fig. 3(a-b)). As for the Bias, the 1-pixel case is closer to 1 for both datasets before 2000 beyond which the 3-pixel case is closer to 1 (Fig. 3c). Hence, it is concluded that the effect of the pixel(s) considered is not systematic. A comparison between the different platforms operation periods reveals minor and non-systematic difference in ACC and HSS throughout the time series. While the consistent slight increasing trend on Bias is notable.

The amount of snow over HKH varies over time during a year. Therefore, the performances of satellite snow datasets on the monthly scale is crucial to know to highlight how they vary at different month of the year. It was assessed by computing validation measures for each month during the experimental period, which are presented in Figs. 4-6.

From the results of ACC derived from AVHRR snow datasets (Fig. 4(a-d)), it can be seen that both the temporal variation trend and the magnitude of ACC show differences from month to month. It is interesting that the results tend to polarize into two groups: the ACC for January through March and the ACC for October through December. Generally, the ACC in former group are generally smaller than those in the latter group. Furthermore, their temporal trends disagree between these two groups. The former group first shows a slight decreasing trend until 1995 and increases gradually from then on until about 2005. After that, it is relatively stable over time. While the latter group show relative large difference within the group regarding the magnitude and temporal trend. The ACC for November is relatively stable in time series. While a slight decreasing trend can be observed in ACC for October until 1990 and form then on an increasing trend appear until 2000. After that, it keeps nearly at a constant. By contrast, the ACC for December first show an increasing trend until 1998 and then present a decreasing trend until 2002, form which it keeps an increasing trend until the end. The ACC for AVHRR raw dataset is over 0.86 for all months and even above 0.90 for the months from October through December. By contrast, AVHRR gap-filled dataset shows a constantly lower ACC, which is distributed above 0.83 but below 0.93 for all months. It is noteworthy that the ACC after 2000 are generally larger and more stable than those before it on the monthly scale (Fig. 4(a-d)), indicating the better accuracy and consistency of the later satellite platforms after 2000. Despite the difference in magnitude of ACC between the two datasets, their temporal trends of ACC agree very well for all months. Regarding the comparison of ACC between case2 and case7, their temporal variation characteristics agree well. But their magnitudes show little differences. The AVHRR raw snow dataset shows slightly higher ACC in case2 and than 7, and it is

especially true before 2000. While the AVHRR gap-filled snow dataset shows slightly higher ACC in case 7 than case2 before 2000.
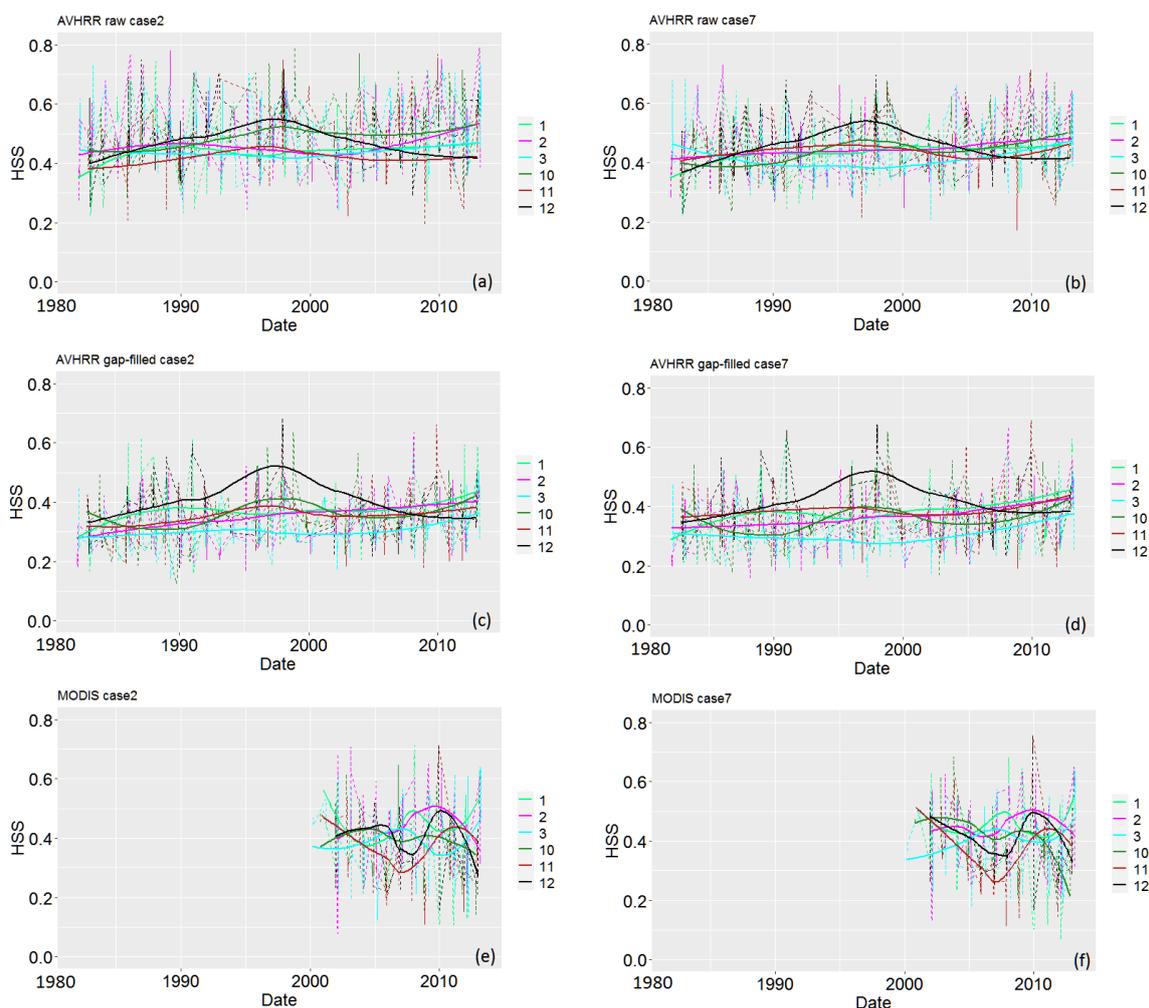

Figure 4. The temporal behavior of ACC for AVHRR raw, AVHRR gap-filled, and MODIS snow data in different months during the snow season.

A similar behavior can be found in the results of HSS derived from AVHRR snow datasets (Fig. 5(a-d)), where a visible divergence appears between the months from January through March and the months from October through December regarding their temporal variation characteristics. The former show relatively smaller temporal variation than the latter throughout the time series. But their magnitudes show relative large difference from month to month. Generally, the HSS is the smallest on March and largest on December for most of the time, and this is particularly true on the results of AVHRR gap-filled snow dataset (Fig. 5(c-d)). While the HSS for other months basically located between the minimum and maximum. This result indicate that during the period of snow ablation, the skill (the correct proportion over the number correct by

chance) decreases. This is understandable since snow cover is often thin and discontinuous during this period. Therefore, the point scale measurement at in situ sites may not be representative of the AVHRR GAC pixel. It is noticeable that from 2005,

375   the HSS of December is even smaller than other months. Instead, the HSS of January is the largest for AVHRR gap-filled dataset and the HSS of October and February is largest for AVHRR raw snow dataset. This result may be associated with the shift of snow cover phenology under the context of global warming. Overall, the HSS between different months is relatively concentrated at the beginning and then spread until 1998 from when they tend to concentrate again until the end. Regarding the comparison of HSS between two datasets, AVHRR raw snow dataset generally display larger HSS than the gap-filled

380   snow dataset, but their temporal variation trend basically agree. As for the comparison between the two cases, most months show similar magnitude and temporal variation trend. But the HSS on November show relative large difference on magnitude for both dataset. This may be attributed to the large surface heterogeneity during this month as the snow cover is more likely to be shallowness and patchiness.

385 **Figure 5**. The temporal behavior of HSS for AVHRR raw, AVHRR gap-filled, and MODIS snow data in different months during the snow season.
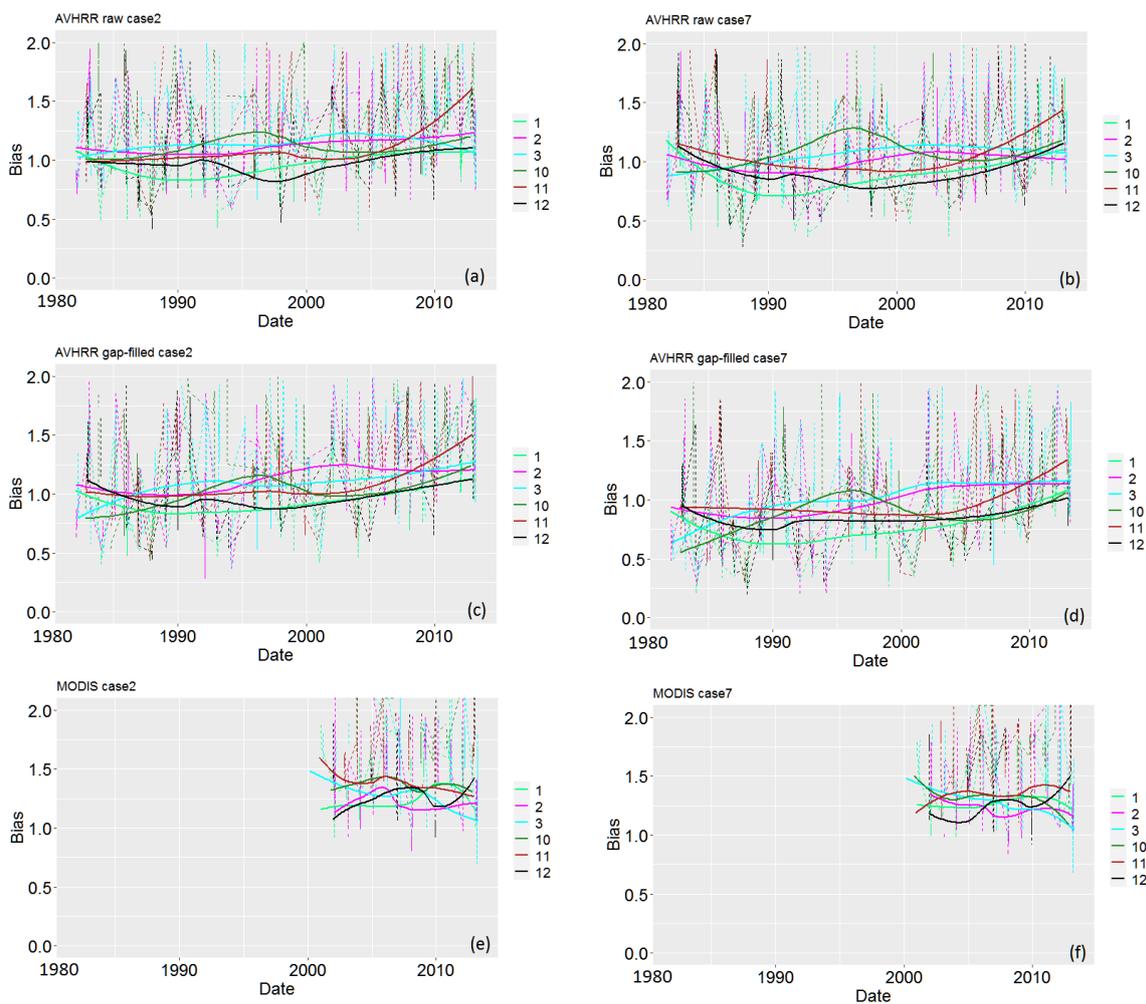

