AGU ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2021WR030895

Uncertainty Characterization of Ground-Based, Satellite, and Reanalysis Snow Depth Products Using Extended Triple Collocation

Dejing Qiao[1,2], Zhen Li[3], Jiangyuan Zeng[4], Shuang Liang[3], Kaighin A. McColl[5], Haiyun Bi[6], Jianmin Zhou[3], and Ping Zhang[3]

[1]College of Surveying and Geo-Informatics, North China University of Water Resources and Electric Power, Zhengzhou, China, [2]College of Geology and Environment, Xi'an University of Science and Technology, Xi'an, China, [3]Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, [4]State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, [5]Department of Earth and Planetary Sciences, School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA, [6]State Key Laboratory of Earthquake Dynamics, Institute of Geology, China Earthquake Administration, Beijing, China

Key Points:

- The ground-based analysis Canadian Meteorological Centre product generally performs the best among three snow depth products in the northern hemisphere
- The skill of three snow depth products is significantly affected by forest cover fraction and surface roughness

Correspondence to:

J. Zeng,
zengjy@radi.ac.cn

Citation:

Qiao, D., Li, Z., Zeng, J., Liang, S., McColl, K. A., Bi, H., et al. (2022). Uncertainty characterization of ground-based, satellite, and reanalysis snow depth products using extended triple collocation. *Water Resources Research*, *58*, e2021WR030895. https://doi.org/10.1029/2021WR030895

Received 23 JUL 2021
Accepted 30 MAR 2022

Author Contributions:

Conceptualization: Dejing Qiao, Jiangyuan Zeng
Formal analysis: Zhen Li, Shuang Liang, Kaighin A. McColl, Haiyun Bi, Jianmin Zhou, Ping Zhang
Investigation: Zhen Li, Shuang Liang, Kaighin A. McColl, Haiyun Bi, Jianmin Zhou, Ping Zhang
Methodology: Dejing Qiao, Jiangyuan Zeng
Resources: Zhen Li, Jiangyuan Zeng
Supervision: Zhen Li, Jiangyuan Zeng
Validation: Dejing Qiao, Jiangyuan Zeng
Writing – review & editing: Dejing Qiao, Jiangyuan Zeng

Abstract The optimal use of hemispheric-scale snow depth (SD) products for various hydrometeorological applications requires a comprehensive assessment of their quality. Most previous validation studies of SD products adopted in situ observations as the ground truth, which may cause representativeness errors due to spatial scale mismatch between point-based ground SD measurements and grid-based SD products. The extended triple collocation (ETC) technique is a powerful tool to estimate the uncertainty of three independent data sets without assuming any one data source is an error-free "truth" reference. This study first used the ETC to assess the uncertainty of three types of hemispheric-scale SD products, including the ground-based analysis Canadian Meteorological Centre (CMC), the satellite-based Advanced Microwave Scanning Radiometer 2 (AMSR2), and the model-based Global Land Data Assimilation System (GLDAS) SD products. Furthermore, the uncertainties of each SD product were analyzed using ETC metrics, that is, the correlation coefficient (R) and error standard deviations (STDs), with respect to several environmental and perturbing factors. Overall, the CMC outperforms the AMSR2 and GLDAS, with a higher R and a smaller STD. Considering multiple environmental and perturbing factors, the poorest performance of the three SD products is mainly found in densely vegetated regions, and they are strongly related to the forest cover fraction and surface roughness. Despite the above factors, the best performance for all three SD products is found over temperate climate regions. The results demonstrate the usefulness of the ETC approach to quantify the uncertainty of SD products particularly in remote regions with sparse in situ measurements.

1. Introduction

Seasonal snow is a major water resource and meltwater from snowpack provides vital water supply around the world (Bormann et al., 2018; Tedesco & Narvekar, 2010). Due to its high solar reflectivity and thermodynamic properties, seasonal snow has substantial influences on the partitioning of energy between the land and atmosphere (Flanner et al., 2011; Lu et al., 2016). Moreover, snow cover has significant implications for the soil freeze-thaw process, soil carbon dioxide efflux, and alpine vegetation growth (Li et al., 2021; Zhang, 2005). A good knowledge of snowpack information is of great significance for various practical applications, such as hydropower, drought and flood predictions, and irrigation scheduling at regional and hemispheric scales (Clark et al., 2011; Dai et al., 2012; Han et al., 2019).

Passive microwave technology has become one of the most viable options for monitoring snow depth (SD) at continental scales due to its high sensitivity to snow permittivity and its 24-hr all-weather coverage capability (Clifford, 2010). In the last few decades, a number of activities have been devoted to obtaining SD by using multichannel and multiband passive microwave instruments (Armstrong & Brodzik, 2002; Chang et al., 1987). The Advanced Microwave Scanning Radiometer 2 (AMSR2) on board the Global Change Observation Mission 1-Water (GCOM-W1) is the successor of the AMSR-E (Cho et al., 2017; Imaoka et al., 2012). Although the current AMSR2 SD algorithm is widely used to estimate global SD, its performance has not been fully evaluated on a hemispherical scale. Understanding the skill and possible error sources of the AMSR2 SD product is

crucial for their utilization in practical applications, and also improves our knowledge on how they can be further improved. In addition to estimating SD from satellites, in situ measurements and land surface models (LSMs) can also provide useful SD information. In situ SD measurements are able to well represent the true SD values at the point scale; therefore, they are usually used for the evaluation of both satellite-based and model-simulated SD data sets. However, in situ SD measurements have limitations in the large-scale spatial representation, particularly for hemispheric-scale data analysis. In response to this challenge, some studies have attempted to use various interpolation techniques to obtain spatial SD from in situ measurements (Erxleben et al., 2002; Hudson & Wackernagel, 2010), but these interpolation approaches are negatively affected by the sparse in situ measurements, especially in high-altitude areas (Dyer & Mote, 2006; Kitaev et al., 2002). The Canadian Meteorological Centre (CMC) generates a daily gridded Northern Hemisphere SD data set, integrating all available ground measurements with a simple snow model (Rutter et al., 2008) to offset the effect of sparse in situ observations. However, this product is still limited by ground observations and complex terrain, and its uncertainties remain to be ascertained. LSMs are considered as another promising tool for producing long-term SD data, and the generation mechanisms are usually based on physical equations that estimate the mass and energy balance among the snow, soil, and atmosphere (Davis et al., 1995). Many LSMs and reanalysis SD data sets are now available, including the Global Land Data Assimilation System (GLDAS; Rodell et al., 2004), the NASA Modern-Era Retrospective Analysis for Research and Applications (MERRA2; 0.625° × 0.5°; Gelaro et al., 2017), the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Reanalysis (ERA-Interim; 0.75° × 0.75°; Dee et al., 2011), and the Japanese 55-year Reanalysis (JRA-55; 0.55° × 0.55°; Kobayashi et al., 2015). GLDAS-2.1 is a newly available GLDAS data set with a high grid resolution (0.25° × 0.25°). The grid resolution of GLDAS-2.1 is consistent with that of the commonly used satellite SD products (e.g., AMSR2). However, uncertainties from the model structure and atmospheric forcing data may introduce errors to the LSM simulations such as the GLDAS SD. All of the above-mentioned products are obtained from measurement systems, such as LSMs, satellite and ground-based observation networks, which are susceptible to numerous error sources, including the difference between the point-scale SD ground observations and the larger-scale SD quantity of relevance to models and satellite observations, sensor calibration, and underlying model assumptions and parametrization. Therefore, characterizing the uncertainty of different SD data sets is critical for their improvement and application in scientific studies.

During the past few decades, extensive validation activities have been carried out to examine the accuracy of satellite-based and model-based SD products at various spatiotemporal scales (Broxton et al., 2016; Kim et al., 2021; Mortimer et al., 2020; Mudryk et al., 2015), and they have mostly used in situ observations as the true values (Larue et al., 2017; Liu et al., 2014). However, assessing the skill of grid-scale satellite- and model-based SD products by using single-point ground SD measurements is challenging due to the inherent spatial mismatch between them as well as the possible errors of the in situ SD data. The triple collocation (TC) approach provides a powerful solution for estimating errors in three independent data sets without requiring the true value (Stoffelen, 1998). It has been widely used to characterize uncertainties in many geophysical parameter products, including soil moisture (Chen et al., 2018), precipitation (Li et al., 2018), sea surface temperature (O'Carroll et al., 2008), land surface freeze/thaw state (Lyu et al., 2018; McColl et al., 2016), and wave height (Liu et al., 2016). More recently, McColl et al. (2014) introduced a variant of TC called the extended triple collocation (ETC). This technique derives an additional performance metric (the temporal correlation) making this approach more appropriate for estimating errors in the geophysical parameters. Although TC/ETC has been recently widely used for characterizing uncertainties in many geophysical parameter products, to the best of our knowledge, it has not been applied to evaluate snow variables (e.g., SD) at a continental scale. Moreover, the performance of the global SD products under different conditions (e.g., land cover, climate, forest cover fraction, snow class, surface roughness, and land cover heterogeneity) has not yet been comprehensively investigated, but is believed to be very helpful for their utilization and improvement. This study gives the first attempt to apply ETC to evaluate the performance of three different types of SD products, namely ground-based analysis SD (CMC), satellite-based SD (AMSR2), and model-simulated SD (GLDAS), on a hemispheric scale, and examines the error characteristics and limitations of these products under various land surface conditions.

The objectives of this study are: (a) investigate hemispheric-scale error patterns of the CMC, AMSR2, and GLDAS using the ETC method; and (b) analyze the skill of the three SD data sets under diverse conditions, including land cover classification, forest cover fraction, snow class, surface roughness, land cover heterogeneity, and climate type. This paper is organized as follows: Section 2 briefly introduces the three SD products from

Table 1
Main Characteristics of the CMC, AMSR2, and GLDAS SD Products

Data sets	Approach	Spatial resolution	Temporal coverage	Reference
CMC	In situ, interpolation, and melt model	24×24 km^2	1998.08–2018.12	Brasnett (1999);Brown et al. (2003)
AMSR2	Snow and brightness temperature difference (TBD) algorithm	$0.25° \times 0.25°$	2012.07–present	Kelly (2009)
GLDAS	Noah-LSMs	$0.25° \times 0.25°$	2000.01–present	Rodell et al. (2004)

ground-based analysis, satellite, and reanalysis SD products as well as the ancillary data. Section 3 describes the data-processing procedures and ETC method. Section 4 presents the results concerning the evaluation of CMC, AMSR2, and GLDAS SD via ETC. The discussions are presented in Section 5. The main conclusions are provided in Section 6.

2. Data

Table 1 presents the main information of the three SD products (CMC, AMSR2, and GLDAS) used in this study. Due to the nature of radiometer observations, AMSR2 products are reliable only in regions with dry snow cover (Stiles & Ulaby, 1980; Walker & Goodison, 1993). Accordingly, the evaluation was performed for the winter season (December–February) from 2012 to 2017.

2.1. Canadian Meteorological Centre

The CMC daily SD analysis data set is generated using in situ daily SD measurements and optimal interpolation with a first-guess field generated from a simple snow accumulation and melt model (Brown et al., 2003). This product uses the statistical interpolation method to obtain spatial SD from in situ observations, including SD from surface synoptic observations, meteorological aviation reports, and special aviation reports from the World Meteorological Organization (WMO) information system. Over regions with sparse in situ observations, a simple snow model is used to generate an optimal interpolation with a first-guess field, which uses precipitation forecasts and screen-level temperature analysis to estimate snowfall and snowmelt for a global domain and assumes persistence of the mass of the snowpack between melting and/or snowfall events (Brasnett, 1999). This data set is available from the National Snow and Ice Data Center (https://nsidc.org/data/NSIDC-0447/versions/1).

2.2. Advanced Microwave Scanning Radiometer 2

The AMSR2 sensor is the successor of the AMSR-E, and it is onboard the JAXA GCOM-W1 satellite launched on May 18, 2012 (Okuyama & Imaoka, 2015). It provides brightness temperature (T_B) ranging from 6.9 to 89.0 GHz at horizontal and vertical polarizations. The current AMSR2 SD algorithm is an evolution of the original AMSR-E SD algorithm (Kelly, 2009), and it is based on the frequency difference method proposed by Chang et al. (1996). The AMSR2 SD algorithm undertakes a forest correction to reduce the uncertainties in the estimates of SD in forested areas. It estimates shallow snow and moderate snow accumulation using the 89 and 37 GHz channels, respectively. For deep snow, the T_B at 19 and 10 GHz are used to retrieve SD (Derksen, 2008). The JAXA provides SD products derived from AMSR2 with two grid resolutions (0.10° and 0.25°), and they can be obtained from the Globe Portal System (G-Portal; https://gportal.jaxa.jp/). To make a consistent and fair comparison, the AMSR2 SD product at 0.25° was used, and its spatial resolution was the same as that of GLDAS.

2.3. Global Land Data Assimilation System

GLDAS 2.1 is the latest version of GLDAS. GLDAS is a combination of LSM and data assimilation techniques used to generate land surface states and fluxes. GLDAS is forced by satellite and ground-based observational data products (Rodell et al., 2004). The GLDAS 2.1 simulation, covering 2000 to the present, is forced with the National Oceanic and Atmospheric Administration (NOAA)/Global Data Assimilation System (GDAS) atmospheric analysis fields (Derber et al., 1991), the spatially and temporally disaggregated Global Precipitation Climatology Project (GPCP) precipitation fields (Adler et al., 2003), and the Air Force Weather Agency's

AGRicultural METeorological modeling system (AGRMET) radiation fields. The SD data set used in this study was from GLDAS 2.1 Noah LSMs, and its spatial and temporal resolution is 0.25° and 3-hr, respectively (see Table 1). This data set is released by the NASA's Hydrology Data and Information Services Center (https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS/GLDAS_NOAH025_3H.2.1/).

2.4. Ancillary Data

To examine the influences of heterogeneous land cover types, the MODIS IGBP land cover classification data set (Friedl et al., 2002), with a spatial resolution of 500 m was used. This study considers only seven land cover classes, including forest, shrublands, savannas, grasslands, croplands, barren, and others. Considering the effect of surface roughness on the snow-depth retrieval modeling, the Global Land One-kilometre Based Elevation Digital Elevation Model (GLOBE DEM) data was adopted to characterize the surface roughness effects (Hastings & Dunbar, 1998). In addition, the snow-climate classification data set with a spatial resolution of 0.5° (Sturm et al., 1995) was used to assess the skill of SD data sets under various snow properties, which has been widely applied in many previous studies (Liu et al., 2014; Sturm et al., 2010). More details of this data set can be found in Sturm et al. (1995). Furthermore, to examine the skill of SD data sets under various climate conditions, the Koeppen-Geiger climate classification data set (Rubel & Kottek, 2010) was used.

