Dear Editor Florent Dominé

Thanks for your understanding and your constructive suggestions on the data sharing policy and platform. We corrected the English website of station data application (http://data.cma.cn/en) in the revised manuscript. And I hope the data sharing policy would be more convenient and open to all over the world in the coming future. The Chinese scientists have being promoted the data sharing to all over the world, including providing English version website and open to all.

Best regards,

Lingmei Jiang, on behalf of the authors

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

Jianwei Yang[1], Lingmei Jiang[1]*, Kari Luojus[2], Jinmei Pan[3], Juha Lemmetyinen[2], Matias Takala[2], Shengli Wu[4]

[1]State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and Aerospace Information Research Institute 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

[2]Finnish Meteorological Institute, PO Box 503, FIN-00101 Helsinki, Finland

[3]State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China

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

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

Abstract. We investigated the potential capability of the random forest (RF) machine learning (ML) model on estimation of snow depth in this work. 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 the 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; and (3) the fitted RF algorithms perform better on temporal than spatial scales, with unbiased root mean square errors (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, indicating better performance than that of 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 for 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 on the QTP. The WESTDC product presented a significant decreasing trend on the 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 Introduction

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

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

Machine learning (ML) has attained outstanding results in the regression estimation of land surface parameters from remotely sensed observations at local and global scales over the past decade (Reichstein et al., 2019). The random forest (RF) is an ensemble method whereby multiple trees are grown from random subsets of predictors, producing a weighted ensemble of trees (Breiman, 2001). RF is also robust against overfitting in the presence of large datasets and increases predictive accuracies over single decision trees (Biau and Scornet, 2016; Tyralis et al., 2019b). Over the last two decades, RF has been one of the most successful ML algorithms for practical applications due to its proven accuracy, stability, speed of processing 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; Reichstein et al., 2019; Tyralis et al., 2019a). Although the RF model can present good results in many research areas, studies on the spatiotemporal prediction of snow depth are few and the potential utility of RF in such studies is unknown.

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

2 Data and Methodology

2.1 Data

(1) Satellite passive microwave measurements

The series of the Special Sensor Microwave/Imager (SSM/I) and Special Sensor Microwave Imager Sounder (SSMIS) instruments has provided continuous T_B measurements at 19.35, 23.235, 37, 85.5 and 91.655 GHz since July 1987. The data are available from the National Snow and Ice Center (https://daacdata.apps.nsidc.org/pub/DATASETS). The SSM/I and SSMIS sensors are suitable for producing a long-term consistent snow depth dataset due to their similar configurations and intersensor calibrations (Armstrong et al., 1994). To avoid the influence of wet snow, only ascending (F08) and descending (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) GHz was ignored (hereafter referred as 19 GHz and 37 GHz, respectively).

(2) In situ measurements

The weather station daily data in China from 1987 to 2018 were provided by the National Meteorological Information Centre, China Meteorology Administration (CMA, http://data.cma.cn/en). The geographical locations of the meteorological stations and the three stable snow cover areas are shown in Fig. 1. The recorded variables include the site name, observation time, geolocation (latitude and longitude), altitude (m), near-surface soil temperature (measured at a 5-cm depth, °C), and snow depth (cm). The sites are not distributed homogeneously, and few are located in inaccessible regions with extreme climates and complex terrain conditions, e.g., the western part of QTP (Fig. 1).

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

(3) Land cover fraction

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

2.2 Methodology

2.2.1 Random forest

RF is an ensemble ML algorithm proposed by Breiman in 2001. It combines several randomized decision trees and aggregates their predictions by averaging in regression (Biau and Scornet, 2016). Generally, approximately two-thirds of the samples (in-bag samples) are used to train the trees and the remaining one-third (out-of-bag samples, OOB) are used to estimate how well the fitted RF algorithm performs. Few user-defined parameters are generally required to optimize the algorithm, such as 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 1000 in the present study since the gain in the predictive performance of the algorithm would be small with the addition of more trees (Probst and Boulesteix, 2018). The default value of *m*try is determined by the number of input prediction variables, usually 1/3 for regression tasks (Biau and Scornet, 2016). The RF regression is insensitive to the quality of training samples and to overfitting due to the large number of decision trees produced by randomly selecting a subset of training samples and a subset of variables for splitting at each tree node (Maxwell et al., 2018). In addition, RF provides an assessment of the relative importance of predictor variables, which have proven to be useful for evaluating the relative contribution of input variables (Tyralis et al., 2019b). Furthermore, the RF model can rapidly trained and is easy to use. In this paper, a randomForest R package (Version 4.6-14) is used for regression (Liaw and Wiener 2002; Breiman et al. 2018).

2.2.2 Feasibility study of the RF model

(1) Selection of predictor variables

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

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

(2) Training sample size

One of the advantages of the RF model is that it can effectively handle small sample sizes (Biau and Scornet et al., 2016). A test was conducted to demonstrate the insensitivity of the RF model to the training sample size. The input predictor variables include geographic location and T_B (Table 2, RF2). The flowchart of the test process is shown in Fig. 4. 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 5,000 to 80,000 (with a step of 5,000) samples selected randomly from data during the period 1987-2004 were used to respectively train the RF models, and a two-year stand-alone dataset from 2005 to 2006 was applied to assess the performance of the trained models. We consider three evaluating indicators (the unbiased root mean square error (RMSE), bias and correlation coefficient) to illustrate the sensitivity of the RF model to the training sample size.

(3) Validation datasets of the fitted RF algorithms

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

2.2.3 Validation of reconstructed snow depth product and trend analysis

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

3 Results

3.1 Sensitivity to training sample size

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

3.2 Validation of the fitted RF algorithms

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

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

3.3 Validation of the reconstructed snow depth product

According to the results in Fig. 6 and Table 4, there are no notable differences in accuracy among the RF2, RF3, RF4 algorithms. In this study, we selected the RF2 algorithm to reconstruct a long-term snow depth dataset (1987 to 2018). We used the independent in situ measurements over the period 1987-2018 from validation stations (Fig. 1) to evaluate this product (hereafter, RF product). Fig. 7 shows the scatter diagrams of estimated vs. measured values for RF and WESTDC products. 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 cm, respectively (Fig. 7a and 7b). The correlation coefficient is 0.65, which is larger than the WESTDC's coefficient of 0.49. Both products particularly underestimate snow depth when snowpack is thicker than 20 cm. The error bar shows that the WESTDC product tends to more seriously underestimate snow depth than do the RF estimates.

To determine the interannual variability in the uncertainty, the time series of assessment indexes, including the unbiased RMSE, bias and correlation coefficient, are shown in Fig. 8. The results show that the RF estimates outperform the WESTDC product with respect to unbiased RMSE and correlation coefficient from season to season. The bias also fluctuates from season to season, ranging from -8 cm to 3 cm (Fig. 8c). There is a slight overestimation during the period 1987-2000, whereas it presents a notable underestimation since 2006. 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.

