The AntSMB dataset: a comprehensive compilation of surface mass balance field observations over the Antarctic Ice Sheet

Yetang Wang¹, Minghu Ding², Carleen H. Reijmer³, Paul C. J. P. Smeets³, Shugui Hou⁴, and Cunde Xiao⁵

¹School of Geography and Environment, Shandong Normal University, Jinan 250014, China
²Tibetan Plateau and Polar Meteorology Institute, Chinese Academy of Meteorological Sciences, Beijing 100081, China
³Institute for Marine and Atmospheric Research Utrecht, Utrecht University, Utrecht, the Netherlands
⁴School of Oceanography, Shanghai Jiao Tong University, Shanghai 200240, China
⁵State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China

Correspondence: Yetang Wang (yetangwang@sdnu.edu.cn) and Cunde Xiao (cdxiao@bnu.edu.cn)

Received: 21 January 2021 – Discussion started: 28 January 2021
Revised: 26 May 2021 – Accepted: 31 May 2021 – Published: 29 June 2021

Abstract. A comprehensive compilation of observed records is needed for accurate quantification of surface mass balance (SMB) over Antarctica, which is a key challenge for calculation of Antarctic contribution to global sea level change. Here, we present the AntSMB dataset: a new quality-controlled dataset of a variety of published field measurements of the Antarctic Ice Sheet SMB by means of stakes, snow pits, ice cores, ultrasonic sounders, and ground-penetrating radar (GPR). The dataset collects 3579 individual multi-year-averaged observations, 687 annually resolved time series from 675 sites extending back over the past 1000 years, and daily resolved records covering 245 years from 32 sites across the whole ice sheet. These records are derived from ice cores, snow pits, stakes/stake farms, and ultrasonic sounders. Furthermore, GPR multi-year-averaged measurements are included in the dataset, covering an area of 22,025 km². This is the first ice-sheet-scale compilation of SMB records at different temporal (daily, annual, and multi-year) resolutions from multiple types of measurement and is available at https://doi.org/10.11888/Glacio.tpdc.271148 (Wang et al., 2021). The database has potentially wide applications such as the investigation of temporal and spatial variability in SMB, model validation, assessment of remote sensing retrievals, and data assimilation. As a case of model estimation, records of the AntSMB dataset are used to assess the performance of ERA5 for temporal and spatial variability in SMB over Antarctica.

1 Introduction

Against the background of rapid global warming, wide international concerns have arisen about changes in the Antarctic Ice Sheet (AIS) mass balance, which positively contributed 14.0 ± 2.0 mm to global sea level rise over 1979–2017 (Rignot et al., 2019). Antarctic mass balance is dependent on the partitioning between ice discharge into the ocean and net snow accumulation at the surface, i.e., surface mass balance (SMB). Recent negative mass balance of the ice sheet reflects larger ice dynamical loss than mass gain from SMB (e.g., Shepherd et al., 2012, 2018). Despite the responsibility of ice discharge for Antarctic mass balance on decadal or longer timescales, considerable inter-annual variability is largely determined by fluctuations in SMB (Rignot et al., 2019). Because annual net mass input into the entire ice sheet through snowfall is equivalent to about a 6 mm global sea level decline (IPCC, 2019), even any small fluctuations in
the Antarctic SMB can result in large variability and trends in global sea level.

SMB is defined as the sum of precipitation, surface and drifting snow sublimation, erosion/deposition caused by drifting snow, and surface meltwater runoff. Since the International Geophysical Year 1957–1958 (IGY), a number of scientific Antarctic traverses/expeditions have been performed with the goal of obtaining SMB measurements by means of stakes, ice cores/snow pits, ultrasonic sounders, or ground-penetrating radar (GPR) (e.g., Isaksson and Melvold, 2002; Mayewski et al., 2005). Due to logistical constraints in the harsh environment, gaps in the spatial coverage of SMB measurements are still large, and long-term samplings are also scarce (Favier et al., 2013). As a result, substantial caveats have been encountered when quantifying SMB at the ice sheet scale by using simple interpolation of these observations (Magand et al., 2007; Genthon et al., 2009). Climate models and various atmospheric reanalysis products provide an important choice of how to assess SMB for large areas. The outputs of regional climate models have been used to calculate ice sheet SMB in recent decades by a wealth of Antarctic mass change estimate studies (e.g., Rignot et al., 2011; Shepherd et al., 2012; Rignot et al., 2019). However, these simulations depend on ground-based observations to improve their accuracy and resolution. Before application, the model’s performance needs to be carefully assessed based on in situ observations, as performed by some previous studies (Medley et al., 2013; Wang et al., 2015; Van Wessem et al., 2018; Agosta et al., 2019; Wang et al., 2020). To improve the ice sheet SMB estimates, field measurements have been used by cross comparison with remotely sensed data (Arthern et al., 2006) or outputs of the climate models (e.g., Monaghan et al., 2006a; Van de Berg et al., 2006; Medley and Thomas, 2019; Wang et al., 2019). Thus, it is still pivotal to compile all available observations from the past to present to better estimate spatial and temporal variability in SMB and to constrain climate models and remote sensing algorithms.

Vaughan and Russell (1997) performed pioneering work to compile all multi-year-averaged SMB field measurement data over the AIS, and this compilation was introduced in detail by Vaughan et al. (1999). However, according to Magand et al. (2007), this dataset includes a lot of unreliable data and should be used with caution. To improve this, Favier et al. (2013) updated the database using the new field measurements carried out during 1999–2012, through quality control proposed by Magand et al. (2007). Recently, several compilations of SMB measurements at an annual resolution have been published (e.g., Mayewski and Dixon, 2013; Altnau et al., 2015; Thomas et al., 2017; Montgomery et al., 2018). In spite of numerous field measurements in these datasets, most cover only a limited area of the AIS. In particular, these datasets missed a large number of annually resolved stake/stake farm observations, such as data from the Japanese Antarctic Research Expedition (JARE), the South Pole, and Vostok. In addition, available SMB measurements derived from GPR are not collected or at least not fully collected in these datasets. Furthermore, all available ultrasonic sounder data from automatic weather stations (AWSs) at daily or higher resolutions have not been compiled until now.

