Evaluation of snow extent time series derived from AVHRR GAC data（1982-2018）in the Himalaya-Hindukush

Xiaodan Wu[1, 2], Kathrin Naegeli[2], Valentina Premier[3], Carlo Marin[3], Dujuan Ma[1], Jingping Wang[1], and Stefan Wunderle[2]

[1]College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, China
[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. The utility of long -term snow cover products lies in their ability to record the real states of earth surface. Although a long-term, consistent snow product derived from ESA CCI+ AVHRR GAC dataset dating back to the 1980s has been generated and released, its accuracy and consistency have not been extensively evaluated. Here, we extensively validate the AVHRR GAC snow cover extent dataset for the mountainous Hindu Kush Himalaya (HKH) region due to its high importance for climate change, impact and adaptation studies. The sensor-to-sensor consistency was first investigated using a snow dataset based on long-term *in situ* stations (1982-2013). Also, this includes a study on the dependence of AVHRR snow cover accuracy related to snow depth. Furthermore, in order to increase the spatial coverage of validation and explore the influences of land cover type, elevation, slope, aspect, and topographical variability on the accuracy of AVHRR snow extent, a comparison with Landsat TM data was included. Finally, the performance of the AVHRR GAC snow cover dataset was also compared to the MODIS (MOD10A1 V006) product. Our analysis shows an overall accuracy of 94% in comparison with *in situ* station data, which is the same with MOD10A1 V006. Using a ±3 days temporal filter caused a slight decrease in accuracy (from 94 to 92%). Validation against Landsat TM data over the area with a great range of conditions (i.e., elevation, topography, and land cover) indicated overall RMSEs of about 13.27% and 16% and overall Biases of about -5.83% and -7.13 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 coverage (1982-2019), 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 exacerbates the effect of global warming through the positive feedback between snow and albedo (Serreze & Francis, 2006). Furthermore, it 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 behaviour is a vital issue in improving weather and climate prediction, supporting water management decisions, and investigating climate change impacts on environmental variables (Arsenault et al., 2014; Sun et al., 2020).

The Hindu Kush Himalaya (HKH), which is often called as the freshwater tower of Asia, comprises the highest concentration of snow outside the polar regions. The snow cover of this area plays a crucial role in the water supply of several major Asian rivers (Immerzeel et al, 2009). On the other hand, the HKH area is of special interest due to large area, rich diversity of climates, hydrology, ecology, and biology (Wester et al., 2019). Variations in snow cover affect the precipitation, near-ground air temperature, and summer monsoon in Eurasia and across the Northern Hemisphere (Hao et al., 2018). Given the fact that HKH is particularly sensitive to climate change and thus shows strong interannual variability, reliable daily snow cover data over a long time series across this area are in great demand.

Optical satellite data provide important data sources for snow cover retrieval through the contrasting spectral behaviour 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 (Marco et al., 2015), MODIS (Riggs et al., 2016a), AVHRR (Shan et al., 2016), VIIRS (Riggs et al., 2016b), 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 ESA Climate Change Initiative (CCI) has emphasized the necessity of generating consistent, high quality long-term datasets over the last 30 years as a timely contribution to the ECV databases. In this demand, a global time series of daily fractional snow cover product has been generated from AVHRR GAC data (Naegeli et al., 2021). This snow dataset is unique as it spans 4 decades and thus provides information about an ECV at climate-relevant time scales.

Nevertheless, there are many factors such as data processing (e.g., calibration, geocoding) and the accuracy of cloud masking, atmospheric constituents, topographic effects, and bidirectional reflectance distribution function (BRDF) and the limitations of snow cover retrieval algorithms influencing the accuracy of AVHRR GAC snow cover extent. Hence, the performance of the AVHRR GAC snow product needs to be extensively evaluated, especially over the HKH which is highly sensitive to climate change. This paper presents the validation of AVHRR GAC snow product over the HKH area during snow seasons. Of particular importance is validating the temporal performance of the product (i.e., different platform operated over the entire datasat period). To this end, the first validation was carried out using 118 *in situ* stations' measurements. The correlation between spatial products and point measurements depends strongly on the selected snow depth. Therefore, the

influence of snow depth on the accuracy of the product was also investigated. Considering that HKH is featured by distinct characteristics of snow cover with shallowness, patchiness, and frequent short duration ephemeral snow (Qin et al., 2006), *in situ* site measurements alone are not enough to characterize its accuracy. A multi-scale validation and comparison strategy is highly needed to assess its accuracy over greater spatial extent and elevation ranges. Within this validation framework, the influences of land cover types, elevations, aspects, slopes, and topographies on the accuracy of AVHRR GAC snow were also explored. Finally, the MODIS snow maps were also introduced to conduct a comparison between the well-validated MODIS product and the new AVHRR GAC snow product. Section 2 describes the study area and data. The validation methodology is explained in Section 3. The performance 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 km^2 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, 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, 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). Snow melt is considered to be a key source of water supply in the HKH range, and the ability of snow products to quantify snow storage and melt is thus critical for the management of water resources (Foster et al 2011).

The validation based on *in situ* stations covers mainly the eastern part of HKH (Fig. 1(a)). To demonstrate the accuracy of AVHRR snow product over the whole area, Landsat data covering the entire region were introduced to conduct a multi-scale validation (Fig. 1(b)). Furthermore, in order to explore its performance in high detail for a great range of conditions (e.g., elevation, topography, and land cover), validation against Landsat TM data was also performed in detail using two tiles of Landsat data (path 140, rows 40 and 41, denoted as "P140-R40/41") (Fig.1(c)), covering a diverse region on the Nepal/Tibet border cantered 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 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).

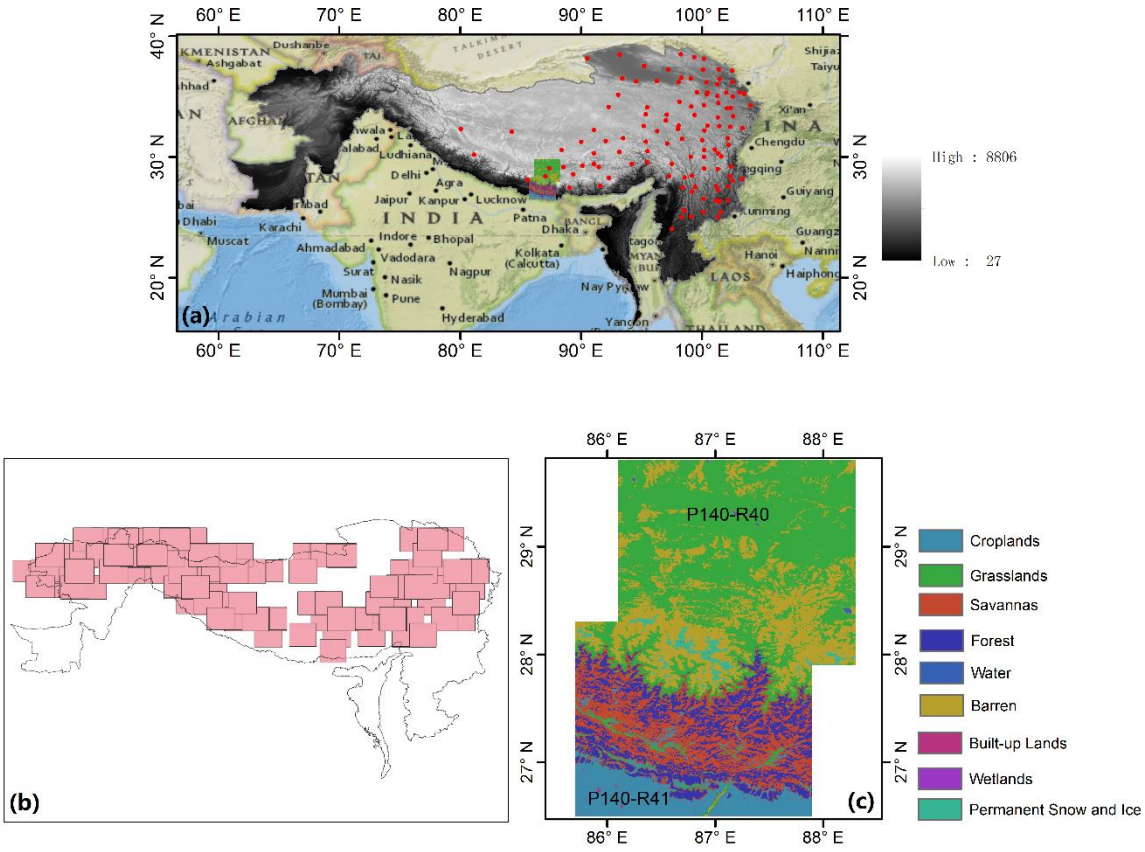

Figure 1. (a) The HKH region and the distribution of *in situ* station locations (red dots); (b) The distribution of Landsat scenes available over the whole HKH; (c) Land cover type of the "P140-R40/41" region of interest which corresponds to Landsat path 140, rows 41 and 42 on the Nepal/Tibet border.