The assessment of snow depth product was performed in three snow cover areas of China. As shown in Fig. 10a, the RF 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 XJ for the RF product, respectively. Fig. 10b shows a notable underestimation and overestimation for the WESTDC product in northern XJ and the QTP, respectively. For the RF product, the bias is close to zero and fluctuates across a relatively narrow range in the three snow cover areas.

Based on the results in Fig. 7, we selected 20 cm as a threshold to assess the performances in deep (> 20 cm) and shallow (≤20 cm) snow cover. The percentage of shallow snow conditions to total samples was approximately 90%. Table 5 displays the comparison between RF estimates and the WESTDC product in the three snow cover areas. The ‘Samples’ row in Table 5 shows the number of samples and the corresponding percentage in each region. Both products present notable underestimation of deep snow cover, with biases of -34.1 cm and -33.8 cm on the QTP for the RF and WESTDC products, respectively. The 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 and -13.2 cm for the WESTDC data. Moreover, the correlation is very poor in deep snow cover, even negative (-0.18) in QTP for the WESTDC product. For shallow snow cover, the RF product is superior to the WESTDC estimates in QTP, with unbiased RMSEs of 3.4 cm (RF) and 5.6 cm (WESTDC). Furthermore, the WESTDC product presents overestimation in QTP, 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 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. However, the RF product tends to overestimate snow depth relative to WESTDC estimates, with higher biases of 1.8 cm and 2.5 cm than WESTDC's 0.5 cm and -0.4 cm in NE and northern XJ, respectively.

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

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

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

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

4 Discussion

4.1 Disadvantages of the RF model

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

The Earth system is interesting because it is nonstationary (especially concerning snow parameters). Generally, snow depth increases at the beginning of winter and then decreases in spring due to melting. Moreover, snow cover has different spatial patterns in various regions, such as generally deep snow in high-latitude and high-elevation areas. In China, there are five climatological snow classes according to the classification by Sturm et al. (1995). Each snow class is defined by an ensemble of snow stratigraphic characteristics, including snow density, grain size, and crystal morphology, which influences the snowpack’s microwave signature (Sturm et al., 2010). These dynamic properties of snow will lead to many cases in which the same satellite T_B corresponds to different snow depths, while the same snow depth is associated with various T_B observations, rendering the fitted RF algorithm suboptimal. 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.

4.2 Influence of predictor variables on the RF model

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

4.3 Potential errors of the reconstructed snow depth

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

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 (< 5 cm), 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/m^3) 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). Second, frozen soil reduces the accuracy of estimates. Both snow and frozen ground are volume-scattering materials, and they have similar microwave radiation characteristics, making them difficult to distinguish. Third, a limiting factor in estimating snow depth for PMW remote sensing is the presence of liquid water. In this study, a snow cover detection method is used to filter out wet snow cover; however, there are still misclassification errors, especially at the end of the winter season (Grody and Basist., 1996; Liu et al., 2018). In such cases, satellite observations are mainly associated with the emissions from the wet surface of the snowpack. Therefore, in wet snow conditions, snow depth retrieval is not possible (Derksen et al., 2010; Tedesco et al., 2016).

5 Conclusions

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

In this study, the fitted RF2 algorithm was used to retrieve a consistent 32-year daily snow depth dataset from 1987 to 2018. Then, an evaluation was carried out using independent reference data from the validation stations during the period 1987-2018. The overall unbiased RMSE and bias were 7.1 cm and -0.05 cm, respectively, outperforming the WESTDC 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 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 were superior to the WESTDC product but still presented an underestimation for deep snow (> 20 cm), with biases of -10.4 cm and -8.9 cm, respectively.

Three long-term (1987-2018) datasets, including ground truth observations, RF estimates and WESTDC product, were applied to analyze the trends of snow depth variation in China. The results suggested that there existed different trends among the three datasets. The overall trend of snow depth in China presented a significant increasing based on the ground truth observations, with a correlation coefficient of 0.57. Moreover, the trend in NE was highly consistent with the overall trend in China, with a correlation coefficient of 0.64. Neither the WESTDC nor the RF product presented significant trends except in QTP. The WESTDC product showed 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.

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

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

Competing interests. The authors declare that they have no conflicts of interest.

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

Code and Data availability. Satellite passive microwave measurements are available for download from https://nsidc.org/ (last access: 21 March 2020). The dataset of daily station snow depth from the China Meteorological Administration (CMA) can be accessed by scientific researchers through the submission of an application (http://data.cma.cn/en, last access: 21 January 2020). The land use/land cover (LULC) map was provided by the Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (http://www.resdc.cn/, last access: 21 May 2019). The WESTDC snow depth product was obtained from the Environmental and Ecological Science Data Center for West China (http://data.casnw.net/portal/, last access: 21 March 2020). The RF snow depth product retrieved in this paper is available to the public (https://figshare.com/articles/RF_based_Longterm_SnowDepth_China_rar/11988027, DOI: 10.6084/m9.figshare.11988027). The RF model code is available in https://CRAN.R-project.org/package=randomForest (version 4.6-14, last access: 28 March 2018).

References

Armstrong, R., Knowles, K., Brodzik, M., and Hardman, M.: DMSP SSM/I-SSMIS Pathfinder Daily EASE-Grid Brightness Temperatures, Version 2. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center, 10.5067/3EX2U1DV3434, 1994.

Bair, E. H., Abreu Calfa, A., Rittger, K., and Dozier, J.: Using machine learning for real-time estimates of snow water equivalent in the watersheds of Afghanistan, The Cryosphere, 12, 1579-1594, 10.5194/tc-12-1579-2018, 2018.

Basang, D., Barthel, K., Olseth, JA.: Satellite and Ground Observations of Snow Cover in Tibet during 2001–2015, Remote Sensing, 9,1201,10.3390/rs9111201, 2017.

Belgiu, M., and Lucian, D.: Random forest in remote sensing: A review of applications and future directions, ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24-31. 10.1016/j.isprsjprs.2016.01.011, 2016.

Biau, G.Ã.Š. and Scornet, E.: A random forest guided tour, TEST, 25, 197–227, 10.1007%2Fs11749-016-0481-7, 2016.

Bormann, K.J., Brown, R.D., Derksen, C., Painter, T.H.: Estimating snow-cover trends from space, Nat. Clim. Chang, 8, 924–928, 2018.

Breiman, L., Cutler, A., Liaw, A., Wiener, M.: randomForest: Breiman and Cutler's Random Forests for Classification and Regression, R package version 4.6-14, 2018. https://CRAN.R-project.org/package=randomForest.

Breiman, L. Random forests. Mach. Learn. 2001, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

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.

Cai, S., Li, D., Durand, M., and Margulis, S.: Examination of the impacts of vegetation on the correlation between snow water equivalent and passive microwave brightness temperature, Remote Sensing of Environment, 193, 244–256, 10.1016/j.rse.2017.03.006, 2017.