In this study, our objective is to generate a comprehensive SMB database for Antarctica, using all available measurements by means of stakes/stake networks, snow pits/ice cores, GPR, and ultrasonic sounders, with control of data quality. This dataset includes SMB measurements at daily, annual, and multi-year resolutions, which can be applied to validation and calibration of climate models and remote sensing, developments of remotely sensed algorithms, examination of spatial and temporal patterns in Antarctic SMB, and estimates of the drivers of SMB changes across multiple scales. As a case of model validation, we make a comparison of the dataset with ERA5 reanalysis.

2 Description of the AntSMB dataset

2.1 Data collections and sources

We compile the dataset of SMB measurements over the AIS by searching the literature and public data portal platforms (e.g., the National Snow and Ice Data Center, NSIDC; PANGAEA; World Data Service for Paleoclimatology, NOAA), by collecting the supplements of publications, and by asking individual data generators to contribute their field measurements by email. If two or more request emails were not replied to, we consider the data unavailable to the public, and thus they are not included in this dataset.

The new data resources of the records in the database include GPR measurements over West Antarctic coastal zones during 2010–2017 (Dattler et al., 2019), over Thwaites Glacier in 2009 (Medley et al., 2013), and between Dome C and Vostok in 2012 (Le Meur et al., 2018) (Fig. 1). They cover an area of 22,025 km². A large number of new stake measurements were acquired by revisiting the traverses from Zhongshan Station to Dome Argus (Dome A) (Ding et al., 2015), from Syowa Station to Dome Fuji (Dome F) (Mohoyama et al., 2015), and between Progress Station and Vostok Station (Khodzher et al., 2014). In addition to new long-term ice core SMB records at the South Pole (Winski et al., 2019), this dataset includes previously published (in figures or tables) but unreleased (in terms of the original data) time series of SMB records from ice cores drilled over the Lambert Glacier basin (Xiao et al., 2001; Li et al., 2009; Ding et al., 2017). Furthermore, an important update of annually resolved SMB data results from the continuous stake network measurements performed at the South Pole, Vostok, and six sites of the traverse between Syowa Station and Dome F. In addition, this is the first public release of the published high-resolution ultrasonic sounder observations on Berkner Island (Reijmer et al., 1999; Reijmer and Van den Broeke, 2003), Dronning Maud Land (DML; Van den Broeke et al., 2004), the East Antarctic Plateau (Reijmer and Van...
den Broeke, 2003), and the Chinese transverse from Zhongshan Station to Dome A (Liu et al., 2019), which are very useful for the investigation of intra-annual and seasonal cycles of SMB. The other records of the database are obtained from existing SMB data compilations, including the multi-year-averaged SMB measurements by Favier et al. (2013) and Wang et al. (2016); time series of ice core records at an annual resolution by Mayewski and Dixon (2013), Altmann et al. (2015), and Thomas et al. (2017); and SMB component measurements over the Antarctic Ice Sheet and Greenland Ice Sheet (SUMup dataset) by Montgomery et al. (2018).

2.2 Selection criteria

In order to establish a comprehensive, complete, and quality-controlled AIS SMB product for a variety of scientific applications, quantitative criteria are designed for record inclusion in the database to center on high-resolution and well-dated records and to optimize data spatial coverage. The criteria are as follows.

Firstly, the records must be published through peer review or publicly available. The duration and temporal resolution of the records vary by measurement type. We select the ultrasonic sounder records with a minimum duration of 1 year. For annually resolved archives (ice core, stake, and stake network measurements), the duration of records included in this dataset should be at least 10 years but smaller than 1000 years. For the multi-year-averaged observations, the included records for the average span more than 3 years, which is the minimum number of years for an accurate estimation of the mean local SMB with an uncertainty of smaller than 10 % (Magand et al., 2007).

Secondly, the essential parameters for each SMB datum are provided, including location, measurement methodology, data time coverage, and references to the primary data sources.

Thirdly, the different kinds of record are quality-checked to the highest degree possible and then included in the dataset. (1) To ensure the multi-year-averaged SMB data are reliable at each site, we select the data determined by the anthropogenic radionuclides and volcanic horizons with errors of smaller than 10 % or stake measurements for more than 3 years, as suggested by Magand et al. (2007). The records with dating based both on stable isotopes and chemical markers and on natural radionuclides are reliable (Magand et al., 2007) and thus included in the dataset. We also include the available GPR-based snow accumulation rate data because their uncertainties can be below 5 % at a firm depth of 10 m and decrease with the increase in the depth after post-processing including interpretation of reflectors, correct density estimates, and proper calibration with ice cores (Spikes et al., 2004; Eisen et al., 2008).

(2) SMB records of annually resolved ice cores should be either cross-dated or layer-counted. Their chronology should include at least two age control points, with one near the youngest part and another near the oldest part of the time series (Stenni et al., 2017). Also, they must be confirmed by the data generator. Furthermore, ice core SMB records are corrected for the impact of firn density and the vertical strain rate profile (Thomas et al., 2017).

(3) The preliminary quality control for AWS snow accumulation data has been performed by the data owner by means of removing the null measurements and physically anomalous snow accumulation data (i.e., data outside of the initial and final accumulation values) (e.g., Braaten, 1997, 2000). Some high-frequency noise still occurs in the AWS snow accumulation data. To reduce the noise, we discard the data points outside of 1 standard deviation of a running daily value as done by Fountain et al. (2010) and Cohen et al. (2013).