Figure 6. The temporal behavior of Bias for AVHRR raw, AVHRR gap-filled, and MODIS snow data in different months during the snow season.

390   Similarly, both the magnitude and the temporal variation characteristics of Bias are not consistent from month to month Fig. 6(a-d). Specifically, the temporal variation of Bias from October through December are generally larger than that of January through March. Although the Bias on November is relatively stable before 2000, a sharp increase appear after that. Regarding the magnitude of the Bias, a visible break can be seen around the year 2005. After 2005, all months show Bias larger than 1 for both datasets in case2 (Fig. 6(a, c)), although the cause of the overestimation of AVHRR snow dataset after

395 2005 is still unknown. Furthermore, the Bias on February and March are relatively stable and even show a slight decreasing trend. While the Bias of other months all present an increasing trend. Before 2005, no systematic bias is found given that the Biases of different months are either slightly larger or smaller than 1. To be specific, the snow cover during December and
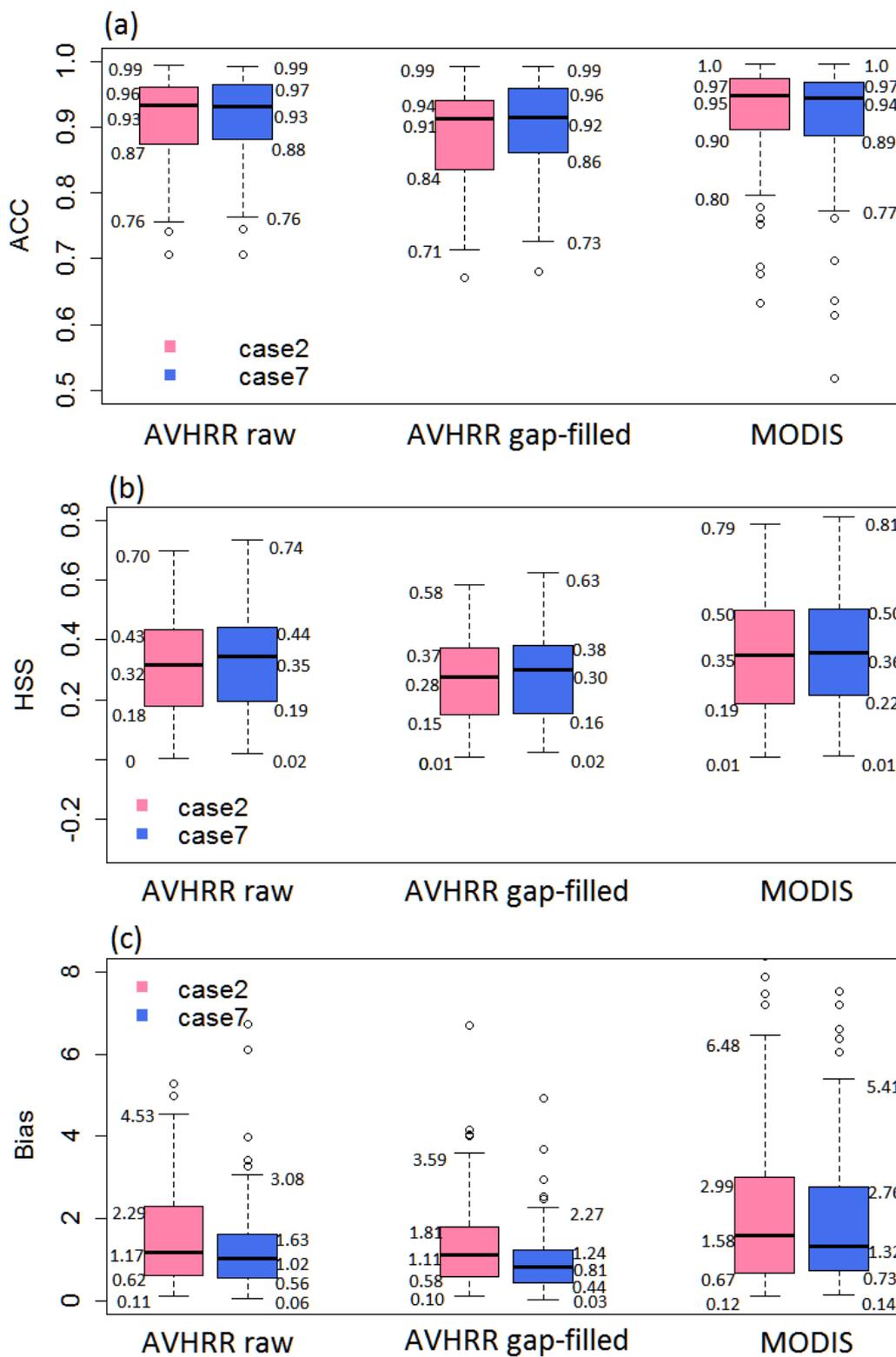
January are underestimated with Bias consistently smaller than 1, while the Bias on November is very close to 1, but the snow cover on February and March seems to be overestimated with Bias larger than 1. This result is understandable because during December and January, snow coverage tends to increase and in situ sites are more likely to be covered by snow, but there is a very strong possibility that the AVHRR GAC pixel contain large areas not covered by snow due to its large spatial coverage. By contrast, during February and March when snow cover is patchy, AVHRR GAC data is more able to detect snow than in situ "point" observation due to its large pixel coverage. From Fig. 6(a-d), it can also be found that both the magnitude and temporal variation characteristics of Bias are not complete coincident between two cases, indicating that the pixel(s) used is an import factor to be considered when estimating Bias.

From the results above, it can be concluded that AVHRR GAC snow dataset performs different at different times of the year, which is mainly related to different amount of snow in HKH. Generally, the agreement between AVHRR snow datasets and in situ data is found to be better for October through December than January through March, but their differences reduce after 2000. It is important to remember that the skill of snow detection decreases during the period of snow ablation. In particular, the AVHRR snow dataset obviously overestimates snow cover in March and February, but underestimates snow cover in December and January.

4.1.3 The spatial divergence of quality indicators

Fig. 7 show the boxplots of validation metric derived from each in situ site, with the aim of revealing their spatial variability. It can be observed that the spatial variability of these validation metrics is widely existed given their dispersed distribution. The maximum of ACC even reach 0.99 for both AVHRR raw and gap-filled snow datasets, while the minimum values are close to 0.7 (Fig. 7a). Similarly, HSS also shows dispersed distribution for both AVHRR snow datasets. The raw dataset ranges from 0 to about 0.7 while the gap-filled dataset ranges from 0 to about 0.6 (Fig. 7b). Likewise, the minimum Bias is located around 0.1 for both AVHRR datasets, while their maximum values range from 2.3 to 4.5, depending on the satellite datasets and different cases (Fig. 7c). These results are understandable because the performance of satellite snow datasets are affected by many factors.

AVHRR raw snow dataset produces better results than gap-filled datasets regarding the values of ACC and HSS, which are consistent with the results in temporal dimension (Fig. 3). The former generally show larger ACC and HSS than the latter in the same case (Fig. 7(a-b)). But the Bias shows a different behavior. AVHRR raw snow dataset tends to overestimate with values deviated from 1 in case2. By contrast, the gap-filled snow dataset tends to underestimate with values deviated from 1 in case7. But the raw snow dataset in case7 and gap-filled snow dataset in case2 show similar values around 1 as well as similar distributions (Fig. 7c). This result indicates again that the effect of which pixel(s) are used cannot be ignored in assessing the Bias of satellite snow dataset.

18

(a)

(b)

(c)

Figure 7. The boxplots of ACC, HSS, and Bias for AVHRR raw, AVHRR gap-filled, and MODIS snow data throughout the
430 sites. The numbers around the plots indicate the minimum, lower quartile, median, upper quartile, maximum of the boxplots,
respectively.

