Supplemental Material

Journal of Hydrometeorology

Using Multisource Satellite Data to Assess Recent Snow-Cover Variability and Uncertainty in the Qinghai–Tibet Plateau

https://doi.org/10.1175/JHM-D-18-0220.1

Supplementary:

1. The four-step cloud removal procedure

1) **Step 1**: use the identification of adjacent two days, involving two filters: a) Adjacent one-day filter: a cloud pixel in day *i* was reclassified to snow if both the previous and next day (*i*−1 and *i*+1) were identified as snow. Likewise, if both days were snow-free, day *i* was considered snow-free; and b) Adjacent two days filter: if one of the day *i*−1 or *i*+1 was cloudy, then check the previous or next two day's (*i*−2/*i*+2) identification. For example, if a cloudy pixel on day *i* was also covered by cloud in day *i*+1 while it had snow in day *i*−*1* and *i*+2, then it was identified as a snow pixel. Note that if both the previous and the next day (day *i*−1 and *i*+1) were cloudy, then proceed to the Step 2.

2) **Step 2**: use the identification of four spatially adjacent pixels. For the four pixels adjacent to a cloudy pixel, if at least three of them had the same identification (snow or snow-free) on the same day, then the cloudy pixel was reclassified to this dominant identification.

3) **Step 3**: create a 'permanent' snow and snow-free mask. The 'permanent' snow/snow-free is defined by the percentage of snow/snow-free days in the total cloudless days after the Steps 1 and 2 were applied. Only the pixels with total cloudless days higher than 50% were used in this method to ensure enough samples were considered. First, we calculated the snow and snow-free days for each season—winter (December, January, February), summer (June–August) and spring/fall (March–May, September–

November) for each year. The snow days in summer and snow-free days in winter were the lowest among seasons (not shown). Then, if a given pixel had snow days higher than 95% of the total cloudless days in summer, it was defined as a 'permanent' snow pixel. The 'permanent' snow-free pixels were identified in the same way with more than 95% snow-free days in winter. Lastly, all the cloud days of the 'permanent' snow/snow-free pixels in the whole year were reclassified to snow/snow-free. The 'permanent' snow estimated by this method is within the same region identified as the 'snow and ice' classification by the MODIS Land Cover Type product (MCD12Q1) (not shown) indicating a conservative and highly reliable estimation.

4) **Step 4**: use a wider temporal window of four days and the closest cloudless identification. If day i was covered with cloud, then use the closest cloudless day's identification in n days before or after day i ($i-n/i+n$, $1 \leqslant n \leqslant 4$) to reclassify the pixel. This step was not applied if both day $i-n$ and $i+n$ were cloudless but had different identification.

2. Details of the random sampling to estimate cloud removal biases

Since three of the four cloud removal methods use the temporal information, the cloud persistence distribution of FY3B was analyzed first to ensure the generated 'fake' cloud were close to the real data. Cloud persistence frequency (F_n) is defined as the occurrence numbers of exactly n consecutive cloudy days in a year for each pixel. Figure 4 shows that the spatial mean F_n decreases with the increasing n, and nine or more days covered by

cloud persistently is rare. Using the observed cloud persistence distribution, the probability (p_n) of day n being covered by cloud when the previous $n-1$ days are cloudy can be written as:

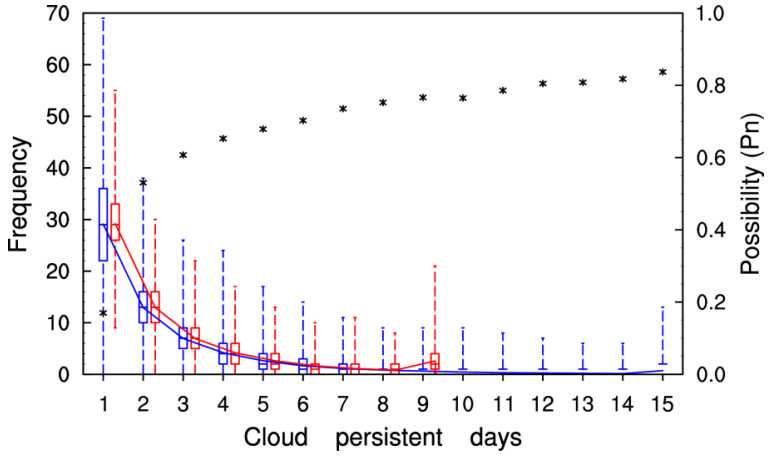
$$p_n = \frac{\sum_{i=n}^{9} F_i}{\sum_{i=n-1}^{9} F_i}, (1 \le n \le 9) \quad (1)$$

