Thank you for providing us valuable suggestions and they do help improve the paper. According to the reviewers' comments, we revised the paper carefully and tried to give satisfactory answers to the reviewers' questions. The corresponding modifications are highlighted with red font in the revised paper.

The summaries of the revision for this paper are as follows:

First, the AVHRR GAC snow dataset have been updated with the final released version, which is based on based on the algorithm SCAMOD (Metsämäki et al. 2015). Furthermore, we extend the description of AVHRR GAC snow retrieval in the revised manuscript.

Second, more in-depth analysis was made regarding the performance of AVHRR GAC snow over different elevation regions (0-200, 200-500, 500-1500, 1500-2500, 2500-3500, 3500-4500, 4500-5500, >5500). Furthermore, the effect of landcover type, slope, aspect, and topographical variability were analyzed for different elevation regions.

Third, the structure of the manuscript has been improved. The accuracy of MODIS based on in situ sites was discussed along with AVHRR in Section 4.1. Furthermore, the comparison between AVHRR GAC and MODIS snow regarding the accuracy and temporal stability is also presented in this section. The comparison between AVHRR GAC snow and MODS snow regarding their absolute values as well as the comparison between AVHRR GAC snow and Landsat snow were presented in Section 4.2 (Comparison based on medium to high resolution data).

Fourth, based on the new results and analyses, more comprehensive conclusions were presented.

Fifth, some details, including figures and grammar, were improved.

For the specific comments for each reviewer, we have made detailed reply as following.

Reviewer2:

General comments

Extensive work has been done regarding a matter of the greatest importance. Being able to have an extensive temporal and spatial serie of snow cover over the Himalayan Hindukush region is a great challenge. But if most of the study shows interesting results, some key components are not considered enough, not discussed enough throughout the paper. What we see are good results hiding probable bad ones, the spatial weight of a "friendlier" topography to AVHRR GAC 4 km compensating on the large scale for the variability of the small and abrupt Himalayan range over the study area. Specifically, sections 2.2, 3.3, 4.2 and 4.3 raised questions, and a conclusion (section 5) that lacks the raised problem. There is a need to go more in depth in the results and discussion of this highly mountainous area, because what is presented does not show confidence into the use of the product over the Himalaya, and by extent over any high mountain range. Even though the rest of the results shows interesting work. This paper is not ready for publication.

Re: Several improvements have been made in the revised manuscript:

First, the AVHRR GAC snow dataset have been updated with the final released version. Because the final AVHRR GAC snow data published and accessible for everyone is different from what we have previously employed in the paper. Our team have improved the retrieval algorithm, because there was a need to retrieve also snow on ground with an identical procedure as for viewable snow. The final AVHRR GAC data dataset (openly accessible here https://catalogue.ceda.ac.uk/uuid/5484dc1392bc43c1ace73ba38a22ac56) in the whole time series was based on the algorithm SCAMOD (Metsämäki et al. 2015). Consequently, many results and conclusions have been reworked.

Second, we have made more in-depth analysis of the performance of AVHRR GAC snow. In particular, the study area has been divided into eight groups according to their elevations (0-200, 200-500, 500-1500, 1500-2500, 2500-3500, 3500-4500, 4500-5500, >5500) in order to take the topography into consideration. Furthermore, the effect of landcover type, slope, aspect, and topographical variability were analyzed for different elevation regions.

Third, the structure of the manuscript has been improved. The accuracy of MODIS based on in situ sites was discussed along with AVHRR in Section 4.1. And the comparison to MODIS regarding the accuracy and temporal stability is also presented in this section.