2.2 AVHRR GAC snow extent retrieval

The AVHRR GAC snow cover extent time series version 1 derived in the frame of the ESA CCI+ Snow project is the most recent long-term global snow cover product available (Naegeli et al., 2021). It covers the period 1982-2019 at a daily temporal and 0.05° spatial resolution. The product is based on the Fundamental Climate Data Record (FCDR) consisting of daily composites of AVHRR GAC data (https://doi.org/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 (Devasthale et al., 2017). Snow cover extent retrieval method was developed and improved based on the ESA GlobSnow approach described by Metsämäki et al. (2015) and complemented with a pre-classification module. Alongside the daily reflectance and brightness temperature information, an excellent cloud mask including pixel-based uncertainty information is provided (Stengel et al., 2017, 2020). All cloud free pixels are then used for

the snow extent mapping, using spectral bands centred at about 630 nm and 1.61 µm (channel 3a or the reflective part of channel 3b), and an emissive band centred at about 10.8 µm. The water bodies, permanent ice bodies and missing values are flagged. SCAmod retrieves both the snow cover on top of the canopy as well as on ground below the canopy by taking the canopy density into account. Here, we focus on the latter variable as this is most suitable for the comparison with *in situ* stations.

To reduce the effect of cloud coverage, a temporal filter of ±3 days of each individual snow cover observation was applied after Foppa and Seiz (2012). The AVHRR GAC FCDR snow cover product comprises only one longer data gap of 92 days between November 1994 and January 1995 resulting in a 99 % data coverage over the entire study period of 38 years. In this study, we will focus on the evaluation of raw daily retrieval of AVHRR GAC snow extent (denoted by "AVHRR_Raw") since additional uncertainty will be introduced with the gap-filling process.

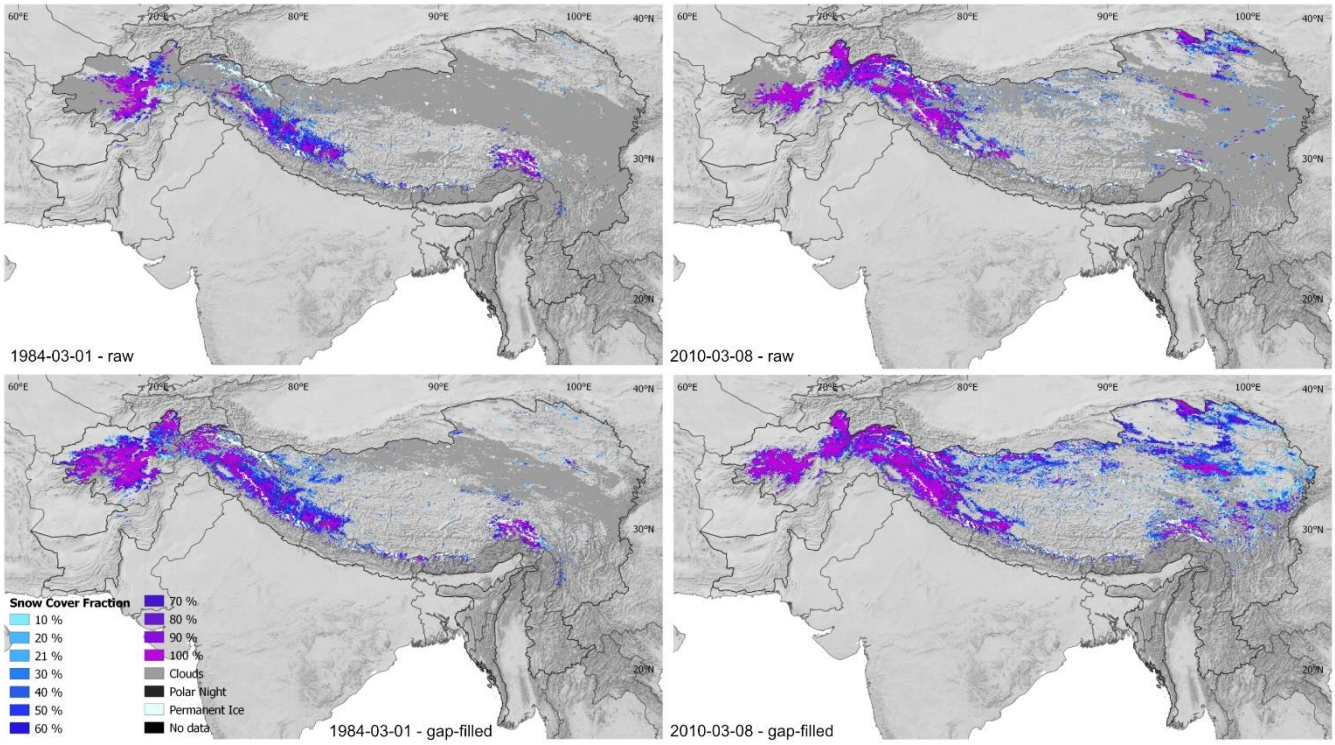

Figure 2. AVHRR GAC raw (top row) and gap-filled (bottom row) snow cover for entire HKH in March 1984 (left) and March 2010 (right).

2.3 Validation data sets

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) measurements (118*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 centimetre. Thus, SD less than 0.5 cm would be labelled as 0 cm in the record. Detailed quality control was made to flag suspicious values. The period from 1982 to 2013 was used to prove the temporal consistency of the AVHRR GAC snow cover extent product.

2.3.2 Landsat TM/ETM data and processing

Landsat data were introduced for two purposes: i) to check the spatial consistency between AVHRR GAC snow and Landsat-based snow based on 197 scenes covering the whole HKH; ii) to explore the influential factors (e.g., elevation, topography, and land cover) on the accuracy of AVHRR GAC snow based on "P140-R40/41". To mitigate the effect of cloud, the validation over "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/). The validation over the whole HKH was restricted to Landsat clear-sky scenes from 1999 to 2018 (197 scenes) (Fig. 1(b)). Level-1 Precision and Terrain corrected ("L1TP") data were selected since they have been radiometrically and geometrically corrected. Following the recommendation of Metsamaki et al. (2015), the fractional snow method by Salomonson and Appel (2006) was employed to generate reference FSC from Landsat TM/ETM imagery. This method is originally designed for MODIS FSC products, with a mean absolute error of less than 10% (Salomonson and Appel, 2004). In this paper, we assumed that such an accuracy can be achieved with higher resolution data. Bands 2 (0.53–0.61 μm) and 5 (1.55–1.75 μm) were used to provide NDSI estimates (Equation (1)), and then the Salomonson and Appel scaling (Equation (2)) is applied. These high-resolution data were then projected to a geographic projection and aggregated to AVHRR GAC pixel scale using the area-weighted average of contributing pixels to 'simulate' the reference FSC estimates at the AVHRR GAC pixel scale.

$$\mathrm{NDSI} = (B2 - B5)/(B2 + B5) \quad (1)$$

$$\mathrm{FSC} = -0.01 + 1.45 \times \mathrm{NDSI} \quad (2)$$

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

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 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 algorithm takes pixel's elevation into account, which is especially important for elevated snow-covered surfaces in spring. In order to avoid spatial scale mismatch between AVHRR and MODIS pixel, MOD10A1 was reprojected to a geographic projection and aggregated to AVHRR GAC pixel scale using the area-weighted average of contributing pixels.

2.3.4 Auxiliary data

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 the accuracy of AVHRR GAC snow extent product. The topographical variability within 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.

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 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. In order to match with the pixel size of AVHRR GAC snow, the MCD12Q1 was resampled to 0.05° spatial resolution with the nearest neighbour interpolation.

3 Methods

AVHRR GAC snow extent was evaluated from several aspects. The validation based on *in situ* sites aims to prove the long-term consistency since *in situ* stations provide valuable long time series measurements. While the comparison with Landsat and MODIS snow is focused on their spatial consistency and the in-depth analysis of influential factors (elevation, topography, and land cover). The validation strategy is briefly summarized in Table 1.

Table 1. Validation strategy: data and purpose.

Data source	Variable(s)	Application purpose	Validation purpose
In situ stations	Snow depth	binary (snow/no snow) validation of snow cover	Temporal consistency but spatially limited, dependence on snow depth

MOD10A1	Snow cover	quantitative comparison of snow cover	Relative comparison to investigate general performance
Landsat east/west	Snow cover	absolute comparison of snow cover extent	Great spatial and patchy temporal coverage
Landsat P140-R40/41	Snow cover	absolute comparison of snow cover extent	Limited spatial and temporal coverage but great variability in elevation, topography and land cover
High and medium resolution satellite images (Landsat, MODIS), SRTM DEM	snow cover, land cover, elevation, topography (aspect, slope)	-	Dependency on land cover type, topography, elevation

3.1 Binary validation 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 provide the only source to validate the time series AVHRR GAC snow extent over this long period. Since there is no reliable way to convert SD to FSC, both FSC and SD information were converted to binary information by applying appropriate 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 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 cantered around this "point" were tested, respectively. This treatment took into consideration the influence of data noise, geometric mismatch, and spatial heterogeneity. Furthermore, 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) were utilized to indicate the quality of the snow product. If both reference data and snow product identified the pixel as snow, it is labelled as a hit (a); if neither of them indicated the pixel as snow, it is labelled 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.

		Reference data (*in situ*)		
Snow product		**yes**	**no**	
(e.g. AVHRR GAC orMODIS)	**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 determined (Equations (3-5)) (Hüsler et al., 2012). ACC denotes the percentage of correctly classified pixels divided by the total number of pixels. ACC values closer to 1 denotes a perfect 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 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 denotes a relatively good score for a reasonably sized sample for the binary forecast (Singh, 2015). The Bias, 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 a value of 1.

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

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

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

The validation follows 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 allows us to check the spatial divergence of accuracy within different sites as well as the effect of land cover on the accuracy of satellite derived snow information. Second, the snow data of all *in situ* sites were combined together for validation of AVHRR GAC snow on the daily basis. In this way, the long-term temporal 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 analysed 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.

