

Response to Reviewer Comments by Divyesh Varade 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.

Thank you for your further comments concerning our manuscript. Those comments are very helpful for revising and improving our paper as well as providing important guidance for our future research. We have studied the comments carefully and have made corrections, which we hope meet with approval. The detailed corrections and the point by point responses to your comments are listed below:

Review #1

General Comments:

1. My primary concern was the selection of the RF algorithm for snow depth estimation. With the broader availability of deep neural networks for similar applications, I am not convinced by the utilization of RF for snow depth estimation. Therefore, the authors should either demonstrate that such advanced methods do not provide a significant improvement in accuracy over RF algorithm or re-do the experiments with these methods. Without this analysis or these experiments, we cannot ascertain the potential of the proposed method, particularly in practical applications.

The experiment with ANN by the authors is appreciated. However, this is not at all convincing, since, we already know from the literature that for medium resolution or coarse resolution imagery, SVM or RF outperforms the ANN.

Response 1: Thank you for your comments. The basic assumption for the widely used empirical or semiempirical inversion algorithms is that TB at K-band (~19 GHz) is weakly affected by the snow on the ground but the Ka-band (~37 GHz) is more sensitive to the snowpack (Chang et al., 1987). However, such empirical and semiempirical retrieval approaches are not robust due to the non-linear relationship between snow depth and TB gradients caused by the confounding effects of snow microstructure, the impact of snowpack stratigraphy, the forest cover, and the saturation for snow conditions exceeding a given threshold (Durand and Margulis, 2007). Machine learning (ML) techniques are powerful tools for snow depth estimation due to their ability to handle non-linearity. There are lots of ML methods (e.g., ANN, SVR) applied into snow depth estimation. Actually, ANNs with many layers are examples of deep learning algorithms (Reichstein et al., 2019). We conducted a comparison of SVM, ANN and RF models for snow depth estimation over China (Fig.1). The results showed that RF still outperforms others. Although SVM performs better than ANN, it takes much more time to train the model. Moreover, to our knowledge, the performance of the RF model in spatiotemporal snow depth estimation has not been assessed to date. Thus, the main aim of this study was to assess the feasibility of the RF model in estimating snow depth, rather than determining which ML technique is most accurate.

We agree with your opinion that deep learning is an advanced technique. As we discussed in the revised manuscript, it is in addressing the shortcomings of traditional machine learning methods that the techniques of deep learning promise breakthroughs. However, any machine learning model has certain advantages. For example, RF regression is insensitive to the quality and size of training samples and tends to process quickly and be easy to use (Biau and Scornet, 2016; Probst and Boulesteix, 2018; Maxwell et al., 2018). The deep learning approach generally requires high-dimensional predictor variables and large sample numbers to train the model. In our view, prior snow knowledge is the most important variable in the ML approach, and the absence of snow parameters (e.g., grain size) does not improve its predictive ability. The reviewer provides a promising approach for snow depth estimation, and our next step will focus on coupling physical process

models with the data-driven deep learning technique (Please refer to the response to “General comments 2” below).

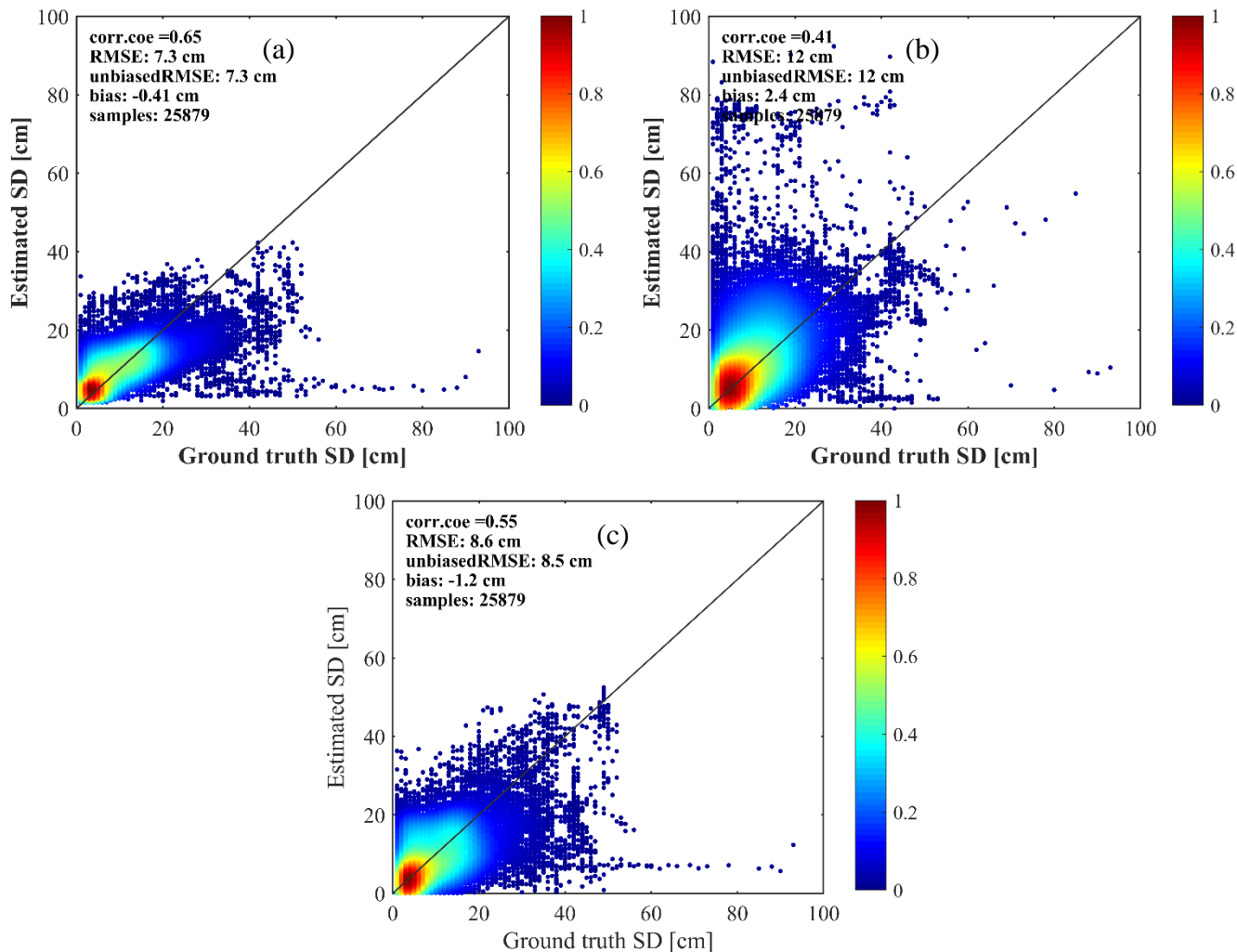


Figure 1. Comparison of (a) RF, (b) ANN, and (c) SVM with respect to snow depth estimation in China.

2. The authors have now discussed the impact of the diurnal changes in the snowpack geophysical properties, including the microstructure, as mentioned in my previous comment. However, this discussion is not comprehensive. Due to the absence of any experiments, this cannot be correlated to the errors observed. Thus, the analysis and the discussion part in this manuscript is still lacking.

For example, the authors have considered in their RF algorithms inputs such as the latitude and the longitude. But why not any snowpack specific parameters. These can be obtained from the ERA5 data and used as inputs to their RF models. The authors then should check the discrepancy in the RMSE based on these parameters.

Another analysis that I find missing is seasonal variation in accuracy, which is very important. The readers should know the difference in the performance of the algorithm in the off-peak snow seasons, i.e., during the early winter and the melt season.

Response 2: Thank you for your comments. We agree with your opinion that integrating a priori knowledge of snowpack into the machine learning models can improve snow depth estimates. We evaluated the ERA5-land hourly snow density product (<https://cds.climate.copernicus.eu/>) with station measurements. Many stations in China measured (observation time 8:00-9:00 a.m.) snow depth and snow pressure every five days during the period 2014-2017; therefore, we collected many in situ snow density observations as the validation dataset. We used this dataset to evaluate ERA5-land hourly snow density data at 8:00 a.m. Fig.2 shows that ERA5 retrievals are poorly correlated with station measurements. Thus, we dismissed this strategy. Actually,

we are attempting to combine physical snow evolution models of snowpack with machine learning techniques. Those physical snow evolution models, such as the Snow Thermal Model (SNTHERM) (Jordan, 1991), SNOWPACK (Lehning et al., 2002a, b), and Crocus (Brun et al., 1989; Vionnet et al., 2012), can be used to simulate snow parameters (e.g., snow grain size, snow density) relatively accurately using time-series meteorological and snowpack measurements. ERA5-land have provided all these forcing data for physical snow evolution models since 1981. It will take some time to conduct this work because such physical models are driven by surface meteorological hourly data (precipitation, air temperature, wind, downward solar, and longwave radiation) and require topography, soil, and vegetation information. Thus, we will show the results in a future publication.

In addition, we discussed this in Section 4.1. “Snow physical snow evolution models, e.g., the Snow Thermal Model (SNTHERM) (Jordan, 1991), SNOWPACK (Lehning et al., 2002a, b), and Crocus (Brun et al., 1989; Vionnet et al., 2012) can be used to simulate snow parameters (e.g., grain size, density) relatively accurately. Thus, integrating a priori knowledge of snowpack into ML techniques has the potential to overcome many limitations that have hindered a more widespread adoption of ML approaches.” (Page 10, Lines 15-18 in the revised manuscript).

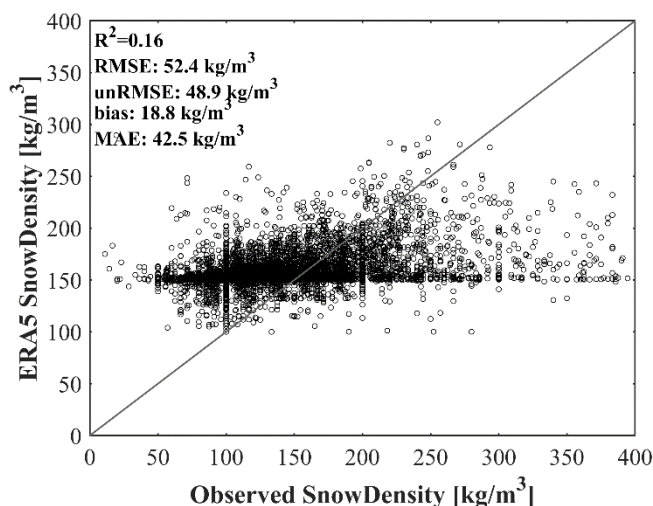


Figure 2. Validation of the ERA5-land snow density product with in situ observations in China.

According to the reviewer’s suggestion, monthly performances of both RF and WESTDC products were analyzed and compared. Fig.3 shows that unRMSEs of both products present increasing trends from the beginning to end of the snow season. This is due to seasonal evolution of snowpack, which leads to uncertainty in snow depth estimates. WESTDC estimates tend to be underestimated in November, December and March, while the RF product is superior to the WESTDC data. The correlation coefficient in January is the highest among snow season months, which is attributable to stable snow cover.

We added the following text in the revised manuscript: “Snow depth estimates with PMW data are usually challenged by the snow metamorphism (e.g., snow grain size). In particular, the large diurnal temperature range in the late snow season leads to rapid snow grain growth (Dai et al., 2012). Fig. 9 presents the monthly performances of both RF and WESTDC products. The RF estimates outperform the WESTDC product in terms of correlation, overall bias and unbiased RMSE. WESTDC estimates tend to be underestimated in November, December and March, while the RF product is superior to the WESTDC data. Due to the influence of the seasonal evolution of snowpack, the unbiased RMSEs of both products present increasing trends from November to March during snow seasons. The correlation coefficient in January is the highest among snow season months, which is attributed to stable snow cover.” (Page 8, Lines 7-13 in the revised manuscript).

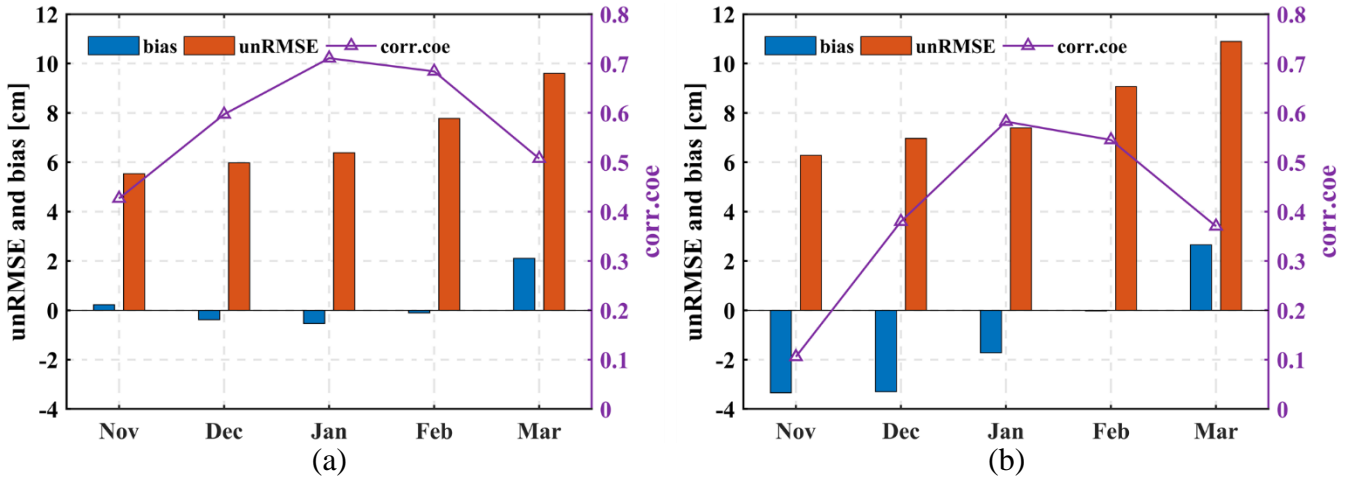


Figure 3. Monthly performances of (a) RF and (b) WESTDC snow depth products.

3. Please mention the upscaling method for 1-km to 25-km LULC map. Additionally, I also mention which latitude and longitudes were used as inputs. I hope these were geographic ones and not projected ones.

Response 3: The 1-km land use/land cover (LULC) data were provided by the Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (<http://www.resdc.cn/>). In a 25-km pixel, there are a total of 625 1-km grids. For example, the number of grassland 1-km grids within a 25-km pixel is n , and the fraction is calculated as:

$$\text{fraction} = \frac{n}{625} * 100\%$$

The latitude and longitude used in this study were geographic coordinates, retrieved with an official procedure (<https://nsidc.org/data/ease/tools>) based on the row and column of EASE-GRID data.

4. I find that the majority of the discussion is based on the comparison between the results from the RF method and the WESTDC. While I believe there should be a better balance in the discussion for comparison with the in-situ measurements.

Response 4: Both the RF and WESTDC products were evaluated using in situ measurements. In Section 3, we presented all the validation results. In Section 4, we discussed the disadvantage (poor performance at the spatial scale) of the RF model, the unnecessary predictor variables (if highly correlated) for the RF model, and the potential errors (overestimation and underestimation) of the RF product. All these discussions were based on the assessment results in Section 3.

5. Another analysis to the manuscript that the authors should add is for the selection of trees. Presently, the authors have fixed this to 1000. However, is there any merit in using this large number? What is the performance for other smaller number of trees, for example, 500, 250, etc. Is the performance significant enough to use such a relatively large number of trees?

Response 5: The default value for n_{tree} is 500 in the RandomForest package. One of the merits of the RF model is rapid training. Moreover, it can avoid overfitting due to the large number of decision trees. Thus, we set n_{tree} to 1000.

6. An important parameter in the discussion is the R-square that is missing?

Response 6: Thank you for your comments. In general, the correlation of in situ measurements and PMW retrievals is not significant (low R-square) due to the non-linear relationship between snow depth and satellite TB gradients. Thus, we dismissed this index in the discussion.

7. Table 5, the results are contradictory to the explanations given by the authors. The authors have mentioned that for shallow snow depth, the PMW could be insensitive. However, Table 5 shows the contradictory, i.e., rather some sensitivity in case of shallow snow depths as compared to deeper snow where there is nearly no correlation. I believe the inconsistency of the results requires a thorough analysis.

Response 7: PMW signals are insensitive to shallow fresh snow (< 5 cm) due to low snow density and snow grain size (Chang et al., 1987; Foster et al., 2005). However, they tend to overestimate snow depth for shallow old snow in the late snow season due to the seasonal evolution of snowpack. This is one of the reasons for the overestimation of shallow snow cover (Table 5). Please refer to the response to “Minor issues 3” below.

Due to snow metamorphism and the ‘saturation effect’, the general tendency for the correlation between snow depth and brightness temperature gradient (TBD_{19V&37V}) is poor for deep (> 20 cm) snow conditions (Fig. 4).