where F_i (F_n when $i=n$) is the frequency of n consecutive cloudy days. The maximum n is selected to be nine because of the low probability of observed cloudy persistence more than nine days. The spatial averaged p_n was obtained by applying Equation 2 to the FY3B (Figure 4). The p_n was then used to generate missing values randomly for the IMS data as 'fake' cloud. Results show that the 'fake' cloud-persistence distribution of IMS closely reproduces the real cloud persistence of FY3B (red line in S-figure 1).

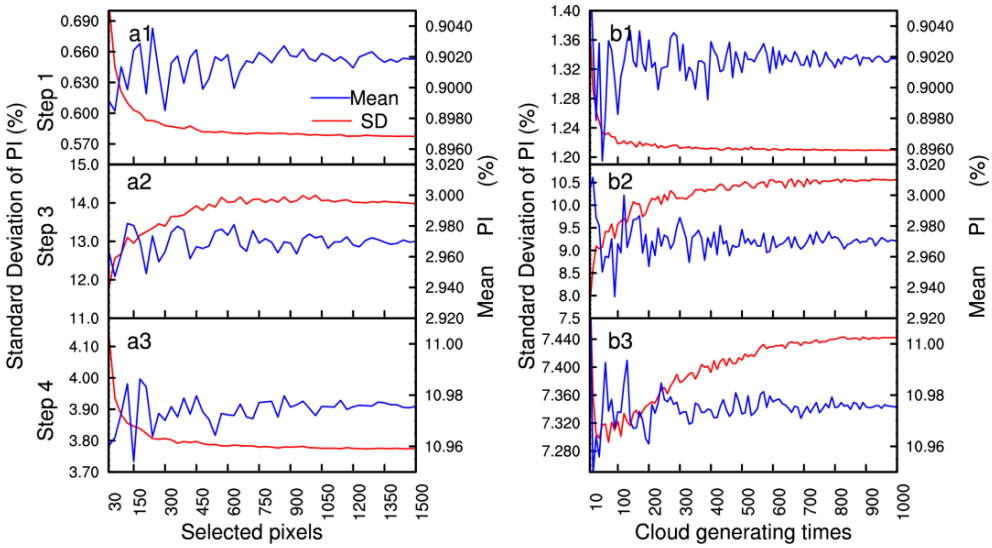
Pixels in the study area were separated into seven groups based on the annual snow days with a 50 snow-day interval: 0–50, 50–100, 100–150, 150–200, 200–250, 250–300 days and more than 300 days. To get robust results, we first selected ***n*** pixels randomly from each of these seven groups, and for each selected pixels, random missing values as 'fake' clouds were generated ***m*** times. Then, two experiments were used to estimate the uncertainties of incorrect percentage (*PI*) caused by the number of samples (the value of ***n*** and ***m***) (S-figure 2). To test parameter sensitivity, we set one of the two parameters to a large value (1500 for ***n*** and 1000 for ***m*** here), and then checked the variations of standard deviation of *PI* and mean *PI* while varying the other parameter.

Results show that the uncertainties of *PI* reduce with increasing ***n*** and ***m*** up to values of 1500 and 1000 respectively, where *PI* is stable and reliable for all cloud-removal steps. Thus, for the final PI estimate, the total number of samples for random sampling is 1,500,000 for each snow-day groups. Figure 5 also shows that the average *PI* is 0.9%, 2.5% and 10.8% for cloud-removed pixels of Step 1, Step 3 and Step 4, respectively.