Despite the awareness of spatial variability of these validation metrics, the degree of variability depends on satellite
dataset, validation cases, and metrics. The ACC of AVHRR gap-filled snow dataset is more divergent than AVHRR raw
snow dataset, but the opposite is true for HSS and Bias. It is obvious that the 1-pixel case show more variable Bias than the
435 3-pixel case, but this is no significant difference between the two cases for ACC and HSS.
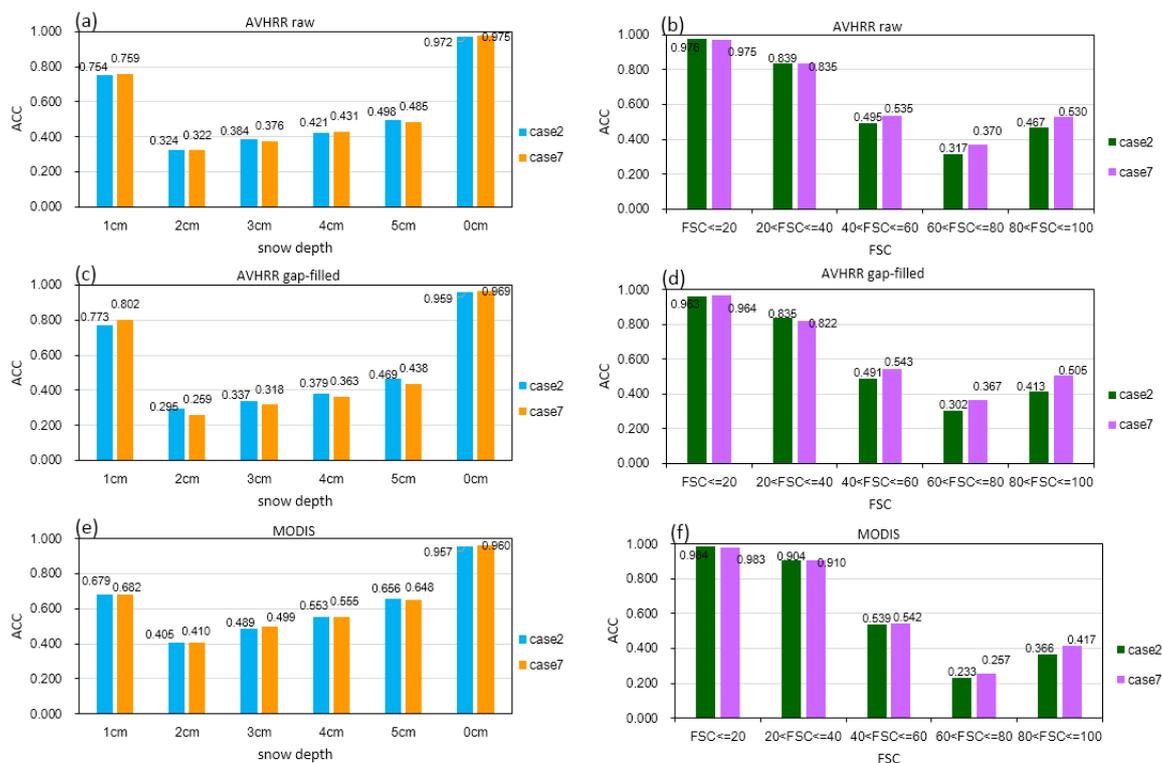
4.1.4 The potential influential factors on accuracy

Following early study (Klein and Barnett, 2003), the effect of SD on the accuracy of satellite snow datasets was evaluated
(Fig. 8(a, c, e)). Observed SD was divided into six categories: SD=0cm, 1cm, 2cm, 3cm, 4cm and 5cm. It is obvious that the
ACC of three satellite snow datasets show similar responses. The highest ACC occurred when SD=0 cm, which is followed
440 by SD=1 cm. When SD>=2 cm, the ACC decreases significantly. The threshold of 2 cm which transform in situ SD
measurements to snow-covered or snow-free information is partly responsible for this result. Another cause of this
phenomenon is the representativeness of the point scale in situ observation, compared with satellite observation on a larger
pixel scale. When SD was less than 2cm, it is more likely that snowfall events only occurred over a limited area of the
satellite pixel. In this condition, satellite snow datasets are more likely to classify the pixel as snow-free, which would
445 increase the agreement between satellite and in situ observations. Despite the decrease of ACC when SD>=2 cm (compared
to SD<2cm), the ACC of various snow datasets clearly show an increasing trend with increasing SD. It is understandable
since with increasing SD, satellite pixel is more likely to be entirely covered by snow, and the agreement between satellite
and in situ observations, as a result, increases. In general, SD was shown to affect the overall agreement of satellite snow
datasets, and their accuracies increase with increasing SD in the situation where in situ site indicates snow-covered
450 information, which is in line with previous studies (Zhou et al., 2013; Wang et al., 2009).

The effect of FSC on the accuracy of satellite snow datasets was checked in Fig. 8(b, d, f). FSC was grouped into five
categories using the ranges of 0-20%, 20%-40%, 40%-60%, 60%-80%, and 80%-100%, respectively. Likewise, the ACC of
three satellite datasets shows similar response to FSC. The highest ACC was found when FSC<=20%, followed by 20 %<
FSC<=40%. This is also partly caused by the threshold (FSC≥50%) applied to FSC maps to transfer fractional to binary
455 snow information and partly caused by the spatial representativeness of in situ sites. When only a small part of the pixel is
covered by snow, in situ sites is more likely covered by very thin snow or not covered by snow. Consequently, in situ sites
are more likely indicate snow-free, which increases the agreement between in situ and satellite observations. In the situation
of 40 %< FSC<=60%, ACC decreases significantly. This occurs because part of satellite data in this group indicate snow
cover using the threshold (FSC≥50%), but there appears a strong possibility that in situ site is not covered by snow or only
460 covered by very thin snow. For the case of 60%<FSC<=80%, all satellite data in this group indicate snow cover but there
remains a very real risk that in situ site is not covered by snow or only covered by very thin snow. As a result, the agreement

between them further decreases. With the further increase of FSC, the possibility that in situ sites indicate snow cover also increases. Thus, ACC increases in the situation of 80 %< FSC<=100. From these results, it is concluded that FSC affects the overall agreement between satellite snow datasets and in situ observations. In the condition where satellite data indicates

465 snow-free, ACC decreases with increasing FSC. By contrast, in the case of satellite data indicating snow-cover, ACC increases with increasing FSC. Nevertheless, it is important to note that the variations of ACC with snow depth and FSC are related to the threshold adopted for transferring SD and FSC to snow-cover or snow-free information.

Figure 8. The variation of ACC with snow depth (left) and FSC (right) for AVHRR raw, AVHRR gap-filled, and MODIS
470 snow data throughout the sites during the experimental period.

Fig. 9 presents the distribution of ACC for three satellite snow datasets against in situ site observations over forest, savannas, grasslands, barren land, and cropland. Although the ACC over cropland is high enough (0.97 and 0.96 for AVHRR raw and gap-filled snow datasets, respectively), this land cover type is not considered in the following analysis, because there is only one in situ site located in cropland area.
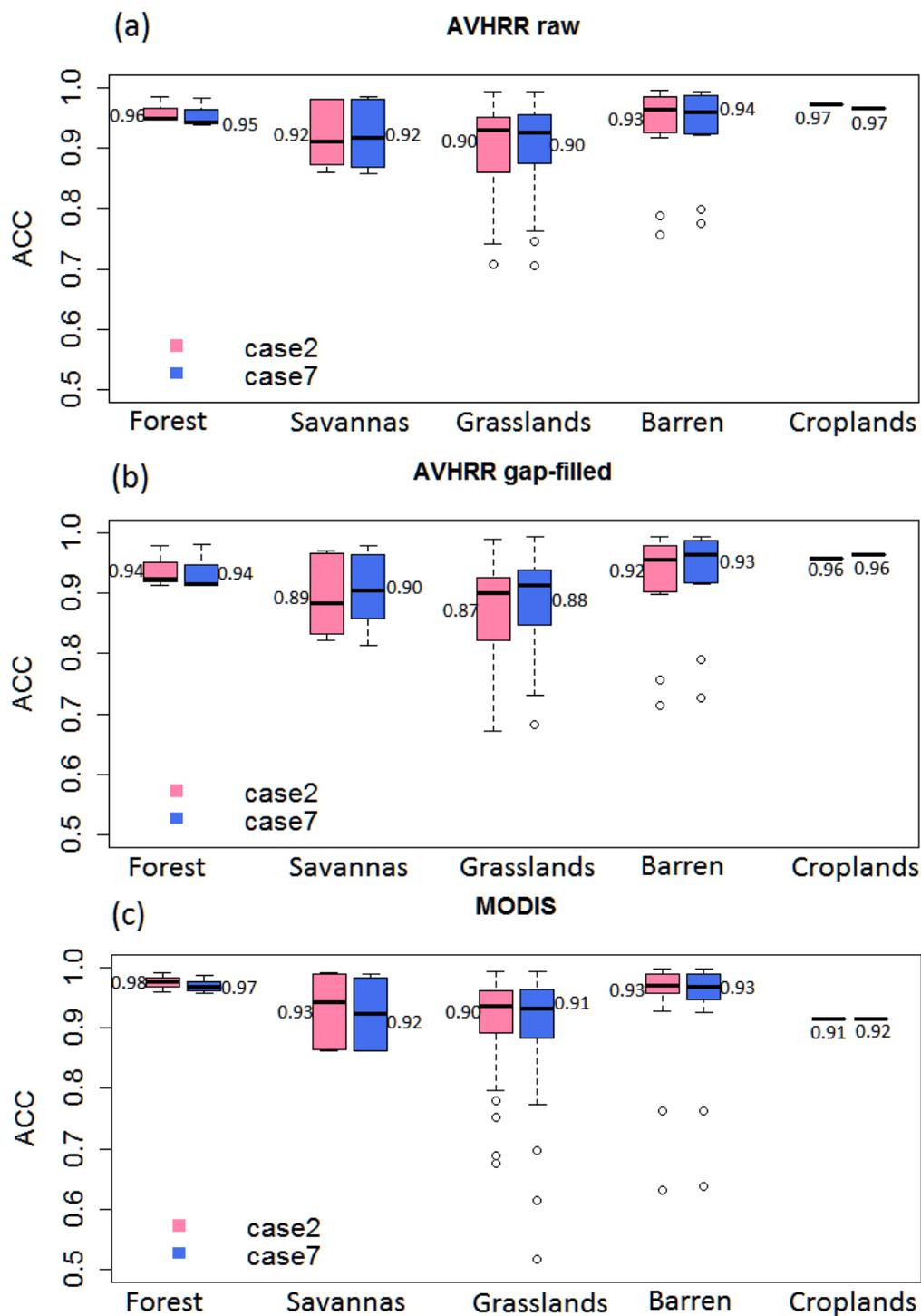
(a) **AVHRR raw**

(b) **AVHRR gap-filled**

(c) **MODIS**

Figure 9. Boxplots of ACC from the direct comparison with *in situ* site observations over different land cover types for AVHRR raw, AVHRR gap-filled, and MODIS snow data. The numbers around the plots are the mean values of the corresponding boxplots. There are 17, 17, 66, 15, 1 sites for forest, savannas, grasslands, barren, and croplands, respectively.

