https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

Open Access Earth System Science Data Discussions

A Six-year circum-Antarctic icebergs dataset (2018-2023)

Zilong Chen[1,2,*], Xuying Liu[3,*], Zhenfu Guan[1], Teng Li[1,2], Xiao Cheng[1,2], Tian Li[4], Yan Liu[5], Qi Liang[1,2], Lei Zheng[1,2], and Jiping Liu[6]

[1]School of Geospatial Engineering and Science, Sun Yat-sen University, Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 51908, China
[2]Key Laboratory of Comprehensive Observation of Polar Environment (Sun Yat-sen University), Ministry of Education, Zhuhai 519082, China
[3]Institute of Artificial Intelligence, Shaoxing University, Shaoxing 312000, China
[4]Bristol Glaciology Centre, School of Geographical Sciences, University of Bristol, Bristol BS8 1SS, UK
[5]State Key Laboratory of Remote Sensing and Digital Earth, College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China
[6]School of Atmospheric Sciences, Sun Yat-sen University, Zhuhai 51908, China
[*]These authors contributed equally to this work.

Correspondence: Teng Li (liteng28@mail.sysu.edu.cn) and Xiao Cheng (chengxiao9@mail.sysu.edu.cn)

Abstract. The distribution of Antarctic icebergs is crucial for understanding their impact on the Southern Ocean's atmosphere and physical environment, as well as their role in global climate change. Recent advancements in iceberg databases, based on remote sensing imagery and altimetry data, have led to products like the BYU/NIC iceberg database, the Altiberg database, and high-resolution SAR-based iceberg distribution data. However, no unified database exists that integrates various iceberg scales and covers the entire Southern Ocean. Our research presents a comprehensive circum-Antarctic iceberg dataset, developed using Sentinel-1 SAR imagery from the Google Earth Engine (GEE) platform, covering the Southern Ocean south of 55°S. A semi-automated classification method that integrates incremental random forest classification with manual correction was applied to extract icebergs larger than 0.04 km^2 , resulting in a dataset for each October from 2018 to 2023. The resulting dataset not only records the geographic coordinates and geometric attributes (area, perimeter, long axis, and short axis) of the icebergs but also provides uncertainty estimates for area and mass. The dataset reveals significant interannual variability in iceberg number and total area-the number of icebergs increased from 33,823 in 2018 to approximately 51,332 in 2021, corresponding to major ice shelf calving events (e.g., the A68a iceberg in the Weddell Sea), followed by a decline in 2022. The annual average total iceberg area is 44,518 $\pm$ 4800 km^2, and the average mass is 8,779 $\pm$ 3,029 Gt. Validation using test set samples and a rolling cross-validation of interannual variability shows that the integrates incremental random forest classification achieves accuracy, recall, and F1 scores exceeding 0.90, and after manual correction, all performance metrics should be even better. Comparisons with existing iceberg products (including the BYU/NIC iceberg database and the Altiberg database) indicate a high consistency in spatial distribution, while our dataset demonstrates clear advantages in terms of spatial coverage, iceberg detection scale, and identification capabilities in regions with dense sea ice. This dataset serves as a novel data resource for investigating the impact of Antarctic icebergs on the Southern Ocean, the mass balance of ice sheets, the mechanisms underlying ice shelf collapse, and the response mechanisms of iceberg disintegration to climate change.

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

Open Access Earth System Science Data Discussions

1 Introduction

Icebergs are large freshwater ice masses that break off from the edges of ice sheets, ice shelves, or glaciers and enter the ocean. They are a critical component in the global climate system (Benn and Åström, 2018). Approximately half of the mass loss from the Antarctic ice sheet is discharged into the Southern Ocean through iceberg calving (Depoorter et al., 2013; Rignot et al., 2013; Liu et al., 2015). Annually, the dissolution of over 100,000 icebergs into the ocean is estimated to introduce a volume of freshwater that, according to certain calculations, exceeds the global annual freshwater consumption (Qadir et al., 2022; Orheim et al., 2023). This resultant freshwater influx plays a critical role in influencing the thermohaline characteristics, heat content, and freshwater balance within the impacted regions of the Southern Ocean (Gladstone et al., 2001; Hammond and Jones, 2016). On the bottom, grounding icebergs can interact with ocean floor and leave scours as a kind of geological record (Dowdeswell and Bamber, 2007; Li et al., 2018; Liu et al., 2021). Additionally, the nutrients carried by icebergs can influence the spatial distribution of primary productivity (Duprat et al., 2016), promoting the development of local ecosystems (Smith et al., 2007; Wu and Hou, 2017; Lin et al., 2024). Furthermore, icebergs pose a potential threat to maritime activities (Bigg et al., 2018), as human activity in the Antarctic region increases, accurate monitoring of iceberg distribution, size, and trajectory prediction has become critical (Evans et al., 2023)

