https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

Open Access Earth System Science Data Discussions

A long-term daily gridded snow depth dataset for the Northern Hemisphere from 1980 to 2019 based on machine learning

Yanxing Hu[1,3], Tao Che[1,2], Liyun Dai[1], Yu Zhu[4], Lin Xiao[5], Jie Deng[1] and Xin Li[2,6]

[1]Key Laboratory of Remote Sensing of Gansu Province, Heihe Remote Sensing Experimental Research Station, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Science, Lanzhou 730000, China
[2]Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Science, Beijing 100101, China
[3]University of Chinese Academy of Science, Beijing 100049, China
[4]Yunnan Key Laboratory of International Rivers and Transboundary Eco-Security, Institute of International Rivers and Eco-Security, Yunnan University, Kunming 650091, China
[5]National Forestry and Grassland Administration Key Laboratory of Forest Resource Conservation and Ecological Safety on the Upper Reaches of the Yangtze River, Sichuan Province Key Laboratory of Ecological Forestry Engineering on the Upper Reaches of the Yangtze River, College of Forestry, Sichuan Agricultural University, Chengdu 611130, China
[6]National Tibetan Plateau Data Center, State Key Laboratory of Tibetan Plateau Earth System, Resources and Environment, Institute of Tibetan Plateau Research, Chinese Academy of Science, Beijing 100101, China

Correspondence: Tao Che (chetao@lzb.ac.cn)

Abstract. A high-quality snow depth product is very import for cryospheric science and its related disciplines. Current long time-series snow depth products covering the Northern Hemisphere can be divided into two categories: remote sensing snow depth product and reanalysis snow depth products. However, existing gridded snow depth products have some shortcomings. Remote sensing-derived snow depth products are temporally and spatially discontinuous and tend to underestimate snow depth, while reanalysis snow depth products have coarse spatial resolutions and great uncertainties. To overcome these problems, in our previous work we proposed a novel data fusion framework based on Random Forest Regression of snow products from Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), Advanced Microwave Scanning Radiometer 2 (AMSR-2), Global Snow Monitoring for Climate Research (GlobSnow), the Northern Hemisphere Snow Depth (NHSD), ERA-Interim, and Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2), incorporated geolocation (latitude and longitude), and topographic data (elevation), which were used as input independent variables. More than 30,000 ground observation sites were used as the dependent variable to train and validate the model in different time period. This fusion framework resulted in a long time series of continuous daily snow depth product over the Northern Hemisphere with a spatial resolution of 0.25°. Here we compared the fused snow depth and the original gridded snow depth products with 13,272 observation sites, showing an improved precision of our product. The evaluation indexes of the fused (best original) dataset yielded a coefficient of determination R^2 of 0.81 (0.23), Root Mean Squared Error (RMSE) of 7.69 (15.86) cm, and Mean Absolute Error (MAE) of 2.74 (6.14) cm. Most of the bias (88.31%) between the fused snow depth and in situ observations was distributed from -5 cm to 5 cm depths. The accuracy assessment of independent snow observation sites – Sodankylä (SOD), Old Aspen (OAS), Old Black Spruce (OBS), and Old Jack Pine (OJP) – showed that the fused snow depth dataset had high precision under snow depths of less than 100 cm with a relatively

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

homogeneous surrounding environment. In the altitude range of 100 m to 2000 m, the fused snow depth had a higher precision, with R^2 varying from 0.73 to 0.86. The fused snow depth had a decreasing trend based on the spatiotemporal analysis and Mann-Kendall trend test method. This fused snow depth product provides the basis for understanding the temporal and spatial characteristics of snow cover and their relation to climate change, hydrological and water cycle, water resource management, ecological environment and snow disaster and hazard prevention. The new fused snow depth dataset is freely available from the National Plateau Data Center (TPDC) and can be downloaded at https://dx.doi.org/10.11888/Snow.tpdc.271701 (Che et al., 2021). This snow depth also can be downloaded at https://zenodo.org/record/6336866#.Yjs0CMjjwzY.

1 Introduction

Seasonal snow is a fundamental component of the global energy and water cycles (Barnett et al., 2005; Bormann et al., 2018). Snowpack is highly vulnerable to climate change, but although snow extent is monitored routinely, the fundamental properties of snow depth is measured remarkably poorly (Pritchard., 2021). The IPCC Special Report on Oceans and Cryosphere in a Changing Climate (SROCC) highlighted key strengths and weaknesses in our understanding of the cryosphere and mass and spatial-temporal variability trends in snow depth over the Northern Hemisphere during the 20th century that have still not been accurately assessed (IPCC, 2019). In contrast to snow cover, reliable quantitative knowledge on snow depth and its trend is lacking (Bormann et al., 2018; Pulliainen et al., 2020). Snow depth is more challenging to monitor than snow cover extent at both global (Zhu et al., 2021) and regional scales because of limited surface observations along with inadequate measurements and methods for retrieval from space (Bormann et al., 2018). Hence, confidence in snow depth trends is much lower than that of snow cover extent.

Currently, snow depth data can be obtained from manual observations, passive microwave brightness temperature, synthetic aperture radars (Lievens et al., 2019), and reanalysis datasets. In situ observations consist of meteorological station data, snow survey data and automatic snow depth observation equipment set up in the field. These datasets, while accurate, are based on single-point observation. Snow surveys are time-consuming and laborious. Some automatic measurement networks provide snow depth information, but are mostly located in nearly flat terrains and seldom cover the highest elevations (Dozier et al., 2016). Snow depth data retrieved based on passive microwave brightness can reveal the spatiotemporal pattern of large areas. Hereafter, these datasets are referred to as remote sensing snow depth products. At present, frequently used remote sensing snow depth products are from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), the Advanced Microwave Scanning Radiometer-2 (AMSR-2), which succeeded AMSR-E, a long time series of the Northern Hemisphere snow depth (NHSD) and the Global Snow Monitoring for Climate Research (GlobSnow). The accuracy of AMSR-E and AMSR-2 is higher in North America than in Eurasia (Xiao et al., 2020). The accuracy of the NHSD snow depth product is high in China but low in Europe, America and Canada, where snow tends to be deep (Hu et al., 2021). GlobSnow incorporated some ground observations and has been recognized as the most accurate remote sensing

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

snow depth product (Larue, et al., 2017; Pulliainen et al., 2020). But the GlobSnow product excludes mountains and the land region north of 35° N. Remote sensing snow depth products tend to saturate in deep snow areas (Xiao et al., 2020). Similarly, reanalysis snow depth products are susceptive to various structural limitations, and some errors caused by uncertainties in the input climate variables mean state (Parker., 2016). Another disadvantage of these reanalysis snow depth products is their low spatial resolution, unsuitable for snow eco-hydrological applications (Dee et al., 2011). These reanalysis snow depth products also include uncertainties caused by the biases in the forcing datasets (Broxton et al., 2016). Additionally, several satellite-derived products do not cover the high latitudes, and atmospheric reanalyses are often inconsistent across these regions (Thackeray et al., 2019). However, some reanalysis snow depth products systematically overestimate snow depth (Mortimer et al., 2020; Xiao et al., 2020). Instead of developing a single "best" snow depth product, an observational ensemble composed of multiple products can provide important information to reduce observational uncertainty. Despite the ongoing challenges, we must continue to assess, intercompare and combine multiple snow depth datasets from independent sources to mitigate uncertainties related to individual products.

Machine learning has emerged as a promising alternative means to solve complex nonlinear problems in the field of geography (Reichstein et al., 2019; Yuan et al., 2020). Machine learning also has incomparable advantages in processing geographical data in the era of big data (Zhang et al., 2019). Some machine learning algorithms have been used to retrieve snow depth (Tedesco et al., 2004) as the performance of Artificial Neural Networks (ANN) has been found to surpass that of traditional retrieval algorithms (Cao et al., 2008; Evora et al., 2008). More advanced machine learning algorithms have been introduced in recent years to retrieve snow depth over large scales. Support Vector Machine (SVM), Random Forest Regression (RFR), and Conventional Neural Networks (CNN) were employed to retrieve snow depth in northern Xinjiang (Liang et al., 2015), Alaska (Wang et al., 2020), China (Yang et al., 2020), and the Northern Hemisphere (Xiao et al., 2018). These results showed that the accuracies achieved by machine learning methods were higher than those from the traditional Spectral Polarization Difference (SPD) algorithm (Chang et al., 1987) and improved retrieval algorithms (Che et al., 2008). Machine learning methods showed great potential for producing long time series of large-scale snow depth datasets (Hu et al., 2021). Hu et al. (2021) innovatively introduced RFR to fuse some gridded snow depth products.

The fused results indicated that machine learning methods could be used to rapidly produce a long times series snow depth product over large areas with high accuracy (Hu et al., 2021). Mudrky et al. (2015) carried out cross-comparisons of snow depth products both globally and regionally and found that they differed by more than 50% (Mudryk et al., 2015). The multiple snow depth products assessment and intercomparisons have typically focused on identifying the best dataset (Mortimer et al., 2020; Xiao et al., 2020). Mortimer et al. (2020) evaluated nine gridded snow depth products over the Northern Hemisphere by using snow course data in Russia, Finland, and Canada. The results indicated that stand-alone passive snow depth products exhibit low spatial and temporal correlations to other products, and GlobSnow snow depth product performed comparably to the reanalysis-based products. Previous assessments of the Northern Hemisphere snow depth products showed GlobSnow, ERA-Interim, and MERRA-2 snow depth products perform better in plain and forest regions. AMSR-E and AMSR-2 showed satisfying performance in shallow snow regions (0~10 cm), while MERRA-2

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

agreed better when snow depth exceeded 50 cm (Xiao et al., 2020). The comprehensive comparison of snow depth products from reanalysis and remote sensing over the Northern Hemisphere has highlighted numerous uncertainties (Mudryk et al., 2015; Mortimer et al., 2020; Xiao et al., 2020). The assessment results showed no clear advantage to using a single type of snow depth product, and even a multi-dataset mean exhibited a great improvement in precision (Mortimer et al., 2020; Mudryk et al., 2015). Only through combined and integrated improvements in remote sensing, modelling, and in situ observations will real progress in snow depth product development be achieved and sustained (Mortimer et al., 2020; Snauffer et al., 2018). Snauffer et al. (2018) directly incorporated the snow depth products and auxiliary information as input variables of the ANN model to obtain a more precise product. This ANN framework was found to have a lower MAE of 47% to 60% than the original individual products, and the fused results indicated that the machine learning algorithm improved the fusion of multi-source datasets. Hu et al. (2021) developed a snow depth fusion algorithm with machine learning methods using several source products over the Northern Hemisphere, showing that machine learning fusion of snow depth products increased estimation accuracy compared with the original snow depth products. These fused frameworks based on machine learning algorithms offer a new perspective for creating long time series of snow depth products over large regions. Machine learning methods not only integrate the advantages of these input snow depth products but also improve their accuracy by using a high number of in situ observations to train the sample data. In Hu et al. (2021), the candidate independent variables for the snow depth RFR model included elevation, longitude, latitude, AMSR-E, AMSR-2, NHSD, GlobSnow, ERA-Interim, and MERRA-2, and land cover type was used to segregate the data into different bins.

In this study, we propose a data fusion framework based on the RFR algorithm (Hu et al., 2021) to derive a comprehensive snow depth product for the Northern Hemisphere from 1980 to 2019. In this framework, geolocation, topographical and as many gridded snow depth products as possible were selected as the input variables. In Sect. 2, we describe the various data in this study and illustrate data pre-processing. Sec. 3 introduces the RFR algorithms, the experimental design, and the accuracy assessment indicators. Detailed comparisons between the fused snow depth dataset and in situ observations and validation works are given in Sect. 4, with subsequent analysis and explanation. Sect. 5 discusses the uncertainty and model generalization capacity and data availability; conclusions and future perspectives are summarized in Sect. 6 and Sect. 7, respectively.

2 Data description

2.1 Gridded snow depth product

The snow depth products used in this study were chosen based on three main criteria: complete the Northern Hemisphere spatial coverage (with the exception of GlobSnow version 2.0, which covers the region 35° N–85° N), continuous availability throughout the satellite era (we used 1980-2019 as our fusion period) and most frequently used by the snow research community (Table 1). Three categories of gridded snow depth products were employed: (1) stand-alone passive microwave retrievals (NHSD, AMSR-E, and AMSR-2), (2) passive microwave estimates incorporated with in situ snow

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

depth measurements (GlobSnow), and (3) snow depth products using some form of reanalysis (ERA-Interim, and MERRA-2).

Table 1. Summary of the original six gridded snow depth products used in this study.

Gridded snow depth products	AMSR-E	AMSR-2	NHSD	GlobSnow	ERA-Interim	MERRA-2
Data ownership	NASA/JAXA	NASA/JAXA	CAS	ESA	ECMWF	NASA
Spatial coverage	0 – 90° N	0 – 90° N	0 – 90° N	0 – 90° N	0 – 90° N	0 – 90° N
Original spatial resolution	0.25° × 0.25°	0.25° × 0.25°	0.25° × 0.25°	25 km × 25 km	0.25° × 0.25°	0.5° × 0.625°
Original temporal resolution	Daily	Daily	Daily	Daily	6 hours	daily
Duration	2003-2011	2013-2019	1980-2019	1980-2019	1980-2019	1980-2019
Parameter transformation	Snow depth	Snow depth	Snow depth	Snow water equivalent / snow density	Snow water equivalent / snow density	Snow depth* × fractional snow cover

Snow depth* represents the average snow depth in the snow-covered area of a 0.5° × 0.625° pixel.