2.3 Types of data measurement collected in the AntSMB dataset

2.3.1 Stakes

Stakes are the easiest and most traditional way to measure SMB. After placing a stake vertically in the snow or ice, relative variations in snow surface heights over a certain period can be determined by repeated measurements of the distance between the top of the stake and the surface. Changes in snow heights are multiplied by snow density to yield the corresponding SMB. This simple method has been widely applied over Antarctica by almost all national glacier surveys. However, in most cases, spatial representativity of a single-stake record is very limited due to large natural spatial variability and small-scale disturbance from post-depositional effects such as the interactions between the stake and local wind. To reduce the related uncertainties, stake lines along a transect or stake farms are often used (e.g., Frezzotti et al., 2005; Kameda et al., 2008; Ding et al., 2011). In particular, these measurements are useful for the investigation of the spatial distribution of SMB at the scale of less than a kilometer.

Given the repeated measurements, stake observations are only performed over easily accessible regions. Due to logistic constraints in the extreme environment of Antarctica, the time span for the measurements usually ranges from 1 year to several years or even more.

2.3.2 Snow pits and ice cores

Snow pits and ice cores are used to construct SMB changes in time by determining the age and density of different layers. The dating is dependent on the different time markers preserved in the column of the snow pit and ice core. The annual layer is dated through the counting of seasonal changes in various parameters including the stratigraphic, oxygen and hydrogen isotopic composition, major chemical ion content, hydrogen peroxide, and electric conductivity. When integrated with the prominent horizons of known age from volcanic or radioactive markers, accuracy of dating is largely
improved and results in time series of annual snow accumulation. Furthermore, the valuable reference horizons can also be used for the estimation of the SMB between horizons.

In Antarctica, counting annual layers based on the seasonal variations in multiparameter records combined with reference horizons can be used to calculate annual SMB in the high accumulation zones. However, seasonal cycles can hardly be identified at regions with low accumulation of smaller than 100 kg m$^{-2}$ yr$^{-1}$, especially for the East Antarctic Plateau. Thus, reference horizons may be the most reliable dating method in the low-accumulation areas, and a mean SMB can only be found between two reference horizons.

2.3.3 GPR

GPR maps firn stratigraphy along a profile from the surface, and the radar-identified firn layers with equal age along the continuous profiles can allow us to gain a detailed insight into SMB patterns. To calculate SMB, the isochronous layers must be well dated, which is usually dependent on the complementary depth–age of highly resolved ice core records along the radar profile. During the past few decades, grounded GPR has been widely used for the estimation of spatial variation in recent and historical SMB over Antarctica (e.g., Frezzotti et al., 2007; Anschütz et al., 2009; Müller et al., 2010). Most recently, the newly developed airborne radar systems have provided the revolutionized SMB measurements over the AIS (Kanagaratnam et al., 2004, 2007). They can robustly resolve the stratigraphy at the shallow (10 m) to intermediate (100 m) depths and hence measure annual and multi-year accumulation rates at the width of hundreds of kilometers along aircraft flight tracks. The systems were first developed by the Center for Remote Sensing of Ice Sheets and flown on the National Aeronautics and Space Administration Operation IceBridge (OIB) campaigns (Leuschen, 2010; Rodriguez-Morales et al., 2014). The AntSMB database contains records of grounded and airborne GPR observations for the 2009–2019 OIB campaigns.

Figure 1. The comprehensive dataset of Antarctic SMB observations. (a) Spatial distribution of available AWS observations. (b) Locations of available annually resolved SMB observations. (c) Locations of available multi-year-averaged SMB field data. Purple five-pointed stars standard for GPR measurements. Black points represent reliable SMB determined by stakes/stake farms and ice cores/snow pits. (d) Location of multi-year-averaged SMB data only during the second half of the 20th century selected for model validation.
Relative to point measurements such as stakes and snow pits/ice cores, the advantage of GPR observations is to yield a more accurate representation of spatial variations in SMB. Furthermore, the radar images of deep internal horizons allow us to quantify long-term variability in SMB. The errors in GPR-based SMB observations are associated with the depth and age of the reflector and extrapolation of density along the radar profile. The resulting uncertainties were estimated to be about 4% of the calculated SMB at a firm depth of 10 m and about 0.5% at the depth of 60 m after the calibration of depth and layer thinning and robust dating (isochronal accuracy of about 1 year) (Spikes et al., 2004).

2.3.4 AWS

In Antarctica, some AWSs equipped with ultrasonic sensors measure snow surface height changes by detecting the vertical distance to the surface. Combined with density observation, snow height changes can be converted to SMB. Despite the occasional poor quality when blowing snow or fog happens, this method can continuously yield SMB records of a high (typically hourly) temporal resolution (van den Broeke et al., 2004; Gorodetskaya et al., 2013), which can be utilized to identify individual accumulation/ablation events, to quantify seasonal cycle of snow accumulation, and also to calculate the surface energy balance coupled with other AWS observations.

In the same way as single ice core or stake observations, AWS measurements represent a single location, and spatial representativity is possibly limited. In addition, after collection of raw snow height data, the temperature-dependent speed of sound correction must be performed. The uncertainty in AWS height measurements is estimated to be ±1 cm or 0.4% of the distance to the surface (Eisen et al., 2008). This means that the measurements are not sufficient to examine the smaller snow accumulation events that usually occur in the interior of the East Antarctic Plateau.

2.4 Structure and metadata

The AntSMB dataset includes three subsets, i.e., (1) multi-year-averaged SMB observations from stakes, ice cores, and GPR measurements; (2) annually resolved SMB measurements from ice cores, stakes, and stake networks; and (3) AWS daily snow height measurements. To facilitate data reuse, subsampling, and reanalysis for scientific research efforts, each record in the three subdatasets includes some essential information, i.e., the names of measurement sites, site locations, measurement method, time coverage of the measurements, and citations. Site locations include latitude, longitude, and surface elevation. Each location of the measurements is in units of decimal degrees relative to the WGS84 ellipsoid. As listed in the dataset’s metadata, measurement techniques include firm/ice cores, snow pits, stakes/stake networks, ultrasonic sounders, and GPR. Table 1 summarizes the essential information for each measurement. Uncertainties in the measurement methods have been discussed in detail by Eisen et al. (2008).

