Investigating the Ability of Multiple Reanalysis Datasets to Simulate Snow Depth Variability over Mainland China from 1981 to 2018

HONGBO ZHANG, a,b,c,d FAN ZHANG, b,e,f TAO CHE, g WEI YAN, h AND MING YE i

a College of Water Resources and Civil Engineering, China Agricultural University, Beijing, China
b Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China
c State Key Laboratory of Cryospheric Science, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou, China
d State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing, China
e CAS Center for Excellence in Tibetan Plateau Earth Sciences, Nanjing, China
f University of Chinese Academy of Sciences, Beijing, China
g Northwest Institute of Eco-Environment and Resources, Lanzhou, China
h School of Geographic Sciences, Xinyang Normal University, Xinyang, China
i Department of Earth, Ocean, and Atmospheric Science, Florida State University, Tallahassee, Florida

(Manuscript received 16 October 2020, in final form 8 September 2021)

ABSTRACT: Though the use of reanalysis datasets to analyze snow changes is increasingly popular, the snow depth variability in China simulated by multiple reanalysis datasets has not been well evaluated. Also, the extent of regional snow depth variability and its driving mechanisms are still unknown. In this study, monthly snow depth observations from 325 stations during the period of 1981–2018 were taken to evaluate the ability of five reanalysis datasets (JRA-55, MERRA-2, GLDAS2, ERA5, and ERA5L) to simulate the spatial and temporal variability of snow depth in China. The evaluation results indicate that MERRA-2 has the lowest root-mean-square deviation of snow depth and a high spatial correlation coefficient with observations. This may be partly related to the high accuracy of precipitation and temperature in MERRA-2. Also, the 31 combinations of the five reanalysis datasets do not yield better accuracy in snow depth than MERRA-2 alone. This is because the other four datasets have larger uncertainty. Based on MERRA-2, four hot-spot regions with significant snow depth changes from 1981 to 2018 were identified, including the central Xinjiang (XJ-C), the southern part of the northeastern plain and mountain (NPM-S), and the southwestern (TP-SW) and southeastern (TP-SE) portions of the Tibetan Plateau. Snow depth changes mostly occurred in spring in TP-SW and winter in XJ-C, NPM-S, and TP-SE. The snow depth increase in XJ-C, NPM-S, and TP-SW is mainly caused by increased seasonal precipitation, while the snow depth decrease in TP-SE is attributed to the combined effects of decreased precipitation and warming temperature in winter.

KEYWORDS: Atmosphere; Snow; Reanalysis data; Climate variability

1. Introduction

Snow affects land surface energy balance due to its high albedo and low thermal conductivity. Snow variability is closely related to regional climate changes such as rapid warming on the Tibetan Plateau (TP) (Guo et al. 2019a; You et al. 2019) and the interannual changes of the East Asian monsoon (Li and Wang 2011; Pu and Xu 2009). Snow is also an important water source in mountainous areas, especially for the arid regions in northwest China (Cao et al. 2018). The analysis of snow variability is essential to studying regional climate changes and managing water resources in China and worldwide. In situ snow measurements (e.g., snow cover and snow depth) fail to represent the spatial variability of snow (Meromy et al. 2013). Recently, remote sensing methods have been widely used to analyze snow variability (Zhang and Ma 2018). Snow cover, snow water equivalent (SWE), and snow depth data have been estimated from optical sensors (Huang et al. 2017; Yu et al. 2016), passive microwaves (Che et al. 2008; Dai et al. 2017; Takala et al. 2011; Wang et al. 2017; X. Xiao et al. 2020), and synthetic aperture radars (Lievens et al. 2019). However, these estimation methods have limitations due to high cloud cover (for optical sensors) (Stillinger et al. 2019) and large retrieval uncertainty under certain conditions (e.g., low snow depth or sensor saturation) (Xiao et al. 2018; H. Zhang et al. 2019). To improve the accuracy and consistency of snow estimations and facilitate the understanding of snow variation processes, several reanalysis datasets have been developed by assimilating snow-cover and snow depth observations (or remote sensing measurements) and/or by simulating snow accumulation, sublimation, and melting processes. For instance, the Global Land Data Assimilation System, version 1 (GLDAS1), reduces snow prediction errors by assimilating MODIS snow-cover data (Rodell and Houser 2004). The European Centre for Medium-Range Weather Forecasts Interim land surface reanalysis (ERA-Interim/Land) dataset reduces the negative bias of snow-covered areas by improving snow schemes (Dutra et al. 2010).
Several snow reanalyses have recently been exploited to analyze snow variability at a regional scale. Mudryk et al. (2015) compared five global SWE datasets [e.g., GLDAS2, ERA-Interim/Land, Modern-Era Retrospective Analysis for Research and Applications (MERRA)], and they found that winter SWE of the Northern Hemisphere decreased from 1981 to 2010. Using snow-cover estimations from seven observation-based global snow datasets (e.g., ERA-Interim/Land, GLDAS2, MERRA) as an observational reference, Thackeray et al. (2016) investigated the snow-cover trends through the simulation on 15 different phase 5 of the Coupled Model Intercomparison Project (CMIP5) models. The reanalysis SWE or snow depth data have also been incorporated into several ensemble datasets (e.g., the Canadian Sea Ice and Snow Evolution Network SWE dataset), and the datasets have been used for analyzing snow variability (Mudryk et al. 2018; You et al. 2020) and evaluating historical snow depth provided by CMIP5 (Li et al. 2018). Since existing reanalysis snow datasets are different (Mudryk et al. 2015), the snow evaluations are different based on the datasets. It is necessary to investigate the underlying reasons for the differences between the datasets so that the datasets can be improved (Wright et al. 2019) and an optimal dataset can be selected to meet specific project needs.