Canovas-Garcia, F., Alonso-Sarria, F., Gomariz-Castillo, F., and Onate-Valdivieso, F.: Modification of the random forest algorithm to avoid statistical dependence problems when classifying remote sensing imagery, Comput. Geosci, 103, 1–11, 10.1016/j.cageo.2017.02.012, 2017.

Chang, A., Foster J., Hall D.: Nimbus-7 derived global snow cover parameters, Annals of Glaciology, 9, 39-44, 10.1017/S0260305500000355, 1987.

Che, T., Dai, L., Zheng, X., Li, X., Zhao, K.: Estimation of snow depth from passive microwave brightness temperature data in forest regions of northeast China, Remote Sensing of Environment, 183, 334–349, 10.1016/j.rse.2016.06.005, 2016.

Che, T., Li, X., Jin, R., Armstrong, R., and Zhang, T.: Snow depth derived from passive microwave remote-sensing data in China, Annals of Glaciology, 49,145-154,10.3189/172756408787814690, 2008.

Che, T., Li, X., Jin, R., and Huang, C.: Assimilating passive microwave remote sensing data into a land surface model to improve the estimation of snow depth, Remote Sensing of Environment, 143, 54-63,10.1016/j.rse.2013.12.009, 2014.

Dahri, Z., Moors, E., Ludwig, F., Ahmad, S., Khan, A., Ali, I., Kabat, P.: Adjustment of measurement errors to reconcile precipitation distribution in the high-altitude Indus basin, Int J Climatol, 38, 1–19, 10.1002/joc.5539, 2018.

Dai, L., Che, T., Ding, Y., and Hao, X.: Evaluation of snow cover and snow depth on the Qinghai–Tibetan Plateau derived from passive microwave remote sensing, The Cryosphere, 11, 1933–1948, 10.5194/tc-11-1933-2017, 2017.

Dai, L., Che, T., Xie, H., and Wu, X.: Estimation of Snow Depth over the Qinghai-Tibetan Plateau Based on AMSR-E and MODIS Data, Remote Sensing, 10, 1989, 10.3390/rs10121989, 2018.

Davenport, I., Sandells, M., and Gurney, R.: The effects of variation in snow properties on passive microwave snow mass estimation, Remote Sensing of Environment, 118, 168–175, 10.1016/j.rse.2011.11.014, 2012.

Derksen, C., Walker, A., and Goodison, B.: Evaluation of passive microwave snow water equivalent retrievals across the boreal forest/tundra transition of western Canada, Remote Sensing of Environment, 96, 315-327, 10.1016/j.rse.2005.02.014, 2005.

Derksen, C., Toose, P., Rees, A., Wang, L., English, M., Walker, A., and Sturm, M.: Development of a tundra-specific snow water equivalent retrieval algorithm for satellite passive microwave data, Remote Sensing of Environment, 114, 1699–1709, 10.1016/j.rse.2010.02.019, 2010.

Derksen. C., and Brown, R.: Spring snow cover extent reductions in the 2008–2012 period exceeding climate model projections, Geophysical Research Letters, 39, 1-6, 10.1029/2012GL053387, 2012.

Dozier, J., Bair, E. H., and Davis, R. E.: Estimating the spatial distribution of snow water equivalent in the world's mountains, WIREs Water, 3, 461-474, doi 10.1002/wat2.1140, 2016.

Dorji, T., Hopping, K., Wang, S., Piao, S., Tarchen, T., and Klein, J.: Grazing and spring snow counteract the effects of warming on an alpine plant community in Tibet through effects on the dominant species, Agric. For. Meteorol, 263, 188–197, 10.1016/j.agrformet.2018.08.017, 2018.

Durand, M., and Margulis, S.: Feasibility test of multifrequency radiometric data assimulation to estimate snow water equivalent, Journal of Hydrometeorology, 7, 443-457, 10.1175/jhm502.1, 2006.

Durand, M., Kim, E., and Margulis, S.: Quantifying uncertainty in modeling snow microwave radiance for a mountain snowpack at the point-scale, including stratigraphic effects, IEEE Trans. Geosci. Remote Sens, 46, 1753–1767, 10.1109/tgrs.2008.916221, 2008.

Fernandes, R., Zhao, H., Wang, X., Key, J., Qu, X., and Hall, A.: Controls on Northern Hemisphere snow albedo feedback quantified using satellite Earth observations, Geophys. Res. Lett, 36, 1–6, 10.1029/2009gl040057, 2009.

Foster, J., Chang, A., Hall D.: Comparison of Snow Mass Estimation From a Prototype Passive Microwave Snow Algorithm, a Revised Algorithm and Snow Depth Climotology, Remote Sensing of Environment, 62, 132–142, 10.1016/S0034-4257(97)00085-0, 1997.

Foster, J., Hall, D., Eylander, J., Riggs, G., Nghiem, S., Tedesco, M., Kim, E., Montesano, P., Kelly, R., Casey, K., and Choudhury, B.: A blended global snow product using visible, passive microwave and scatterometer satellite data, International Journal of Remote Sensing, 32, 41 1371-1395, 10.1080/01431160903548013, 2011.

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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.

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

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

Li, Q., Kelly, R.: Correcting Satellite Passive Microwave Brightness Temperatures in Forested Landscapes Using Satellite Visible Reflectance Estimates of Forest Transmissivity, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10, 3874-3883, 10.1109/JSTARS.2017.2707545, 2017.

Liaw, A., and Wiener, M.: Classification and regression by randomForest, R News, 2, 18–22, 2002.

Liu, X., Jiang, L., Wu, S., Hao, S., Wang, G., and Yang, J.: Assessment of Methods for Passive Microwave Snow Cover Mapping Using FY-3C/MWRI Data in China, Remote Sensing, 10, 524-539, 10.3390/rs10040524, 2018.

Liu, X., Jiang, L., Wang, G., Hao, S., and Chen, Z.: Using a Linear Unmixing Method to Improve Passive Microwave Snow Depth Retrievals, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens, 11, 4414–4429, 10.1109/PIERS.2016.7735542, 2018.

Maxwell, A., Warner, T., and Fang, F.: Implementation of machine-learning classification in remote sensing:

An applied review, Int. J. Remote Sens, 39, 2784–2817, 2018.

Mann, H. B.: Nonparametric tests against trend, Econometrica 13, 245–259, 1945.

Milan, G., and Slavisa, T.: Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia, Global Planet Change, 100, 172-182, 10.1016/j.gloplacha.2012.10.014, 2013.

Meløysund, V., Bernt, L., Karl, V., and Kim R.: Predicting snow density using meteorological data, Meteorological Applications, 14, 413–423, 10.1002/met.40, 2007.