Among the three subdatasets, the number of data in the multi-year mean SMB subdataset is largest, including unique measurements by radar isochrones with a coverage of 22,025 km², 2276 stake measurements, and 1303 ice core and snow pit observations (Fig. 2). The majority of these observations are derived from airborne snow radar measurements in the coastal zone of West Antarctica and the Antarctic Peninsula, the Ronne Ice Shelf, the South Pole, and Pine Island and Thwaites glaciers (Medley et al., 2013, 2014; Dattler et al., 2019). GPR data from two transects in East Antarctica account for 30% of all measurements in this subdataset. In most cases, SMB values come from original measurements. However, for the Japanese traverse route from Syowa Station to Dome F, we updated multi-year-averaged SMB by combining new stake surface height measurements during 2007–2013 with the improved snow density data from Wang et al. (2015).

The annually resolved SMB subdataset contains 687 time series of records, of which 79 records come from the compilation of ice core snow accumulation by Thomas et al. (2017), and 26 from the shallow-firm-core records in DML.

Figure 2. Bar charts indicating the number of the different types of measurement technique in the SMB observation dataset. The left bar demonstrates the distribution of approaches with the exclusion of high-resolution snow accumulation measurements and GPR measurements, and the area covered by GPR measurements is shown in the right bar.
Table 1. Brief description of metadata fields used in the Antarctic SMB observation database.

<table>
<thead>
<tr>
<th>Column</th>
<th>Name of field in database</th>
<th>Description</th>
<th>Format</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site name</td>
<td>Geo_siteName</td>
<td>Name of the site</td>
<td>Number code</td>
<td>Number code</td>
</tr>
<tr>
<td>Dataset ID</td>
<td>DatasetName</td>
<td>Specific identifier assigned to all SMB records from a given site and publication</td>
<td>Number code</td>
<td>Number code</td>
</tr>
<tr>
<td>Latitude</td>
<td>Geo_latitude</td>
<td>Latitude of the site</td>
<td>WGS84</td>
<td>Decimal degrees (−90 to 90°)</td>
</tr>
<tr>
<td>Longitude</td>
<td>Geo_longitude</td>
<td>Longitude of the site</td>
<td>WGS84</td>
<td>Decimal degrees (−180 to 180°)</td>
</tr>
<tr>
<td>Elevation</td>
<td>Geo_elevation</td>
<td>Elevation of the site</td>
<td>Height above the EGM geoid</td>
<td>m above sea level</td>
</tr>
<tr>
<td>Variable name</td>
<td>SMB</td>
<td>SMB in millimeters of water equivalent per year</td>
<td>Mass loss is defined as negative</td>
<td>kg m(^{-2}) yr(^{-1})</td>
</tr>
<tr>
<td>Method</td>
<td>Method</td>
<td>How each measurement was collected</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Start date</td>
<td>MinYear</td>
<td>Start date of measurement, i.e., minimum (oldest) year of each SMB record</td>
<td>Number</td>
<td>Year</td>
</tr>
<tr>
<td>End date</td>
<td>MaxYear</td>
<td>End date of measurement, i.e., maximum (more recent) year of each SMB record</td>
<td>Number</td>
<td>Year</td>
</tr>
<tr>
<td>Citation</td>
<td>Citation</td>
<td>Citation for the first publication presenting the SMB record</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

collected by Altnau et al. (2015). Continuous stake surface height measurements at a sub-annual resolution since the 1970s are available for the transverse route from Syowa Station to Dome F (Motoyama et al., 2015). We converted the measurements to SMB for the subdataset by multiplying snow height changes by snow density estimated from Wang et al. (2015).

AWS snow accumulation data are measured by the determination of the variations in the vertical heights between the sensor and snow using surface ultrasonic height rangers. The measurements are performed at 32 sites, of which 10 are located at DML, 7 at the Ross Ice Shelf, and 4 along the Chinese transverse route from Zhongshan Station to Dome A.

3 Spatial and temporal analysis of the AntSMB dataset

3.1 Spatial coverage of SMB records

The comprehensive observed SMB database collects SMB field data at the daily, annual, and multi-year scales from the whole AIS. Spatial distribution of the records is uneven within Antarctica. AWS snow accumulation measurements cover a wide range of areas, including the coastal zones of East Antarctica (McMorrow et al., 2001) and West Antarctica (van Lipzig et al., 2004) with high snow accumulation, the dry East Antarctic Plateau (Reijmer and Broeke, 2003; van den Broeke et al., 2004), the Ross Ice Shelf (Cohen et al., 2013), Berkner Island, the Lambert Glacier drain, and the McMurdo Dry Valleys (Doran et al., 2002) (Fig. 1a). Availability of time series from the annual resolution SMB subdataset is rich for West Antarctica, DML, Berkner Island, and the traverse from Syowa Station to Dome F (Fig. 1b). However, large parts of the Antarctic interior with low snow accumulation remain undocumented, which is easily understood because the seasonal stratigraphy in ice cores is almost unavailable at the regions with an accumulation of smaller than 100 kg m\(^{-2}\) yr\(^{-1}\) (Frezzotti et al., 2007, 2013). Compared with SMB compilation by Favier et al. (2013), spatial coverage of the multi-year SMB subdataset has greatly improved, especially for West Antarctica and the Antarctic Peninsula. Despite the improvement of spatial distribution, SMB records are still poor for the region from the Ronne Ice Shelf via the South Pole to Dome C and for the coastal zone of East Antarctica.