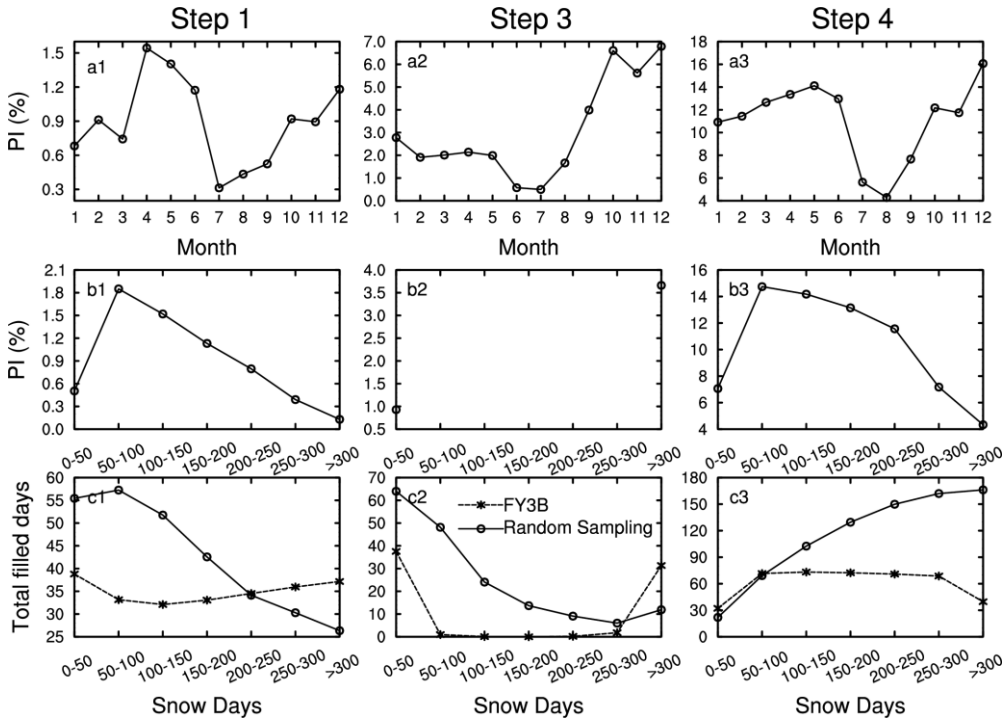
S-figure 3 shows that Step 1, 3 and 4 have lowest *PI* in July and August, and higher *PI* in snow melting time (April, May) and snow on-set time (October to December). The pixels with 50–100 snow days have the highest *PI*, while the smallest and largest snow-day groups have the lowest *PI*. Figure 6 (bottom row) also shows that the 'fake' cloud filled days of the random sampling is higher than the real-data cloud removed days (from FY3B) generally. Therefore, it is quit necessary to derive the real biases of cloud removal by multiplying *PI* to their real cloud removed percentages as shown in Table 3.



S-figure 1 The cloud persistence distribution of 0.01° FY3B (blue) and the randomly generated 'fake'-cloud persistence using p_n (red). The spatial averaged p_n is labeled in dots in the right axis.



S-figure 2 Standard deviation of *PI* (red) and mean *PI* (blue) versus the selected pixels (***n***) of each snow-day level (50<, 50–100, 100–150, 150–200, 200–250, 250–300, >300 snow days) when the cloud generating times of each pixel (***m***) is set to 1000 (a1–a3); and versus the cloud generating times of each pixel (***m***) when the selected pixels of each snow-day level (***n***) is set to 1500 (b1–b3) for Step 1, Step 3 and Step 4.



S-figure 3 Mean *PI* averaged in each month (a1–a3) and at seven snow-day levels (b1–b3); and the total cloud filled days of random sampling (c1–c3, red) and the real cloud removed days of FY3B (c1–c3, blue) in 2015 for cloud removal Step 1, Step 3, and Step 4.