3. Methodology

3.1. Data Pre-Processing

Since the CMC, AMSR2, and GLDAS data sets are distributed on different grids, the nearest neighbor approach was used to re-project the CMC SD data sets onto a uniform georeferenced 0.25° × 0.25° grid, the same as that of AMSR2 and GLDAS SD product (Draper et al., 2013; Zeng et al., 2015). Compared with other resampling methods (e.g., bilinear interpolation, cubic convolution, and majority resampling), the advantage of the nearest neighbor interpolation method is that it is the fastest resampling method and minimizes changes to pixel values since no new values are created, and it has been widely used in the resampling of SD data (Liu et al., 2013; Xiao et al., 2020) as well as other geophysical parameter products, such as soil moisture (Zeng et al., 2020) and snow cover (Drusch et al., 2004). In addition, melt water in the snowpack can result in an increase of T_B, particularly at frequencies above 30 GHz (Foster et al., 2005), leading to the difficulty in extracting SD information by microwave radiation under wet snow conditions. To minimize these effects, only nighttime data (e.g., AMSR2 descending data and GLDAS data obtained from the average values at 00:00, 03:00, and 06:00) in the winter season (December–February) were selected in this study (Dai et al., 2017; Mudryk et al., 2015). In this study, we used a spatial averaging method to aggregate the 500 m MODIS IGBP land cover to the SD coarse resolution of 0.25°, and then the dominant land cover within each 0.25° grid cell was adopted as the land cover for that grid cell. The original spatial resolution of the snow-climate classification data set and Koeppen-Geiger climate classification data set is 0.5°, which were resampled to 0.25° by using the nearest neighbor interpolation method. This study investigated only seasonal snow over land and the regions of open water and permanent snow and ice were excluded based on the MODIS IGBP classification. Moreover, there are no snow values in the tropical climate regions, and therefore these regions are also excluded in the study.

3.2. Extended Triple Collocation

We used ETC to assess the error structure of the three independent SD data sets (CMC, AMSR2, and GLDAS). TC is a statistical method that estimates the standard deviation (STD) (Stoffelen, 1998) and correlation coefficient (McColl et al., 2014) without requiring the assumption that a given reference data set is free of errors (Gruber et al., 2016; Khan et al., 2018; Larue et al., 2017; McColl et al., 2014).

We briefly review the derivation of the ETC estimation equations provided in McColl et al. (2014). First, ETC assumes an error model, and its standard form is as follows:

$$X_i = \alpha_i + \beta_i T + \varepsilon_i \tag{1}$$

where X_i is a measurement from the ith SD products ($i = 1$ is AMSR2, $i = 2$ is GLDAS, and $i = 3$ is CMC). T represents the true SD value. α_i and β_i are the systematic additive and multiplicative biases of ith data set with

respect to the true value, respectively. ε_i represents the zero-mean random error. This is a standard error model that is often used (sometimes implicitly) in validation studies that do not use TC or ETC. Note that the error model does not assume that errors have zero mean, since it includes the term α_i, which does not need to be zero. Equation 1 is identical to the error model $X_i = \beta_i T + \hat{\varepsilon}_i$ where $\hat{\varepsilon}_i$ represents random error with non-zero mean. This model is also capable of representing systematic spatial differences in errors (e.g., differences in errors between land cover types or climate regimes).

The covariance between different SD products is:

$$\begin{aligned} Cov(X_i, X_j) &= E(X_i X_j) - E(X_i)\,E(X_j) \\ &= \beta_i \beta_j \sigma_T^2 + \beta_i Cov(T, \varepsilon_j) + \beta_j Cov(T, \varepsilon_i) + Cov(\varepsilon_i, \varepsilon_j) \end{aligned} \tag{2}$$

where $\sigma_T^2 = Var(T)$. It is assumed that the errors from the independent sources are uncorrelated with each other ($Cov(\varepsilon_i, \varepsilon_j) = 0, i \neq j$) and the true value ($Cov(T, \varepsilon_i) = 0$). While these assumptions are unlikely to be exactly correct, they are not unique to TC validation studies, and will be discussed in more detail below. Thus, Equation 2 can be reduced to:

$$C_{ij} = Cov(X_i, X_j) = \begin{cases} \beta_i \beta_j \sigma_T^2 (i \neq j) \\ {\beta_i}^2 \sigma_T^2 + \sigma_{\varepsilon_i}^2 (i = j) \end{cases} \tag{3}$$

We now have six equations, given by the six unique terms in the 3×3 covariance matrix (C_{11}, C_{12}, C_{13}, C_{22}, C_{23}, C_{33}), but more than six unknowns ($\beta_1, \beta_2, \beta_3, \sigma_{\varepsilon_1}, \sigma_{\varepsilon_2}, \sigma_{\varepsilon_3}, \sigma_T$). Therefore, the system cannot be solved for the unknowns in its current form. However, by defining $\theta_i = \beta_i \sigma_T$, Equation 3 can be rewritten as:

$$C_{ij} = \begin{cases} \theta_i \theta_j (i \neq j) \\ {\theta_i}^2 + \sigma_{\varepsilon_i}^2 (i = j) \end{cases} \tag{4}$$

We now have six equations and six unknowns ($\theta_1, \theta_2, \theta_3, \sigma_{\varepsilon_1}, \sigma_{\varepsilon_2}, \sigma_{\varepsilon_3}$), meaning the system can be solved for the unknowns. Note that this transformation allows us to solve for θ_i and σ_{ε_i}, but not for β_i and σ_T (which have both been absorbed into the combined term θ_i), and not for α_i (which cancel out and do not appear in the above equation). The resulting expression for the STD (i.e., σ_ε) is:

$$\sigma_\varepsilon = \begin{bmatrix} \sqrt{C_{11} - \frac{C_{12} C_{13}}{C_{23}}} \\ \sqrt{C_{22} - \frac{C_{12} C_{23}}{C_{13}}} \\ \sqrt{C_{33} - \frac{C_{13} C_{23}}{C_{12}}} \end{bmatrix} \tag{5}$$

From Equation 5, using the definition of the correlation and covariance, θ_i can be obtained:

$$\theta_i = R_{T,Xi} \sqrt{C_{ii}} \tag{6}$$

The correlation coefficient (R_{T,X_i}) between T and X_i can be solved from Equation 6; it can be shown that:

$$R_{T,X} = \pm \begin{bmatrix} \sqrt{\frac{C_{12} C_{13}}{C_{11} C_{23}}} \\ sign(C_{13} C_{23}) \sqrt{\frac{C_{12} C_{23}}{C_{22} C_{13}}} \\ sign(C_{12} C_{23}) \sqrt{\frac{C_{13} C_{23}}{C_{33} C_{12}}} \end{bmatrix} \tag{7}$$

The $R_{T,X}$ value denotes the temporal correlation between the product and the unknown truth, which refers to the commonly used Pearson's correlation coefficient (Beck et al., 2021; Chen et al., 2018; McColl et al., 2014). We

emphasize that while ETC is not able to directly estimate α_i and β_i, this does not prevent the estimation of R and STD.

A full derivation of the ETC estimation equations has been provided, including its assumptions; however, we emphasize that ETC does not make significant new assumptions compared with a more standard validation analysis (Gruber et al., 2016). For example, consider a study that does not use TC, but instead assesses the performance of a model of SD by estimating the correlation coefficient between the model's outputs and a ground truth measurement of SD, and interprets a higher correlation coefficient as meaning better model performance. Using the same error model for consistency (Equation 1), the estimated correlation coefficient $\hat{R}$ between the ground truth measurement ($k = 1$) and the model ($k = 2$) is then (Equation 9 of Gruber et al. (2016))

$$\hat{R} = \frac{C_{12}}{\sqrt{C_{11}C_{22}}} = \frac{\beta_1\beta_2\sigma_T^2 + \beta_2\mathrm{Cov}(T,\varepsilon_1) + \beta_1\mathrm{Cov}(T,\varepsilon_2) + \mathrm{Cov}(\varepsilon_1,\varepsilon_2)}{\sqrt{\left(\beta_1^2\sigma_T^2 + 2\beta_1\mathrm{Cov}(T,\varepsilon_1) + \sigma_{\varepsilon_1}^2\right)\left(\beta_2^2\sigma_T^2 + 2\beta_2\mathrm{Cov}(T,\varepsilon_2) + \sigma_{\varepsilon_2}^2\right)}} \tag{8}$$

If the ground truth measurement is assumed to be free of errors ($\alpha_1 = 0$, $\beta_1 = 1$, and $\sigma_{\varepsilon_1}^2 = 0$), then this implies that $\mathrm{Cov}(\varepsilon_1,\varepsilon_2) = \mathrm{Cov}(T,\varepsilon_1) = 0$, just as assumed in triple collocation. In addition, if a higher correlation coefficient is interpreted as meaning better model performance, then this interpretation requires the implicit assumption that $\mathrm{Cov}(T,\varepsilon_2) = 0$, just as in TC. If this assumption is not made, Equation 8 shows that a high correlation coefficient could simply be an artifact of strong correlation between errors in the model and the true SD (i.e., large $\mathrm{Cov}(T,\varepsilon_2)$). In summary, the error assumptions made in TC are not new, and are routinely made in validation studies that do not use TC.

The main challenge in implementing ETC, rather than a conventional validation analysis, is in finding a third estimate of the geophysical variable of interest (in this study, SD) that is relatively independent of the other two estimates. Fortunately for this study, three such estimates exist for SD, but this may not be the case for other geophysical variables.

3.3. Application of ETC

Before we use the ETC method to calculate the errors in each SD product, we discarded values where SD is zero for all three products (e.g., low-latitude snow-free regions such as the barren land region below 30°N). However, for midlatitude to high-latitude regions, if the value of a product is occasionally 0, while that of the other two products are not 0, we have also performed ETC method in this case.

Several assumptions are made in the ETC method in order to make the system solvable, which have been discussed in detail in previous studies (e.g., Gruber et al., 2016). As mentioned in the previous section, most of the assumptions discussed here are not unique to TC; they are also assumed in traditional validation analyses, albeit implicitly. As a prerequisite of ETC, three independent data sets should be prepared (corresponding to the assumption regarding zero correlation between the errors of the three data sets). The smaller the correlation between errors in different products, the smaller the error of ETC results (McColl et al., 2014; Yilmaz & Crow, 2014). Thus, the choice of data sets needs to be carefully considered before applying the ETC approach. The reasons of selecting three SD products in this study were as follows: first, three products need to meet the fundamental assumption of the ETC method that the three products should be independent. At present, the acquisition of SD data is mainly through in situ measurement, passive microwave remote sensing observation and model simulation. Therefore, though many SD products are available, three data sets including the ground-based analysis (CMC), the satellite-based (AMSR2), and the model-based (GLDAS) SD products were selected to form the triplets required by the ETC method in this study. Their measurement technique and retrieval concept are fundamentally different, which suggests that the selected data sets in this study are likely to meet the zero error correlation requirements of the ETC method (Dorigo et al., 2010; Scipal et al., 2008). For the assumption concerning zero correlation between errors and the true signal (i.e., error orthogonality), Yilmaz and Crow (2014) pointed out that the effect of violating this assumption on the error variance estimates is negligible, because error non-orthogonality is dampened by the application of rescaling parameters or even compensated if the magnitude of nonorthogonality is approximately the same for all considered data sets. In terms of assumptions concerning the stationarity of the signal, a violation of signal stationarity does not affect ETC analysis per se, as it considers temporally collocated triplets, which differ from a hypothetical stationary mean and variance by the same magnitude. A violation of

Table 2
The Surface Roughness and GSI Can Be Divided Into Five Groups (From I to V, With Increasing Intensity) (Ma et al., 2019)

DEM_SD value ($\log_e$)	Pixel type	GSI value	Pixel type	Climate classes	Snow classes
$Log(DEM_{SD})\in[-3.367,1.042)$	Rou-I	GSI∈[0,0.164]	GSI-I	Arid	Tundra
$Log(DEM_{SD})\in[1.042,2.832)$	Rou-II	GSI∈[0.164,0.304]	GSI-II	Temperate	Taiga
$Log(DEM_{SD})\in[2.832,4.223)$	Rou-III	GSI∈[0.304,0.444]	GSI-III	Cold	Maritime
$Log(DEM_{SD})\in[4.223,5.812)$	Rou-IV	GSI∈[0.444,0.585]	GSI-IV	Polar	Prairie
$Log(DEM_{SD})\in[5.812,7.203]$	Rou-V	GSI∈[0.585,0.878]	GSI-V		Alpine
					Ephemeral

Note. The climate is divided into four classes based on the Koeppen-Geiger climate classification, and the seasonal snow cover of the Northern Hemisphere is categorized into six classes based on the Sturm snow-climate classification system.

stationarity of the errors would not harm the reliability of the estimated average error variance per se (Crow et al., 2005). In the ETC method, it is assumed that the error of a single product from different ETC triplets is stable. That is to say, if changing one product in the ETC triplet, the error of another two products will remain consistent. This have been demonstrated by previous studies (e.g., Dorigo et al., 2010), and they found that the errors estimated for another two products are hardly influenced by the choice of the third independent data set. In terms of the zero-mean error (ε_i in Equation 1), it should be noted that ETC can only estimate the random error component, but not the biases with respect to the truth. Apart from the abovementioned assumptions, the ETC needs temporally collocated data of the same geophysical variable to derive the STD and correlation coefficient. That is to say, for a given grid at a specific time, the ETC was applied only when all three data sets (triplet) were available. Previous studies found that a minimum number of 100 triplets is sufficient to apply the ETC approach (Miralles et al., 2011; Scipal et al., 2008), and if there are not enough triplets, the ETC results will be biased (McColl et al., 2014). In our study, we used SD data sets during the winters from 2012 to 2017, which meets the number of triplets (>100) required for ETC calculation.

3.4. Proxies for Error Analysis

In this study, three proxies were adopted to investigate the influences of the forest cover fraction, surface roughness, and land cover heterogeneity on SD data sets. The forest cover fraction (*ff*) was calculated from the IGBP land cover data set, which was used to parameterize the effect of forest (Che et al., 2016; Lee et al., 2015). Note that other forest cover fraction products are also available to the public (e.g., Hansen et al., 2013). The Gini-Simpson index (GSI; Simpson, 1949) was calculated based on the land cover types from the IGBP land cover classification data sets, which was used to describe the land cover heterogeneity (Ma et al., 2019). The surface roughness within each pixel of the SD product was obtained from the GLOBE DEM (Kim et al., 2015; Ma et al., 2019). The three proxies can be obtained through:

$$ff = \sum_i f_i \tag{9}$$

$$Roughness = \log_e(DEM\_SD) \tag{10}$$

$$GSI = 1 - \sum_i p_i^2 \tag{11}$$

where f_i is the proportion of pixels belonging to the IGBP forest class (i.e., evergreen needle leaf, evergreen broadleaf, deciduous needle leaf, deciduous broadleaf, and mixed forest); *DEM_SD* is the STD of 625 1-km DEM cells; *Roughness* denotes the surface roughness proxy used in this study; p_i represents the proportion of pixels belonging to IGBP class *i*; and GSI denotes the landscape heterogeneity proxy. The categorized principles of *Roughness* and GSI are listed in Table 2. The main climate and snow classes, which are based on the Sturm snow-climate classification system and Koeppen-Geiger climate classification, are presented in the same table.