3.2 Temporal variability in the SMB records

The records in the comprehensive SMB dataset cover different time spans, ranging from a minimum of 1 year to a maximum of 1000 years. The covered time periods are closely associated with the measurement method. AWS provides very high resolution measurements of snow height changes, but the records generally span only a few years (1–18 years).
Although a significant advantage of ice cores is to record SMB changes over long time spans, it is difficult to perform these observations at a high spatial density. Stake farms are the easiest method to observe SMB, but continuous measurements are available for between several years and tens of years, largely due to the logistic constraints in the extreme Antarctic environment. GPR can detect the local SMB from the last tens of years to about 1000 years along continuous profiles of the snowpack. The temporal resolution of GPR measurements is dependent on the age estimates of reflection horizons, and the resulting records in our dataset range from decadal to centennial.

For the annually resolved SMB subdataset, of 183 time series from ice core and stake network measurements, 47 span from 1801 onwards (Fig. 3a). The number of time series peaks during the early 2000s when ice cores were retrieved in DML (Altnau et al., 2015) and the West AIS (Mayewski and Dixon, 2013). Prior to 1800, the number of time series decreases greatly, with only 10 with a duration beyond the past 500 years and 5 beyond the past 1000 years (Fig. 3a). The sharp decline since the mid-2000s results from a lack of coring efforts. Annually resolved stake measurements cover the past 40 years, peaking from the mid-1990s to the early 2000s (Fig. 3b).

For the multi-year-averaged SMB subdataset, 83% of the records with the exception of radar measurements cover less than 20 years, and 43% span less than 5 years. Figure 3c presents the distribution of years when these records were measured from 1950 onwards. The distribution of the measurements is relatively even until the 1990s, when the number of samples increases. The temporal coverage of radar observations ranges from 25 to 185 years.

### 4 Inter-comparison of the different types of SMB measurement

The dataset compiles the different types of SMB measurement including ice cores/snow pits, stakes, ultrasonic sensors, and GPR approaches. It is critical to investigate if the resulting data have systematic discrepancies due to the distinct measurement methods. In particular, the measurements by ice cores, stakes, and ultrasonic sensors are performed at the centimeter scale, whereas GPR samples are at the meter scale. Despite the scale difference, GPR measurements that are near 100 year averaged agree well with 5-year-averaged single-stake measurements at the corresponding locations along the transect near Talos Dome, with differences of around 10% (Frezzotti et al., 2007). Given that no existing observed SMB dataset can be used as an independent reference of the different types of Antarctic multi-year-averaged SMB observation, the inter-comparison of SMB determined by different methods at the same or near the same locations is made, as presented in Fig. 4. The locations are mainly distributed near Talos Dome, along a transect from Terra Nova Bay to Dome C, on western DML, and at Dome F and Dome A. It is clear that despite the differently averaged time coverage, the different SMBs provide a reasonable match with each other, with the largest discrepancy of less than 20%, which is consistent with previous, similar intercomparison (e.g., Vaughan et al., 2004; Frezzotti et al., 2005; Anschütz et al., 2009).

### 5 Comparison with the previous AIS SMB observation datasets

Here, we present an unprecedentedly comprehensive compilation of SMB observations at the daily, annual, and multi-year scales. For the compilation of the multi-year-averaged SMB ground-based observations including stakes/stake farms, snow pits/ice cores, and GPR, we apply the same quality control criteria as used in the compilation by Favier et al. (2013) and updated by Wang et al. (2016). Compared with the dataset, our compilation greatly improved the data spatial coverage by updating records using more recently published observations along the margins of West Antarctica; across Marie Byrd Land and the Antarctic Peninsula; around the South Pole; between Dome C and Vostok; and along the transects of Progress Station–Vostok Station, Dumont d’Urville–Dome C, Talos Dome, etc. We also updated SMB records along the transects of Zhongshan Station–Kunlun Station and Syowa Station–Dome F based on recent revisiting measurements. In particular, our dataset provides the first comprehensive compilation of grounded and airborne GPR measurements.

In terms of the collections of annual-resolution SMB measurements from ice cores, the SUMup dataset focused on the limited ice core records at West Antarctica from the US International Trans-Antarctic Scientific Expedition during the early 2000s (US ITASE; Mayewski and Dixon, 2013), several cores drilled in 2010–2011 (Medley et al., 2013), and records over DML and Berkner Island from the European Project for Ice Coring in Antarctica (EPICA; Oerter, 2008a–b). The Antarctica 2k database constructed by Thomas et al. (2017) included 80 ice core records spanning at least 30 years, and shorter and other ground-based measurement records were omitted. However, our AntSMB dataset focuses on the collection of annually resolved snow accumulation records from different kinds of measurement covering the whole ice sheet. As a result, this dataset contains 175 annually resolved ice core snow accumulation records, 8 stake network measurements covering at least 10 years, and 512 time series of continuous stake measurements spanning more than 18 years.

Previous SMB compilations centered on glaciological observations on annual and longer timescales (e.g., Vaughan et al., 1999; Frezzotti et al., 2013; Thomas et al., 2017), which are useful for the examination of trends and large-scale variability in AIS snow accumulation. Nevertheless, they do not
Figure 3. (a) Availability of records excluding the stake measurements along the traverse from Syowa Station to Dome F in the annually resolved SMB subdatabase over time during the past 1000 years. (b) Time coverage of the stake measurements along the traverse from Syowa Station to Dome F. (c) Histograms indicating the date taken of the multi-year-averaged SMB subdataset only including ice core and stake measurements.