Based on long-term and continuous snow observations, a comprehensive validation and comparison of multiple reanalysis snow datasets is important for characterizing the uncertainty in snow reanalyses and evaluating the applicability of the reanalyses for snow trend investigation. There has been a plethora of studies that evaluated snow cover, snow depth, and SWE based on various reanalyses (Brown et al. 2010; Khan et al. 2008; Peings et al. 2013). Compared to remote sensing data of snow cover, in situ observations of snow depth and SWE have significantly longer records and are more suitable for calculating snow mass and volume (Dong 2018). Mortimer et al. (2020) recently evaluated SWE in the Northern Hemisphere with six reanalysis datasets and three satellite-based products. Their study was based on in situ measurements mainly from Russia, Scandinavia, and Canada. However, this study did not include the measurements from the stations in China. The snow depth observations in China have a large coverage area and a long observation period, so they are useful for evaluating reanalysis snow datasets and trend analysis that requires long-term and continuous records. Several studies have used in situ observations to analyze snow depth trends and their driving factors in China (Huang et al. 2019; Ma et al. 2020; X. Zhang et al. 2021). For example, based on long-term daily snow depth observations, Huang et al. (2019) found an overall increasing trend of snow depth in China from 1952 to 2012. However, because there are fewer observation stations in high-elevation areas such as the TP (Bian et al. 2020), the spatial distribution of the stations may affect the effectiveness of snow depth variability analysis. Meanwhile, the lack of observations of important related variables (e.g., snowfall and snowmelt) also limits the understanding of the mechanisms of snow depth changes. Therefore, it is necessary to identify optimal snow reanalyses that are suitable for analyzing snow depth variability and its driving factors in China.

Currently, only a few studies have evaluated snow depth and its variability in China (including spatial patterns and temporal trends) by simulations on reanalysis datasets (Wegmann et al. 2017; L. Xiao et al. 2020). L. Xiao et al. (2020) compared the snow depth accuracy of three remote sensing products and two reanalysis datasets over the Northern Hemisphere. They found that, in China, MERRA-2 has a lower root-mean-square deviation (RMSD) and a higher correlation coefficient (R) than other remote sensing data, such as the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), AMSR2, and the Global Snow Monitoring for Climate Research (GlobSnow). Orsolini et al. (2019) used snow depth observations from 33 stations of the Chinese Meteorological Administration (CMA) over the TP to evaluate the accuracy of four global reanalysis snow datasets. They found that the most recent ECMWF product (i.e., ERA5) and the Japanese 55-year Reanalysis (JRA-55) overestimate snow depth due to excessive snowfall. Also, they found that it is difficult to validate a snow depth trend because the observations used in their study only span five years from 2009 to 2013. Bian et al. (2019) evaluated 14 SWE products (including nine reanalysis datasets) based on the SWE observations from 40 stations of the TP for the seven years from 2003 to 2009. They pointed out that none of the SWE products could represent the spatiotemporal patterns of SWE well.

The main objectives of this study are as follows: 1) to evaluate the accuracy of snow depth by simulations on five reanalysis snow datasets; 2) to investigate the suitability of five reanalysis snow datasets for analyzing spatial patterns and temporal trends of snow depth in China; 3) to determine the optimal reanalysis dataset for analyzing regional snow depth changes and the underlying mechanisms. The remainder of this paper is organized as follows. Section 2 describes the reanalysis datasets and the observations of snow depth obtained from 1981 to 2018 from 325 weather stations in China. To the best of our knowledge, this is the first time that such a large number of observations are used to evaluate the spatial patterns and temporal trends of reanalysis snow depth in China. Also, the methods used in the study are introduced in this section. Section 3 discusses the validation and comparison results of the reanalysis snow depth and its ability to simulate snow depth variability. Finally, section 4 concludes this paper.

2. Data and method

a. In situ observations of snow depth and other meteorological variables

The observations were selected from more than 900 CMA stations based on two criteria: 1) for any month, the number of days with missing data does not exceed two; 2) there are consecutive monthly mean snow depth records without missing observations for at least 30 years from 1981 to 2018. The 30-y criterion was selected following the research work conducted by Zhou et al. (2018). In this way, the monthly mean snow depth observations from 325 CMA stations in China (Fig. 1) were compiled and used as the ground truth in this study. The monthly mean snow depth values were calculated by aggregating daily snow depth that was manually measured with a ruler at 0800 local time (CMA 2017). Monthly mean snow
depth was used as the primary variable for evaluating snow reanalyses. Following the research work conducted by Wei and Dong (2015), the annual and seasonally averaged snow depths were used to analyze temporal trends, though the maximum snow depth is often used to analyze snow climatology (Wegmann et al. 2017). Besides, monthly total precipitation and monthly average air temperature data were calculated from the daily observations of each station, and these data were used to analyze the error sources of the snow depth reanalysis datasets.

b. Five reanalysis datasets

This study evaluated monthly mean snow depth from five major reanalysis datasets, including JRA-55, MERRA-2, GLDAS2, ERA5, and ERA5/Land (ERA5L). Other reanalysis datasets (e.g., ERA-Interim and MERRA) were not selected by this study, because they are no longer updated. The selected reanalysis datasets use data assimilation techniques to provide better estimations by optimally integrating land surface models (LSMs) and different types of observations. The LSMs are driven by various meteorological forcings (e.g., precipitation, air temperature, and radiation), and they have different degrees of complexity for different reanalysis datasets.

ERA5 exploits the LSM of the Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land (H-TESSEL) to simulate snow accumulation and melting with a single snow layer (de Rosnay et al. 2014). The outputs of the LSM-based snow model were taken as the background values. Then these values were further adjusted by using a two-dimensional “Optimal Interpolation” method (Brasnett 1999) that efficiently assimilates the remote sensing snow-cover information from the Interactive Multisensor Snow and Ice Mapping System (IMS) product and the snow depth observations from the synoptic (SYNOP) network (the stations from China were not included) (De Rosnay et al. 2015). ERA5L is a recent reanalysis dataset that reruns the land component of ERA5 with improved spatial resolution (Muñoz Sabater 2019; Muñoz Sabater et al. 2017). Although ERA5L does not involve data assimilation, it uses atmospheric forcings from ERA5 as model inputs. Then, the inputs are corrected with lapse rates to account for the elevation differences between the grids of ERA5 and ERA5L. Thus, the observations assimilated by ERA5 can indirectly affect the simulated snow depth of the ERA5L. JRA-55 employs an offline version of the Simple Biosphere Model to simulate snow-mass changes with a single snow layer (Kobayashi et al. 2015). The simulated snow depth was corrected once a day by assimilating snow depth observations and microwave satellite data of the snow cover. The assimilated observations include those from the CMA stations across China (Kobayashi et al. 2015). GLDAS2 adopts the Noah LSM with a single snow layer (Rodell et al. 2004), which is established by the observation-based “Global Meteorological Forcing Dataset from Princeton University” (Broxton et al. 2016; Rodell et al. 2004; Sheffield et al. 2006). Also, GLDAS2 does not involve snow data assimilation (Bian et al. 2019). For MERRA-2, the “Catchment model” with a three-layer snow scheme was used (Stieglitz et al. 2001). Although the meteorological observations such as precipitation, were assimilated (Gelaro et al. 2017), there was no snow data assimilation in MERRA-2 (Orsolini et al. 2019). The detailed information of the five reanalysis datasets is summarized in Table 1.