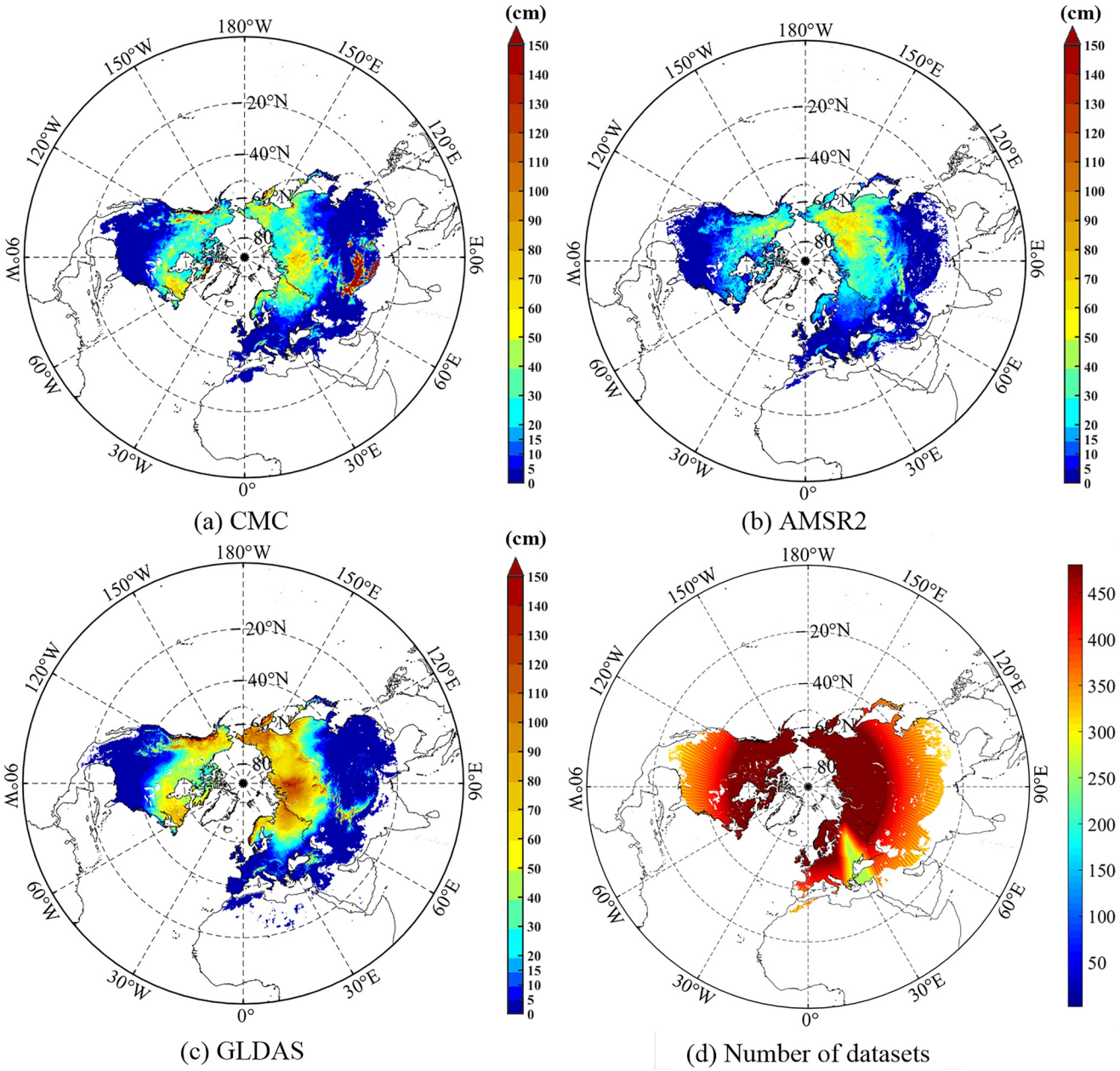

Figure 1. Spatial distributions of averaged SD values during the winters from 2012 to 2017 for (a) CMC, (b) AMSR2, (c) GLDAS, and (d) the number of triplets used for calculation.

4. Results

4.1. Spatial Analysis of SD Products at the Hemisphere Scale

Three daily average SD were calculated by using the data from 2012 to 2017 during the Northern Hemisphere winters (December-February). The spatial distributions of the daily mean winter SD values are illustrated in Figures 1a–1c), and Figure 1d shows the spatial pattern of the number of triplets used for calculation. It should be noted that the green patch in Figure 1d is due to the data missing of AMSR2 SD products. The spatial patterns of the mean SD from the CMC, AMSR2, and GLDAS display visible differences in the Northern Hemisphere. These disparities are mainly caused by different acquisition methods of the three SD products (see Section 2). The GLDAS exhibits much higher values than those of CMC and AMSR2 in high-latitude regions (above 40°N),

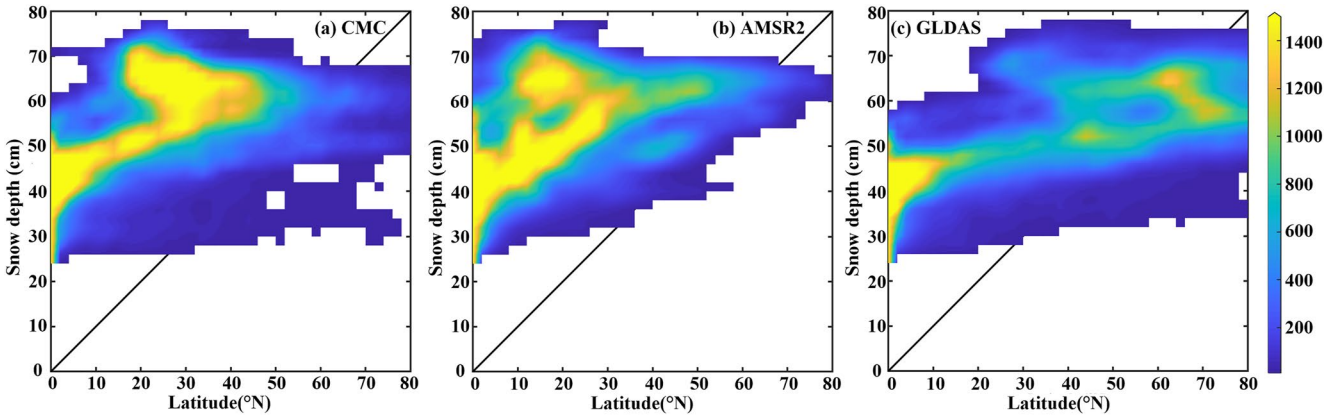

Figure 2. The density scatterplot of the latitude and the SD data sets from (a) CMC, (b) AMSR2, and (c) GLDAS.

while the CMC shows abnormally high values in parts of the Tibetan Plateau, indicating that the use of different approaches to generate the products may lead to inconsistent SD values in the Northern Hemisphere.

As shown in Figure 2, the three SD values show increasing trends as the latitude increases. However, there are still some differences among the three SD data sets at different latitudes. In the latitude of 30°–50°N, the CMC and GLDAS SD is mainly distributed below 10 cm, while some of the AMSR2 SD are distributed in 10–20 cm. In the latitude of 60° –70°N, most of the SD values are distributed in 20–40 cm for CMC, 10–20 cm for AMSR2, and 60–75 cm for GLDAS. Generally, they clearly reveal high SD values at latitudes above 40°N. Higher SD values are also found in the Tibetan Plateau, except for AMSR2. In high-latitude regions, the SD values of the three products exhibit higher SD values over western Siberia and north-eastern Canada. CMC SD shows much higher values over the Tibetan Plateau than do the AMSR2 and GLDAS SD products. AMSR2 exhibits greater SD values over north-eastern Siberia, while the GLDAS product presents significantly higher SD values over Siberia and North America.

4.2. Comparison of SD Products Using ETC Statistics

ETC analysis is employed to characterize uncertainties in the three SD data sets at the hemispheric scale without the true value. The comparison results using ETC are shown in Figure 3.

The CMC performs the best in capturing the evolution of SD (the median $R = 0.77$), higher than GLDAS ($R = 0.69$) and AMSR2 ($R = 0.48$). It is interesting to note that both CMC and GLDAS share a similar distribution of R and perform reasonably well in latitudes above 40°N (e.g., Siberia, Europe, and North America), while the lower R values are distributed in Central Asia. The CMC adopts a great deal of ground observations distributed in Siberia, Europe, and North America (Brasnett, 1999), which results in the good performance of the CMC in these regions. In terms of the GLDAS 2.1, it considers the dynamic SD adjustments based on snow compaction over time, which is helpful to accurately capture the variability in the SD (Koren et al., 1999; Rodell et al., 2004). Compared to both CMC and GLDAS, the AMSR2 presents noticeably lower correlation values in the Northern Hemisphere. Passive microwave observations are easy to get saturated in deep snow conditions (i.e., brightness temperature no longer decreases as SD increases; Takala et al., 2009). In addition, the AMSR2 is not able to characterize the SD variability in complex mountainous regions (e.g., the Tibetan Plateau). The reason for the poor performance of passive SD algorithms is related to many factors (e.g., topography, forest cover, and wind speed; Smith & Bookhagen, 2016).

Figures 3d–3f show the spatial patterns of the STD values obtained from the ETC analysis. In general, the STD map is consistent with the results of the correlation coefficient. The larger STD values of the CMC are mainly found in complex mountainous regions (e.g., Tibetan Plateau and the Rocky Mountains), which may be due to limited ground observations in such regions (Liu et al., 2014). The AMSR2 exhibits larger STD values at latitudes above 40°N, and the GLDAS demonstrates larger STD values in northern Siberia, North America, and Pamir Mountains in central Asia.

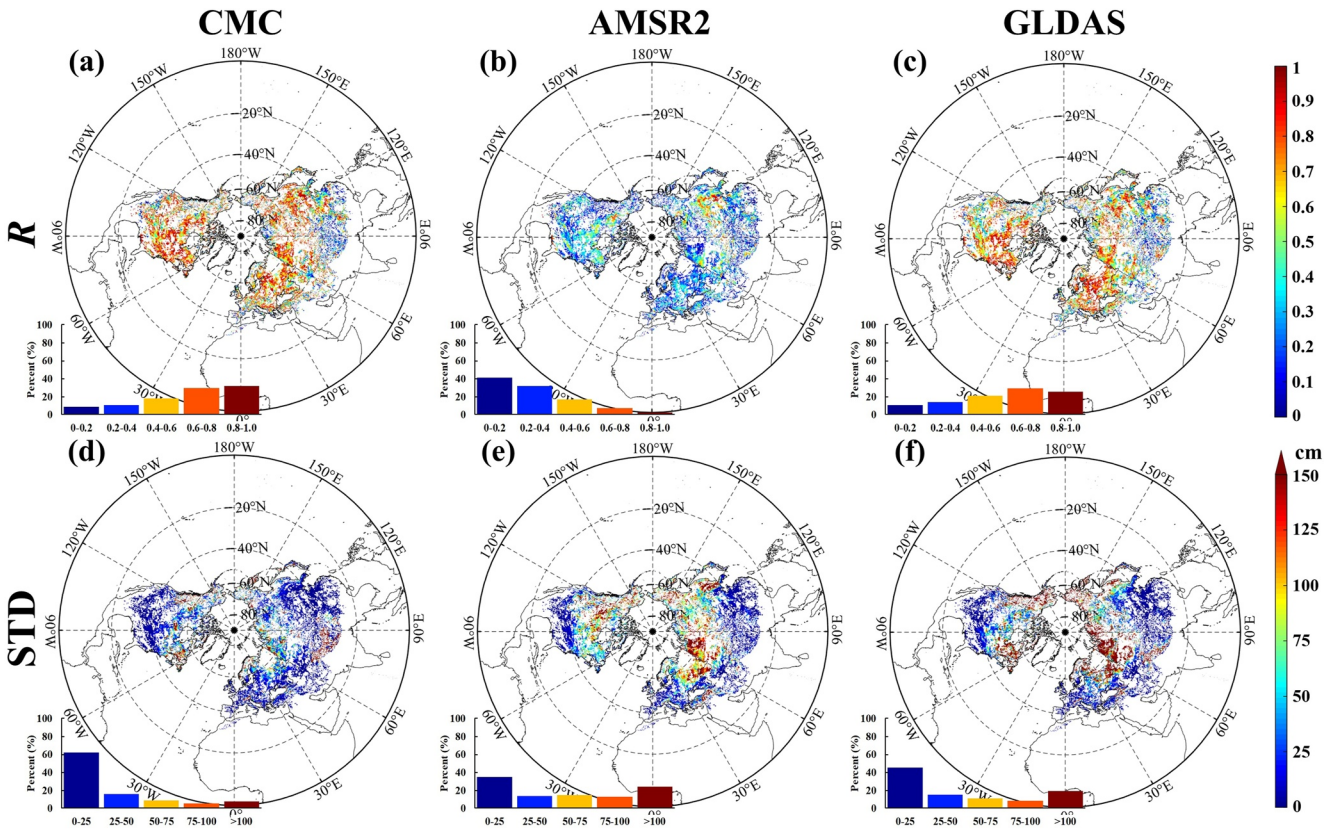

Figure 3. The spatial distribution of *R* and STD for CMC (left column), AMSR2 (middle column), and GLDAS (right column) during the winters from 2012 to 2017 by using the ETC method. The percentage of *R* and STD in different intervals is presented in each subgraph.

4.3. Analysis of Environmental and Perturbing Factors

4.3.1. Impacts of Land Cover

Because different land cover types will affect snow cover redistribution and microwave radiation, it is crucial to investigate the performance of SD products under various land cover types (Tennant et al., 2017). The statistical indicators of the three SD products were calculated and illustrated in Figure 4 and Table 3 for different land cover types.

In terms of capturing the variation of SD, the CMC product outperforms the other two SD products under all studied land cover types, followed by GLDAS and AMSR2, shown in Figure 4a. The CMC exhibits relatively low *R* values in barren regions. One possible reason is that in situ measurements are limited in these regions. The climate in these regions is warmer, which leads to unstable snow cover. For the AMSR2 SD, the lower time-series correlations are mainly found in forests, shrublands, and savannas. The CMC is also superior to the other two products with regard to the STD. All three SD products exhibit the smallest STD values in barren regions. Overall, AMSR2 SD generally performs worse in vegetated regions than in non-vegetated regions (Cho et al., 2020; Lee et al., 2015; Wang et al., 2019). GLDAS has a larger STD in shrublands and savannas regions.

4.3.2. Impacts of Forest Cover Fraction

Aside from assessing the influences of various land cover types on the SD estimates, the response of SD performance to different forest cover fractions was investigated, as shown in Figure 5. Table 4 summarizes the error metrics of the three SD products under different forest cover fractions. In forested areas, the vegetation affects not only the snowfall interception and retention of the surface but also the upwelling microwave radiation from the snowpack (Foster et al., 1991; Lundquist et al., 2013; Tong et al., 2010).