Stand-alone passive microwave snow depth products

AMSR-E and AMSR-2 snow depth products are produced by the National Aeronautics and Space Administration (NASA) and Japan Aerospace Exploration Agency (JAXA), while NHSD snow depth was released by the Chinese Academy Science (CAS). AMSR-E, AMSR-2, and NHSD snow depth products did not rely on any external in situ snow measurements, only on passive microwave brightness data. The retrieval algorithm used in producing the AMSR-E and AMSR-2 snow depth products was derived by Kelly (2009), which was an improved Chang algorithm (Chang, 1987) that takes the fractional forest cover into consideration. The AMSR-E and AMSR-2 brightness temperatures were calibrated based on the Special Sensor Microwave Imager/Sounder (SSMI/S) to ensure data consistency. In these two snow depth products, if the snow was detected by the passive microwave brightness temperature, the value of snow depth was set to 5 cm. Spatiotemporal interpolation was employed to fill the striped gaps in the original AMSR-E and AMSR-2 snow depth products to produce a full space coverage product. The accuracy of the AMSR-E and AMSR-2 snow depth products were more accurate in North America, while the error was larger in Eurasia. However, the second version of the AMSR-E snow depth product, produced by NASA, used the ANN algorithm to improve the accuracy (Tedesco and Jeyaratnam, 2016).

The NHSD snow depth product was derived based on multiple-sensor inter-calibrated passive microwave brightness temperature data (Dai et al., 2015) using an improved Chang algorithm (Che et al., 2008). Through extensive in situ

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

observations, a dynamic coefficient was applied to the Northern Hemisphere to produce a more reliable and time-consistent snow depth product. The validation of in situ observation sites showed that the relative deviation of the NHSD snow depth product was less than 30% (Che et al., 2019). The NHSD snow depth product has higher accuracy in China than in Europe, America and Canada where the values of snow depth are deeper. Due to the saturation effect of the passive microwave brightness temperature, all remote sensing snow depth products tend to underestimate when the snow depth is deeper than 50 cm.

Passive microwave incorporated in situ observation snow depth product

The GlobSnow snow depth product, released from the European Space Agency (ESA), provides the snow water equivalent value and snow density at the pixel scale over the Northern Hemisphere.

The GlobSnow snow depth product included some in situ observations (Takala et al., 2011) and is currently considered the most accurate remote sensing snow depth product (Pulliainen et al., 2020). However, this product excludes mountain areas and only includes areas north of 35° N. The algorithm of the GlobSnow snow depth dataset was used to best estimate snow packs roughly 0.05 m–1.00 m deep (Luojus et al., 2021). The GlobSnow snow depth product shows high accuracy in Eurasia and low precision in regions of North America and the Tibetan Plateau. The global gridded GlobSnow depth product, also released by ESA, marks snow depth values in mountain areas as missing values. A recent study, attempted to apply ANN to improve the accuracy of the GlobSnow snow depth product (Venäläinen et al., 2021).

Reanalysis snow depth products

The ERA-Interim snow depth product from the fourth generation of reanalysis (Dee et al., 2011) was generated and released from the European Centre for Medium-Range Weather Forecasts (ECMWF). The latest ERA-Interim reanalysis dataset was updated until 31 August 2019, after which the dataset was replaced by ERA5. For this study, the snow water equivalent and snow density from ERA-Interim were downloaded at a six-hour time step, then processed as a daily snow depth dataset.

The MERRA-2 reanalysis snow depth dataset (Gelaro et al., 2017) was released by the Global Modelling and Assimilation Office of NASA (GMAO). In the process of the MERRA-2 reanalysis dataset, the Catchment land surface model (Gelaro et al., 2017) was employed to improve the quality of the data. In the study, the average snow depth of the snow-covered area in a pixel and the fractional snow cover in same pixel were used to calculate the average snow depth of the pixel. The original spatial resolution of the MERRA-2 reanalysis dataset was 0.5° × 0.625°, which was subsequently resampled to 0.25° × 0.25° spatial resolution by nearest interpolation. This ensured that the original gridded snow depth products used in this study had the same spatial resolution of 0.25° × 0.25° and daily temporal resolution from 1980 to 2019.

2.2 Ground-based snow depth measurement

In situ snow depth datasets included meteorological station observations from China and Russia, snow survey data from Russia and Global Historical Climatology Network (GHCN)–daily snow depth. These four categories of snow depth datasets were used as the ground-truth values for training and validating the accuracy of the proposed fused framework. In this study, the time series of these in situ observations were obtained from 1980 to 2019.

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

Daily snow depth data of China were acquired from the national meteorological information centre of the Chinese Meteorological Administration, with 945 stations used in this study. This dataset offers several meteorological parameters, i.e., snow depth, snow pressure, location, and elevation of the station sites. Daily snow depth data and snow pressure per five days data were manually observed at 8:00 am. with the same ruler. These snow depth data were calibrated and quality controlled under rigorous data standards before they were uploaded on the national meteorological data platform.

A daily snow depth dataset from Russia between 1980 and 2019 was collected from the Russian meteorological centre. This dataset also contained snow depth, location, and elevation of the station sites, and the quality checked field was also conducted. Some anomalous snow depth records were eliminated in this dataset during the data pre-processing phase.

Snow survey data of Russia were also accessed through the Russian meteorological centre. Several snow depth variables are included in this dataset, i.e., maximum, average, and minimum snow depth, snow density every five days to ten days from September to May. This dataset was also conducted according to the value of snow depth.

The GHCN daily snow depth dataset was the most comprehensive for recording global snow depth observations. This dataset also contained snow depth and elevation of in situ observation sites. Through the inter-annual consistency and climatological outlier check, some errors or abnormal values were excluded. Finally, 41261 in situ station sites were retained for the random forest fused training and validation.

In this study, seven long-term snow depth datasets from reference in situ sites were chosen to evaluate the machine learning fused snow depth dataset (Ménard et al., 2019). These datasets were derived from the Earth System Model-Snow Model Intercomparison Project (ESM-SnowMIP) (Krinner et al., 2018). These in situ observation sites offer observation periods for seven to twenty years. These in situ observations were vital for developing snow depth retrieved model and quantitative evaluation of the uncertainty of the model. To objectively and accurately evaluate the fused snow depth dataset, these independent in situ sites were used to test our data.

2.3 Auxiliary Data

Auxiliary data mainly included land use/cover data and elevation. Forest can cause snow depth to be underestimated in remote sensing-derived products, and bare land can impact wind-blown snow causing deeper snow depth in some areas. To improve the overall fusion accuracy by training and verifying the model under different vegetation types covers, the land of the Northern Hemisphere land was divided into 5 categories (forest, grassland, shrub, bare-land, and unclassified) (Fig. A1.). We assumed that the major land cover type at a spatial resolution of 0.25° did not change over the 40-year period from 1980 to 2019. The snow depth data within one land cover type should be systematically more consistent and may show less variability than taking the snow depth data as a whole (Ntokas et al., 2021).

Global Multi-resolution Terrain Elevation data 2010 (GMTED2010) is an update of GTOPO30 released by the United States Geological Survey (USGS). Elevation data are available at three spatial resolutions (Danielson and Gesch, 2011). In this study, the elevation data with a spatial resolution of 30 arc-seconds was resampled into a new spatial resolution of 0.25° × 0.25°. Elevation data at a resolution of 0.25° were used as an input variable.

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

3 Methodology and the accuracy assessment indices

As described in previous work (Hu et al., 2021), three machine learning methods, i.e., ANN, SVR, and RFR, were employed to try fuse snow depth datasets at the Northern Hemisphere. This work showed that the random forest fusion framework has the best performance for generating a long time series snow depth dataset. Detailed information about the random forest algorithm can be found in Belgiu and Drăguț. (2016) and information on the random forest fusion framework was referenced from Hu et al. (2021).

3.1 Description of the fused data preparation process

In this study, a snow hydrology season corresponding to year Y was defined as starting on September 1st of year Y-1 and ending on May 31st of year Y. The dates were selected because the Northern Hemisphere winter season spans from September until June over the new year (Venäläinen et al., 2021). As the snowpack evolves, the snow depth tends to first increase and then decrease. A complete snow hydrology year includes three snow phases (snow accumulation, snow stabilization, and snow melting). To reduce the errors caused by seasonal variations, one snow hydrology year was divided into three snow seasons: autumn (September to November), winter (December to February) and spring (March to May). In this study, these three snow phases are only used to split the snow depth dataset and were not used as input variables. To improve the accuracy of the RFR model in different snow seasons, different models were established. According to the three snow seasons and five land cover types (Fig. A1), 15 models can be employed to train and verify the model. According to existing accuracy assessment of several snow depth products (Xiao et al., 2020) and the testing of the data fusion based on machine learning (Hu et al, 2021), the random forest fusion framework displayed a better performance on large scale snow depth combination. Based on our previous work, we used the proposed random forest data fusion framework to generate a high precision daily snow depth dataset in the Northern Hemisphere from 1980 to 2019.

Because the original gridded snow depth products had various time durations, the period 1980 to 2019 was divided into three parts. The first period was from 1980 to 2002, and 2012 to 2013, the input gridded snow depth products were NHSD, GlobSnow, ERA-Interim, and MERRA-2 products. The second period include five input gridded snow depth products, i.e., AMSR-E, NHSD, GlobSnow, ERA-Interim, and MERRA-2, from 2003 to 2012. The last period also contained five input gridded snow depth products, i.e., AMSR-2, NHSD, GlobSnow, ERA-Interim, and MERRA-2, from 2013 to 2019. In different periods, different selected datasets were extracted for the random forest fusion model training and prediction. In every stage, a “leave-one-year-out” cross-validation was employed to generate the fused snow depth dataset. Lastly, by integrating the data from these three stages, the daily fused snow depth for the last 40 years was obtained.

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

Open Access Earth System Science Data Discussions

3.2 Accuracy and uncertainty assessment

We evaluated the accuracy between the fused snow depth, the original gridded snow depth products, and the in situ observations. The coefficients of determination (R^2), root mean square error (RMSE), mean absolute error (MAE) and the bias of in situ observations and snow depth products (BIAS) were calculated to assess these snow depth products.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(S_i - S_i)^2}{\sum_{i=1}^{n}(S_i - \overline{S})^2},$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(S_i - S_i)^2}{n}},$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|S_i - S_i\right|,$$

$$BIAS = S_i - S_i$$

where n is the number of sample pixels, S_i represents the in situ observation snow depth, while S_i denotes the original gridded snow depth or fused snow depth values of the i-th pixel, respectively. $\overline{S}$ represents the mean value of in situ observations of n pixels.

Furthermore, to detect the variation trend of the fused snow depth dataset from 1980 to 2019, We used the Mann-Kendall trend analysis method.

4 Description, accuracy assessment and trend analysis of the fused snow depth dataset

4.1 Description of the fused snow depth dataset

The duration of the fused snow depth dataset is from 1980 to 2019, and temporal resolution is daily. This daily snow depth dataset contains January 1 to May 31 and September 1 to December 31 of 1981 to 2018 exclude June 1 to August 31, and the dataset for 1980 includes September 1 to December 31 while for 2019 contains January 1 to May 31. During June to August, snow cover existed in limited areas over the Northern Hemisphere, and the presence of liquid water content with snowpack can lead to large error of snow depth derived from passive microwave remote sensing (e.g. AMSR-E, AMSR2, NHSD and GlobSnow). In fact, the GlobSnow gridded snow depth products are severely missing in June to August. The data missing also severely influence the accuracy of the fused snow depth, so this fused data did not cover the temporal phase of June to August. The area of Greenland and Iceland were excluded in all gridded snow depth dataset, and Svalbard was excluded in GlobSnow snow depth product, so these three areas were also excluded in the fused snow depth dataset. The original gridded

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

snow depth datasets of NHSD and GlobSnow were produced by the passive microwave brightness temperature from SMMR (Scanning Multichannel Microwave Radiometer) in 1980 to 1987. Brightness temperatures form SMMR were collected every other day, and the scanning width is much narrower than SSM/I (Special Sensor Microwave/Imager), thus, large data missing exist. Snow depth product of NHSD and GlobSnow were also inevitably have a large number of missing data, resulting in the similar missing in the fused snow depth dataset. In addition, as daily detections of AMSR-E and AMSR2 do not fully cover the extent of the Northern Hemisphere, stripe gaps southward of 55°N can be found in the daily images. Although the original gridded snow depth products were interpolated using adjacent days, a few stripe gaps in low-middle latitude areas were still presented in some days. The fused snow depth dataset still missed a little data in low-middle latitude areas on some days from 1988 to 2019. The fused snow depth dataset missed values owe to the limitations of the machine learning fused framework when the original gridded snow depth product missed values in some areas.