Generally, there are two ways to compare the reanalysis data and station observations. One way is to directly compare the station observation with the corresponding gridcell values (Wang and Zeng 2012; Xie et al. 2014), and the other way is to interpolate the reanalysis data at the station location (H. Zhang et al. 2021). Following the research work conducted by Bracegirdle and Marshall (2012) and Wang et al. (2019), bilinear interpolation was used in this study to validate all the reanalysis variables. For each station, the values obtained from the four nearest reanalysis blocks were interpolated at the station location, and the interpolated values were compared with the corresponding station observations. Although the bilinear interpolation can partly alleviate the scaling issue by accounting for the offset between a station location and the corresponding block center of the station, it may introduce additional errors (Wang and Zeng 2012). This problem is discussed later in this paper.

c. Evaluating reanalysis snow depth and identifying the optimal dataset

In this study, the RMSD and Pearson correlation coefficient R were taken as two primary metrics to evaluate the accuracy of the reanalysis snow depth. To show the spatial distribution of snow depth accuracy, the reanalysis dataset was validated by calculating the RMSD and R values for each station using monthly snow depth observations. Meanwhile, the Taylor diagram (Taylor 2001) was also used to show the combined information of RMSD and R to present the overall accuracy comparison results among the five reanalysis datasets based on monthly snow depth data from all 325 stations.

The performance of the reanalysis datasets in simulating the spatial and temporal variability of snow depth was determined by calculating spatial correlations and comparing temporal trends.
Following the research work of Lievens et al. (2019), the spatial correlation was obtained by calculating the $R$ values between the multiyear average snow depth from the observations and reanalysis data at the 325 stations. The temporal trend was obtained through a linear regression between the annual mean snow depth and the time series, and the significance level ($P$ value) of the trend was calculated based on the $t$ test. To calculate the annual mean snow depth, a snow year was defined as the period from 1 September to 31 August of the following year (Ke et al. 2016). The snow depth trend was validated by calculating the RMSD between the trends of the observations and the trends of the reanalysis. The sign (+ or −) of the trend may affect the validity of trend analysis results, and the reliability of trend analysis relies on statistical significance. Thus, a consistency index (CI) is defined as

$$CI = \frac{N_{inc} + N_{dec} + N_{no}}{N_{tot}},$$

(1)

where $N_{inc}$ is the number of stations where both the reanalysis and observed trends increase significantly ($P < 0.05$); $N_{dec}$ is the number of stations where both trends decrease significantly; $N_{no}$ is the number of stations where both trends show insignificant variations; and $N_{tot}$ is the total number of stations. A higher CI value indicates better performance in trend simulation. It should be noted that “unbalanced” samples may impact the value of CI in specific situations. For example, if there are a large number of insignificant variations, the ability of CI to capture significantly decreasing/increasing trends is reduced.

Previous studies have indicated that an ensemble-based method using multiple reanalysis datasets can potentially improve snow reanalysis because the method considers the differences in model structure (or complexity) and the uncertainty under different conditions of the datasets (Mortimer et al. 2020; Mudryk et al. 2015; Rutter et al. 2009). To investigate whether an optimal dataset or an ensemble mean of multiple datasets is suitable for the simulation of snow depth changes in China, a detailed comparison between all the possible combinations of the five reanalysis datasets was conducted by creating 31 groups of ensemble datasets. For each group, the output from the ensemble datasets was calculated as an equally weighted average of all the reanalysis datasets of the group (Mudryk et al. 2015).

d. Analysis of error sources in forcing data

Considering the generation process of snow reanalyses, errors in snow depth may be caused by two factors, that is, LSMs driven by forcing data and the snow data assimilation process. DeRosnay et al. (2014) found that, though most LSMs used in reanalysis can capture the dynamics of snow cover, the accuracy of snow-mass (including snow depth) simulation is largely affected by the uncertainty in meteorological inputs and imperfect snow schemes. Precipitation and air temperature are two key meteorological inputs to snow models (Zhang et al. 2015), and LSMs are generally sensitive to precipitation (Mortimer et al. 2020; Orsolini et al. 2019). Thus, the errors in precipitation and air temperature (e.g., the RMSD between the precipitation/air temperature observed from stations and those from reanalysis data) may be major error sources of reanalysis snow depth. Here, the error source analysis is mainly based on the correlation analysis between the errors of snow depth and those of precipitation and air temperature. Other possible error sources, including differences in snow schemes and snow data assimilation, are discussed in section 3e. Similar to the case of snow depth, the “errors” of precipitation and air temperature were calculated by comparing the station records with the corresponding reanalysis outputs obtained by bilinear interpolation. To reduce the scaling effects on air temperature, a constant temperature lapse rate of 0.65°C (100 m)$^{-1}$ was used to correct the temperature in the vertical direction from a grid block to a station point before the bilinear interpolation. Meanwhile, the Shuttle Radar Topography Mission (SRTM) data with an original resolution of about 90 m were resampled (or averaged) to yield a coarser reanalysis grid to provide the elevation information for all the five reanalysis datasets.

e. Analysis of snow depth changes across mainland China and its driving mechanisms

The best dataset among the five reanalysis datasets was selected to analyze the snow depth trend in mainland China. For each reanalysis grid, the snow depth trend was calculated using the algorithm described in section 2c. The statistically significant ($P < 0.05$) changes in annual mean snow depth were considered as “hot spot” regions of snow depth changes. Meanwhile, the corresponding changes in air temperature, precipitation and snowfall data were also compared with the changes in snow depth to analyze the possible change mechanisms. Besides, correlation analysis was conducted between the meteorological variables and snow depth. Additionally, linear regression analysis was used to quantify the contribution of different meteorological variables (or their combinations) to snow depth change by calculating the $R^2$ value of the model following the work conducted by Zhang et al. (2020).