Last but not the least, we would like to point out that different reference data sources (i.e., in situ, Landsat, MODIS snow) were employed in this paper in order to assess the performance of AVHRR GAC snow from different perspectives (absolute accuracy, sensor-to-sensor consistency, spatial distribution of accuracy, as well as the influential factors on accuracy). It is true that the spatial weight of Landsat and MODIS snow to AVHRR GAC pixel scale may compensate for the variability of the small and abrupt Himalayan range. Nevertheless, from the perspective of validation of AVHRR GAC snow, the spatial aggregation of Landsat and MODIS snow is necessary, because there is significant spatial scale mismatch between AVHRR GAC and Landsat as well as MODIS. The evaluation of AVHRR GAC snow is an urgent need since this dataset is unique as it spans 4 decades and thus provides information about an ECV at climate-relevant time scales.

- Specific comments

Section 2.2

There is a lack of basic informations about the product used, we should have more to get a better hold of it. Channels used to build the product, spatial resolution, … Even though your work was not to build it, it would be of importance to have some basic informations about it. You describe MODIS and Landsat products more in depth than the product you actually want to validate.

Re: In the revised manuscript, we gave a detailed description of AVHRR GAC snow as "*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." in **Section 2.2**.

I understand the advantage of using AVHRR GAC because of the temporal resolution and time serie, **but I am not convinced the 4km spatial resolution to be an adequate choice for those regions with very high topographic variability.** Studies generally tend to discard AVHRR for the specific purpose you try to use it for (snow cover in mountainous regions), so you really need more justification to convince of the interest of such a low resolution product compared to the high topographic variability within a pixel. (an exemple here : Sharma, V. et al (2014). Topographic controls on spatio-temporal snow cover distribution in Northwest Himalaya.International journal of remote sensing, 35(9), 3036-3056.)

Re: We would like to point out the AVHRR GAC snow is a global product for all land areas, excluding Antarctica and Greenland ice sheets. It provides daily products for the period 1982-2019, which is very important for climate-relevant studies. In fact, this is the best spatial resolution over such a long-time scale at daily resolution available. It is important to note that the validation over HKH is a typical representation of its performance over mountainous area.

The HKH was selected as the study area partly because of its particular sensitiveness to climate change and thus reliable daily snow cover data across this area are in great demand, and partly because this area is featured by rich diversity of climates, hydrology, ecology, biology, and topography. Then it provides a favorable condition to explore the influential factors (e.g., elevations, landcover type, slope, aspect, and topographical variability) on the accuracy.

You use The fractional snow retrieval method by Salomonson and Appel (2006), at a different spatial resolution that the modis 500m product spatial resolution it was developed for. Both for AVHRR GAC at 4 km, and Landsat at 30 m. Were there any issues changing the scale of the spatial resolution for the application of this methodology ? would be interesting to discuss this matter

Re: It is important to note that the final AVHRR GAC snow adopted by the revised manuscript is different from what we have previously employed. Our team have improved the retrieval algorithm, because there was a need to retrieve also snow on ground with an identical procedure as for viewable snow. The final algorithm was developed and improved based on the ESA GlobSnow approach described by Metsämäki et al. (2015) and complemented with a pre-classification module.

Although the method by Salomonson and Appel (2006) is originally designed for MODIS FSC products with a mean absolute error of less than 10% (Salomonson and Appel, 2004). We assumed that such an accuracy can be achieved with higher resolution data in this paper. This treatment follows the recommendations of (Metsamaki et al., 2015), which applied fractional snow method by Salomonson and Appel (2006) to Landsat data for the evaluation of coarse-pixel snow extent products.

In order to clarify this point, we have added the sentences as "*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.*" in Section 2.3.2 in the revised manuscript.

You don't discuss much topography… 0m to 8000m in the   areas you use, gonna have a huge impact on SD , especially as viewed by satellites