Metsämäki, S., Pulliainen, J., Salminen, M., Luojus, K., Wiesmann, A., Solberg, R., Böttcher, K., Hiltunen, M., and Ripper, E.: Introduction to GlobSnow Snow Extent products with considerations for accuracy assessment, Remote Sensing of Environment, 156, 96–108, 10.1016/j.rse.2014.09.018, 2015.

Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M., Papritz, A.: Evaluation of digital soil mapping approaches with large sets of environmental covariates, Soil, 4, 1, 10.5194/soil-4-1-2018, 2018.

Orsolini, Y., Wegmann, M., Dutra, E., Liu, B., Balsamo, G., Yang, K., de Rosnay, P., Zhu, C., Wang, W., Senan, R., and Arduini, G.: Evaluation of snow depth and snow cover over the Tibetan Plateau in global reanalyses using in situ and satellite remote sensing observations, The Cryosphere, 13, 2221–2239, 10.5194/tc-13-2221-2019, 2019.

Pan, J., Durand, M., Vander Jaqt, B., and Liu, D.: Application of a Markov Chain Monte Carlo algorithm for snow water equivalent retrieval from passive microwave measurements, Remote Sensing of Environment, 192, 150-165, 10.1016/j.rse.2017.02.006, 2017.

Picard, G.: Simulation of the microwave emission of multi-layered snowpacks using the dense media radiative transfer theory: The DMRT-ML model, Geosci. Model Develop. Discuss., 6, 3647–3694, 2012.

Prasad, A., Iverson, L., and Liaw, A.: Newer classification and regression tree techniques: bagging and random forests for ecological prediction, Ecosystems, 9, 181-199, 10.1007/s10021-005-0054-1, 2006.

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, 2018.

Pulliainen, J., Grandell, J., and Hallikainen, M.: HUT snow emission model and its applicability to snow water equivalent retrieval, IEEE Trans. Geosci. Remote Sens, 37, 1378–1390,10.1109/36.763302, 1999.

Pulliainen, J.: Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations, Remote Sens. Environ, 101, 257–269, 10.1016/j.rse.2006.01.002, 2006.

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 Microwave Soil Moisture Product Using Random Forest Adopting AMSR-E/AMSR2 Brightness Temperature and SMAP over the Qinghai–Tibet Plateau, China, Remote Sensing, 11, 683, 10.3390/rs11060683, 2019.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat.: Deep learning and process understanding for data-driven Earth system science, Nature 566, 195–204, 2019.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

List of Tables and Figures

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 D: 18:20	A: 17:17 D: 05:17	A: 17:58 D: 05:58	A: 17:31 D: 05:31
Frequency & footprint (GHz): (km × km)		19.35: 45×68 23.235: 40×60 37: 24×36 85.5: 11×16		19.35: 42×70 23.235: 42×70 37: 28×44 91.655: 13×15

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}	snow depth	land cover types: grassland, cropland, bare land, shrubland, forest
RF2	T_{B19V}, T_{B37V}, Latitude, Longitude		
RF3	T_{B19V}, T_{B37V}, Latitude, Longitude, Elevation		
RF4	T_{B19V}, T_{B37V}, Latitude, Longitude, Elevation, Land cover fraction		

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

Name	Test1 (OOB)		Test2 (temporal subset)		Test3 (spatiotemporal 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

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

Name	Test1 (OOB)			Test2 (temporal subset)			Test3 (spatiotemporal 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

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	
Snow Depth (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	
Snow Depth (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%)

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

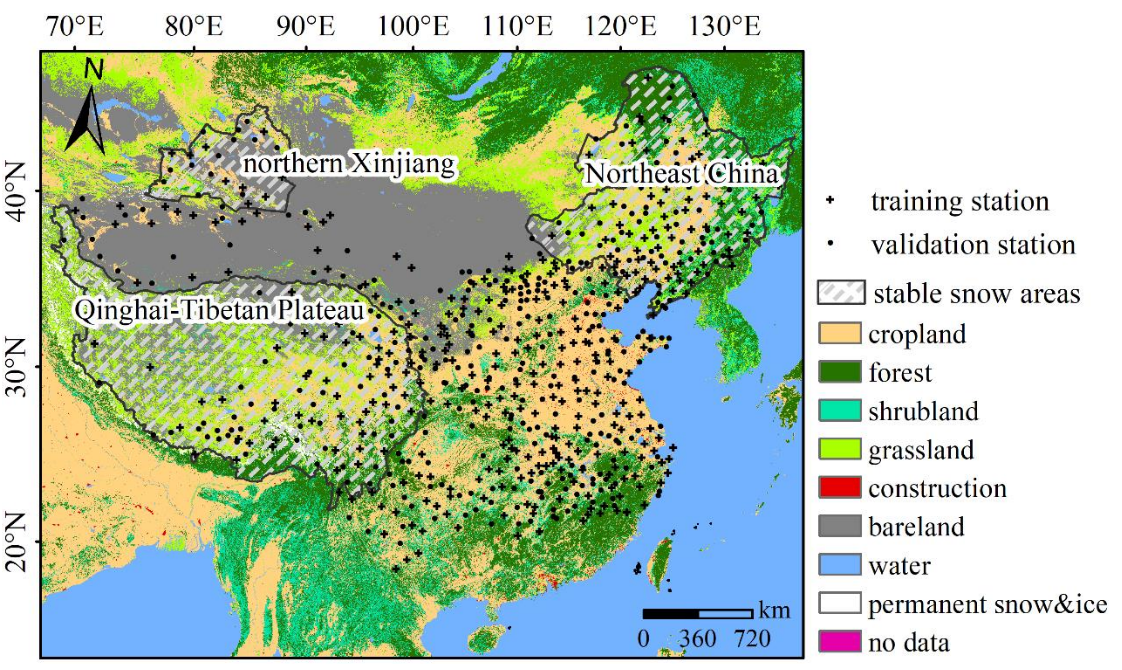

Figure 1. Spatial distribution of the weather stations and land cover types in the study area. There are three stable snow cover areas in China: Northeast China (NE), northern Xinjiang (XJ) and the Qinghai-Tibetan Plateau (QTP).