Figure 4. Inter-comparison between different types of SMB measurement including AWSs, snow pits/ice cores, and GPR at 42 locations. The measurements are mainly distributed near Talos Dome, along a transect from Terra Nova Bay to Dome C, on western Dronning Maud Land, and at Dome F and Dome A.

give insight into SMB changes at much shorter timescales, such as at synoptic scales and following accumulation events. AWSs provide high-resolution (typically hourly) snow accumulation measurements, which is an advantage in quantifying the seasonal cycle of SMB and in examining the synoptic sources of individual accumulation events, relative to the other methods such as snow pits, ice cores, and stakes. Snow accumulation data from individual or several AWSs in the different sectors of Antarctica have been published by some previous studies (e.g., Reijmer and Van den Broeke, 2003; Thiery et al., 2012; Cohen et al., 2013; Thomas and Bracegirdle, 2015). However, these data have been not well compiled until now. Our dataset is the first attempt to collect all AWS snow accumulation measurements in Antarctica.
6 Comparison with ERA5

6.1 ERA5 output

Reanalysis utilizes a large number of observations assimilated into a numerical model to generate a spatially and temporally complete state of the atmosphere. Because the main assimilated data are atmospheric and oceanic measurements, reanalysis outputs are not entirely subject to the density of surface observations and thus have the potential to provide important information over the regions with few or even no surface observations. Recent studies have revealed that European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim) is likely to be the best or among the best reanalysis dataset for the representation of inter-annual variability in Antarctic precipitation (e.g., Bromwich et al., 2011; Wang et al., 2016).

ERA5 is the fifth-generation ECMWF reanalysis product produced by the Integrated Forecasting System (IFS) Cy41r2 operational from 2016 (Hersbach et al., 2020). Compared with ERA-Interim (∼80 km and 60 pressure levels), a major advantage of ERA5 is much higher horizontal and vertical resolutions (∼31 km and 137 pressure levels, respectively) and more enhanced outputs (hourly). Furthermore, IFS Cy41r2 includes a more advanced 4DVar assimilation scheme together with uncertainty estimation, and many more observations are assimilated. Detailed improvements can be found in Hersbach et al. (2020). This reanalysis dataset has replaced ERA-Interim, for which updates were stopped in August 2019. Here, our main objective is to know if ERA5 is able to provide a good SMB compared to the AntSMB observational dataset. Despite the recent release of ERA5 data extending back to 1950, we only use the outputs for the 1979–2018 period due to the spurious shift in reanalysis outputs in 1979 largely caused by changes in the number of assimilated observations (e.g., Zhang et al., 2018; Huai et al., 2019; Wang et al., 2020).

6.2 A subset of data used for the comparison with ERA5

6.2.1 Multi-year-averaged SMB observations

Given that the output of climate models centers on climate information since 1979 in Antarctica, it is necessary to define a special dataset for the model comparison. To match with the coverage period of the models, we only retain observations starting from 1950 onwards in the multi-year-averaged SMB subdatabase. In particular, we discard observations starting in the 1950–1978 period with a time coverage of no more than 10 years. Because blowing-snow processes are not schematized by ERA5, 190 measurements in blue-ice regions (SMB values < 0) are excluded. Finally, GPR measurements covering an area of 15 638 km² and 3184 multi-year-averaged observations are left for model–observation comparison.

6.2.2 Annually resolved SMB observations

To estimate the temporal performance of ERA5 for snow accumulation, we use the records covering at least 10 years from the annually resolved SMB subdatabase starting from 1979. This results in 159 time series of annually resolved SMB. The representativeness of SMB measurements at a simple site for a region is influenced by local noise from the interaction between wind and the local snow surface, especially in the regions with accumulation rates of less than 120 kg m⁻² yr⁻¹ (e.g., Frezzotti et al., 2005, 2007; Ding et al., 2011). This can be confirmed by the fact that on the DML plateau, ERA5-simulated individual annual SMBs highly correlate with each other (r > 0.70), but time series of SMB records from different ice cores are poorly correlated, even from the same drilling site. As a result, the relationships between ice core records and the corresponding ERA5 simulations at the drilling core location are variable, including significantly negative, positive, and insignificant correlations (Fig. 5a). Various linear relationships between the simulated and observed time series are also found over Berkner Island and the Ronne Ice Shelf with a high density of cores, with r values ranging from −0.35 to 0.67. In the two areas, the difference in the standard deviation of annual SMB values of individual ice cores is large, even for the records from the same locations (Fig. 5c). At the South Pole, ERA5 shows a significant correlation (r = 0.68; p < 0.05) with stake farm measurements but fails to do so with the individual ice core records (Fig. 5a). A main possibility is that SMB derived from stake networks is less noisy through the removal of small-scale spatial variability by using the average of many stakes together. At most of the sites, the standard deviation of annual ice core records is larger than the corresponding ERA5 simulations for their overlapping periods (Fig. 5d).

To reduce local noise and better assess the performance of ERA5, we first average the individual observation records in the same grid cell and then stack the averaged time series in the same geographic region. If there are ice core records and stake farm observations in the same location, the measurements of stake farms are utilized. Because the sites at the top of ice domes likely have minor local noise (Monaghan et al., 2006b), the four time series of ice core records from the ice domes are not discarded in the estimate. Following Frezzotti et al. (2007, 2013), a single ice core site with accumulations of more than 700 kg m⁻² yr⁻¹ allow the determination of annual SMB at a ±10% accuracy, which corresponds to the accuracy derived from the instrumental measurement, and hence the corresponding ice core records are retained. After the composite and filtering, 48 locations or regions with annually resolved SMB are left to compare with ERA5 simulations.
Figure 5. Spatial distribution of the correlation coefficients (a) between annually resolved SMB observation and ERA5 simulations for their overlapping period (circle: ice core; diamond: stake network) and (b) between averaged observed time series in the same location/region and the corresponding simulations from ERA5; (c) standard deviation of annual observed SMB and ERA5 simulations; (d) the ratios of the standard deviation of annual SMB from ERA5 simulations to observations.