The format of the daily fused snow depth was GeoTiff, which was properly set the projection with EPSG:4326. Spatio resolution of this snow depth is 0.25° × 0.25°. This gridded snow depth of every day was stored in a GeoTiff file. This fused snow depth dataset also adopted centimetre as unit of the snow depth to match other snow depth products, and NoData_Value is represented by -9999. We adopted the YYYY-MM-DD as the name format of the daily fused snow depth dataset, and for each year was stored in the same folder from 1980 to 2019. The name of every daily snow depth was "ML_NHSD_YYYY-MM-DD.tif", ML represents machine learning, NHSD denotes snow depth of the Northern Hemisphere, YYYY stands for the year 1980 to 2019, MM means the month from January to December excluding June to August, and DD represents date. This snow depth dataset can be drag-and-drop into ArcMAP or QGIS to view and process.

4.2 Inter-comparison between the fused snow depth dataset and original snow depth products based on in situ observations

From 1980 to 2019, several original gridded snow depth products, the fused snow depth dataset and in situ observations were cross-verified at pixels scale. In the process of model training, one point every ten points was retained for independent validation by sparse sampling. Because these points did not participate in the model training phase, they can be considered as independent reference points. The results indicate that the correlations between the commonly used gridded snow depth products and the in situ observation values were not particularly good, while there was good consistency between the fused snow depth product and the in situ observations. The coefficient of determination increased from the best 0.23 (GlobSnow snow depth product) in the original gridded snow depth products to 0.81 (fused snow depth product) (Fig. 1). The accuracy of the fused data was strongly improved in terms of RMSE and MAE, which were both reduced by about two times.

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

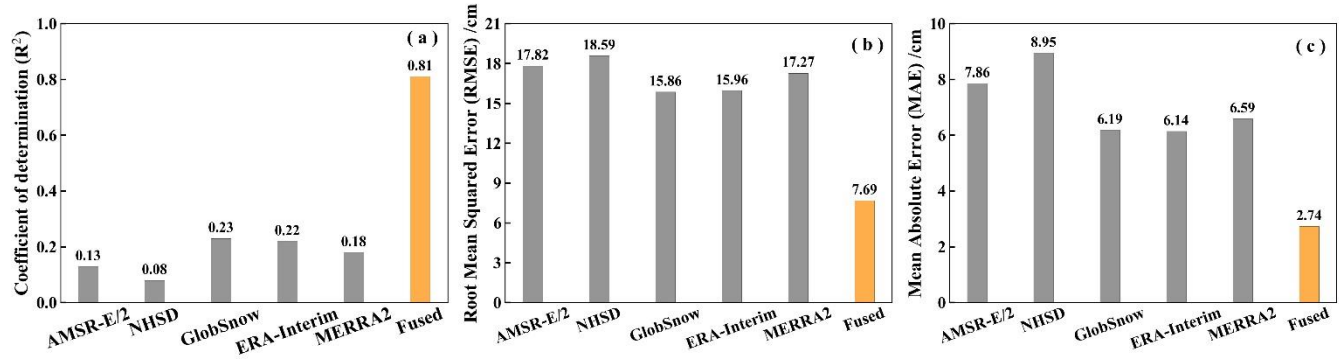


Figure 1. Accuracies of the fused snow depth dataset for six gridded snow depth products compared with in situ observations. The raw gridded snow depth products are displayed in grey, while the fused snow depth dataset was shown in orange.

As can be seen from the scatter density map (Fig. 2), the goodness of fit between the fused snow depth dataset and the in situ observations was high. When the snow depth values were less than 50 cm, most of these values were distributed near the 1:1 line. It can also be seen from this figure that the fused snow depth dataset showed good accuracy no matter when the snow depth values were shallow or deep. It also can be found that AMSR-E/AMSR-2, NHSD, and GlobSnow remote sensing snow depth products were largely underestimated when the snow depth value was greater than 50 cm. Although ERA-Interim has a good capture of the deep snow depth values, its overestimation and underestimation were obvious. MERRA-2, as a reanalysis snow depth product, was not very accurate, and there were many points of underestimating and overestimating disorderly distribution.

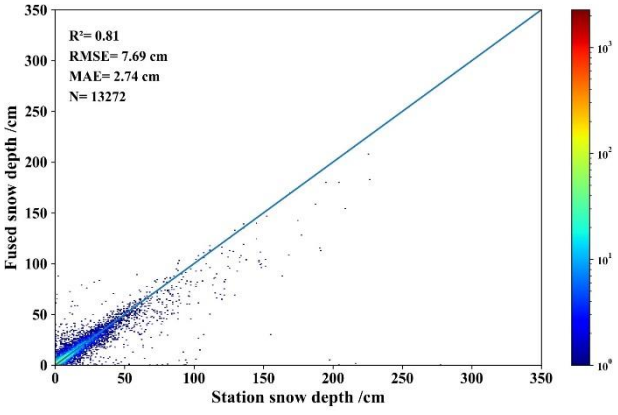


Figure 2. Scatterplots between snow depth and in situ observations between 1980 and 2019.

The BIAS of the fused data was mostly between -5 cm and 5 cm, which accounted for 88.31%, indicating that the fused snow depth dataset had high accuracy (Fig. 3). This also indicated that the consistency between the fused snow depth and ground station observations was very good over the entire Northern Hemisphere. Fused snow depth data were prone to overestimation and underestimation in areas where the in situ observations were sparse. The accuracy of the proposed

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

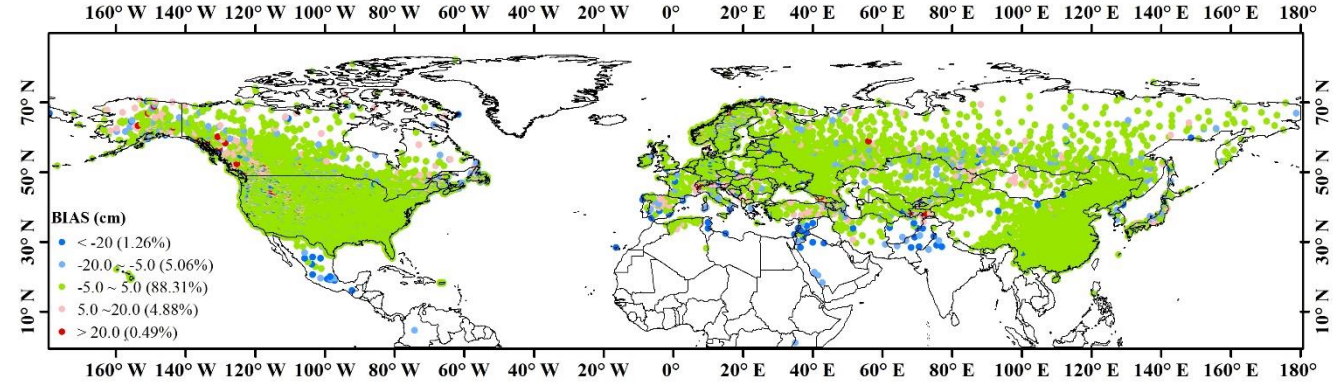


Figure 3. Spatial distribution of the average BIAS between fused snow depth dataset and in situ observations. The numbers outside the parentheses represent the intervals of BIAS while the numbers between parentheses denote the percentage of each interval in the legend.

4.3 Accuracy assessment of the fused snow depth dataset at several independent in situ snow depth observation sites

In Sect. 4.1, the accuracy assessment of the fused snow depth was conducted using ground station observations, which may reduce the credibility of this product since some station observations were already used in the model train. To increase the reliability of the fused snow depth dataset, some independent in situ observations were introduced to validate our snow depth product. In this study, seven reference snow observation sites, which were chosen for evaluating models participating in the ESM-SnowMIP (Krinner et al., 2018), were employed to test the accuracy of the fused snow depth dataset in these regions. Detailed information on these selected sites is shown in Table A1. The different time series of the fused snow depth dataset were extracted based on the time series of these in situ observations. From the accuracy validation of these seven sites, Sodankylä (SOD) had the highest assessment accuracy with the R^2, RMSE, MAE, and BIAS of 0.89, 8.7 cm, 6.4 cm, and 3.9 cm, respectively (Table 2).

Table 2. Accuracy metrics at several in situ observation sites. R^2: coefficient of determination, RMSE: root mean square error, MAE: mean absolute error, N: Number. NAN means that R^2 was not calculated because the errors were too large.

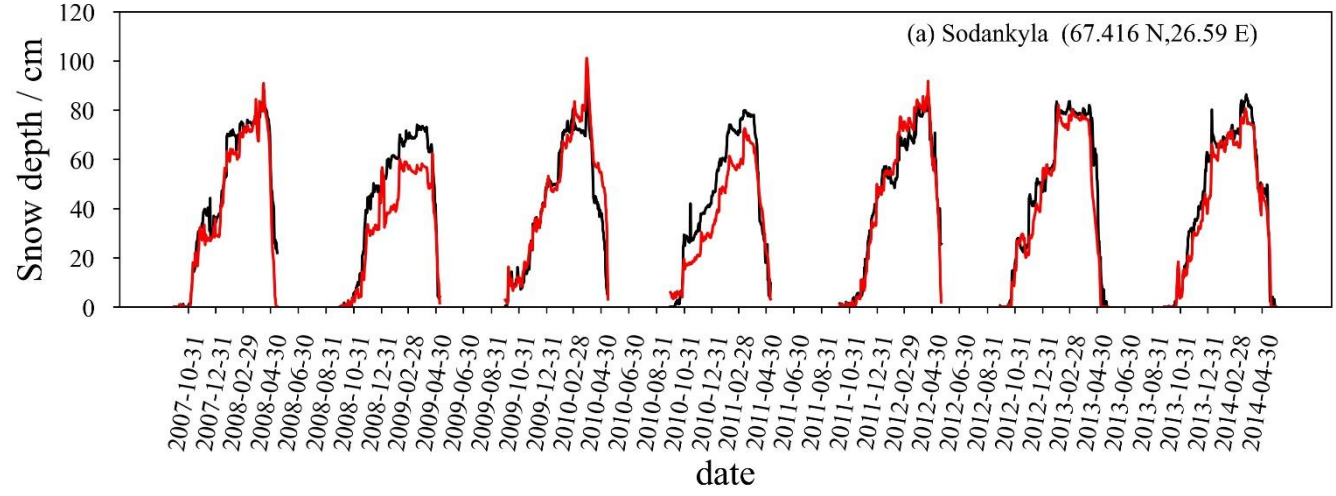
Sites name	R^2	RMSE / cm	MAE / cm	BIAS / cm	N
(a) Sodankylä (SOD)	0.89	8.7	6.4	3.9	1611
(b) Old Aspen (OAS)	0.76	8.2	4.4	-3.4	2673
(c) Old Black Spruce (OBS)	0.76	6.8	4.5	4.0	2456
(d) Old Jack Pine (OJP)	0.77	7.6	4.5	1.1	3512
(e) Senator Beck Basin Study Area (SBBSA)	0.22	56.5	42.2	-37.8	2365

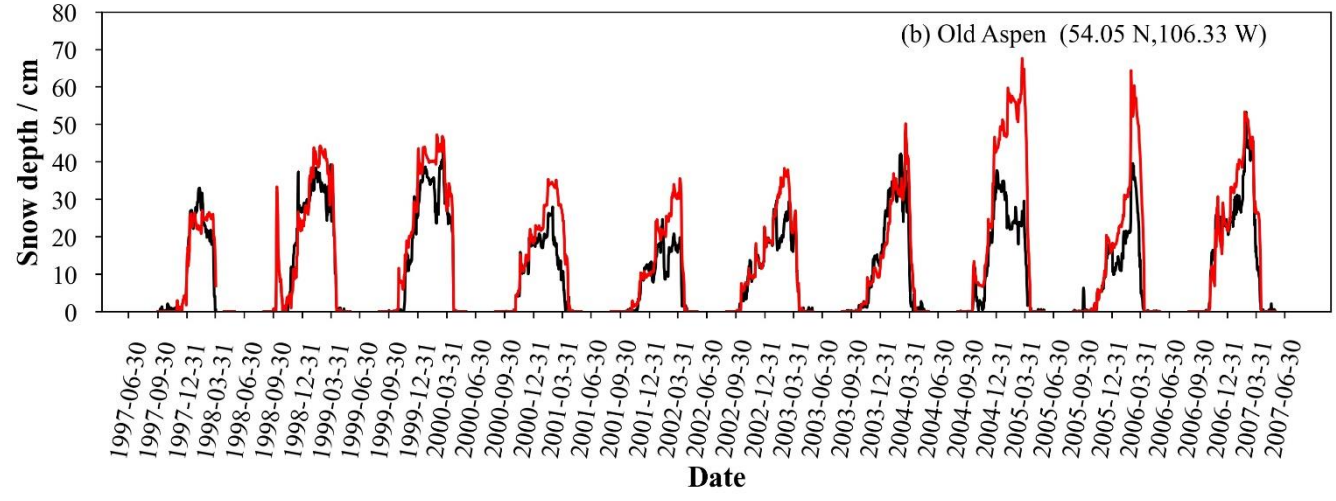
https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