---

**Table 1. Summary of the five reanalysis datasets.**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Snow scheme</th>
<th>Available time</th>
<th>Spatial resolution (lon x lat)</th>
<th>Assimilating snow depth in China</th>
</tr>
</thead>
<tbody>
<tr>
<td>JRA-55</td>
<td>Simple Biosphere Model, 1-layer</td>
<td>1958–current</td>
<td>$0.5625\degree \times 0.5625\degree$</td>
<td>Yes</td>
</tr>
<tr>
<td>MERRA-2</td>
<td>Catchment model, 3-layer</td>
<td>1980–current</td>
<td>$0.625\degree \times 0.5\degree$</td>
<td>No</td>
</tr>
<tr>
<td>GLDAS2</td>
<td>Noah model, 1-layer</td>
<td>1948–2010a</td>
<td>$0.25\degree \times 0.25\degree$</td>
<td>No</td>
</tr>
<tr>
<td>ERA5</td>
<td>H-TESSEL model, 1-layer</td>
<td>1979–current</td>
<td>$0.25\degree \times 0.25\degree$</td>
<td>No</td>
</tr>
<tr>
<td>ERA5/Land</td>
<td>H-TESSEL model, 1-layer</td>
<td>1981–current</td>
<td>$0.1\degree \times 0.1\degree$</td>
<td>No</td>
</tr>
</tbody>
</table>

*aGLDAS-2.1 is used for extending the GLDAS-2 data during 2011–18.*
3. Results and discussions

a. Validation of monthly snow depth

The comparison results of the overall accuracy of monthly snow depth obtained from the five reanalysis datasets are shown in Fig. 2. MERRA-2 and JRA-55 achieved the best accuracy with RMSD values as low as 1.57 and 1.72 cm and \( R \) values as high as 0.81 and 0.82, respectively. ERA5 and ERA5L obtained the worst accuracy with their RMSD values of 5.98 and 6.58 cm and \( R \) values of 0.43 and 0.50, respectively. Since MERRA-2 and JRA-55 had similar accuracy, a further comparison was conducted on the 325 station-based RMSD (and \( R \)) values by using a paired \( t \) test with the Holm correction (Holm 1979) calculated by following the method of Zhang et al. (2020). The comparison results show that there was no significant difference between MERRA-2 and JRA-55 in terms of RMSD and \( R \) values (Fig. S1 in the online supplemental material). Besides, it is indicated that the RMSD values of GLDAS2 were not significantly different from those of MERRA-2 or JRA-55, but the \( R \) values of GLDAS2 were significantly lower than those of MERRA-2 and JRA-55 (Fig. S1).

The spatial distribution of the validation accuracy of all 325 CMA stations is illustrated in Fig. 3. Following the work conducted by G. Zhang et al. (2019), a regional comparison was conducted by dividing mainland China into six regions: TP, Xinjiang (XJ), Inner Mongolia and Loess Plateau (ILP), northeastern plain and mountain (NPM), eastern plain (EP), and Yunnan and Guizhou Plateau (YGP). The results for the six regions were consistent with the comparison results discussed above. For almost all regions, MERRA-2 and JRA-55 had the highest accuracy that was indicated by the highest \( R \) and the lowest RMSD values. The results for TP were consistent with those of Orsolini et al. (2019), which concluded that JRA-55 and MERRA-2 are the two best reanalysis snow depth datasets for TP. Similarly, Bian et al. (2019) pointed out that MERRA-2 had the best accuracy for SWE of the TP. It should be noted that the between-region comparison results based on \( R \) were different from those based on RMSD. As shown in Fig. 3f, the TP generally had the lowest \( R \) among the six regions, while TP, ILP, EP, and YGP had lower RMSDs than XJ and NPM if only JRA-55, MERRA-2 and GLDAS2 were considered. Surprisingly, both ERA5 and ERA5L had very large RMSDs in TP, whereas their RMSDs in the other regions were significantly smaller. The large errors in TP also had significant impacts on their overall performance. If the stations in TP were excluded, the RMSD values of ERA5 and ERA5L would respectively decrease from 5.98 to 3.40 cm and 6.58 to 4.09 cm, while the \( R \) values of ERA5 and ERA5L would respectively increase from 0.43 to 0.66 and 0.50 to 0.71; meanwhile, ERA5 and ERA5L would still have lower accuracy than JRA-55, MERRA-2, and GLDAS2, with the RMSD (\( R \)) values of 1.74 (0.82), 1.64 (0.81) and 2.45 cm (0.62), respectively. Besides, the validation of monthly results shows that the months with larger errors (RMSD > 2 cm) are clustered around winter (December–February) when the snow depth is relatively large (Fig. 4). In general, JRA-55 (Fig. 4a), MERRA-2 (Fig. 4b), ERA5 (Fig. 4d), and ERA5L (Fig. 4e) overestimated the monthly snow depth to different extents with mean monthly mean deviations (MDs) of 0.21, 0.03, 1.32, and 1.49 cm, respectively; GLDAS2 underestimated monthly snow depth with a mean monthly MD of −0.09 cm. To minimize the impacts of the magnitude of snow depth on the snow depth accuracy, normalized RMSD was further used, and it was obtained by dividing the RMSD by the average observed snow depth. In terms of seasons, the normalized RMSDs were relatively high in the winter months but low in summer months for all the five reanalysis datasets (Fig. S2). This result is consistent with the temporal pattern of RMSDs. In terms of regional accuracy (Fig. S3), the normalized RMSDs were the highest in TP, and the lowest in NPM for all the datasets, except for GLDAS2. This result is consistent with the regional comparison in terms of \( R \) (Fig. 3f).

b. Spatial and temporal variability

There are large differences in the performance of simulating the spatial variability of snow depth on the five reanalysis
datasets. For annual snow depth, JRA-55 and MERRA-2 achieved the best performance with spatial R values as high as 0.90 and 0.88, respectively (Fig. 5). The spatial R value of GLDAS2 was moderate (0.70), whereas those of ERA5L and ERA5 were as low as 0.54 and 0.42, respectively. The monthly snow depth results further demonstrate that MERRA-2 had a more stable performance than JRA-55 for simulating the spatial variability of snow depth. Though MERRA-2 and JRA-55 had similar spatial R values in cold months (November–April), MERRA-2 had higher spatial R than JRA-55 in warm months (May–October). Especially, for the “shoulder” months (May and October), the spatial R values of MERRA-2 (0.74 in May and 0.77 in October) were higher than those of JRA-55 (0.54 in May and 0.77 in October). A further investigation indicates that JRA-55 overestimated snow depth more frequently in May and October than MERRA-2 (Fig. S4). This is partly because JRA-55 assimilates certain microwave satellite snow data [i.e., the data from the Special Sensor Microwave Imager (SSM/I) and the Special Sensor Microwave Imager/Sounder (SSMI/S)] (Kobayashi et al. 2015) with a coarse resolution and large uncertainty in China (Dai et al. 2015). Since the snow depths of the summer months (June–August) and September are very low (Fig. 4f), a large number of extremely low or zero snow depth values can introduce a large uncertainty in the correlation analysis.