6.3 Spatial performance of ERA5 output

A comparison of the density distribution of ERA5 precipitation minus evaporation \((P - E)\) with the filtered multi-year-averaged SMB observations reveals that the multi-year-averaged dataset is representative of the high accumulation zone but not for the bins with accumulation rates of 100–300 kg m\(^{-2}\) yr\(^{-1}\) over the West AIS (Fig. 6a). However, this dataset represents the entire \((P - E)\) spectrum of the model over the East AIS (Fig. 6b). As shown in Fig. 6c and d, the dataset also represents well the samples’ elevation distribution of SMB in relation to the West AIS and East AIS, especially between 200 and 1000 m elevations where it was not correctly sampled by the SMB observation dataset compiled by Favier et al. (2013).

ERA5 reveals large spatial gradients of snow accumulation over the AIS (Fig. 7a), with values higher than 1000 kg m\(^{-2}\) yr\(^{-1}\) at the margins and lower values (less than 30 kg m\(^{-2}\) yr\(^{-1}\)) in the hinterland of the East Antarctic Plateau. There is a very high correlation between ERA5 output and the observed SMB \((R^2 = 0.93\) and \(p < 0.01\)) which is calculated based on the logarithm of SMB values, due to the lognormal SMB distributions). The major spatial pattern of ERA5 simulations is in good agreement with the multi-year observations (Fig. 7a). Dry biases occur in most sites of the inland East AIS and the Ross Ice Shelf and around Byrd Station, whereas wet biases occur in East AIS margins and part of sites in West AIS coastal regions (Fig. 7b). The mean bias accounts for 6.6 % of the average of observed SMB, which is slightly higher than regional climate models (MAR and RACMO2.3p2) (Agosta et al., 2019; van Wessem et al., 2018). It is obvious that ERA5 robustly captures the sharp decrease in SMB with elevation (Fig. 7c). Compared with observation in each 200 m elevation bin, ERA5 is slightly wet below 1600 m in elevation, whereas dry biases occur in inland Antarctica with elevations above 3000 m.

6.4 Temporal performance of ERA5 output

A recent study showed that ERA5 presents relatively good skills for representing snow accumulation changes on the synoptic timescale, observed at the AWSs over the Ross Ice Shelf and along the traverse route from Zhongshan Station to Dome A, with 56 %–88 % of extreme snowfall events cap-
Figure 6. Relative frequency of ERA5 $P - E$ field data and gridded averaged records from the multi-year-averaged SMB subdatabase, with a bin range of $50 \text{ kg m}^{-2} \text{ yr}^{-1}$ on (a) the West Antarctic Ice Sheet (WAIS) and (b) the East Antarctic Ice Sheet (EAIS). ERA5 field data are bilinearly interpolated over a 30 km Cartesian grid. We average SMB for each $30 \times 30 \text{ km}$ grid cell (values from points located in the same grid cell are averaged), and then the number of grid cells in each bin is calculated. The contribution of the area of the elevation bin for ERA5 grid cells containing measurements and the entire elevation range to (c) the WAIS and (d) the EAIS. The 200 m elevation bins are used.

Figure 7. (a) Spatial distribution of ERA5 mean precipitation minus evaporation (approximated as SMB) for the period 1979–2018, and multi-year mean SMB measurements averaged on ERA5 grid cells ($30 \text{ km} \times 30 \text{ km}$). (b) ERA5 minus observed SMB averaged on the ERA5 grid cells. (c) Multi-year-averaged observations and ERA5 simulations, binned in 200 m elevation intervals. The number of ERA5 grid cells with in situ measurements in each elevation bin is shown by the blue line (right axis). Red lines denote SMB averaged by ERA5 for grid cells including observations within each elevation bin. Green lines denote SMB averaged by ERA5 over each whole bin of elevations.
Figure 8. Cumulative daily snow accumulation and snowfall over time for each station over Dronning Maud Land (a–k). (l) Spatial distribution of the AWS stations (notice that AWS3 station records are not included due to a number of missing data).

...tured (Liu et al., 2019). Given that these AWS observations are included in our AntSMB dataset, to avoid repetition, here we make a comparison between cumulative daily snowfall from ERA5 and the corresponding accumulation records from 11 AWS observations over DML (Fig. 8). Obviously, gaps in the AWS records occur in most stations because of the problems of sensors or data transmission. Snow accumulation decreases in the daily cumulative AWS records and reflects the important role of drifting snow, compaction, sublimation, or even ablation in the accumulation changes. Despite the noise of these post-deposition processes, stepped increases are observed for both ERA5 snowfall and AWS snow accumulation at each station in Fig. 8. Furthermore, the occurrence of large snowfall events is broadly consistent with the corresponding large accumulation events at all stations. These suggest, in spite of the limits of AWS measurements due to the complex impacts of post-deposition noise, they are very useful for evaluating synoptic changes in the precipitation from reanalysis products or climate models.

The correlation coefficients ($r$) between ERA5 simulations and SMB observations at 48 locations are shown in Fig. 5b. Significant and high correlations are observed at two out of five sites over the Antarctic Peninsula, with $r$ values of more than 0.7 ($p < 0.05$). Over the WAIS, ERA5 simulations are correlated significantly with observations ($r > 0.45; p < 0.05$) at 14 out of 18 sites, and correlation coefficients exceed 0.8 at 5 sites, suggesting relatively good skills of simulated records in capturing the observed annual variability in accumulation rates. Significant and positive correlations are present over the plateau and western and eastern coastal areas of DML. Correlation coefficients show a large fraction of inter-annual changes (> 70%) in SMB observations at Law Dome of Wilkes Land. The performances are relatively good for the South Pole, Vostok, and Talos Dome. However, no significant or negative correlations are observed at the sites...
of the Lambert basin, Princess Elizabeth Land, the middle DML coast, and Adélie Land.

To further assess the temporal performance of ERA5, we use the continuous time series of stake measurements along the JARE traverse route from Syowa Station to Dome F. These stake measurements are divided into four subgroups, as carried out for this traverse route by Wang et al. (2015). Stake measurements in each subgroup are stacked and then compared with the composites of ERA5 simulations for the respective subgroup (Fig. 9). ERA5 overestimates the observed SMB in the coastal and katabatic regions but underestimates that in the inland plateau region. The modeled records match particularly well with observations in the coastal, higher katabatic, and inland plateau regions, with higher $r^2$ values of $>0.5$. Observed SMB at the lower katabatic region is simulated well by the reanalysis dataset.