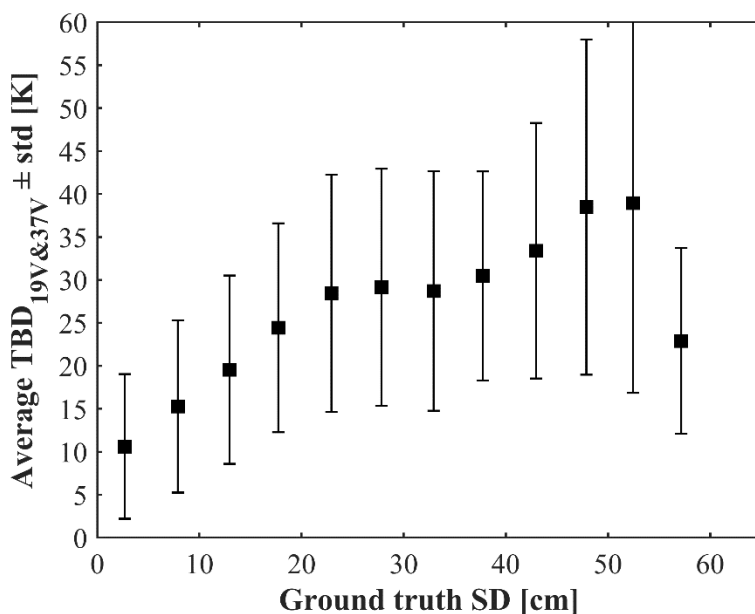


Figure 4. Correlation between brightness temperature gradient and in situ snow depth measurements.

Minor issues:

1. Please check Figure 4 for correctness. Is the validation dataset correct? This is very confusing from the authors response.

Response 1: We apologize for the confusion. Training datasets for RF models were selected randomly from station records during the period 1987-2004. We applied the same validation dataset (2005-2006) to evaluate the well-trained RF models (Fig. 5).

We rewrote this part as follows: “To ensure a sufficient number of samples, all station records (approximately 100,000 samples) from 1987 to 2006 were used to analyze the sensitivity of the RF model to the training sample size. A total of 5000 to 80,000 (with a step of 5000) samples were selected randomly from data during the period 1987-2004 and used to train the RF models, and a two-year stand-alone dataset from 2005 to 2006 was applied to assess the performance of the well-trained model” (Page 6, Lines 1-4 in the revised manuscript).

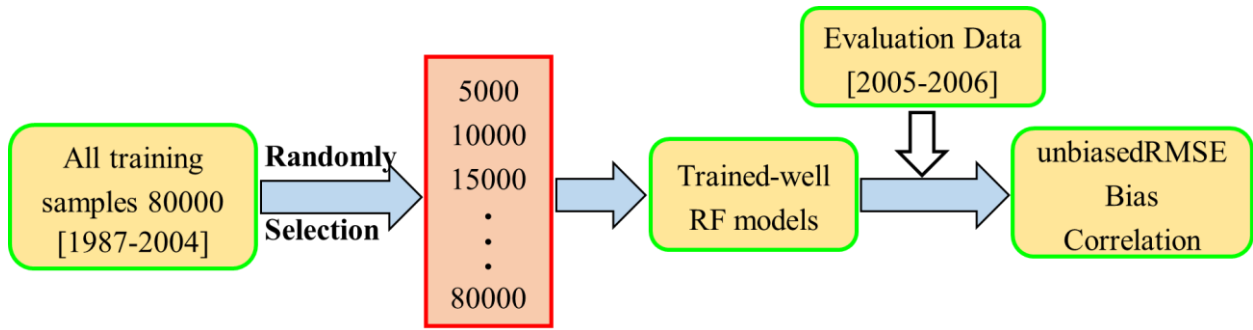


Figure 5. The test process flowchart for the sensitivity of the RF model to the training sample size.

2. Some of the colors used by authors are very poor. Figure 6 the turquoise text, Figure 9, the orange legends and lines.

Response 2: Corrected.

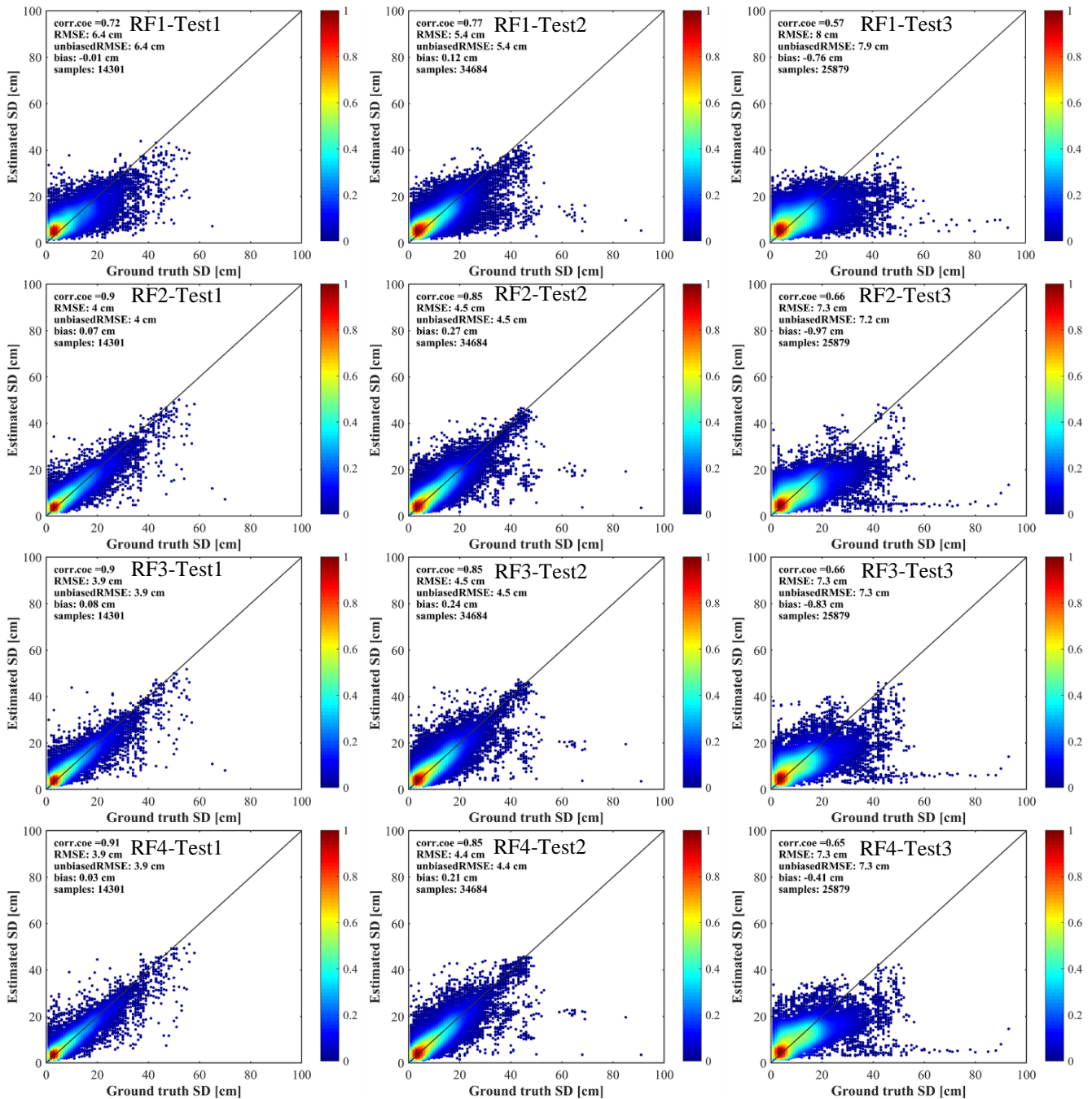


Figure 6. The color-density scatterplots of the estimated snow depth with four fitted RF algorithms and the ground truth snow depth. The four trained RF algorithms (RF1, RF2, RF3, RF4) were evaluated with three validation datasets (Test1, Test2, Test3).

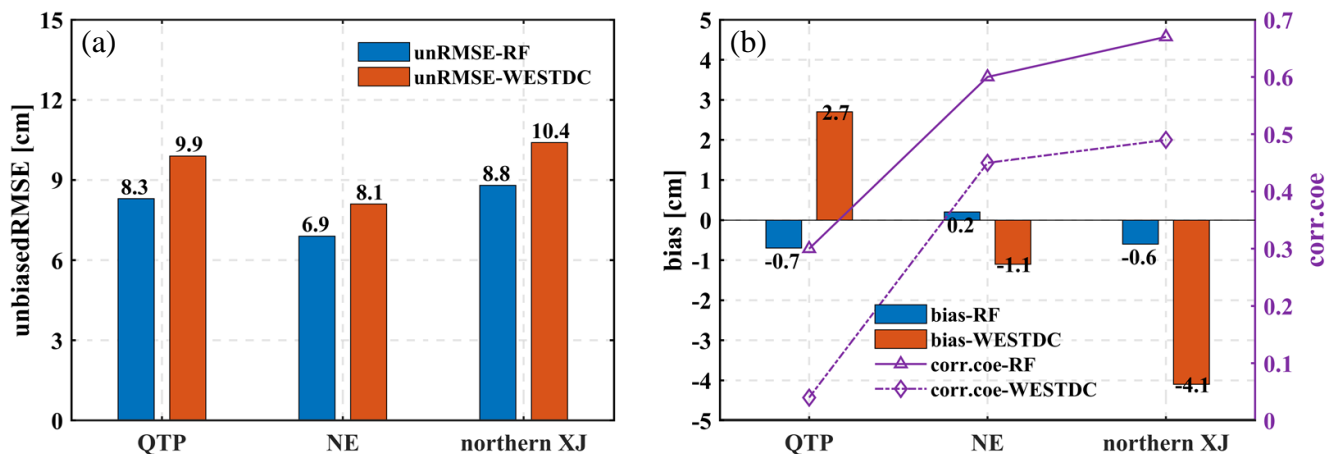


Figure 7. The validation of the RF and WESTDC snow depth products in three stable snow cover areas over China with respect to (a) the unbiased RMSE and (b) bias and the correlation coefficient.

3. Although I meant with my comment to include a discussion, the MEMLS simulation experiment is appreciated. However, the parameters used in the MEMLS simulation are not correct in the response by the authors. How often do we have a snowpack representing fresh snow characteristics? This is very seldom. Thus, there is no point showing an experiment with a snow density of 100 kg/cubic m corresponding to fresh snow. Instead, the authors should have selected something between 200-350 kg/cubic m for the simulation, which corresponds to old snow.

Response 3: Thank you for your constructive comments. The RF product presents an overestimation for shallow snowpack. One of the reasons is the strong scattering effect caused by snow metamorphism, especially for old shallow snow. Table 1 shows that the accuracy in the late snow season is poor. This is due to the seasonal snow evolution.

We rewrote this part as follows: “Table 5 indicates that there is overestimation in NE and northern XJ for shallow snow cover, which may be due to the following reasons. First, the PMW signals are insensitive to thin snow cover (< 5cm), especially for fresh snow with low snow density and snow grain size, which generally results in underestimation (Foster et al., 2005). In contrast, it tends to overestimate snow depth for shallow old snow in the late snow season due to the seasonal evolution of snowpack. For example, the large diurnal temperature range in the late snow season tends to subject the snowpack to frequent freeze-thaw cycles and leads to rapid snow grain (~2 mm) and snow density (200-350 kg/m³) growth and consequently a high T_B difference (Meløysund et al., 2007; Durand et al., 2008; Yang et al., 2015; Dai et al., 2017). Thus, the overall bias and unbiased RMSE for shallow snowpacks (< 10 cm) present increasing trends from November to March in NE and northern XJ” (Page 11, Lines 14-21 in the revised manuscript).

Table 1. Summary of monthly performances of the RF product in NE and northern XJ.

Month	NE				
	November	December	January	February	March
corr.coe	0.32	0.41	0.40	0.23	0.08
bias (cm)	2.33	2.19	2.93	4.74	7.97
unRMSE (cm)	3.66	3.69	4.16	5.24	6.16

northern XJ					
Month	November	December	January	February	March
corr.coe	0.20	0.27	0.40	0.20	0.08
bias (cm)	3.68	3.35	2.97	5.65	10.60
unRMSE (cm)	4.49	4.77	4.61	6.83	7.09

The microwave emission model of layered snowpack (MEMLS) was applied to simulate the T_B with varying snow parameters. Generally, the snow density ($< 100 \text{ kg/m}^3$) and snow grain size (correlation length $< 0.2 \text{ mm}$) are small for shallow, new snow cover ($< 5 \text{ cm}$), while these values are $200\text{-}350 \text{ kg/m}^3$ and $1\text{-}2 \text{ mm}$ for snow density and grain size, respectively. Fig. 8 shows the sensitivity of snow depth to T_B at 36 GHz for fresh and old snowpacks. The passive microwave signals are insensitive to the new snow cover. In contrast, there is notable attenuation to microwave radiation at the 36.5-GHz channel for old snow; therefore, snow depth is overestimated.

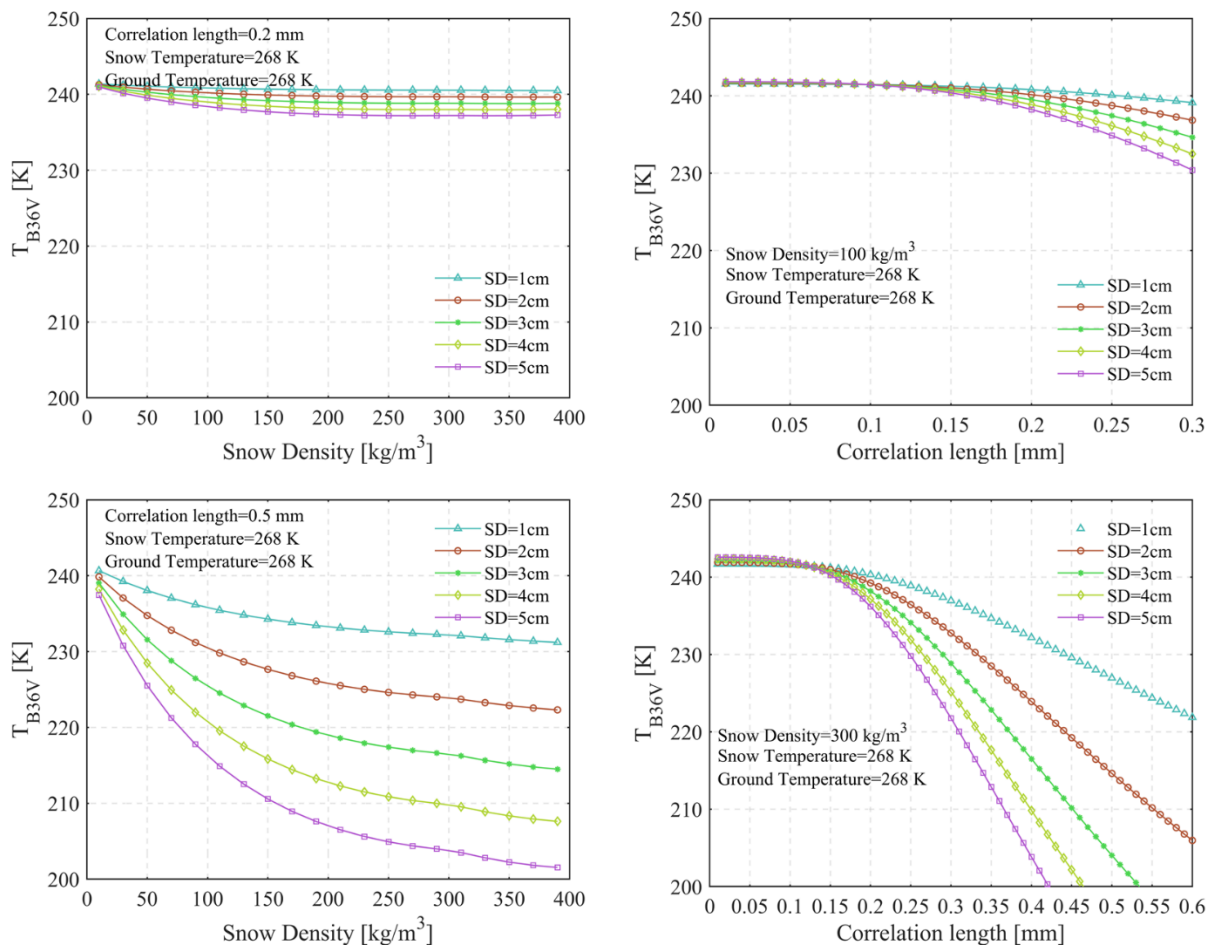


Figure 8. The sensitivity of snowpack stratigraphy to the passive microwave brightness temperature simulated with the MEMLS model. Snow grain size is generally 4-5 times greater than the correlation length in the MEMLS model (Mätzler, 2002). Top: fresh snow; Bottom: old snow.

Response to Reviewer Comments by Review #2 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.

Many thanks for the very positive comments on the revised version of the manuscript. We have studied the comments carefully and have made corrections, and responses are provided in blue font below.

Referee #2

The manuscript has been improved considerably. The authors had to address multiple conflicting reviewers' comments and in my opinion they did so fine. I have some minor comments remaining to be addressed:

General evaluation

1. Perhaps the manuscript emphasizes the use of random forests, however given the limited comparison with other algorithms in both the original and revised manuscript, the manuscript should emphasize the construction of the new dataset.