Finally, in order to check the relative performance of AVHRR GAC snow to well-used MODIS product, MOD10A1 V006 was also evaluated with *in situ* station data following the same method. It is expected that the major difference in their performance 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 root mean square error (RMSE), mean bias (mBias) and the coefficient of correlation (R) were derived through the scene-by-scene comparison.

3.2 Multi-scale validation based on high-resolution snow cover maps

In order to evaluate AVHRR GAC snow at a broader spatial scale, Landsat TM/ETM aggregated FSC was used as the reference. Snow free values are treated as 0% snow and fully snow-covered pixel is assigned 100% snow. The validation was conducted from two aspects: i) one is based on 197 scenes covering the whole HKH in order to increase the spatial coverage of validation. ii) the other is based on 46*2 scenes over 'P140-R40/41' in order to make a detailed analysis of the influential factors (e.g., elevation, topography, and land cover) on the accuracy AVHRR snow dataset.

4 Results and discussion

4.1 The validation based on *in situ* data

4.1.1 Snow depths and pixel 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 Bias, ACC and HSS with all the threshold combinations (Table 2) are shown in Fig. 3.

As shown in Fig. 3(a), a SD threshold of 2 cm (case2) maximizes the overall accuracy of the AVHRR GAC snow cover dataset. With the further increase of SD threshold, the AVHRR GAC snow will be detected to be seriously overestimated. This indicates the presence of snow can be best detected by the AVHRR GAC dataset for *in situ* snow depth measurement of 2 cm. Furthermore, the increasing rate of ACC and decreasing rate of HSS are the highest between the 1 and 2 cm SD threshold

and flattens for greater SD thresholds. When it comes to the influences of geometric mismatch or spatial heterogeneity (center pixel versus 3 × 3 pixels, Table2), they show significant effects on both the magnitude and the variation trend of these accuracy indicators. But such effects are not fixed and vary with satellite datasets and accuracy indicators (Fig. 3). For this reason, we chose case2 (SD>=2cm, center pixel) as the optimum threshold combination for the evaluation of AVHRR GAC snow dataset. For consistency, this choice was also used for comparison of the performance of MODIS with *in situ* data.

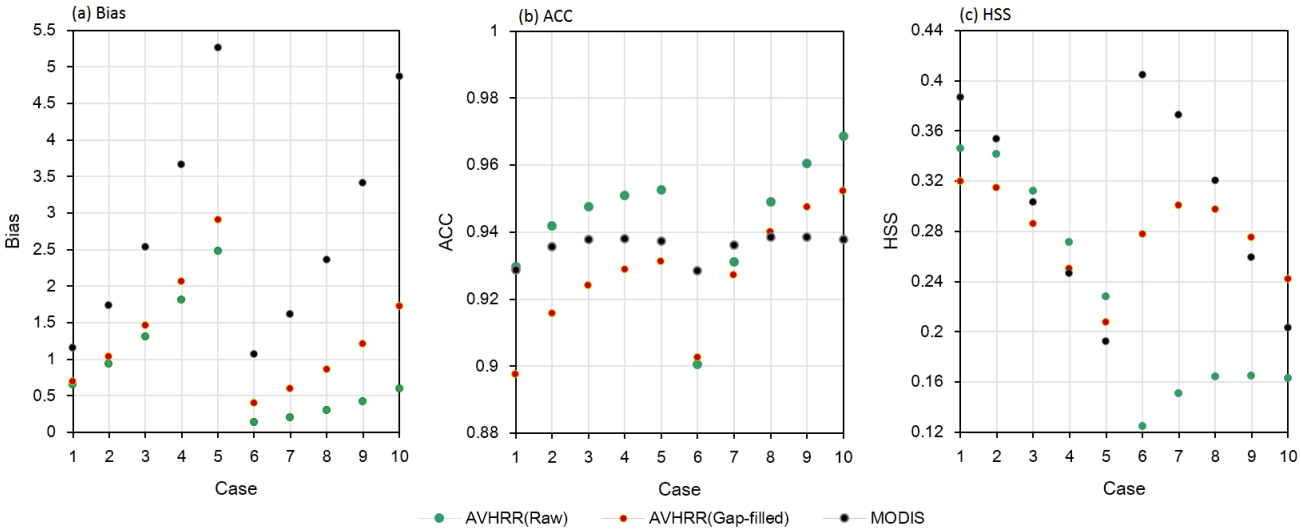

Figure 3. Sensitivity analysis of product accuracy related to snow depth thresholds of the *in situ* station data. The overall error is a spatiotemporally integrated statistical measure.

As seen from Fig. 3(a), AVHRR snow datasets show distinct advantages over MODIS snow regarding the Bias value. The former show Bias of 0.94 and 1.03 for the AVHRR raw snow and gap-filled snow respectively. While the latter is seriously overestimated with the Bias of 1.74. Nevertheless, the three datasets show comparable ACC, with the values of 0.94, 0.92, and 0.94 for AVHRR raw snow, AVHRR gap-filled snow, and MODIS snow datasets, respectively. The HSS of the three datasets are reasonable, with the values larger than 0.3. MODIS snow shows the largest HSS of 0.35, followed by AVHRR raw snow, with a HSS of 0.34. The AVHRR gap-filled snow cover dataset ranks last, with the smallest HSS of 0.31. From the above results, it can be found that AVHRR raw dataset performs slightly better than AVHRR gap-filled dataset with respect to the agreement with *in situ* sites and the algorithm performance (skill). This is reasonable since additional uncertainty was introduced in the gap-filling process. For this reason, we will only focus on AVHRR raw snow for further analysis. Generally, AVHRR raw snow is comparable with MODIS snow when ACC and HSS are focused.

4.1.2 The temporal consistency of quality indicators

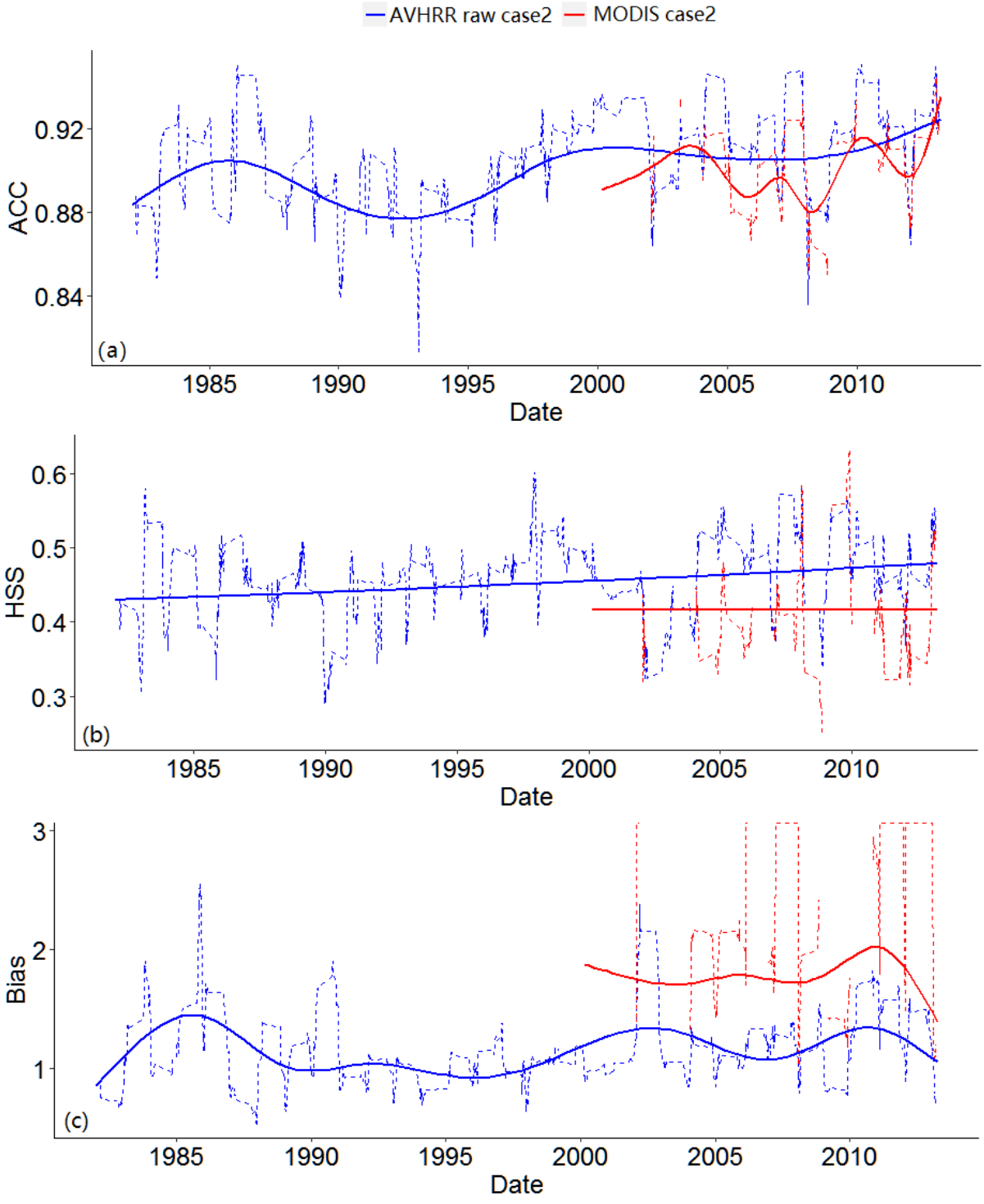

Figure 4. The time series (denoted by the dashed lines) of ACC, HSS, and Bias for AVHRR raw and MODIS snow data during the investigated period. A simple moving average with box dimension n=30 was applied to the time series in order to reduce the noise and uncover patterns in the data. The solid lines represent the fitted trend of these accuracy indicators.

