

Interactive comment on “Real-Time Snow Depth Estimation and Historical Data Reconstruction Over China Based on a Random Forest Machine Learning Approach” by Jianwei Yang et al.

Divyesh Varade (Referee)

varadedm@gmail.com

Received and published: 19 November 2019

Snow depth estimates are significant for the assessment of the hydrological potential of the snowpack. The application of machine learning tools provides us with a means to derive new depth estimates from a trained model. The methods for the modeling of snow depth using remote sensing data are predominantly based on passive microwave data with much higher repeatability and spatial coverage than InSAR data, rendering such analysis suitable for the monitoring of the snow accumulation. I thus, consider this work to be significant.

Overall, the manuscript is organized and written neatly and represented in a well-

C1

structured manner. The language is mostly appropriate except for a few sentences which are not easily understandable. There are some claims and statements made by the authors that lack references or evidence.

This work is appreciable in the extent of the analysis performed by the authors, in particular for the time series evolution of the snow depth in some of the major provinces in China. However, the manuscript also presents some weaknesses in the methodology, experiments, and particularly the validation. Specific comments are as follows.

Major issues

1. The authors have not clearly stated the novelty of their proposed method. In my opinion, the novelty of the proposed method is in the design of the regression model using the Random Forests i.e. the step -1 in Figure 3 and its application for the modeling of snow depth. The other steps are similar to the methodology proposed in – Jiang, L., Wang, P., Zhang, L. et al. *Sci. China Earth Sci.* (2014) 57: 1278. <https://doi.org/10.1007/s11430-013-4798-8>

2. Why the Random Forest is used, in contrast to better alternatives such as deep neural networks? The authors claim that RF is superior to SVM and ANN, is there any documented evidence regarding RF to be superior to SVM or ANN in link with modeling of geophysical parameters similar to snow depth? Deep learning for classification and regression has been found very useful in recent literature. What is the reason that the authors use RF instead of deep neural networks? Please provide evidence for this or perform additional experiments to prove that RF-based estimates are superior to SVM, ANN, and deep NN based estimates.

3. In both cases, steps 1 and 3, the authors use only a single year data for validation. This neither provides enough points for validation nor any comprehensive inferences from the validation results.

4. The datasets used for training and testing have some issues. The authors have

C2

shown how the actual depth has varied through the years 1987-2019. But for training only data till 2004 was used. The trends from Figures 10 and 11 show a marginal decrease in the mean snow depth. Would it not be better to use data from every two year or alternate year for training the RF. Similarly for testing, the authors use data from the only year 2012-13 for model testing and 2017-18 for testing the final results. This is not sufficient to develop a comprehensive interpretation of the results.

5. In section 3.2, the correlation coefficient is 0.77. Is this satisfactory enough to be used to generate the reference dataset from the RF model? A majority of data are below 10 cm snow depth, then an error of 4.5 cm is significantly high. To have a better understanding of the modeled results, it is vital that we observe the accuracy for the points of higher snow depth also. Particularly, when there is a very high snow depth different for the regions QTP and the others. The validation should be carried out for these regions separately. I suggest the authors show a histogram of the data and also carry out a separate fit for points of snow depth >10cm or perform a case by case fit with respect to the study area. A significant concern is that in the case of shallow snow (<10cm), is the brightness temperature actually representative of the contributions from the shallow snowpack or the underlying ground. This requires further investigations. This is important since the bulk of the data is within the 0-10 cm range. Another concern is that there are very few points with snow depth >40cm. In several locations in the Himalayas, the peak snow depth is usually around 1m or more. Thus, the applicability of the proposed method or the transferability of the proposed method to other areas, in these cases, is in question.

6. The authors observed higher errors for shallow snow depth, but the manuscript lacks any discussion on the contributions from the underlying ground layer to the passive microwave brightness temperature in case of shallow snow depth. The authors have simply added some references. A discussion is required in the manuscript on the sensitivity of snowpack thickness and stratigraphy towards the passive microwave brightness temperature.