Based on this, the dataset should be uploaded online and made available to the readers of the Journal.

Response 1: We appreciate the reviewer's suggestions. Accurate retrieval of snow depth from passive microwave (PM) measurements has been challenging due to the confounding effects of snow microstructure, the impact of snowpack stratigraphy, the forest cover, the liquid water in the snowpack, the coarse spatial resolution, and the saturation effect (Durand and Margulis, 2007). Thus, the relationship between snow depth and satellite observation is clearly non-linear for all frequencies. Machine learning (ML) techniques are powerful tools for establishing the non-linear relations between independent variables and the target variable. Random forest approach is one of the most successful ML algorithms due to its ability to handle non-linearity (Rodriguez-Galiano et al., 2012; Belgiu et al., 2016; Maxwell et al., 2018; Bair et al., 2018). In addition, the assessment of RF algorithm performance in snow depth estimation has not been made to date for China.

We would like to share this dataset online. Actually, we have already prepared the reconstructed dataset from 1987 to 2018. We also wrote a descriptive document to introduce this dataset. Snow depth data are publicly available at the following link: <https://figshare.com/s/60c2e9db228130e6768b>.

2. Furthermore, the discussion in Section 4.1 should be modified to incorporate elements of stationarity of the earth's system that allow the use of random forests. Stationarity should not be confused with the concept of a constant system and is not directly related to the concept of dynamic system (which is used in Section 4.1).

Response 2: We agree with your opinion. We have changed the sentence to “however, these studies aim to obtain spatial predictions of elements of stationarity in the Earth system, e.g., soil types and soil properties” in the revised manuscript.

3. Lastly, while the revised manuscript includes an extended presentation of the properties of the implemented algorithms, it has not been made explicit that the properties of the constructed dataset largely depend on the implemented algorithm (albeit this can be deduced by the manuscript), while reporting RMSE (and other metrics) values does not fully address issues related to the properties of the dataset.

Response 3: Thank you for your comments. The widely used inversion algorithms for snow depth estimation are based on empirical or semiempirical relationships between snow depth and brightness temperature (TB) gradients. The basic assumption for these algorithms is that TB at K-band (~19 GHz) is weakly affected by the snow on the ground but the Ka-band (~37 GHz) is more sensitive to the snowpack. However, accurate retrieval of snow depth from PM data has been challenging due to many factors, e.g., the confounding effects of snow microstructure, the impact of snowpack stratigraphy, the forest cover, and the saturation for snow conditions exceeding a given threshold. Thus, such empirical and semiempirical retrieval approaches are not robust due to the non-linear relationship between snow depth and TB gradients. Random forest approach is one of the most powerful ML tools to handle non-linearity relationship between snow depth and satellite observation, which makes retrievals improved.

Meanwhile, we have explicitly described the properties of the implemented RF model in the introduction.

Procedure code: a version 4.6-14 randomForest R package;

Model parameters: two user-defined parameters: $mtry = 1000$; $mtry = 2$;

Predictor variables: satellite vertical polarization brightness temperature observations at 19 GHz and 37 GHz (TB_{19V} , TB_{37V}), Latitude, Longitude;

Target Variable: snow depth;

Training samples: a total of 28602 during the period 2012-2014;

Validation dataset: spatially independent reference data during the period 1987-2018.

Dear Editor, dear reviewers,

I would like to thank the two reviewers for their positive and constructive comments on our paper. Below, we list the changes in the revised manuscript.

List of changes (page and line numbers refer to the revised manuscript):

- **p.6, lines 1-4:** Rewrote the description of the flowchart as: “All station records (approximately 100,000 samples) from 1987 to 2006 were used to analyze the sensitivity of the RF model to the training sample size. A total of 5,000 to 80,000 (with a step size of 5,000) samples were selected randomly from the data during the period 1987-2004 and were used respectively to train the RF models, and a two-year stand-alone dataset from 2005 to 2006 was applied to assess the performances of the well-trained models.”
- **p.8, lines 7-14:** Added the analysis of monthly performances of the RF product constructed in this study: “Snow depth estimates with PMW data are usually challenged by the snow metamorphism (e.g., snow grain size). In particular, the large diurnal temperature range in the late snow season leads to a rapid snow grain growth (Dai et al., 2012). Fig. 9 presents the monthly performances of both RF and WESTDC products. The RF estimates outperform the WESTDC product in terms of correlation, overall bias and unbiased RMSE. WESTDC estimates tend to be underestimated in November, December and March, while the RF product is superior to the WESTDC data. Due to the influence of the seasonal evolution of snowpack, the unbiased RMSEs of both products present increasing trends from November to March during the snow seasons. The correlation coefficient in January is the highest among snow season months, which is attributed to stable snow cover.”
- **p.10, lines 6-8:** Replaced “static” with “stationarity” and “not static but dynamic” with “non-stationary”.
- **p.10, lines 15-19:** Added: “Physical snow evolution models, e.g., the Snow Thermal Model (SNTHERM) (Jordan, 1991), SNOWPACK (Lehning et al., 2002a, b), and Crocus (Brun et al., 1989; Vionnet et al., 2012), can be used to simulate snow parameters (e.g., grain size, density) relatively accurately. Thus, integrating a priori knowledge of snowpack into ML techniques has the potential to overcome many limitations that have hindered a more widespread adoption of ML approaches.”
- **p.11, lines 15-21:** Rewrote: “First, the PMW signals are insensitive to thin snow cover (< 5cm), especially for fresh snow with low snow density and snow grain size, which generally results in underestimation (Foster et al., 2005). In contrast, it tends to overestimate snow depth for shallow old snow in the late snow season due to the seasonal evolution of snowpack. For example, the large diurnal temperature range in the late snow season tends to subject the snowpack to frequent freeze-thaw cycles and leads to rapid snow grain (~2 mm) and snow density (200-350 kg/m³) growth and consequently a high T_B difference (Meløysund et al., 2007; Durand et al., 2008; Yang et al., 2015; Dai et al., 2017). Thus, the overall bias and unbiased RMSE for shallow snowpacks (< 10 cm) present increasing trends from November to March in NE and northern XJ (Table 6).”
- **p.12, line 24:** Added: “Combining the physical snow evolution model (e.g., SNOWPACK) with the ML method will be the focus of future work.”

- **p.13, line 33:** Added: “Brun, E., Martin, E., Simon, V., Gendre, C., and Coleou, C.: An Energy and Mass Model of Snow Cover Suitable for Operational Avalanche Forecasting, *Journal of Glaciology*, 35, 333–342, 10.1017/S0022143000009254, 1989.”
- **p.15, line 18:** Added: “Jordan, R.E. 1991.: A One-Dimensional Temperature Model for a Snow Cover: Technical Documentation for SNTHERM.89; U.S. Army Cold Regions Research and Engineering Laboratory: Hanover, NH, USA.”
- **p.15, lines 30-34:** Added: “Lehning, M., Bartelt, P., Brown, B., Fierz, C., Satyawali, P.: A physical SNOWPACK model for the Swiss avalanche warning part II. Snow microstructure, *Cold Reg. Sci. Technol*, 35(3), 147–167, 10.1016/S0165-232X(02)00073-3, 2002a. Lehning, M., Bartelt, P., Brown, B., Fierz, C.: A physical SNOWPACK model for the Swiss avalanche warning: Part III: meteorological forcing, thin layer formation and evaluation, *Cold Reg. Sci. Technol*, 35(3):169–184, 10.1016/S0165-232X(02)00072-1, 2002b.”
- **p.17, line 37:** Added: “Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Le Moigne, P., Martin, E., Willemet, J.-M.: The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2, *Geosci. Model Dev*, 5, 773–791, 2012.”
- **p.19, line 5:** Added: “Table 6”
- **p.22, line 6:** Redrawn: “Fig. 6”
- **p.24, line 2:** Redrawn: “Fig. 9”
- **p.24, line 6:** Added: “Fig. 10”

Snow Depth Estimation and Historical Data Reconstruction Over China Based on a Random Forest Machine Learning Approach

Jianwei Yang¹, Lingmei Jiang¹, Kari Luo², Jinmei Pan³, Juha Lemmetyinen², Matias Takala², Shengli Wu⁴

¹State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and the Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences, Beijing Engineering Research Center for Global Land Remote Sensing Products, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

²Finnish Meteorological Institute, Helsinki Fi00101, Finland

³State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China

⁴National Satellite Meteorological Center, China Meteorological Administration, Beijing 100081, China

Corresponding Author: Lingmei Jiang (jiang@bnu.edu.cn)

Abstract. We studied whether the random forest (RF) machine learning (ML) model could be used to retrieve snow depth. Four combinations composed of critical predictor variables were used to train the RF model. Then, we utilized three validation datasets from out-of-bag (OOB) samples, a temporal subset and a spatiotemporal subset to verify the fitted RF algorithms. The results indicated the following: (1) the accuracy of RF model is greatly influenced by geographic location, elevation, and land cover fractions; (2) however, the redundant predictor variables (if highly correlated) slightly affect the RF model; (3) the fitted RF algorithms perform better in temporal scale than in spatial scale, with unbiased RMSEs of ~4.4 cm and ~7.3 cm, respectively. Finally, we used the fitted RF2 algorithm to retrieve a consistent 32-year daily snow depth dataset from 1987 to 2018. This product was evaluated against the independent station observations during the period 1987-2018. The mean unbiased RMSE and bias were 7.1 cm and -0.05 cm, respectively, outperforming the former snow depth dataset (8.4 cm and -1.20 cm) from the Environmental and Ecological Science Data Center for West China (WESTDC). Although the RF product was superior to the WESTDC dataset, it still underestimated deep snow cover (> 20 cm), with biases of -10.4 cm, -8.9 cm and -34.1 cm in Northeast China (NE), northern Xinjiang (XJ) and the Qinghai-Tibetan Plateau (QTP), respectively. Additionally, the long-term snow depth datasets (station observations, RF estimates and WESTDC product) were analyzed in terms of temporal and spatial variations over China. On a temporal scale, the ground truth snow depth presented a significant increasing trend from 1987 to 2018, especially in NE. However, the RF and WESTDC products displayed no significant changing trends except in QTP. The WESTDC product presented a significant decreasing trend in QTP, with a correlation coefficient of -0.55, whereas there were no significant trends for ground truth observations and the RF product. For the spatial characteristics, similar trend patterns were observed for RF and WESTDC products over China. These characteristics presented significant decreasing trends in most areas and a significant increasing trend in central NE.

1 **1 Introduction**

2 Seasonal snow covers a considerable portion of the land surface in the Northern Hemisphere during winter and has a
3 significant effect on the Earth's radiation balance and surface-atmosphere interaction due to its high albedo and low thermal
4 conductivity (Fernandes et al., 2009; Derksen et al., 2012; Kevin et al., 2017; Dorji et al., 2018; Bormann et al., 2018). Snow
5 depth is a crucial parameter for climate studies, hydrological applications and weather forecasts (Foster et al., 2011; Takala et
6 al., 2017; Tedesco et al., 2016; Safavi et al., 2017). For these applications, long time series are needed to conduct meaningful
7 statistics on trends and variability. Fortunately, passive microwave (PMW) signals can penetrate snow cover and provide snow
8 depth estimates through volume scattering of snow particles in dry snow conditions. PMW remote sensing also has the
9 advantage of sensing without depending on solar illumination and weather conditions (Chang et al., 1987; Foster et al., 2011).
10 In addition, there exists a long historical record of spaceborne PMW data dating back to 1978, allowing us to study seasonal
11 snow climatological changes (Takala et al., 2011; Santi et al., 2012). These advantages make snow depth estimation from
12 satellite PMW remote sensing an attractive option.

13 Diverse methods have been proposed to retrieve snow depth from PMW observations. The most widely used inversion
14 algorithms were based on empirical relationships between satellite brightness temperature (T_B) gradient and snow depth
15 (Chang et al., 1987; Foster et al., 1997; Derksen et al., 2005; Che et al., 2008; Kelly et al., 2003; Kelly et al., 2009; Jiang et al.,
16 2014). However, these algorithms are not always reliable in all regions due to the fixed empirical constants (Derksen et al.,
17 2010; Davenport et al., 2012; Che et al., 2016; Yang et al., 2019). Subsequently, more advanced algorithms that use
18 theoretical or semi-empirical radiative transfer models were developed (Jiang et al., 2007; Takala et al., 2011; Picard et al.,
19 2012; Lemmetyinen et al., 2015; Metsämäki et al., 2015; Tedesco et al., 2016; Pan et al., 2017; Saberi et al., 2017); however,
20 these complicated algorithms are computationally expensive and require complex ancillary data to provide accurate
21 predictions. These factors restrict the applications of these algorithms on a global scale. Improving the performance of PMW
22 retrieval algorithms through data assimilation has also been investigated (Durand et al., 2006; Tedesco et al., 2010; Che et al.,
23 2014; Huang et al., 2017). The widely used and operational assimilation system combines synoptic weather station data with
24 satellite PMW radiometer measurements through the snow forward model (Helsinki University of Technology snow emission
25 model, HUT), and it provides long-term snow water equivalent data from 1979 to the present in the Northern Hemisphere (>
26 35° N) (Pulliainen et al., 1999; Pulliainen., 2006; Takala et al., 2011). However, the coverage of this product does not include
27 the Qinghai-Tibetan Plateau (QTP), which is one of three stable snow cover areas in China.

28 Machine learning (ML) has attained outstanding results in the regression estimation of land surface parameters from
29 remotely sensed observations at local and global scales over the past decade (Reichstein et al., 2019). The random forest (RF)
30 is an ensemble method whereby multiple trees are grown from random subsets of predictors, producing a weighted ensemble

1 of trees (Breiman, 2001). RF is also robust against overfitting in the presence of large datasets and increases predictive
2 accuracies over single decision trees (Biau and Scornet, 2016; Tyralis et al., 2019b). Over the last two decades, RF has been
3 one of the most successful ML algorithms for practical applications due to its proven accuracy, stability, speed of processing
4 and ease of use (Rodriguez-Galiano et al., 2012; Belgiu et al., 2016; Maxwell et al., 2018; Bair et al., 2018; Qu et al., 2019;
5 Reichstein et al., 2019; Tyralis et al., 2019a). Although the RF model can present good results in many research areas, studies
6 on the spatio-temporal prediction of snow depth are few and the potential utility of RF in such studies is unknown.

7 The primary objectives of this study are to assess the feasibility of the RF model in estimating snow depth, to determine
8 whether the inclusion of auxiliary information (geolocation, elevation and land cover fraction) contributes to the improvement
9 of RF, and eventually to develop a time series (1987 to 2018) of snow depth data in China and analyze the trends in annual
10 mean snow depth. To complete the feasibility study of the RF model, we designed four RF algorithms trained with different
11 combinations of predictor variables and validated them using temporally and spatially independent reference data. To our
12 knowledge, this type of assessment of RF algorithm performance has not been made to date for China. The data and
13 methodology are described in Section 2. Section 3 presents the results regarding the feasibility study of the RF model, the
14 validation of the snow depth product reconstructed with the RF algorithm and the trend analysis of snow depth. The results are
15 discussed in Section 4, and conclusions are given in Section 5.

16 **2 Data and Methodology**

17 **2.1 Data**

18 (1) Satellite passive microwave measurements

19 The series of the Special Sensor Microwave/Imager (SSM/I) and Special Sensor Microwave Imager Sounder (SSMIS)
20 instruments has provided continuous T_B measurements at 19.35, 23.235, 37, 85.5 and 91.655 GHz since July 1987. The data
21 are available from the National Snow and Ice Center (<https://daacdata.apps.nsidc.org/pub/DATASETS>). The SSM/I and
22 SSMIS sensors are suitable for producing a long-term consistent snow depth dataset due to their similar configurations and
23 intersensor calibrations (Armstrong et al., 1994). To avoid the influence of wet snow, only ascending (F08) and descending
24 (F11, F13 and F17) overpass data were used (Table 1). In this study, the difference between 19.35 (36.5) GHz and 18.7 (37)
25 GHz was ignored (hereafter referred as 19 GHz and 37 GHz, respectively).

26 (2) In situ measurements

27 The weather station daily data in China from 1987 to 2018 were provided by the National Meteorological Information Centre,
28 China Meteorology Administration (CMA, <http://data.cma.cn/en>). The geographical locations of the meteorological stations
29 and the three stable snow cover areas are shown in Fig. 1. The recorded variables include the site name, observation time,

1 geolocation (latitude and longitude), altitude (m), near-surface soil temperature (measured at a 5-cm depth, °C), and snow
2 depth (cm). The sites are not distributed homogeneously, and few are located in inaccessible regions with extreme climates
3 and complex terrain conditions, e.g., the western part of QTP (Fig. 1).

4 Quality control was conducted prior to using the data for developing and validating the retrieval algorithm. The first step
5 was to select the records where the near-surface soil temperature was lower than 0 °C. The second step was to remove the sites
6 if the areal fraction of the open water exceeded 30% within a satellite pixel. Finally, the 683 stations were randomly divided
7 into two roughly equal-sized parts (Fig. 1). The snow depth observations from training stations (342 sites) together with
8 satellite T_B and other auxiliary data can be used to train the RF model. The measurements from validation stations (341 sites),
9 as independent data spatially, can be applied to validate the fitted RF algorithm. Fig. 2 shows the histograms of snow depth
10 observations from training and validation stations during the period 2012-2018. Ninety percent of the samples range from 1
11 cm to 25 cm. The maximum values of the snow depth extend to approximately 50 cm. However, the number of such cases is
12 small and is therefore not evident in Fig. 2.