From Fig. 4, it can be seen that the inter-annual variability of these accuracy metrics is evident, especially for ACC and HSS. In the time series of AVHRR GAC snow, ACC is basically distributed between about 88% and 92% (Fig. 4(a)). An obvious increase in ACC can be observed from 1982 to 1985, followed by a decrease in ACC from 1985 to 1992. Then an increasing trend of ACC occurs from 1992 to 2000. From 2000 to 2010, ACC is relatively stable with time. But after 2010, an increasing trend reappears. Differently from the previous assessments, the ACC of AVHRR snow datasets at the beginning of the time series (1982-2000) is slightly worse than the end of the time series (2000-2013) regarding the magnitude of ACC and its temporal consistency. The HSS shows a different behaviour compared to ACC (Fig. 4(b)), which increases slightly and monotonously from 0.45 at the beginning to about 0.48 at the end of the time series. This further indicates that the performance of AVHRR snow continues to improve with time. Nevertheless, the improvements of the performance of AVHRR GAC snow do not occur in the Bias (Fig. 4(c)). The Bias shows the best performance from 1990 to 2000, with relatively stable values around 1. But during other time periods, relatively large fluctuations appear, and it generally overestimates snow during these periods.

As shown in Fig. 4, it can be seen MODIS snow is inferior to AVHRR GAC snow regarding the magnitude of ACC and its temporal consistency. Furthermore, its HSS is consistently smaller than that of AVHRR GAC snow. Nevertheless, its temporal stability is slightly better than AVHRR GAC snow since the HSS of MODIS almost keeps constant over time. When it comes to Bias, MODIS snow shows more serious overestimation than AVHRR GAC snow, but comparable temporal stabilities to the latter.

In order to highlight the performance of AVHRR GAC snow in different months, the temporal variations of ACC, HSS, and Bias over different months are presented (Fig. 5). From Fig. 5(a), it can be seen that ACC of AVHRR GAC snow is over 0.85 for all months and even above 0.90 for October. Nevertheless, both the temporal variation trend and the magnitude of ACC show differences from month to month. It is clear that the AVHRR GAC snow shows highest ACC in October and lowest ACC in January, but the temporal stability of ACC is best in November and worst in January and December. It is interesting that the results tend to polarize into two groups: ACC for January through March and ACC for October through December. Generally, ACC in the former group is generally smaller than those in the latter group. It is noteworthy that ACC after 2000 is generally larger and more stable than those in earlier years on the monthly scale (Fig. 5(a)), indicating the better accuracy and consistency of the younger satellite platforms after 2000. Compared to AVHRR GAC snow, ACC of MODIS snow consistently shows large temporal variations for all months, and there is no month that shows advantages over others regarding the magnitude and temporal stability of ACC (Fig. 5(b)).

The HSS for different months are larger than 0.4 throughout the time series, but large differences of the magnitude and temporal stability exist between different months (Fig. 5(c)). Similar to ACC, the AVHRR GAC snow generally show the largest HSS in October for most of the time. Furthermore, the HSS in October shows a similar temporal variation trend with

the overall temporal trend of HSS in Fig. 4(b). Among all the months, the HSS in December shows the largest temporal variations, featured by the highest HSS from 1990 to 2000 and the lowest HSS from 2005 to the end. The HSS in January through March show relatively smaller temporal variations than those in October through December. Regarding the magnitude of HSS, the different rank of these months during different periods may be associated with the shift of snow cover phenology due to interannual variability intensified by global warming. Unlike AVHRR GAC snow, MODIS snow shows larger HSS in January and February (Fig. 5(d)). Furthermore, the temporal variations of HSS are more significant than AVHRR GAC during the same period.

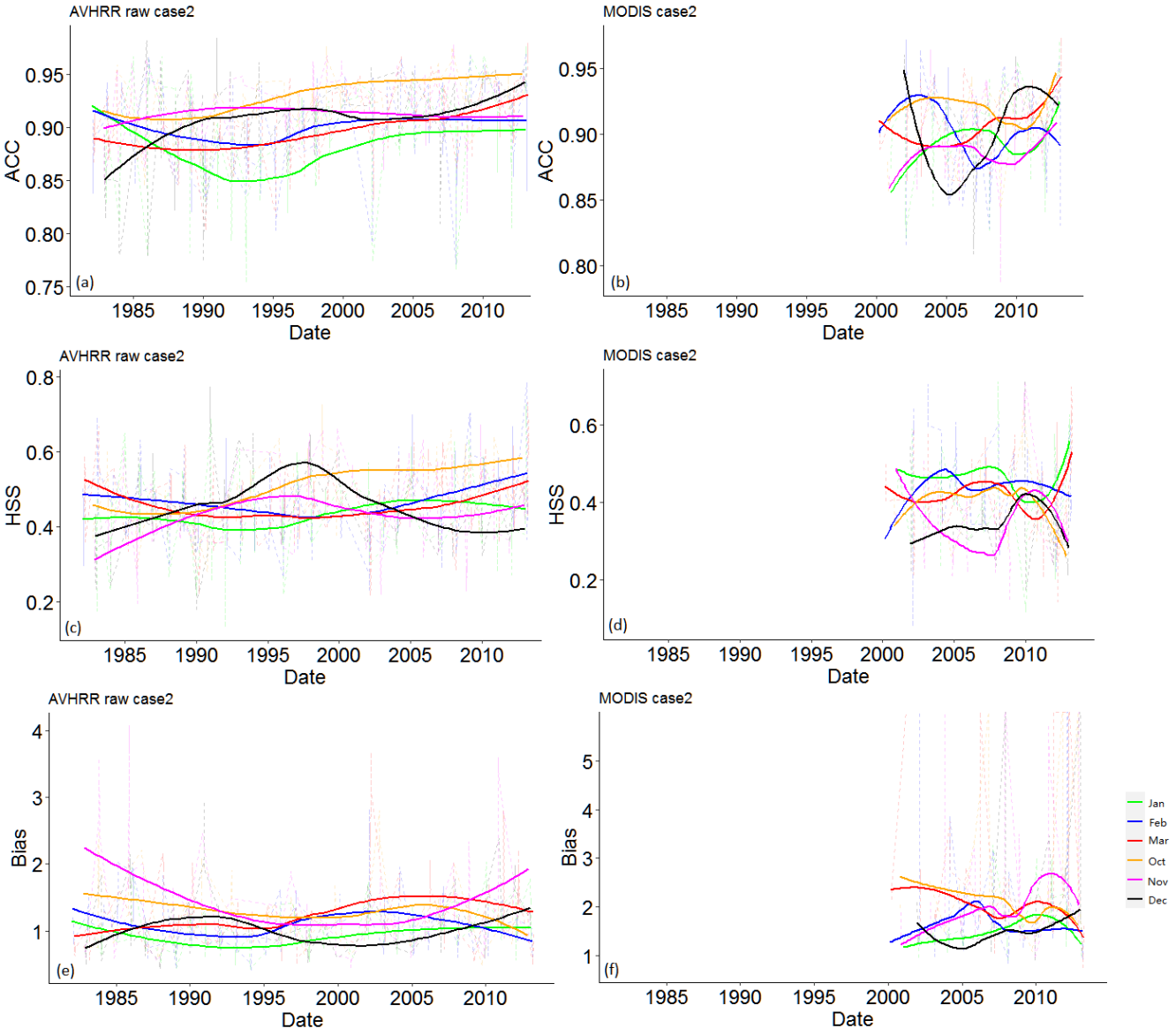

Figure 5. The temporal behavior of ACC for AVHRR raw and MODIS snow data in different months (indicated by the different colored lines) during the snow season. The solid-colored lines represent the fitted trend of these accuracy indicators for different months.

Although the AVHRR GAC snow shows the best performance in October regarding the magnitude of ACC and HSS, it shows serious overestimation in this month (Fig. 5(e)). In particular, AVHRR GAC snow generally overestimates snow in February, March, October, and November. By contrast, it either slightly overestimates or underestimates snow in December and January, with the Bias distributed around 1. This result is understandable because during December and January, snow coverage tends to be dense and spatially continuous, which results in unbiased estimation. By contrast, during February, March, October, and November, snow cover tends to be patchy, and AVHRR GAC data is more able to detect snow than *in situ* "point" observation due to its large pixel coverage. MODIS snow consistently overestimates snow in different months and shows larger temporal variations than AVHRR GAC snow (Fig. 5(f)).

From the results above, it can be concluded that AVHRR GAC snow dataset performs variable throughout the course of the year, which may be related to the different amounts of snow in HKH. Generally, the magnitudes of ACC and HSS are largest in October and smallest in January. But the temporal stability of ACC is best in November and worst in January and December. While that of HSS is worst in December. The results of Bias provide different perspectives for the performance of AVHRR GAC snow. It generally overestimates snow in February, March, October, and November. By contrast, unbiased estimation is likely to occur in December and January. Compared to AVHRR snow datasets, the inter-annual variability of ACC, HSS, and Bias of MODIS snow product in different months are generally stronger (Fig. 5).