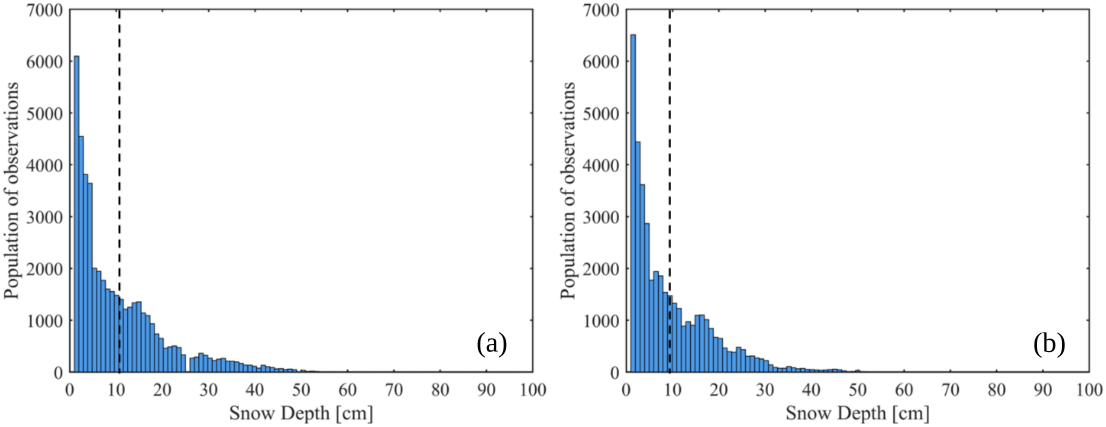

Figure 2. Histograms of snow depth observations from (a) training and (b) validation stations. The average values (black dashed lines) are equal to 10.5 cm and 9.8 cm, respectively.

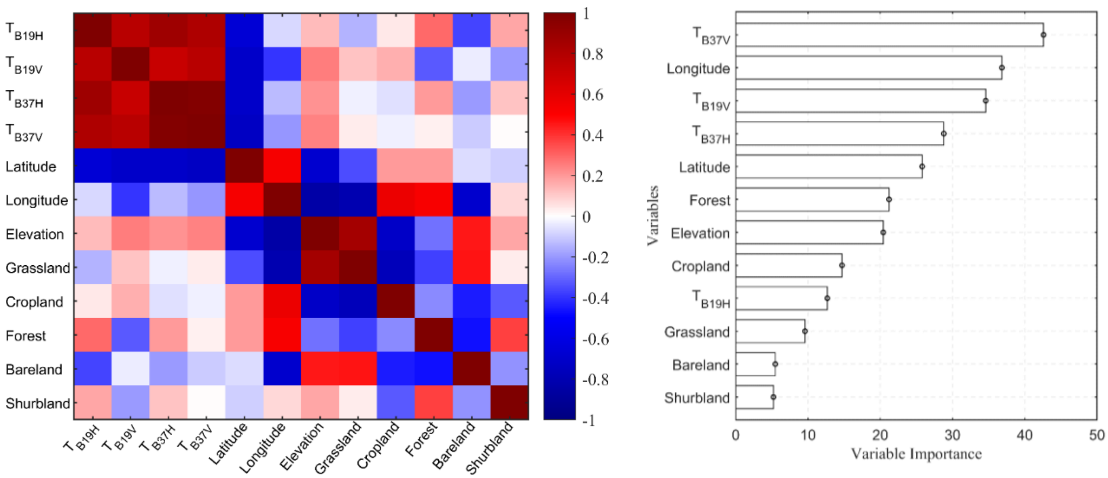

Figure 3. Correlations between the predictor variables (left) and the ranking of variable importance (right). The importance of variables, referred to as Mean Decrease Accuracy (MDA) in the RF model, is obtained by averaging the difference in out-of-bag error estimation before and after the permutation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

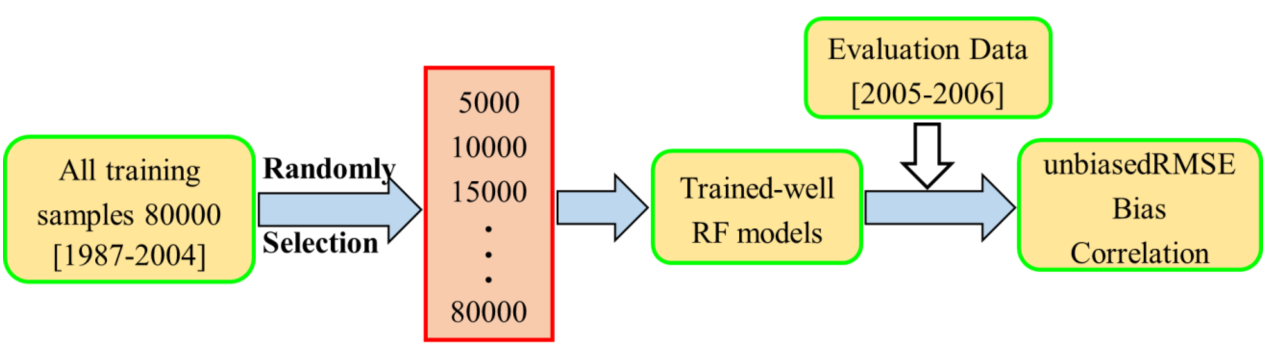

Figure 4. The test process flowchart for the sensitivity of the RF model to the training sample size.

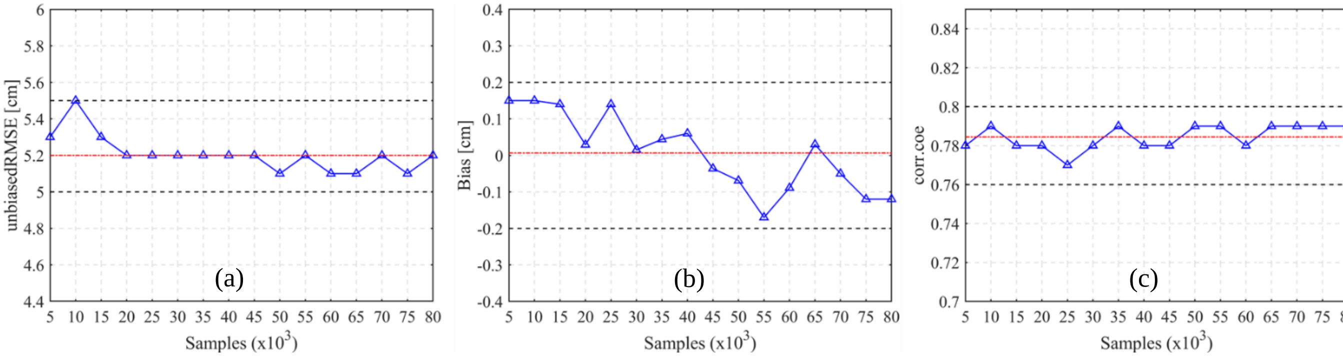

Figure 5. Trends of (a) unbiased RMSE, (b) bias and (c) correlation coefficient with increasing training sample size.