Generally, ACC shows similar responses to land cover type between AVHRR raw and gap-filled snow datasets (Fig.
480 9(a-b)). The highest agreement of these land cover types is found in forest, with mean value larger than 0.94 for both datasets in two cases. Furthermore, the ACC over forest is centralized distributed. This is unexpected given the well-known issues of identifying snow in forested areas using optical satellite data. One possible reason is the sampling technique of in situ site and precision within the forest. The other reason may be that the forest in this region is not very dense. The land cover type followed is barren land and savannas. The former show mean ACC of 0.93 and 0.94 in case2 and case7 respectively for
485 AVHRR raw snow dataset, and mean ACC of 0.92 and 0.93 in case2 and case7 respectively for AVHRR gap-filled snow data. While the latter present mean ACC of 0.92 in both case for AVHRR raw snow dataset, and mean ACC of 0.89 and 0.90 in case2 and case7 respectively for AVHRR gap-filled snow dataset. Moreover, the distribution of ACC over savannas is more dispersed than that of barren land. The land cover type ranked last in terms of ACC is grassland, where mean ACC of AVHRR raw dataset are 0.90 in both cases for AVHRR raw snow dataset, and 0.87 and 0.88 in case2 and case7 respectively
490 for AVHRR gap-filled snow dataset. The wide spread of ACC over grassland is partly caused by the much larger number of in situ sites over it.
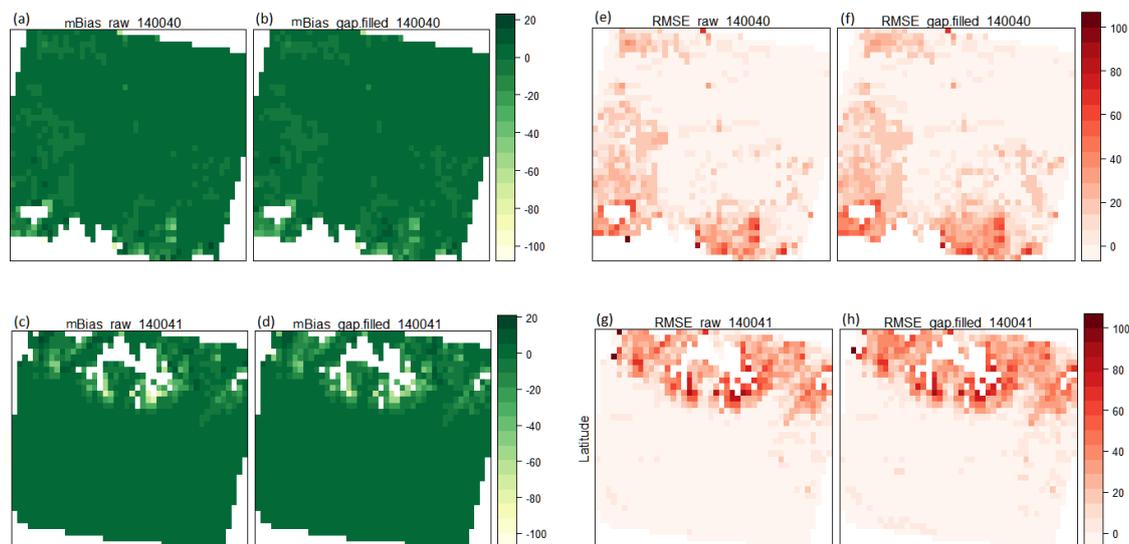
4.2 The comparison with high-resolution data

High resolution images were used here to avoid the limitations of in situ observations and allow a more precise description of errors than binary metrics. The spatial accuracy was assessed on the pixel basis based on Landsat5 TM data time series
495 over the areas covered by 'P140-R40/41' (Fig. 10). Large spatial variations can be observed between different areas in both mBias and RMSE. Here, histogram was employed to show the distribution of mBias and RMSE in different levels (Fig. 11), and the overall accuracy indicators were summarized in Table 4.
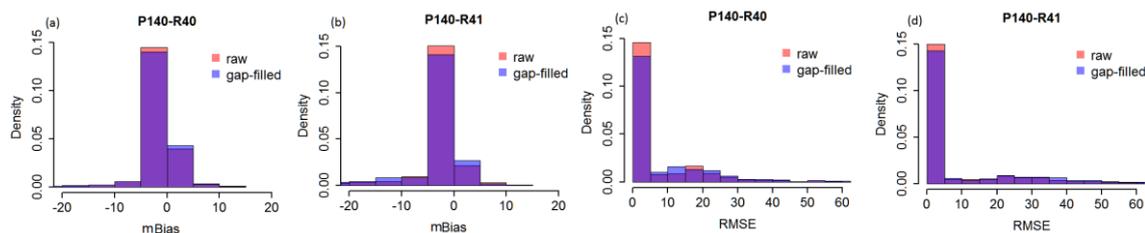
It can be seen that the mBias distribution is non-symmetrical towards negative values (Fig. 11(a-b)). Although the maximum value of mBias can even larger than 20%, most of them are distributed within the range of ±5% with the overall
500 mBias of -1.065% and -1.78% for AVHRR raw and gap-filled snow datasets, respectively. This result indicates that AVHRR snow datasets, on average, slightly underestimate snow covered areas with regards to the Landsat5 TM data. This can be explained by the fact that direct coarse resolution FSC is more likely to be lower than the FSC aggregated from high-resolution FSC. Because high-resolution data is able to pick up snow in one pixel, which is too little to create enough snow signals in coarse resolution pixels but will show up in the aggregated FSC (Singh et al., 2014; Jain et al., 2008). Despite the
505 wide spread of RMSE (Fig. 11(c, d)), the RMSEs gained from pixel-to-pixel comparison are mostly distributed within the range of 0-5% with the overall values of 12.593% and 16.172% for AVHRR raw and gap-filled datasets, respectively. It is thus concluded that AVHRR GAC snow datasets are accurate enough with good closeness with reference dataset estimated from high-resolution Landsat5 TM data. Despite the consistent distribution of mBias and RMSE between these two tiles (Fig.

11), the accuracies of two datasets seem better over "P140-R40" than "P140-R41", indicated by the RMSE and mBias in
Table 4. The slight difference of their accuracies is mainly related to other factors, such as land cover types (Fig. 1b) and
topography. AVHRR raw and gap-filled snow datasets show similar results, with raw dataset slightly more accurate with
smaller RMSE of 12.593% (vs. 16.172%) and mBias of -1.065% (vs. -1.78%) as well as a higher R of 0.462 (vs. 0.456).

Figure 10. The spatial accuracy (mBias and RMSE) of AVHRR raw and AVHRR gap-filled snow data with Landsat-5 TM
snow data on the pixel basis by combining all the valid data in the time dimension for "P140-R40" and "P140-R41",
respectively.

Figure 11. The distribution of mBias and RMSE for "P140-R40" and "P140-R41", respectively. For histograms, the heights
of the bars indicate the density. In this case, the area of each bar is the relative frequency, and the total area of the histogram
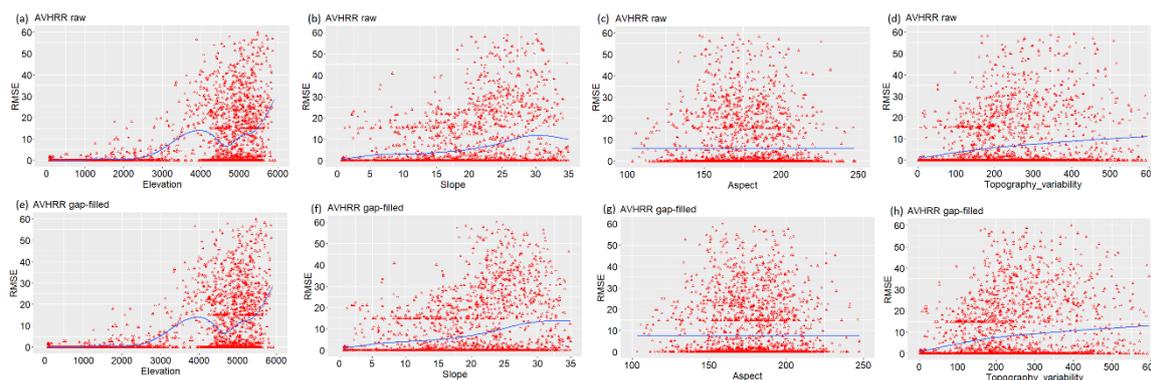is equal to 1.

Table 4. Summary of accuracies of AVHRR raw and AVHRR gap-filled snow data with Landsat5 TM snow over "P140-R40" and "P140-R41".

	AVHRR gap-filled snow			AVHRR raw snow		
	RMSE	mBias	R	RMSE	mBias	R
P140-R40	13.808	-1.018	0.465	11.004	-0.642	0.457
P140-R41	18.352	-2.59	0.453	14.139	-1.53	0.469

Overall	16.172	-1.78	0.456	12.593	-1.065	0.462

Since there is a slight difference of accuracies over different regions (Table 4), the influential factors on the accuracy of AVHRR GAC snow dataset were explored here. Fig. 12 displays the scatterplots of RMSE against elevation, slope, aspect, and topographical variability for both datasets. It is obvious that RMSEs of two datasets show similar responses to these factors. Generally, a considerable number of RMSEs increase with elevation, although there are still a large number of RMSEs approaching 0 (Fig. 12(a, e)). This is unexpected as coarse-pixel satellite snow products generally perform better in higher elevations due to the continuous and thick snow cover. The larger RMSEs in higher elevations are partly caused by the large values of FSC themselves and partly caused by the cloud effects given that the probability of cloud rises with rising altitude in mountain areas. When slope is considered as a criterion, AVHRR GAC snow datasets tend to give better results over the areas with smaller slopes (Fig. 12(b, f)). This is understandable because satellite data suffer from topographic effects in steep mountain area. Regarding the effect of aspect, although there is not a clear trend of RMSEs with aspect, the large RMSEs are mostly located around 180 ° (south facing slope) (Fig. 12(c, g)). This is attributed to the fact that during winter months, south facing slopes receive significantly more radiant energy, providing unfavorable environment for snow accumulation. Thus, snow cover in south facing slopes is more likely to be shallowness and patchiness, reducing the accuracy of AVHRR GAC snow datasets. As seen from Fig. 12(d, h), there is an increasing trend of RMSE against topographical variability, indicating its significant effect on the accuracy of AVHRR GAC snow datasets. This is because the rugged relief can lead to shadowing effects, resulting in different degree of surface information loss between high-resolution satellite data and coarse resolution satellite data.