13 (3) Land cover fraction

14 A 1-km land use/land cover (LULC) map derived from the 30-m Thematic Mapper (TM) imagery classification was provided
15 by the Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (<http://www.resdc.cn/>). The
16 map was recalculated as the areal percentages of each land cover type in the 25-km grid cells. In this study, the fractions of
17 grassland, bareland, cropland, forest, and shrubland were calculated as predictor variables of the RF model. To avoid the
18 influence of water bodies and construction, the record was used only if the total fraction was greater than 60%.

19 2.2 Methodology

20 2.2.1 Random forest

21 RF is an ensemble ML algorithm proposed by Breiman in 2001. It combines several randomized decision trees and aggregates
22 their predictions by averaging in regression (Biau and Scornet, 2016). Generally, approximately two-thirds of the samples
23 (in-bag samples) are used to train the trees and the remaining one-third (out-of-bag samples, OOB) are used to estimate how
24 well the fitted RF algorithm performs. Few user-defined parameters are generally required to optimize the algorithm, such as
25 the number of trees in the ensemble (n_{tree}) and the number of random variables at each node (m_{try}). The n_{tree} is set equal to
26 1000 in the present study since the gain in the predictive performance of the algorithm would be small with the addition of
27 more trees (Probst and Boulesteix, 2018). The default value of m_{try} is determined by the number of input prediction variables,
28 usually 1/3 for regression tasks (Biau and Scornet, 2016). The RF regression is insensitive to the quality of training samples
29 and to overfitting due to the large number of decision trees produced by randomly selecting a subset of training samples and a
30 subset of variables for splitting at each tree node (Maxwell et al., 2018). In addition, RF provides an assessment of the relative

1 importance of predictor variables, which have proven to be useful for evaluating the relative contribution of input variables
2 (Tyralis et al., 2019b). Furthermore, the RF model can rapidly trained and is easy to use. In this paper, a randomForest R
3 package (Version 4.6-14) is used for regression (Liaw and Wiener 2002; Breiman et al. 2018).

4 **2.2.2 Feasibility study of the RF model**

5 (1) Selection of predictor variables

6 The possible predictor variables used include geographic location (longitude, latitude), elevation, land cover fractions
7 (grassland, cropland, bareland, shrubland and forest) and multi-channel brightness temperatures. All available channels on the
8 SSM/I and SSMIS are listed in Table 1. The 23 GHz channel is sensitive to water vapor and not surface scattering, which
9 introduces uncertainty to the estimation process (Ji et al., 2017). The 85 (91) GHz channel is seriously influenced by the
10 atmosphere (Kelly et al., 2009; Xue et al., 2017). Typically, the lower frequency (19 GHz) is used to provide a background T_B
11 against which the higher frequency (37 GHz) scattering-sensitive channels are used to retrieve snow depth. The mixed-pixel
12 problem is the dominant limitation on snow depth estimation accuracy (Derksen et al., 2005; Jiang et al., 2014; Roy et al.,
13 2014; Cai et al., 2017; Li et al., 2017). The satellite pixel usually covers several land cover types due to a coarse footprint.
14 Thus, the land cover fractions were included as possible predictor variables. Previous studies have shown that geographic
15 location and elevation indeed contribute to improving ML model performance (Bair et al., 2018; Qu et al., 2019).

16 To determine a suitable selection rule for training samples, we selected four combinations of predictor variables from
17 training stations (Fig. 1) during the period 2012-2014 to train the RF algorithms. Table 2 presents a detailed description of the
18 four selection rules of training samples. The correlations between the predictor variables and the variable importance metrics
19 are shown in Fig. 3. The T_B measurements at horizontal polarization (H-pol) are highly correlated (correlations higher than 0.9)
20 with observations at vertical polarization (V-pol). Moreover, according to their ranking of the predictor variables, the channels
21 of V-pol are more relevant to the independent variable (snow depth) than are the H-pol channels. Therefore, the RF1 algorithm
22 was trained with only two channels' T_B measurements at V-pol. The ranking of variables' importance in Fig. 3 indicates that
23 the geographic location is more important than elevation to snow depth. Thus, the geographic location and elevation were
24 included in the predictor variables of RF2 and RF3, respectively. Fig. 3 also shows that the correlations between T_B and land
25 cover fraction are relatively low. Thus, we will validate whether the inclusion of land cover fraction would increase the
26 performance of the fitted RF4 algorithm.

27 (2) Training sample size

28 One of the advantages of the RF model is that it can effectively handle small sample sizes (Biau and Scornet et al., 2016). A
29 test was conducted to demonstrate the insensitivity of the RF model to the training sample size. The input predictor variables
30 include geographic location and T_B (Table 2, RF2). The flowchart of the test process is shown in Fig. 4. To ensure a sufficient

1 number of samples, all station records (approximately 100,000 samples) from 1987 to 2006 were used to analyze the
2 sensitivity of the RF model to the training sample size. A total of 5,000 to 80,000 (with a step of 5,000) samples selected
3 randomly from data during the period 1987-2004 were used to respectively train the RF models, and a two-year stand-alone
4 dataset from 2005 to 2006 was applied to assess the performance of the trained models. We consider three evaluating
5 indicators (the unbiased root mean square error (RMSE), bias and correlation coefficient) to illustrate the sensitivity of the RF
6 model to the training sample size.

7 (3) Validation datasets of the fitted RF algorithms

8 We conducted three tests to verify the fitted RF algorithms (Table 3). The same training samples (same algorithms) were used
9 for the three tests but with different validation datasets. In Test1, the validation data were from OOB samples. This
10 preliminary assessment generally offers a simple way to adjust the parameters of the RF model. However, the OOB errors
11 should be used with caution because its samples are not independent at temporal and spatial scales. In Test2, we applied
12 independent reference data during the period 2015-2018 to assess the accuracy of the temporal prediction of fitted algorithms.
13 Although this dataset is composed of observations from training stations in Fig. 1, it is temporally independent of the training
14 samples (2012-2014). Generally, the RF model cannot extrapolate outside the training range (Hengl et al., 2018). Thus, in
15 Test3, a spatially independent dataset from validation stations during the period 2015-2018 was used to assess the accuracy of
16 spatio-temporal prediction. The unbiased RMSE, bias and correlation coefficient are used for the assessment of the predictive
17 performance of the fitted algorithms.

18 2.2.3 Validation of reconstructed snow depth product and trend analysis

19 The reconstructed long-term snow depth dataset was evaluated by the stand-alone ground truth measurements over the period
20 1987-2018 from the validation stations (Fig. 1). The reconstructed product was also compared with the static linear-fitting
21 algorithm developed by fitting 19 and 37 GHz with the snow depth measurements with a constant empirical coefficient over
22 China (Che et al., 2008). The daily snow depth data were obtained from the Environmental and Ecological Science Data
23 Center for West China (<http://westdc.westgis.ac.cn>) (hereafter, WESTDC product). Then, the spatiotemporal patterns of snow
24 depth were analyzed in Northeast China (NE), northern Xinjiang (XJ), and the QTP. The slope method (regression) was
25 employed to analyze the snow depth variation trend at the temporal scale (Huang et al., 2019). To show the spatial distribution
26 of snow depth variation, the Mann-Kendall test (significance levels of $\alpha=0.05$) was used to analyze the trends of changes in
27 China (Mann., 1945; Kendall et al., 1975; Milan, 2013). To ensure the presence of dry snow cover, the reconstruction periods
28 are the main snow winter season (January, February, March, November, and December).

29 3 Results

30 3.1 Sensitivity to training sample size

1 The sensitivity of the RF model toward the training sample size was evaluated to confirm the appropriate number of training
2 samples. Fig. 5 displays the accuracy according to unbiased RMSE, bias, and correlation coefficient. These accuracy indexes
3 show slight fluctuations when the number of training sample increases from 5000 to 80,000. Fig. 5a shows that the unbiased
4 RMSE ranges from 5.1 cm to 5.5 cm with increasing training samples. Fig. 5c shows that the correlation coefficient is as high
5 as 0.79 and becomes stable when the samples are up to 30,000. According to the sensitivity analysis, the number of training
6 samples has less influence on the prediction accuracy of the RF model. This test is very helpful for us to determine the number
7 of training samples because of the limited number of training samples over the period 2012-2014. We selected all available
8 samples (28,602) from training stations (Fig. 1) during the period 2012-2014 to train the RF models in Table 2.

9 **3.2 Validation of the fitted RF algorithms**

10 The fitted RF algorithms were evaluated by three validation datasets as shown in Table 3. The color-density scatterplots of the
11 measured snow depth versus the retrieved snow depth are presented in Fig. 6. For all fitted RF algorithms (RF1, RF2, RF3 and
12 RF4), notable differences in accuracy were revealed through the validation of three datasets (Table 4). Generally, the
13 validation with OOB samples presented higher overall accuracy than the other two datasets. This result, however, does not
14 demonstrate that the fitted RF algorithm performs well in snow depth estimation. The assessments in Test2 (temporal subset)
15 and Test3 (spatio-temporal subset) demonstrate that the temporal prediction of the RF model outperforms the spatio-temporal
16 prediction, with unbiased RMSEs of 4.4-5.4 cm and 7.2-7.9 cm, respectively.

17 Comparing the validation results of RF1, RF2, RF3 and RF4, we find that the inclusion of auxiliary information indeed
18 improved the performance of the fitted RF algorithms (Fig. 6). For Test1(OOB), the unbiased RMSE decreased from 6.4 cm to
19 3.9 cm with increasing predictor variables of auxiliary information, while the correlation coefficient increased from 0.72 to
20 0.90 (Table 4). For Test2(temporal subset), the unbiased RMSE decreased from 5.4 cm to 4.4 cm and the correlation
21 coefficient increased from 0.77 to 0.85 (Table 4). There was a slight improvement in spatio-temporal prediction when
22 including the auxiliary information, with the unbiased RMSE ranging from 7.9 cm to 7.3 cm (Table 4).

23 **3.3 Validation of the reconstructed snow depth product**

24 According to the results in Fig. 6 and Table 4, there are no notable differences in accuracy among the RF2, RF3, RF4
25 algorithms. In this study, we selected the RF2 algorithm to reconstruct a long-term snow depth dataset (1987 to 2018). We
26 used the independent in situ measurements over the period 1987-2018 from validation stations (Fig. 1) to evaluate this product
27 (hereafter, RF product). Fig. 7 shows the scatter diagrams of estimated vs. measured values for RF and WESTDC products.
28 The overall accuracy of the RF product is higher than that of the WESTDC estimates, with unbiased RMSEs of 7.1 cm and 8.5
29 cm, respectively (Fig. 7a and 7b). The correlation coefficient is 0.65, which is larger than the WESTDC's coefficient of 0.49.

1 Both products particularly underestimate snow depth when snowpack is thicker than 20 cm. The error bar shows that the
2 WESTDC product tends to more seriously underestimate snow depth than do the RF estimates.

3 To determine the interannual variability in the uncertainty, the time series of assessment indexes, including the unbiased
4 RMSE, bias and correlation coefficient, are shown in Fig. 8. The results show that the RF estimates outperform the WESTDC
5 product with respect to unbiased RMSE and correlation coefficient from season to season. The bias also fluctuates from
6 season to season, ranging from -8 cm to 3 cm (Fig. 8c). There is a slight overestimation during the period 1987-2000, whereas
7 it presents a notable underestimation since 2006. Snow depth estimates with PMW data are usually challenged by the snow
8 metamorphism (e.g., snow grain size). In particular, the large diurnal temperature range in the late snow season leads to a
9 rapid snow grain growth (Dai et al., 2012). Fig. 9 presents the monthly performances of both RF and WESTDC products. The
10 RF estimates outperform the WESTDC product in terms of correlation, overall bias and unbiased RMSE. WESTDC estimates
11 tend to be underestimated in November, December and March, while the RF product is superior to the WESTDC data. Due to
12 the influence of the seasonal evolution of snowpack, the unbiased RMSEs of both products present increasing trends from
13 November to March during the snow seasons. The correlation coefficient in January is the highest among snow season months,
14 which is attributed to stable snow cover.

15 The assessment of snow depth product was performed in three snow cover areas of China. As shown in Fig. 10a, the RF
16 data are superior to the WESTDC estimates, with the unbiased RMSEs of 8.3 cm, 6.8 cm and 8.8 cm in QTP, NE and northern
17 XJ for the RF product, respectively. Fig. 10b shows a notable underestimation and overestimation for the WESTDC product in
18 northern XJ and the QTP, respectively. For the RF product, the bias is close to zero and fluctuates across a relatively narrow
19 range in the three snow cover areas.

20 Based on the results in Fig. 7, we selected 20 cm as a threshold to assess the performances in deep (> 20 cm) and shallow
21 (≤ 20 cm) snow cover. The percentage of shallow snow conditions to total samples was approximately 90%. Table 5 displays
22 the comparison between RF estimates and the WESTDC product in the three snow cover areas. The ‘Samples’ row in Table 5
23 shows the number of samples and the corresponding percentage in each region. Both products present notable underestimation
24 of deep snow cover, with the biases of -34.1 cm and -33.8 cm in QTP for the RF and WESTDC products, respectively. The
25 biases are -10.4 cm and -8.9 cm for the RF product in NE and northern XJ, respectively, whereas the same biases are -11.8 cm
26 and -13.2 cm for the WESTDC data. Moreover, the correlation is very poor in deep snow cover, even negative (-0.18) in QTP
27 for the WESTDC product. For shallow snow cover, the RF product is superior to the WESTDC estimates in QTP, with
28 unbiased RMSEs of 3.4 cm (RF) and 5.6 cm (WESTDC). Furthermore, the WESTDC product presents overestimation in QTP,
29 with a bias of 4.0 cm that is much higher than the RF’s bias of 0.6 cm. The unbiased RMSEs of the RF product are 5.4 cm and
30 6.1 cm in NE and northern XJ for shallow snow cover, respectively, lower than the WESTDC’s values of 6.5 cm and 7.4 cm.

1 However, the RF product tends to overestimate snow depth relative to WESTDC estimates, with higher biases of 1.8 cm and
2 2.5 cm than WESTDC's 0.5 cm and -0.4 cm in NE and northern XJ, respectively.

3 **3.4 Spatial-temporal analysis of snow depth in three snow cover areas**

4 The trend analysis of snow depth was conducted based on ground truth observations, the RF dataset and the WESTDC product
5 during the period 1987-2018. The time series of yearly mean snow depth in different regions over China is shown in Fig. 11.
6 The red, green and blue solid lines represent yearly mean snow depth in northern XJ, NE and QTP, respectively. The black
7 solid line displays the overall mean snow depth in China. Fig. 11a shows that the ground truth snow depth in China presents a
8 significant increasing trend from 1987 to 2018, with a correlation coefficient of 0.57. The trend in NE is highly consistent with
9 the overall trend over China, with a correlation coefficient of 0.64 (Fig. 11a). Although there are increasing trends in northern
10 XJ and QTP, the correlation coefficients are lower than 0.40, not significant (Fig. 11a). Fig. 11b and 11c show the time series
11 of yearly mean snow depth from the RF and WESTDC products, respectively. Neither of these values present significant
12 trends. In the QTP, the WESTDC product presents a significant decreasing trend, with a correlation coefficient of -0.55 (Fig.
13 11c). Snow depth in northern XJ is the greatest among three snow cover areas, and snow cover in the QTP is very shallow,
14 approximately 5 cm (Fig. 11a and 11b). With respect to magnitude and change trends, the ground truth observations and RF
15 estimates in this study are consistent.

16 Fig. 12 shows the spatial patterns of snow depth variation based on the RF and WESTDC products. Only the area with
17 continuous snow depth measurements from 1987 to 2018 is shown in Fig. 12. The two products show similar patterns in the
18 most areas over China. There are notable trend differences between RF and WESTDC products in the northeast of QTP and
19 western NE. The RF product presents an increasing trend in the northeast of QTP, whereas a significant decreasing trend is
20 presented for the WESTDC product (Fig. 12a and 12b). In the western NE, there is a significant increasing for the RF product
21 but no significant trend for WESTDC data.

22 Based on the comparison of trends in Fig. 12 and available station observations in Fig. 1, we selected two specific areas
23 (black and green grids in Fig. 12) to test the changing trend. Fig. 13 shows the trends of snow depth based on the station
24 observations (black solid line), RF estimates (red solid line) and WESTDC product (blue solid line). The ground truth snow
25 depth presents a significant increasing trend in the specific area of NE, with a high correlation coefficient of 0.75 (Fig. 13a).
26 The RF product shows a significant increasing trend, which is consistent with the ground truth data (Fig. 12a and Fig. 13a).
27 Fig. 13b shows that WESTDC product displays a decreasing trend in the selected area of QTP, while station observations and
28 RF estimates present no significant trends.

