Responses to Referee 1:

Thank you for your positive comments and helpful suggestions. This document provides point-by-point answers to your remarks. We have worked in particular on:

- describing the discrimination algorithm of snow and cloud cover clearly.
- emphasizing the dataset and weakening the methodology in the Conclusion

In the document, lines in **bold** echo your comments for ease of reading, the revised manuscript with new elements in green.

We sincerely hope that these corrections can reach your expectations.

This paper proposes an a long-term AVHRR snow cover extent product from 1981 until 2019 over China. The product has the spatial resolution of 5-km and the daily temporal resolution, and is a completely gap-free product, which is produced through quality control, cloud detection, snow discrimination, and gap-filling. The validations based on ground measurement and Landsat-5 snow maps both demonstrate its higher accuracy than that of the JASMES AVHRR product. As a long-term record, the dataset will provide a valuable data source for analyzing the influence of climate changes on the cryosphere on multiple time scales. The need for such a dataset is well justified and the authors cite ample relevant literature. The paper is basically well-written and presented.

1. it is best to delete "Using the Google Earth Engine (GEE) platform" in the first sentence because GEE is just the platform of producing the product, it is not the main contribution of the study. However, I suggest to add a subsection in the section of "data and preprocessing" to describe the computing platform and the reason of choosing GEE.

Response: Thanks for this helpful suggestion. We have deleted "Using the Google Earth Engine (GEE) platform" in the first sentence, described the advantages of the GEE platform, and explained why the GEE platform is used.

A long-term AVHRR snow cover extent (SCE) product from 1981 until 2019 over China has been generated by the snow research team in the Northwest Institute of Eco-Environment and Resources (NIEER), Chinese Academy of Sciences.

Google established the Google Earth Engine (GEE) cloud computing platform from 2012. GEE enables academics to quickly access massive amounts of remote sensing data without downloading it, which could support scientific analysis and visualization of geospatial datasets with petabyte-scale (Gorelick，2012). In this study, all AVHRR SR V4 images were processed by GEE cloud platform.

New reference: Gorelick, N.: Google Earth Engine. Gebruiker Woody Bousson/kladblok, 2012.

2. why not add the dataset of 2020 year?

Response: Thanks for your suggestion. In the study, the NOAA Climate Data Record (CDR) of AVHRR Surface Reflectance Version 4 (AVHRR SR V4) was used as primary input data. The AVHHR CDR V4 used in our work is available from 1981-06-24T 00:00:00 to 2019-05-16T 00:00:00. Our work also indicated that the low quality of AVHHR after 2018 has a noticeable impact on the accuracy of snow cover extent products. This dataset developed mainly focused on the data before 2000. After 2019, the AVHRR CDR V4 product will no longer be produced due to satellite data quality issues. To extend the time series, we prefer to choose newly launched satellite sensors, like MODIS or NPP products.

3. Is it sufficient enough to validate so big product using just eight Landsat-5 images?

Response: We added the description of the three major snow-covered regions (Northern Xinjiang, Northeast China, and Qinghai-Tibet Plateau) of China in figure 1. In addition, we added two new validation images in Northeast China, and removed a similar validation image. A total of 9 Landsat-5 images distributed in the three major snow-covered regions were selected as validation data. And the detailed information of each verification image is shown in new table 9 (All serial number of images are rearranged). The main factors for snow cover detection are the type of land cover and topography. We chose three sceneries in Northern Xinjiang, one scene is located in the flat area of Altay (C8), one scene is in the Altai Mountains (C7), and the other scene is in the Tianshan Mountains (C9). The Northeast China, snow area is relatively flat. The snow is mainly distributed in forest areas, cultivated land and grassland, so we choose two forest area images (C1, C3), located in Greater Khingan Range area and Small Khingan Range area, respectively, with one farmland (C2) and grassland in Inner Mongolia (C4). The snow cover of the Qinghai-Tibet Plateau is mainly distributed in mountainous areas. And the verification data we selected are mainly distributed in mountainous areas (C5, C6). The main types of land cover are grassland and bare land. Considering the above factors, a total of 9 Landsat-5 validation images are representative and can meet the requirements.