In terms of *R* values, the impacts of forest cover fraction on the three SD products are similar compared with the results of land cover. Generally, CMC and GLDAS have better performance in capturing the SD dynamics

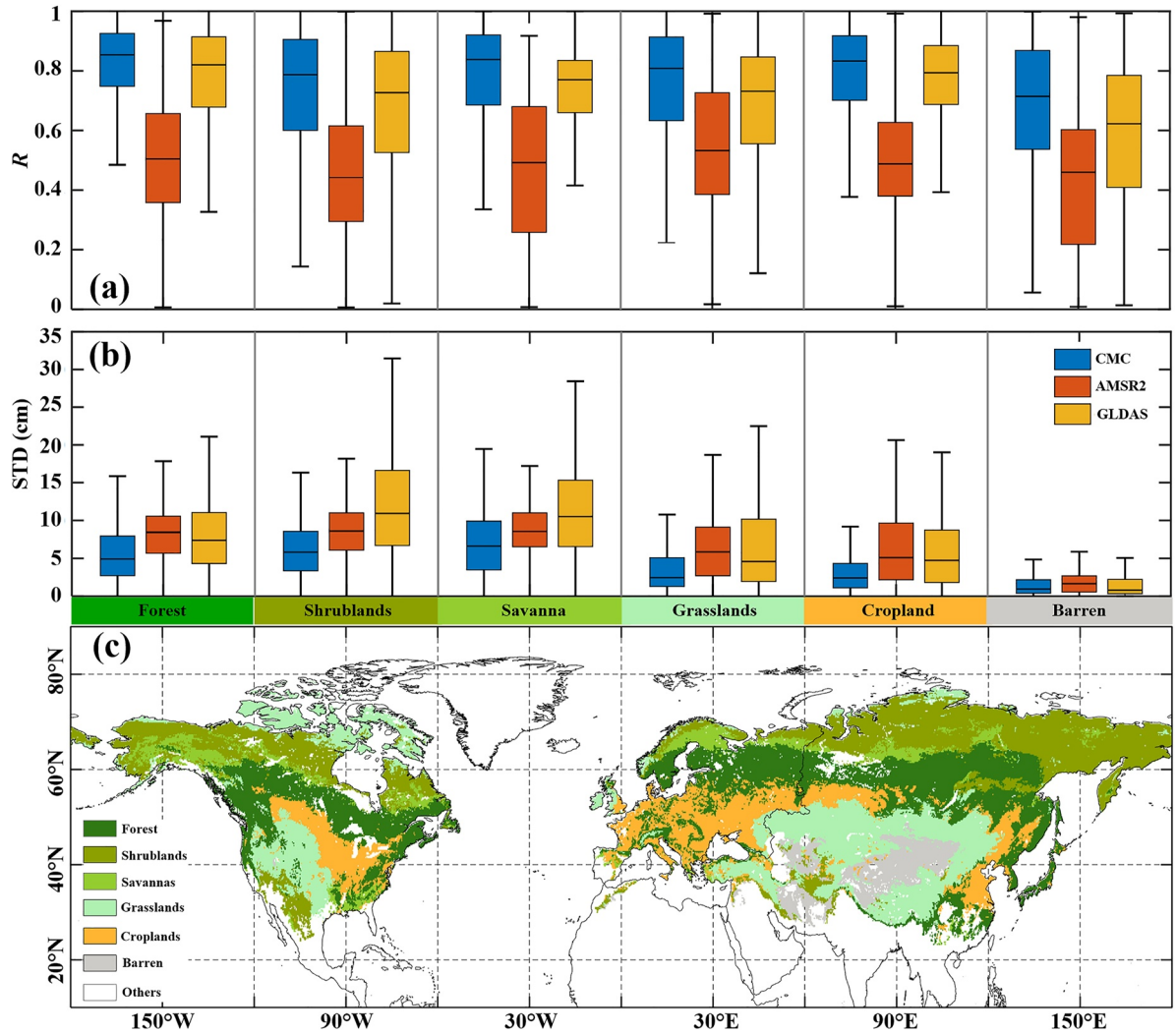

Figure 4. Comparison of (a) *R* and (b) STD for the three SD products (CMC, AMSR2, and GLDAS) considering various land cover types, (c) denotes the MODIS IGBP land cover. In each box, the central mark indicates the median, the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively, and the whiskers extend to the most extreme data points.

compared to AMSR2. The worsened capability of AMSR2 in capturing the SD trends in forest regions is consistent with Lemmetyinen et al. (2009). The large fluctuations of correlation coefficient of AMSR2 SD may be caused by diverse forest types. The forest cover fractions used in this study include five IGBP forest types (the Evergreen Needleleaf Forest, Evergreen Broadleaf Forest, Deciduous Needleleaf Forest, Deciduous Broadleaf Forest, and Mixed Forest), which have different effects on microwave emission, snowfall interception, and retention (De Roo et al., 2007; Pomeroy et al., 1998). To capture the absolute value of the ground SD, the STD values of the three products present continuously increasing trends with increasing forest cover fraction, exhibiting a median STD ranging from 2.3/5.2/4.3 cm for CMC/AMSR2/GLDAS in sparsely forested areas ($0 < ff < 15\%$) to 5.1/8.2/8.3 cm in densely forested areas ($60\% < ff < 100\%$). The results indicate that the forest cover fraction exerts little effect on the SD trend but significantly affects the absolute values of the SD products.

4.3.3. Impacts of Snow Classes

Because the snow properties exert important influences on the SD estimates, the performance of the three SD products with different snow classes was analyzed, and the results are shown in Figure 6 and Table 5. With regards to the skills of the three SD products under the various snow classes, the time series correlations of the three products given in Figure 6a suggest that the three SD products have different capacities to monitor variations of the ground SD. In general, the CMC and GLDAS have similar performance in terms of *R* under various

Table 3
Error Metrics (the Median R and STD) of the Three SD Products Based on the ETC Method Under Different Land Cover Types

Land cover types	Pixels	CMC		AMSR2		GLDAS	
		R	STD (cm)	*R*	STD (cm)	*R*	STD (cm)
Forest	41,893	**0.81**	5.0	0.48	8.0	**0.77**	8.5
Shrublands	33,798	0.73	7.9	0.46	9.7	0.68	13.3
Savannas	15,304	0.77	8.0	0.47	8.7	0.74	10.9
Grasslands	29,784	0.75	4.4	**0.54**	6.5	0.68	6.9
Croplands	28,013	0.79	2.5	0.51	5.5	0.76	5.0
Barren	26,704	0.55	**1.1**	0.43	**2.2**	0.52	**1.2**

Note. The pixels in the table mean the total number of pixels used to determine the ETC-based *R* and STD. The best performance of each error metric is highlighted in bold.

snow classes, and both are superior to the AMSR2 product. Under most snow classes, the CMC outperforms the other two products, with higher *R* and smaller STD values. The three SD products all display the best performance under ephemeral snow with the lowest STD values, which may be due to the thin snow cover with a relatively warm climate. The relative STD can be used as a supplement to the original STD to quantify the performance of different SD products under different ground conditions (with different SD ranges). However, since the true (reference) SD data at the hemispheric scale are not available, this metric cannot be obtained in the study. The three SD products also show low STD values under prairie snow conditions, with an average STD of 4.9/5.6/5.3 cm for CMC/AMSR2/GLDAS, respectively, which may be due to the sparse vegetation cover in such areas (Liu et al., 2014). Discrepancies between the ground SD and estimated values from the CMC, AMSR2, and GLDAS products are large under tundra snow conditions, with larger average STD of 7.2, 8.5, and 13.4 cm, respectively.

4.3.4. Impacts of Surface Roughness

Surface roughness has important effects on physical snow properties and snow distribution (Gharaei-Manesh et al., 2016). Table 6 summarizes the error metrics of the three SD products under different surface roughness. Figure 7 displays the *R* and STD values of the three SD products under different surface roughness. The performance of the three SD products generally decreases with the increase in surface roughness. For CMC, larger *R* values are found over relatively smooth surfaces (0.78 for Rou II and 0.83 for Rou III, and the classification of surface roughness can be seen in Table 2). Similar to CMC, the performance of GLDAS also shows the highest median *R* in low roughness regions with *R* value of 0.78 in Rou I and 0.77 in Rou II. It is hard to satisfactorily capture the evolution of SD under high surface roughness conditions in all SD products. This difficulty may be because all approaches cannot characterize the dramatic spatial changes in complex mountainous regions (Li et al., 2012; Pulvirenti et al., 2011). Compared

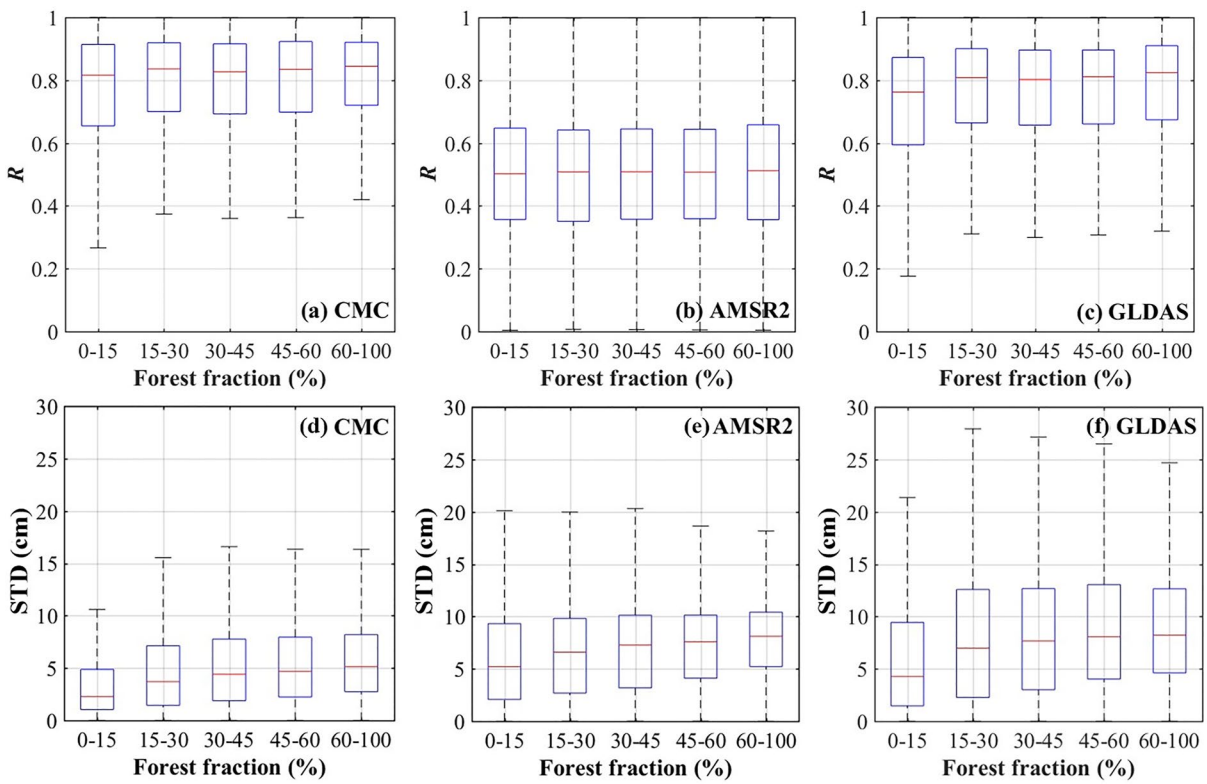

Figure 5. Comparison of *R* and STD for (a, d) CMC, (b, e) AMSR2, and (c, f) GLDAS SD considering various forest cover fractions.

Table 4
Error Metrics (the Median R and STD) of the Three SD Products Based on the ETC Method Under Different Forest Cover Fractions

		CMC		AMSR2		GLDAS	
Forest fraction(%)	Pixels	*R*	STD (cm)	*R*	STD (cm)	*R*	STD (cm)
0–15	88,240	0.82	**2.3**	0.50	**5.2**	0.76	**4.3**
15–30	9,724	0.84	3.7	0.50	6.6	0.81	7.0
30–45	8,313	0.83	4.4	**0.51**	7.3	0.80	7.7
45–60	6,373	0.83	4.6	0.50	7.6	0.81	8.1
60–100	69,633	**0.85**	5.1	**0.51**	8.2	**0.82**	8.3

Note. The pixels in the table mean the total number of pixels used to determine the ETC-based *R* and STD. The best performance of each error metric is highlighted in bold.

with CMC and GLDAS, the AMSR2 SD products show lower time-series correlation values.

As shown in Figure 7 and Table 6, the AMSR2 and GLDAS products exhibit smaller STD values under low surface roughness conditions, with median STD values of 1.3 and 1.4 cm in Rou-I, respectively. Similarly, with increasing roughness, the accuracy of CMC degrades, exhibiting a median STD ranging from 1.0 cm in Rou-I to 3.7 cm in Rou-V.

4.3.5. Impacts of Land Cover Heterogeneity

Various land cover types have different microwave radiation properties. The spatial resolution of current SD data sets is very coarse (e.g., 0.25°), and the landscape in one SD model grid or satellite pixel is hard to be regarded as homogeneous. Land cover heterogeneity may cause uncertainties in SD estimates (Roy et al., 2014). As mentioned in Section 3, the GSI was adopted to characterize the heterogeneous effects of the land surface (Ma et al., 2019). The correlation coefficients and STD of the three SD products under different land cover heterogeneity are shown in Figure 8 and Table 7.

In terms of *R* value under different land cover heterogeneity, the stable performance of the three products is illustrated in Figure 8. AMSR2 exhibits much lower correlation values than both CMC and GLDAS. It is also noted that CMC outperforms the other two products concerning STD, as shown in Figure 8. For the CMC SD product, it is visually seen that the STD values slightly increase with increasing land cover heterogeneity (Figure 8a). Neither AMSR2 nor GLDAS presents noticeable variation trends in terms of their STD values, which are similar to the correlation. Based on these results, the land cover heterogeneity exerts smaller effects on the accuracy of SD products than other factors.

4.3.6. Impacts of Climate Types

To investigate the skill of the three SD products under various climate regions, the *R* and STD values were calculated using the ETC method, as shown in Figure 9 and Table 8. Under most climate conditions, the CMC exhibits the best performance among the three SD products, with the highest *R* and smallest STD. Relatively better performance of the CMC, AMSR2, and GLDAS SD products is achieved under temperate climate conditions, with a median STD of 1.5, 1.9, and 1.7 cm, respectively. The next best performance is found in arid conditions, in which the CMC outperforms the AMSR2 product, with median STD values of 1.2. This result may be due to the shallow snow cover in the above two climate regions. AMSR2 is unable to detect shallow snow (<5 cm) and tends to underestimate in deep snow (>30 cm; Zhang et al., 2016). Nevertheless, the unstable capabilities of AMSR2 and GLDAS in terms of tracking the variations in SD are observed under cold and polar conditions in Figure 9.

Table 5
Error Metrics (the Median R and STD) of the Three SD Products Based on the ETC Method Under Different Snow Classes

		CMC		AMSR2		GLDAS	
Snow classes	Pixels	*R*	STD (cm)	*R*	STD (cm)	*R*	STD (cm)
Tundra	32,398	0.74	7.2	0.49	8.5	0.72	13.4
Taiga	33,384	**0.83**	6.2	**0.55**	9.8	0.79	10.0
Maritime	10,307	0.81	6.5	0.45	5.7	0.80	9.1
Ephemeral	64,578	0.66	**1.9**	0.43	**3.1**	0.63	**2.2**
Prairie	25,880	0.74	4.9	0.50	5.6	0.71	5.3
Alpine	8,689	0.82	5.9	0.46	8.7	**0.82**	9.1

Note. The pixels in the table mean the total number of pixels used to determine the ETC-based *R* and STD. The best performance of each error metric is highlighted in bold.