Re: In the revised manuscript, we have discussed the topography in depth by dividing the *in situ* sites into five groups according to their elevations. The detailed analysis was added in **Section 4.1.4** as "*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.*".
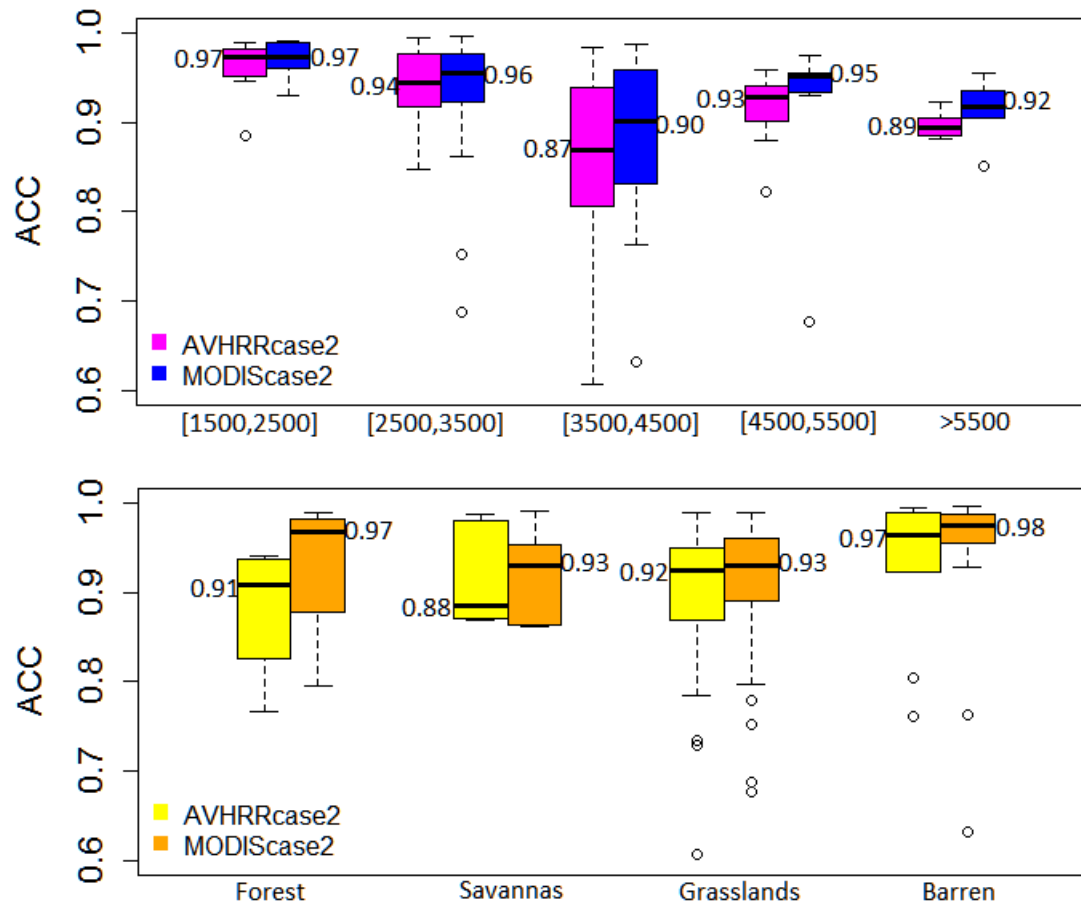
Figure 8. Boxplots of ACC from the direct comparison with *in situ* site observations over different elevation regions (a) and land cover types (b) for AVHRR raw and MODIS snow data.

Section 3.3

By resampling and projecting modis FSC 500m to AVHRR GAC 4km pixels in order to compare their absolute values, you clearly loose a lot of information in the highly variable topographic areas. But you don't discuss / show a comparative analysis of MODIS at 500m versus MODIS at 4 km to assess the accuracy of the resampleded product. Might be of importance especially in the more mountainous areas.

Re: We admit that resampling and projecting modis FSC 500m to AVHRR GAC 4km will loss lot of information in the highly variable topographic areas. Nevertheless, as we have explained before, the spatial aggregation of MODIS snow to AVHRR GAC snow pixel is necessary in validation, because there is significant spatial scale mismatch between them.