(f) Swamp Angel Study Plot (SASP)	NAN	130.2	110.3	-110.3	2378
(g) Weissfluhjoch (WFJ)	NAN	129.1	103.7	-103.6	5443

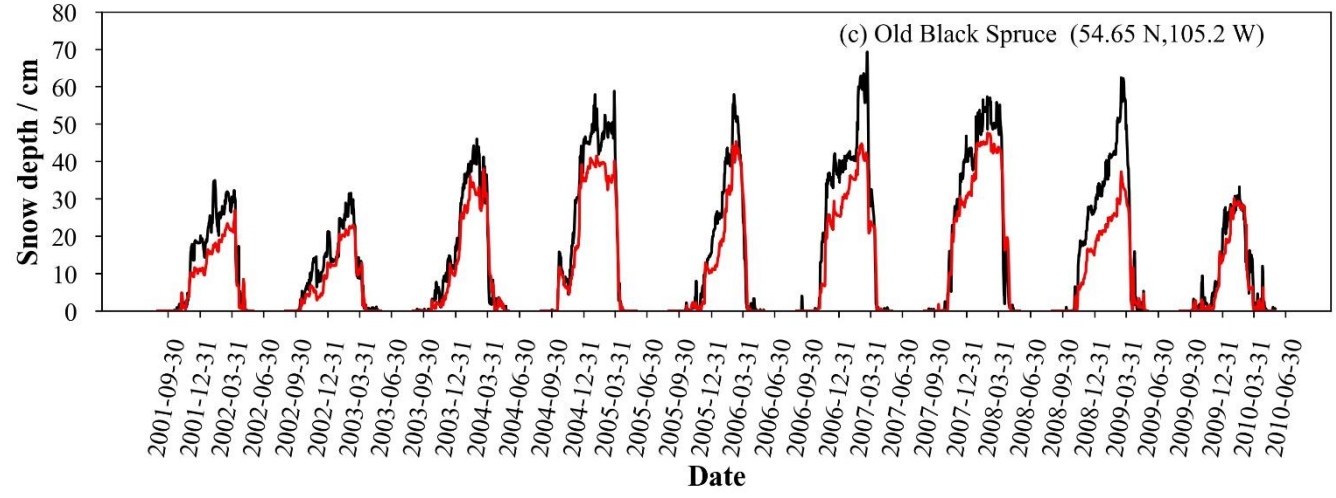
Although, the assessment accuracies of Old Aspen (OAS), Old Black Spruce (OBS), and Old Jack Pine (OJP) were not as high as those of Sodankylä, their accuracies were still relatively high compared with those of other gridded snow depth products. The fused snow depth can accurately estimate deeper snow, i.e., deeper than 50 cm (Fig. 4), capturing the maximum value and inflection point. Senator Beck Basin Study Area (SBBSA), Swamp Angel Study Plot (SASP), and Weissfluhjoch (WFJ), which had deeper snow depths in winter and spring, did not have a good performance. In these three in situ sites, the fused snow depth product showed a prominent underestimation, especially from November to April. Because the period from November to May is an important period for snow accumulation, stabilization, and ablation, the fused snow depth dataset is not suitable for these regions. Compared with the original gridded snow depth in SBBSA, SASP, and WFJ, there was a greater underestimation across all products. From these assessment sites, in a homogeneous environment and relative lower elevation of an 0.25° pixel, the fused snow depth showed a better performance. At sites in relatively complex environments, the fused snow depth product had a poor performance. Based on the KML file provided by Ménard et al. (2019), we found that SBBSA and SASP are located in the same basin. SASP is in a sheltered area and is surrounded by a sub-alpine forest. In this complex environment, the surrounding forest made it easier to store more snowpack; also, this site was better suited for measuring precipitation. In the original gridded snow depth products, remote sensing snow depths were shallower than in situ observations. The SBBSA site is located at the top of the mountain, with an elevation of more than 3700 m. The land cover type of this pixel is short grass, while the land cover of a 0.25° pixel is complex, with grass, bare rock and forest. In a 0.25° pixel, this site only represents a small region; the elevation range was varied from 2700 m to 3900 m. These two sites can be used to observe the snowpack characteristics in a basin but can not represent the large area snow depth. The WFJ site is located on a hillside at an altitude of about 2540 m. The major landcover type is grassland in one pixel, but this site has a higher altitude. The elevation increased from 800 to 2600 m over one pixel, so this site also has a complex environment. During the winter months, deeper snow builds up at this altitude. These results indicate that the accuracy of the fused snow depth dataset was heavily dependent on the input gridded snow depth products. Additionally, the snow depths changed too rapidly to be accurately captured by these products. In other words, the fused snow depth dataset had higher accuracy at the snow depths of less than 100 cm.

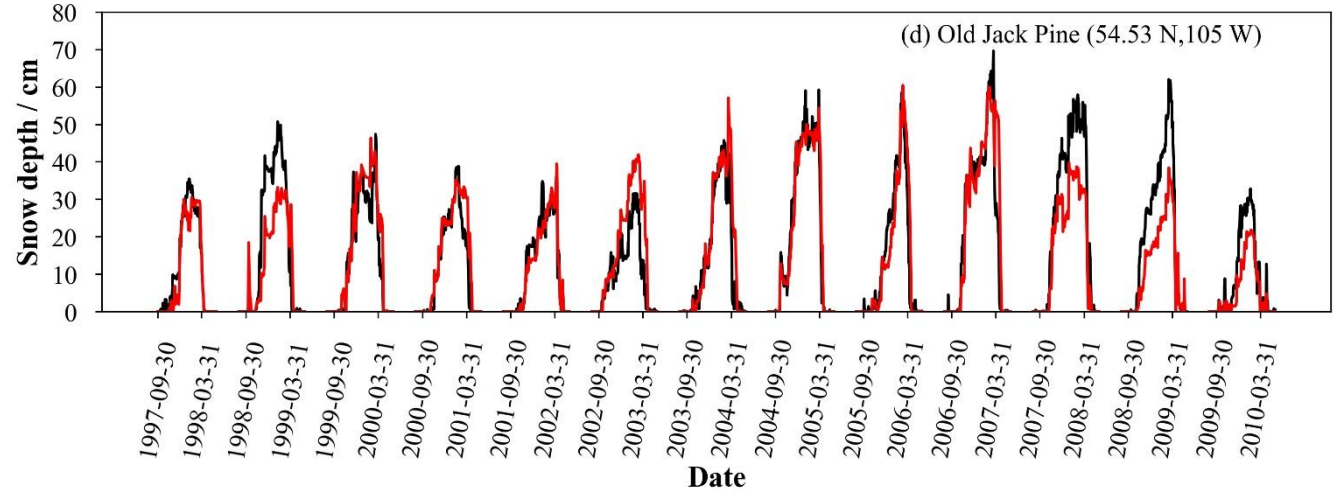
https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022



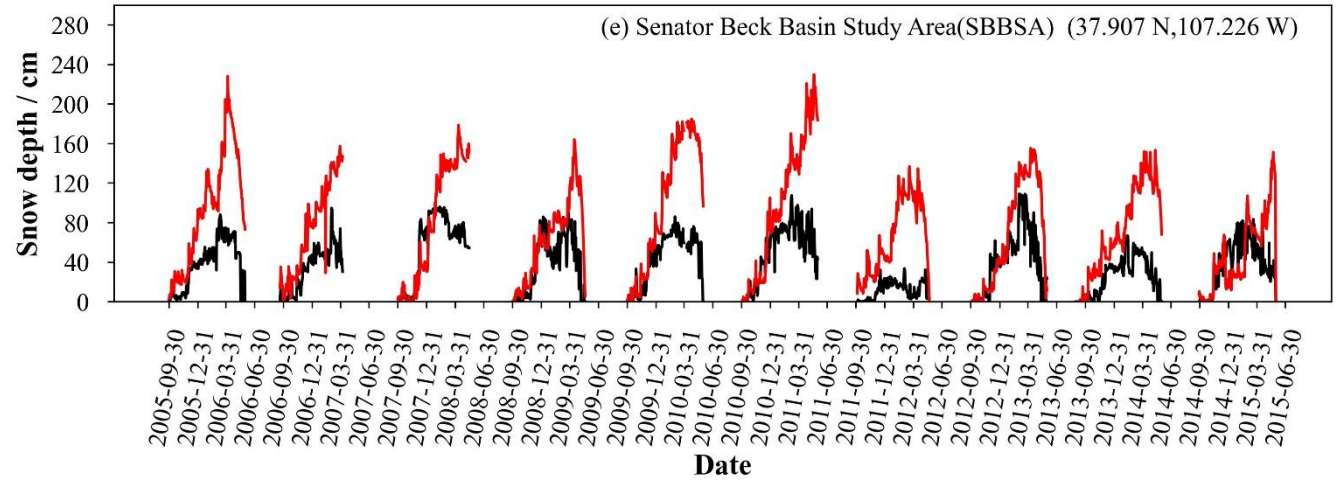


https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

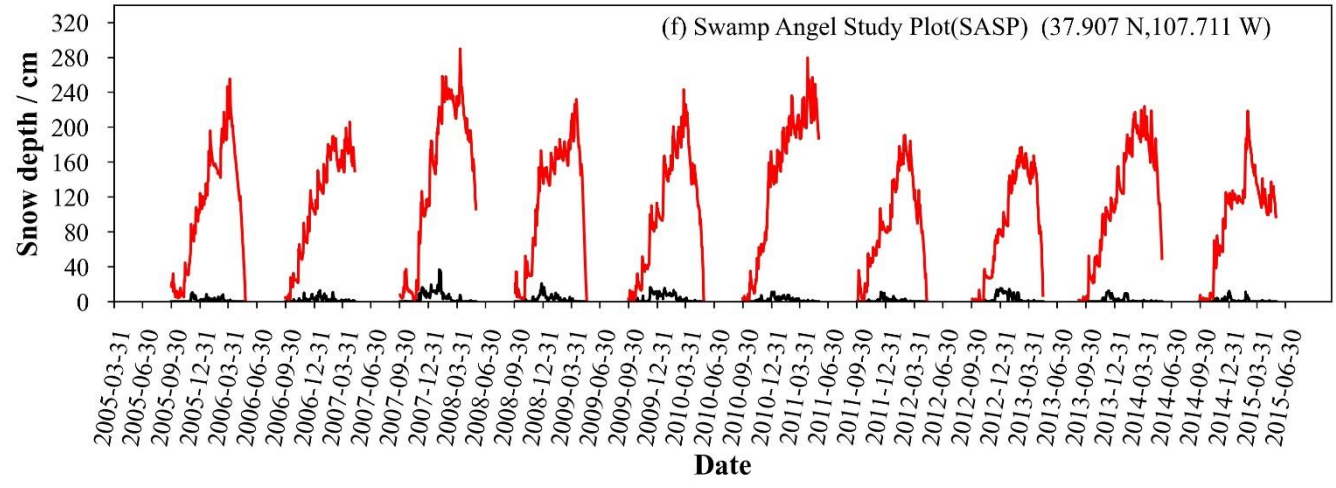


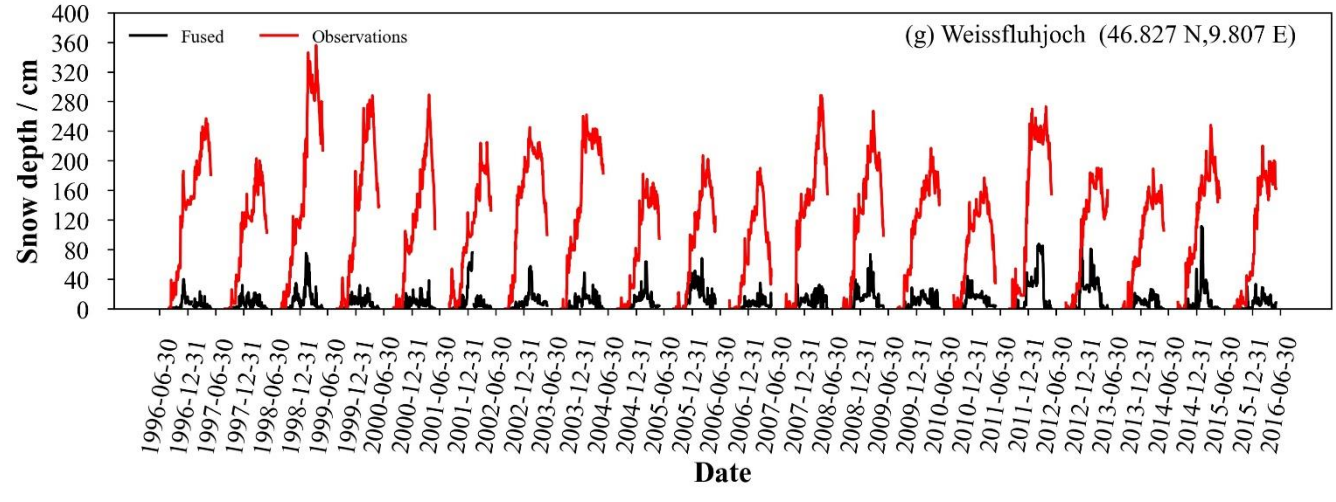


https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022



360





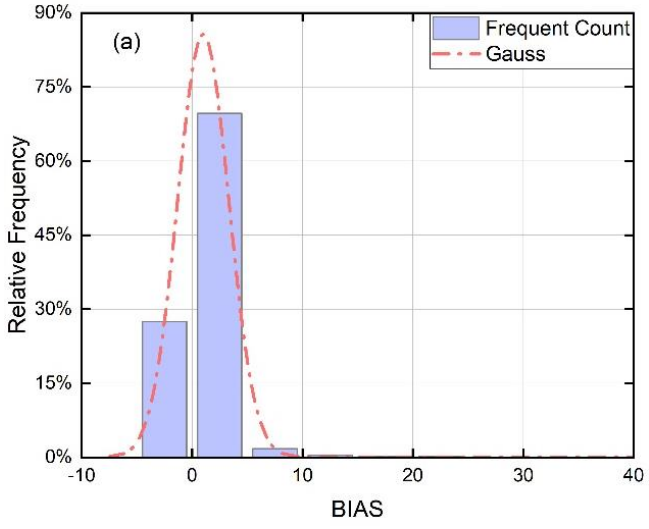
https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