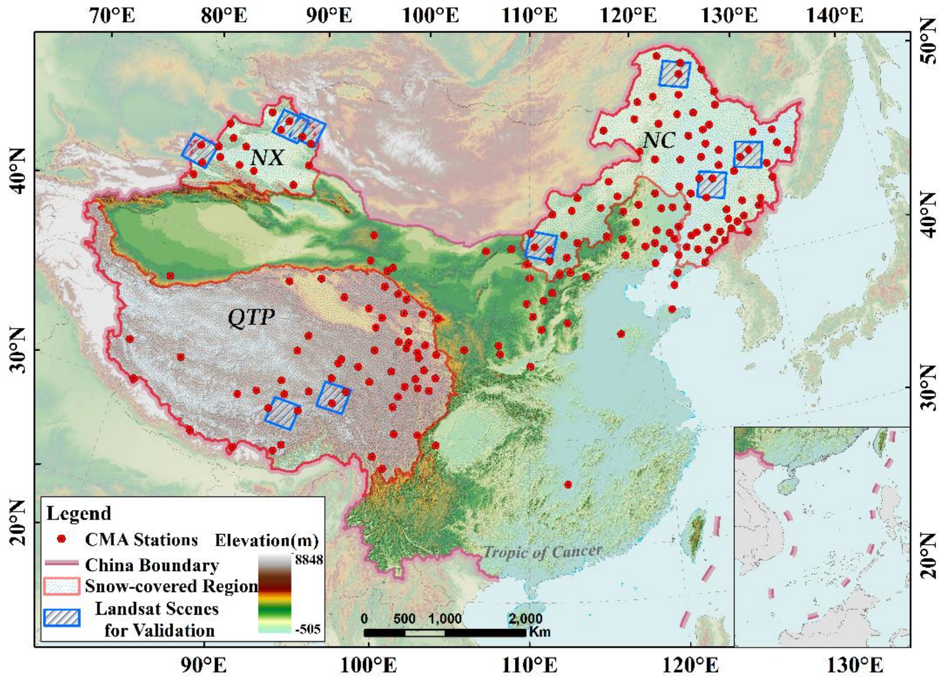

Figure 1: The geographic location of study area and the spatial distribution of major snow-covered regions, climate stations and Lansat-5 validation dataset. The elevation data were derived from Shuttle Radar Topography Mission (SRTM).

Another validation with reference to higher-resolution snow maps derived from Landsat-5 Thematic Mapper (TM) images demonstrates an overall accuracy of 87.3%, a producer's accuracy of 86.7%, a user's accuracy of 95.7%, and a Cohen's kappa value of 0.695.

A total of 9 Landsat-5 TM snow maps were used as the validation dataset (Fig.1). The training and validating samples were evenly distributed across China's major seasonally snow-covered regions (including North Xinjiang, Northeast China, and the Qinghai-Tibet Plateau) to ensure reliability and representativeness.

In the study, 9 Landsat-5 snow maps were used to further evaluate the NIEER AVHRR product. Table

9 gives the validation results of our maps versus the Landsat-5 TM SCE maps. The OA was as high as 87.3%. The high UA and low PA revealed that the product has a slight tendency to underestimate the snow cover extent. The CK value (0.695) of the 'area to area' method also demonstrated 'substantial' agreement, which was close to that of ground measurements validation (0.717).

Figure 8 further displays three intuitional examples demonstrating the detailed difference between NIEER AVHRR SCE maps and Landsat-5 SCE reference maps. The three images (serial number "C1, C5, and C8") were located in Northeast China, the Qinghai-Tibet Plateau, and North Xinjiang, respectively.