There were large errors (>0.20 cm decade$^{-1}$) in the temporal trends of snow depth (Fig. 6). While GLDAS2 obtained the lowest RMSD value in snow depth trends on the five datasets, the value was as high as 0.20 cm decade$^{-1}$ (Fig. 6h). Such accuracy may not be acceptable for the accurate prediction of snow depth trends considering that the average of observed trends is only 0.10 cm decade$^{-1}$. However, the CI values of the four reanalysis snow depth datasets (except ERA5) are larger than 0.7 (Fig. 6g), indicating that the four datasets can successfully predict the correct sign (+, −, or insignificant) of the snow variability for 70% of the time. Since ERA5 (Fig. 6e) predicted more significantly decreasing trends of snow depth than the observed (Fig. 6a), ERA5 achieved the lowest CI and the largest RMSD (Fig. 6h). Among the five reanalysis datasets, MERRA-2 was the best for simulating the temporal trends of snow depth because it obtained significantly higher CI values and its RMSD value was slightly higher than that of GLDAS2 (Fig. 6h).
c. The relations between precipitation/air temperature accuracy and snow depth accuracy

Because the snow depth obtained from the five datasets showed significantly larger errors in winter than in other periods of the snow season (Fig. 4), the error source analysis was conducted only for the three winter months. In a snow model, precipitation and air temperature determine the amount of snowfall that can be accumulated as snow and control snow melting (Weiland et al. 2015). Several model parameters related to precipitation and air temperature are also crucial for snow modeling. One such parameter is the precipitation gradient that depicts the precipitation amount in different elevation bands, and the other parameter is the critical temperature for distinguishing rainfall from snowfall (Zhang et al. 2015). The errors in precipitation/temperature were then compared with those in snow depth to analyze the relationship between them. Considering the representativeness of elevation and region, four stations were selected as example stations, including the Wudaoliang station in TP (with the highest elevation of 4612 m), the Bayanbulak station (with an elevation of 2458 m) in XJ, the Yumen station (with an elevation of 1526 m) in ILP, and the Jianping station (with an elevation of 462 m) in NPM (Fig. 1). The four sample stations were used to evaluate the relationship between the “month by month” errors in precipitation (or air temperature) and that in snow depth, as shown in Fig. 7 and S5–S7. The month-by-month error is the difference between the reanalysis data and the station observations for a given month. For a given station, the impacts of precipitation and air temperature on the snow depth were examined.

Fig. 4. (a)–(e) Monthly validation accuracy of the five reanalysis snow depth datasets. All the monthly results from the 325 CMA stations are averaged. The green bar represents RMSE, and the blue point represents MD. (f) Plot of the multiyear average monthly snow depth observed at stations.
and air temperature accuracy vary with different datasets. This is because the accuracy of snow reanalyses can also be affected by other factors (e.g., the data assimilation process and model structure), and more discussions are provided in section 3e. The overestimation of precipitation coincides with the overestimation of snow depth at the Wudaoliang station for ERA5 (Fig. 7g) and ERA5L (Fig. 7i). Although most of the precipitation differences are randomly distributed around zero, the outliers in Fig. 7e indicate that many snow depth underestimations in GLDAS2 could be related to substantial underestimations of precipitation. Similar precipitation underestimations were also observed at the Yumen station for MERRA-2 (Fig. S6c) and GLDAS2 (Fig. S6e). The precipitation errors were positively correlated with the snow depth errors at these stations. Similarly, air temperature overestimations (underestimations) can decrease (increase) the snowfall amount and/or facilitate (impede) snowmelt, and therefore may lead to underestimation (overestimation) of the snow depth. Hence, temperature errors were negatively correlated with snow depth errors. For example, the frequent snow depth

![Figure 6](https://example.com/figure6.png)

**Fig. 6.** Validation and comparison of the temporal trends of snow depth (mostly during 1981–2018) simulated on the five reanalysis datasets. The temporal trends of snow depth from (a) observations, (b) JRA-55, (c) MERRA-2, (d) GLDAS2, (e) ERA5, and (f) ERA5L. Summary of the (g) CI and (h) RMSD of the five reanalysis snow depth datasets calculated by comparing with observed trends from 325 stations.
underestimation by MERRA-2 at the Bayanbulak station is closely related to the overestimation of temperature (Fig. S5c). Similarly, the large snow depth overestimations by ERA5L at the Jianping station are consistent with the underestimations of air temperature (Fig. S7j).

To investigate the strength of the relationship between the accuracy of precipitation and temperature and that of the snow depth estimation of China, the correlation analysis for the four example stations was extended for all 325 stations, and the results are provided in the supplemental material (Figs. S8 and S9). For the four reanalysis datasets (except JRA-55), there were significant correlations between the errors in snow depth and the errors in precipitation and air temperature for more than half of the stations. While related studies have discussed the importance of precipitation accuracy for snow modeling or snow data reanalysis (Han et al. 2019; Orsolini et al. 2019; Weiland et al. 2015), our results indicate that accurate air temperature is also important, and it should be improved in the future. This is especially the case for JRA-55, ERA5, and ERA5L, because the impacts of air temperature errors are larger than those of precipitation errors for the three reanalysis datasets. It is not surprising that errors in forcing data have moderate effects on the snow depth errors of JRA-55 because JRA-55 has already assimilated snow depth observations in China.