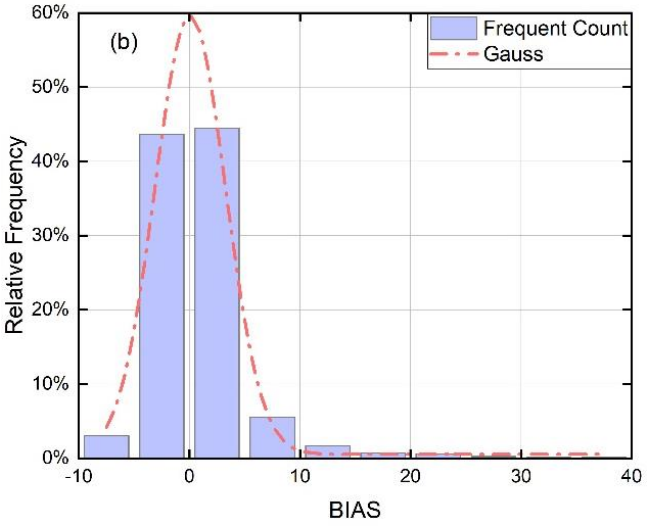
Figure 4. Fused (black line) and in situ (red line) snow depths at seven independent observation sites. (a) Sodankyla observation site in Finland (Essery et al. (2016)); (b) Old Aspen Black site in Canada (Bartlett et al. (2006)); (c) Old Black Spruce site in Canada (Bartlett et al. (2006)); (d) Old Jack Pine site in Canada (Bartlett et al. (2006)); (e) Senator Beck Basin Study Area site in the United States of America (Landry et al. (2014)); (f) Swamp Angel Study Plot site in the United States of America (Landry et al. (2014)) and (g) Weissfluhjoch site in Switzerland (Wever et al. (2015)).

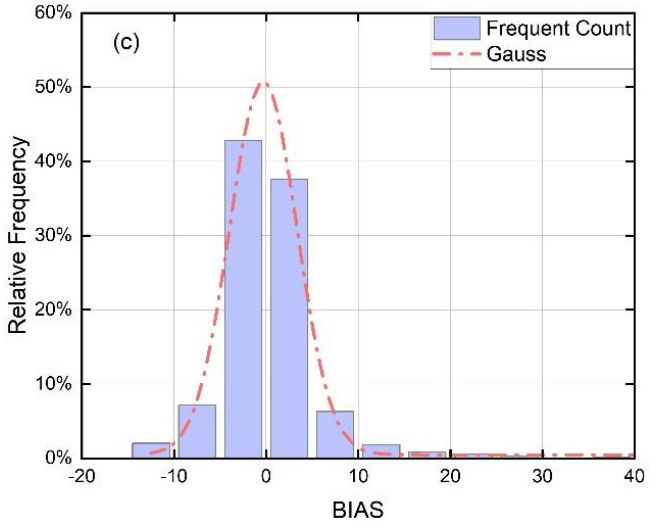
4.4 Accuracy evaluation of the fused snow depth dataset at different levels of snow depth values

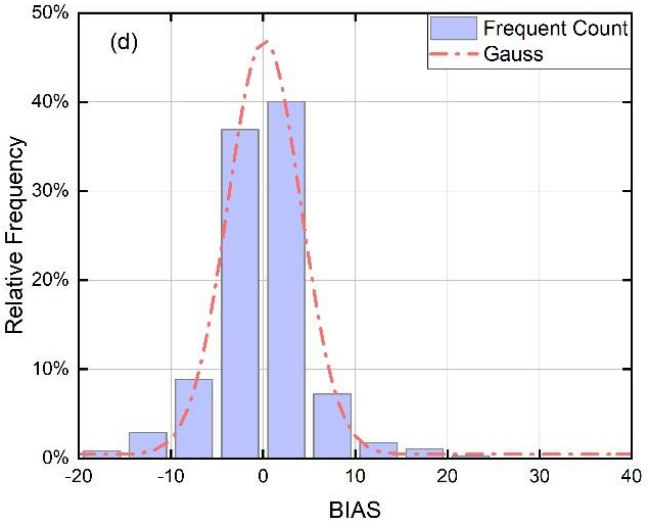
We compared the accuracy of the fused snow depth product at different levels based on selected in situ snow depth values. The fused snow depth values were divided into eight intervals: 0-5 cm, 5-10 cm, 10-15 cm, 15-20 cm, 20-30 cm, 30-40 cm, 40-50 cm, and above 50 cm. In the interval of 0-5 cm, the relative frequency of BIAS between -5 cm to 5 cm was more than 90 %, and it indicated that the fused snow depth had a slightly overestimated trend. In the range of 5-10 cm, the BIAS also showed slight overestimation, but the relative frequency of the BIAS between -5 to 5 cm was also more than 80 %. The distribution charts of relative frequency in the intervals of 10-15, 15-20, and 20-30 cm depths were similar. These frequency distribution charts had a near normal distribution, and the relative frequency of the BIAS between -5 to 5 cm were all higher than 70 %. This result indicated that the errors of the fused snow depth were still relatively small. In the intervals 30-40 cm and 40-50 cm, the trend distribution of the relative frequency was similar, the frequency of the BIAS between -10 to 10 cm was higher. This indicated that errors increase with snow depth. In the last interval, the in situ snow depth observations were larger than 50 cm; the relative distribution plots showed an underestimation of the fused snow depth product. Although the intervals of high snow depth were underestimated, the relative frequency of BIAS near 0 was relatively high. Overall, Fig. 5 showed that the error of the fused snow depth dataset was small and the accuracy was high.

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022









https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

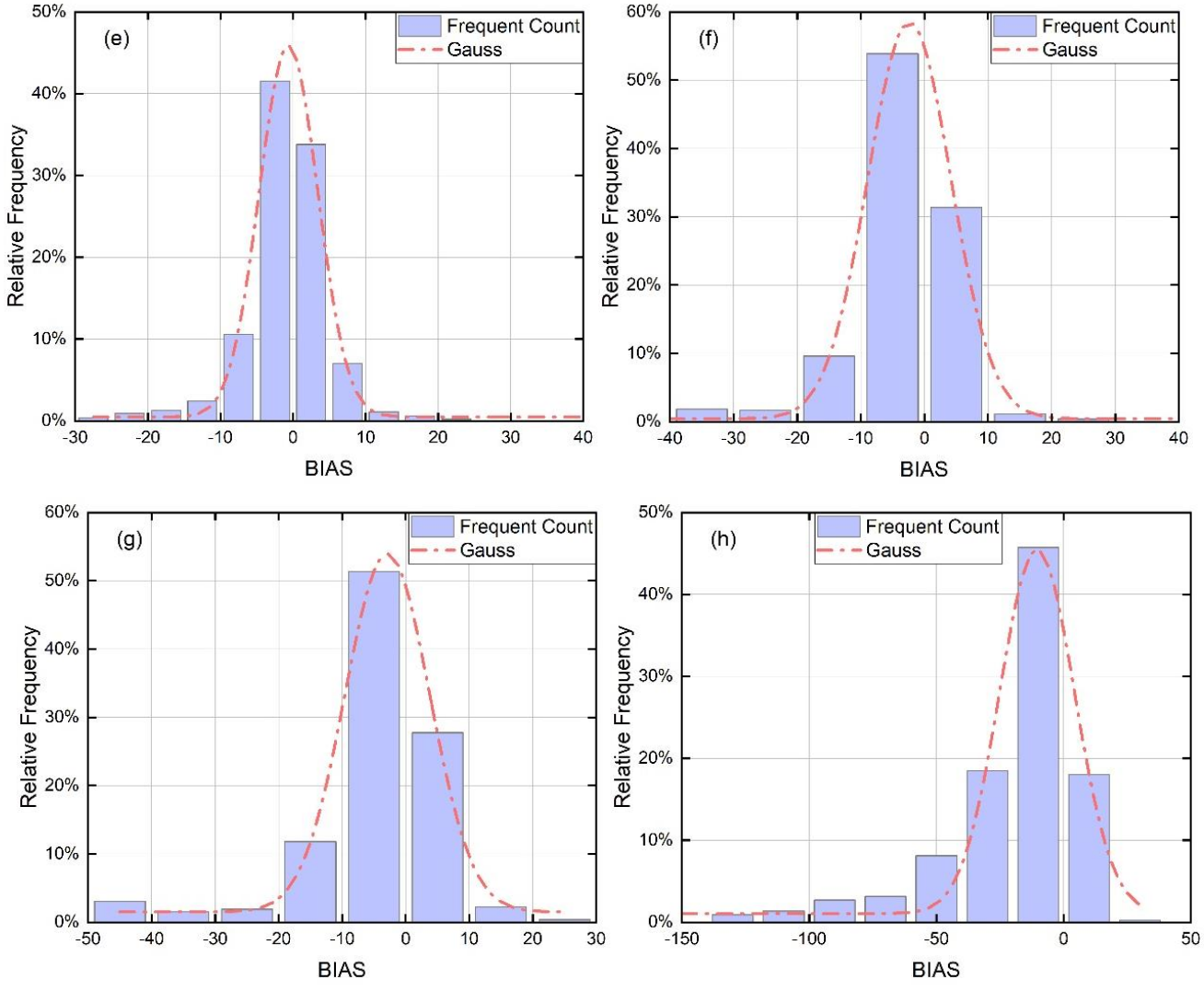


Figure 5. The frequency histogram of BIAS between the fused snow depth dataset and selected in situ observations. a, b, c, d, e, f, g, and h correspond to snow depths of the in situ observations of 0-5 cm, 5-10 cm, 10-15 cm, 15-20 cm, 20-30 cm, 30-40 cm, 40-50 cm, and above 50 cm, respectively.

4.5 Accuracy assessment of the fused snow depth dataset at different elevations

We investigated the accuracy of the fused snow depth dataset at different elevation intervals (Table. 3). The accuracy of the fused snow depth data had poor precision at altitudes less than 100 m or greater than 2000 m. In the elevation interval greater than 100 m and less than 1000 m, the accuracy of the fused snow depth was high and the consistency of the fused snow depth data was better. In the range of 1000 m to 1500 m, both snow depth and error of the fused dataset were greater. In the

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

interval range of 1500 m to 2000 m, the accuracy of the fused snow depth dataset was highest, with R^2 reaching 0.86. In the range of the elevation greater than 2000 m, the fused accuracy was higher than at the range of elevation less than 100 m.

Table 3. Accuracy assessment between the fused snow depth dataset and in situ observations at different elevations.

Range of elevation	R^2	RMSE / cm	MAE / cm
0~100 m	0.55	8.55	1.79
100~500 m	0.79	5.37	1.88
500~1000 m	0.79	6.42	2.44
1000~1500 m	0.73	13.29	4.11
1500~2000 m	0.86	9.84	3.98
>2000 m	0.69	16.45	5.97

R^2: coefficient of determination, RMSE: root mean square error, MAE: mean absolute error.

4.6 Spatial and temporal patterns of the snow depth between 1980 and 2019

We separated the Northern Hemisphere into North America and Eurasia to evaluate the temporal change in average snow depth from 1980 to 2019. There was an overall trend of decrease followed by a slowly increase in both regions, and the minimum value of snow depth occurred in the 2012 snow hydrology season. From 2013 to 2019, the average snow depth in North America and Eurasia were relatively smooth. From 1980 to 2019, the decreasing trend of snow depth in North America was stronger than in Eurasia (Fig. 6 (a)).

The spatial distribution of the multi-year average fused daily snow depth dataset from 1980 to 2019 is shown in Fig. 6(b). The regions with the deepest snow depth values are distributed in the western Siberian plain, Rockies, and Alps. The average snow depth values of the central Siberia Plateau were about 20 to 30 cm, which was shallower than in the eastern Siberian Mountain and western Siberia plain. The average snow depth distribution in Canada increased gradually from west to east, from 10 cm to 40 cm. From this spatial pattern, we also found that most low latitude regions (< 40° N) had snow depth of less than 5.0 cm. The snow depth of the Tibetan Plateau was also shallower; only small parts near the eastern part of Qinghai-Tibet Plateau had snow depth of about 5 to 10 cm. There were fewer in situ observations to train the random forest data fusion framework, resulting in a lower snow depth accuracy in this region.