The current databases on the distribution of Antarctic icebergs, as shown in Table 1, are primarily categorized into four types: (1) Ship-based observations, such as the SCAR International Iceberg Database (Orheim et al., 2023), compiled and published by the Norwegian Polar Institute (NPI) and the Scientific Committee on Antarctic Research (SCAR), which records 323,520 icebergs and serves as an important historical dataset. However, it is only confined to shipping lanes, not fully representing the Antarctic iceberg's spatial distribution and its interannual changes; (2) Low-resolution satellite imagery-based databases, with the National Ice Center (NIC) and Brigham Young University (BYU) Antarctic Iceberg Database as a notable example (Long et al., 2002; Stuart and Long, 2011a, b). Budge and Long (2018) consolidated these databases to offer iceberg location, length, and area data, but they are restricted to larger icebergs (length>5km) due to the limitations of low-resolution imagery; (3) Satellite radar altimetry-based databases, like the Altiberg database from the French Research Institute for Exploitation of the Sea (Tournadre et al., 2012, 2015, 2016, 2024). This database is effective at detecting icebergs in open waters, but in complex scene, such as areas with dense ice or high iceberg concentrations, it becomes challenging to extract accurate iceberg information from the altimetric waveforms; (4) High-resolution SAR data-derived products. Wesche and Dierking (2015) extracted icebergs larger than 0.3 km^2 in the Antarctic coastal region using Radarsat-1 circum-Antarctic mosaic images. Barbat applied a random forest algorithm to Radarsat circum-Antarctic mosaic images from 1997, 2000, and 2008 to obtain iceberg distributions for the corresponding years (Barbat et al., 2019a); (5) circum-Antarctic iceberg calving dataset. This dataset was derived from continuous optical (MODIS and Landsat-8) and radar (Envisat ASAR and Sentinel-1) satellite observations and was released by Qi et al. (2021). The product provides detailed information on each calving event, including time, area, size, thickness, etc., but it only focused on the transient icebergs just calved from ice shelves therefore lacking the spatial distribution across the open ocean. All above data products primarily cover the Antarctic coastal region, and the published datasets are not

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

Open Access Earth System Science Data Discussions

Table 1. Overview of Antarctic Iceberg Datasets.

Iceberg dataset	Time scale	Iceberg size range	Satellite data(sensors)
The SCAR International Iceberg Database	1982-2010	>10m	-
USNIC Antarctic Iceberg Data	1978-Present	>18 km	SAR, visible, and infrared remotely sensed imagery
BYU Antarctic Iceberg Tracking Database	1978 & 1992-Present	>5 km	SASS, ERS-1/2, NSCAT, QuikSCAT, ASCAT, OSCAT, SeaWinds, NIC (multiple sensors)
Altiberg	1992-2023	Determined by the resolution of the satellite altimeter	ERS1/2, Topex, Poseidon, Jason1/2/3, Envisat, Cryosat, Cryosat SAR, Cryosat SARIN, ALTIKA, HY-2A/B/C, Sentinel-3(A&B) PLRM, Sentinel-3(A&B) SAR, Geosat
Qi et al., 2021	2005-2020	>1 km	Envisat ASAR, Sentinel-1 SAR, MODIS, Landsat 8 OLI

real-time monitoring results, but rather used for historical scientific research. In summary, there is currently no comprehensive iceberg database covering multiple scales and the entire Southern Ocean has been established to date.

High-precision, large-scale, and long-term continuous remote sensing observations of circum-Antarctic iceberg distribution not only characterize the spatiotemporal patterns of iceberg occurrence but also provide critical data for elucidating the mechanisms of iceberg formation and evolution, ice-shelf dynamics, and their complex interactions with climate change. In this study, we leveraged the Google Earth Engine (GEE) platform to acquire Sentinel-1 SAR mosaic imagery and applied an incremental random forest classification combined with manual correction to identify Antarctic icebergs larger than 0.04 km^2, extracting each iceberg's outline, location, area, mass, and associated uncertainty. Based on these results, we constructed a circum-Antarctic iceberg distribution dataset covering each October from 2018 to 2023 and conducted a comprehensive analysis of the spatiotemporal characteristics of iceberg distribution over this six-year period. To ensure the reliability of the dataset, we performed an internal accuracy validation of the classifier and conducted external validation by comparing our results with existing iceberg databases and data products.

2 Data

To identify circum-Antarctic icebergs, we utilized the European Space Agency (ESA) Sentinel-1 C-band SAR Ground Range Detected (GRD) data. Given the extensive coverage of the data, we chose the Extra Wide (EW) swath mode, which provides a spatial resolution of 40 m. The Sentinel-1 data offers various band combinations based on different polarization modes (e.g., VV, HH, VV + VH, and HH + HV), with HH polarization being the primary mode available in polar regions (Koo et al., 2023; Ferdous et al., 2018). Therefore, only HH polarization band images were used for analysis.