The accuracy of snow depth, precipitation, and air temperature in winter from the five datasets was compared across all six regions, and the comparison results are shown in Fig. S10. It should be noted that the accuracy (i.e., RMSD of snow depth) shown here is different from that in Fig. 3f because the latter is based on the annual data. For winter months, although JRA-55 assimilated snow depth, it had slightly lower accuracy in snow depth than MERRA-2, with mean RMSD values of 1.91 and 0.68 cm in TP, and 1.73 and 1.23 cm in ILP, respectively. One possible reason is that JRA-55 had a substantially larger RMSD value in precipitation since it does not assimilate precipitation observations (Orsolini et al. 2019). Another reason is that the stations considered by JRA-55 for snow data assimilation may not include all the stations used in this study. For the other four datasets, MERRA-2 and GLDAS2 achieved better accuracy than ERA5 and ERA5L in precipitation and air temperature in most regions and better snow depth accuracy. The large regional differences in snow depth accuracy (Fig. 3) may be correlated with the spatial pattern of errors in precipitation and air temperature. MERRA-2 generally overestimates air temperature in mainland China. The largest RMSD in snow depth occurred in XJ, where a large overestimation of air temperature coincided with an obvious underestimation of snow depth (Fig. S10b). Similar to MERRA-2, GLDAS2 also overestimated air temperature, with the largest RMSD in XJ. However, GLDAS2 only slightly underestimated the snow depth in XJ, and the underestimation is potentially correlated with the influence of precipitation overestimation shown in Fig. S10c. GLDAS2 underestimated the snow depth across the six regions. This is consistent with the finding of Broxton et al. (2016) that GLDAS2 substantially underestimated snow mass in North America. ERA5 and ERA5L obtained similar regional results (Figs. S10d,e), with the largest RMSD in snow depth in XJ. Although ERA5 and ERA5L only slightly underestimated the snow depth in XJ, the underestimation is potentially correlated with the influence of precipitation overestimation shown in Fig. S10c. GLDAS2 underestimated the snow depth across the six regions. This is consistent with the finding of Broxton et al. (2016) that GLDAS2 substantially underestimated snow mass in North America. ERA5 and ERA5L obtained similar regional results (Figs. S10d,e), with the largest RMSD in snow depth in XJ.
This result is consistent with the fact that ERA5 and ERA5L had similar errors of precipitation and air temperature in the four regions in comparison with the other three datasets. It is noted that large precipitation errors but small snow depth errors appear in EP and YGP. This is mainly because the average snow depth in EP and YGP is the lowest among the six regions, given that EP and YGP are warmer with less snowfall and have shorter snow seasons.

Our study shows that ERA5 and ERA5L had the lowest snow depth accuracy on the five datasets. However, a recent evaluation conducted by Mortimer et al. (2020) found that ERA5 had better accuracy than GLDAS2 and MERRA-2 in simulating SWE of the Northern Hemisphere. The difference between the findings may be caused by several factors. Mortimer et al. (2020) evaluated SWE rather than snow depth, and the validation data from China was not included in this study. Also, Mortimer et al. (2020) did not include mountainous areas in their analysis, whereas our findings show that ERA5 and ERA5L obtained substantially low snow depth accuracy in TP as discussed above. Besides, our study also finds that ERA5 and ERA5L obtained significantly larger RMSD values of the temporal snow depth trends than the other three datasets (Fig. 6h). This may be explained by the fact that ERA5 and ERA5L assimilated IMS snow-cover data and the IMS data are only available after 2004. As indicated by Mortimer et al. (2020), the IMS data discontinuity around 2004 causes ERA5 (and ERA5L) to be unsuitable for trend analysis. This may also affect the CI values of ERA5 and ERA5L. Though the CI value of ERA5 was only 0.3 (Fig. 6g), it would increase to 0.58 for the period of 1981–2003 and 0.93 for the period of 2004–18, if the two periods before and after 2004 were considered separately.

The comparison results of the 31 combinations of the reanalysis snow depth datasets are summarized in Fig. 8. The combinations of JRA-55, MERRA-2, and GLDAS2 obtained similarly good accuracy in monthly snow depth. However, the accuracy decreased when ERA5 or ERA5L was added to the combinations (Fig. 8a). Compared to MERRA-2 (referred to as combination “M”) with a mean RMSD of 0.95 cm, only two combinations achieved higher accuracy (i.e., lower RMSD). One combination is JRA-55 and MERRA-2 (referred to as combination “JM”) with a mean RMSD of 0.88 cm, and the other combination is JRA-55, MERRA-2, and GLDAS2 (referred to as “JMG”) with a mean RMSD of 0.89 cm. Since the difference of RMSD-based accuracy between “M,” “JM,” and “JMG” was small, a multiple comparison was conducted on the RMSD values from the 325 stations using a paired t test with Holm correction. Figure 8b shows that there was no significant difference in the accuracy between “M,” “JM,” and “JMG.” The accuracy comparison in simulated snow depth trends shows that compared to MERRA-2 with a CI value of 0.85, no combination achieved a higher CI value (Fig. 8c). Thus, for the five reanalysis datasets, the use of a simple ensemble mean did not yield significantly better accuracy than using MERRA-2 alone for snow depth simulation in China. This is because that MERRA-2 had the best accuracy in almost all regions (Fig. 3). Also, in comparison with MERRA-2, JRA-55, and GLDAS2 had lower accuracy in several regions, making the ensemble mean uncertain. Therefore, there is no need to use an ensemble mean of the five datasets to analyze the spatiotemporal changes in snow depth in China.

Although MERRA-2 seemed near “top” in terms of all or most evaluation metrics, regions, or months, it should be noted that there is no best reanalysis dataset with the highest score in all the evaluation metrics. For example, MERRA-2 and JRA-55 performed equally well in terms of the RMSD and R values of snow depth, and GLDAS2 obtained the lowest RMSD value in predicting snow depth trends. MERRA-2 is the only dataset with a relatively good prediction accuracy for both snow depth and spatiotemporal changes. For example, JRA-55 obtained a significantly lower spatial R of snow depth than MERRA-2 in shoulder seasons; also, GLDAS2 obtained significantly lower overall R (Fig. 2) and spatial R (Fig. 5) of snow depth than those of MERRA-2. Because MERRA-2 achieved better accuracy in both the spatial R and the CI value, MERRA-2 was used to analyze the spatiotemporal changes of snow depth.

d. Snow depth changes over the Chinese mainland and the driving factors

Figure 9a shows the snow depth trends calculated from the MERRA-2 dataset. While there were no significant changes in snow depth over most areas of mainland China, this study identified four hot-spot areas with significantly increasing or decreasing changes in annual mean snow depth. As shown in Fig. 9a, the hot-spot areas are the central part of XJ (XJ-C), the southern part of NPM (NPM-S), the southwestern part (TP-SW), and the southeastern part of TP (TP-SE). For the seasonal changes shown in Figs. 9b–e, only the significant changes of winter snow depth were found for XJ-C, NPM-S, and TP-SE, indicating that the significant changes in annual snow depth in these areas are mainly caused by winter snow depth. The situation for TP-SW was complex because the snow depth in all seasons except for winter increased significantly. The changing rates of seasonal snow depth at the four hot-spot areas are summarized in Fig. 9f to further illustrate their seasonal contributions to snow depth changes. The results indicate that winter snow depth contributes the most to annual snow depth change for all the hot-spot areas, except TP-SW where the spring snow depth change was dominant over a year.