Figure 12. The variation of RMSE with elevation, slope, aspect, and topographical variability. The results have combined data over "P140-R40" and "P140-R41". The blue line denotes the fitting line.

Another influence factor this is known to impact satellite snow cover detection is land cover type. The boxplots of RMSE of pixels over different land cover types are given in Fig. 13. The distribution of RMSE responses similarly to land cover types between the two datasets. It is obvious that AVHRR GAC snow datasets perform better in some land covers than in others. Specifically, in forest, savannas, croplands, and built-up lands, the errors in snow mapping are very low with RMSE basically concentrated near 0. In grassland, errors become much larger. And the largest errors occur in barren lands.

25

Furthermore, RMSEs show larger spatial variability over grassland and barren lands, demonstrating that the accuracy of AVHRR GAC snow datasets over these two land cover types is more likely to be affected by other factors. One reason for the wide distribution of RMSEs over these two land cover types is their larger number of pixels than other land cover types. Combining the results of Fig. 13 and Fig. 9, it can be found that validation using different and independent reference datasets indicates a good performance of AVHRR GAC snow datasets in forests, savannas, and croplands, and a slightly lower accuracy in grasslands. The results based on in situ data and high resolution data basically agree with an exception in barren lands. This is partly due to significantly different number of samples in the boxplots and partly due to the different mechanism of calculation of reference data as well as validation metrics.
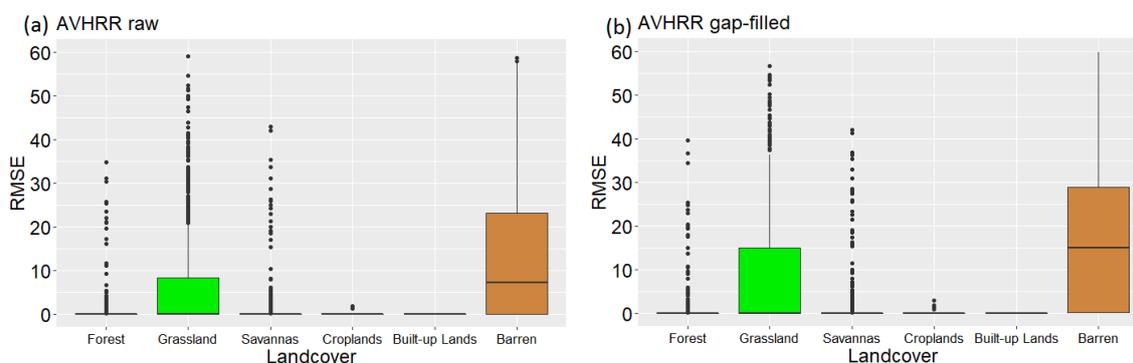

Figure 13. The variation of RMSE with land cover types. The results have combined data over "P140-R40" and "P140-R41". There are 326, 1211, 446, 202, 6, 605 pixels for forest, grassland, savannas, croplands, built-up lands, and barren, respectively.

In order to describe the variation of accuracy of AVHRR GAC snow datasets throughout the time series (1985-2012), the day-by-day results of RMSE, mBias, and R are therefore presented in Fig. 14. The RMSE and mBias derived in scene-by-scene comparison still indicate a reasonable accuracy of both datasets, with the former distributed between 5% and 20% and the latter distributed between 0 and -5% (Fig. 14(a and b)). This result of mBias has confirmed once more that AVHRR GAC snow datasets are generally underestimated in these areas. Both the temporal variation trends of RMSE and mBias indicate that accuracies of both datasets gradually increase over time. The RMSEs drop from about 20% at the beginning to about 10% in the end. And the mBiases drop from about -2.5% at the beginning to about 0 in the end. The results of mBias disagree with those based on in situ site (Fig. 3c). The disagreement is partly caused by different mechanism of calculation of reference data and validation metrics, and partly caused by the effect of spatial representativeness errors of in situ sites.

The day-by-day results reveal that AVHRR raw snow dataset is more accurate than the gap-filled snow dataset. Both the magnitude of RMSE and mBias of the former are consistently smaller than those of the latter in the whole time series (Fig. 14(a and b)). This result agrees well with above results. The difference of RMSE and mBias between the two areas ("P140-R40/41") stand out again in the day-by-day comparison. Both datsets show better performance over the area covered by "P140-R40" than those over the area covered by "P140-R41". The former consistently display smaller RMSE and mBias than the latter for both datasets in the whole time series. This result demonstrate again that in the validation based on high-

26

575 resolution datasets, validation results depends the location of the area. Therefore, larger areas are desired in this kind of
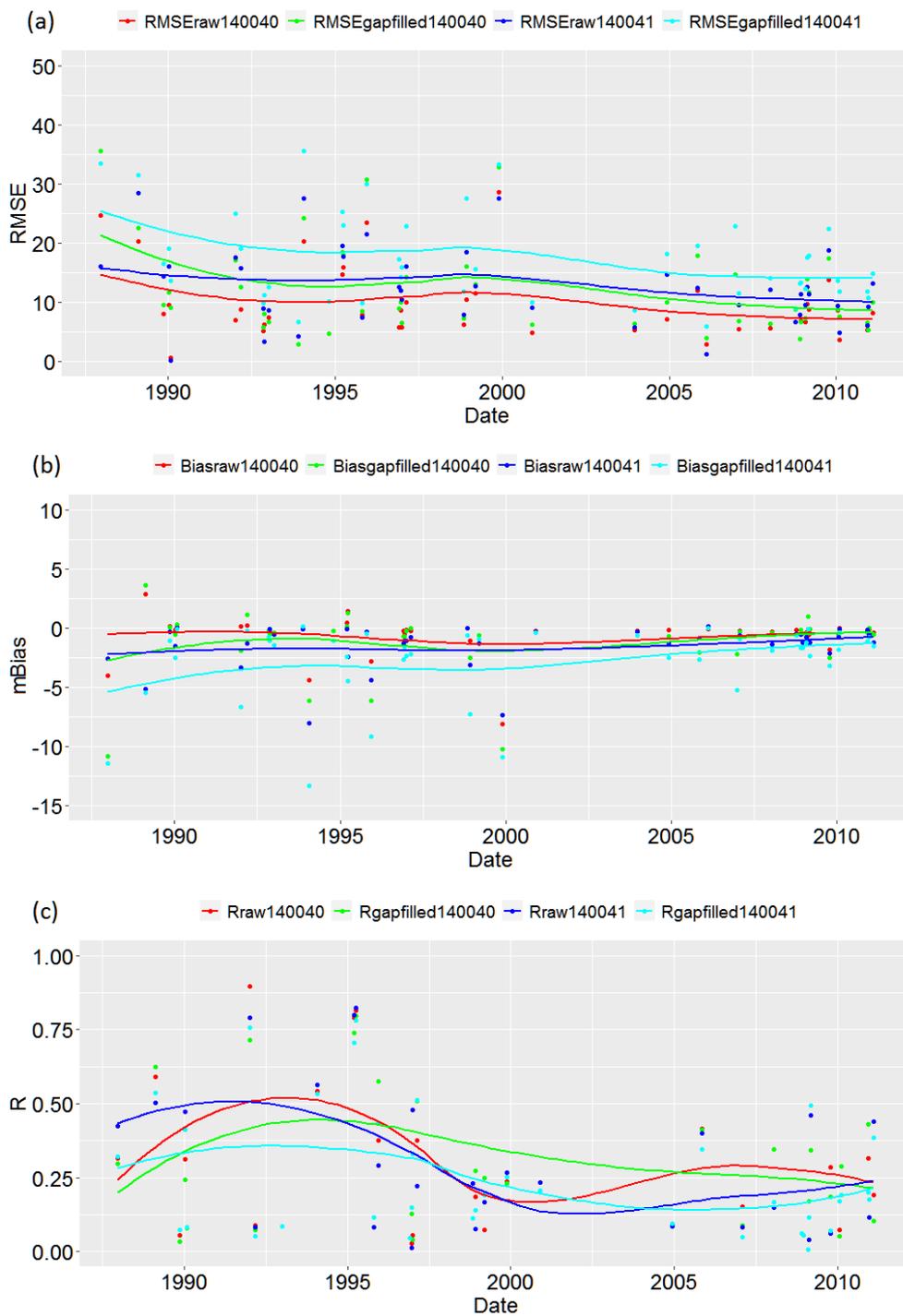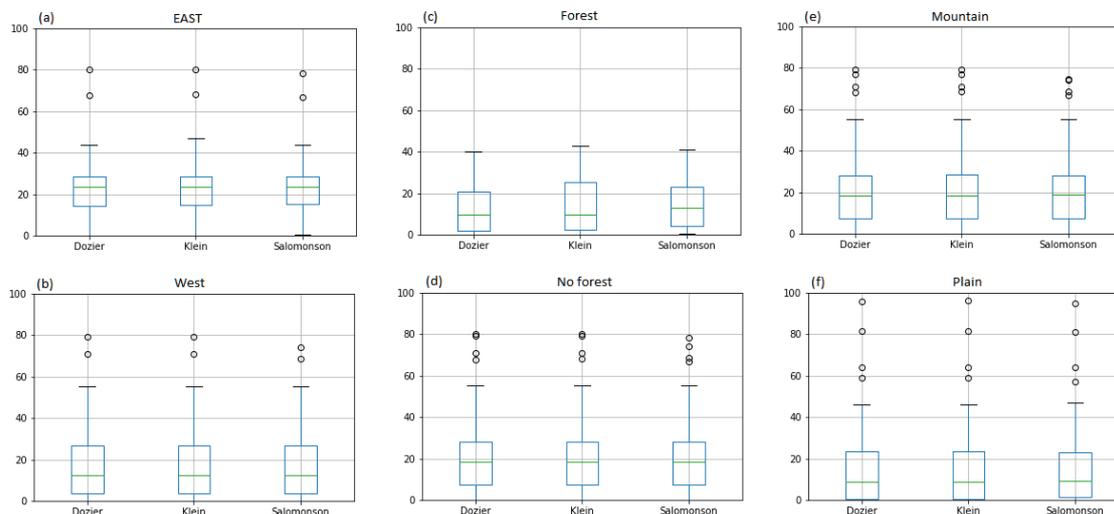validation.