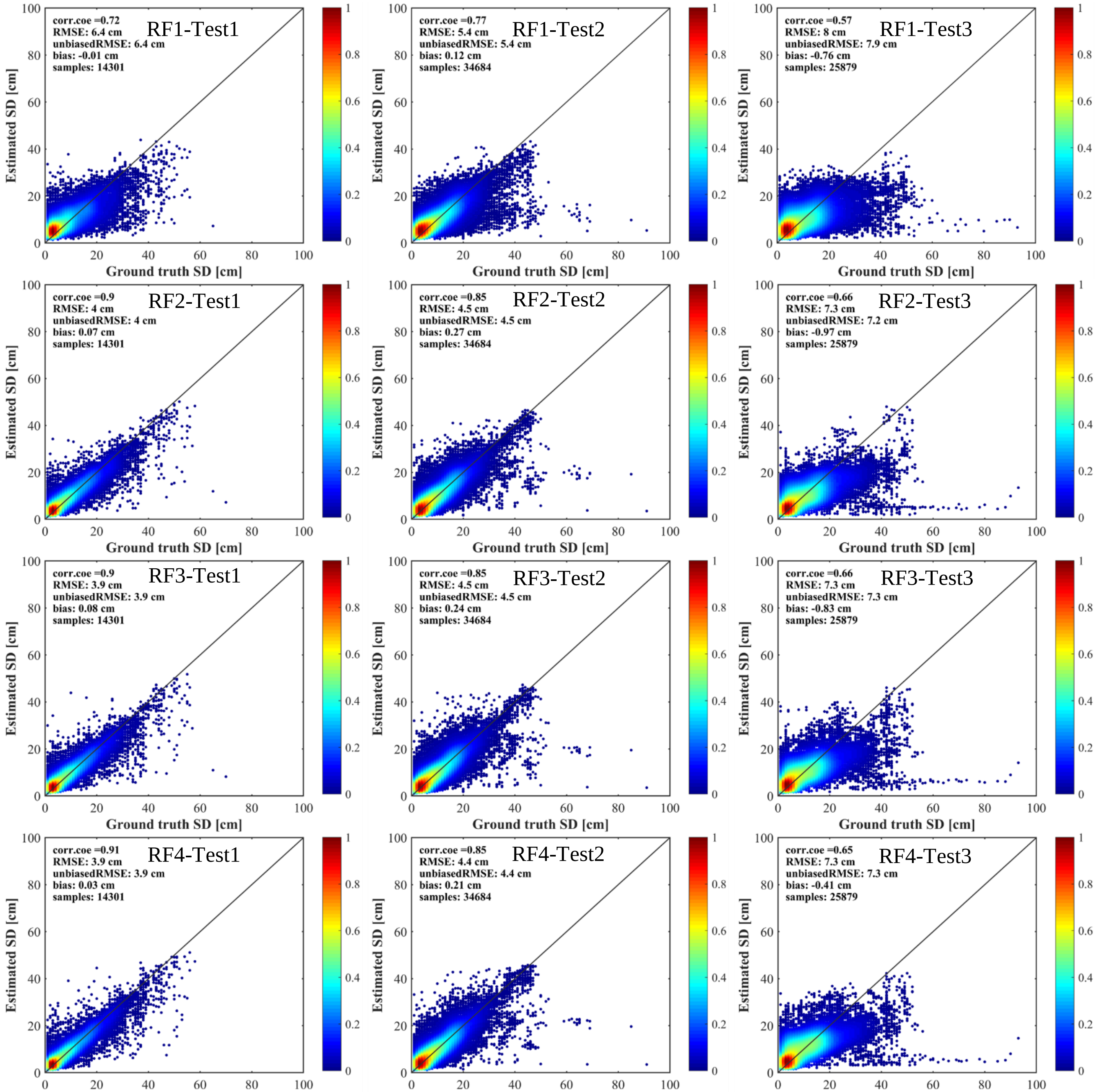

Figure 6. The color-density scatterplots of the estimated snow depth with four fitted RF algorithms and the ground truth snow depth. The four trained RF algorithms (RF1, RF2, RF3, RF4) were evaluated with three validation datasets (Test1, Test2, Test3).

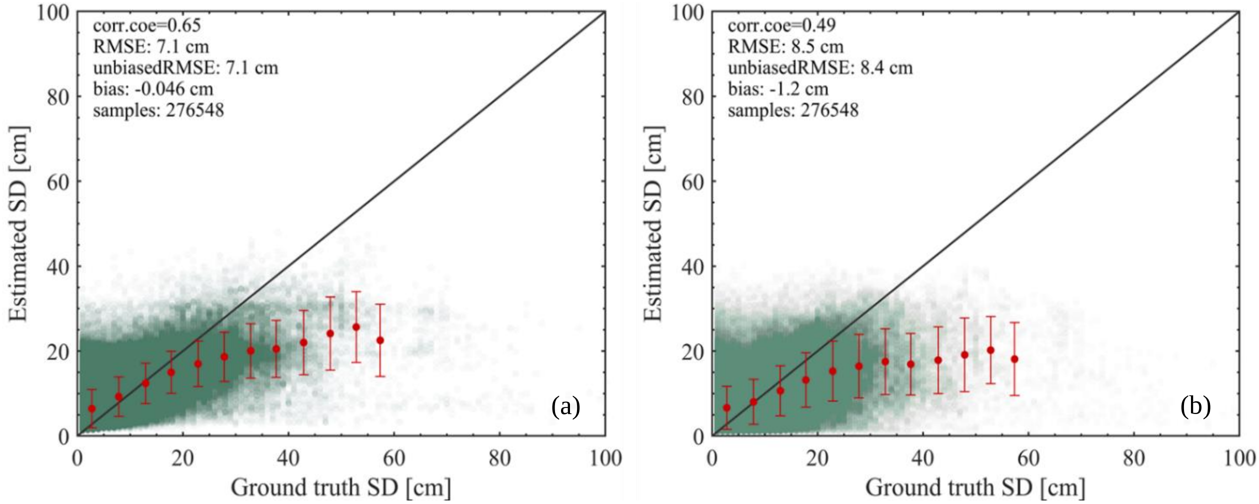

Figure 7. Scatterplots of the estimated snow depth and the ground truth observation for (a) RF and (b) WESTDC products.