C3

7. Page 12, L25-27: Does this mean 3-10 samples in (3x25)x(3x25) sq. km area? This is not clear to me. I think the authors are referring to measurements from field campaigns or weather stations as samples. In this case, the number of samples is very small per the averaging window. Please provide references for this.

8. Figures 9a and 9b. There are very few samples used for validation in these figures. Further, these samples are discontinuous (Figure 9a) and therefore, this should not be used as the basis for ascertaining the performance of the proposed method, since due to the distribution of the points, it is expected that the fit will provide better results. The authors may perform other significance tests such as Neman's test, but the fact remains that the validation data is not really comprehensive. The data shown in Figure 9b is much better for assessment, as it is continuous. But why only 10 points? Earlier it was shown that several ground stations exist in the area. I suggest the authors also use data from other years in their validation scheme, as the results shown at present are not convincing. Why is the modeled snow depth showing very less sensitivity between 20-40cm (nearly constant) and again afterward? This is an issue that requires investigation.

9. In section 4.5, the selection of sample size for training and testing is reversed. Since the MEMLS requires auxiliary information, which is seldom available, the training samples should be much less than the validation samples. This validation strategy is not convincing. From the discrepancy in the training and testing samples, it is already expected that the model accuracy would be high. Minor issues

Finally, some minor corrections in language are also required. Some specific comments are as follows. The authors should check the manuscript for similar errors.

Page 02, L7: " the Himalayas during. . .". The Himalayan ranges are very long and are shared by several countries. Please specify which Himalayan ranges the authors are referring to here. I do not agree with the statement that mean snow depth is maximum in Xinjiang for the entire Himalayan range. Please provide references for this.

C4

Page 02: L8-11: These are documented facts in literature for several other locations, however. Thus, the authors should strictly restrict their inferences to their own findings and not speculate. Thus, here the sentence should be specific to the study area in the manuscript.

Page 02, L11-13: The sentence “In conclusion. . .” is not clear. Please rephrase.

Page 02, L24: “mean snow density”. I believe the authors are here referring to mean stratigraphic snow density”. Please correct this.

Page 03, L17-18: “however, these. . .”. Is there any evidence that the RTM based methods are computationally more expensive than machine learning-based methods. In my opinion, both depend on the selection of the parameters. For example, an RF with substantial input and a high number of trees may be as expensive computationally. If there is no documented evidence on this, please remove this statement.

Page 06, L 28-29: “The lack of ..”. This is only true when observed from the perspective of the spatial resolution of 25x25km. However, the topography of these pixels is not shown to the readers. The authors should show a high-resolution map of these pixels in Figure 1 also.

Page 11, L 11-13: Please correct the range as 200~350 kg/m³ and provide a reference, for example- Meløysund, Vivian, Bernt Leira, Karl V. Høiset, and Kim R. Lisø. 2007. “Predicting snow density using meteorological data.” *Meteorological Applications* 14 (4): 413–23. doi:10.1002/met.40.

Page 17, L20: “The snowpack is set ..”. This should be the snowpack is assumed to comprise a single layer indicating a semi-infinite medium. This is a common assumption in electromagnetic modeling of the snowpack. Please add references to this.

Figure 1: This needs to be revised. Firstly, the authors use 3 areas for their study which have not been shown on the large map. Secondly, the two pixels mentioned previously should be shown at a higher resolution. Third, write in captions what the color bar

C5

represents, is it elevation? Finally, the pixels shown should also have a lat-long grid and scale bar.

Figure 7: Why is the number of points and their locations changing in the maps showing stations. I believe this should remain fixed irrespective of the month. If there is no snow at some of the stations which have been omitted, these should be shown with either a different symbol or a color.

Figure 8: The images are distorted. It appears as if they were stretched manually to fit some size.

Figure 9/Table 4 and several other instances: The R² and R, i.e. the determination coefficient and the correlation coefficient, respectively, are two different parameters and have been used interchangeably with similar symbols in the manuscript, which makes it difficult to judge the accuracy of the results.

Interactive comment on The Cryosphere Discuss., <https://doi.org/10.5194/tc-2019-161>, 2019.

C6