29 **4 Discussion**

1 **4.1 Disadvantages of the RF model**

2 The RF technique is already used to generate temporal and spatial predictions. Generally, the RF model cannot extrapolate
3 outside the training range (Hengl et al., 2018). Fig. 6 and Table 4 indicate that the spatial predictions of fitted RF algorithms
4 are more biased than are the temporal predictions. Thus, the transferability of a fitted RF algorithm to other areas is in question.
5 Several studies (Prasad, Iverson & Liaw, 2006; Hengl et al., 2017; Vaysse & Lagacherie, 2015; Nussbaum et al., 2018) have
6 proven that RF is a promising technique for spatial prediction; however, these studies aim to obtain spatial predictions of
7 elements of stationarity in the Earth system, e.g., soil types and soil properties.

8 What makes the Earth system interesting is that it is non-stationary (especially concerning snow parameters). Generally,
9 snow depth increases at the beginning of winter and then decreases in spring due to melting. Moreover, snow cover has
10 different spatial patterns in various regions, such as generally deep snow in high-latitude and high-elevation areas. In China,
11 there are five climatological snow classes according to the classification by Sturm et al. (1995). Each snow class is defined by
12 an ensemble of snow stratigraphic characteristics, including snow density, grain size, and crystal morphology, which
13 influences the snowpack's microwave signature (Sturm et al., 2010). These dynamic properties of snow will lead to many
14 cases in which the same satellite T_B corresponds to different snow depths, while the same snow depth is associated with
15 various T_B observations, rendering the fitted RF algorithm suboptimal. Physical snow evolution models, e.g., the Snow
16 Thermal Model (SNTHERM) (Jordan, 1991), SNOWPACK (Lehning et al., 2002a, b), and Crocus (Brun et al., 1989; Vionnet
17 et al., 2012), can be used to simulate snow parameters (e.g., grain size, density) relatively accurately. Thus, integrating a priori
18 knowledge of snowpack into ML techniques has the potential to overcome many limitations that have hindered a more
19 widespread adoption of ML approaches.

20 **4.2 Influence of predictor variables on the RF model**

21 Fig. 6 and Table 4 indicate that the inclusion of correlated predictor variables has a very slight influence in the predictive
22 performance. Geographic location contributes to improving the RF model's temporal and spatio-temporal estimates, and the
23 inclusion of both elevation and land cover fraction does not further improve the performance of the fitted models (Fig. 6). This
24 is because elevation is highly correlated (correlations higher than 0.9) with geographic location (Fig. 3). Fig. 3 also indicates
25 that the correlation between longitude or elevation and land cover type (e.g., grassland, cropland, forest and bareland) is
26 significant. However, this correlation does not mean that the effects of elevation and land cover fraction on fitted RF model
27 can be ignored. We tested the RF algorithms trained with T_B and elevation or land cover fraction data. The results (not shown
28 here) indicate that these auxiliary data do improve the performance of the fitted algorithms. Strongly correlated variables have
29 a very slight influence on the predictive performance of the RF model (Boulesteix et al. 2012). Therefore, in some cases, a few
30 representative predictor variables should be selected.

1 4.3 Potential errors of the reconstructed snow depth

2 Fig. 7 indicates that the RF model does not fully solve the overestimation and underestimation problems. For deep snow (> 20
3 cm), the biases are up to -8.9 cm and -10.4 cm in NE and northern XJ, respectively. Deep snow conditions account for roughly
4 10% of all training samples (Fig. 2). The estimates for deep snow cover in the QTP exhibit a large bias of -34.1 cm. Fig. 6 also
5 illustrates that the fitted RF algorithms have no predictive ability for extremely deep snow conditions, especially in QTP. We
6 checked the training data and found that the extreme high snow depth data (> 60 cm) occurred in QTP. However, the number
7 of such cases is very small. In addition, the station measurements are point values while the satellite grids have a spatial
8 resolution of 25 km × 25 km. Thus, the representativeness of these data is questionable. Snow depth estimation in the
9 mountains remains a challenge (Lettenmaier et al., 2015; Dozier et al., 2016; Dahri et al., 2018). Numerous studies have been
10 conducted on the snow cover over the QTP and have indicated that the snow cover in the Himalayas is higher than elsewhere,
11 ranging from 80% to 100% during the winter (Basang et al., 2017; Hao et al., 2018). Additionally, Dai et al. (2018) showed
12 that deep snow (greater than 20 cm) was mainly distributed in the Himalayas, Pamir, and Southeastern Mountains. Thus, the
13 RF product produced in this paper has poor performance in QTP for the deep snow cover.

14 Table 5 indicates that there is overestimation in NE and northern XJ for shallow snow cover, which may be due to the
15 following reasons. First, the PMW signals are insensitive to thin snow cover (< 5cm), especially for fresh snow with low snow
16 density and snow grain size, which generally results in underestimation (Foster et al., 2005). In contrast, it tends to
17 overestimate snow depth for shallow old snow in the late snow season due to the seasonal evolution of snowpack. For example,
18 the large diurnal temperature range in the late snow season tends to subject the snowpack to frequent freeze-thaw cycles and
19 leads to rapid snow grain (~2 mm) and snow density (200-350 kg/m³) growth and consequently a high T_B difference
20 (Meløysund et al., 2007; Durand et al., 2008; Yang et al., 2015; Dai et al., 2017). Thus, the overall bias and unbiased RMSE
21 for shallow snowpacks (< 10 cm) present increasing trends from November to March in NE and northern XJ (Table 6). Second,
22 frozen soil reduces the accuracy of estimates. Both snow and frozen ground are volume-scattering materials, and they have
23 similar microwave radiation characteristics, making them difficult to distinguish. Third, a limiting factor in estimating snow
24 depth for PMW remote sensing is the presence of liquid water. In this study, a snow cover detection method is used to filter
25 out wet snow cover; however, there are still misclassification errors, especially at the end of the winter season (Grody and
26 Basist., 1996; Liu et al., 2018). In such cases, satellite observations are mainly associated with the emissions from the wet
27 surface of the snowpack. Therefore, in wet snow conditions, snow depth retrieval is not possible (Derksen et al., 2010;
28 Tedesco et al., 2016).

29 5 Conclusions

1 The present study analyzed the application of the RF model to snow depth estimation at temporal and spatial scales.
2 Temporally and spatially independent datasets were applied to verify the fitted RF algorithms. The results suggested that the
3 accuracy of fitted RF algorithms was greatly influenced by auxiliary data, especially the geographic location. However, the
4 inclusion of strongly correlated predictor variables (elevation and land cover fraction) did not further improve the RF
5 estimates. Therefore, in some cases, a few representative predictor variables should be selected. Due to naive extrapolation
6 outside the training range, the transferability of a fitted RF algorithm at the temporal scale was better than that in spatial terms,
7 e.g., with unbiased RMSEs of 4.5 cm and 7.2 cm for the RF2 algorithm, respectively.

8 In this study, the fitted RF2 algorithm was used to retrieve a consistent 32-year daily snow depth dataset from 1987 to
9 2018. Then, an evaluation was carried out using independent reference data from the validation stations during the period
10 1987-2018. The overall unbiased RMSE and bias were 7.1 cm and -0.05 cm, respectively, outperforming the WESTDC
11 product (8.4 cm and -1.20 cm). In QTP, the unbiased RMSE and bias of RF estimates for shallow (≤ 20 cm) snow cover were
12 3.4 cm and 0.59 cm, respectively, much lower than WESTDC's 5.6 cm and 4.02 cm. In NE and northern XJ, RF estimates
13 were superior to the WESTDC product but still presented an underestimation for deep snow (> 20 cm), with biases of -10.4
14 cm and -8.9 cm, respectively.

15 Three long-term (1987-2018) datasets, including ground truth observations, RF estimates and WESTDC product, were
16 applied to analyze the trends of snow depth variation in China. The results suggested that there existed different trends among
17 the three datasets. The overall trend of snow depth in China presented a significant increasing based on the ground truth
18 observations, with a correlation coefficient of 0.57. Moreover, the trend in NE was highly consistent with the overall trend in
19 China, with a correlation coefficient of 0.64. Neither the WESTDC nor the RF product presented significant trends except in
20 QTP. The WESTDC product showed a significant decreasing trend in QTP, with a correlation coefficient of -0.55, whereas
21 there were no significant trends for ground truth observations and the RF product.

22 As discussed in Section 4, our reconstructed snow depth estimates are still challenged by several problems, e.g.,
23 underestimation for deep snow. Additional prior knowledge of snow cover, such as snow cover fraction, snow density, and
24 snow grain size, is necessary to improve the RF model. [Combining the physical snow evolution model \(e.g. SNOWPACK\)](#)
25 with the ML method will be the focus of future work. Furthermore, the mass balance approaches, e.g., the Parallel Energy
26 Balance model, will be used to improve the snow depth retrievals in high-altitude areas. In addition, although our results
27 indicate that the RF method is a promising potential tool for snow depth estimation, there are a few pitfalls such as the risk of
28 naive extrapolation and poor transferability in spatial terms limiting its application in spatio-temporal dynamics. It is in
29 addressing these shortcomings that the techniques of deep learning promise breakthroughs. We are attempting to operate the
30 Deep Neural Networks (DNN) model to overcome the limitations of traditional ML approaches.

31

1 *Author contributions.* L. Jiang conceived and designed the study; J. Yang produced the first draft of the manuscript, which was
2 subsequently edited by J. Lemmetyinen, K. Luoju, L. Jiang and J. Pan; and K. Luoju, M. Takala, S. Wu, J. Pan and J. Yang
3 contributed to the analytical tools and methods.

4
5 *Competing interests.* The authors declare that they have no conflicts of interest.
6

7 *Acknowledgments.* This study was supported by the Science and Technology Basic Resources Investigation Program of China
8 (2017FY100502) and the National Natural Science Foundation of China (41671334). The authors would like to thank the
9 China Meteorological Administration, National Geomatics Center of China, National Snow and Ice Data Center and NASA's
10 Earth Observing System Data and Information System for providing the meteorological station measurements, land cover
11 products and satellite datasets.

12 *Data availability.* Satellite passive microwave measurements are available for download from <https://nsidc.org/>. The in situ
13 measurements provided by the China Meteorology Administration (CMA) and Chinese Snow Survey (CSS) project are not
14 available to the public due to legal constraints. The land use/land cover (LULC) map was provided by the Data Center for
15 Resources and Environmental Sciences, Chinese Academy of Sciences (<http://www.resdc.cn/>). The daily snow depth product
16 was obtained from the Environmental and Ecological Science Data Center for West China (<http://westdc.westgis.ac.cn>).

17 **References**

- 18 Armstrong, R., Knowles, K., Brodzik, M., and Hardman, M.: DMSM SSM/I-SSMIS Pathfinder Daily EASE-Grid Brightness
19 Temperatures, Version 2. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active
20 Archive Center, 10.5067/3EX2U1DV3434, 1994.
- 21 Bair, E. H., Abreu Calfa, A., Rittger, K., and Dozier, J.: Using machine learning for real-time estimates of snow water
22 equivalent in the watersheds of Afghanistan, *The Cryosphere*, 12, 1579-1594, 10.5194/tc-12-1579-2018, 2018.
- 23 Basang, D., Barthel, K., Olseth, J.A.: Satellite and Ground Observations of Snow Cover in Tibet during 2001–2015, *Remote
24 Sensing*, 9, 1201, 10.3390/rs9111201, 2017.
- 25 Belgiu, M., and Lucian, D.: Random forest in remote sensing: A review of applications and future directions, *ISPRS Journal
26 of Photogrammetry and Remote Sensing*, 114, 24-31. 10.1016/j.isprsjprs.2016.01.011, 2016.
- 27 Biau, G.Ã.Š. and Scornet, E.: A random forest guided tour, *TEST*, 25, 197–227, 10.1007%2Fs11749-016-0481-7, 2016.
- 28 Bormann, K.J., Brown, R.D., Derksen, C., Painter, T.H.: Estimating snow-cover trends from space, *Nat. Clim. Chang*, 8, 924–
29 928, 2018.
- 30 Breiman, L., Cutler, A., Liaw, A., Wiener, M.: randomForest: Breiman and Cutler's Random Forests for Classification and
31 Regression, R package version 4.6-14, 2018. <https://CRAN.R-project.org/package=randomForest>.
- 32 Breiman, L. Random forests. *Mach. Learn.* 2001, 45, 5–32, <https://doi.org/10.1023/A:1010933404324>, 2001.
- 33 Brun, E., Martin, E., Simon, V., Gendre, C., and Coleou, C.: An Energy and Mass Model of Snow Cover Suitable for
34 Operational Avalanche Forecasting, *Journal of Glaciology*, 35, 333–342, 10.1017/S002214300009254, 1989.
- 35 Cai, S., Li, D., Durand, M., and Margulis, S.: Examination of the impacts of vegetation on the correlation between snow water
36 equivalent and passive microwave brightness temperature, *Remote Sensing of Environment*, 193, 244–256,
37 10.1016/j.rse.2017.03.006, 2017.
- 38 Canovas-Garcia, F., Alonso-Sarria, F., Gomariz-Castillo, F., and Onate-Valdivieso, F.: Modification of the random forest
39 algorithm to avoid statistical dependence problems when classifying remote sensing imagery, *Comput. Geosci*, 103, 1–11,
40 10.1016/j.cageo.2017.02.012, 2017.

- 1 Chang, A., Foster J., Hall D.: Nimbus-7 derived global snow cover parameters, *Annals of Glaciology*, 9, 39-44,
2 10.1017/S0260305500000355, 1987.
- 3 Che, T., Dai, L., Zheng, X., Li, X., Zhao, K.: Estimation of snow depth from passive microwave brightness temperature data
4 in forest regions of northeast China, *Remote Sensing of Environment*, 183, 334–349, 10.1016/j.rse.2016.06.005, 2016.
- 5 Che, T., Li, X., Jin, R., Armstrong, R., and Zhang, T.: Snow depth derived from passive microwave remote-sensing data in
6 China, *Annals of Glaciology*, 49,145-154,10.3189/172756408787814690, 2008.
- 7 Che, T., Li, X., Jin, R., and Huang, C.: Assimilating passive microwave remote sensing data into a land surface model to
8 improve the estimation of snow depth, *Remote Sensing of Environment*, 143, 54-63,10.1016/j.rse.2013.12.009, 2014.
- 9 Dahri, Z., Moors, E., Ludwig, F., Ahmad, S., Khan, A., Ali, I., Kabat, P.: Adjustment of measurement errors to reconcile
10 precipitation distribution in the high-altitude Indus basin, *Int J Climatol*, 38, 1–19, 10.1002/joc.5539, 2018.
- 11 Dai, L., Che, T., Ding, Y., and Hao, X.: Evaluation of snow cover and snow depth on the Qinghai–Tibetan Plateau derived
12 from passive microwave remote sensing, *The Cryosphere*, 11, 1933–1948, 10.5194/tc-11-1933-2017, 2017.
- 13 Dai, L., Che, T., Xie, H., and Wu, X.: Estimation of Snow Depth over the Qinghai-Tibetan Plateau Based on AMSR-E and
14 MODIS Data, *Remote Sensing*, 10, 1989, 10.3390/rs10121989, 2018.
- 15 Davenport, I., Sandells, M., and Gurney, R.: The effects of variation in snow properties on passive microwave snow mass
16 estimation, *Remote Sensing of Environment*, 118, 168–175, 10.1016/j.rse.2011.11.014, 2012.
- 17 Derksen, C., Walker, A., and Goodison, B.: Evaluation of passive microwave snow water equivalent retrievals across the
18 boreal forest/tundra transition of western Canada, *Remote Sensing of Environment*, 96, 315-327,
19 10.1016/j.rse.2005.02.014, 2005.
- 20 Derksen, C., Toose, P., Rees, A., Wang, L., English, M., Walker, A., and Sturm, M.: Development of a tundra-specific snow
21 water equivalent retrieval algorithm for satellite passive microwave data, *Remote Sensing of Environment*, 114, 1699–
22 1709, 10.1016/j.rse.2010.02.019, 2010.
- 23 Derksen, C., and Brown, R.: Spring snow cover extent reductions in the 2008–2012 period exceeding climate model
24 projections, *Geophysical Research Letters*, 39, 1-6, 10.1029/2012GL053387, 2012.
- 25 Dozier, J., Bair, E. H., and Davis, R. E.: Estimating the spatial distribution of snow water equivalent in the world's mountains,
26 *WIREs Water*, 3, 461-474, doi 10.1002/wat2.1140, 2016.
- 27 Dorji, T., Hopping, K., Wang, S., Piao, S., Tarchen, T., and Klein, J.: Grazing and spring snow counteract the effects of
28 warming on an alpine plant community in Tibet through effects on the dominant species, *Agric. For. Meteorol*, 263, 188–
29 197, 10.1016/j.agrformet.2018.08.017, 2018.
- 30 Durand, M., and Margulis, S.: Feasibility test of multifrequency radiometric data assimilation to estimate snow water
31 equivalent, *Journal of Hydrometeorology*, 7, 443-457, 10.1175/jhm502.1, 2006.
- 32 Durand, M., Kim, E., and Margulis, S.: Quantifying uncertainty in modeling snow microwave radiance for a mountain
33 snowpack at the point-scale, including stratigraphic effects, *IEEE Trans. Geosci. Remote Sens*, 46, 1753–1767,
34 10.1109/tgrs.2008.916221, 2008.
- 35 Fernandes, R., Zhao, H., Wang, X., Key, J., Qu, X., and Hall, A.: Controls on Northern Hemisphere snow albedo feedback
36 quantified using satellite Earth observations, *Geophys. Res. Lett*, 36, 1–6, 10.1029/2009gl040057, 2009.
- 37 Foster, J., Chang, A., Hall D.: Comparison of Snow Mass Estimation From a Prototype Passive Microwave Snow Algorithm,
38 a Revised Algorithm and Snow Depth Climatology, *Remote Sensing of Environment*, 62, 132–142,
39 10.1016/S0034-4257(97)00085-0, 1997.
- 40 Foster, J., Hall, D., Eylander, J., Riggs, G., Nghiem, S., Tedesco, M., Kim, E., Montesano, P., Kelly, R., Casey, K., and
41 Choudhury, B.: A blended global snow product using visible, passive microwave and scatterometer satellite data,
42 *International Journal of Remote Sensing*, 32, 41 1371-1395, 10.1080/01431160903548013, 2011.