4.1.3 The spatial consistency of quality indicators

Fig. 6 show the boxplots of validation metric derived from each *in situ* station, with the aim of revealing their spatial variability. It can be observed that the spatial variability of these validation metrics widely exists given their dispersed distribution. The maximum of ACC even reaches 0.99 for AVHRR snow datasets, while the minimum values are close to 0.76 (Fig. 6(a)). Similarly, HSS also shows a dispersed distribution for AVHRR snow datasets. The AVHRR raw dataset ranges from 0.2 to 0.39 with min-max values of 0.01 to about 0.68 (Fig. 6(b)). Likewise, the Bias is located around 0.51-1.6 with min-max values of 0.05 and 2.89 for AVHRR dataset (Fig. 6(c)). These results are understandable because the performance of satellite snow datasets is affected by many factors. Despite the awareness of spatial variability of these validation metrics, the degree of variability depends on satellite datasets and metrics. The HSS and Bias of MODIS snow dataset are more divergent than AVHRR raw snow dataset (Fig. 6).

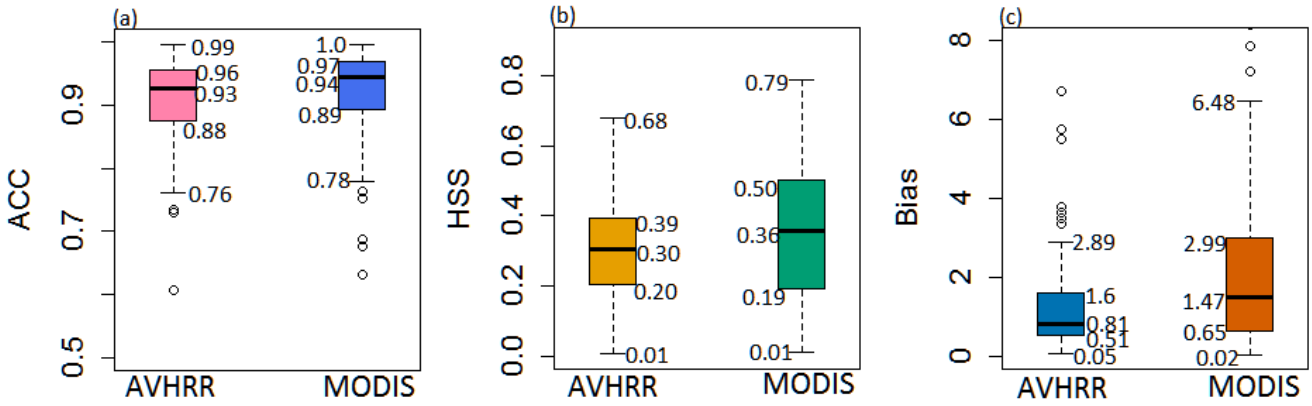

Figure 6. The boxplots of ACC (a), HSS (b), and Bias (c) for AVHRR raw and MODIS snow data throughout the sites. The numbers around the plots indicate the minimum, lower quartile, median, upper quartile, maximum of the boxplots, respectively.

4.1.4 The potential influential factors on accuracy

Following the early study (Klein and Barnett, 2003), the effect of SD on the accuracy of satellite snow datasets was evaluated (Fig. 7(a)). Observed SD was divided into six categories: SD=0cm, 1cm, 2cm, 3cm, 4cm and 5cm. It is obvious that the ACC of the two satellite snow datasets based on AVHRR and MODIS show similar responses. The highest ACC occurred when SD=0 cm, which is followed 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 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 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. 7(b). 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 AVHRR and MODIS snow datasets shows a 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 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 are more likely covered by very thin snow or not covered by snow. Consequently, *in situ* sites are more likely to 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 indicates snow cover using the threshold (FSC≥50%), but there appears a strong possibility that *in situ* site is not covered by snow or only 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 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.

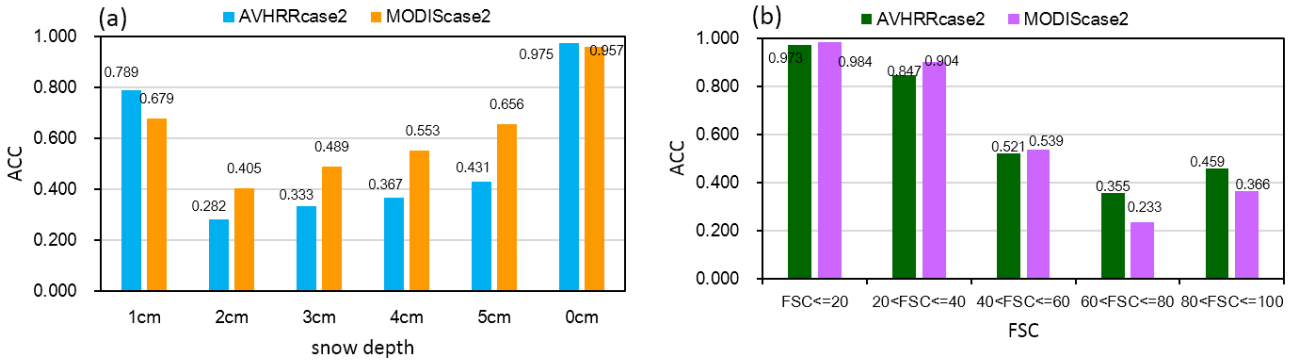

Figure 7. The variation of ACC with snow depth (left) and FSC (right) for AVHRR raw and MODIS snow data throughout the sites during the experimental period.

As seen from Fig. 7(a), 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 very thin snow conditions, AVHRR snow datasets are less misclassified than MODIS snow product. But in contrast, in snow-covered conditions, 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. From Fig. 7(b), it becomes apparent 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 phenomenon is related to the different degrees of spatial representativeness of *in situ* sites relative to different pixel scales.

Figs. 8(a-b) present the distribution of ACC for two satellite snow datasets against *in situ* site observations over different elevation regions (five classes) and landcover types (four types), respectively. It is generally thought that coarse-pixel satellite snow products perform better in higher elevations due to the continuous and thick snow cover (Yang et al., 2011). Nevertheless, the ACC over HKH shows different phenomena. The two satellite snow products consistently show larger ACC over slightly lower elevations than those over higher elevations. Nevertheless, an exception can be found in the elevation region of [3500

m, 4500 m], where the ACC of the two dataset is the lowest over the whole HKH. Furthermore, the ACC over these elevation regions is the most divergent, demonstrating that the accuracy of snow product within this range is more likely to be affected by other factors. It is noteworthy that MODIS snow product slightly outperforms AVHRR snow dataset over different elevation regions. This is reasonable since the spatial scale mismatch between *in situ* and satellite-based observations is greater for the AVHRR snow datasets than for the MODIS snow dataset.

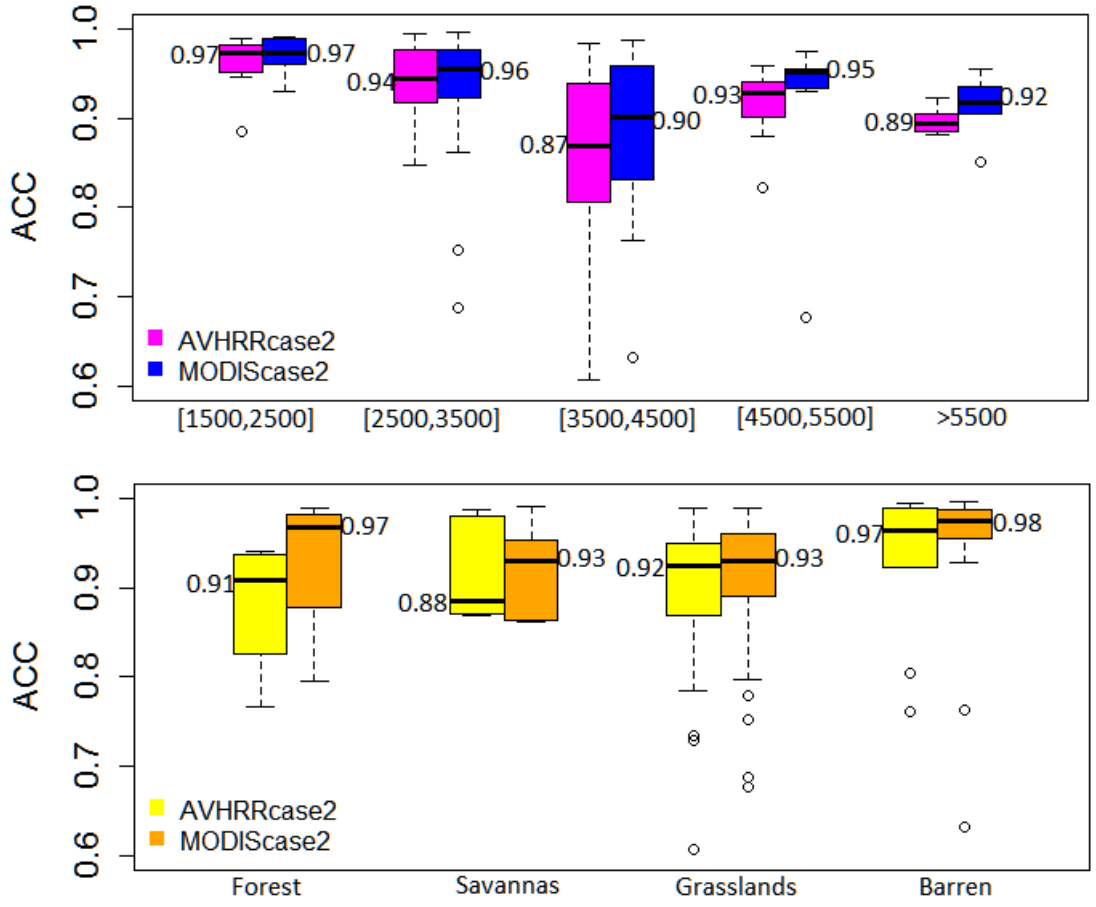

Figure 8. Boxplots of ACC from the direct comparison with *in situ* site observations over different elevation regions (a) and land cover types (b) for AVHRR raw and MODIS snow data.