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

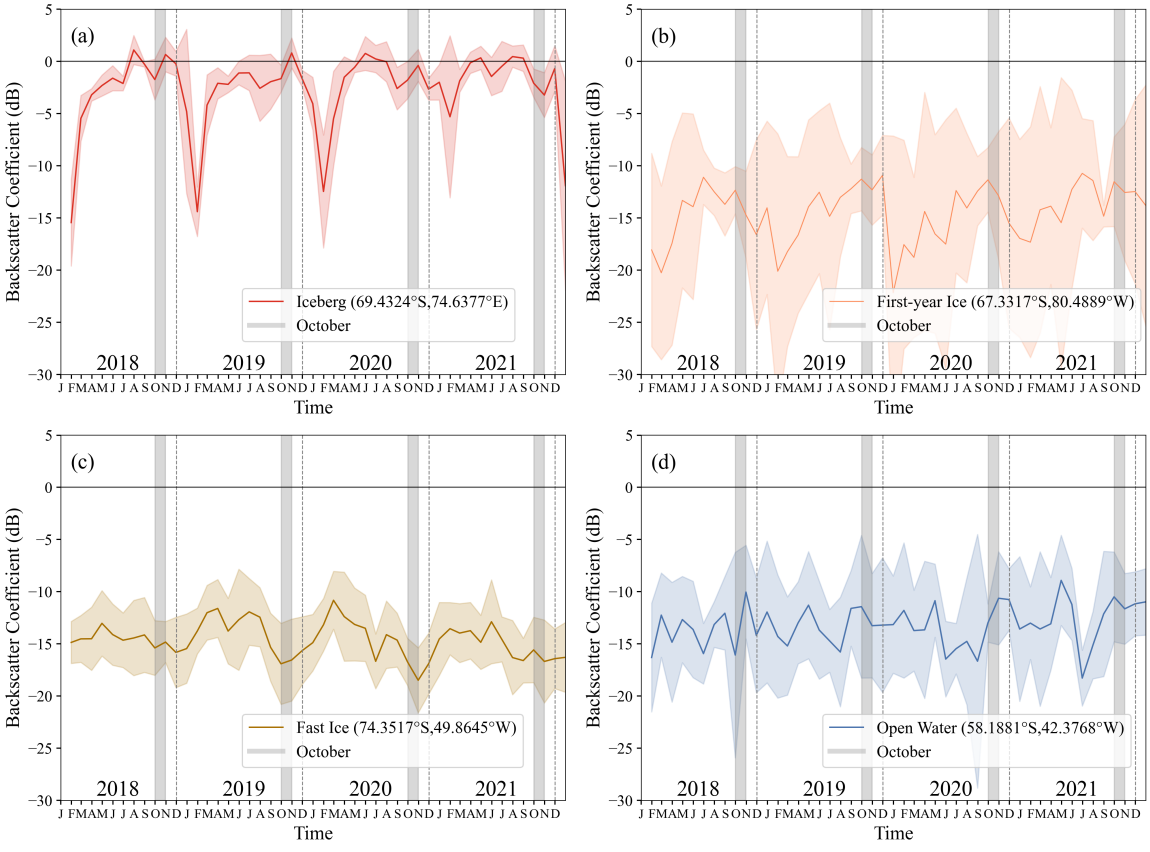

Figure 1. Time series of backscatter coefficients for typical Antarctic surfaces from 2018 to 2021: (a) iceberg, (b) first-year ice, (c) fast ice, and (d) open water. The lines show monthly average backscatter coefficients. Shaded regions represent uncertainty intervals based on data standard deviation. Gray-highlighted areas indicate the selected months (October of each year).

To optimize iceberg detection, we analyzed the backscatter characteristics of typical Antarctic oceanic features under HH polarization across different seasons (Fig. 1). As noted by Drinkwater et al. (1995) in their study of sea ice in the Weddell Sea, distinct differences in backscatter coefficients exist between various oceanic features. For instance, rough and undisturbed first-year ice, second-year ice, and other ice types exhibit unique reflective properties, which become more pronounced with seasonal and environmental changes. Environmental factors such as temperature and heat flux cause significant variation in backscatter coefficients. By comparing the interannual backscatter coefficient trends of typical Antarctic oceanic features, it was found that from June to October, the backscatter coefficient of icebergs is significantly higher than that of fast ice, first-year ice, and open water (Wesche and Dierking, 2012, 2015; Mazur et al., 2017), especially in October when the backscatter coefficient of fast ice reaches its annual minimum, providing optimal conditions for distinguishing icebergs from other oceanic features. Based on the above analysis, we selected Sentinel-1 SAR data in October for each year.

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

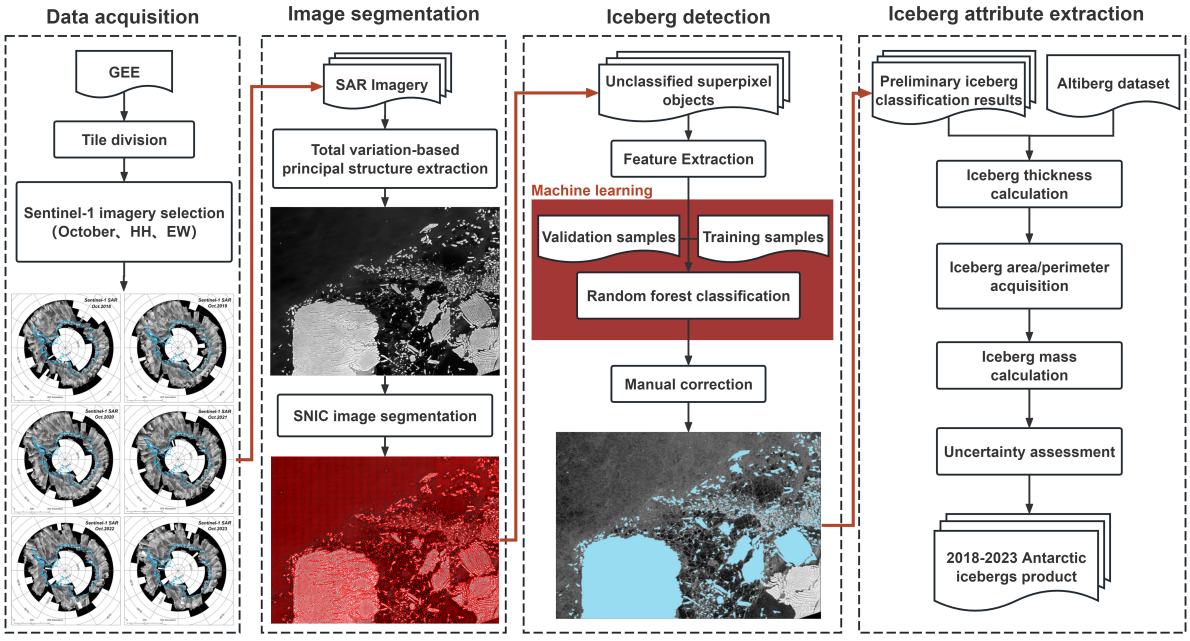

Figure 2. Flowchart of our methodology to obtain the 2018-2023 Antarctic iceberg product.