As suggested by the reviewer, we have made a comparative analysis of MODIS at 500m versus MODIS at 4 km to assess the accuracy of the resampled product using several clear-sky Landsat TM data over the study area. The evaluation results were summarized in the following table. It can be found that compared to the 500 m MODIS snow, the 4 km MODIS snow consistently shows a higher accuracy at different times of the year.

Table. Summary of accuracies of 500m MODIS and the aggregated MODIS at 4 km pixel scale with Landsat5 TM snow over "P140-R40" through scene-by-scene comparison.

	500 m MODIS snow			4 km MODIS snow		
	RMSE(%)	mBias(%)	R(%)	RMSE(%)	mBias(%)	R(%)
2010.01.19	13.88	-3.52	0.54	9.38	-3.23	0.74
2010.02.04	15.26	-4.17	0.50	10.74	-3.96	0.69
2010.12.05	16.00	-4.44	0.48	10.45	-4.13	0.69
2010.12.21	16.51	-5.01	0.45	11.68	-4.85	0.63

Section 4.2

What your results are showing, essentially, is that using AVHRR in high mountainous areas with strong variations of slope, elevation, … within a pixel, just doesn't work. RMSE values are over the roof (fig. 12), which is also quite explicit in fig. 10. You get overall good results for most of the study area, but not for the Himalayan range, and therefore cannot write that the results are good there as well.

Re: This part has been rephrased in the revised manuscript.

You need a more specific analysis to show that your results are good enough, putting both of the Himalayan part of the tile and the further Tibetan highland part under the same "mountainous area" category doesn't make sense as their geographical caracteristics aren't alike.