Despite the effect of elevation on ACC, it was not treated when we explore the effect of landcover type on ACC (Fig. 8(b)). Because the number of *in situ* sites over different landcover types and different elevation regions are very limited. For AVHRR GAC snow, the highest agreement with *in situ* measurements is found in barren, followed by grasslands and savannas. Although nearly half of the *in situ* sites over forest show ACC larger than 0.91, substantial numbers of *in situ* stations show relatively low ACC over forest. This indicates that the well-known issues of identifying snow in forested areas using optical satellite data are not fully resolved in AVHRR GAC snow. It is interesting to find that MODIS snow product remains its

superiority over different land cover types, and its advantage becomes more pronounced over forest and savannas. The different performance between AVHRR snow and MODIS snow is partly caused by their individual accuracy and partly caused by the different effects of spatial scale mismatch between *in situ* and satellite-based observations.

4.2 Comparison based on medium to high resolution data

4.2.1 Quantitative comparison to MOD10A1

In order to investigate the absolute difference between AVHRR GAC and MODIS snow, we compared them on the pixel basis following the cross-validation framework. The indicators of RMSE, mean Bias (mBias), and correlation coefficient (R) are used to reveal their difference and consistency. The scene-by-scene comparison was made over the region "P140-R40/41" throughout the snow season of 2012 and 2013. As shown in Fig. 9, the highest density is between 0 and -5 for mBias and 0 and 10 for RMSE. Only a small part of the scenes show relatively large mBias of 15% and RMSE of 30%. Their overall mBias are very small, with the value of 0.06% and 0.94% for AVHRR raw and gap-filled snow datasets, respectively. The overall RMSEs are 12.8% and 17.0% for AVHRR raw and gap-filled snow datasets, respectively. Furthermore, the spatial distribution characteristics of FSC indicated by AVHRR GAC snow basically agree with that of MODIS snow given the overall R is 0.63 and 0.53 for AVHRR raw and gap-filled snow datasets, respectively.

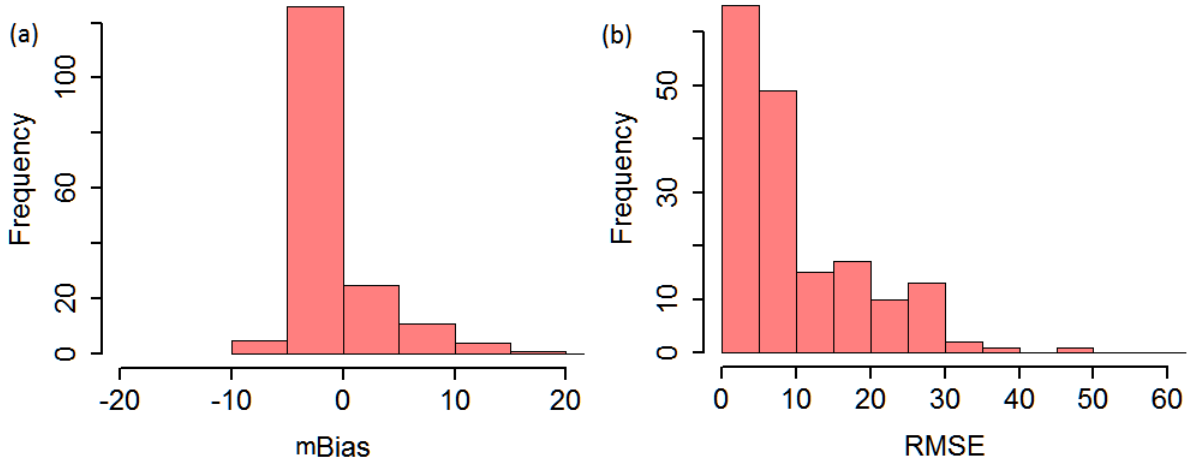

Figure 9. The distribution of mBias (a) and RMSE (b) derived from the scene-by-scene comparison between AVHRR GAC raw snow datasets and MODIS snow products over the region "P140-R40/41" throughout the snow season between 2012 and 2013.

4.2.2 Spatial consistency of snow cover extent

In order to avoid the spatial limitations of the *in situ* stations, the comparison between AVHRR raw snow datasets and Landsat data was also carried out over the whole extent of HKH. The RMSE, mBias, and R in different conditions are summarized in Table 4. RMSE is generally less than 23% in different conditions with an overall RMSE of 22.31%. The mBias

still indicates an underestimation of AVHRR snow datasets, with the overall mBias of -2.96%. The consistency between Landsat and AVHRR snow is good, with the overall R of 0.82. The best performance of AVHRR GAC snow is observed in plain, with the smallest RMSE of 18.2% and mBias of -1.65, as well as the largest R of 0.90. By contrast, the largest RMSE of 22.9% and mBias of -3.18% appear in mountain areas. When it comes to consistency, the worst performance occurs in forests, with the lowest R of 0.57.

Table 4. Summary of accuracies of AVHRR raw snow data with Landsat5 TM snow data over the whole HKH.

	RMSE (%)	**mBias (%)**	**R**
Plain	18.20	-1.65	0.90
Mountain	22.90	-3.18	0.80
Forest	20.41	-2.17	0.57
Global	22.31	-2.96	0.82

In order to explore the performance of AVHRR GAC snow in high details for a great range of conditions, the spatial accuracy was assessed on the pixel basis based on Landsat5 TM data time series over the areas covered by 'P140-R40/41'. AVHRR snow datasets systematically underestimate snow covered areas with regards to the Landsat5 TM data (Table 5). 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). The accuracy of AVHRR GAC snow is different over the two areas, with better performance over "P140-R40" than "P140-R41" (Table 5). AVHRR raw snow shows a higher accuracy with a smaller RMSE of 11.39% (vs. 15.08%) and mBias of -4.19% (vs. -7.64%) over "P140-R40" than "P140-R41". Similar results can be also seen in AVHRR gap-filled snow, with a smaller RMSE of 13.40% (vs. 18.37%) and mBias of -4.94% (vs. -9.46%) over "P140-R40" than "P140-R41". When the two areas are combined together, the AVHRR GAC snow presents overall RMSEs of 13.27% and 16% and mBias of -5.83% and -7.13 for the raw and gap-filled datasets respectively over the highly variable region (e.g., elevation, topography, and land cover). From Table 5, it is clear that AVHRR raw snow shows a higher accuracy than the gap-filled snow dataset. But their overall consistency with Landsat TM snow is comparable (0.46 and 0.47 for raw and gap-filled snow dataset).

Table 5. 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	13.40	-4.94	0.52	11.39	-4.19	0.49
P140-R41	18.37	-9.46	0.43	15.08	-7.64	0.44
Overall	16.00	-7.13	0.47	13.27	-5.83	0.46

4.2.3 Pixel-based comparison and potential influential factors on accuracy

Both the land cover types and topography are highly heterogeneous over the HKH. Here, the sub-region "P140-R40/R41" was chosen to investigate the influential factors (i.e., elevations, landcover type, slope, aspect, and topographical variability) on the accuracy of AVHRR GAC snow dataset (Fig. 10). From Fig. 10(a), it can be seen that RMSE shows a strong positive response to elevations. But an exception can be found within the region of [3500 m, 4500 m], where RMSE shows a clear decrease but also the greatest spread. This occurs because the accuracy of AVHRR GAC snow is not merely influenced by elevations. Over the flat areas ([0 m, 200 m]) and hills ([200 m, 500 m]), the highest density of RMSE is distributed between 0 and 5%. Over lower and medium height mountains ([500 m, 2500 m]), the highest density of RMSE is distributed between 0 and 10%. With the further increase of elevation (i.e., [2500 m, 3500 m]), more than half of the pixels show RMSE larger than 10%. Nevertheless, over the elevation region of [3500 m, 4500 m], more than half of the pixels show small RMSEs of less than 5%. But the maximum RMSE can reach 45% over this region. Over the elevation region of [3500 m, 4500 m], the highest density of RMSE is lower than 15%. And the RMSE increases significantly over the extreme high area (>5500 m). This finding is inconsistent with Yang et al. (2011) who consider that coarse-pixel satellite snow products generally perform better in higher elevations due to the continuous and thick snow cover. The larger RMSEs in the highest elevations are partly caused by the large values of FSC themselves, partly caused by the roughness, topographic effects, and shadows, and partly caused by the cloud effects given that the probability of cloud rises with rising altitude in mountain areas.