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

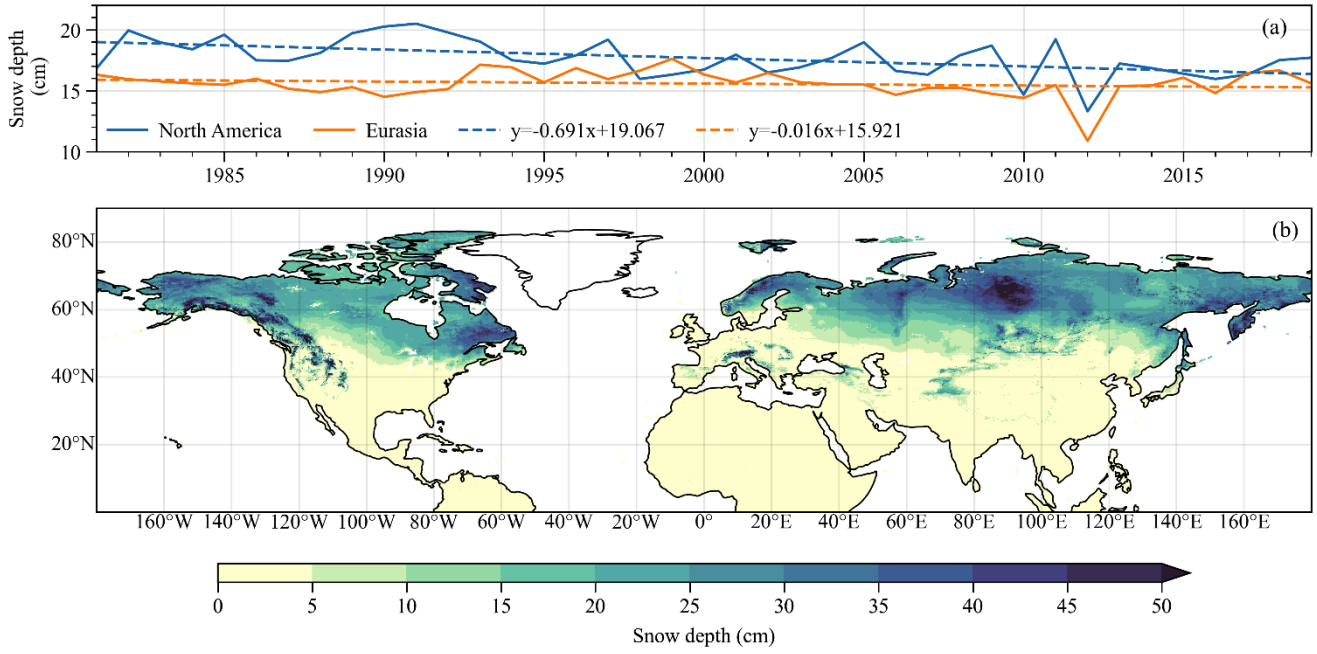


Figure 6. Spatial and temporal distribution of multi-year average snow depth over the Northern Hemisphere from 1980 to 2019. (a) Time series of average snow depth over North America and Eurasia; (b) spatial distribution of multi-year average snow depth.

The spatial distribution of average snow depth in autumn, winter, and spring was roughly similar to the overall spatial pattern, with more pronounced distribution areas with high value (Fig. 7). High average snow depths were found in western Siberian plain, eastern European plain, the farthest east of Canada, Rockies, and Alps. Seasonally, the mean snow depth in Autumn was roughly below 5 cm, which was significantly lower than that in Winter and Spring, but all three times were relatively smooth, indicating that it was reasonable to divide the snow hydrological year into three different snow seasons, and to calculate the snow depth at different levels was more reasonable and precise.

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

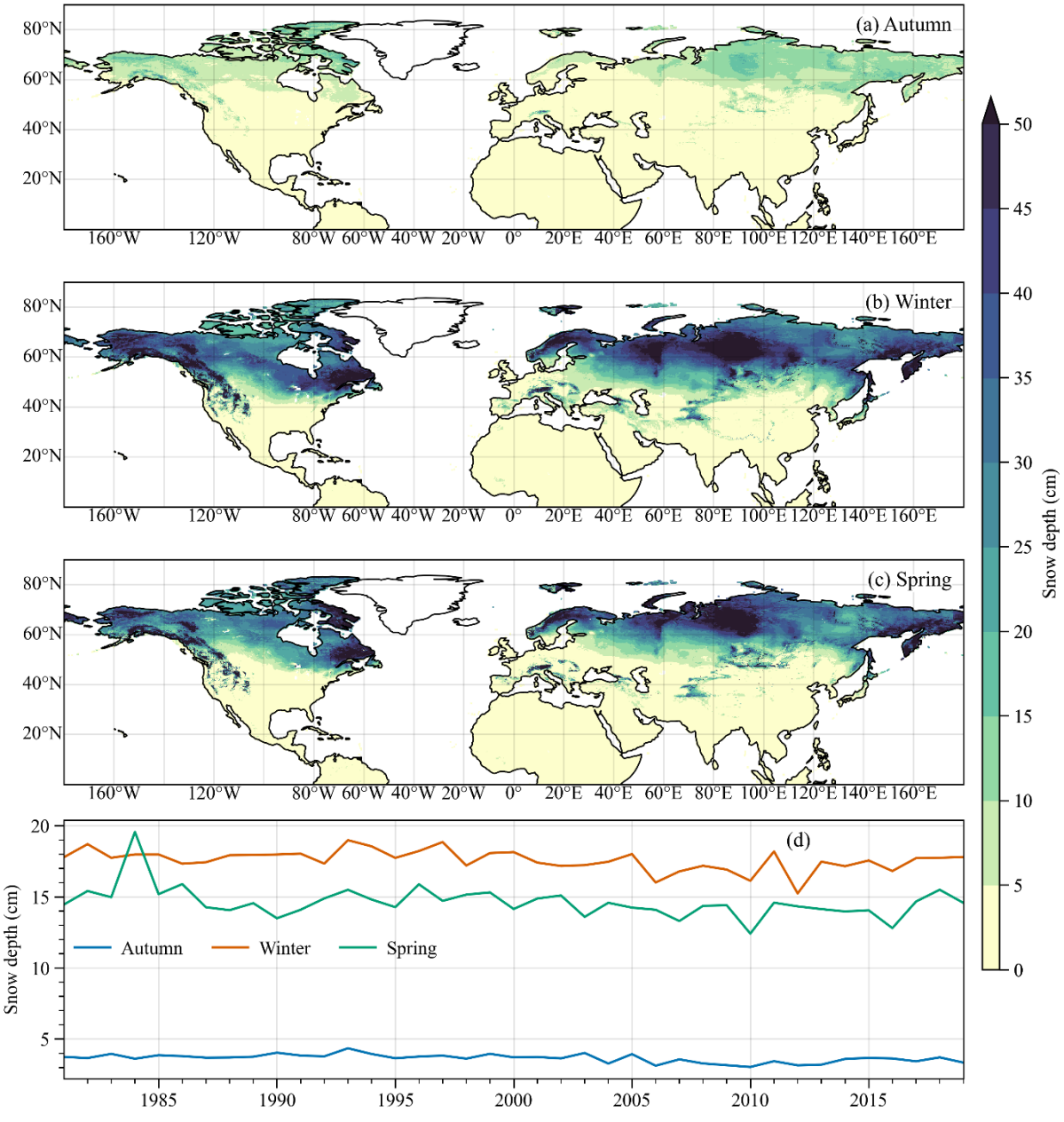


Figure 7. Spatial and temporal distribution of multi-year average snow depth in different snow seasons over the Northern Hemisphere from 1980 to 2019. (a) Spatial pattern of multi-year average snow depth in Autumn (September to November); (b) spatial pattern of multi-year average snow depth in Winter (December to February); (c) spatial pattern of multi-year average snow depth in Spring (March to May), and (d) time series of change of the average snow depth over the Northern Hemisphere by snow seasons.

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

4.7 Comparison of the changing trend of the fused snow depth based on the Mann-Kendall trend test

The changing trend in the annual average snow depth of the Northern Hemisphere from 1980 to 2019 was detected by the Mann-Kendall trend test method. From 1980 to 2019, the test value of fused snow depth product was -3.28, and showed a significantly decreasing trend at the significance test level of 0.05.

The trend analysis result indicated that 71 % area had a decreasing trend while 29 % of the area showed an increasing trend. According to the Mann-Kendall trend monitoring algorithm, the areas of significant increase were mainly in the West Siberian plain, East Siberian, and Rockies (Fig. 8). These areas also passed the 0.05 significance test. The regions with significant decreases were mainly distributed in northern and eastern Canada, Sweden, and Finland. Most parts of these regions also passed the 0.05 significance test. In the regions below 40° N, the change trend was not very significant.

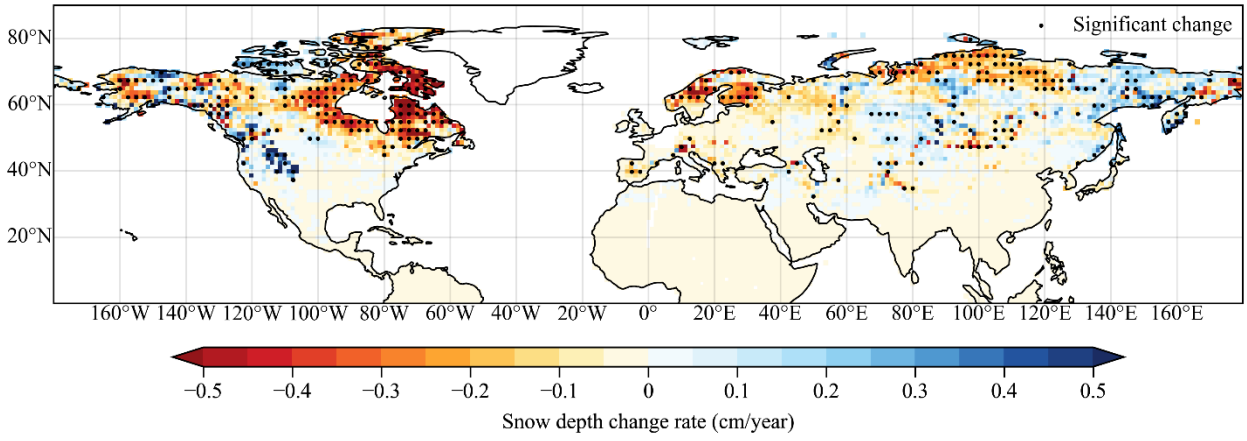


Figure 8. Spatial distribution of snow depth changed response times over the past four decades by Mann-Kendall trend test. The black dots in the graph indicated these regions passed the significance test of 0.05.

5 Discussion

5.1 Limitations of the fused snow depth dataset

Previous assessments determined that the original gridded snow depth products contain some errors (Xiao et al., 2020). Based on the limitations of our proposed data fusion framework (Hu et al., 2021), creating a seamless snow depth product over the Northern Hemisphere requires that all original gridded snow depth products have non-missing snow depth values. During pre-processing, the AMSR-E and NHSD snow depth products were spatio-temporal interpolated to remove striping. The GlobSnow snow depth product only covered the area north of 35° N. In this study, we assumed that land cover change at a spatial resolution of 25 km was negligible, so the land cover was kept constant. Lastly, the accuracy of the fused framework depends heavily on the training samples. Most snow depths from the meteorological stations were small, and only a small number of deeper snow values were derived from snow field surveys. The quality and accuracy of the fused

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

snow depths in regions with deeper snow were poor. To overcome this deficiency, more snow surveys data should be included as input variables.

Because most machine learning methods are currently black-box models, they have no physical mechanism to constrain their estimates. In our fused framework, the input variable selection is based on experience and previous studies. Even so, the fused results indicate that the approach used in this study is rational. For example, in different snow seasons, the absolute snow depth differed considerably, suggesting that it was reasonable to fuse snow depth datasets in different snow seasons. Although current machine learning algorithms are not coupled with physical processes, some physical mechanisms are being depicted in the resulting products. In a future study, we will not only pursue integrated physical process constraints in machine learning algorithms (Karniadakis et al., 2021), but also use these machine learning results to explain physical mechanisms and significant trends.

5.2 Generalization ability of the proposed random forest regression model

To evaluate the stability of the random forest model and validation of the results, we used a combination of different spatial positions in the training samples (same time), different times of training samples and verification samples (same spatial positions), and different time and spatial positions of the training and verification samples. As expected, the highest accuracy was obtained when training and predicting using time samples in the same region and different years. Because of the generalization ability of machine learning, it was a more reasonable choice to use all the data in the Northern Hemisphere for model training. For example, to achieve the highest accuracy, we used pixels on the same position, snow hydrologic year data from 2002 to 2003 as the training data, and snow hydrologic year data from 2003 to 2004 as the verification sample. For pixels in different positions, the accuracy is not as high as when the model uses the same position and time, whether they are the same in time (different space) or different in time and space. It is also recognized that machine learning models are challenging to generalize in different spatial locations. The proposed RFR model structure is expected to be applicable to every area in the world. However, new training is advisable. The significance score increases even when only one variable is eliminated. Thus, more input variables help widen the spread of the ensemble (Ntokas, et al., 2021). Therefore, adding more variables could further increase the model's reliability. In the future, an important direction is to improve the random forest fused model architectures so the obtained model can be transferred across different geographic areas.