Re: In order to take the topography into consideration, the study area have been divided into eight groups according to their elevations (0-200, 200-500, 500-1500, 1500-2500, 2500-3500, 3500-4500, 4500-5500, >5500) (Fig. 10). Furthermore, the effect of landcover type, slope, aspect, and topographical variability were analyzed for different elevation regions (Fig. 10).
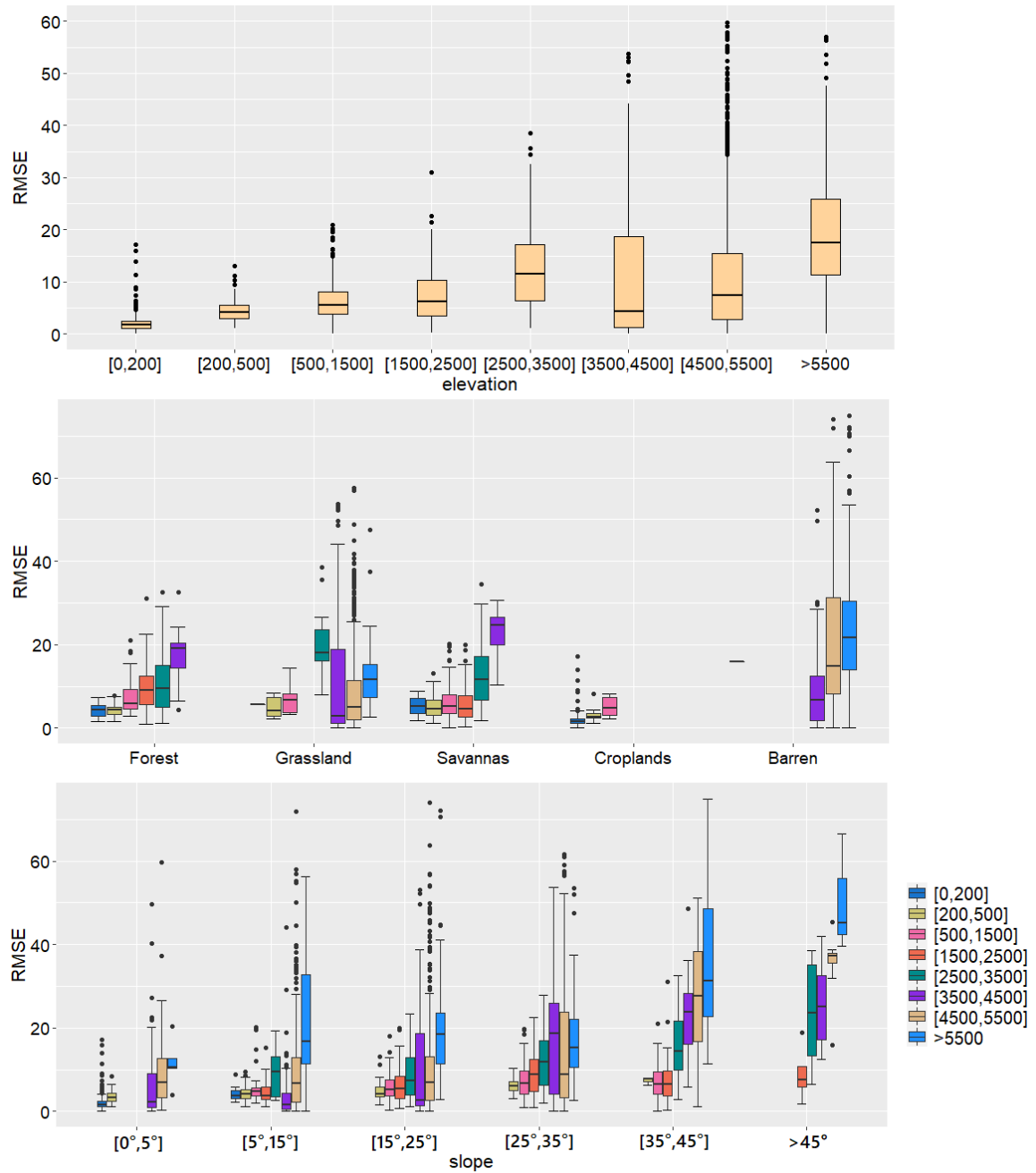
Figure 11. The variation of RMSE in dependence of a) elevation, b) land cover, c) slope, d) aspect and e) topographical variability. The different colors refer to the 8 different elevation classes. The plots show combined results of "P140-R40" and "P140-R41".
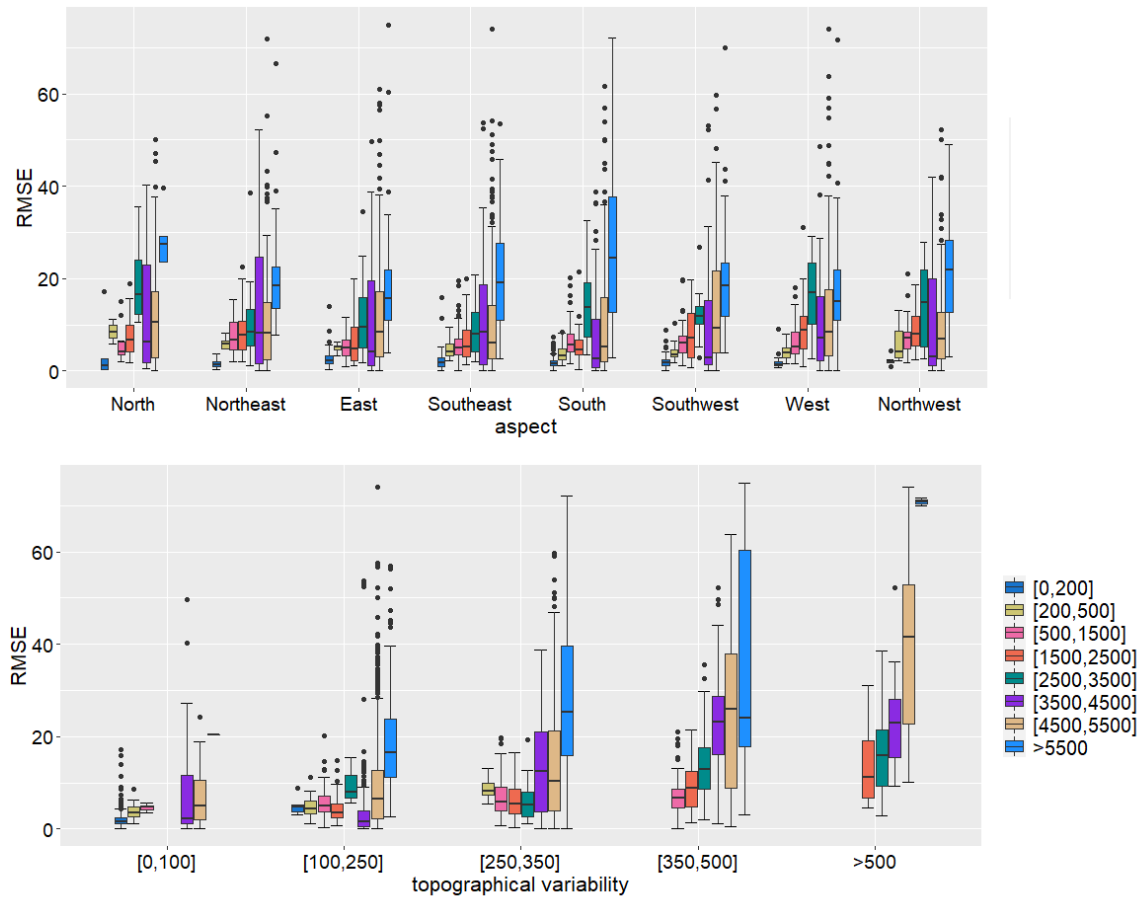
Figure 11. continued.