3 Method

The semi-automated workflow for extracting Antarctic icebergs using machine learning is shown in Fig. 2 and consists of four subprocesses: (1) Data acquisition, (2) Image segmentation, (3) Iceberg detection, and (4) Iceberg attribute extraction. In this section, we will provide the technical methods and details for each subprocess.

3.1 Data acquisition

GEE is a cloud-based platform developed by Google for the visualization and analysis of geospatial data. Through GEE, users can easily access a wide area of satellite remote sensing datasets (Gorelick et al., 2017; Amani et al., 2020). The Sentinel-1 SAR data provided by GEE have been pre-processed to remove thermal noise, apply radiometric calibration, and perform terrain correction, resulting in GRD backscatter coefficient images (expressed in dB). Given the vast extent of the Southern Ocean, this study divides the region south of 55°S into 5° × 5° tiles, resulting in a total of 430 tiles annually. For each tile, we retrieved Sentinel-1 SAR HH-polarization data from the EW swath mode acquired in October of each year between 2018 and 2023 (Fig. 3), and mosaicked the data chronologically to create monthly composite images. We delineated the effective observation area for each year and determined the intersection and union of these areas across the different years. The intersection of the effective observation ranges over six years has reached 16.67 million km^2, nearly covering the sea regions where icebergs might exist, thereby providing data support for obtaining the distribution of circum-Antarctic icebergs. In the subsequent analysis of

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

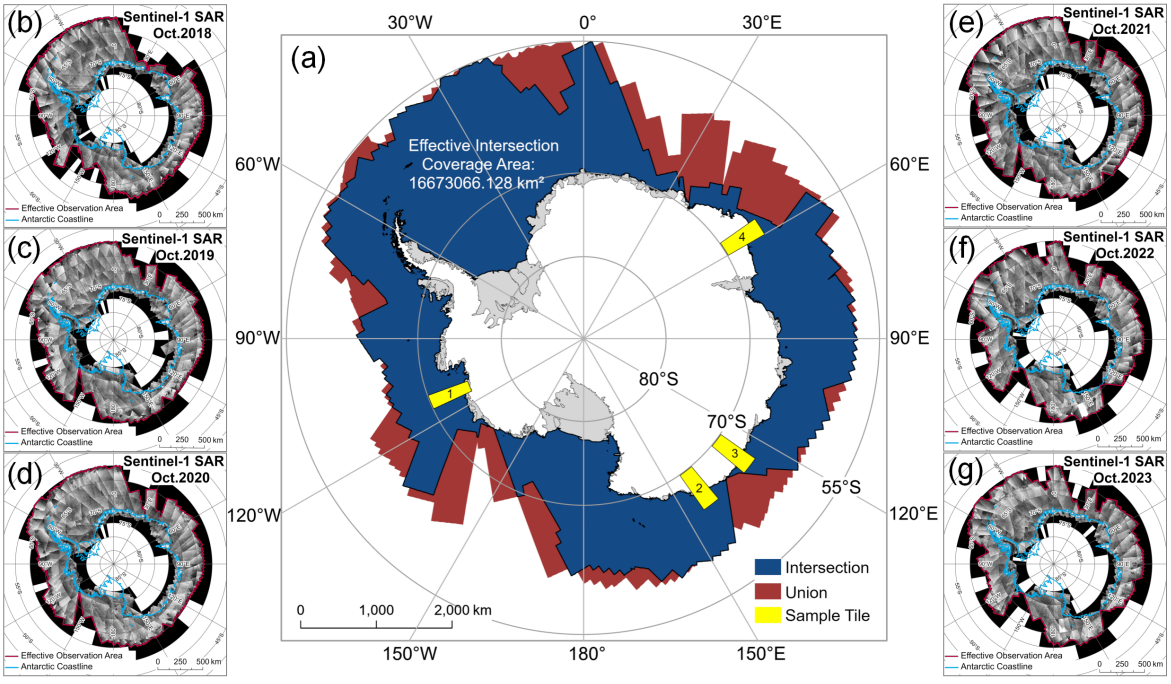

Figure 3. Circum-Antarctic Sentinel-1 SAR Data. The left and right columns display the Sentinel-1 mosaic images acquired from 2018 to 2023 on the GEE platform. The blue line delineates the coastline, while the red line indicates the valid observation boundaries. The central map illustrates the intersection and union of the observation areas over the six-year period, along with the four selected 5°×5° tile sample areas.

annual variation, we primarily focused on comparing icebergs within the intersecting observation areas across years, in order to identify trends in iceberg numbers and distribution. We emphasized this comparison in the consistent dimension, ensuring that the trends we observed were on an equal footing and thus more reliably indicative of actual changes in the iceberg population. Furthermore, to quantitatively assess the issues of misclassification, omission, iceberg merging, and contour deviations in the iceberg dataset, we selected four 5°×5° tile sample areas with low ocean current speeds and slow iceberg drift (as indicated by the yellow regions in Fig. 3). These sample areas effectively reflect the uncertainties in iceberg detection under complex ocean conditions and thus serve as representative of the overall detection performance of the entire dataset.