Considering the limitations of point-to-area validation, the overall OA, PA, UA, and CK values were 87.3%, 86.7%, 95.7%, and 0.695, respectively, using Landsat-5 TM area-to-area, which showed the same trend of accuracy as the point validation.

Table 9 The accuracy of NIEER AVHRR SCE maps versus Landsat-5 TM SCE maps. C1~C9 denotes the different Landsat-5 TM SCE.

Path/row	Serial number	Date	Cloud percentage	Snow percentage	OA	PA	UA	CK
116028	C1	19970312	2.0%	77.2%	87.9%	88.3%	95.9%	0.678
118029	C2	20161109	0.2%	88.0%	84.5%	82.8%	99.5%	0.519
121024	C3	20160319	2.0%	96.4%	98.1%	100.0%	98.1%	1
127031	C4	20180130	1.1%	45.3%	82.0%	63.0%	96.1%	0.626
135038	C5	19961109	1.0%	66.5%	79.5%	81.0%	87.9%	0.552
137039	C6	19961123	2.0%	50.7%	78.2%	65.7%	88.5%	0.566
142027	C7	19870323	0.0%	96.1%	97.2%	100.0%	97.2%	0.036
143027	C8	20051110	2.0%	48.6%	93.1%	86.7%	99.8%	0.863
147029	C9	20160222	1.1%	89.0%	90.6%	91.4%	98.0%	0.587
		Total			87.3%	86.7%	95.7%	0.695

4.Line 17, NIEER is the abbreviation of the authors institute, "the new NIEER product" is difficult to project to the new AVHRR snow cover extent product, because many products are produced by the institute.

Response: Thanks for your helpful suggestion. Referring to NASA NSIDC MODIS series, we plan to publish some snow products, including MODIS SCE, AVHRR SCE, and MODIS FSC. We revised "the new NIEER product" by "the NIEER AVHHR SCE products" to distinguish from the other products.

The NIEER AVHHR SCE product has the spatial resolution of 5-km and the daily temporal resolution, and is a completely gap-free product, which is produced through a series of processes such as the quality control, cloud detection, snow discrimination and gap-filling (GF).

5.Line 27, "nearly 40%" should be changed to precise number.

Response: Thanks for the careful check. According to our validation results, we have changed “nearly 40%” into “60.8%”.

For example, the overall accuracy of our products was 15% higher than the well-known JAMES AVHRR product. The omission error dropped from 60.8% to 19.7%, the commission error dropped from 31.9% to 21.3%, and the CK value increased by more than 114 percent.

6.Line 37, "these decades" refers to which decades? The citing references was published in 2005 and 2018.

Response: Thanks for your helpful suggestion. We have replaced “over these decades” with “over the past several decades”. Generally speaking, over these days, years or decades indicate during the past several days, years or decades.

With the continuous warming of the global climate, snow cover on the Earth has been shrinking evidently over the past several decades (Barnett et al., 2005; Bormann et al., 2018).

7.Line 38, "is not only of particular importance for climate research but also an indispensable

indicator of climate change" needs to be rephrased.

Response: Thanks for your suggestion. We improved this sentence, as marked as red in lines 38-39.

Therefore, long-term snow cover data are particularly important for climate research and are also a crucial indicator of climate change.

8.Line 56, the abbreviation SCE has been described in the last paragraph.

Response: Thanks for the helpful suggestion, and we deleted it.

The Japan Aerospace Exploration Agency (JAXA) recently issued the long-term SCE product JASMES with a spatial resolution of 5 km throughout the Northern Hemisphere.

9.Line 99, the sensor attenuation refers to AVHRR? Please specify it. Then, why the AVHRR sensor attenuation requires the algorithm to choose different TM images? And how different?