Figure 14. The RMSE, mBias, and R of AVHRR raw and AVHRR gap-filled snow data based on Landsat 5 TM snow data as a function of all the available dates in "P140-R40" and "P140-R41", respectively.

580   Unlike RMSE and mBias, the correlation coefficient R between AVHRR GAC snow dataset and Landsat5 TM data show a relatively large temporal variation for both datasets over the two areas. On average, the R after 2000 are significantly lower than those before it. And it is not very high ($0.125 \le R \le 0.5$ depending on the time) over these two areas, indicating that the consistency of the spatial distribution characteristics of FSC is not very good. A relative large difference can be observed in the magnitude of R between these two areas, and this is especially true for AVHRR gap-filled snow dataset. The

585   area covered by "P140-R40" generally shows larger R than the area covered by "P140-R41", again demonstrating that the validation based on high-resolution datasets depends on validation region. Therefore, a more comprehensive validation by expanding the comparison area is needed.

In order to give more comprehensive validation results based on high-resolution datasets, the comparison between AVHRR snow datasets and Landsat data was also carried out over the whole extent of HKH. The RMSE, mBias, and R in

590   different conditions are summarized in Table 5. RMSE is generally less than 25% for both datasets in different conditions with an overall RMSE of 22.91% and 24.65% for AVHRR raw and gap-filled snow datasets, respectively. The mBias still indicates underestimation of AVHRR snow datasets, and the degree of underestimation is smallest in forest, but with small R of 0.32 and 0.42 for AVHRR raw and gap-filled snow datasets, respectively. The mountain areas show a larger mBias and a slightly smaller R relative the to the plain areas. The overall mBias are -6.74% and -7.56% and the overall R are 0.83 and 0.8

595   for AVHRR raw and gap-filled snow datasets, respectively, again demonstrating the slightly better accuracy of AVHRR raw snow datset.

Table 5. Summary of accuracies of AVHRR raw and AVHRR gap-filled snow data with Landsat5 TM snow data over the whole HKH.

	RMSE(gap-filled)%	mBias(gap-filled)%	R(gap-filled)	RMSE (raw)%	mBias (raw)%	R (raw)
Plain	21.92	-5.47	0.87	20.22	-5.16	0.89
Mountain	24.98	-7.82	0.79	23.25	-6.96	0.82
Forest	23.23	-4.55	0.42	19.42	-0.92	0.32
Global	24.65	-7.56	0.8	22.91	-6.74	0.83

Figure 15. The boxplots of RMSE for AVHRR gap-filled snow data based on Landsat5 TM snow data derived with the methods proposed by Dozier and Painter (2004), Klein et al. (1998), Salomonson and Appel (2006) respectively in different conditions over the whole HKH.

In order to show the effect of FSC retrieval method based on high-resolution data, a simple comparison between different methods was made by summarizing RMSEs in the form of boxplots (Fig. 15). Several conditions including east vs. west, forest vs. no forest, mountain vs. plain, were considered here. Similar results in terms of the magnitude and distribution of RMSEs can be found between different methods. It is interesting that the eastern area shows larger RMSEs than western areas, but the latter shows a larger spatial variability (Fig. 15(a and b)). Consistent with previous results, forest shows smaller RMSEs and spatial variability compared to no forest areas. As expected, the mountain areas show larger RMSEs and spatial variability than plain area.

4.3 The performance relative to MODIS snow product

The performances of AVHRR GAC snow datasets were compared with MOD10A1 V006 in terms of their overall accuracy relative to in situ sites, temporal consistency of accuracy over time, spatial variability of accuracy, the accuracy under different conditions (SD, FSC, and land cover types), as well as the difference of their absolute values.

As indicated by Fig. 2, the ACC and HSS of MODIS snow product are both located between AVHRR raw and gap-filled snow dataset in case2 and case7, but closer to AVHRR raw dataset. It is thus clear that AVHRR raw dataset performs slightly better while AVHRR gap-filled dataset performs worse than MODIS snow product with respect to the agreement with in situ sites and the algorithm performance (skill). Nevertheless, both AVHRR raw and gap-filled datasets show distinct advantage over MODIS snow product when Bias is focused. MODIS snow product presents serious overestimation errors in two cases while AVHRR snow datasets show unbiased results in the same cases.
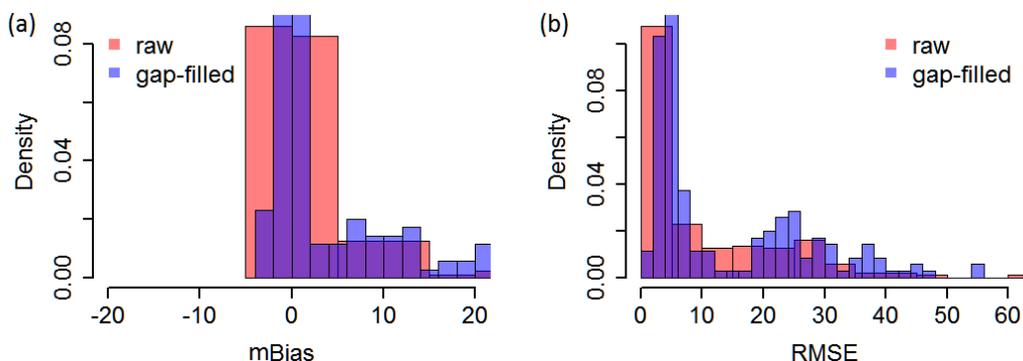
29

620   Regarding the temporal behavior of three snow datasets (Fig. 3), the ACC of MODIS snow product reveals slightly
larger variability during the year 2000-2013, but its value is still basically distributed between AVHRR raw and gap-filled
datasets (Fig. 3a). Unlike AVHRR snow datasets, the ACC of MODIS snow in two cases are almost equal over time series,
meaning that its accuracy is less affected by spatial heterogeneity or geometric mismatch. However, the HSS of MODIS
snow show large difference between case2 and case7. The former show larger temporal variations than AVHRR snow
625   datasets, while the latter keep constant over time (Fig. 3b). The magnitude of HSS of MODIS basically located between
AVHRR raw and gap-filled datasets. However, the Bias of MODIS snow product exhibits different temporal variation
characteristics compared to ACC, which almost keep constant over time (Fig. 3c). Therefore, we can conclude that AVHRR
snow datasets display a more stable behavior regarding the accuracy but less stable regarding the Bias than MODIS snow
products over the same period. As for the performance of MODIS snow products in different months of the year, the inter-
630   annual variability of ACC (Fig. 4(e and f)), HSS (Fig. 5(e and f)), and Bias (Fig. 6(e and f)) all become more significant
compared to Fig. 3(a-c). Furthermore, the temporal variations of these metrics on the monthly scale are generally more
significant for MODIS snow product than AVHRR snow datasets (Figs. (4-6)).

    When it comes to the spatial variability of the performance of MODIS and AVHRR snow datasets, the ACC of the
former are more concentrated than the latter (Fig. 7a). Nevertheless, the HSS and Bias of the former show a more variable
635   distribution than the latter (Fig. 7(b and c)). Hence, we can believe that the accuracy of AVHRR snow datasets shows a
larger spatial variability than MODIS snow product, but the HSS and Bias of AVHRR snow exhibit a smaller spatial
variability than the latter.

    Seen from Fig. 8(a, c, e), we can find that the accuracy of AVHRR snow datasets is larger than MODIS snow product
when SD<=1cm, but consistently smaller than MODIS snow at each SD when SD>=2cm. This means that in snow-free or
640   very thin snow conditions, AVHRR snow datasets are less misclassified than MODIS snow product. By contrast, in snow-
covered condition, although the three datasets all reveal an increase in ACC with increasing SD, MODIS snow product is
more reliable and correctly classified. The discrepancies between them are mainly resulted from the different spatial scale of
the pixel. By comparing Fig. 8f with Fig. 8(b and d), it can be found that accuracies of AVHRR snow datasets are slightly
lower than MODIS product for each level of FSC when FSC<=60%, but larger than MODIS product when FSC>60%. This
645   phenomenon is related to the different degrees of spatial representativeness of in situ sites relative to different pixel scales.

    MODIS snow product presents a similar response to land cover types as AVHRR snow datasets (Fig. 9). It performs
best over forest with a high agreement even up to 0.97 and very narrow distribution, followed by barren land and savannas.
However, over grasslands, its agreement decreases, together with a very dispersed distribution (Fig. 9c). Generally, AVHRR
snow datasets and MODIS snow product present comparable ACC over each land cover type.

Figure 16. The distribution of mBias and RMSE derived from the scene-by-scene comparison between AVHRR GAC snow datasets and MODIS snow products over the region "P140-R40/41" throughout the snow season between 2012 and 2013.

Table 6. Overall accuracies of AVHRR raw and AVHRR gap-filled snow data in comparison with MODIS snow over "P140-R40/41".