1 Grody, N., Basist, A.: Global identification of snow cover using SSM/I measurements, *IEEE Trans. Geosci. Remote Sens.*, 34,
2 237–249, 10.1109/36.481908, 1996.

3 Hao, S., Jiang, L., Shi, J., Wang, G., Liu, X.: Assessment of MODIS-Based Fractional Snow Cover Products Over the Tibetan
4 Plateau, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 99, 1-16,
5 10.1109/JSTARS.2018.2879666, 2018.

6 Hengl, T. et al.: SoilGrids250m: global gridded soil information based on machine learning. *PLoS ONE* 12, e0169748, 2017.

7 Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B.: Random forest as a generic framework for predictive
8 modeling of spatial and spatio-temporal variables, *PeerJ*, 10.7717/peerj.5518, 2018.

9 Huang, C., Newman, A., Clark M., Andrew, W., and Zheng, X.: Evaluation of snow data assimilation using the Ensemble
10 Kalman Filter for seasonal streamflow prediction in the Western United States, *Hydrology and Earth System Sciences*, 21,
11 635-650, 10.5194/hess-21-635-2017, 2017.

12 Ji, D.B., Shi, J.C., Xiong, C., Wang, T.X., Zhang, Y.H.: A total precipitable water retrieval method over land using the
13 combination of passive microwave and optical remote sensing, *Remote Sensing of Environment*, 191, 313-327, 2017.

14 Jiang, L., Shi, J., Tjuatja, S., Dozier, J., Chen, K., and Zhang, L.: A parameterized multiple-scattering model for microwave
15 emission from dry snow, *Remote Sensing of Environment*, 111, 357-366, 10.1016/j.rse.2007.02.034, 2007.

16 Jiang, L., Wang, P., Zhang, L., Yang, H., Yang, J.: Improvement of snow depth retrieval for FY3B-MWRI in China, *Science
17 China: Earth Sciences*, 44,531-47, 10.1007/s11430-013-4798-8,2014.

18 [Jordan, R.E. 1991.: A One-Dimensional Temperature Model for a Snow Cover: Technical Documentation for SNTHERM.89;
19 U.S. Army Cold Regions Research and Engineering Laboratory: Hanover, NH, USA.](#)

20 Kelly, R., Chang, A., Leung, T., and Foster, L.: A prototype AMSR-E global snow area and snow depth algorithm, *IEEE
21 Transactions on Geoscience and Remote Sensing*, 41, 230 - 242, 10.1109/TGRS.2003.809118, 2003.

22 Kelly, R.: The AMSR-E Snow Depth Algorithm: Description and Initial Results, *Journal of The Remote Sensing Society of
23 Japan*, 29, 307-317, 10.11440/rssj.29.307, 2009.

24 Kendall, M. G.: *Rank Correlation Methods*, Griffin, London, 1975.

25 Kevin, J., Kotlarski, S., Scherrer, S., and Schär, C.: The Alpine snow-albedo feedback in regional climate models, *Climate
26 Dynamics*, 48, 1109–1124, 10.1007/s00382-016-3130-7, 2017.

27 Kühnlein, M., Appelhans, T., Thies, B. & Nauss, T.: Improving the accuracy of rainfall rates from optical satellite sensors
28 with machine learning—a random forests-based approach applied to MSG SEVIRI, *Remote Sens. Environ.*, 141,129–143,
29 2014.

30 [Lehning, M., Bartelt, P., Brown, B., Fierz, C., Satyawali, P.: A physical SNOWPACK model for the Swiss avalanche warning
31 part II. Snow microstructure, *Cold Reg. Sci. Technol.*, 35\(3\), 147–167, 10.1016/S0165-232X\(02\)00073-3, 2002a.](#)

32 [Lehning, M., Bartelt, P., Brown, B., Fierz, C.: A physical SNOWPACK model for the Swiss avalanche warning: Part III:
33 meteorological forcing, thin layer formation and evaluation, *Cold Reg. Sci. Technol.*, 35\(3\):169–184,
34 10.1016/S0165-232X\(02\)00072-1, 2002b.](#)

35 Lemmetyinen, J., Derksen, C., Toose, P., Proksch, M., Pulliainen, J., Kontu, A., Rautiainen, K., and Seppänen, J.: Hallikainen,
36 M. Simulating seasonally and spatially varying snow cover brightness temperature using HUT snow emission model and
37 retrieval of a microwave effective grain size, *Remote Sensing of Environment*, 156, 71–95, 10.1016/j.rse.2014.09.016,
38 2015.

39 Lettenmaier, D., Alsdorf, D., Dozier, J., Huffman, G., Pan, M., and Wood, E.: Inroads of remote sensing into hydrologic
40 science during the WRR era, *Water Resour. Res.*, 51, 7309-7342, 10.1002/2015WR017616, 2015.

- 1 Li, Q., Kelly, R.: Correcting Satellite Passive Microwave Brightness Temperatures in Forested Landscapes Using Satellite
2 Visible Reflectance Estimates of Forest Transmissivity, *IEEE Journal of Selected Topics in Applied Earth Observations
3 and Remote Sensing*, 10, 3874-3883, 10.1109/JSTARS.2017.2707545, 2017.
- 4 Liaw, A., and Wiener, M.: Classification and regression by randomForest, *R News*, 2, 18–22, 2002.
- 5 Liu, X., Jiang, L., Wu, S., Hao, S., Wang, G., and Yang, J.: Assessment of Methods for Passive Microwave Snow Cover
6 Mapping Using FY-3C/MWRI Data in China, *Remote Sensing*, 10, 524-539, 10.3390/rs10040524, 2018.
- 7 Liu, X., Jiang, L., Wang, G., Hao, S., and Chen, Z.: Using a Linear Unmixing Method to Improve Passive Microwave Snow
8 Depth Retrievals, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 11, 4414–4429, 10.1109/PIERS.2016.7735542, 2018.
- 9 Maxwell, A., Warner, T., and Fang, F.: Implementation of machine-learning classification in remote sensing:
10 An applied review, *Int. J. Remote Sens.*, 39, 2784–2817, 2018.
- 11 Mann, H. B.: Nonparametric tests against trend, *Econometrica* 13, 245–259, 1945.
- 12 Milan, G., and Slavisa, T.: Analysis of changes in meteorological variables using Mann-Kendall and Sen’s slope estimator
13 statistical tests in Serbia, *Global Planet Change*, 100, 172-182, 10.1016/j.gloplacha.2012.10.014, 2013.
- 14 Meløysund, V., Bernt, L., Karl, V., and Kim R.: Predicting snow density using meteorological data, *Meteorological
15 Applications*, 14, 413–423, 10.1002/met.40, 2007.
- 16 Metsämäki, S., Pulliainen, J., Salminen, M., Luoju, K., Wiesmann, A., Solberg, R., Böttcher, K., Hiltunen, M., and Ripper, E.:
17 Introduction to GlobSnow Snow Extent products with considerations for accuracy assessment, *Remote Sensing of
18 Environment*, 156, 96–108, 10.1016/j.rse.2014.09.018, 2015.
- 19 Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M., Papritz, A.: Evaluation of
20 digital soil mapping approaches with large sets of environmental covariates, *Soil*, 4, 1, 10.5194/soil-4-1-2018, 2018.
- 21 Orsolini, Y., Wegmann, M., Dutra, E., Liu, B., Balsamo, G., Yang, K., de Rosnay, P., Zhu, C., Wang, W., Senan, R., and
22 Arduini, G.: Evaluation of snow depth and snow cover over the Tibetan Plateau in global reanalyses using in situ and
23 satellite remote sensing observations, *The Cryosphere*, 13, 2221–2239, 10.5194/tc-13-2221-2019, 2019.
- 24 Pan, J., Durand, M., Vander Jaqt, B., and Liu, D.: Application of a Markov Chain Monte Carlo algorithm for snow water
25 equivalent retrieval from passive microwave measurements, *Remote Sensing of Environment*, 192, 150-165,
26 10.1016/j.rse.2017.02.006, 2017.
- 27 Picard, G.: Simulation of the microwave emission of multi-layered snowpacks using the dense media radiative transfer theory:
28 The DMRT-ML model, *Geosci. Model Develop. Discuss.*, 6, 3647–3694, 2012.
- 29 Prasad, A., Iverson, L., and Liaw, A.: Newer classification and regression tree techniques: bagging and random forests for
30 ecological prediction, *Ecosystems*, 9, 181-199, 10.1007/s10021-005-0054-1, 2006.
- 31 Probst, P., and Boulesteix, A.: To tune or not to tune the number of trees in random forest, *J. Mach. Learn. Res.*, 18, 1–18,
32 2018.
- 33 Pulliainen, J., Grandell, J., and Hallikainen, M.: HUT snow emission model and its applicability to snow water equivalent
34 retrieval, *IEEE Trans. Geosci. Remote Sens.*, 37, 1378–1390, 10.1109/36.763302, 1999.
- 35 Pulliainen, J.: Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne
36 microwave radiometer data and ground-based observations, *Remote Sens. Environ.*, 101, 257–269,
37 10.1016/j.rse.2006.01.002, 2006.
- 38 Qu, Y., Zhu, Z., Chai, L., Liu, S., Montzka, C., Liu, J., Yang, X., Lu, Z., Jin, R., Li, X., Guo, Z., and Zheng, J.: Rebuilding a
39 Microwave Soil Moisture Product Using Random Forest Adopting AMSR-E/AMSR2 Brightness Temperature and
40 SMAP over the Qinghai–Tibet Plateau, China, *Remote Sensing*, 11, 683, 10.3390/rs11060683, 2019.
- 41 Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat.: Deep learning and process
42 understanding for data-driven Earth system science, *Nature* 566, 195–204, 2019.

1 Rodriguez-Galiano, V., Ghimire, B., Rogan, J., Chica-Olmo, M., and Rigol-Sanchez, J.: An assessment of the effectiveness of
2 a random forest classifier for land-cover classification, *ISPRS J. Photogramm. Remote Sens.*, 67, 93–104,
3 10.1016/j.isprsjprs.2011.11.002, 2012.

4 Roy, A., Royer, A., and Hall R.: Relationship Between Forest Microwave Transmissivity and Structural Parameters for the
5 Canadian Boreal Forest, *IEEE Geoscience and Remote Sensing Letters*, 11, 1802-1806,10.1109/LGRS.2014.2309941,
6 2014.

7 Saberi, N., Kelly, R., Toose, P., Roy, A., and Derksen, C.: Modeling the observed microwave emission from shallow
8 multi-layer tundra snow using DMRT-ML, *Remote Sensing*, 9, 1327, 10.3390/rs9121327, 2017.

9 Safavi, H., Sajjadi, S., and Raghobi, V.: Assessment of climate change impacts on climate variables using probabilistic
10 ensemble modeling and trend analysis, *Theoretical and Applied Climatology*, 130, 635–653, 10.1007/s00704-016-1898-3,
11 2017.

12 Santi, E., Pettinato, S., Paloscia, S., Pampaloni, P., MacElloni, G., and Brogioni, M.: An algorithm for generating soil
13 moisture and snow depth maps from microwave spaceborne radiometers: HydroAlgo, *Hydrology and Earth System
14 Sciences*, 16, 3659–3676, 10.5194/hess-16-3659-2012, 2012.

15 Sturm, M., Holmgren, J., Liston, G.E.: A seasonal snow cover classification system for local to global applications, *J. Clim.*, 8,
16 1261–1283, 1995.

17 Sturm, M., and Wagner, A.M.: Using repeated patterns in snow distribution modeling: An arctic example, *Water Resour. Res.*,
18 46, 65–74, 2010.

19 Takala, M., Luojus, K., Pulliainen, J., Lemmetyinen, J., Juha-Petri, K., Koskinen, J., and Bojkov, B.: Estimating northern
20 hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and
21 ground-based measurements, *Remote Sensing of Environment*, 115, 3517-3529, 10.1016/j.rse.2011.08.014, 2011.

22 Takala, M., Ikonen, J., Luojus, K., Lemmetyinen, J., Metsämäki, S., Cohen, J., Arslan, A., and Pulliainen J.: New Snow Water
23 Equivalent Processing System With Improved Resolution Over Europe and its Applications in Hydrology, *IEEE Journal
24 of Selected Topics in Applied Earth Observations and Remote Sensing*, 10, 428-436, 10.1109/JSTARS.2016.2586179,
25 2017.

26 Tedesco, M., and Narvekar, P.: Assessment of the NASA AMSR-E SWE product, *IEEE Journal of Selected Topics in Applied
27 Earth Observations and Remote Sensing*, 3, 141-159, 10.1109/jstars.2010.2040462, 2010.

28 Tedesco, M., and Jeyaratnam, J.: A new operational snow retrieval algorithm applied to historical AMSR-E brightness
29 temperatures, *Remote Sensing*, 8, 1037, 10.3390/rs8121037, 2016.

30 Tyralis, H., Papacharalampous, G., and Langousis, A.: A Brief Review of Random Forests for Water Scientists and
31 Practitioners and Their Recent History in Water Resources, *Water*, 11, 910, 2019a.

32 Tyralis, H., Papacharalampous, G., and Tantane, S.: How to explain and predict the shape parameter of the generalized
33 extreme value distribution of streamflow extremes using a big dataset, *Journal of Hydrology*, 574, 628–645,
34 10.1016/j.jhydrol.2019.04.070, 2019b.

35 Vaysse, K., and Lagacherie, P.: Evaluating digital soil Mapping approaches for mapping GlobalSoilMap soil properties from
36 legacy data in Languedoc-Roussillon (France), *Geoderma Regional*, 4, 20-30, 10.1016/j.geodrs.2014.11.003, 2015.

37 Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Le Moigne, P., Martin, E., Willemet, J.-M.: The detailed snowpack
38 scheme Crocus and its implementation in SURFEX v7.2, *Geosci. Model Dev*, 5, 773–791, 2012.

39 Xue, Y., and Forman, B.A.: Atmospheric and Forest Decoupling of Passive Microwave Brightness Temperature Observations
40 Over Snow-Covered Terrain in North America, *IEEE Journal of Selected Topics in Applied Earth Observations &
41 Remote Sensing*, 10, 3172–3189, 2017.

1 Yang, J., Jiang, L., Ménard, C., Luo, J., Lemmetyinen, J., and Pulliainen, J.: Evaluation of snow products over the Tibetan
2 Plateau, *Hydrol. Processes*, 29, 3247–3260, 10.1002/hyp.10427, 2015.

3 Yang, J., Jiang, L., Wu, S., Wang, G., Wang, J., and Liu, X.: Development of a Snow Depth Estimation Algorithm over China
4 for the FY-3D/MWRI, *Remote Sensing*, 11, 977, 10.3390/rs11080977, 2019.

5 Zhong, X., Zhang, T., Kang, S., Wang, K., Zheng, L., Hu, Y., and Wang, H.: Spatiotemporal variability of snow depth across
6 the Eurasian continent from 1966 to 2012, *The Cryosphere*, 12, 227-245, 10.5194/tc-12-227-2018, 2018.

7 Ziegler, A., König, I.R.: Mining data with random forests: Current options for real-world applications, *Wiley Interdiscip. Rev.*
8 *Data Min. Knowl. Discov.* 4, 55–63, 10.1002/widm.1114, 2014.

9

10 **List of Tables and Figures**

11 Table 1. Summary of the main passive microwave remote sensing sensors.

Sensor	SSM/I			SSMIS
Satellite	DMSP-F08	DMSP-F11	DMSP-F13	DMSP-F17
On Orbit time	1987-1991	1991-1995	1995-2008	2006-present
Passing Time	A: 06:20	A: 17:17	A: 17:58	A: 17:31
	D: 18:20	D: 05:17	D: 05:58	D: 05:31
Frequency & footprint (GHz): (km × km)		19.35: 45×68		19.35: 42×70
		23.235: 40×60		23.235: 42×70
		37: 24×36		37: 28×44
		85.5: 11×16		91.655: 13×15

12

13 Table 2. A detailed description of the input predictor variables based on four selection rules of the training sample.