To investigate the driving factors of snow depth change in the four hot-spot regions, the trends of precipitation, snowfall, and air temperature were investigated on the MERRA-2 dataset, and the results are shown in Fig. S11. In XJ-C, the winter precipitation and snowfall significantly increased during 1981–2018, and the winter air temperature did not show a significant variation trend, resulting in a significantly increasing trend in snow depth. This was also true for NPM-S. For TP-SW, there was also a significantly increasing trend in spring precipitation, leading to a significant increase in the spring snow depth. The insignificant warming weakened the increase in snowfall compared to the increase in precipitation. There were combined effects of precipitation and air temperature in the TP-SE because winter air temperature significantly increased and winter precipitation and snowfall significantly decreased during 1981–2018. These two factors resulted in a significantly
decreasing trend in snow depth in the TP-SE. Such results are supported by linear regression analysis, which quantifies the proportion of variance in snow depth explained by models considering different factors (Fig. S12). Precipitation is a major factor in snow depth changes in all hot-spot areas because the $R^2$ values between precipitation and snow depth were always above 0.7, and the regression model using precipitation as the sole predictor could explain at least 49% of the variance in snow depth. In comparison with precipitation, snowfall had a higher $R^2$ with snow depth, and the models based on snowfall obtained larger $R^2$ values. Although air temperature had negligible effects on the snow depth of XJ-C, it had a relatively high value of $R$ with snow depth in all the other hot-spot areas. In particular, the linear regression models that considered precipitation (or snowfall), air temperature, and their interaction showed significantly higher $R^2$ values than the regression models that only considered precipitation (or snowfall). TP-SE is most sensitive to air temperature warming because the $R^2$ obtained from the model based on precipitation greatly increased from 0.55 to 0.85 after air temperature was further considered.

Previous studies found that snow depth generally increased along the direction of north of 40°N and decreased along the south of 40°N (Huang et al. 2020). This is partly consistent with
our finding that the two hot-spot areas, XJ-C and NPM-S, in northern China exhibited increasing trends and the TP-SE in the south showed the opposite trends. Based on daily snow depth observation from about 1500 stations, X. Zhang et al. (2021) recently found that the annual cumulative amount of snow depth during 1960–2014 showed increasing trends in northeastern and northwestern China and a decreasing trend in eastern TP. These findings are also consistent with ours. Due to the sparse stations, the snow depth trends of TP-SW are unclear. Bian et al. (2020) recently investigated the temporal changes of multiple snow-related indices for the period of 1980–2018, including the annual mean snow depth over TP based on MERRA-2, JRA-55, and a microwave satellite snow depth dataset. They did not find any significant trend in annual snow depth from MERRA-2 in either TP-SW or TP-SE. It should be noted that the snow year (August–July) defined in their study is different from that defined here (September–August) because the study areas are different. Meanwhile, based on the JRA-55 dataset, Bian et al. (2020) found a decreasing trend in annual mean snow depth in TP-SE especially in spring, but no significantly increasing trend in TP-SW. Such discrepancy between the two snow reanalyses may be attributed to the large uncertainty in meteorological forcing in TP-SW where there are few stations. Since the errors of precipitation and air temperature of MERRA-2 were relatively high in TP compared with other regions (Fig. S10b), there may be increased uncertainty in the snow depth trends in TP-SW obtained from the MERRA-2 dataset. Several recent studies have found that the air temperature in southwestern TP showed a cooling trend after 2000 (Guo et al. 2019a,b). This is consistent with our finding that the air temperature in TP-SW showed an insignificantly warming trend (1981–2018), yielding a
relatively weak effect on snow depth changes. Although snow depth may also be affected by other factors such as radiation and wind speed, all regression models considering precipitation and air temperature can explain at least 60% of the snow depth variance, indicating less contribution of other factors (Fig. S12).

e. Uncertainties and limitations

As shown in Fig. 4f, for most stations, the snow depths in warmer months (i.e., June–September) were close to zero, which may impact the evaluation accuracy. After these months were excluded, the overall RMSD values of all the five reanalysis datasets increased (Fig. S13). However, in terms of RMSD, the relative order of the datasets did not change. The RMSD value of MERRA-2 was still the lowest, although it increased from 1.57 to 2.81 cm.

In addition to the errors in precipitation and air temperature, data assimilation and model structure may also affect the snow depth predictions obtained from the reanalysis datasets. For example, JRA-55 overestimated (with an MD of 15.4 mm) the precipitation in TP, and the overestimation was similar to that of ERA5 (with an MD of 17.7 mm). However, JRA-55 had a significantly smaller overestimation (MD = 1.3 cm) of the snow depth in TP than ERA5 (MD = 15.9 cm), because JRA-55 assimilates the snow depth observations of TP. The assimilation process for snow depth could weaken the impact of forcing data. This could be partly correlated with the fact that compared with the other datasets, JRA-55 involves fewer stations showing significant correlations between snow depth errors and precipitation/temperature errors (Figs. S8 and S9). Although ERA5 and ERA5L have the highest spatial resolutions, they have significantly lower snow depth accuracy than the other three datasets. A possible reason is that ERA5 and ERA5L neither assimilated snow depth observations nor corrected precipitation or air temperature forcing with the observations in China. Besides, the reason for MERRA-2 achieving the best snow depth accuracy could be partly explained by the differences in model structure. For example, MERRA-2 runs an LSM with three snow layers, making MERRA-2 more complex and verisimilar than other datasets with a single snow layer.

