

[essd-2018-144] - Theia Snow collection: high resolution operational snow cover maps from Sentinel-2 and Landsat-8 data

Please note that text citations are reported in italics.

GENERAL COMMENTS

The authors introduce a freely-available collection of snow products derived from Sentinel-2 and Landsat-8 data through a robust retrieval algorithm, which allows overcoming well-known limitations affecting the snow detection. The newly-developed algorithm is described in detail and its potential is tested and discussed through several analyses. The satellite-based dataset presented in this study is of critical importance and great usefulness not only for the scientific research community, but also for users interested in operational applications.

I have identified some issues which deserve only minor revisions of the manuscript before its publication.

SPECIFIC COMMENTS

Here are listed my main comments:

1. The manuscript contains several self-explanatory figures, which make it easier to read and understand the text. However, further references to figures need to be introduced, since for some of them the reference is even missing.
2. The manuscript comprehensively describes the newly-developed algorithm and the accuracy analysis of the resulting satellite products. However, a description of the satellite datasets used in this study (i.e., Sentinel-2 and Landsat-8) is missing. It would be of key interest to introduce the spectral bands of these satellite imagery to support the description of LIS algorithm.
3. A further section introducing the evaluation metrics should be added to better enable the assessment of the results.
4. Although the title mentions both Sentinel-2 and Landsat-8 data, references to Landsat-8 dataset are generally missing in the manuscript. Indeed, both methodology and results are actually referred to Sentinel-2 imagery.
5. The final section 6 includes information of critical importance, which should be reported and discussed in the text. Please consider my comments on Section 6.

Please consider replacing *sen2cor* with *Sen2Cor* throughout the whole text.

Section 1 (Introduction)

Pag. 1, l. 20: Please report the reference within brackets.

Pag. 1, l. 22: Please replace *12h* with 12 hours.

Pag. 2, l. 5: Please consider reporting the full name of the plan.

Pag. 2, l. 10: Please report the reference within brackets.

Pag. 2, l. 14: Please consider replacing *5 day* with 5-days.

Pag. 2, l. 19, 21: Please explicitly define and explain what level-1C and level-2A products are (i.e., TOA, BOA).

Pag. 2, l. 21: Please specify the starting sensing date of the processed L2A product provided by ESA.

Pag. 2, l. 24: What are the main limitations of ESA's L2A algorithm for snow detection?

Pag. 2, l. 28: Please consider explicitly specifying here that Theia collection relies on a combination of these two satellite datasets. Currently this information is more clearly reported only in the Conclusions (see main remark n°5 in General comments).

Pag. 2, l. 28: Please consider rephrasing the sentence as "... at 20-m and 30-m resolution, respectively...".

Pag. 2, l. 30: Please report here the reference for the Normalized Difference Snow Index.

Pag. 2, l. 31: Please close the bracket.

Section 2 (Algorithm)

Please add in the text references to Figures 1 and 2, which need to be commented (see main remark n°1 in General comments).

Please refer to Table 1 whenever an algorithm parameter is introduced.

Generally, in Section 2 the authors do not explain how the parameters values have been defined and optimized. Please add this missing information in the algorithm description.

Section 2.1 (Scope)

Currently some limitations of the ESA's processor are discussed within the description of LIS algorithm, for instance the critical issues affecting the cloud mask (i.e., thin clouds, edges of snow cover). In order to improve the readability of the manuscript and to properly highlight the potential of LIS algorithm, the authors could clearly identify and explain in this section what the main limitations of the ESA's processor are, before describing their algorithm.

Section 2.2 (Input)

In this section the input data are listed, but the data processing to derive input *L2A product* is not discussed. This section should also briefly explain how the correction of atmospheric and terrain slope effects is performed.

Pag. 3, l. 13: Please consider replacing *remote sensing images* with *remotely-sensed images*.

Section 2.3 (Pre-processing)

Pag. 3, l. 22: Please specify the original spatial resolution of the red and green bands. It would be useful to introduce a table reporting the spectral bands and their central wavelengths of the original satellite data (see main remark n°2 in General comments).

Pag. 3, l. 23: Please spell out *SRTM*.

Section 2.4.1 (Snow detection)

Are the cloud-contaminated pixels preliminary flagged and neglected in this first procedure? If so, please specify this information.

Section 2.5 (Snowline elevation)

Please consider renumbering this Section as 2.4.2.

Pag. 4, l. 15: How are the false snow detections defined? What is the reference dataset to identify false snow detections?