	AVHRR gap-filled snow			AVHRR raw snow		
	RMSE	mBias	R	RMSE	mBias	R
Overall	19.024	2.832	0.433	14.636	1.564	0.504

In order to show the absolute difference between AVHRR GAC and MODIS snow, RMSE and mBias derived from scene-by-scene comparison over the region "P140-R40/41" throughout the snow season between 2012 and 2013 were presented in the form of histograms (Fig. 16). It can be seen that AVHRR GAC snow datasets are slightly overestimated relative to MODIS snow given the non-symmetrical distribution of mBias towards positive values. But their overall mBiases are very small, with the values of 1.564% and 2.832% for AVHRR raw and gap-filled snow datasets, respectively (Table 6). The RMSEs are mainly distributed within the range of 1-10%, with the overall values of 14.636% and 19.024% for AVHRR raw and gap-filled snow datasets, respectively. Nevertheless, the spatial distribution characteristics of FSC are not in good agreement, given that the R between them are not very high (0.504 and 0.433 for AVHRR raw and gap-filled snow datasets respectively).

5 Conclusion

Snow cover extent is a key component of climatic variability. This paper introduces a new version of snow extent dataset derived from AVHRR GAC data and assesses the spatial and temporal accuracy of the dataset over complex terrain. Compared to other AVHRR snow extent products, this dataset is designed to provide global snow extent with consistent performance across the whole suite of AVHRR sensors, which is considered a major step toward a detailed snow climatology on the global scale.

The Cryosphere
Discussions
Open Access
EGU

A comprehensive validation was carried out using different and independent reference datasets. First, the validation was performed with in situ SD measurements and was provided in terms of snow detection or misclassification frequencies covering the timespan (1982-2013), with more emphasis on the sensor-to-sensor consistency. The optimum threshold to transform in situ SD to snow-covered information was determined through sensitivity analysis, which has been proved to be 675   2 cm in this study. With consideration of the effect of spatial heterogeneity and geometric mismatch, two validation cases (1 pixel and 3×3 pixels centered on in situ site) were carried out. Second, Landsat5 TM scenes, with a high spatial resolution, provide absolute snow cover information to investigate small quantitative inconsistencies. 30 years (1984-2013) of Landsat5 TM FSC data for two different regions were used to enable a spatio-temporal appreciation of the AVHRR GAC detection errors. Third, the daily 500m MODIS snow product was evaluated using in situ measurements to be compared with AVHRR 680   GAC snow dataset in order to assess its performance from a relative perspective. Furthermore, their absolute difference was also quantified.

Validated against in situ site observations, the overall ACC of AVHRR raw snow dataset was about 94% for the two cases. The use of temporal filter caused a slight reduction in ACC of AVHRR gap-filled snow dataset, with overall values around 92% in two cases. The ACC of MODIS is situated between AVHRR raw and gap-filled snow data, which equal to 685   93.6% in both cases. AVHRR raw and gap-filled snow datasets are very close in Bias, with values around 1 in two cases. Nevertheless, MODIS snow shows obvious overestimation errors with Bias of 1.740 and 1.613 in case2 and case7, respectively. When validated against Landsat5 TM images over the region of "P140-R40/41", the overall RMSEs are 12.593% and 16.172% and the overall mBiases are -1.065% and -1.78% for AVHRR raw and gap-filled snow datasets, respectively. But the R for two datasets are low (0.462 and 0.456 respectively for AVHRR raw and gap-filled datasets). When the whole 690   HKH was considered, the RMSE, on average, is generally less than 25% for both datasets in different conditions, with the overall RMSEs of 22.91% and 24.65% and R of 0.83 and 0.8 for AVHRR raw and gap-filled snow datasets, respectively.

Regarding the temporal behavior of AVHRR GAC snow datasets, the sensor-to-sensor consistency was found to differ slightly and unsystematically in ACC and HSS throughout the time series. While the consistent slight increasing trend on Bias is noteworthy. By contrast, the validation results based on Landsat5 TM data suggest a better performance with later 695   sensors, as indicated by the temporal variation trends of RMSE and Bias. The disagreement between them is partly related to different mechanism of calculation of reference data and validation metrics, and partly related to the effect of spatial representativeness errors of in situ sites. It is important to point out that the different performance of AVHRR GAC snow datasets in different months is mainly caused by variable amount of snow. Particularly, AVHRR snow dataset obviously overestimates snow cover in March and February but underestimates snow cover in December and January. During the 700   period of snow ablation, the skill of snow detection decreases. The validation results with two independent reference data both show considerable spatial variabilities, indicating the effect of other factors (e.g., SD, FSC, land cover type, elevation, slope, aspect, and topographical variability).

Generally, in the situation of snow cover, accuracy of satellite snow datasets increases with increasing SD and FSC. By contrast, in snow-free condition, accuracy decreases with increasing SD and FSC. In terms of how errors relate to land cover

705 type, validation via two independent reference datasets indicates a good performance of AVHRR snow datasets in forests, savannas, and croplands, and a slightly lower accuracy in grasslands. Higher elevation is more likely to experience larger errors, which is partly caused by the large values of FSC themselves and partly caused by the cloud effects. Better accuracy is generally observed in areas with smaller slopes since the topographic effects become significant in steep mountain areas. The errors in south facing slopes are generally larger than other aspects, because snow cover is more likely to be shallowness

710 and patchiness, reducing the accuracy of AVHRR GAC snow datasets when compared with Landsat5 TM data. The errors of AVHRR snow datasets tend to increase with topographical variability. This is because the shadowing effects caused by rugged relief would result in different degree of surface information loss between high resolution Landsat5 TM data and coarse resolution AVHRR GAC data.

When it comes to the performance relative to MODIS snow products, AVHRR raw dataset performs slightly better

715 while AVHRR gap-filled dataset performs worse than MODIS snow product with respect to the agreement with in situ sites and the algorithm performance (skill). Nevertheless, both AVHRR snow datasets show distinct advantages over MODIS snow product focusing on Bias. Regarding the temporal and spatial behaviors, different results appear in the two dimensions. In the temporal dimension, AVHRR snow datasets display a more stable behavior regarding the accuracy but less stable regarding the Bias than MODIS snow products. But in the spatial dimension, AVHRR snow datasets show a larger spatial

720 variability of accuracy but a smaller spatial variability of HSS and Bias than MODIS snow products. The absolute differences between AVHRR GAC and MODIS snow datasets were still reasonable, with the overall RMSE of 14.636% and 19.024%, mBias of 1.564% and 2.832%, and R of 0.504 and 0.433 for AVHRR raw and gap-filled snow datasets, respectively.

This study represents the first validation of daily AVHRR GAC snow extent in the mountain region. Although the

725 reference datasets (i.e., in situ sites, high resolution satellite data) have their own limitations and flaws, our results still encourage the compilation of a consistent, complete, long time series snow extent dataset from historical AVHRR GAC data. This study characterizes the product performance with distinct accuracy parameters from different perspectives, thus contributes to the ongoing efforts to improve the performance of existing snow products by enhancing our knowledge of the thematic and absolute accuracy of current products.

730 **6 Author contributions**

Xiaodan Wu was responsible for the main research ideas and writing the manuscript. Kathrin Naegeli and Carlo Marin contributed to the data collection. Stefan Wunderle contributed to the manuscript organization. All the authors thoroughly reviewed and edited this paper.

The Cryosphere
Discussions

Open Access

EGU

7 Competing interests

735   The authors declare that they have no conflict of interest.

8 Data availability

The in situ snow depth data are provided by the China Meteorology Administration (CMA) (https://data.cma.cn/en), which are not available to the public due to legal constraints on the data's availability. The Landsat collection 1 Tier 1 data from Landsat 4 to Landsat 8 are publicly available via https://glovis.usgs.gov/. The MOD10A1 dataset are provided by NASA

740   National Snow and Ice Data Center Distributed Active Archive Center (NSIDC DAAC), which is registered here https://doi.org/10.5067/MODIS/MOD10A1.006. It was downloaded using Google Earth Engine via the code presented at https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD10A1. The AVHRR GAC FCDR is registered here doi:10.5676/DWD/ESA_Cloud_cci/AVHRR-PM/V003 and can be accessed via https://public.satproj.klima.dwd.de/data/ESA_Cloud_CCI/CLD_PRODUCTS/v3.0/L3U/AVHRR-PM/ (Stengel et al. 2020).

745   The landing page points to additional documentation and data download sites.

9 Acknowledgements

References

Anderson, K., Fawcett, D., Cugulliere, A., Benford, S., Jones, D., and Leng, R.: Vegetation expansion in the subnival Hindu Kush Himalaya, Glob. Chang Biol., 26(3), 2020.

Arsenault, K. R., Houser, P. R., and De Lannoy, G. J.: Evaluation of the MODIS snow cover fraction product, Hydrol.

755   Processes, 28(3): 980-998, 2014.

Bookhagen, B., and Burbank, D. W.: Topography, relief, and TRMM-derived rainfall variations along the Himalaya, Geophys. Res. Lett., 33, L08405, 2006.

Brown, R. D., and Mote, P. W.: The response of Northern Hemisphere snow cover to a changing climate, J. Climate, 22(8), 2124-2145, 2009.

760   Crawford, C. J.: MODIS Terra Collection 6 fractional snow cover validation in mountainous terrain during spring snowmelt using Landsat TM and ETM+, Hydrol. Processes, 29(1), 128-138, 2015.

Cracknell, A.P.: Advanced very high resolution radiometer avhrr, Crc Press, 1997.

Devasthale, A. et al.: PyGac: An open-source, community-driven Python interface to preprocess nearly 40-year AVHRR Global Area Coverage (GAC) data record, Quarterly, 11, 3-5, 2017.

765   Dozier, J., and Painter, T. H.: Multispectral and hyperspectral remote sensing of alpine snow properties, Annu. Rev. Earth Pl. Sc., 32(1), 465-494, 2004.

Fletcher, C. G., Kushner, P. J., Hall, A., and Qu, X.: Circulation responses to snow albedo feedback in climate change, Geophys. Res. Lett., 36(9), 1–4, 2009.

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X.: MODIS Collection 5
770   global land cover: Algorithm refinements and characterization of new datasets, Remote Sens. Environ., 114, 168–182, 2010.

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, 2012.

Gascoin, S., Grizonnet, M., Bouchet, M., Salgues, G. and Hagolle, O.: Theia Snow collection: high-resolution operational
775   snow cover maps from Sentinel-2 and Landsat-8 data, Earth Syst. Sci. Data, 11(2), 2019.

Guangwei, C.: Summary reports of workshops on biodiversity conservation in the Hindu Kush-Himalayan ecoregion, In Biodiversity in the eastern Himalayas: conservation through dialogue, ICIMOD, 2002.