Response: Thanks for your excellent question. Although AVHHR CDR products have been calibrated, the reflectance test of the Antarctic ice sheet at Dome-C (75°6’ 0” S, 123°21’0” E) reveals that the attenuation is relatively low in the visible band but high in the near-infrared band, which leads to different product quality. The results are shown in figure (1), (2) and (3). Considering the difference in image data before and after 2000, we trained two sets of corresponding decision trees using different TM images, respectively.

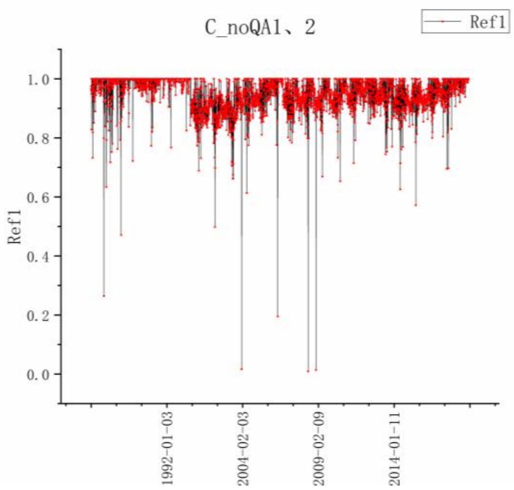

Figure (1) The reflectance of band 1 from AVHRR SR V4 at Dome-C on the Antarctic ice sheet from 1981 to 2019.

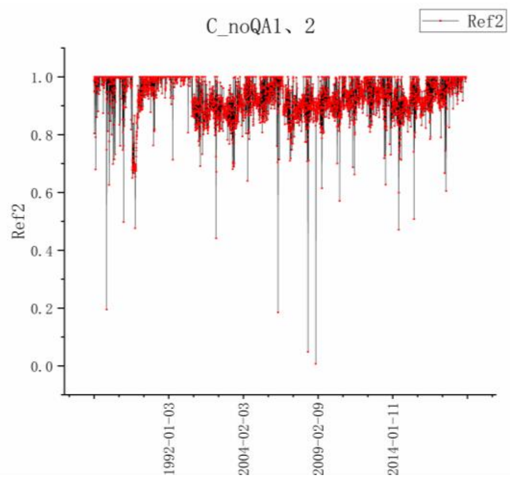

Figure (2) The reflectance of band 2 from AVHRR SR V4 at Dome-C on the Antarctic ice sheet from 1981 to 2019.

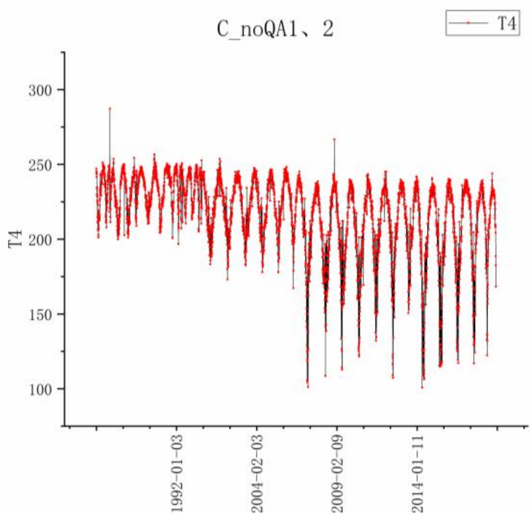

Figure (3) The brightness temperature of BT11 from AVHRR SR V4 at Dome-C on the Antarctic ice sheet from 1981 to 2019.

10. Line 103, where is the main seasonally snow-covered areas in China? Please specify it. This means move the content at Line 122 here.

Response: Thanks for this insightful suggestion, and we have added the content of major seasonally snow-covered regions in Line 102-104. We have also deleted the description of the three major snow regions in lines 120 to 121.

In order to ensure reliability and representativeness, the training and validating samples were evenly distributed in three major seasonally snow-covered regions across China, including North Xinjiang, Northeast China, and the Qinghai-Tibet Plateau.

The available CMA stations were evenly distributed across the three major seasonally snow-covered regions in China.