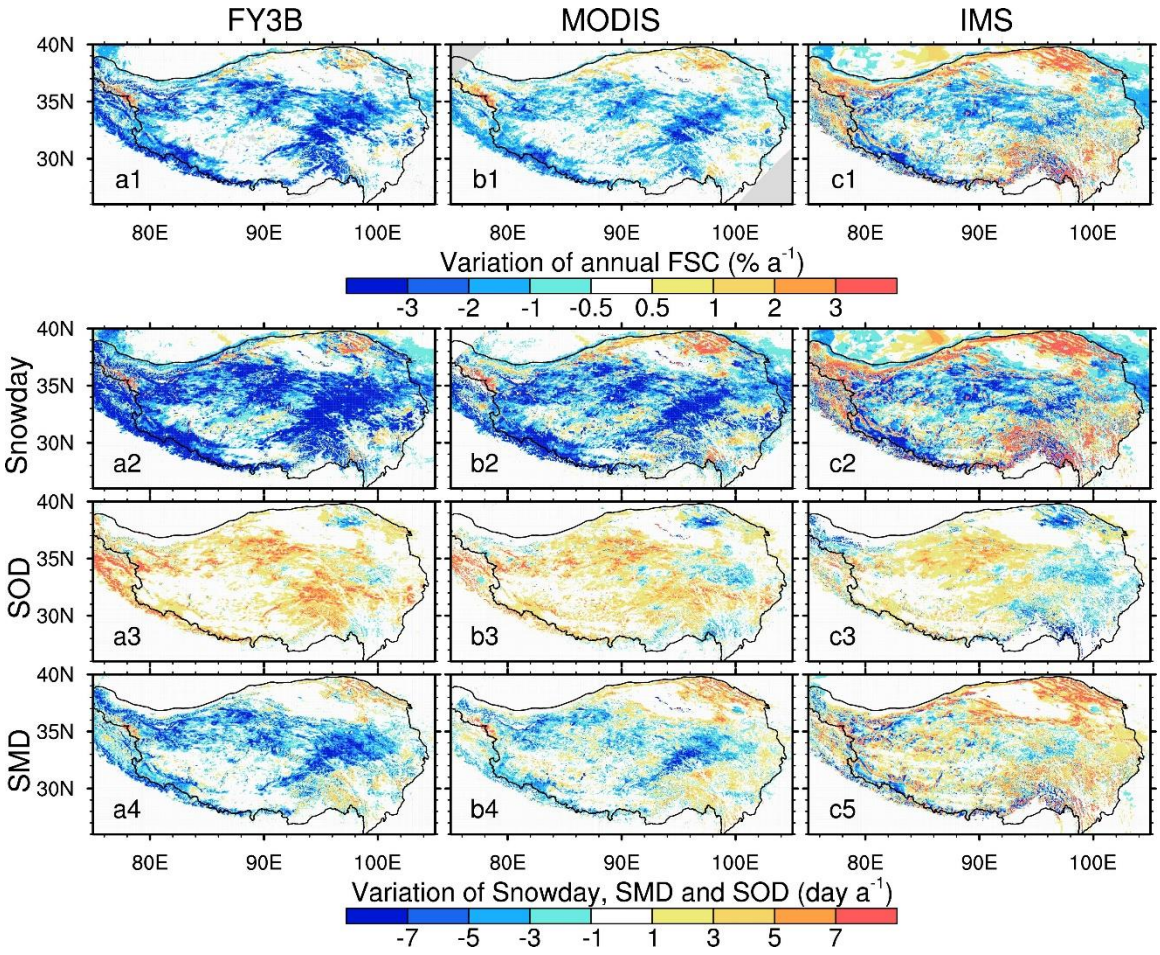
3. Snow-cover variations in 2012-2017

This part is an additional analysis for the variations of three snow-cover products, FY3B, MODIS and IMS. Since three data sets have different observation periods, we only analyzed their overlapping period in 2012-2017. For recent six years, three data sets (FY3B, MODIS and IMS) measure a decreasing annual FSC in most regions except for the Arkin and Karakoram Mountain (S-figure 4). Negative variation of snow duration show similar spatial distribution pattern with that of FSC, which is caused by the delayed snow onset with increasing SOD (S-figure 4 a3–c3) and the earlier snow melting with decreasing SMD