3.2 Image segmentation

3.2.1 Total Variation-based principal structure extraction (TV) algorithm for Sentinel-1 images smoothing

Due to the presence of background features such as sea ice and sea water, the edges and shapes of icebergs in SAR images can be unclear. To address this issue, we applied a Total Variation-based principal structure extraction (TV) algorithm (Xu et al., 2012), which separates the SAR images into two layers: a background texture layer and a primary structure layer that represents

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

the shape characteristics of the ocean surface. By extracting the primary structure layer, we were able to enhance the visibility of the iceberg edges and improve the accuracy of contour detection. The TV algorithm is particularly effective when the size of the background textures differs substantially from that of the primary structures, as it preserves the image edges and clarifies the boundaries. The results (Fig. 4) show that the TV algorithm successfully reduced background interference, retaining only the main contours of the icebergs, which made the iceberg bodies and boundaries much distinct. Even in complicated scene (Fig. 4c) or for small icebergs only a few hundred meters in size (Fig. 4b), the algorithm was able to effectively extract their contours.

3.2.2 Simple Linear Iterative Clustering (SLIC) image segmentation

We applied the Simple Linear Iterative Clustering (SLIC) algorithm for superpixel segmentation on the smoothed SAR images to avoid noise amplification and reduce computational complexity that may arise from using individual pixels during the subsequent Random Forest (RF) classification (Mazur et al., 2017; Karvonen et al., 2022; Koo et al., 2023). A superpixel is defined as a small, contiguous cluster of adjacent pixels that share similar backscatter characteristics, effectively representing a meaningful image region rather than individual pixels. By grouping pixels with similar backscatter characteristics into small, connected clusters, referred to as "superpixels", we not only improved classification efficiency but also significantly decreased the computational burden during the classification process (Achanta et al., 2012). The results of superpixel segmentation on the SAR images used in this study are shown in Fig. 4, with superpixel outlines displayed independently and not combined. Compared to the original image, the SLIC algorithm effectively delineates the boundaries of oceanic features and adapts well to different categories.

Given the large volume of image data and the spatial variability of iceberg distribution, we adopted a two-stage segmentation approach. In the first stage, we performed coarse segmentation using larger superpixels (40 $\times$ 40 pixels). For superpixels exhibiting histograms with multiple peaks, we then applied finer segmentation using smaller superpixels (5 $\times$ 5 pixels). This approach ensures that the smallest detectable iceberg has a length greater than 200 m or an area larger than 0.04 km^2.

3.3 Iceberg detection

3.3.1 Feature extraction

After image segmentation, we extracted features for each superpixel object based on the segmentation labels applied to the original, unprocessed image. These features were then used to construct a feature set for classification. In conjunction with manual interpretation, a sample set was created for the subsequent classification process. The extracted object features were categorized into three types: Statistical features, histogram-based features, and texture features, resulting in a total of 24 features. A detailed description and explanation of these features can be found in Appendix A.

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

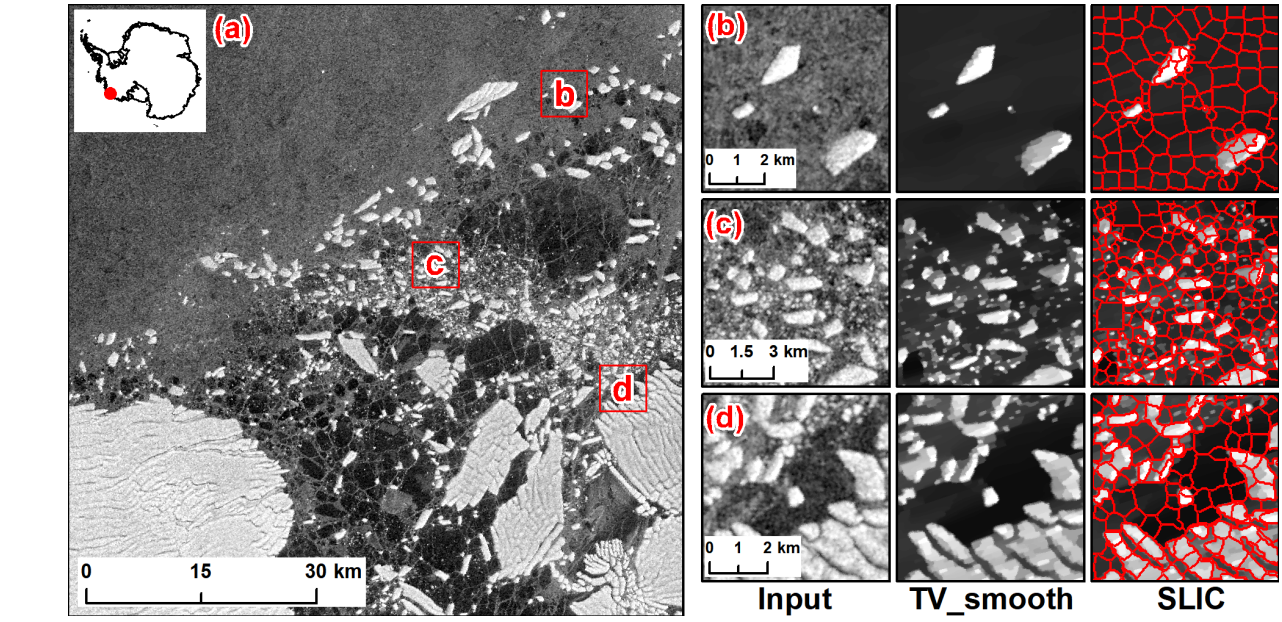

Figure 4. The results of the TV algorithm and SLIC segmentation on SAR imagery are shown in the following panels. Panel (a) provides an overview of the study area, with three representative sub-regions highlighted. Panels (b–d) show enlarged views of these sub-regions, presenting the original SAR image, the denoised output from the TV-smoothing algorithm, and the segmented image generated by the SLIC algorithm, respectively.