Overall, ERA5 fits inter-annual variability in observed SMB acceptably at most sites over the AIS, and this reveals much of atmospheric circulation is represented by this reanalysis product. Nevertheless, its performance is limited at some sites of the Lambert basin, inland West Antarctica, and parts of the East Antarctic coast. This may result from the unresolved processes in ERA5 such as drifting snow, the limited performance of ERA5 for the storm frequency related to synoptic-scale circulations, and sublimation because of circulation variations. Detailed interpretation of uncertainty in ERA5 is beyond the scope of this study.

### 7 Data availability

The comprehensive SMB observation dataset is available through A Big Earth Data Platform for Three Poles. The dataset can be downloaded from https://doi.org/10.11888/Glacio.tpdc.271148 (Wang et al., 2021). In this repository, the three subdatasets included in the entire dataset are provided in Excel spreadsheet format together with metadata files.

### 8 Discussion and conclusions

The dataset provides an unprecedentedly comprehensive compilation of SMB observations, with better spatial coverage than that of previous studies. In particular, our compilation greatly improves spatial density of measurements in the 200–1000 m elevations which are not correctly sampled by the dataset from Favier et al. (2013). However, there is a clear need to increase the spatial density of annually resolved SMB measurements over the inland East AIS, daily SMB observations over West Antarctica, and the 90–170° E sector of East Antarctica.

This dataset can be used to estimate the temporal and spatial changes in the AIS SMB. A temporally homogeneous climatology of SMB for the second half of the 20th century may be obtained by temporal rescaling of the multiyear-averaged SMB subdataset against ERA5 outputs as performed by Medley et al. (2019) and Wang et al. (2019). The available syntheses of time series of records from the annually resolved SMB subdataset will allow the investigation of regional snow accumulation changes during the past several decades or centuries (Kaspari et al., 2004; Frezzotti et al., 2013; Altnau et al., 2015; Thomas et al., 2017). The combination of the annual SMB subdataset with reanalysis products or the outputs of regional climate models can generate gridded datasets to better constrain the temporal and spatial

---

**Figure 9.** The left map shows the locations of stake measurements along the traverse between Syowa Station and Dome F and the regional boundaries. The four right charts show the comparison of the inter-annual variability in spatially averaged stake measurements and snow accumulation simulated by ERA5 for (a) the coastal region, (b) lower katabatic region, (c) upper katabatic region, and (d) inland plateau region.
variability in AIS SMB at the different scales (Monaghan et al., 2006a; Medley et al., 2019; Wang et al., 2019). The availability of AWS snow height measurements will allow insights into synoptic and seasonal patterns of SMB, which are vital for model estimation and ice core dating studies.

In the current study, we have made a comparison between observation data and ERA5 output. As a result, in spite of discrepancies in magnitude, ERA5 represents spatial variations in SMB observations well and captures a large proportion of the inter-annual variability. Similarly, this dataset can be used to evaluate the quality of other atmospheric reanalyses, and regional or global climate models such as JRA-55, MERRA-2, RACMO2.3, MAR, and CESM. Moreover, a high spatial density of stake and GPR measurements along several transections from coasts to inland is included in the dataset, which correctly samples the actual distribution of SMB and thus allow us to provide stringent constraints on the models in these specific regions. Annually resolved SMB observations in the database are also likely to be used as an important input of data assimilation for paleoclimate reconstructions (Dalaiden et al., 2020). The dataset is of vital importance for improvement of remote sensing algorithms for Antarctic snow accumulation/snowfall rates, such as the CloudSat 2C-SNOW-PROFILE product (Palmero et al., 2014; Behrangi et al., 2016).

The scientific community is expected to apply this dataset to Antarctic hydrological studies, model–data inter-comparison, and remotely sensed algorithm developments. The cryospheric community is also encouraged to further share their SMB observation data to update this dataset in the future.

**Supplement.** The supplement related to this article is available online at: https://doi.org/10.5194/essd-13-3057-2021-supplement.

**Author contributions.** YW designed and constructed the AntSMB dataset. MD, CHR, PCJPS, and CX contributed observed data to the dataset. YW wrote the manuscript, which was edited by MD, CHR, PCJPS, SH, and CX. All authors contributed to discussion and revision of the manuscript and approved the final version.

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

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

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

**Acknowledgements.** Thanks to Cécile Agosta, the anonymous reviewer, and the editor Tao Che for providing thoughtful and constructive comments and advice to greatly improve the quality of the manuscript.

**Financial support.** Funding for this work was provided by the Strategic Priority Research Program of the Chinese Academy of Sciences (grant no. XDA19070103), the National Natural Science Foundation of China (grant no. 41971081), the Project for Outstanding Youth Innovation Team in the Universities of Shandong Province (grant no. 2019KJH011), and the Outstanding Youth Fund of Shandong Provincial Universities (grant no. ZR2016JL030).

**Review statement.** This paper was edited by Tao Che and reviewed by Cécile Agosta and one anonymous referee.

**References**


Frezzotti, M.,2011.


Otter, H.: Annual means of δ18O and accumulation rates of snow pit DML78S05_13, https://doi.org/10.1594/PANGAEA.708115, 2008c.


Otter, H.: Annual means of δ18O and accumulation rates of snow pit DML82S05_17, https://doi.org/10.1594/PANGAEA.708119, 2008g.

Otter, H.: Annual means of δ18O and accumulation rates of snow pit DML83S05_18, https://doi.org/10.1594/PANGAEA.708120, 2008h.

Otter, H.: Annual means of δ18O and accumulation rates of snow pit DML84S05_19, https://doi.org/10.1594/PANGAEA.708121, 2008i.