Given the considerable effect of elevation on the accuracy of AVHRR GAC snow, the regions 'P140-R40/41' are divided into eight groups according to their elevations (Fig. 10(b)). From Fig. 10(b), it can be seen that the RMSE is rising with elevation in each individual land cover type. Nevertheless, an exception can be found in grasslands, which show the largest RMSE over the region [2500 m, 3500 m]. And the RMSEs decrease significantly over the region [3500 m, 4500 m]. Over the flat areas ([0 m, 200 m]), AVHRR snow mapping accuracy is the best in croplands and the worst in barren. And the accuracy is slightly better in forest than in savannas. Moreover, the accuracy is most spatially stable in grasslands given the centralized distribution of RMSE. When it comes to hills ([200 m, 500 m]), croplands still show the best accuracy, followed by the forest and grasslands, and savannas rank last. As the elevation increases to [500 m, 1500 m], croplands still show the best accuracy. By contrast, grasslands show the worst accuracy. Savannas show smaller RMSE than forests. With the further increase of elevations ([2500 m, 3500m]), only grassland, savannas, and forests appear. The best performance occurs in forests, followed by savannas, and grasslands rank last. Over the high mountain area ([3500 m, 4500 m]), savannas present the largest RMSE, followed by forest. And the grasslands show the largest spatial variations within this range. With the further increase of elevation (>4500 m), only grasslands and barren appear, and the former shows better accuracy than the latter with regard to the magnitude of RMSE and its spatial variations. Therefore, we can conclude that the performance of AVHRR GAC snow over different landcover types depends mainly on elevations. Its accuracy is generally good in croplands since it is distributed only within the region of [<1500 m]. The accuracy of barren is generally not good because it is merely distributed within the range of [>3500 m]. Forests and savannas basically show comparable overall accuracy. The accuracy of grasslands shows a

different response to elevations, which is the worst over regions [2500 m, 3500 m]. Its accuracy is comparable to other land cover types over relatively low elevations (<1500 m) and outperforms barren over high elevations (> 3500 m).

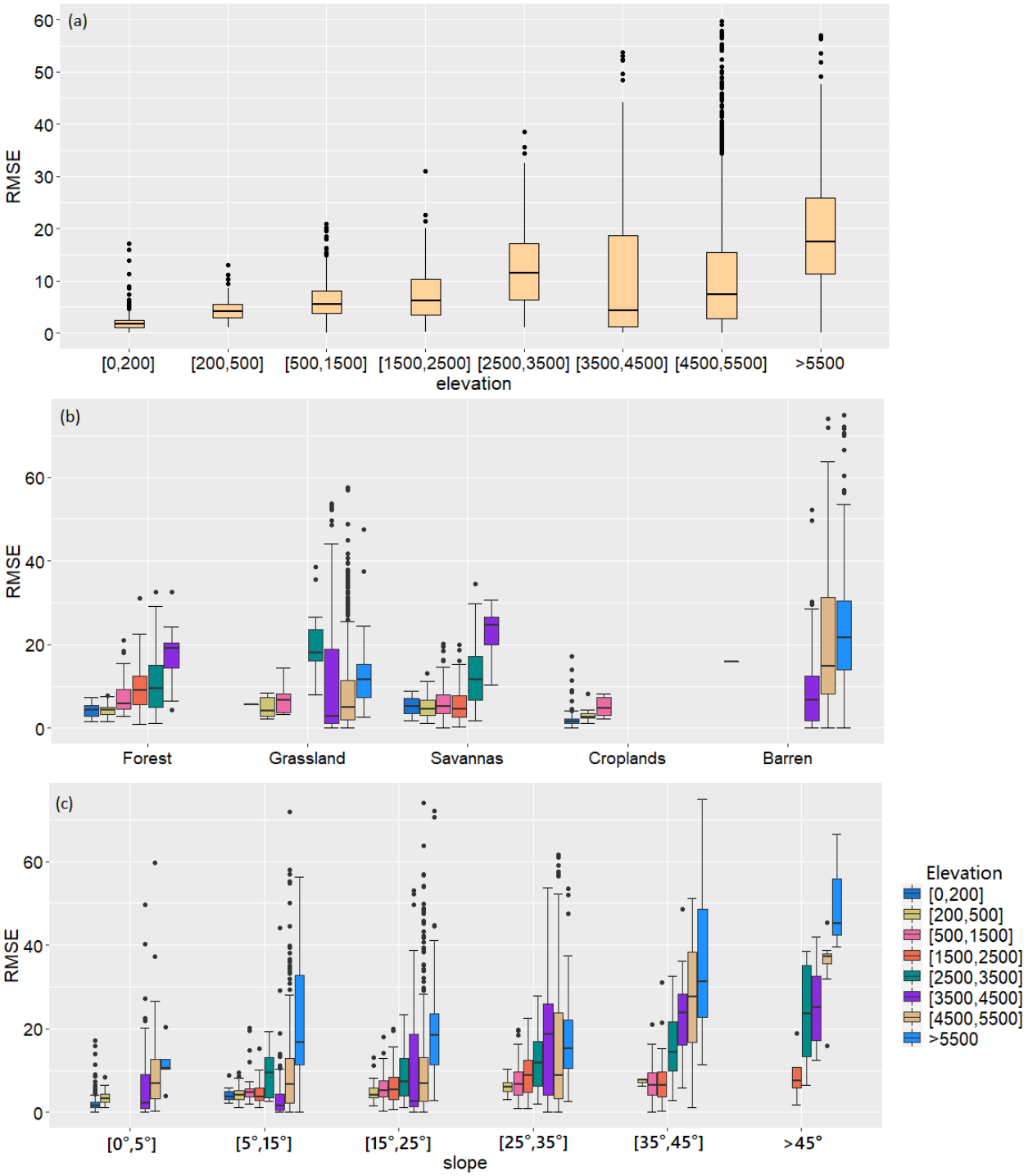

Figure 10. The variation of RMSE in dependence of a) elevation, b) land cover, c) slope, d) aspect and e) topographical variability. The different colors refer to the 8 different elevation classes. The plots show combined results of "P140-R40" and "P140-R41".

The effect of slope on the accuracy of AVHRR GAC snow datasets is clearly shown in Fig. 10(c). Better results tend to appear over the areas with smaller slopes. The RMSE over different elevation regions generally shows an increasing trend with slope. Nevertheless, there are two outliers over the regions [1500, 2500] and extremely high areas (>5000). In the former region, the RMSEs with slopes ranging from 25° to 35° are slightly larger than those with slopes ranging from 35° to 45°. In the latter region, there is no increase of RMSE when slopes increase from [15°, 25°] to [25°, 35°]. This occurs because the accuracy of AVHRR GAC snow is affected by many factors. In fact, the effect of slope on snow mapping accuracy is understandable since the topographic effects tend to be significant in a steep mountain area.

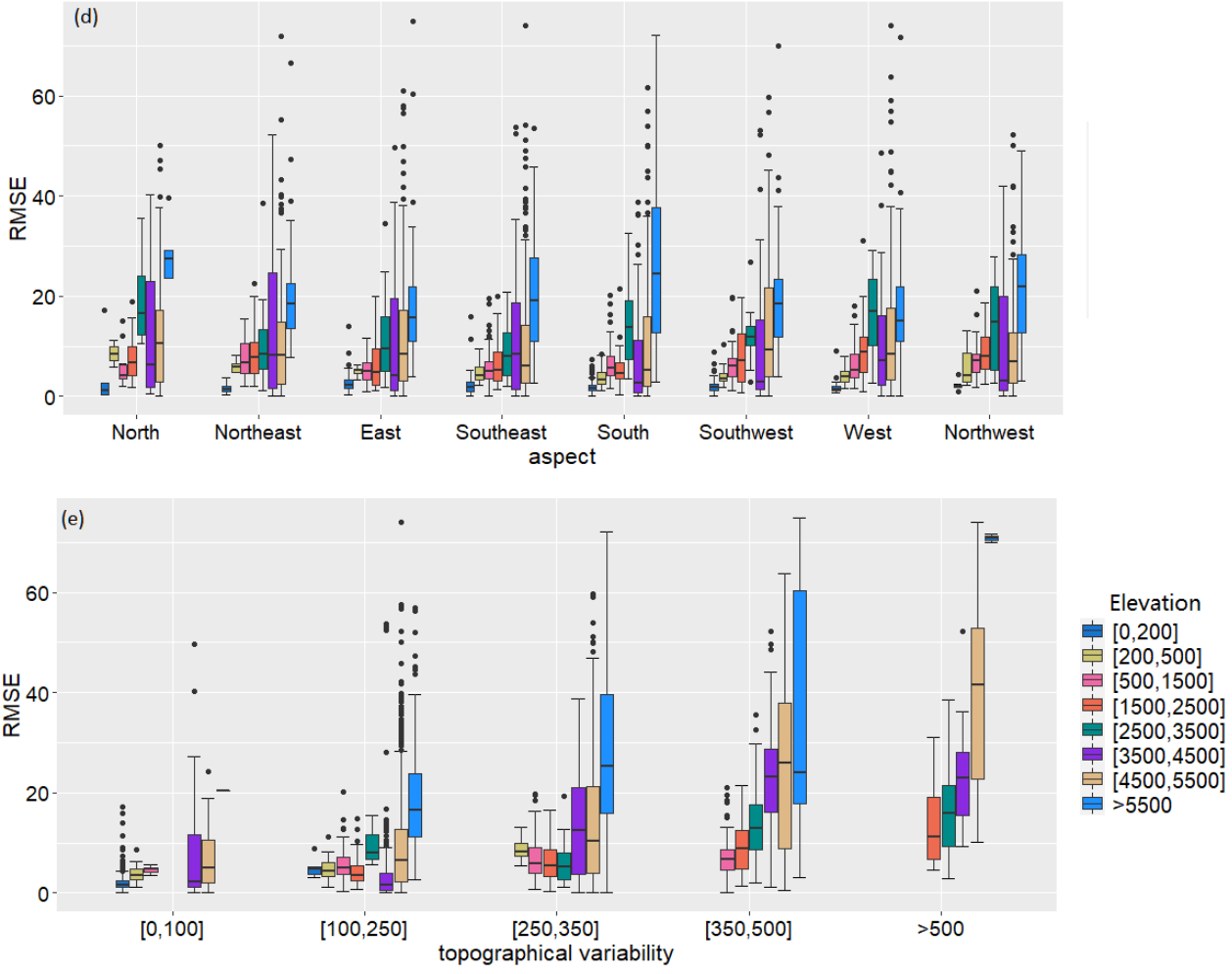

Figure 10. continued.