In order to make an in-depth analysis of the performance of AVHRR GAC snow over different conditions, we have added a new section (4.2.3 Pixel-based comparison and potential influential factors on accuracy) in the revised manuscript as follows:

*"**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). …… 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.…… 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).

The effect of slope on the accuracy of AVHRR GAC snow datasets is clearly shown in Fig.

10(c). …… 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.

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.……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).……… 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. "

Section 5

Conclusion reflects the study, showing overall convincing results over the study area but lacking in depth work regarding the Himalayan range and the errors that go with. Work is in my opinion not complete if you want to include highly variable topographic areas in the spectrum of your validation.

Re: As we explained before, more in-depth analyses was made regarding the effect of topography in the revied manuscript. We have added the analyses and discussion about the influence of landcover type, slope, aspect, and topographical variability over different elevation regions (0-200, 200-500, 500-1500, 1500-2500, 2500-3500, 3500-4500, 4500-5500, >5500) in **Section 4.2.3**.

Furthermore, more comprehensive conclusions were presented in **Section 5** based on the new analyses as "*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).*".

- Technical corrections

Beware of the over use of logical connections at the beginning of your sentences. Makes the manuscript sometimes hard to read.

Re: Regarding the writing issues, this paper has been polished by a native English speaker. Furthermore, we have also made a very thorough check of the new manuscript.

Section 4.3

Difficult to follow the discussion in a reader point of view, having to scroll back and forth to get to the plots. Would have been easier to read if the MODIS comparison was discussed along with the AVHRR, as a comparison to MODIS is one of the key point of your study to validate the AVHRR product.

Re: As suggested by the reviewer, we have adjusted the structure of the manuscript as follows:

First, the accuracy of MODIS based on in situ sites was discussed along with AVHRR in Section 4.1. Furthermore, the comparison between AVHRR GAC and MODIS snow regarding the accuracy and temporal stability is also presented in this section.

Second, the comparison between AVHRR GAC snow and MODS snow regarding their absolute values as well as the comparison between AVHRR GAC snow and Landsat snow were presented in **Section 4.2** (Comparison based on medium to high resolution data). The former is in Section 4.2.1 (Quantitative comparison to MOD10A1). While the latter was displayed in Section 4.2.2 (Spatial consistency of snow cover extent).

Third, we have added a new section (**4.2.3** Pixel-based comparison and potential influential factors on accuracy) to analyze and discuss the influence of landcover type, slope, aspect, and topographical variability over different elevation regions.