The Cryosphere Discuss.,
https://doi.org/10.5194/tc-2019-161-RC5, 2019

The Cryosphere
Discussions

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.

Tomasz Berezowski (Referee)

tomberez@eti.pg.edu.pl

Received and published: 28 November 2019

Dear Authors, I found that the second point of my review did not paste fully into the online form. Here is the complete point. Sorry for the inconvenience:

The most important issue is that the validation of the RF and the pixel based snow depths is not fair. This is because stations used for validation are only a temporal subsample of the training station set. The spatial sub-sampling was not conducted, i.e., stations for all geographic locations were used for training and validation. This is a very important problem, because latitude and longitude are the third and fourth important predictors in the model, nearly as important as the Tb. The RF model cannot know

values of this predictors already during training, because the validation does not make sense. Therefore, the errors reported in this study are very optimistic (underestimated) and should be recalculated using 50% of the stations (not the data, i.e. spatial not temporal subset) which were not used to train the RF model. The same 50% subset should be used to validate the pixel-based method,

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