3.3.2 Incremental random forest classification

In this study, we employed an ensemble incremental Random Forest (RF) classifier (Zhou, 2012) to identify Antarctic circum-
140 polar icebergs. The process consisted of two main steps: (1) Using the training and validation sample sets, we evaluated the classification performance of various feature combinations, optimized the parameters of each RF classifier, determined their weights and classification thresholds, and constructed the ensemble classifier; (2) For each tile, we performed incremental RF training and classification on the superpixel objects, enabling automated iceberg detection.

Construction of Incremental random forest classifiers

145 Based on the Sentinel-1 SAR data, we manually selected approximately 2,000 sample points each year, with roughly half representing icebergs and the other half non-icebergs. The sample set was then divided into three subsets: an initial training set, a validation set, and a test set, in a 6:2:2 ratio. The training set was used to train the RF classifier, the validation set was used to evaluate the model's performance and optimize parameters, and the test set was used for final evaluation of the model's generalization ability and reliability.

150 We developed four RF classifiers: RF1, based on statistical features; RF2, based on histogram features; RF3, based on texture features; and RF4, based on a combination of all features. Out-of-bag (OOB) error analysis was performed to optimize both

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

the number of decision trees and feature selection for each classifier. Using the validation set, we calculated key performance metrics, including classification accuracy, precision, recall, and F1 scores. Based on these results, we assigned weights to each classifier and constructed an ensemble RF model. The model's performance was then further evaluated using precision-recall (P-R) curves and receiver operating characteristic (ROC) curves, which helped determine the optimal classification threshold for distinguishing icebergs from non-icebergs.

Automated Antarctic iceberg identification

After constructing the ensemble RF classifier, we predicted all the superpixels within each $5° \times 5°$ tile. Given the complexity of the data within each tile, image segmentation typically produces tens of thousands to hundreds of thousands of superpixels that require classification. Given the limited size of the initial training sample and the potential variation in iceberg and non-iceberg characteristics across different tiles, we adopted an incremental Random Forest approach for each tile. This method uses Mahalanobis distance to allow the classifier to adaptively learn and better match local data characteristics.

The process began by training RF1–RF4 using the initial training set, which were then combined into an ensemble classifier to generate the initial classification results for the tile. Then, we randomly selected an equal number of iceberg and non-iceberg samples from the newly identified objects to expand the training set. Based on feature importance ranking, we selected the most significant three features to construct the feature space for icebergs and non-icebergs. Subsequently, we calculated the mean (μ) and standard deviation (σ) of the distances between iceberg samples and the center of the iceberg, as well as the mean distance from non-iceberg samples to the iceberg center. If the mean distance from non-iceberg samples to the iceberg center exceeds $\mu + \sigma$, or the iteration count did not exceed five, we retrained the classifier with the incremental samples. The iteration limit of five was determined through multiple experiments. The incremental learning process terminates when either the conditions were not met or the iteration limit was reached. The predicted iceberg results from the final iteration were then taken as the final classification results for that tile.

3.3.3 Manual correction

The automatically classified superpixels labels identifying icebergs were used to generate iceberg outlines based on the geographic coordinates of the SAR images. These outline vectors were then manually refined in ArcMap 10.8 software interactively to ensure they accurately represent the true shapes of the icebergs as observed in the corresponding SAR image. Manual correction addressed three main issues: (1) the automatic detection process still resulted in misclassifications and missed icebergs; (2) some iceberg contours were incomplete at the tile boundaries; and (3) due to the mosaic nature of the tiles, some fast-moving icebergs were detected in multiple segments. The results for the four sample areas after incremental random forest classification and manual correction are shown in Fig. 5.