11. Line 124, the usage of the snow depth data for the proposed study (line 127) should be used as the first sentence.

Response: Thanks for your helpful suggestion, and we updated it as you suggested.

Che et al. (2008) and Dai et al. (2015) generated snow-depth data by using an inter-sensor calibration of multiple satellites' passive-microwave observations, which provides daily, 0.25-degree snow-depth data for China from 1979 to 2020. And this data set of long-term daily snow depth in China is available at http://data.tpdc.ac.cn.

12. Line 128, the cloud/snow confusion is generated by averaging the hourly ERA 5 land climate reanalysis dataset?

Response: Thanks for your helpful suggestion. Sorry! there is an obvious clerical error. The word “generated” in the sentence should be deleted.

We used the land surface temperature (LST) daily product to alleviate the cloud/snow confusion by averaging the hourly ERA 5 land climate reanalysis dataset on the GEE platform (Muñoz Sabater, 2019).

13. Line 157, did you use the NDVI? If not, the formula (1) is not needed.

Response: Thanks for your question. We do use NDVI in the cloud detection algorithm, as listed in Table 4. The cloud detection method considers elevation, NDVI, the difference of reflectance, and brightness temperature. The cloud detection method could be grouped into A and B. A included 4 subcategories and B with 7 subcategories. Among of them, B7 adopts NDVI when identifying the non-snow category.

14. Line 230, I suggest to add Huang as one of the co-authors.

Response: Thanks for your advice. We just referenced a small part of the algorithm proposed by Huang (Huang et al., 2018). Huang was not involved in our work.

15.Line 267-268, how to determine these thresholds?

Response: A good question. Elevation classification refers to the MODIS global snow cover algorithm from NDIDC (Riggs et al., 2006). And the land surface temperature (LST) is obtained from our actual training samples. We count the distribution of BT11 (bright temperature in snow and no snow samples (Figure (4)), and use the BT11 condition (T4 <274.8 K) that can identify 95% of snow samples as the threshold, so we set one of the LST thresholds of the temperature mask in the paper to 275 K. The definition of 281 K for warm snow is derived from a large number of snow samples observed on the Tibetan Plateau.

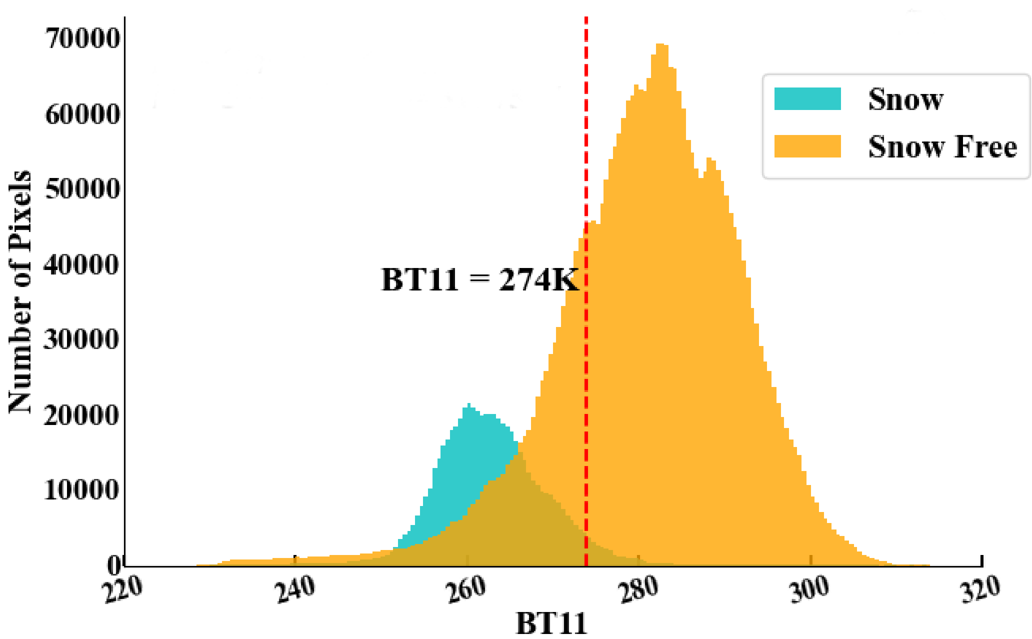

Figure (4) The histogram of snow and snow-free pixels on BT 11 from AVHRR. The cyan samples are snow and the yellow samples are snow free.