Pag. 4, l. 18: Sentinel-2 tile is 100 km².

Pag. 4, l. 20: Please specify how snow fraction is evaluated (fraction evaluated over the total number of snow and no-snow pixels?). Furthermore, how is defined the threshold of snow fraction f_t ?

Pag. 4, l. 23-24: How are defined d_z and f_s ?

Section 2.6 (Cloud mask processing)

Please consider renumbering this Section as 2.4.3.

Pag. 5, l. 15: Please either specify how factor r_f is defined or neglect this information by just reporting the corresponding reference.

Pag. 6, l. 2, 6: Please specify how r_D and r_B are defined.

Figure 3: Please introduce MUSCATE in text before this figure.

Section 2.7 (Implementation)

Please consider renumbering this Section as 2.4.4.

Section 3 (Data description)

Pag. 9, l. 3: The number of selected tiles is already reported in pag. 8, l. 15.

Pag. 12, l. 12: Please specify how the pixels classified as no-snow are defined. Are all snow- and cloud-free pixels classified as no-snow ones? Does the no-snow class include water pixels? Here it should be also mentioned the uncertainty in classifying vegetated pixels as snow or no-snow ones.

Pag. 12, l. 13: Please specify what pixels are classified as no-data.

Please switch Figure 5 and 6 since they are mentioned reversely in the text.

Section 4.1 (Method)

Since in this section the authors refer only to Sentinel-2 (e.g., pag. 13, l. 28; Table 3), they should explicitly specify that the evaluation is based only on Sentinel-2 data, as clearly stated in Section 6 (see main remark n°5 in General comments).

As this manuscript is focused on the mapping of snow cover, it would be more proper to limit the analysis over the snow period, not up to the end of summer season. For instance, Table 3 reports a selected pair of snow products in August, which is not of high interest for the main purpose of this research. Furthermore, it is noteworthy that, when assessing the contingency matrix, the analysis of these snow-free scenes could affect the representativeness of the results, as shown by the significantly higher number of correct negatives in Table 4.

Section 4.1.1 (Comparison with in situ snow depth measurements)

Please add in the text references to Figures 7 and 8, which need to be commented (see main remark n°1 in General comments).

Pag. 15, l. 2: How is defined the *first tile*? Are the Sentinel-2 data of the second tile neglected even when sensed in a different day with respect to the selected first tile? If so, it would entail a loss of useful remotely-sensed snow data.

Section 4.1.2 (Comparison with snow maps of higher spatial resolution)

Pag. 16, l. 11: Please remove *the*.

Pag. 16, l. 11: Could the authors mention the classification algorithms that have been tested and provide their references?

Section 4.2.1 (Comparison with in situ snow depth measurements)

To improve the assessment of the results, please report how the evaluation metrics are defined (see main remark n°3 in General comments).

Section 4.2.2 (Comparison with snow maps of higher spatial resolution)

While in Section 4.1.2 the authors explain that “*The Theia and sen2cor snow maps were compared ...*”, here Table 5 shows only the results related to Sentinel-2 dataset. Once again, please specify that the evaluations are based on Sentinel-2 data, as currently reported in Section 6 (see main remark n°5 in General comments).

Table 5: Please provide the definition of F1, FPR and FNR, how they have been evaluated and their optimal values (see main remark n°3 in General comments).

Pag. 19, l. 4-5: Please consider moving the definition of “*false positive*” and “*false negative*” pixels in Section 4.2.1, where these terms are introduced for the first time.

Section 5 (Discussion)

Pag. 21, l. 10: Please provide a reference, if available.

Pag. 21, l. 25: Please consider removing *of* preceding *the snow cover*.

Pag. 21, l. 26: According to my previous comment on Section 2, it would be useful to explain how the algorithm parameters have been defined.

Section 6 (Conclusion)

In this final section the authors provide useful information of key importance, which should be provided throughout the manuscript:

1. In the Introduction it should be explicitly specified that most snow products are derived from Sentinel-2 and only recently from an integration of Sentinel-2 and Landsat-8 data. As well, at the beginning of Section 4 the authors should explicitly state that the *evaluation was based only on Sentinel-2 products*.
2. In the section describing the methodology it should be properly explained that the analyses have been performed by considering the snow cover detected over the satellite field of view, and not at ground level (no correction of vegetation masking effect).
3. The condition of clear-sky should be explicitly mentioned as one criteria for the selection of the scenes to be analysed.