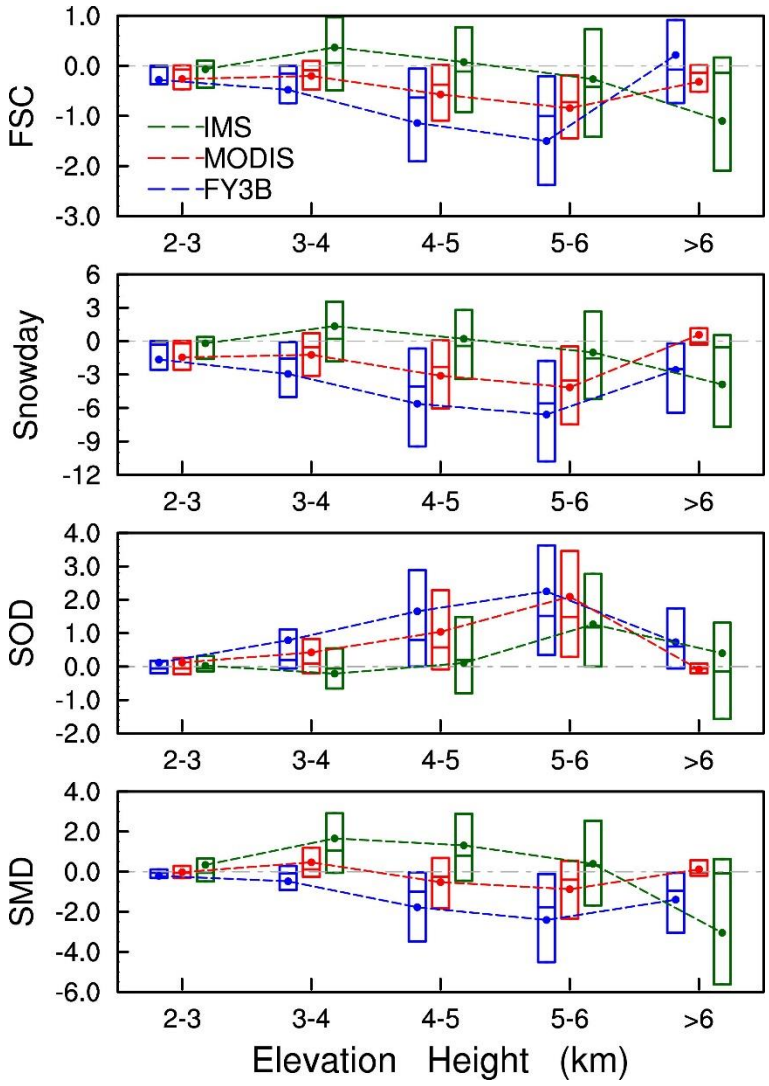
(S-figure 4 a4–c4). Despite the similarity among three data sets, IMS has more snow extent with positive variations than MODIS and FY3B, such as the eastern QTP and the Nychen Tanggula Mountain region. Meanwhile, the SMD increases for IMS snow cover at most areas, while it decreases for MODIS and FY3B, indicating a high uncertainty of SMD variations.

For each elevation zone, decreasing annual FSC and snow days are common features revealed by at least two data sets (S-figure 5). However, IMS has positive variations of annual FSC and snow days in 3–5 km elevation zones. An elevation dependence is also found for the variation trends of annual FSC and snow duration below 6 km. S-figure 5 shows that higher elevation zones have higher decreasing rates of annual FSC and snow days. The 5-6 km elevation zone has the most severe drop in annual FSC and snow duration with delayed snow onset and earlier melting.

For uncertainties, the same as Figure 10, S-figure 5 also shows that higher elevation zones have higher uncertainties below 6 km. For example, S-figure 5 shows that the differences of mean FSC variation rates between data sets increase from 0.21 to 1.32 % a^{-1} with increasing elevation zones. The uncertainties of snow-duration trends show consistent elevation distribution with that of annual FSC trends. S-figure 5 also shows that all data sets have positive trends of SOD indicating a delaying trend of snow-onset date with low uncertainties. However, the snow melting is earlier for MODIS and FY3B but delayed for IMS snow cover (S-figure 5). This indicates a higher uncertainty of snow-melting trends than snow-onset trends.



S-figure 4 Spatial distribution of the linear-regression slopes of annual FSC (% a^{-1}), snow days (day a^{-1}), SOD and SMD (day a^{-1}) in 2012-2017 for 0.04° FY3B, MODIS and IMS snow-cover data.



S-figure 5 Box plots of the linear-regression slopes of annual FSC (% a^{-1}), snow days (day a^{-1}), SOD and SMD (day a^{-1}) at five elevation zones in 2012–2017 for 0.04° FY3B, MODIS and IMS snow-cover data. The top and bottom points of the boxes are the first and the third quartile, and the horizontal short lines are the median value. The mean values are labeled in dash lines with solid dots.