4.4. ETC Performance Rankings

As discussed in Section 4.3, the three data sets show different performance and error characteristics under different land surfaces and climate zones. Figure 10 shows the best performance of three SD products at the hemispheric scale in terms of *R* and STD obtained from the ETC statistical results. The blue, red, and yellow parts represent the best results concerning *R* (Figure 10a) and STD (Figure 10b) for the CMC, AMSR2, and GLDAS, respectively. The results indicate the great complementary potential to integrate the three different products to form a more reliable hemispheric SD product with increasing *R* and decreasing STD values. As shown in Figure 10a, the CMC product is ranked first in most cases, accounting for 50% of the pixels with viable ETC results. This result implies that CMC has relatively better performance in capturing the trends in most regions of the northeast hemisphere. Compared to CMC, GLDAS performs best in 40.2% of the available pixels, and it has a similar spatial pattern of the highest *R*-value,

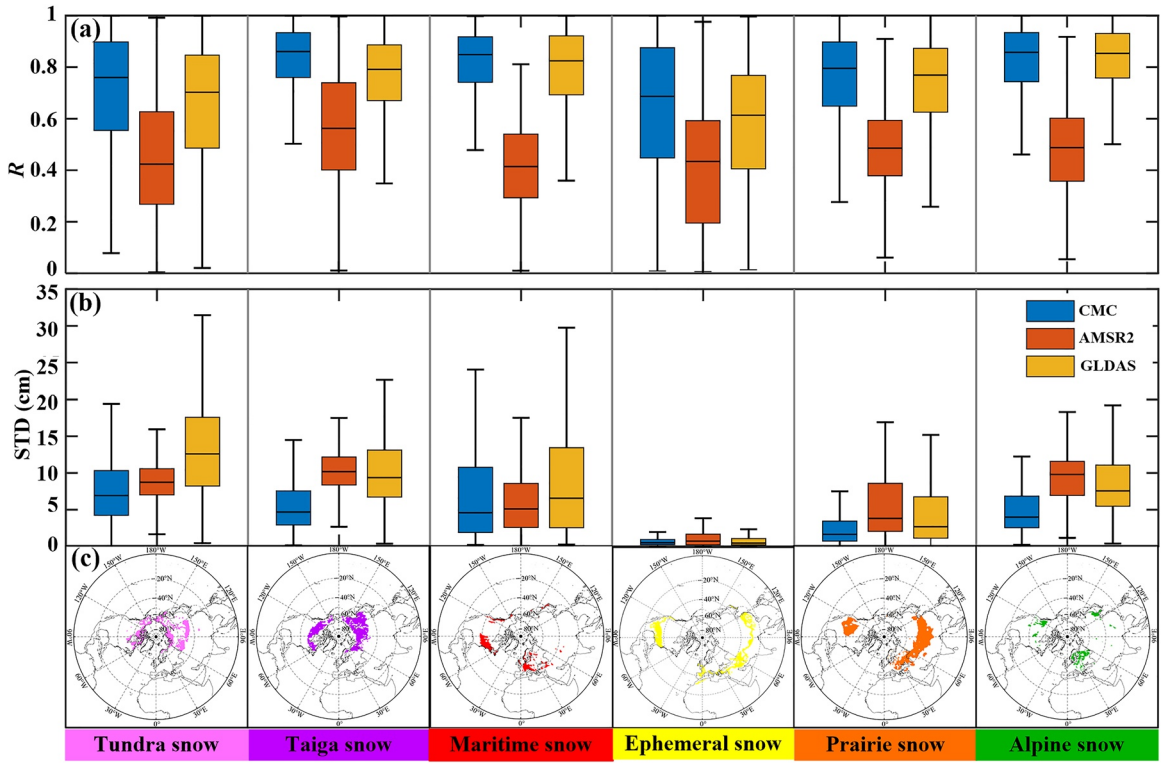

Figure 6. Comparison of (a) *R* and (b) STD for the three SD products (CMC, AMSR2, and GLDAS) considering various snow classes, (c) represents the six snow classes based on the Sturm snow-climate classification system.

except in the low latitudes (<40°N). The AMSR2 product has displayed the fewest pixels (only 9.8%) in terms of the highest ETC-based *R* value. Similarly, in Figure 10b, it is noticed that the CMC with the lowest STD values, accounts for 60% of the pixels with viable ETC results. The percentage of the other two SD products is relatively small (around 20%). This finding from the ETC results reveals the great potential to integrate the three different SD data sets to form a more reliable hemispheric SD product.

5. Discussion

To date, various SD products are available at large scales, including ground-based analysis data sets, passive microwave products, and LSMs data sets. Although they can be applied to various hydrometeorological applications, each product that derives from different methods has its own strengths and weaknesses. The significance of uncertainty analysis of SD products generally includes two aspects: one is that the evaluation results can facilitate the potential users to understand the status of the products and thus better use them for practical applications, and the other is that the results can be served as a feedback to the algorithm/model developers for further algorithm/model refinement.

Table 6
Error Metrics (the Median R and STD) of the Three SD Products Based on the ETC Method Under Different Surface Roughness

		CMC		AMSR2		GLDAS	
SurfaceRoughness	Pixels	*R*	STD (cm)	*R*	STD (cm)	*R*	STD (cm)
Rou-I	3,429	0.77	**1.0**	0.47	**1.3**	**0.78**	**1.4**
Rou-II	42,784	0.78	1.1	**0.53**	2.1	0.77	1.5
Rou-III	66,109	**0.83**	1.2	0.51	2.4	0.76	1.7
Rou-IV	59,695	0.73	1.5	0.48	2.7	0.70	1.9
Rou-V	9,411	0.71	3.7	0.47	3.1	0.63	4.4

Note. The pixels in the table mean the total number of pixels used to determine the ETC-based *R* and STD. The best performance of each error metric is highlighted in bold.

In this study, the ETC-based results show that the CMC ground-based analysis product has better performance than AMSR2 and GLDAS. The main advantage of CMC is that it contains the snow data of the ground observation sites, resulting in its good performance in data-rich regions. Nevertheless, its accuracy is limited by the spatial distribution of the ground sites. The large uncertainties of CMC SD mainly occur in the complex mountainous regions. This is because the meteorological stations are mostly distributed in flat and low-altitude regions. Point-based observations are highly representative only if SD has smooth temporal and spatial variation (Che et al., 2016). Sparse

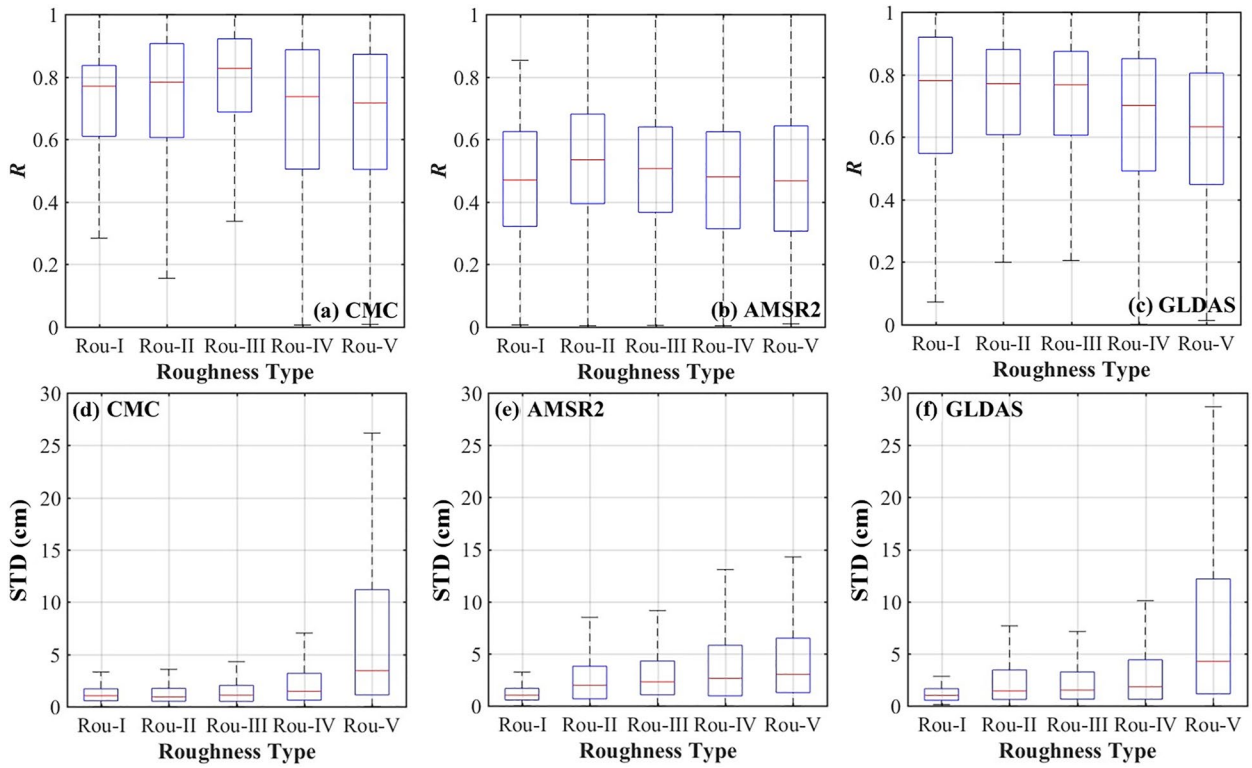

Figure 7. Comparison of *R* and STD for (a, d) CMC, (b, e) AMSR2, and (c, f) GLDAS SD products with regard to various surface roughness conditions.

ground measurements cannot capture the spatial and temporal features of SD under high surface roughness conditions. In addition, the CMC has a larger STD in barren region. This may be because the surface of barren region is relatively flat (i.e., the surface roughness is small), and the landscape heterogeneity is small compared with other land cover types that leads to the smaller STD in this region. Issues with the CMC SD estimates in forest regions may be because most meteorological stations are distributed in open and flat regions. A variety of passive microwave SD products have been released based on different algorithms and sensors. The method of AMSR2 SD product used in this study only considers the relationship between SD and passive microwave brightness temperature. Based on the findings of this study, the accuracy of AMSR2 SD is affected by many factors, which is consistent with previous studies (Cho et al., 2020; Dawson et al., 2018). The AMSR2 SD has larger uncertainty in cold and polar climate conditions. The saturation of passive microwave signals over deep snow is also a limitation of AMSR2 SD product (Zhang et al., 2016). Moreover, different land cover types have different effects on the accuracy of AMSR2 SD. The forest coverage is one of the main factors that influence the performance of AMSR2 SD (Derksen et al., 2005). This is because forest inventory variables (e.g., canopy closure and stem volume) are more likely to attenuate microwave emission from snow surfaces and emit radiation itself, and the snow information contained in the brightness temperature received by the radiometers is reduced accordingly, which are hard to model accurately in an inversion scheme (De Roo et al., 2007; Yang et al., 2015). Although the AMSR2 SD algorithm introduces forest cover fraction, its influences cannot be completely removed. One possible reason is that the statistical relationship between forest cover fraction and SD retrieval underestimation can be ambiguous at the satellite footprint (Derksen et al., 2005). In addition, surface roughness has a strong influence on the accuracy of the AMSR2 SD products. One possible reason is that surface elevation influences the atmospheric contribution to satellite measurements, and surface roughness adds its own contribution to the snow emission, resulting in reduced sensitivity of snow

Table 7
Error Metrics (the Median R and STD) of the Three SD Products Based on the ETC Method Under Different Land Cover Heterogeneity

		CMC		AMSR2		GLDAS	
GSIType	Pixels	*R*	STD (cm)	*R*	STD (cm)	*R*	STD (cm)
GSI-I	69,483	0.81	**3.1**	0.48	**7.6**	0.76	**6.3**
GSI-II	20,379	0.83	4.2	0.49	7.9	0.80	7.3
GSI-III	22,336	0.83	4.3	0.50	**7.6**	0.80	7.2
GSI-IV	33,870	0.82	4.3	0.49	**7.6**	0.81	7.2
GSI-V	35,578	**0.85**	4.9	**0.52**	7.7	**0.83**	7.8

Note. The pixels in the table mean the total number of pixels used to determine the ETC-based *R* and STD. The best performance of each error metric is highlighted in bold.

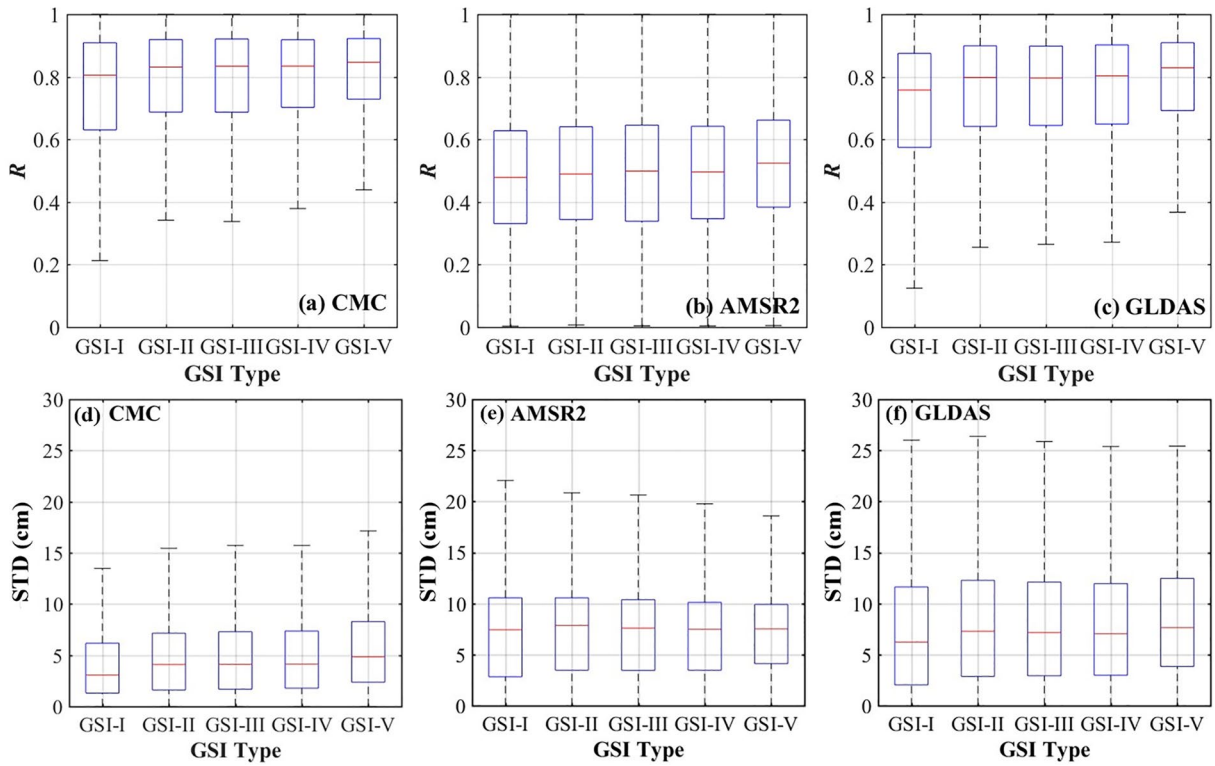

Figure 8. Comparison of *R* and STD for (a, d) CMC (b, e) AMSR2, and (c, f) GLDAS SD products with regard to different GSI groups.

emissivity to SD (Savoie et al., 2009). The LSM takes into account the energy exchange process among snow, soil, and atmosphere, as well as the physical change process inside the snowpack. The accuracy of the SD derived from Noah LSM (GLDAS SD) largely depends on the quality of atmospheric forcing data, model structure, and the selected snow physical process model (e.g., the model structure of snow accumulation and ablation; Ansari et al., 2019; Broxton et al., 2016; Pomeroy et al., 1998). Different land cover would change the cumulative ablation of snow and the energy exchange process between the surface and snow, which in turn has great impact on both accumulation and ablation processes. GLDAS does not take into account the impact of surface complexity on snow evolution and energy exchange processes, resulting in relatively large uncertainties of GLDAS. GLDAS has non-negligible uncertainty in densely forested areas, and this may be because the effect of forest cover is not well considered in GLDAS (Ek et al., 2003). In terms of the effects of snow classes, three SD products have the larger STD in tundra snow cover, and the reason may rise from a variety of reasons. First, the tundra snow cover is characterized by thin, cold, and wind snow, which was usually found above or north of tree line (Sturm et al., 1995). Additionally, the snow cover composition in this region is an ice crystal bottom layer with a composite wind-blown snow shell on it (Sturm et al., 1995). The complicated snow cover form may increase the difficulty of accurate inversion of models and remote sensing as well.