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

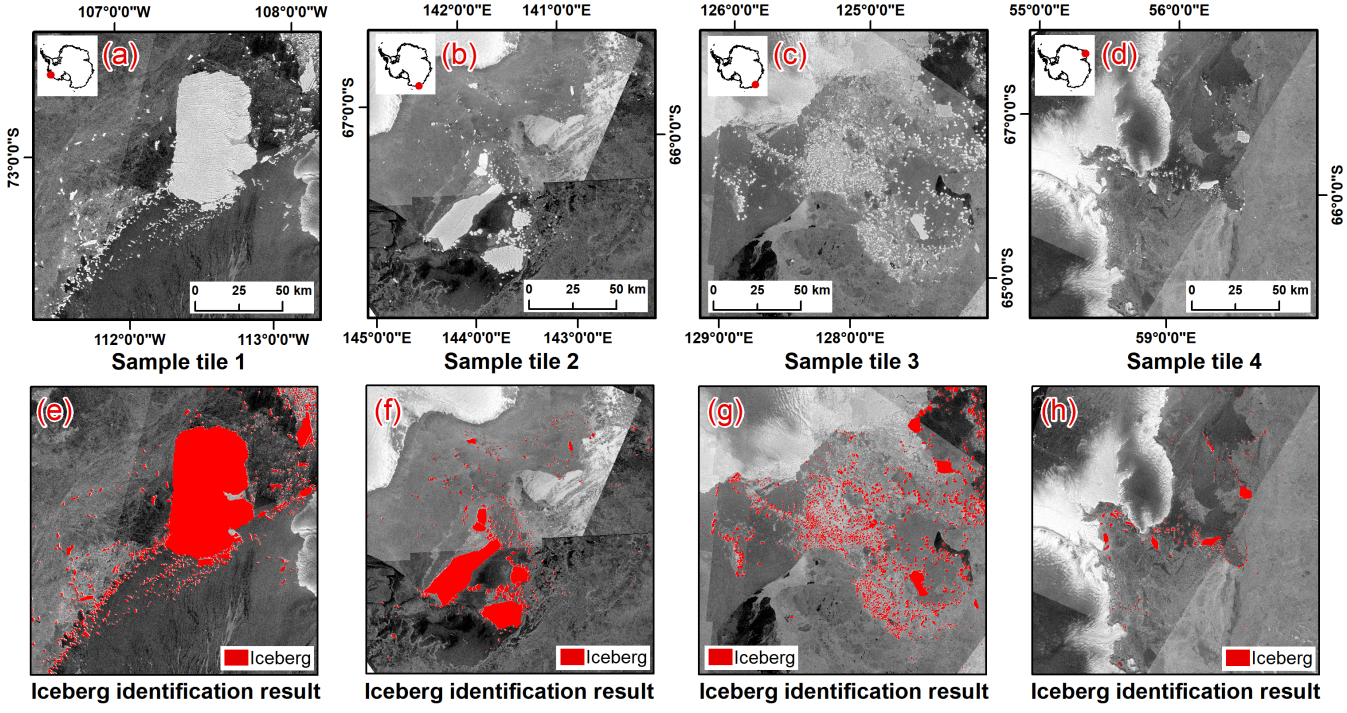

Figure 5. Iceberg identification results, panels (a)–(d) display the Sentinel-1 SAR images from the sample areas, while panels (e)–(h) present the classification results derived from these images using an incremental random forest classification supplemented with manual corrections. In these panels, the red vectors denote icebergs.

3.4 Iceberg attribute extraction

For each iceberg, key attributes such as area, perimeter, long axis, short axis, average thickness, mass, and the associated uncertainties for these parameters were calculated. This section outlines the methods used to derive these iceberg attributes and assess the uncertainties involved.

After obtaining the iceberg outline vector data, we calculated the area (km^2) and perimeter (km) of each iceberg under the Antarctic Polar Stereographic projection (EPSG: 3031). Based on the iceberg area data, we estimated the total mass of icebergs in the circumpolar region for each year. Using 19,945 CryoSat-2 SARIn data points from the Altiberg database recorded after 2018 (Tournadre et al., 2024), the average iceberg freeboard was found to be 40 m. Assuming average densities of 850 kg/m^3 for icebergs and 1,025 kg/m^3 for seawater (Silva et al., 2006), the average iceberg thickness ($\overline{H}$) was estimated to be approximately 232 m, based on Archimedes' principle of buoyancy. Using the total area of the icebergs, average thickness, and average density, the total mass (M) of icebergs with an area greater than 0.04 km^2 was calculated using Equation (1):

$$M = A_{Iceberg} \times \overline{H} \times \rho_{Iceberg} \quad (1)$$

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025


Open Access Earth System Science Data Discussions

Due to the diverse shapes of icebergs, we used the principal orientation method to determine their geometric characteristics. First, we calculated the centroid of the iceberg's geometry, which serves as its geometric center. Then, we applied Principal Component Analysis (PCA) to the iceberg's boundary points to determine the directions of its principal axes. The first principal component corresponds to the long axis of the iceberg, while the second principal component corresponds to the short axis. Next, we projected the boundary points along the long axis and computed the projection length in this direction to obtain the length of the iceberg's long axis, and we used the same method to obtain the length of the short axis.

3.5 Uncertainty assessment

According to Equation (1), calculating the total mass of icebergs involves several sources of uncertainty, including errors in measuring iceberg area, uncertainties in thickness estimation, and deviations in the assumed density of the iceberg. In this section, we assess the primary uncertainties encountered in extracting iceberg attributes.

3.5.1 Iceberg area uncertainty

The uncertainty in iceberg area measurement primarily arises from two independent factors: (1) the spatial resolution of SAR imagery; and (2) the detection errors introduced during iceberg identification (e.g., misclassification, omission, or merging of iceberg targets). The uncertainty due to image resolution (U_1) can be approximated as the product of the total iceberg perimeter and the pixel size of the imagery, that is, we estimate the area uncertainty from the pixel error along the iceberg boundaries using Equation (2):

$$U_1 = P \times \Delta x \tag{2}$$

Where P is the total perimeter of all icebergs each year (km), and Δx is the spatial resolution of the imagery, which is 0.04 km.

The second source of uncertainty (U_2) primarily arises from errors in iceberg classification and extraction, such as omissions, false detections, erroneous merging (i.e., mistakenly detecting adjacent icebergs as a single object), and contour deviations. To quantitatively evaluate this component, we acquired mosaic images in the Interferometric Wide (IW) swath mode (with a spatial resolution of 20 m) from four 5°×5° sample tile areas, while ensuring that, in iceberg-dense areas, the time interval between the IW mode images and the EW mode images (with a spatial resolution of 40 m) did not exceed 10 days. In each sample tile area, we manually digitized iceberg outlines from high-resolution IW images to construct a reference dataset representing the "true" iceberg count and area, and then compared it with the dataset obtained from EW mode imagery using an incremental random forest algorithm supplemented with manual corrections. As shown in Table 2, the comparison results indicate that in the most complex sample area, the relative error in total iceberg area reached up to 3.15%. For a conservative estimation of uncertainty, we adopt 4% as the parameter—i.e., the uncertainty due to detection errors is calculated by multiplying the annual total iceberg area by 4%.