Regarding the limitations of this study, the fused snow depth product showed poor performance for deeper snow, mainly because the amount of available training data was low for these extreme values, as found in previous studies.

6 Data availability

The fused snow depth data presented here are freely available for download from the National Tibetan Plateau Data Center (TPDC) (https://dx.doi.org/10.11888/Snow.tpdc.271701 (Che et al., 2021)), and this snow depth dataset can also be download at https://zenodo.org/record/6336866#.Yjs0CMjjwzY. The accuracy evaluation data are available from

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

https://doi.org/10.1594/PANGAEA.897575. These validation points have different time series from 1994 to 2016, and each validation site provides a file in netCDF format. The comparison between the fused snow depth dataset and theses validation sites is based on the time duration of these sites.

7 Conclusions

We presented a new fused daily gridded database of snow depth for the Northern Hemisphere from 1980 to 2019 at a resolution of 0.25° (~ 25 km). This dataset fused an unprecedented amount of in situ measurements and multisource gridded snow depth products, using a machine learning algorithm. Comparing AMSR-E/AMSR-2, NHSD, GlobSnow, MERRA-2, ERA-Interim, and the fused snow depth datasets with in situ observation snow depths, the results shows that the original gridded snow depth have weak correlations, while the fused snow depth dataset has a good consistency with in situ measurements. The fused (best original) dataset has a coefficient of determination R^2 of 0.81(0.23), Root Mean Squared Error (RMSE) of 7.69 (15.86), and Mean Absolute Error (MAE) of 2.74 (6.14). Precision validation via several independent in situ observations indicates that fused snow depth data are likely more accurate than those of previous studies. Therefore, this new fused snow depth dataset not only provides information about snow depth and its variation over the Northern Hemisphere but also presents potential value for hydrological and water cycle studies related to seasonal snowpacks. More importantly, the fused snow depth dataset can be used for water resource management, environmental monitoring, and early warning of snowfall.

Although the proposed machine leaning fused framework performs well in the snow depth product fusion over the Northern Hemisphere, some drawbacks and limitations still need to be overcome, especially on the input variables selection and machine learning algorithm optimisation. In our future work, deep learning algorithms will be introduced to fuse snow depth datasets in conjunction with exploring relevant physical information about the snowpack. We will update the fused snow depth dataset when new gridded snow depth products become available. For example, ERA5 snow depth product will be incorporated in our proposed framework.

Author contributions. TC designed the proposed model. YH and LX deal with the data. YZ improved the python code for the Mann-Kendall trend test. YH produced the first draft of the manuscript, which was subsequently edited by TC, LD, YZ, LX, JD, and XL. All authors contributed to the analysis and interpretation of the results and to improve this paper.

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

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

Special issue statement. This article is part of the special issue "Extreme environment datasets for the three poles". It is not associated with a conference.

Acknowledgements. The authors would like to thank the National Tibetan Plateau Data Center (TPDC) for NHSD, NASA and JAXA for AMSR-E, AMSR-2 and MERRA-2 snow depth products, ESA for GlobSnow snow water equivalent, and ECMWF for MERRA-2 snow product. We thank Dipesh Rupakheti for editing some grammatical errors and the ESM-SMIP for offering the independent validation data. The authors also thank the national meteorological information center for providing the snow depth of China's meteorological stations, Russian meteorological center for providing the snow depth values of Russian meteorological stations and snow survey dataset, and the National Oceanic and Atmospheric Administration (NOAA) for providing GHCN snow depth dataset. We also acknowledge the Earth big data cloud service platform which supports by CASEarth.

Financial support. This study was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (no. XDA19070101), the National Science Fund for Distinguished Young Scholars (no. 42125604), the National Nature Science Foundation of China (no. 41771389 and 42001289) and the CAS 'Light of West China' Program (E029070101).

Review statement. This paper was edited by XXX and reviewed by XXX anonymous referees.

Appendix

Table A1. Independent validation sites (Based on Ménard et al., 2019).

Site name	Short name	Latitude (°)	Longitude (°)	Elevation (m)	Vegetation Type	Reference
Sodankylä	SOD	67.416	26.59	179	Clearing (short heather and lichen) in coniferous forest	Essery et al. (2016)
Old Aspen	OAS	54.05	-106.33	600	21 m high aspen forest. Thick understory of 2 m high hazelnut. Winter stem area ~ 1, summer3.7-5.2	Bartlett et al. (2006)
Old Black Spruce	OBS	54.65	-105.2	629	12 m high black	Bartlett et al. (2006)

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

					spruce forest. Sparse understory. Leaf area index 3.5-3.8	
Old Jack Pine	OJP	54.53	-105	579	14 m high forest. Sparse understorey. Leaf area index 2.5-2.6	Bartlett et al. (2006)
Senator Beck Basin Study Area (SBBSA)	SBBSA	37.907	-107.226	3714	Alpine tundra	Landry et al. (2014)
Swamp Angel Study Plot (SASP)	SASP	37.907	-107.711	3371	Clearing (short grass) in subalpine forest	Landry et al. (2014)
Weissfluhjoch	WFJ	46.827	9.807	2536	Barren	Wever et al. (2015) WSL (2017)

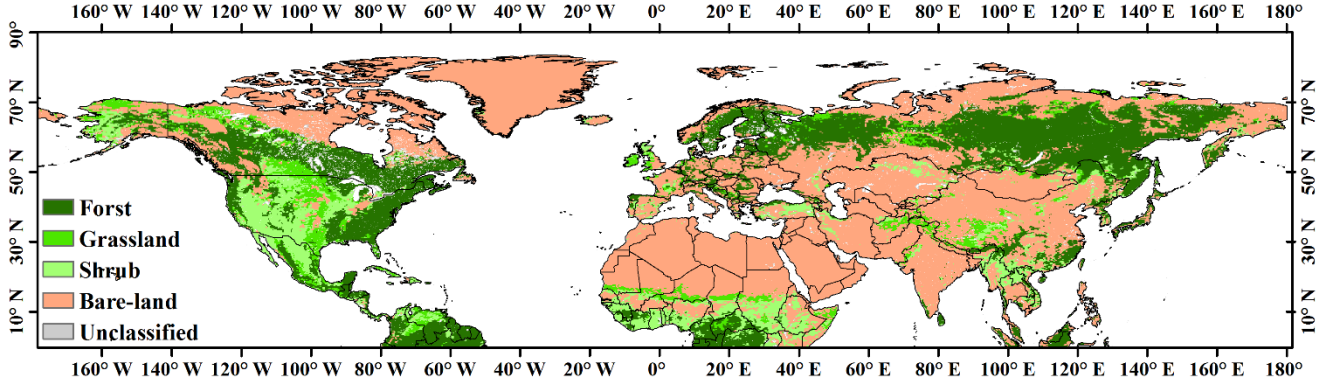


Figure A1. Land cover classification of the Northern Hemisphere based on GlobCover 2009. The land cover types were reclassified into five categories: forest, grassland, shrub, bare-land, and unclassified.

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

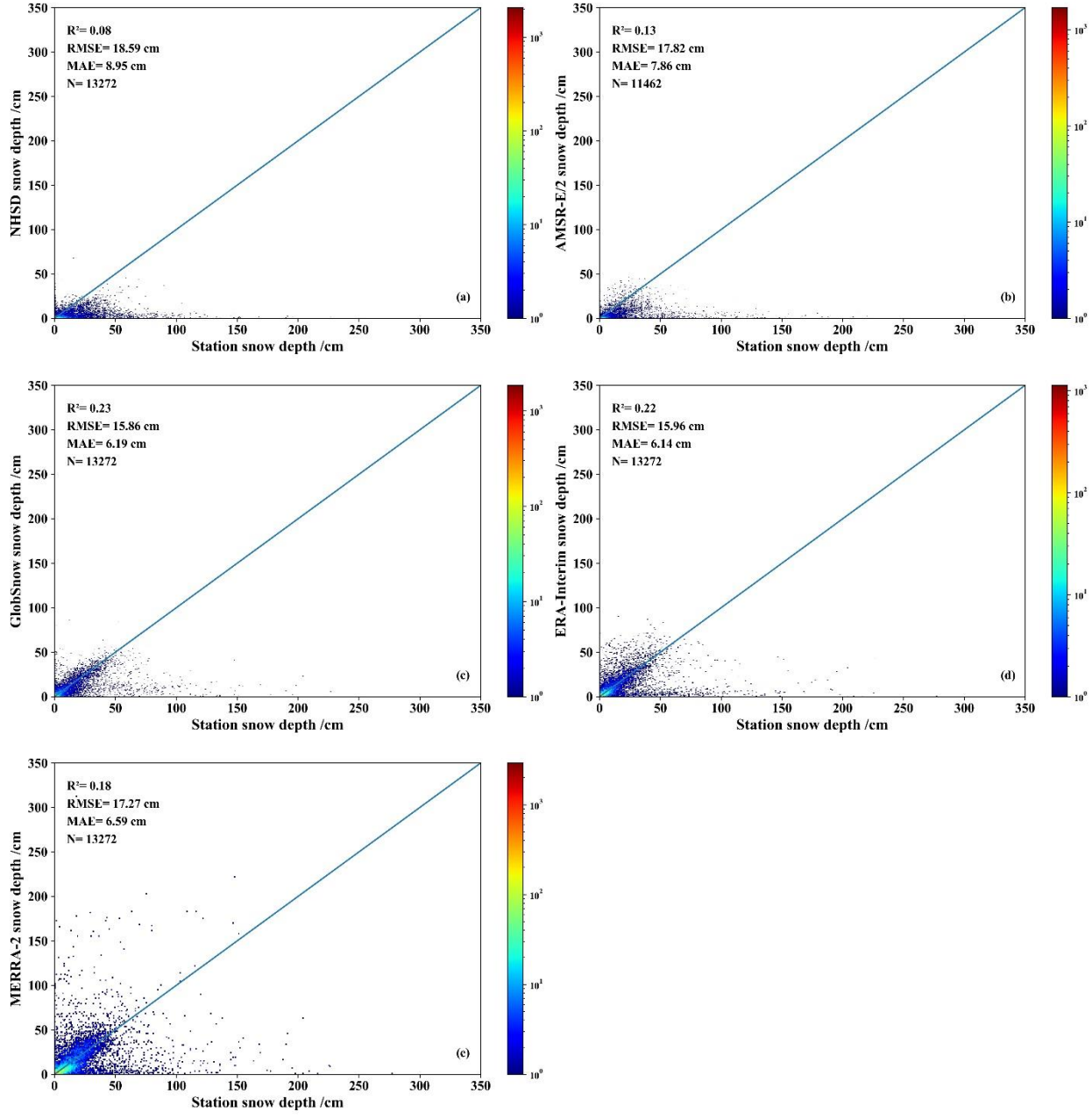


Figure A2. Scatterplots of the original gridded snow depth values with selected in situ observations. (a) NHSD snow depth product, (b) AMSR-E/2 snow depth product, (c) GlobSnow snow depth, (d) ERA-Interim snow depth product, and (e) MERRA-2 snow depth product.

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

References

545 Barnett, T. P., Adam, J. C., and Lettenmaier, D. P.: Potential impacts of a warming climate on water availability in snow-dominated regions, Nature, 438, 303-309, https://doi.org/10.1038/nature04141, 2005.

Bartlett, P. A., MacKay, M. D., and Verseghy, D. L.: Modified Snow Algorithms in the Canadian Land Surface Scheme: Model Runs and Sensitivity Analysis at Three Boreal Forest Stands, Atmos. Ocean, 43, 207–222, https://doi.org/10.3137/ao.440301, 2006.

550 Belgiu, M., and Drăguț, L.: Random forest in remote sensing: A review of applications and future directions, ISPRS J. Photogramm. Remote Sens., 114, http://dx.doi.org/10.1016/j.isprsjprs.2016.01.011, 2016.

Bormann, K.J, Brown, R.d, Derksen, C., and Painter, T.H: Estimating snow-cover trends from space, Nat. Clim. Chang., 8, 924-928, https://doi.org/10.1038/s41558-018-0318-3, 2018.

Broxton, P. D, Zeng, X., and Dawson, N.: Why do global reanalyses and land data assimilation products underestimate snow
555 water equivalent, Journal. Hydrometeorol., https://doi.org/10.1175/JHM-D-16-0056.1, 2016.

Cao, Y., Yang, X., and Zhu, X.: Retrieval snow depth by artificial neural network methodology from integrated AMSR-E and In-situ data—A case study in Qinghai-Tibet Plateau, Chin. Geogra. Sci., 18, 356-360, https://doi.org/10.1007/s11769-008-356-2, 2008.