Additionally, the users should pay more attention to interpreting the absolute STD of the SD products in the study. Particularly in some cases, the low STD values of SD products may be due to the shallow snow in certain regions. Therefore, using absolute STD to present the uncertainty of SD products under different conditions (with different SD ranges) may be not intuitive sometimes. The normalized STD (i.e., absolute STD divided by mean SD) seems to be a better statistical metric. However, the calculation of normalized STD often needs a true (or a recognized reliable reference) SD value (e.g., the ground SD measurements). Selecting different mean SD values will lead to different results and implications. Here, we calculated two normalized STD divided by two different mean SD. One is divided by each mean SD estimates of CMC, AMSR2, and GLDAS, respectively, shown in Figure 11, and the other is divided by the mean SD estimates of the three SD products, displayed in Figure 12. It can be seen that there are some notable differences between the two results, especially in the magnitude and the location of the largest STD. Since the hemispheric-scale true value is not available (the ETC does not need true value for calculation), we kept the presentation of absolute STD values of the SD products in this study.

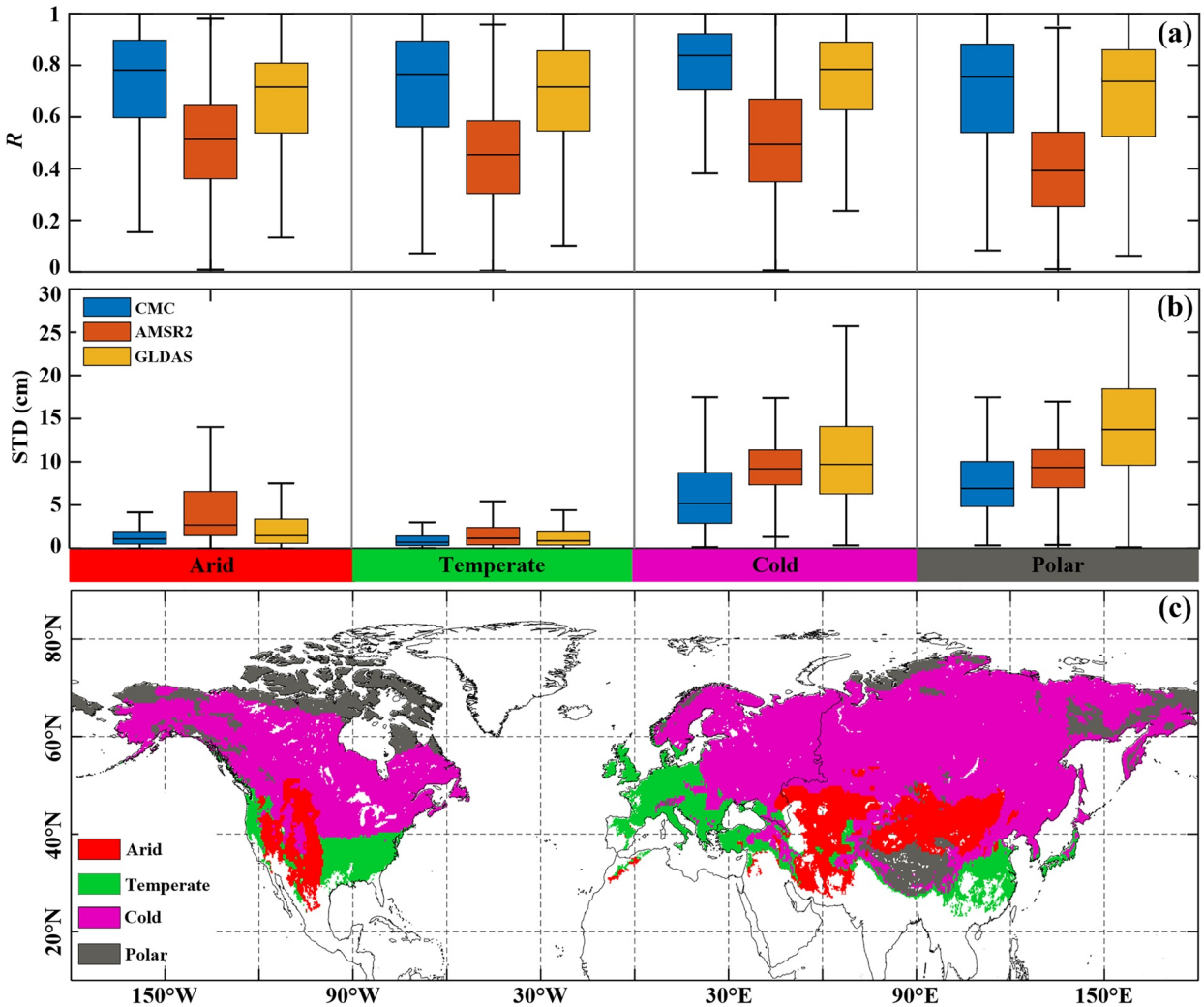

Figure 9. Comparison of (a) *R* and (b) STD for the three SD products (CMC, AMSR2, and GLDAS) considering various climate classes, (c) represents the four climate classes based on the Koeppen-Geiger climate classification.

Table 8
Error Metrics (the Median R and STD) of the Three SD Products Based on the ETC Method Under Different Climate Classes

Climate classes	Pixels	CMC *R*	CMC STD (cm)	AMSR2 *R*	AMSR2 STD (cm)	GLDAS *R*	GLDAS STD (cm)
Arid	44,563	0.76	**1.2**	0.50	2.6	0.69	2.6
Temperate	22,928	0.79	1.5	0.45	**1.9**	0.76	**1.7**
Cold	79,194	**0.85**	5.3	**0.51**	9.0	**0.83**	9.1
Polar	17,802	0.72	6.5	0.44	6.2	0.72	10.5

Note. The pixels in the table mean the total number of pixels used to determine the ETC-based *R* and STD. The best performance of each error metric is highlighted in bold.

In summary, the SD products derived from ground-based analysis, passive microwave, and LSM have their own advantages and limitations. For instance, the CMC has the highest correlations and the lowest STD in most regions of the northeast hemisphere, but it still struggles to estimate SD over complex mountainous regions (e.g., the Tibetan Plateau). AMSR2 shows lower accuracy under vegetation-covered areas in terms of STD, while CMC shows better performance in these regions. Hence, according to the spatial distribution of uncertainty of the three SD products based on the ETC method (e.g., Figure 10), the three SD data sets can be fused to generate a higher quality and more complete SD product by using the merging approaches proposed in previous studies (Gruber et al., 2017; Yilmaz et al., 2012). Meanwhile, the results indicate that the three SD products have different performance under different variables (e.g., land cover, climate, surface roughness, etc.), which can provide a reference for the users to facilitate their applications and improve the SD products. Although this study systematically evaluated the performance of three different types of northern hemisphere SD products, these results focused on the winter period (December, January, and February) due to the difficulty in extracting SD information by microwave radiation

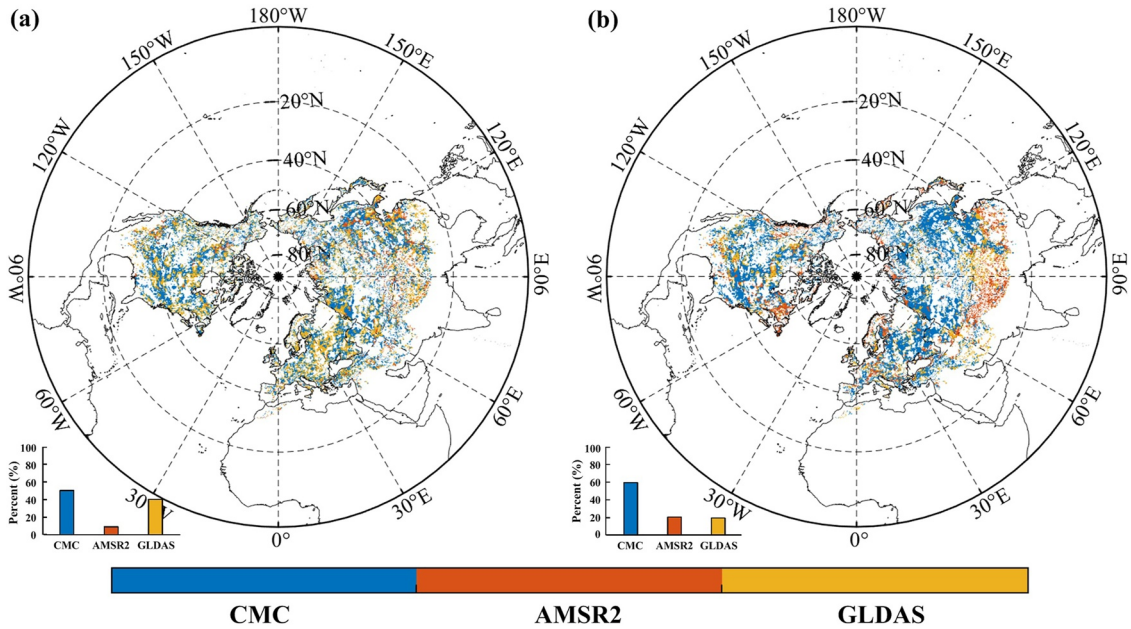

Figure 10. The SD products (CMC, AMSR2, and GLDAS) with the highest ETC-based (a) *R* and (b) STD values. The blue, red, and yellow parts indicate the best performance in terms of each error metric for CMC, AMSR2, and GLDAS, respectively. The white pixels indicate insignificant results (p value≥0.05).

under wet snow conditions, which resulted in a lack of analysis for the melting periods. It also should be noted that though ETC method is an effective technique to calculate the STD and *R* metrics of various geophysical data sets, it cannot obtain the bias component which is also an important metric for SD products. Future efforts will focus on in-depth investigation of the physical reasons for different behaviors of the SD products at diverse ground conditions, to offer guidance for further improvement of the SD products.

6. Conclusions

Knowledge on the uncertainties of SD products is important for their improvement and utilization. This study gives the first attempt to systematically evaluate three different SD products (i.e., CMC, AMSR2, and GLDAS) on a hemispheric scale covering diverse ground conditions by using the ETC technique. The effects of various

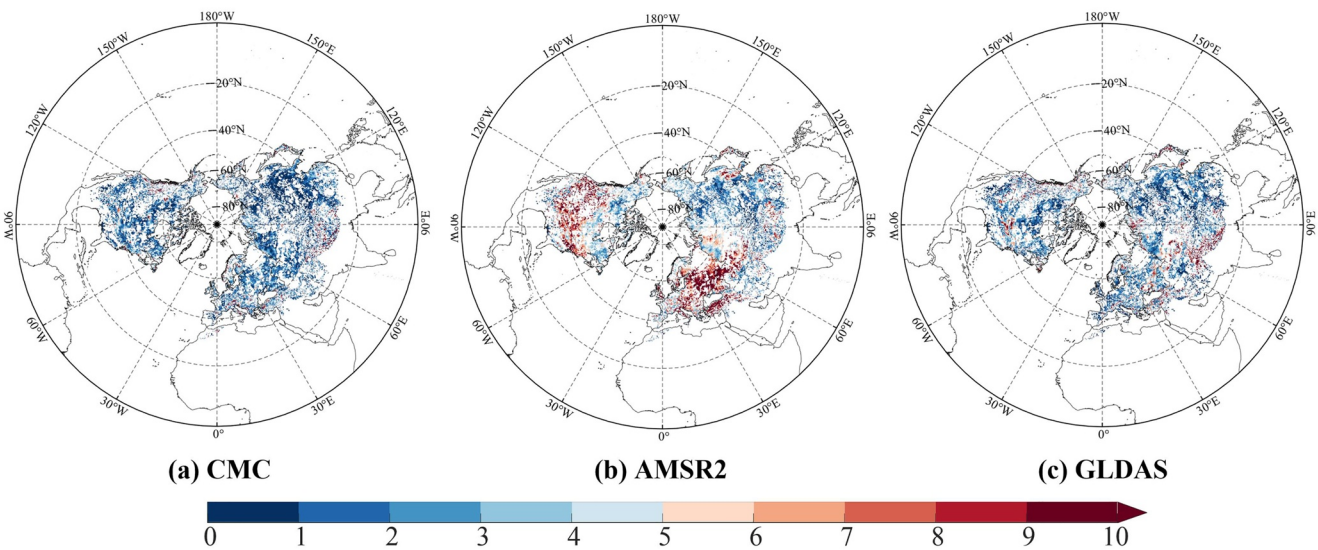

Figure 11. The spatial patterns of the ETC-based STD divided by the mean SD from the corresponding SD data set: (a) CMC, (b) AMSR2, and (c) GLDAS.