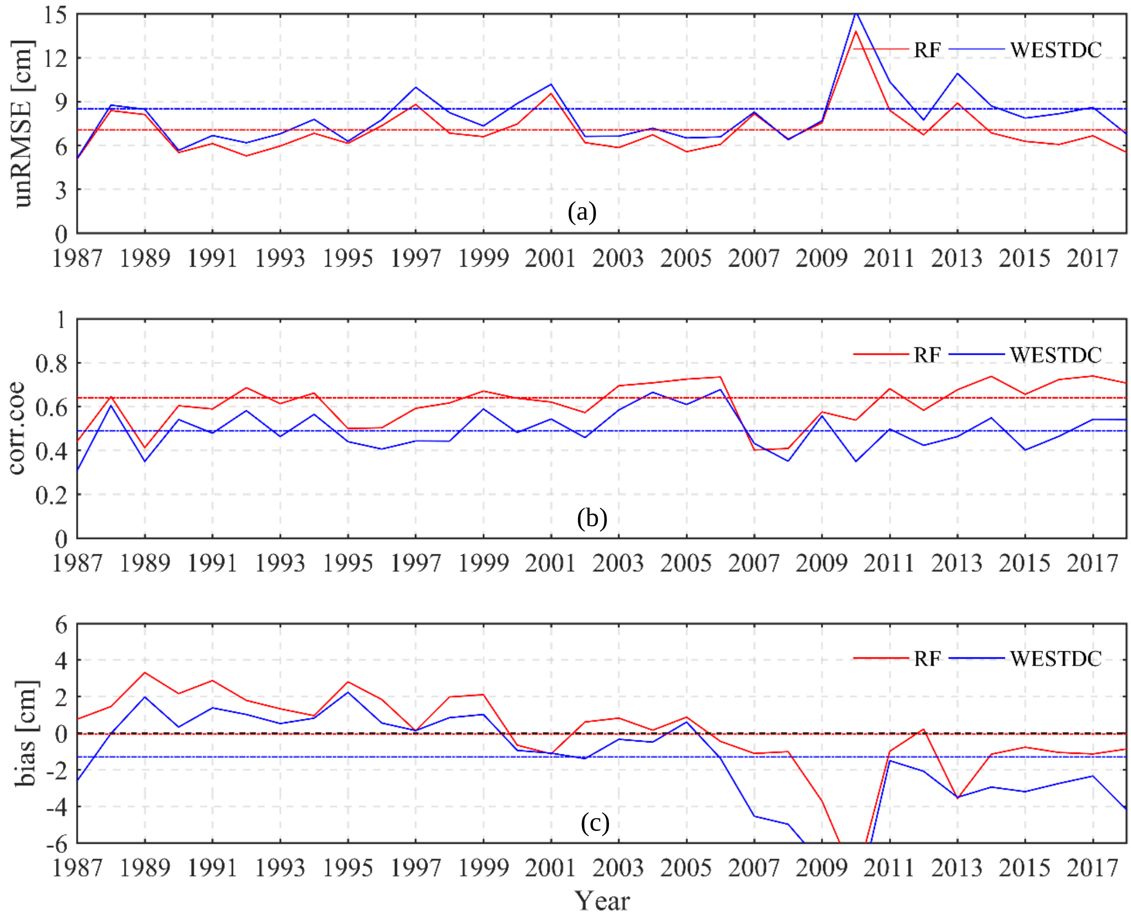

Figure 8. Time series of (a) unbiased RMSE (unRMSE), (b) correlation coefficient (corr.coe) and (c) bias for RF and WESTDC products. The colorful dashed lines represent mean values of assessment indexes.

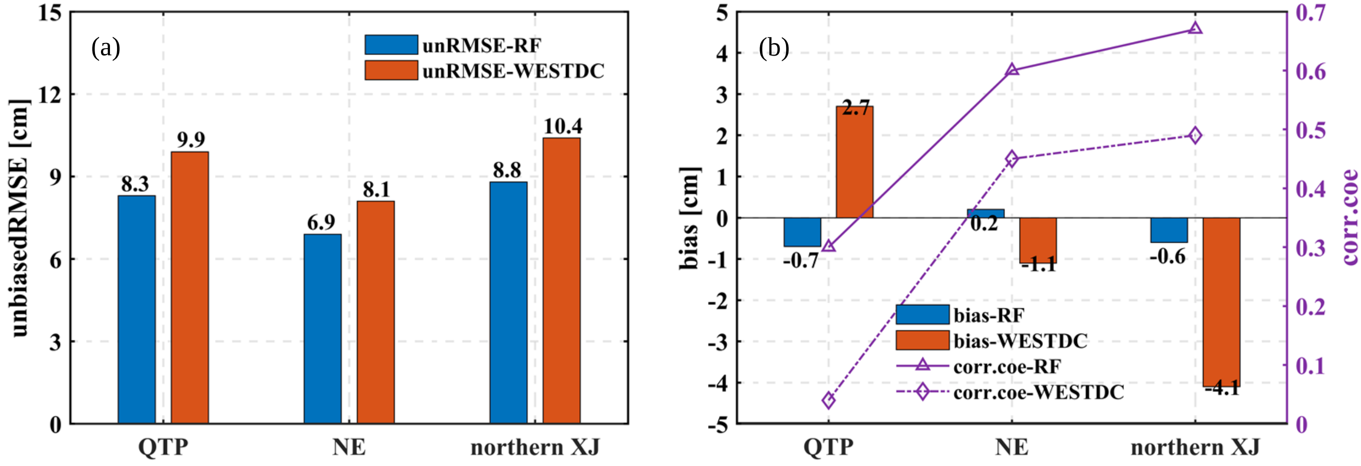

Figure 9. The validation of RF and WESTDC snow depth products in three stable snow cover areas over China with respect to (a) the unbiased RMSE, (b) bias and correlation coefficient.

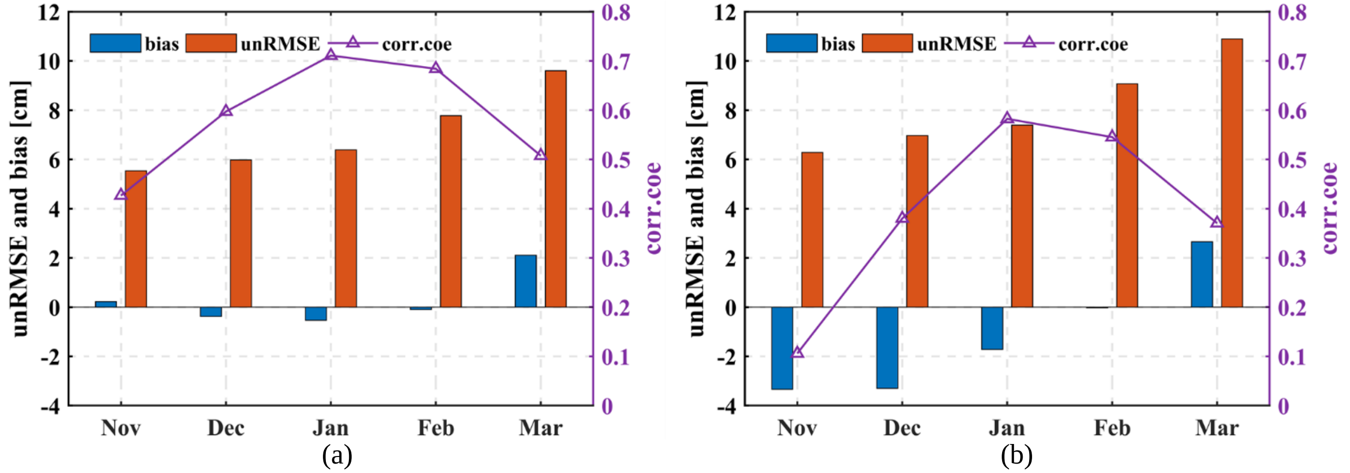

Figure 10. Monthly performances of (a) RF, and (b) WESTDC snow depth products. Nov: November; Dec: December; Jan: January; Feb: February; Mar: March.