Figure 6a shows that most of the observed snow depth trends in China are insignificant. This result is consistent with previous observation-based studies (Huang et al. 2019) that used different time series of data. The limited number of significant trends may degrade the capability of CI for evaluating the trend simulation on different reanalysis datasets. However, the consistency of insignificant trends between the reanalysis and observations is also important for comprehensively capturing the correct signs of snow depth trends. For example, JRA-55 (Fig. 6b), MERRA-2 (Fig. 6c), GLDAS2 (Fig. 6d), and ERA5L (Fig. 6f) successfully captured most of the insignificant trends, whereas ERA5 mistakenly took many insignificant trends as significantly decreasing trends (Fig. 6e).

For the high-elevation areas (e.g., the TP), large uncertainty was caused by the scaling issue, because the complex topography cannot be well depicted by the coarse reanalysis grids (Lievens et al. 2019). To investigate this, the absolute differences between the elevations of the station and the grid center were calculated. Then, a correlation analysis was conducted between the elevation difference and validation accuracy based on the data of 325 stations. The absolute R values of JRA-55, MERRA-2, and GLDAS were respectively only 0.12, 0.02, and 0.05, indicating very small scaling effects. However, the absolute R values of ERA5 and ERA5L were respectively 0.50 and 0.37, indicating larger scaling effects. Thus, a more efficient method that can obtain the altitudinal distribution of snow depth within the coarse grid should be developed to compare the reanalysis snow depth with station observations in the future. Besides, the scaling issue was partly alleviated by the application of bilinear interpolation. For most datasets, the interpolated snow depth showed a slightly higher validation accuracy than the original one. The average RMSD values of JRA-55, MERRA-2, GLDAS2, and ERA5L respectively decreased by 0.072, 0.014, 0.006, and 0.001 cm, although that of ERA5 increased by 0.009 cm. In addition to the uncertainty in meteorological forcing, the deficiency of the model structure may also affect the accuracy of snow depth. For example, blowing snow sublimation is estimated to account for approximately 30% of global surface sublimation (Déry and Yau 2002) and account for approximately 24% of the total snowfall in an alpine catchment in China (Zhou et al. 2014). However, LSMs generally do not consider blowing snow (Mortimer et al. 2020). This is the case for Noah LSM (Koren et al. 1999), which is driven by GLDAS2. The other four reanalysis datasets tested here also did not consider blowing snow (Orsolini et al. 2019). The problem of scarce observations in alpine regions may also introduce uncertainty to our analysis, especially for TP where only a few stations are available and no station is located above the elevation of 5000 m.

Although MERRA-2 achieved the best consistency index for simulated snow depth trends, uncertainty still existed in the gridded analysis of snow depth changes due to the errors in the reanalysis forcing and model deficiency. It should also be noted that although most of the signs of the trends at stations were successfully predicted by MERRA-2, the magnitude of snow depth trends may not be well simulated by the reanalysis data according to the validation results shown in Fig. 6h. Future studies should thus improve the accuracy of meteorological forcing and develop models at finer resolutions to simulate the snow depth changes in China more accurately.

4. Conclusions

This study evaluated the accuracy of monthly snow depth estimated on five reanalysis datasets and examined the ability of these datasets to simulate the spatial and temporal variability of snow depth in China. Based on the long-term (mostly during 1981–2018) consecutive monthly snow depth observations from 325 stations, the snow depth trends of China simulated by multiple reanalysis datasets were evaluated and compared. It is found that MERRA-2 and JRA-55 had the highest accuracy in terms of monthly snow depth. The two datasets obtained similar RMSD and R values for snow depth, although only JRA-55 assimilated snow depth observations. This result may be correlated with the fact that MERRA-2 has better accuracy in precipitation and has a more complex snow
modeling scheme. Because MERRA-2 has a higher spatial correlation coefficient and consistency index value, it can simulate the spatial patterns and temporal trends of snow depth better. Besides, the discrepancy between the snow depth accuracy of the five reanalysis datasets is related to their different snow modeling schemes and their differences in the accuracy of precipitation and air temperature inputs. For most of the reanalysis datasets, the errors in both precipitation and air temperature were significantly correlated with the errors in snow depth. This was observed at more than half of the stations. In addition, the reanalysis datasets with lower accuracy in snow depth also had lower accuracy in precipitation and air temperature.

Meanwhile, it is suggested that, over China, an ensemble mean of multiple reanalysis datasets does not yield higher accuracy in snow depth or its trend than using MERRA-2 alone. This is possibly due to the larger uncertainty in the individual datasets used in the ensemble predictions. The four hot-spot areas (XI-C, NPM-S, TP-SW, and TP-SE) were detected on the MERRA-2 dataset, and regional and seasonal differences were observed. Precipitation is the major factor driving snow depth changes in China, and temperature warming also significantly affects the snow depth in TP-SE. Considering the significant impacts of precipitation and air temperature errors on snow depth simulation, more accurate meteorological forcing data need to be obtained to further improve the accuracy of reanalysis snow simulation. The findings of this study are significant for selecting optimal reanalysis snow depth datasets and understanding regional snow depth changes and their controlling mechanisms across mainland China.

Acknowledgments. The authors declare that they have no conflict of interest. This study is supported by the National Natural Science Foundation of China (Grants 41988101-03 and 41701079), State Key Laboratory of Cryospheric Science, Northwest Institute of Eco-Environment and Resources, Chinese Academy Sciences (Grant SKLCS-OP-2020-13), State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute (Grant 2019kmm02), the Talent Cultivation and development Support Program by China Agricultural University (Li Sien), and the Natural Science Foundation of Jiangsu Province (Grant BK20181059).

Data availability statement. All the reanalysis data are publicly available and can be freely downloaded from the internet. The meteorological observations (including monthly snow depth, precipitation, and air temperature) can be obtained from China Meteorology Administration (http://data.cma.cn) by access requests.

REFERENCES
——, ——, ——, and M. Dalhousi, 2015: Snow data assimilation at ECMWF. ECMWF Newsletter, No. 143, ECMWF, Reading, United Kingdom, 26–31, https://doi.org/10.21957/klpxq6x.


Muñoz Sabater, J., 2019: First ERA5-Land dataset to be released this spring. *ECMWF Newsletter*, No. 159, ECMWF, Reading, United Kingdom, 8–9.


Wang, C., R. M. Graham, K. Wang, S. Gerland, and M. A. Granskog, 2019: Comparison of ERA5 and ERA-Interim near-surface air temperature, snowfall and precipitation over Arctic sea ice: Effects on sea ice thermodynamics and...