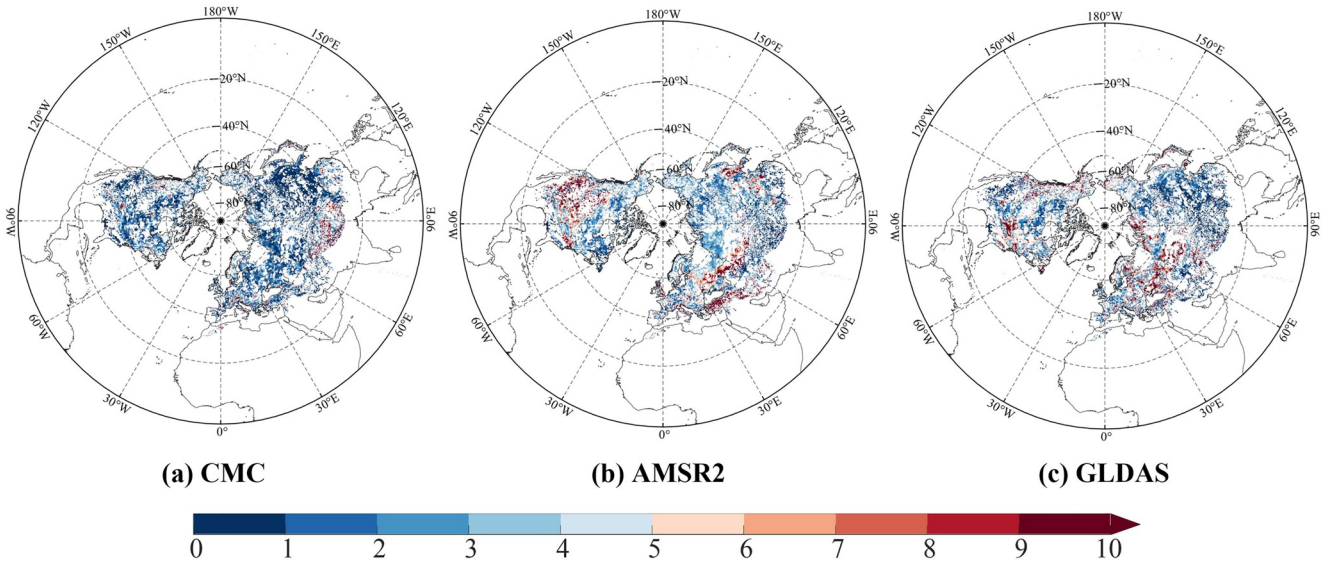

Figure 12. The spatial patterns of the ETC-based STD divided by the mean SD of three SD data sets: (a) CMC, (b) AMSR2, and (c) GLDAS.

environmental and perturbing factors, including land cover, forest cover fraction, snow class, surface roughness, land cover heterogeneity, and climate, on the performance of the three SD products are comprehensively investigated and analyzed.

Generally, the CMC SD product performs the best among the three products, with a hemispheric-scale averaged *R* of 0.73 and STD of 4.8 cm. The accuracy of CMC still needs to be improved in regions with complex terrain (e.g., the Tibetan Plateau), where the ground measurements are very limited. All three products achieve higher absolute accuracy in barren regions, exhibiting averaged STD values of 1.1/2.2/1.2 cm for CMC/AMSR2/GLDAS, respectively. It is also found that forest cover exerts significant impacts on SD products. In particular, an increase in the forest cover fraction causes the performance of CMC, AMSR2, and GLDAS worsen, with median STD ranging from 2.3/5.2/4.3 cm in sparsely forested areas to 5.1/8.2/8.3 cm in densely forested areas, respectively. The complementarity among the three SD products is further demonstrated by considering diverse snow class conditions. The three SD products exhibit the best performance in ephemeral snow regions, while the worst performance is mainly found in tundra, taiga, and maritime snow class regions. The performance of the three SD products is affected by the surface roughness. With increasing surface roughness, the accuracy of the three SD products worsens, exhibiting median STD ranging from 1.0/1.3/1.4 cm in Rou-I to 3.7/3.1/4.4 cm in Rou-V, respectively. This result reveals that the uncertainty of SD products is associated with topographical complexity. In contrast, in terms of land cover heterogeneity, the median STD value of CMC exhibits a slight increasing trend from 3.1 cm (GSI-I) to 4.9 cm (GSI-V), while AMSR2 and GLDAS do not show a significant variation trend in STD values. Considering a variety of climate types, the best performance for all three products is found over temperate climate regions.

Data Availability Statement

The authors also thank the researchers and their teams for providing the data used in this study. The CMC SD data set is available at https://nsidc.org/data/NSIDC-0447/versions/1. The JAXA SD data set can be obtained via ftp.gportal.jaxa.jp. The GLDAS-2.1 SD data set can be downloaded at https://disc.gsfc.nasa.gov/datasets. The MODIS IGBP land cover classification data sets are available at https://lpdaac.usgs.gov/products/mcd12c1v006/. The GLOBE DEM data are obtained at https://www.ngdc.noaa.gov/mgg/topo/globe.html. The other processed auxiliary data have been submitted to Dryad repository, but the data is currently in process (https://datadryad.org/stash/share/nWhnV7k_EMHLqCz_EQXokkYMcYj_8vXw6TbfHJRZu2w), and the data are also available at the current link (https://drive.google.com/drive/folders/1ZXfD3Jy7XTfG7I43GNbhiUd2-yq38LIo?usp=sharing) for the purposes of peer review only.

Acknowledgments
This study was supported by the National Key Research and Development Program of China under Grant 2018YFA0605403 and the Second Tibetan Plateau Scientific Expedition and Research Program (STEP) under Grants 2019QZKK0206 and 2019QZKK0201, and the Natural Science Foundation of China (Grant nos. 41971317 and 42071084). Jiangyuan Zeng was supported by the Youth Innovation Promotion Association CAS (Grant no. 2018082). The authors thank all the reviewers for their helpful reviews and suggestions that have greatly improved the study.

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P. P., Janowiak, J., et al. (2003). The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). *Journal of Hydrometeorology*, *4*, 1147–1167. https://doi.org/10.1175/1525-7541(2003)004<1147:tvgpcp>2.0.co;2

Ansari, H., Marofi, S., & Mohamadi, M. (2019). Topography and land cover effects on snow water equivalent estimation using AMSR-E and GLDAS Data. *Water Resources Management*, *33*, 1699–1715. https://doi.org/10.1007/s11269-019-2200-0

Armstrong, R. L., & Brodzik, M. J. (2002). Hemispheric-scale comparison and evaluation of passive-microwave snow algorithms. *Annals of Glaciology*, *34*, 38–44. https://doi.org/10.3189/172756402781817428

Beck, H. E., Pan, M., Miralles, D. G., Reichle, R. H., Dorigo, W. A., Hahn, S., et al. (2021). Evaluation of 18 satellite-and model-based soil moisture products using in situ measurements from 826 sensors. *Hydrology and Earth System Sciences*, *25*(1), 17–40. https://doi.org/10.5194/hess-25-17-2021

Bormann, K. J., Ross, D. B., Derksen, C., & Painter, T. H. (2018). Estimating snow-cover trends from space. *Nature Climate Change*, *8*, 924–928. https://doi.org/10.1038/s41558-018-0318-3

Brasnett, B. (1999). A global analysis of snow depth for numerical weather prediction. *Journal of Applied Meteorology*, *38*, 726–740. https://doi.org/10.1175/1520-0450(1999)038<0726:agaosd>2.0.co;2

Brown, R. D., Brasnett, B., & David, R. (2003). Gridded North American monthly snow depth and snow water equivalent for GCM evaluation. *Atmosphere-Ocean*, *41*(1), 1–14. https://doi.org/10.3137/ao.410101

Broxton, P. D., Zeng, X., & Dawson, N. (2016). Why do global reanalyses and land data assimilation products underestimate snow water equivalent? *Journal of Hydrometeorology*, *17*(11), 2743–2761. https://doi.org/10.1175/jhm-d-16-0056.1

Chang, A. T. C., Foster, J. L., & Hall, D. K. (1987). Nimbus-7 SMMR derived global snow cover parameters. *Annals of Glaciology*, *9*, 39–44. https://doi.org/10.1017/s0260305500000355

Chang, A. T. C., Foster, J. L., & Hall, D. K. (1996). Effects of forest on the snow parameters derived from microwave measurements during the BOREAS Winter Field Campaign. *Hydrological Processes*, *10*(12), 1565–1574. https://doi.org/10.1002/(sici)1099-1085(199612)10:12<1565::aid-hyp501>3.0.co;2-5

Che, T., Dai, L., Zheng, X., Li, X., & Zhao, K. (2016). Estimation of snow depth from passive microwave brightness temperature data in forest regions of northeast China. *Remote Sensing of Environment*, *183*(15), 334–349. https://doi.org/10.1016/j.rse.2016.06.005

Chen, F., Crow, W. T., Bindlish, R., Colliander, A., Burgin, M. S., Asanuma, J., & Aida, K. (2018). Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation. *Remote Sensing of Environment*, *214*(1), 1–13. https://doi.org/10.1016/j.rse.2018.05.008

Cho, E., Jacobs, J. M., & Vuyovich, C. M. (2020). The value of long-term (40 years) airborne gamma radiation SWE record for evaluating three observation-based gridded SWE data sets by seasonal snow and land cover classifications. *Water Resources Research*, *56*(1), e2019WR025813. https://doi.org/10.1029/2019wr025813

Cho, E., Tuttle, S. E., & Jacobs, J. M. (2017). Evaluating consistency of snow water equivalent retrievals from passive microwave sensors over the North Central U. S. SSM/I vs. SSMIS and AMSR-E vs. AMSR2. *Remote Sensing*, *9*(5), 465. https://doi.org/10.3390/rs9050465

Clark, M. P., Hendrikx, J., Slater, A. G., Kavetski, D., Anderson, B., & Cullen, N. J. (2011). Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review. *Water Resources Research*, *47*(7), W07539. https://doi.org/10.1029/2011wr010745

Clifford, D. (2010). Global estimates of snow water equivalent from passive microwave instruments: History, challenges and future developments. *International Journal of Remote Sensing*, *31*(14), 3707–3726. https://doi.org/10.1080/01431161.2010.483482

Crow, W. T., Koster, R. D., Reichle, R. H., & Sharif, H. O. (2005). Relevance of time-varying and time-invariant retrieval error sources on the utility of spaceborne soil moisture products. *Geophysical Research Letters*, *32*, L24405. https://doi.org/10.1029/2005gl024889

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

Dai, L. Y., Che, T., Wang, J., & Zhang, P. (2012). Snow depth and snow water equivalent estimation from AMSR-E data based on a priori snow characteristics in Xinjiang, China. *Remote Sensing of Environment*, *127*, 14–29. https://doi.org/10.1016/j.rse.2011.08.029

Davis, R., McKenzie, J. C., & Jordan, R. (1995). Distributed snow process modelling: An image processing approach. *Hydrological Processes*, *9*(8), 865–875. https://doi.org/10.1002/hyp.3360090804

Dawson, N., Broxton, P., & Zeng, X. (2018). Evaluation of remotely sensed snow water equivalent and snow cover extent over the contiguous United States. *Journal of Hydrometeorology*, *19*(11), 1777–1791. https://doi.org/10.1175/jhm-d-18-0007.1

De Roo, R., Chang, A., & England, A. (2007). Radiobrightness at 6.7-, 19-, and 37-GHz downwelling from mature evergreen trees observed during the Cold Land Processes Experiment in Colorado. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(10), 3224–3229. https://doi.org/10.1109/tgrs.2007.900688

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Derber, J. C., Parrish, D. F., & Lord, S. J. (1991). The new global operational analysis system at the National Meteorological Center. *Weather and Forecasting*, *6*, 538–547. https://doi.org/10.1175/1520-0434(1991)006<0538:tngoas>2.0.co;2

Derksen, C. (2008). The contribution of AMSR-E 18.7 and 10.7 GHz measurements to improved boreal forest snow water equivalent retrievals. *Remote Sensing of Environment*, *112*(5), 2701–2710. https://doi.org/10.1016/j.rse.2008.01.001

Derksen, C., Walker, A. E., & Goodison, B. (2005). Evaluation of passive microwave snow water equivalent retrievals across the boreal forest/tundra transition of Western Canada. *Remote Sensing of Environment*, *96*(3–4), 315–327. https://doi.org/10.1016/j.rse.2005.02.014

Dorigo, W. A., Scipal, K., Parinussa, R. M., Liu, Y. Y., Wagner, W., De Jeu, R. A., & Naeimi, V. (2010). Error characterisation of global active and passive microwave soil moisture datasets. *Hydrology and Earth System Sciences*, *14*(12), 2605–2616. https://doi.org/10.5194/hess-14-2605-2010

Draper, C., Reichle, R., Jeu, R. D., Naeimi, V., Parinussa, R., & Wagner, W. (2013). Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. *Remote Sensing of Environment*, *137*, 288–298. https://doi.org/10.1016/j.rse.2013.06.013

Drusch, M., Vasiljevic, D., & Viterbo, P. (2004). ECMWF's global snow analysis: Assessment and revision based on satellite observations. *Journal of Applied Meteorology*, *43*(9), 1282–1294. https://doi.org/10.1175/1520-0450(2004)043<1282:egsaaa>2.0.co;2

Dyer, J. L., & Mote, T. L. (2006). Spatial variability and trends in observed snow depth over North America. *Geophysical Research Letters*, *33*. https://doi.org/10.1029/2006gl027258

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., et al. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research: Atmospheres*, *108*(D22), 8851. https://doi.org/10.1029/2002jd003296

Erxleben, J., Elder, K., & Davis, R. (2002). Comparison of spatial interpolation methods for estimating snow distribution in the Colorado Rocky Mountains. *Hydrological Processes*, *16*, 3627–3649. https://doi.org/10.1002/hyp.1239

Flanner, M. G., Shell, K. M., Barlage, M., Perovich, D. K., & Tschudi, M. A. (2011). Radiative forcing and albedo feedback from the northern hemisphere cryosphere between 1979 and 2008. *Nature Geoscience*, *4*, 151–155. https://doi.org/10.1038/ngeo1062

Foster, J. L., Chang, A., Hall, D., & Rango, A. (1991). Derivation of snow water equivalent in boreal forests using microwave radiometry. *Arctic*, *44*, 147–152. https://doi.org/10.14430/arctic1581

Foster, J. L., Sun, C., Walker, J. P., Kelly, R., Chang, A., Dong, J., & Powell, H. (2005). Quantifying the uncertainty in passive microwave snow water equivalent observations. *Remote Sensing of Environment*, *94*(2), 187–203. https://doi.org/10.1016/j.rse.2004.09.012

Friedl, M. A., Mclver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., et al. (2002). Global land cover mapping from MODIS: Algorithms and early results. *Remote Sensing of Environment*, *83*(1–2), 287–302. https://doi.org/10.1016/s0034-4257(02)00078-0

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The Modern-Era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, *30*, 5419–5454. https://doi.org/10.1175/jcli-d-16-0758.1

Gharaei-Manesh, S., Fathzadeh, A., & Taghizadeh-Mehrjardi, R. (2016). Comparison of artificial neural network and decision tree models in estimating spatial distribution of snow depth in a semi-arid region of Iran. *Cold Regions Science and Technology*, *122*, 26–35. https://doi.org/10.1016/j.coldregions.2015.11.004

Gruber, A., Dorigo, W. A., Crow, W., & Wagner, W. (2017). Triple collocation-based merging of satellite soil moisture retrievals. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(12), 6780–6792. https://doi.org/10.1109/tgrs.2017.2734070

Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., & Wagner, W. (2016). Recent advances in (soil moisture) triple collocation analysis. *International Journal of Applied Earth Observation and Geoinformation*, *45*, 200–211. https://doi.org/10.1016/j.jag.2015.09.002

Han, P., Long, D., Han, Z., Du, M., Dai, L., & Hao, X. (2019). Improved understanding of snowmelt runoff from the headwaters of China's Yangtze river using remotely sensed snow products and hydrological modeling. *Remote Sensing of Environment*, *224*, 44–59. https://doi.org/10.1016/j.rse.2019.01.041

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., et al. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850–853. https://doi.org/10.1126/science.1244693