Name	Predictor Variables	Target	Note
RF1	T_{B19V} , T_{B37V}		land cover types:
RF2	T_{B19V} , T_{B37V} , Latitude, Longitude	snow	grassland, cropland,
RF3	T_{B19V} , T_{B37V} , Latitude, Longitude, Elevation	depth	bareland, shrubland,
RF4	T_{B19V} , T_{B37V} , Latitude, Longitude, Elevation, Land cover fraction		forest

14

15 Table 3. Summary of three tests of the fitted RF algorithms in Table 2.

Name	Test1 (OOB)		Test2 (temporal subset)		Test3 (spatio-temporal subset)	
training	training stations	2012-2014	training stations	2012-2014	training stations	2012-2014
	samples	28602	samples	28602	samples	28602
validation	training stations	2012-2014	training stations	2015-2018	validation stations	2015-2018
	samples	14301	samples	34684	samples	25879

16

17 Table 4. Accuracy of four snow-depth retrieval models with unbiased RMSE, bias and correlation coefficient.

Name	Test1 (OOB)			Test2 (temporal subset)			Test3 (spatio-temporal subset)		
	unRMSE	bias	corr.coe	unRMSE	bias	corr.coe	unRMSE	bias	corr.coe
RF1	6.4	-0.01	0.72	5.4	0.12	0.77	7.9	-0.76	0.57
RF2	4.1	0.07	0.90	4.5	0.27	0.85	7.2	-0.97	0.66
RF3	3.9	0.08	0.90	4.5	0.24	0.85	7.3	-0.83	0.66

RF4 3.9 0.03 0.91 4.4 0.21 0.85 7.3 -0.40 0.65

1
2
3

Table 5. Comparison between RF estimates and WESTDC product in three stable snow cover areas for deep (> 20 cm) and shallow (≤ 20 cm) snow cover.

RF product						
Regions	QTP		NE		northern XJ	
SnowDepth (cm)	≤ 20	> 20	≤ 20	> 20	≤ 20	> 20
corr.coe	0.30	0.06	0.49	0.17	0.48	0.31
bias (cm)	0.59	-34.12	1.79	-10.38	2.52	-8.85
unRMSE (cm)	3.43	20.70	5.36	7.00	6.12	9.62
Samples	15503 (96.4%)	583 (3.6%)	151939 (87.3%)	22168 (12.7%)	32468 (69.8%)	14051 (30.2%)

WESTDC product						
Regions	QTP		NE		northern XJ	
SnowDepth (cm)	≤ 20	> 20	≤ 20	> 20	≤ 20	> 20
corr.coe	0.16	-0.18	0.37	0.03	0.34	0.16
bias (cm)	4.02	-33.78	0.47	-11.75	-0.39	-13.22
unRMSE (cm)	5.60	21.62	6.47	9.10	7.35	11.30
Samples	15503 (96.4%)	583 (3.6%)	151939 (87.3%)	22168 (12.7%)	32468 (69.8%)	14051 (30.2%)

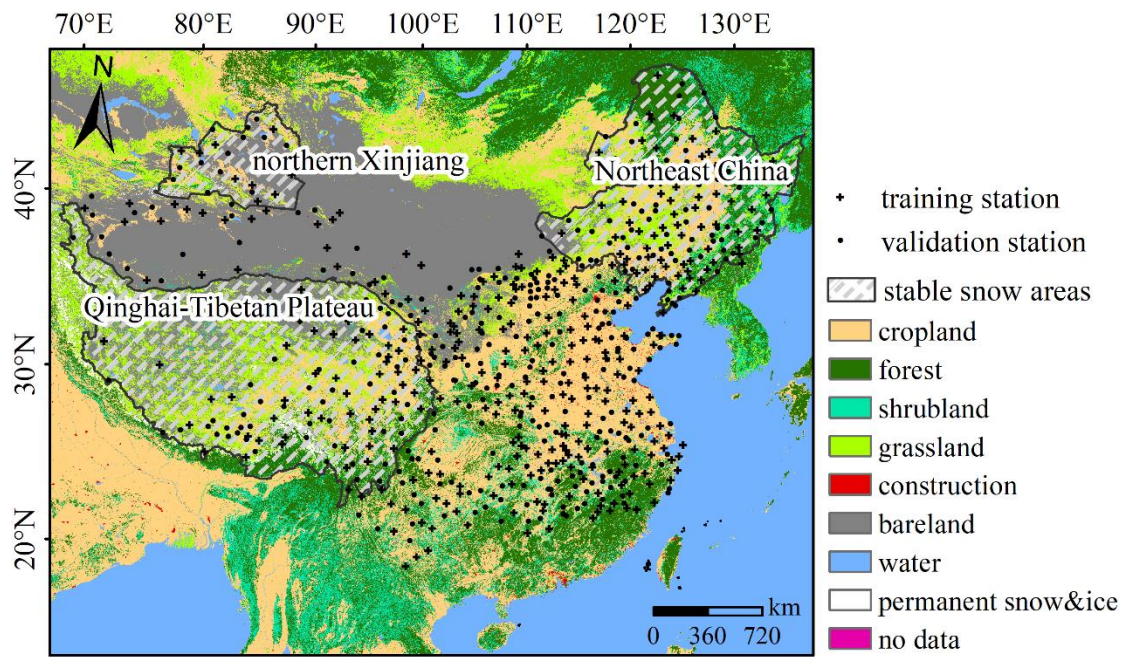
4
5

Table 6. Summary of monthly performances of the RF product in NE and northern XJ.

NE					
Month	November	December	January	February	March
corr.coe	0.32	0.41	0.40	0.23	0.08
bias (cm)	2.33	2.19	2.93	4.74	7.97
unRMSE (cm)	3.66	3.69	4.16	5.24	6.16

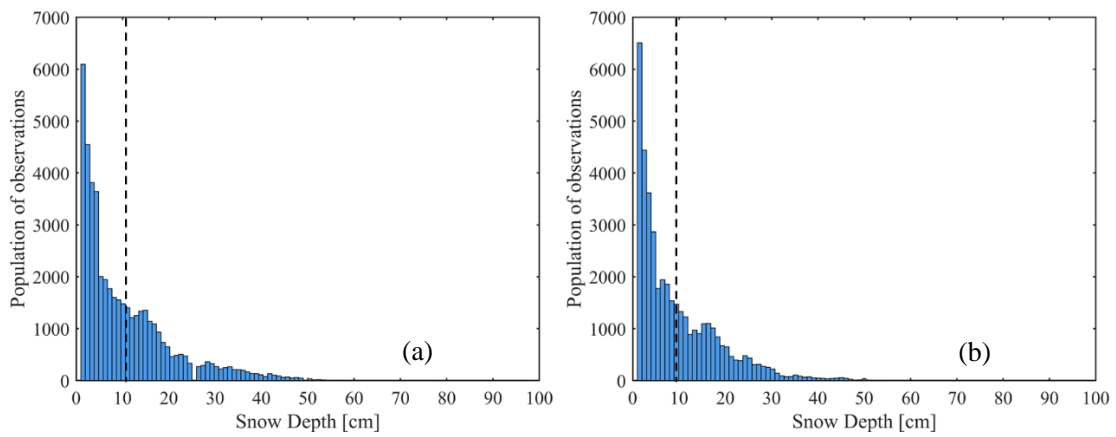
northern XJ					
Month	November	December	January	February	March
corr.coe	0.20	0.27	0.40	0.20	0.08
bias (cm)	3.68	3.35	2.97	5.65	10.60
unRMSE (cm)	4.49	4.77	4.61	6.83	7.09

6



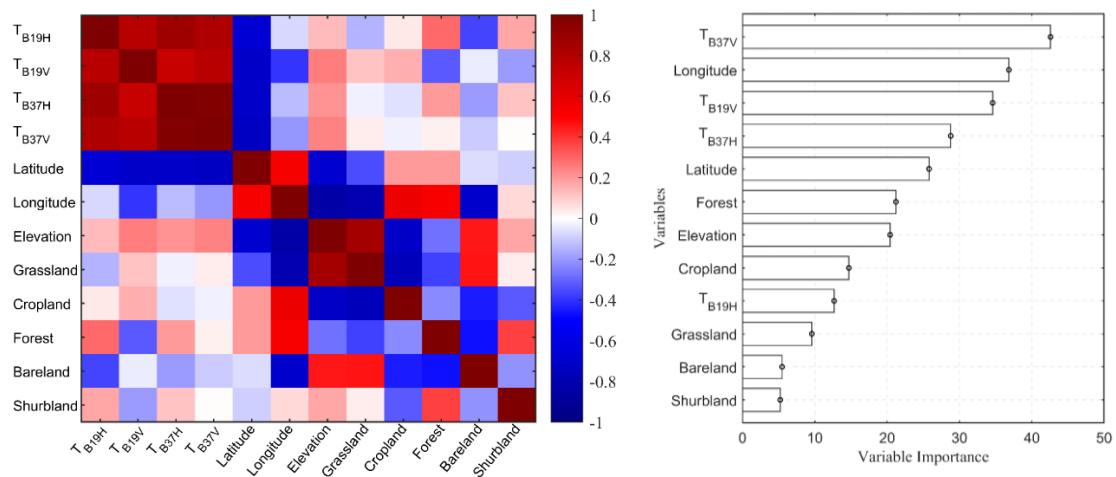
1
2
3
4

Figure 1. Spatial distribution of the weather stations and land cover types in the study area. There are three stable snow cover areas in China: Northeast China (NE), northern Xinjiang (XJ) and the Qinghai-Tibetan Plateau (QTP).



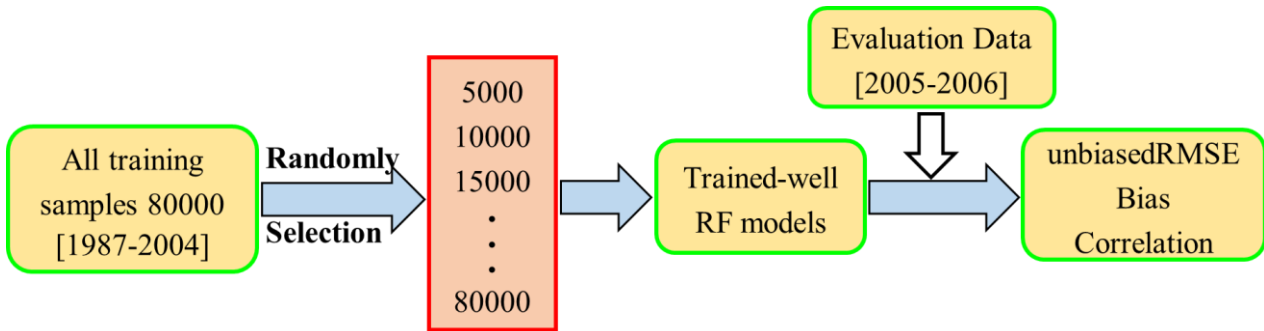
5
6
7
8

Figure 2. Histograms of snow depth observations from (a) training and (b) validation stations. The average values (black dashed lines) are equal to 10.5 cm and 9.8 cm, respectively.



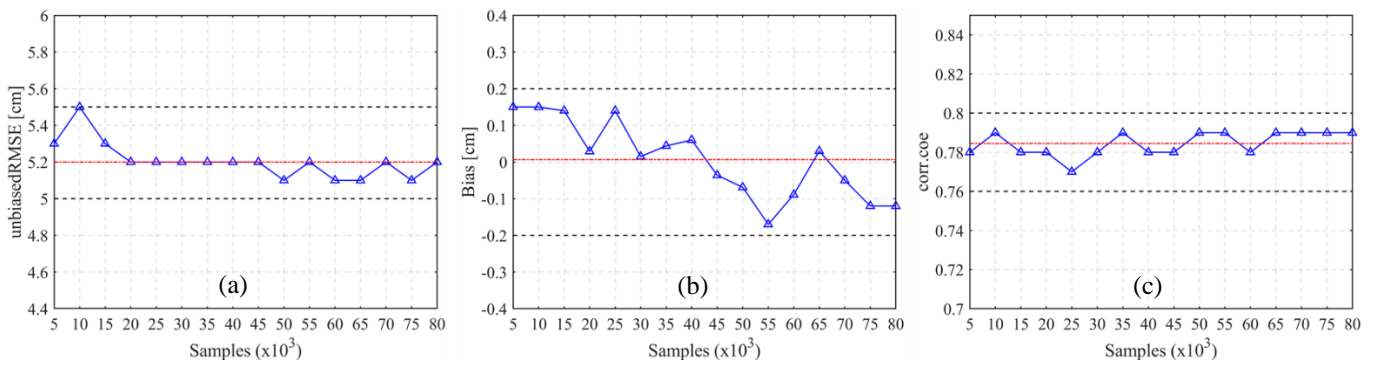
9

1 Figure 3. Correlations between the predictor variables (left) and the ranking of variable importance (right). The
 2 importance of variables, referred to as Mean Decrease Accuracy (MDA) in the RF model, is obtained by averaging the
 3 difference in out-of-bag error estimation before and after the permutation over all trees. The larger the MDA, the greater
 4 the importance of the variable is.



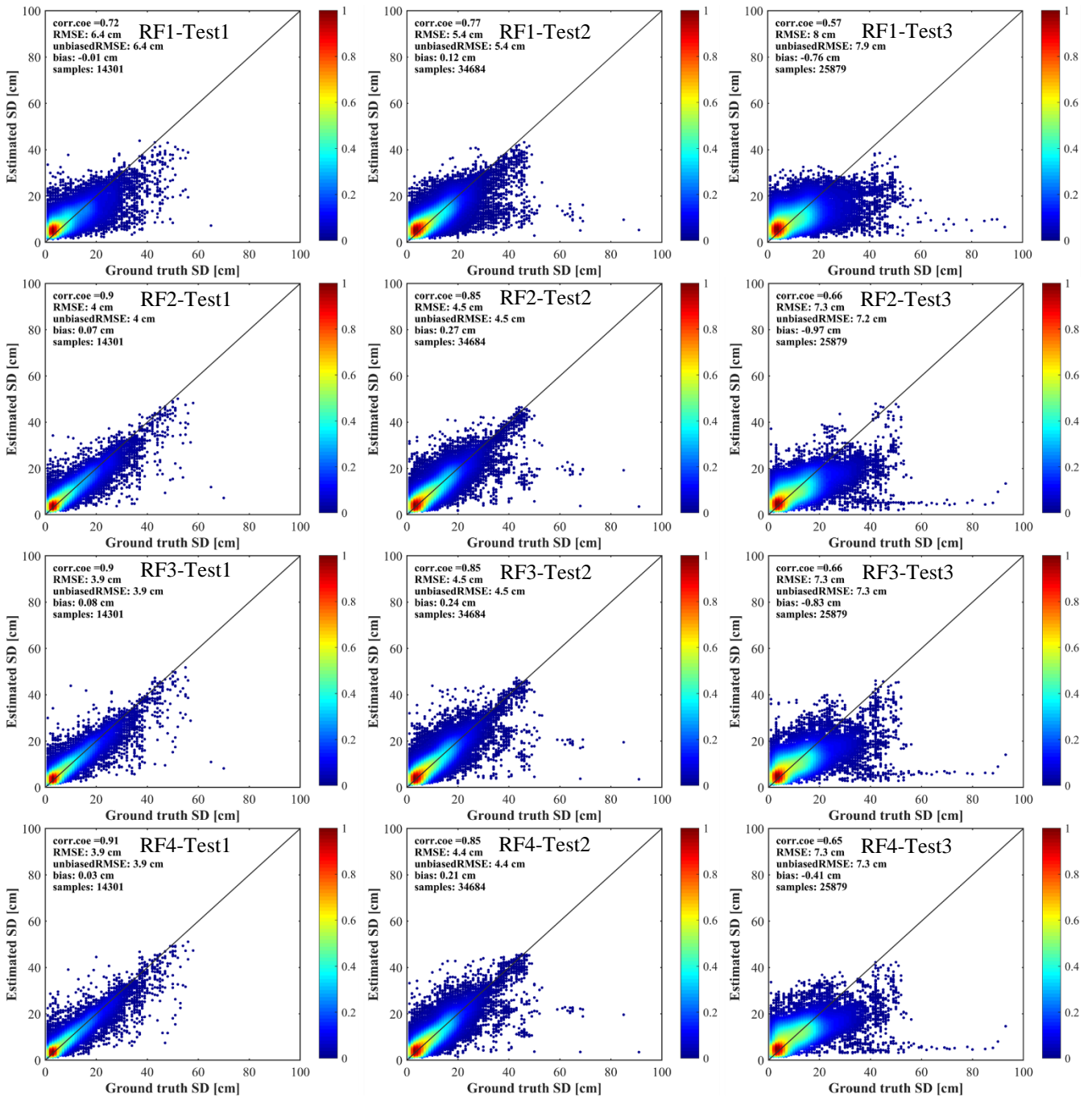
5
6

Figure 4. The test process flowchart for the sensitivity of the RF model to the training sample size.



7
8

Figure 5. Trends of (a) unbiased RMSE, (b) bias and (c) correlation coefficient with increasing training sample size.