Reference: Riggs, G.A., Hall, D.K., & Salomonson, V.V. (2006). MODIS Snow products user guide to collection 5. In. http://modis-snow-ice.gsfc.nasa.gov/?c=userguide

16. Line 291, what did cause the poor quality of raw satellite?

Response: A good question. The design of satellite sensor had limited presence life. There are inconsistencies in the calibration coefficients due to sensor changes at a later stage, which are the major reasons for the deterioration of data quality.

17. Line 353, "poorly" is not good to describe the others' product.

Response： Thanks for your helpful suggestion. The “poorly” has been revised to “much worse”.

The JASMES SCE products performed much worse, with total OA, PA, UA, and CK values amounting to 71.8%, 39.2%, 68.1%, and 0.321, respectively.

18. Line 375, what is the meaning of "GF"? Why not delete it?

Response： Thanks for pointing this out. GF is the abbreviation of gap-filling. We have modified the abstract and body where gap-filled first appears.

Line 17-20：The NIEER AVHHR SCE product has the spatial resolution of 5-km and the daily temporal resolution, and is a completely gap-free product, which is produced through a series of processes such as the quality control, cloud detection, snow discrimination and gap-filling (GF).

Line 79-80：(5) Improved gap-filling (GF) strategies are adopted to obtain complete snow coverage.

19.Line 379 and 380, why not list the names according to the increase of the values?

Response: Thanks, it has been revised according to the order.

The values in the product are classified as non-snow (0), snow from AVHRR (1), snow from HMRF (2), snow from SD (3), water (4), and filling value (255).

20. Line 383, as an ESSD paper, the conclusion should focus on the data, not the method.

Response: Thanks for your helpful suggestion. The first paragraph of section "conclusions" has been revised according to the suggestion. We focused on the product and weakened the method.

Line428: In this study, a daily AVHRR SCE product with a spatial resolution of 5 km across China from 1981 to 2019 has been generated by the snow research team in the NIEER, Chinese Academy of Sciences. The NIEER AVHRR SCE product used a multi-level decision tree algorithm for cloud and snow discrimination and an improved GF technique. The product was validated using snow depth measurements provided by the China Meteorological Administration and higher spatial resolution SCE maps derived from Landsat-5 TM.

21.For the other comments:

Line 22: "the producer's accuracy was 81.0% the user's accuracy was 81.3%, and the Cohen's kappa value was 0.717" --> "the producer's accuracy is 81.0%, the user's accuracy is 81.3%, and the Cohen's Kappa value is 0.717".

Line 47: "throughout China" should be deleted, because they are global products.

Line 52: "The polar orbit meteorological satellite of" should be deleted.

Line 84: “processing” -> “preprocessing”.

Line 101: “true values” -> “ground truth”.

Line 137: "between" should be deleted.

Line 139: “forth” -> “fourth”.

Line 175: "previous" -> "the previous". It is the same in line 272.

Line 219: "algorithm above" -> "above-mentioned algorithm", and "first" should be deleted.

Line 270: “Methodology” -> “Metrics”.

Line 315: delete "new".

Line 319: delete "the".

Line 321: “edges” -> “boundaries”.

Line 327 and 328: “snow-covered areas” -> “snow cover”.

Response: All the above comments have been revised in the text of the new manuscript. Thank you again for all of the valuable comments or suggestions.