Hastings, D. A., & Dunbar, P. (1998). Development & assessment of the global land one-km base elevation digital elevation model (GLOBE). *Group*, *4*(6), 218–221

Hudson, G., & Wackernagel, H. (2010). Mapping temperature using kriging with external drift: Theory and an example from scotland. *International Journal of Climatology*, *14*, 77–91

Imaoka, K., Maeda, T., Kachi, M., Kasahara, M., Ito, N., & Nakagawa, K. (2012). Status of AMSR2 instrument on GCOM-W1. *Proceedings of SPIE*, 852815. https://doi.org/10.1117/12.977774

Kelly, R. (2009). The AMSR-E snow depth algorithm: Description and initial results. *Journal of the Remote Sensing Society of Japan*, *29*(1), 307–317

Khan, M. S., Liaqat, U. W., Baik, J., & Choi, M. (2018). Stand-alone uncertainty characterization of GLEAM, GLDAS and MOD16 evapotranspiration products using an extended triple collocation approach. *Agricultural and Forest Meteorology*, *252*, 256–268. https://doi.org/10.1016/j.agrformet.2018.01.022

Kim, R. S., Kumar, S., Vuyovich, C., Houser, P., Lundquist, J., Mudryk, L., et al. (2021). Snow Ensemble Uncertainty Project (SEUP): Quantification of snow water equivalent uncertainty across North America via ensemble land surface modeling. *The Cryosphere*, *15*(2), 771–791. https://doi.org/10.5194/tc-15-771-2021

Kim, S., Liu, Y. Y., Johnson, F. M., Parinussa, R. M., & Sharma, A. (2015). A global comparison of alternate AMSR2 soil moisture products: Why do they differ? *Remote Sensing of Environment*, *161*, 43–62. https://doi.org/10.1016/j.rse.2015.02.002

Kitaev, L., Kislov, A., Krenke, A., Razuvaev, V., Martuganov, R., & Konstantinov, I. (2002). The snow cover characteristics of northern Eurasia and their relationship to climatic parameters. *Boreal Environment Research*, *7*, 437–445.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, *93*, 5–48. https://doi.org/10.2151/jmsj.2015-001

Koren, V., Schaake, J., Mitchell, K., Duan, Q. Y., Chen, F., & Baker, J. M. (1999). A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. *Journal of Geophysical Research: Atmospheres*, *104*(D16), 19569–19585. https://doi.org/10.1029/1999jd900232

Larue, F., Royer, A., Seve, D. D., Langlois, A., Roy, A., & Brucker, L. (2017). Validation of GlobSnow-2 snow water equivalent over Eastern Canada. *Remote Sensing of Environment*, *194*, 264–277. https://doi.org/10.1016/j.rse.2017.03.027

Lee, Y. K., Kongoli, C., & Key, J. (2015). An in-depth evaluation of heritage algorithms for snow cover and snow depth using AMSR-E and AMSR2 measurements. *Journal of Atmospheric and Oceanic Technology*, *32*(12), 2319–2336. https://doi.org/10.1175/jtech-d-15-0100.1

Lemmetyinen, J., Derksen, C., Pulliainen, J., Strapp, W., Toose, P., Walker, A., et al. (2009). A comparison of airborne microwave brightness temperatures and snowpack properties across the boreal forests of Finland and Western Canada. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(3), 965–978. https://doi.org/10.1109/tgrs.2008.2006358

Li, T., Chen, Y. Z., Han, L. J., Cheng, L. H., Lv, Y. H., Fu, B. J., et al. (2021). Shortened duration and reduced area of frozen soil in the Northern Hemisphere. *The Innovation*, *2*(3), 100146. https://doi.org/10.1016/j.xinn.2021.100146

Li, D., Durand, M., & Margulis, S. A. (2012). Potential for hydrologic characterization of deep mountain snowpack via passive microwave remote sensing in the Kern River basin, Sierra Nevada, USA. *Remote Sensing of Environment*, *125*, 34–48. https://doi.org/10.1016/j.rse.2012.06.027

Li, C., Tang, G., & Hong, Y. (2018). Cross-evaluation of ground-based, multi-satellite and reanalysis precipitation products: Applicability of the Triple Collocation method across Mainland China. *Journal of Hydrology*, *562*, 71–83. https://doi.org/10.1016/j.jhydrol.2018.04.039

Liu, Q., Babanin, A. V., Guan, C., Zieger, S., Sun, J., & Jia, Y. (2016). Calibration and validation of HY-2 altimeter wave height. *Journal of Atmospheric and Oceanic Technology*, *33*, 919–936. https://doi.org/10.1175/jtech-d-15-0219.1

Liu, J., Li, Z., Huang, L., & Tian, B. S. (2014). Hemispheric-scale comparison of monthly passive microwave snow water equivalent products. *Journal of Applied Remote Sensing*, *8*, 084688. https://doi.org/10.1117/1.jrs.8.084688

Liu, Y., Peters-Lidard, C. D., Kumar, S., Foster, J. L., Shaw, M., Tian, Y., & Fall, G. M. (2013). Assimilating satellite-based snow depth and snow cover products for improving snow predictions in Alaska. *Advances in Water Resources*, *54*, 208–227. https://doi.org/10.1016/j.advwatres.2013.02.005

Lu, H., Wei, W. S., Liu, M. Z., Han, X., Li, M., & Hong, W. (2016). Variations in seasonal snow surface energy exchange during a snowmelt period: An example from the Tianshan Mountains, China. *Meteorological Applications*, *23*, 14–25. https://doi.org/10.1002/met.1511

Lundquist, J. D., Dickerson-Lange, S. E., Lutz, J. A., & Cristea, N. C. (2013). Lower forest density enhances snow retention in regions with warmer winters: A global framework developed from plot-scale observations and modeling. *Water Resources Research*, *49*(10), 6356–6370. https://doi.org/10.1002/wrcr.20504

Lyu, H., McColl, K. A., Li, X., Derksen, C., Berg, A., Black, T. A., et al. (2018). Validation of the SMAP freeze/thaw product using categorical triple collocation. *Remote Sensing of Environment*, *205*, 329–337. https://doi.org/10.1016/j.rse.2017.12.007

Ma, H. L., Zeng, J. Y., Chen, N. C., Zhang, X., Cosh, M., & Wang, W. (2019). Satellite surface soil moisture from SMAP, SMOS, AMSR2 and ESA CCI: A comprehensive assessment using global ground-based observations. *Remote Sensing of Environment*, *231*, 111215. https://doi.org/10.1016/j.rse.2019.111215

McColl, K. A., Roy, A., Derksen, C., Konings, A. G., Alemohammed, S. H., & Entekhabi, D. (2016). Triple collocation for binary and categorical variables: Application to validating landscape freeze/thaw retrievals. *Remote Sensing of Environment*, *176*, 31–42. https://doi.org/10.1016/j.rse.2016.01.010

McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., & Stoffelen, A. (2014). Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. *Geophysical Research Letters*, *41*, 6229–6236. https://doi.org/10.1002/2014gl061322

Miralles, D. G., De Jeu, R. A., Gash, J. H., Holmes, T. R., & Dolman, A. J. (2011). Magnitude and variability of land evaporation and its components at the global scale. *Hydrology and Earth System Sciences*, *15*, 967–981. https://doi.org/10.5194/hess-15-967-2011

Mortimer, C., Mudryk, L., Derksen, C., Luojus, K., Brown, R., Kelly, R., & Tedesco, M. (2020). Evaluation of long-term northern hemisphere snow water equivalent products. *The Cryosphere*, *14*(5), 1579–1594. https://doi.org/10.5194/tc-14-1579-2020

Mudryk, L. R., Derksen, C., Kushner, P. J., & Brown, R. (2015). Characterization of northern hemisphere snow water equivalent datasets, 1981–2010. *Journal of Climate*, *28*, 8037–8051. https://doi.org/10.1175/jcli-d-15-0229.1

O'Carroll, A. G., Eyre, J. R., & Saunders, R. W. (2008). Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. *Journal of Atmospheric and Oceanic Technology*, *25*, 1197–1207.

Okuyama, A., & Imaoka, K. (2015). Intercalibration of advanced microwave scanning radiometer-2 (AMSR2) brightness temperature. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(8), 4568–4577. https://doi.org/10.1109/tgrs.2015.2402204

Pomeroy, J. W., Parviainen, J., Hedstrom, N., & Gray, D. M. (1998). Coupled modelling of forest snow interception and sublimation. *Hydrological Processes*, *12*, 2317–2337. https://doi.org/10.1002/(sici)1099-1085(199812)12:15<2317::aid-hyp799>3.0.co;2-x

Pulvirenti, L., Pierdicca, N., & Marzano, F. S. (2011). Prediction of the error induced by topography in satellite microwave radiometric observations. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(9), 3180–3188. https://doi.org/10.1109/tgrs.2010.2096514

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C., et al. (2004). The global land data assimilation system. *Bulletin of the American Meteorological Society*, *85*, 381–394. https://doi.org/10.1175/bams-85-3-381

Roy, A., Royer, A., & Hall, R. J. (2014). Relationship between forest microwave transmissivity and structural parameters for the Canadian boreal forest. *IEEE Geoscience and Remote Sensing Letters*, *11*(10), 1802–1806. https://doi.org/10.1109/lgrs.2014.2309941

Rubel, F., & Kottek, M. (2010). Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorologische Zeitschrift*, *19*, 135–141. https://doi.org/10.1127/0941-2948/2010/0430

Rutter, N., Cline, D., & Li, L. (2008). Evaluation of the NOHRSC snow model (NSM) in a one-dimensional mode. *Journal of Hydrometeorology*, *9*, 695–711. https://doi.org/10.1175/2008jhm861.1

Savoie, M. H., Armstrong, R. L., Brodzik, M. J., & Wang, J. R. (2009). Atmospheric corrections for improved satellite passive microwave snow cover retrievals over the Tibet Plateau. *Remote Sensing of Environment*, *113*(12), 2661–2669. https://doi.org/10.1016/j.rse.2009.08.006

Scipal, K., Holmes, T., deJeu, R., Naeimi, V., & Wagner, W. (2008). A possible solution for the problem of estimating the error structure of global soil moisture data sets. *Geophysical Research Letters*, *35*(24), L24403. https://doi.org/10.1029/2008gl035599

Simpson, E. H. (1949). The measurement of diversity. *Nature*, *163*, 688. https://doi.org/10.1038/163688a0

Smith, T., & Bookhagen, B. (2016). Assessing uncertainty and sensor biases in passive microwave data across High Mountain Asia. *Remote Sensing of Environment*, *181*, 174–185. https://doi.org/10.1016/j.rse.2016.03.037

Stiles, W. H., & Ulaby, F. T. (1980). The active and passive microwave response to snow parameters: 1. Wetness. *Journal of Geophysical Research: Oceans*, *85*(C2), 1037–1044. https://doi.org/10.1029/jc085ic02p01037

Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *Journal of Geophysical Research*, *103*, 7755–7766. https://doi.org/10.1029/97jc03180

Sturm, M., Holmgren, J., & Liston, G. E. (1995). A seasonal snow cover classification system for local to global applications. *Journal of Climate*, *8*, 1261–1283. https://doi.org/10.1175/1520-0442(1995)008<1261:asccs>2.0.co;2

Sturm, M., Taras, B., Liston, G. E., Derksen, C., Jonas, T., & Lea, J. (2010). Estimating snow water equivalent using snow depth data and climate classes. *Journal of Hydrometeorology*, *11*, 1380–1394. https://doi.org/10.1175/2010jhm1202.1

Takala, M., Pulliainen, J., Metsamaki, S. J., & Koskinen, J. T. (2009). Detection of snowmelt using spaceborne microwave radiometer data in Eurasia from 1979 to 2007. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(9), 2996–3007. https://doi.org/10.1109/tgrs.2009.2018442

Tedesco, M., & Narvekar, P. S. (2010). Assessment of the NASA AMSRE SWE product. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *3*(1), 141–159. https://doi.org/10.1109/jstars.2010.2040462

Tennant, C. J., Harpold, A. A., Lohse, K. A., Godsey, S. E., Crosby, B. T., Larsen, L. G., et al. (2017). Regional sensitivities of seasonal snowpack to elevation, aspect, and vegetation cover in Western North America. *Water Resources Research*, *53*(8), 6908–6926. https://doi.org/10.1002/2016wr019374

Tong, J., Déry, S. J., Jackson, P. L., & Derksen, C. (2010). Testing snow water equivalent retrieval algorithms for passive microwave remote sensing in an alpine watershed of Western Canada. *Canadian Journal of Remote Sensing*, *36*(1), S74–S86. https://doi.org/10.5589/m10-009

Walker, A. E., & Goodison, B. E. (1993). Discrimination of a wet snow cover using passive microwave satellite data. *Annals of Glaciology*, *17*, 307–311. https://doi.org/10.1017/s026030550001301x

Wang, Y., Huang, X., Wang, J., Zhou, M., & Liang, T. (2019). AMSR2 snow depth downscaling algorithm based on a multifactor approach over the Tibetan Plateau, China. *Remote Sensing of Environment*, *231*, 111268. https://doi.org/10.1016/j.rse.2019.111268

Xiao, L., Che, T., & Dai, L. (2020). Evaluation of remote sensing and reanalysis snow depth datasets over the Northern Hemisphere during 1980–2016. *Remote Sensing*, *12*(19), 3253. https://doi.org/10.3390/rs12193253

Yang, J., Jiang, L., Menard, C. B., Kari, L., Lemmetyinen, J., & Pulliainen, J. (2015). Evaluation of snow products over the Tibetan Plateau. *Hydrological Processes*, *29*, 3247–3260. https://doi.org/10.1002/hyp.10427

Yilmaz, M. T., & Crow, W. T. (2014). Evaluation of assumptions in soil moisture triple collocation analysis. *Journal of Hydrometeorology*, *15*, 1293–1302. https://doi.org/10.1175/jhm-d-13-0158.1

Yilmaz, M. T., Crow, W. T., Anderson, M. C., & Hain, C. (2012). An objective methodology for merging satellite-and model-based soil moisture products. *Water Resources Research*, *48*(11), W11502. https://doi.org/10.1029/2011wr011682

Zeng, J., Chen, K. S., Cui, C., & Bai, X. (2020). A physically based soil moisture index from passive microwave brightness temperatures for soil moisture variation monitoring. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(4), 2782–2795. https://doi.org/10.1109/tgrs.2019.2955542

Zeng, J., Li, Z., Chen, Q., Bi, H., Qiu, J., & Zou, P. (2015). Evaluation of remotely sensed and reanalysis soil moisture products over the Tibetan Plateau using in-situ observations. *Remote Sensing of Environment*, *163*, 91–110. https://doi.org/10.1016/j.rse.2015.03.008

Zhang, T. J. (2005). Influence of the seasonal snow cover on the ground thermal regime: An overview. *Reviews of Geophysics*, *43*(4), RG4002. https://doi.org/10.1029/2004rg000157

Zhang, R., Liang, T., Feng, Q., Huang, X., Wang, W., Xie, H., & Guo, J. (2016). Evaluation and adjustment of the AMSR2 snow depth algorithm for the Northern Xinjiang Region, China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(9), 3892–3903.