1
2
3
4
5

Figure 6. The color-density scatterplots of the estimated snow depth with four fitted RF algorithms and the ground truth snow depth. The four trained RF algorithms (RF1, RF2, RF3, RF4) were evaluated with three validation-test datasets (Test1, Test2, Test3).

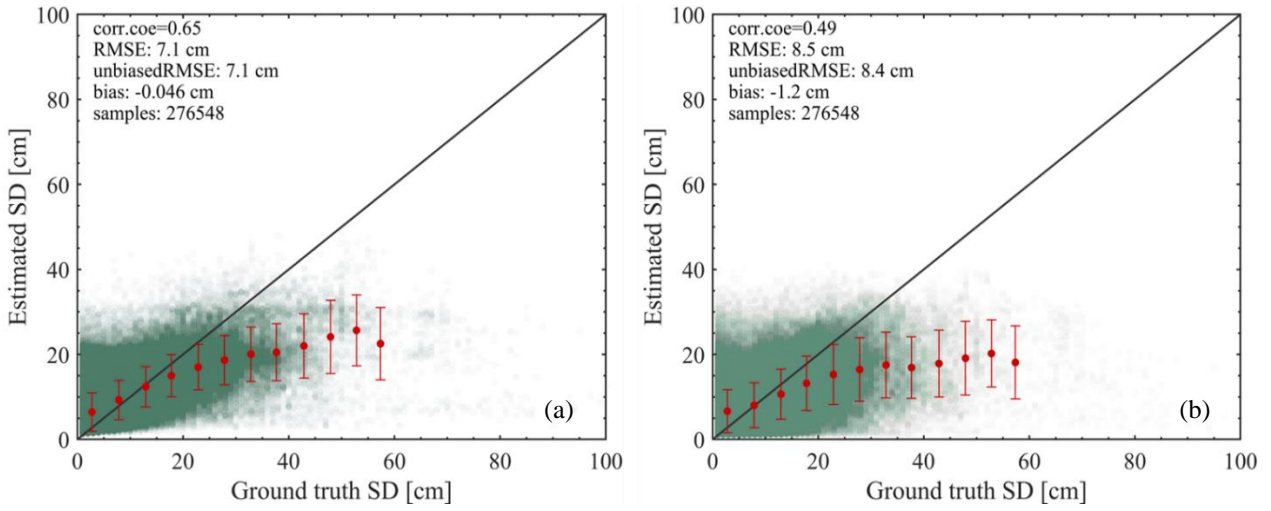


Figure 7. Scatterplots of the estimated snow depth and the ground truth observation for (a) RF and (b) WESTDC products.

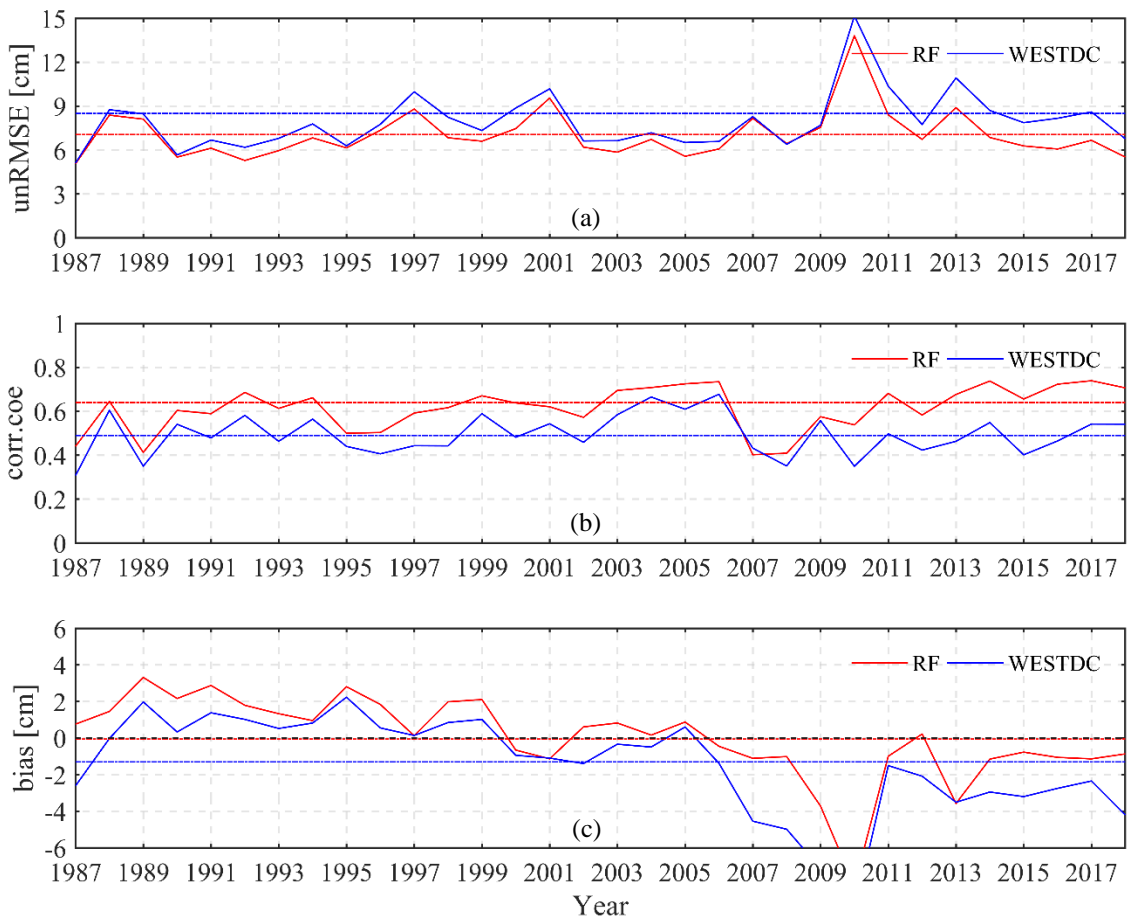


Figure 8. Time series of (a) unbiased RMSE (unRMSE), (b) correlation coefficient (corr.coe) and (c) bias for RF and WESTDC products. The colorful dashed lines represent mean values of assessment indexes.

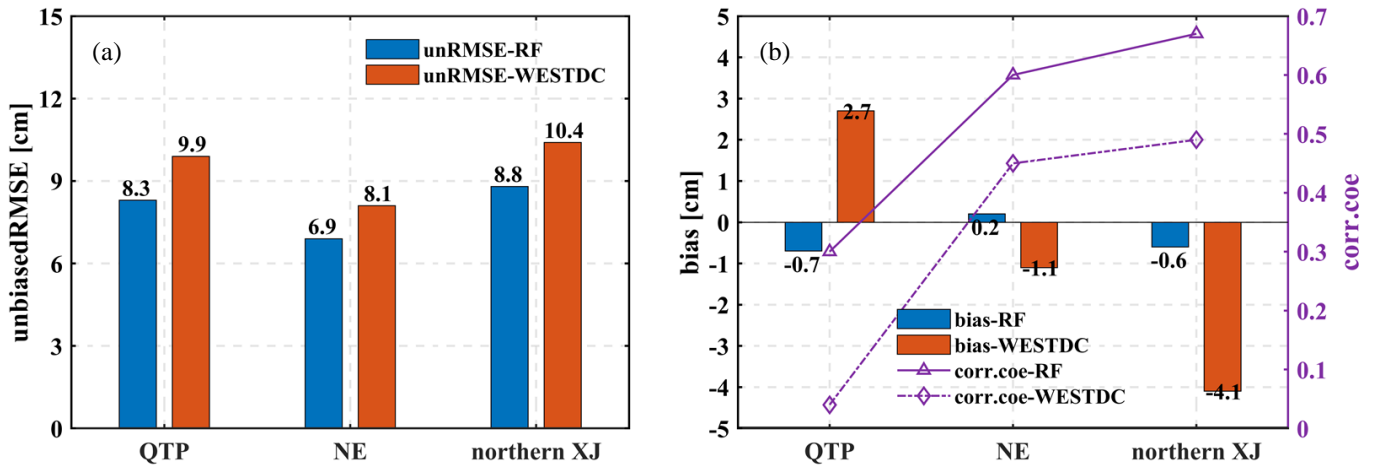


Figure 9. The validation of RF and WESTDC snow depth products in three stable snow cover areas over China with respect to (a) the unbiased RMSE, (b) bias and correlation coefficient.

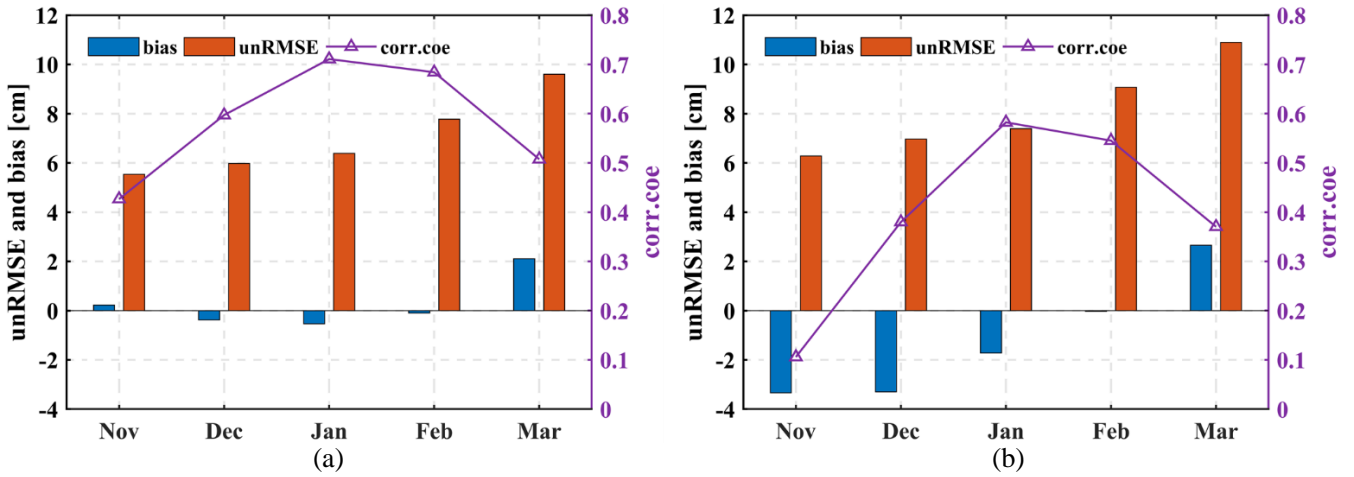


Figure 10. Monthly performances of (a) RF, and (b) WESTDC snow depth products. Nov: November; Dec: December; Jan: January; Feb: February; Mar: March.

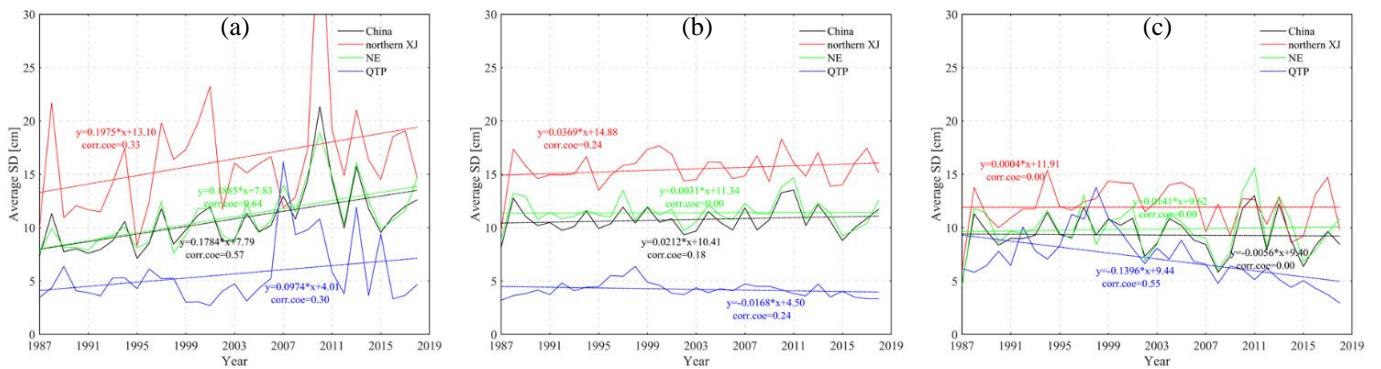
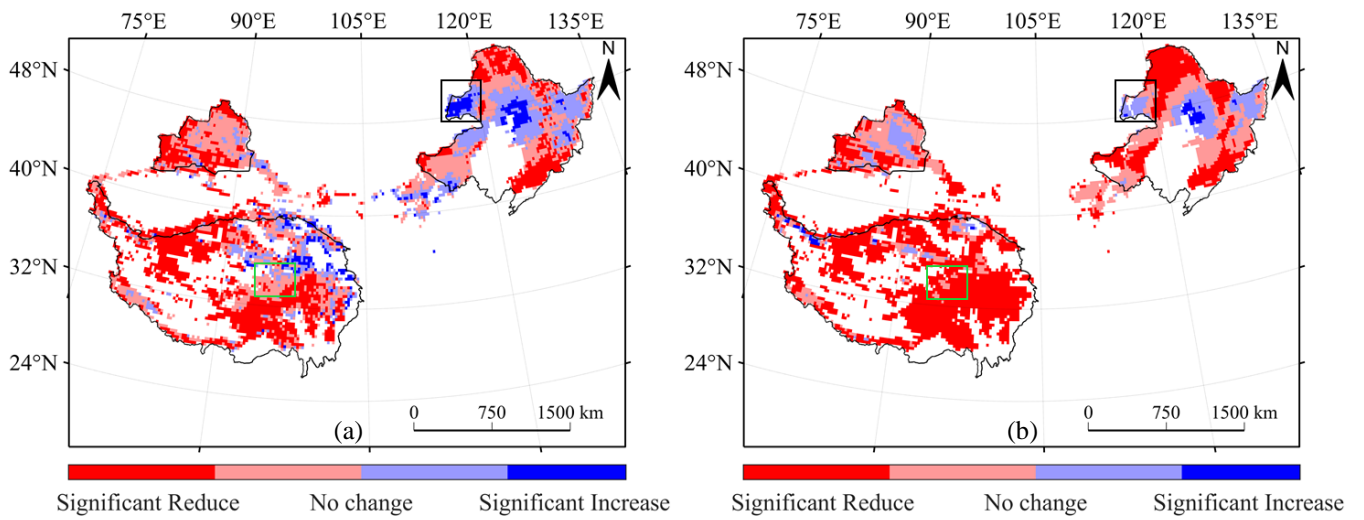
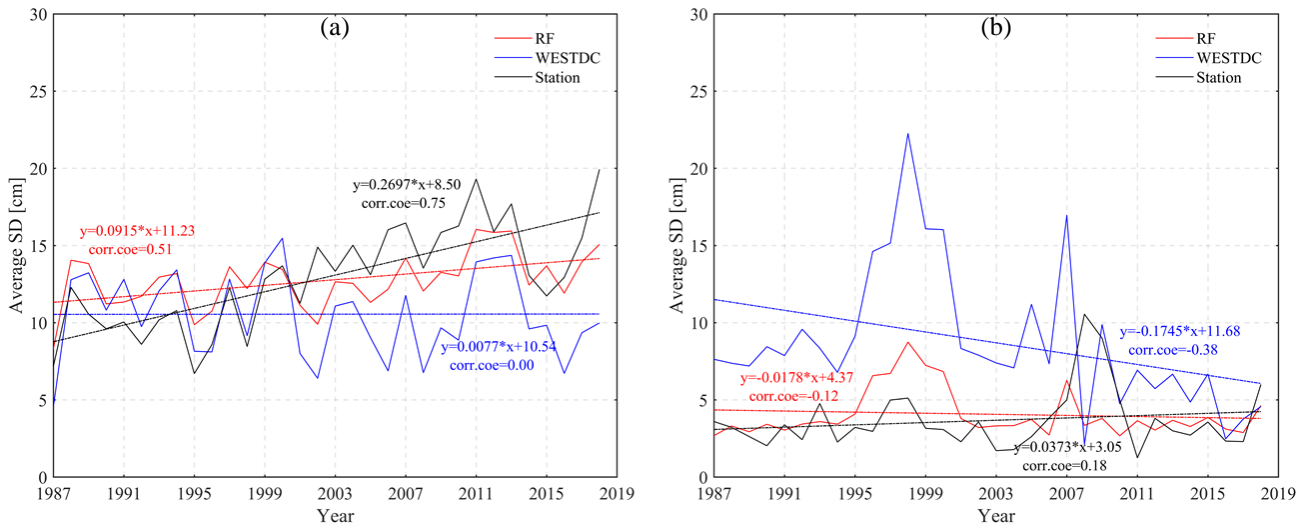


Figure 11. Trend analysis of snow depth based on (a) station observations, (b) RF estimates, and (c) WESTDC product in three stable snow cover areas of China. The correlation is statistically significant at the 0.05 level.



1
2
3
4
Figure 12. Trend analysis of snow depth during the period 1987-2018: (a) RF product; (b) WESTDC data. Light red and light blue represent no significant trend changes.



5
6
7
Figure 13. Comparison of changing trends of snow depth between RF estimates and WESTDC product in specific areas of (a) NE and (b) QTP.