Hall, D. K. , Riggs, G. A. , Salomonson, V. V. , Digirolamo, N. E. , and Bayr, K. J.: Modis snow-cover products, Remote Sens. Environ., 83(1-2), 181-194, 2002.

780   Hall, D. K., and Riggs, G. A.: Accuracy assessment of the MODIS snow products, Hydrol. Processes, 21(12), 1534-1547, 2007.

Hall, D. K., and Riggs, G. A.: MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid, Version 6. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: https://doi.org/10.5067/MODIS/MOD10A1.006, 2016.

785   Hao, S. , Jiang, L. , Shi, J. , Wang, G., and Liu, X.: Assessment of modis-based fractional snow cover products over the tibetan plateau, IEEE J. Sel. Topics Appl. Earth Observ., PP(2), 1-16, 2018.

Hao, X., Luo, S., Che, T., Wang, J., Li, H., Dai, L., Huang, X., and Feng, Q.: Accuracy assessment of four cloud-free snow cover products over the Qinghai-Tibetan Plateau, Int. J. Digit. Earth, 12(4), 375-393, 2019.

Heidinger, A. K., Frey, R., and Pavolonis, M.: Relative merits of the 1.6 and 3.75 μm channels of the AVHRR/3 for cloud
790   detection, Can. J. Remote Sens., 30(2), 182-194, 2004.

Hori, M., Sugiura, K., Kobayashi, K., Aoki, T., Tanikawa, T., Kuchiki, K., Niwano, M., and Enomoto, H.: A 38-year (1978–2015) Northern Hemisphere daily snow cover extent product derived using consistent objective criteria from satellite-borne optical sensors, Remote Sens. Environ., 191, 402-418, 2017.

Hüsler, F., Jonas, T., Wunderle, S., and Albrecht, S.: Validation of a modified snow cover retrieval algorithm from historical
795   1-km AVHRR data over the European Alps, Remote Sens. Environ., 121, 497-515, 2012.

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), 133-152, 2011.

Huang, Y., Liu, H., Yu, B., Wu, J., Kang, E.L., Xu, M., Wang, S., Klein, A., and Chen, Y.: Improving MODIS snow products with a HMRF-based spatio-temporal modeling technique in the Upper Rio Grande Basin, Remote Sens. Environ., 204, 568-582, 2018.

IPCC, Climate Change 2013, The Physical Science Basic. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge and New York.

Jain, S. K., Goswami, A., and Saraf, A. K.: Accuracy assessment of MODIS, NOAA and IRS data in snow cover mapping under Himalayan conditions, Int. J. Remote. Sens., 29(20): 5863–5878, 2008.

KELLY, R.: Development of a prototype amsr-e global snow area and snow volume algorithm, IEEE Trans. Geosci. Remote Sens., 41, 230-242, 2003.

Key, J. R. , Mahoney, R. , Liu, Y. , Romanov, P. , Tschudi, M. , Appel, I., et al.: Snow and ice products from suomi npp viirs, J. Geophys. Res. Atmos., 118(23), 2013.

Klein, A. G. , and Barnett, A. C.: Validation of daily modis snow cover maps of the upper rio grande river basin for the 2000-2001 snow year, Remote Sens. Environ., 86(2), 162-176, 2003.

Klein, A. G., Hall, D. K., and Riggs, G. A.: Improving snow-cover mapping in forests through the use of a canopy reflectance model, Hydrol. Processes, 12, 1723-1744, 1998.

Liu, X., Jiang, L., Wu, S., Hao, S., Wang, G., and Yang, J.: Assessment of methods for passive microwave snow cover mapping using FY-3C/MWRI data in China, Remote Sens., 10(4), p.524, 2018.

Marchane, A., Jarlan, L., Hanich, L., Boudhar, A., Gascoin, S., Tavernier, A., Filali, N., Le Page, M., Hagolle, O., and Berjamy, B.: Assessment of daily MODIS snow cover products to monitor snow cover dynamics over the Moroccan Atlas mountain range, Remote Sens. Environ., 160, pp.72-86, 2015.

Mir, R.A., Jain, S.K., Saraf, A.K. and Goswami, A.: Accuracy assessment and trend analysis of MODIS-derived data on snow-covered areas in the Sutlej basin, Western Himalayas, Int. J. Remote. Sens., 36(15), pp.3837-3858, 2015.

Ning, W., Rawat, G. S., and Sharma, E.: High-altitude ecosystem interfaces in the Hindu Kush Himalayan region, 2014.

Parajka, J., and Blöschl, G.: Validation of MODIS snow cover images over Austria, Hydrol. Earth Syst. Sci. Disc., 3(4), 1569-1601, 2006.

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. Sci., 16(7), pp.2365-2377, 2012.

Qin, D., Liu, S., and Li, P.: Snow cover distribution, variability, and response to climate change in western China, J. Climate, 19:1820–1833, 2006.

Qiu, J.: Trouble in Tibet: Rapid changes in Tibetan grasslands are threatening Asia's main water supply and the livelihood of nomads, Nature, 529, 142–145, 2016.

The Cryosphere
Discussions

Open Access

EGU

830 Riggs, G. A., Hall, D. K., and Román, M. O.: MODIS Snow Products Collection 6 User Guide, Aug Available:
      https://modis-snow-ice.gsfc.nasa.gov/uploads/C6_MODIS_ Snow_User_Guide.pdf, 2016.

Rosenthal, W., and Dozier, J.: Automated mapping of montane snow cover at subpixel resolution from the landsat thematic
      mapper, Water Resour. Res., 32, 1996.

Salomonson, V. V., and Appel, I.: Development of the Aqua MODIS NDSI fractional snow cover algorithm and validation
835   results, IEEE Trans. Geosci. Remote Sens., 44, 1747–1756, 2006.

Serreze, M. C., and Francis, J. A.: The polar amplification debate. Clim. Change, 76, 241 -264, 2006.

Siljamo, N., and Hyvärinen, O.: New Geostationary Satellite–Based Snow-Cover Algorithm, J. Appl. Meteorol. Climatol.,
      50(6), 1275-1290, 2011.

Simpson, J. J., Stitt, J. R., and Sienko, M.: Improved estimates of the areal extent of snow cover from AVHRR data, J.
840   Hydrol., 204(1-4), 1-23, 1998.

Singh, S., K., Rathore B., P. et al.: Snow cover variability in the himalayan–tibetan region, Int. J. Climatol., 34: 446-452,
      2014.

Singh, D.: Re: What value of Heidke Skill Score is practically good for categorical precipitation forecast? And what is the
      same   for   avalanche   forecast?   Retrieved   from:
845   https://www.researchgate.net/post/What_value_of_Heidke_Skill_Score_is_practically_good_for_categorical_precipitati
      on_forecast_And_what_is_the_same_for_avalanche_forecast/54e75aa0d3df3e2a468b464b/citation/download, 2015.

Solberg, R., Wangensteen, B., Metsämäki, S., Nagler, T., Sandner, R., Rott, H., et al.: GlobSnow snow extent product guide
      product version 1.0. Tech. rep., ESA Globsnow, 2010.

Stengel, M. et al.: Cloud property datasets retrieved from AVHRR, MODIS, AATSR and MERIS in the framework of the
850   Cloud _ cci project, Earth Syst. Sci. Data, 9, 881–904, 2017.

Stengel, M., Sus, O., Stapelberg, S., Schlundt, C., Poulsen, C., and Hollmann, R.: ESA Cloud Climate Change Initiative
      (ESA Cloud_cci) data: Cloud_cci AVHRR-AM L3C/L3U CLD_PRODUCTS v2.0, Deutscher Wetterdienst (DWD),
      https://doi.org/10.5676/DWD/ESA_Cloud_cci/AVHRR-AM/V002, 2017.

Stengel, M. et al.: Cloud_cci Advanced Very High Resolution Radiometer post meridiem (AVHRR-PM) dataset version 3:
855   35-year climatology of global cloud and radiation properties. Earth Syst. Sci. Data, 12, 41–60, 2020.

Sun, Y., Zhang, T., Liu, Y., Zhao, W., and Huang, X.: Assessing Snow Phenology over the Large Part of Eurasia Using
      Satellite Observations from 2000 to 2016, Remote Sens., 12, 2060, 2020.

Tedesco, M.: Remote sensing of the cryosphere, John Wiley & Sons, 2014.

Wester, P. , Mishra, A. , Mukherji, A., and Shrestha, A. B.: The Hindu Kush Himalaya Assessment: Mountains, Climate
860   Change, Sustainability and People, 2019.

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

WMO.: Review on remote sensing of the snow cover and on methods of mapping snow. In 14th Session of the WMO
    Commission for Hydrology (Vol. CHy-14, p. 26 p.). Geneva, Switzerland: World Meteorological Organization (WMO),
865   2012.

Wunderle, S., Gross, T., and Hüsler, F.: Snow extent variability in Lesotho derived from MODIS data (2000–2014), Remote
    sens., 8(6), 448, 2016.

Xiao, X., Zhang, T., Zhong, X., Shao, W., and Li, X.: Support vector regression snow-depth retrieval algorithm using
    passive microwave remote sensing data, Remote Sens. Environ., 210, 48-64, 2018.

870 Yang, J., Jiang, L., Ménard, C. B., Luojus, K., Lemmetyinen, J., and Pulliainen, J.: Evaluation of snow products over the
    Tibetan Plateau, Hydrol. Process, 29(15), 3247-3260, 2015.

You, Q. L., Ren, G. Y., Zhang, Y. Q., Ren, Y. Y., Sun, X. B., Zhan, Y. J., ... and Krishnan, R.: An overview of studies of
    observed climate change in the Hindu Kush Himalayan (HKH) region, Adv. Clim. Change Res., 8(3), 141-147, 2017.

Zhang, H., Zhang, F., Zhang, G., Che, T., Yan, W., Ye, M. and Ma, N.: Ground-based evaluation of MODIS snow cover
875   product V6 across China: Implications for the selection of NDSI threshold, Sci. Total Environ., 651, pp.2712-2726,
    2019.

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, 2013.