Regarding the effect of aspect (Fig. 10(d)), there is not a clear trend of RMSEs with aspect over the regions less than 5000 m. Nevertheless, over the areas higher than 5500 m, the RMSE first show a clear decreasing trend and then a clear increasing trend when aspect changes from north facing slope to south facing slope, and vice versa. Moreover, the maximum RMSE can even reach 70% over south facing slope, which is larger than that of north facing slope (~40%). This is attributed to the fact that during winter months, south facing slopes receive significantly more radiant energy, providing unfavourable 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.

From Fig. 10(e), it can be found that there is only small topographical variability over the regions with low elevations (<500). The RMSEs of these regions with different elevations generally show an increasing trend with 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 degrees of surface information loss between high-resolution satellite data and coarse resolution satellite data. Furthermore, the increasing trend is more significant over the regions with large elevations. It is noteworthy that there are also several outliers that do not show a clear increasing trend. For instance, over the elevation region of [2500 m, 3500 m], even a decrease of RMSE can be observed when topographical variability increases from [100, 250] to [250, 350]. This is due to the fact that the topographical variability is just one of the factors influencing the accuracy of AVHRR GAC snow.

From the results above, we can conclude that the accuracy of AVHRR GAC snow is closely related to elevations, slopes, and topographical variability. And the negative influence of these factors on snow mapping accuracy is more significant over regions with high elevations. The effect of aspect can be ignored over the regions less than 5500 m. But for the areas higher than 5500 m, the accuracy first increases and then decreases gradually from the north facing slope to the south facing slope and vice versa. The effect of landcover type on snow mapping accuracy is related to elevations.

5 Conclusions

In this study, the ESA CCI+ Snow project AVHRR GAC snow cover extent product was evaluated using different reference datasets. 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 validation was conducted from two aspects. First, more than 30 years *in situ* measurements over 118 stations were employed to assess the sensor-to-sensor consistency. Second, medium to high resolution data (i.e., MODIS and Landsat snow) were introduced to provide great spatial coverage and investigate the general performance of AVHRR GAC snow. Furthermore, an in-depth analysis was made over the area with a great range of conditions (e.g., elevation, topography, and land cover) in order to explore the influential factors on the performance of AVHRR GAC snow.

Validated against *in situ* station observations, the overall ACC of AVHRR raw snow dataset was about 94%, which is the same as for MOD10A1. The use of temporal filter caused a slight reduction in ACC of AVHRR gap-filled snow dataset, with

overall values around 92%. AVHRR GAC raw snow is slightly underestimated, with the bias of 0.94. Based on the observations of all *in situ* sites, we obtain HSS=0.34 for AVHRR GAC snow, which is also comparable to the one by MOD10A1 (HSS=0.35). When validated against Landsat5 TM images over the whole HKH, the RMSE and R are 22.31% and 0.82, respectively. But for the highly variable sub-region "P140-R40/R41", AVHRR GAC snow presents RMSEs of 13.27% and 16% and mBias of -5.83% and -7.13 for the raw and gap-filled datasets, respectively. Their consistency with Landsat snow is reduced, with a relatively low R of 0.46 and 0.47 for raw and gap-filled snow datasets, respectively.

Regarding the temporal consistency of AVHRR GAC snow datasets, the sensor-to-sensor consistency was found to differ slightly and unsystematically in ACC and Bias throughout the time series. While the consistent slight increasing trend on HSS is noteworthy. It is important to point out that the different performance of AVHRR GAC snow datasets in different months is mainly caused by the variable amount of snow. Particularly, the performance of AVHRR GAC snow is worst in January and best in October regarding the magnitude of ACC and HSS. But when temporal stability of accuracy was considered, it performs best in November and worst in January and December regarding the ACC. While that of HSS is worst in December. The results of Bias provide different perspectives for the performance of AVHRR GAC snow. It generally overestimates snow on February, March, October, and November, which is strongly linked to the patchiness of the snow cover that isn't captured by the *in situ* data. By contrast, unbiased estimation is likely to occur in December and January when the snow cover is most continuous over greater areas.

The validation results with two independent reference data (i.e., *in situ* and Landsat) 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 snow-covered situations, the accuracy of satellite snow datasets increases with increasing SD and FSC. By contrast, in snow-free condition, accuracy decreases with increasing SD and FSC. Furthermore, the accuracy of AVHRR GAC snow is closely related to elevations, slopes, and topographical variability. And the negative influence of these factors on snow mapping accuracy is more significant over regions with high elevations. The RMSE over different elevation regions generally shows an increasing trend with slope. The effect of aspect can be ignored over the regions less than 5500 m. But over the areas higher than 5500 m, the accuracy first increases and then decreases gradually from north facing slope to south facing slope and vice versa. The effect of landcover type on snow mapping accuracy is related to elevations. Its accuracy is generally good in croplands since it is distributed only within the region of [<1500 m]. The accuracy of barren is generally not good because it is merely distributed within the range of [>3500 m]. Forests and savannas basically show comparable overall accuracy. The accuracy of grasslands shows different responses to elevations, which is the worst over regions [2500 m, 3500 m]. Its accuracy is comparable to other land cover types over relatively low elevations (<1500 m) and outperforms barren over high elevations (> 3500 m).

When it comes to the performance relative to MODIS snow products, AVHRR raw snow is comparable with MODIS snow when ACC and HSS are focused. Nevertheless, it shows distinct advantages over MODIS snow product focusing on Bias. Regarding the temporal and spatial behaviours, different results appear in the two dimensions. In the temporal dimension, AVHRR snow datasets display a more stable behaviour regarding the ACC but less stable regarding the HSS than MODIS

snow products. But in the spatial dimension, AVHRR snow datasets show a comparable spatial 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 12.8% and 17.0%, mBias of 0.06% and 0.94%, and R of 0.63 and 0.53 for AVHRR raw and gap-filled snow datasets, respectively.

This study represents the first validation of the unique daily AVHRR GAC snow extent spanning 4 decades over the HKH. Although the 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.

6 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 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 SCFG is openly accessible here https://catalogue.ceda.ac.uk/uuid/5484dc1392bc43c1ace73ba38a22ac56.

7 Author contributions

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

8 Competing interests

The authors declare that they have no conflict of interest.

9 Acknowledgements

The authors are grateful to the ESA CCI (Climate Change Initiative) cloud project team to make the data sets available for this study. The *in situ* data were provided by the China Meteorology Administration (CMA) (https://data.cma.cn/en).

10 Financial support

This work was jointly supported by the National Natural Science Foundation of China (Grant No. 42071296 and 41801226) and the European Space Agency (ESA) Snow Climate Initiative (CCI+) project (grant 4000124098/18/I-NB).

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. 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.

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.

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.

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.

Foster, J. L. , Hall, D. K. , Eylan De R, J. B. , Riggs, G. A. , Nghiem, S. V. , & Te De Sco, M. , et al. (2011). A blended global snow product using visible, passive microwave and scatterometer satellite data. International Journal of Remote Sensing, 32(5-6), 1371-1395.

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X.: MODIS Collection 5 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 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., 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.

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, S. , Jiang, L. , Shi, J. , Wang, G. , & Liu, X. . (2018). Assessment of modis-based fractional snow cover products over the tibetan plateau. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, PP(2), 1-16.

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.

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 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.

Immerzeel, W. W. , Droogers, P. , SMD Jong, & Bierkens, M. . (2009). Large-scale monitoring of snow cover and runoff simulation in himalayan river basins using remote sensing. Remote Sensing of Environment, 113(1), 40-49.

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.

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.

Marco, T. , & Jeyavinoth, J. . (2016). A new operational snow retrieval algorithm applied to historical amsr-e brightness temperatures. Remote Sensing, 8(12), 1-25.

Metsamaki, S., Pulliainen, J., Luojus, &K., et al.: Introduction to globsnow snow extent products with considerations for accuracy assessment, Remote Sens. Environ., 156, pp. 96-108, 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.

Naegeli, K.; Neuhaus, C.; Salberg, A.-B.; Schwaizer, G.; Wiesmann, A.; Wunderle, S.; Nagler, T. (2021): ESA Snow Climate Change Initiative (Snow_cci): Daily global Snow Cover Fraction - snow on ground (SCFG) from AVHRR (1982 - 2019), version1.0. NERC EDS Centre for Environmental Data Analysis, 12 May 2021. doi:10.5285/5484dc1392bc43c1ace73ba38a22ac56.

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.

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, 2016a.

Riggs, G. A., Hall, D.K. and Román, M.O.: VIIRS snow products user guide for Collection 1 (C1), available at: http://modissnow-ice.gsfc.nasa.gov/?c=userguides, last access: 8 March 2017, 2016b.

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 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.

Shan, L. U. , Oki, K. , & Omasa, K. (2016). Mapping snow cover using avhrr/ndvi 10-day composite data. Journal of Agricultural Meteorology, 60(6), 1215-1218.

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. 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: https://www.researchgate.net/post/What_value_of_Heidke_Skill_Score_is_practically_good_for_categorical_precipitation_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 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: 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 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), 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.

Yang, G. , Ning, L. , and Yao, T. Evaluation of a cloud-gap-filled modis daily snow cover product over the pacific northwest usa. J. Hydrol., 404(3-4), 157-165, 2011.

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 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.