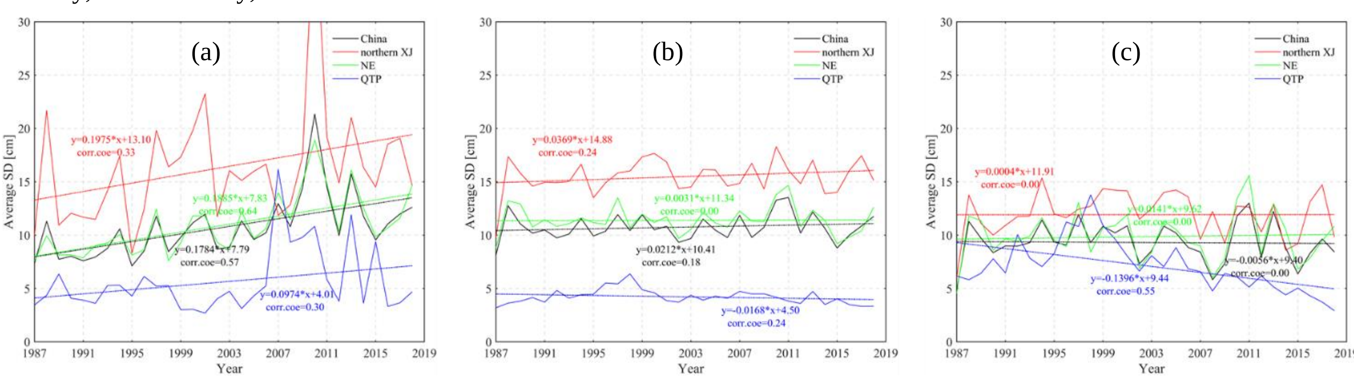

Figure 11. Trend analysis of snow depth based on (a) station observations, (b) RF estimates, and (c) WESTDC product in three stable snow cover areas of China. The correlation is statistically significant at the 0.05 level.

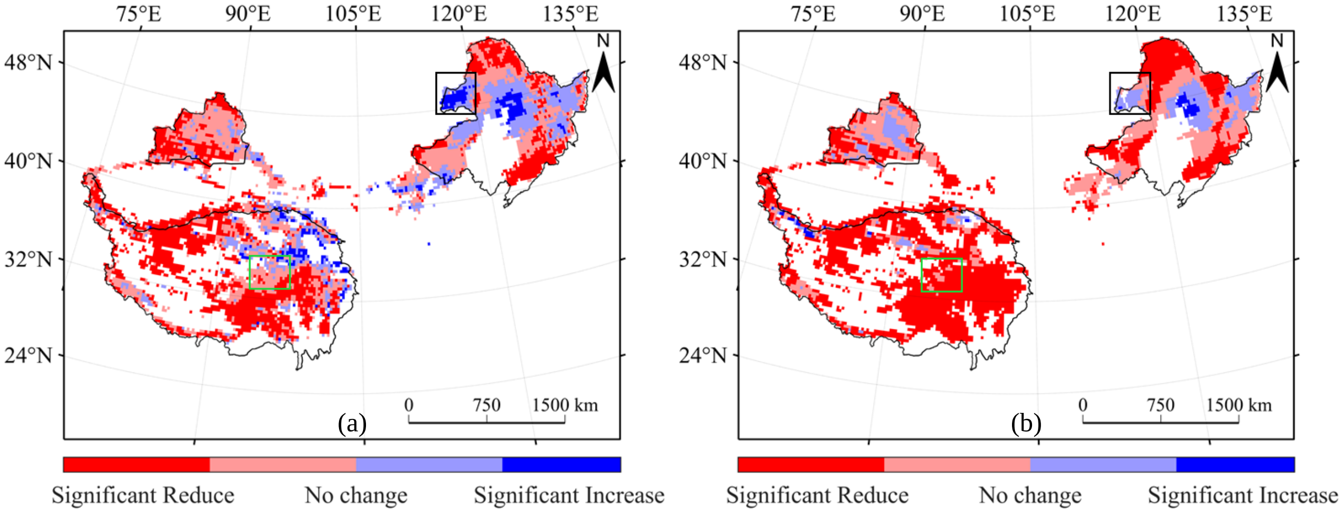

Figure 12. Trend analysis of snow depth during the period 1987-2018: (a) RF product; (b) WESTDC data. Light red and light blue represent no significant trend changes.

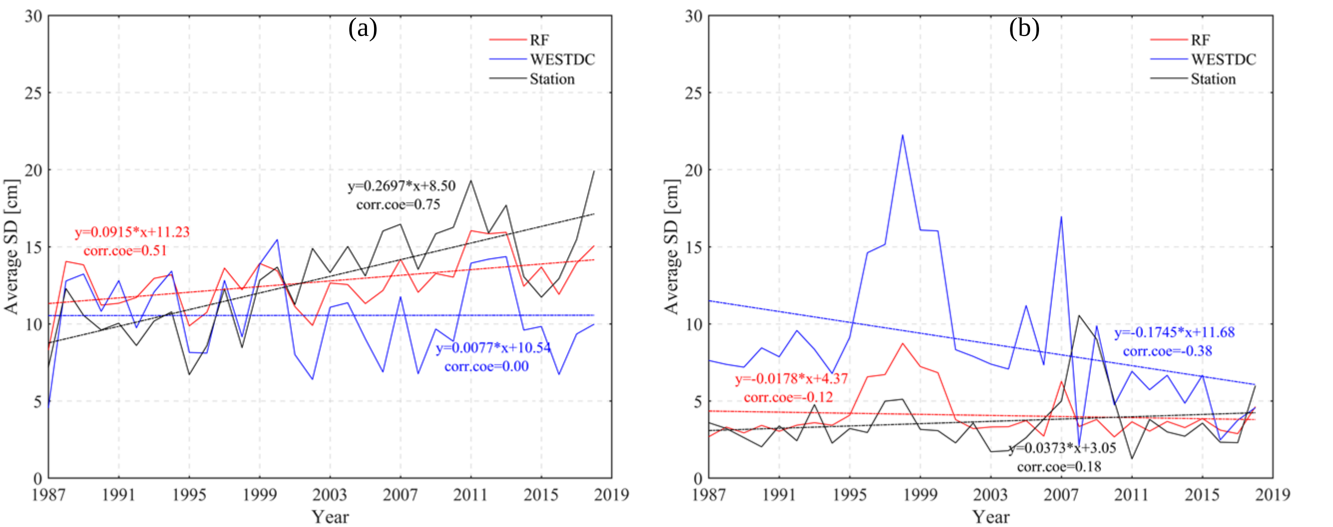

Figure 13. Comparison of changing trends of snow depth between RF estimates and WESTDC product in specific areas of (a) NE and (b) QTP.