Chang, A. T. C., Foster, J. L., and Hall, D. K.: Nimbus-7 SMMR derived global snow cover parameters, Ann. Glaciol., 9,
560 39-44, 1987.

Che, T., Li, X., and Dai, L.: Long-term series of daily global snow depth (1979-2017), National Tibetan Plateau Data Center, https://doi.org/10.11888/Snow.tpdc.270925, 2019.

Che, T., Hu, Y., Dai, L., and Xiao. L.: Long-term series of daily snow depth dataset over the Northern Hemisphere based on machine learning (1980-2019), National Tibetan Plateau Data Center, https://doi.org/10.11888/Snow.tpdc.271701, 2021.

565 Che, T., Li, X., Jin, R., Armstrong, R., and Zhang, T.: Snow depth derived from passive microwave remote-sensing data in China, Ann. Glaciol., 49, 145-154, 2008.

Danielson, J. J., and Gesch, D.B.: Global multi-resolution terrain elevation data 2010 (GMTED 2010), US Geological Survey: Reston, VA, USA, 2011.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo,
570 G., Bauer, P., Bechtold, P., Beljaars, A. C. M., Berg, L. v d., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kallberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. -J., Park, B. -K., Peubey, C., Rosnay, P. d, Tavolato, C., Thépaut, J. -N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc, 137, 553-597, https://doi.org/10.1002/qj.828, 2011.

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

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

Evora, N. D., Tapsoba, D., and Sève, D. D.: Combining artificial neural network models, geostatistics, and passive microwave data for snow water equivalent retrieval and mapping, IEEE T. Geosci. Remote, 46, https://doi.org/10.1109/TGRS.2008.916632, 2008.

Essery, R., Kontu, A., Lemmetyinen, J., Dumont, M., and Ménard, C. B.: A 7-year dataset for driving and evaluating snow models at an Arctic site (Sodankylä, Finland), Geosci. Instrum. Method. Data Syst., 5, 219–227, https://doi.org/10.5194/gi-5-219-2016, 2016.

Gelaro, R., McCarty, W., Suarez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., Silva, A. d., Gu, W., Kim, G. K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J Climate, 30, 5419-5454, https://doi.org/10.1175/Jcli-D-16-0758.1, 2017.

Hu, Y., Che, T., Dai, L., and Xiao, L.: Snow depth fusion based on machine learning methods for the Northern Hemisphere, Remote Sens., 13, 1250, https://doi.org/10.3390/rs13071250, 2021.

Kelly, R.: The AMSR-E snow depth algorithm: description and initial results, J. Remote Sens. Soc. Jpn., 29, 307-317, 2009.

Karniadakis, G. E., KevrekidiS, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang L.: Physics-informed machine learning, Nat. Rev. Phys., https://doi.org/10.1038/s42254-021-00314-5, 2021.

Krinner, G., Derksen, C., Essery, R., Flanner, M., Hagemann, S., Clark, M., Hall, A., Rott, H., Brutel-Vuilmet, C., Kim, H., Ménard, C. B., Mudryk, L., Thackeray, C., Wang, L., Arduini, G.,Balsamo, G., Bartlett, P., Boike, J., Boone, A., Chéruy, F., Colin,J., Cuntz, M., Dai, Y., Decharme, B., Derry, J., Ducharne, A.,Dutra, E., Fang, X., Fierz, C., Ghattas, J., Gusev, Y., Haverd, V.,Kontu, A., Lafaysse, M., Law, R., Lawrence, D., Li, W., Marke,T., Marks, D., Ménégoz, M., Nasonova, O., Nitta, T., Niwano,M., Pomeroy, J., Raleigh, M. S., Schaedler, G., Semenov, V.,Smirnova, T. G., Stacke, T., Strasser, U., Svenson, S., Turkov,D., Wang, T., Wever, N., Yuan, H., Zhou, W., and Zhu, D.: ESM-SnowMIP: assessing snow models and quantifying snow-related climate feedbacks, Geosci. Model Dev., 11, 5027–5049, https://doi.org/10.5194/gmd-11-5027-2018, 2018.

Landry, C. C., Buck, K. A., Raleigh, M. S., and Clark, M. P.: Mountain system monitoring at Senator Beck Basin, San Juan Mountains, Colorado: A new integrative datasource to develop and evaluate models of snow and hydrologic processes, Water Resour. Res., 50, 1773–1788, https://doi.org/10.1002/2013WR013711, 2014.

Larue, F., Royer, A., Sève, D. D., Langlois, A., Roy, A., and Brucker, L.: Validation of GlobSnow-2 snow water equivalent over Eastern Canada, Remote Sens. Environ., 194, 264-277, http://dx.doi.org/10.1016/j.rse.2017.03.027, 2017.

Liang, J., Liu, X., Huang, K., Li, X., Shi, X., Chen, Y., and Li, J.: Improved snow depth retrieval by integrating microwave brightness temperature and visible/infrared reflectance, Remote Sens. Environ., 156, https://dx.doi.org/10.1016/j.rse.2014.10.016, 2015.

Lievens, H., Demuzere, M., Marshall, H.-P., Reichle, R. H., Brucker, L., Brangers, I., Rosnay, P.d., Dumont, M., Girotto, M., Immerzeel, W.W., Jonas, T., Kim, E.J., Koch, I., Marty, C., Salornata, T., Schöber, J., and Lannoy, G.J.M.D.: Snow depth

variability in the Northern Hemisphere mountains observed from space, Nat. Commun., 10, 1-12, https://doi.org/10.1038/s41467-019-12566-y, 2019.

Luojus, K., Pulliainen, J., Takala, M., Lemmetyinen, J., Mortimer, C., Derksen, C., Mudryk, L., Moisander, M., Hiltunen, M., Smolander, T., Ikonen, J., Cohen, J., Salminen, M., Norberg, J., Veijola, K., and Venäläinen, P.: GlobSnow v3.0 Northern Hemisphere snow water equivalent dataset, Sci. Data, 8, 163, https://doi.org/10.1038/s41597-021-00939-2, 2021.

Ménard, C. B., Essery, R., Barr, A., Bartlett, A., Derry, J., Dumont, M., Fierz, C., Kim, H., Kontu, A., Lejeune, Y., Marks, D., Niwano, M., Raleigh, M., Wang, L., and Wever, N.: Meteorological and evaluation datasets for snow modelling at 10 reference sites: description of in situ and bias-corrected reanalysis data, Earth Syst. Sci. Data, 11, 865-880, https://doi.org/10.5194/essd-11-865-2019, 2019.

Mortimer, C., Mudryk, L., Derksen, C., Luojus, K., Brown, R., Kelly, R., and Tedesco, M.: Evaluation of long-term Northern Hemisphere snow water equivalent products, The Cryosphere, 14, 1579-1594, https://doi.org/10.5194/tc-14-1579-2020, 2020.

Mudryk, L. R, Derksen, C., Kushner, P. J., and Brown, R.: Characterization of Northern Hemisphere snow water equivalent datasets, 1981-2010, J Climate, 28, 8037-8051, https://doi.org/10.1175/JCLI-D-15-0229.1, 2015.

Ntokas, K. F. F., Odry, J., Boucher, M. -A., and Garnaud, C.: Investigating ANN architectures and training to estimate snow water equivalent from snow depth. Hydrol. Earth Syst. Sci., 25, 3017-3040, https://doi.org/10.5194/hess-25-3017-2021, 2021.

IPCC. IPCC Special Report on the Ocean and Cryosphere in a Changing Climate. https://www.google.com/search?sxsrf=ALeKk02egQANAcAkBMg_uQmHCmSrkkQ48g:1621681027673&q=Geneva&stick=H4sIAAAAAAAAAOPgEQz9U3MC5PK1aCsCwNjLSMMsqt9JPzc3JSk0sy8_P084vSE_MyqxJBnGKrjNTElMLSxKKS1KJihZz8ZLDwIlY299S81LLEHayMAE4-RWAAAA&sa=X&ved=2ahUKEwjhld_MkN3wAhUz7XMBHe26A70QmxMoATAzegQIJhAD.

Parker W.: Reanalyses and observations: What's the difference?, Bull. Amer. Meteor. Soc., 97, 1565-1572, https://doi.org/10.1175/BAMS-D-14-00226.1, 2016..

Pritchard, H.: Global data gaps in our knowledge of the terrestrial cryosphere, Front. Clim., 3: 689823. doi:10.3389/fclim.2021.689823, 2021.

Pulliainen, J., Luojus, K., Derksen, C., Mudryk, L., Lemmetyinen, J., Salminen, M., Ikonen, J., Takala, M., Cohen, J., Smolander, T., and Norberg, J.: Patterns and trends of Northern Hemisphere snow mass from 1980 to 2018, Nature, 581, 294-198, https://doi.org/10.1038/s41586-020-2258-0, 2020.

Santi, E., Brogioni, M., Leduce-Leballeur, M., Macekkoni, G., Montomoli, F., Pampaloni, P., Lemmetyinen, J., Cohen, J., Rott, H., Nagler, T., Derksen, C., King, J., Rutter, N., Essery, R., Menard, C., Sandells, M., and Kern, M.: Exploiting the ANN potential in estimating snow depth and snow water equivalent from the airborne SnowSAR data at X and Ku bands, IEEE T. Geosci. Remote., https://doi.org/10.1109/TGRS.2021.3086893, 2021.

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

Snauffer, A. M., Hsieh, W. W., Cannon, A. J., and Schnorbus, M. A: Improving gridded snow water equivalent products in British Columbia, Canada: multi-source data fusion by neural network models, The Cryosphere, https://doi.org/10.5194/tc-12-891-2018, 2018.

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J.P., 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 Sens. Environ., 115, https://doi.org/10.1016/j.rse.2011.08.014, 2011.

Tedesco, M., and Jeyaratnam, J.: A new operational snow retrieval algorithm applied to historical AMSR-E brightness temperatures, Remote Sens., https://doi.org/10.3390/rs8121037, 2016.

Tedesco, M., Pulliainen, J., Takala, M., Hallikainen, M., and Pampaloni, P.: Artificial neural network-based techniques for the retrieval of SWE and snow depth from SSM/I data, Remote Sens. Environ., https://doi.org/10.1016/j.rse.2003.12.002, 2004.

Thackeray, C. W., Derksen, C., Fletcher, C. G, and Hall, Alex.: Snow and climate: feedbacks, drivers, and indices of change, Curr. Clim. Change Rep., 5, https://doi.org/10.1007/s40641-019-00143-w, 2019.

Venäläinen, P., Luojus, K., Lemmetyinen, J., Pulliainen, J., Moisander, M., and Takala, M.: Impact of dynamic snow density on GlobSnow snow water equivalent retrieval accuracy, The Cryosphere, 15, 2969-2981, https://doi.org/10.5194/tc-15-2969-2021, 2021.

Wang, J., Yuan, Q., Shen, H., Liu, T., Li, T., Yue, L., Shi, X., and Zhang, L.: Estimating snow depth by combining satellite data and ground-based observations over Alaska: A deep learning approach, J. Hydrol., 585, https://doi.org/10.1016/j.jhydrol.2020.124828, 2020.

Wever, N., Schmid, L., Heilig, A., Eisen, O., Fierz, C., and Lehning, M.: Verification of the multi-layer SNOWPACK model with different water transport schemes, The Cryosphere, 9, 2271–2293, https://doi.org/10.5194/tc-9-2271-2015, 2015.

WSL Institute for Snow and Avalanche Research SLF: Weissfluhjoch dataset for ESM-SnowMIP, WSL Institute for Snow and Avalanche Research SLF. https://doi.org/10.16904/16, 2017.

Xiao, L., Che T., and Dai, L.: Evaluation of remote sensing and reanalysis snow depth datasets over the Northern Hemisphere during 1980-2016, Remote Sens., 12, 3253, https://doi.org/10.3390/rs12193253, 2020.

Xiao, X., Zhang, T., Zhong, X., Shao, W., and Li, X.: Support vector regression snow-depth retrieval algorithm using passive microwave remote sensing data, Remote Sens. Environ., 210, https://doi.org/10.1016/j.rse.2018.03.008, 2018.

Yang, J., Jiang, L., Luojus, K., Pan, J., Lemmetyinen, J., Takala, M., and Wu, S.: Snow depth estimation and historical data reconstruction over China based on a random forest machine learning approach, The Cryosphere, 14, 1763-1778, https://doi.org/10.5194/tc-14-1763-2020, 2020.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., and Zhang, L.: Deep learning in environmental remote sensing: Achievements and challenges, Remote Sens. Environ., https://doi.org/10.1016/j.rse.2020.111716, 2020.

https://doi.org/10.5194/essd-2022-63
Preprint. Discussion started: 28 March 2022

Zhang, B., Benediktsson, J. A., Liu, B., Zou, L., Li, J., and Plaza, A.: Remotely sensed big data: evolution in model development for information extraction, Proc. IEEE., 107, 2294-2301, https://doi.org/10.1109/IPROC.2019.2948454, 2019.

680 Zhu, J., Tan, S., Tsang, L., Kang, D-H., and Kim, E.: Snow water equivalent retrieval using active and passive microwave observations, Water Resour. Res., 57, https://doi.org/10.1029/2020WR027563, 2021.