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

Open Access Earth System Science Data Discussions

The uncertainty in the total annual iceberg area (U_A) can be calculated using the error propagation law, as shown in Equation (3):

$$U_{\mathrm{A}} = \sqrt{U_1^2 + U_2^2}. \tag{3}$$

It should be noted that for an individual iceberg, its area uncertainty is solely determined by the image resolution (U_1), since an iceberg is either correctly extracted or not detected at all; whereas for the total annual iceberg area, both (U_1) and (U_2) must be considered, and the overall error is calculated using Equation (3).

Table 2. Validation of iceberg detection in four sample regions. Iceberg counts from EW and IW imagery, detection errors (inaccurate outlines, merged and missed icebergs), average missed iceberg area, total iceberg areas, and relative area uncertainty (%) are presented.

Region	EW Count	IW Count	Inaccurate Outlines	Merged Icebergs	Missed Icebergs	Avg. Missed Area (km^2)	EW Total Area (km^2)	IW Total Area (km^2)	Area Uncertainty
1	683	728	6	13	24	0.138	637.19	637.22	≈0
2	695	816	12	5	103	0.142	1340.06	1353.48	1.00%
3	3151	3575	25	13	401	0.164	1895.16	1954.79	3.15%
4	583	681	9	8	86	0.126	296.73	305.09	2.82%

3.5.2 Iceberg thickness uncertainty

The uncertainty in thickness estimation primarily arises from errors in measuring the iceberg's freeboard height and deviations in the assumptions of physical parameters. Using the CryoSat-2 SARIn data from the aforementioned Altiberg database, we calculated the standard deviation of the iceberg's freeboard height to be 13 m, which results in an uncertainty of 76 m in the estimated iceberg thickness.

In previous studies, a thickness of 250 m was commonly adopted for mass estimations of icebergs (Wesche and Dierking, 2015; Rackow et al., 2017; Barbat et al., 2019a). Gladstone based on a comprehensive analysis of observational data from Antarctic icebergs, established a size classification system for icebergs ranging from 60 to 2200 m in length (Gladstone et al., 2001), noting that the thickness increases with size, culminating in a maximum thickness of 250 m for icebergs exceeding 500 m in length. In this study, we utilized the mean freeboard height of icebergs measured by CryoSat-2 SARIn from the Altiberg database to determine the thickness of the icebergs. Given that the minimum identifiable area of icebergs in our study is 0.04 km^2, a significant number of icebergs fall within the 200 to 500-meter length range, corresponding to classes 3 to 4 as defined by Gladstone et al. (2001), with thicknesses ranging from 133 to 175 m. Based on this analysis, we selected an average thickness of 232m for the icebergs, which is slightly below the upper limit of 250 m proposed by Gladstone et al. (2001), and this choice is considered reasonable.

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

3.5.3 Iceberg mass uncertainty

Since the total mass of the iceberg derived from three independent and unrelated components, we employed a synthetic standard uncertainty to assess its uncertainty (Qi et al., 2021), which is calculated using Equation (4):

$$U_M = M \times \sqrt{\frac{U_A^2}{A_{Iecberg}^2} + \frac{U_H^2}{\overline{H}^2} + \frac{U_\rho^2}{\rho_{Iceberg}^2}} \tag{4}$$

Where M is the total mass of icebergs each year, $A_{Iceberg}$, $\overline{H}$ and $\rho_{Iceberg}$ represent the total area, average thickness, and density of icebergs each year, respectively. U_A, U_H and U_ρ represent the uncertainties in the total area, thickness, and the density of icebergs, respectively. $\rho_{Iceberg}$ and U_ρ are set to 850 kg/m^{3}and 5 kg/m^3 (Griggs and Bamber, 2011).

4 Validation and uncertainty

4.1 Accuracy assessment of Antarctic iceberg identification algorithm

Using a test set of approximately 400 manually selected samples per year, we evaluated the performance of the automated classification results. The results show that the automatic classification algorithm achieved high accuracy in identifying circum-Antarctic icebergs from 2018 to 2023 (Table 3), with all metrics exceeding 0.9, indicating excellent classification performance.

To assess the model's performance and adapt to the time-series nature of the data while minimizing the risks of overfitting and data leakage, we employed a rolling-window validation method for time-series cross-validation (Table 4). Specifically, in the first iteration, the model was trained on 2018 data and tested on 2019 data. In the second iteration, the model was trained on data from 2018 and 2019, and tested on 2020 data, and so forth. These results show that, as the training data accumulated each year, the model maintained high classification performance across the test data from different years. Notably, the recall rate consistently remained above 0.95, demonstrating strong stability and robustness in iceberg detection. Additionally, the fluctuations in accuracy (ACC) and F1 scores were minimal, further confirming the reliability of the model.

After classifier performance evaluation, our data product incorporates a manual correction step in addition to the machine learning-based automated iceberg detection (see Sect.3.3.3). By visually inspection and manually correcting the automated classification results, we further reduced instances of false positives and false negatives. As a result, the final iceberg data product demonstrates even higher precision across various accuracy metrics.

4.2 Attribute uncertainties of Icebergs

Based on a comparison of the results from four sample areas (Table 2), we found that iceberg omissions are relatively severe, resulting in an underestimation of the total iceberg amount by approximately 3%-15%. However, the missed icebergs are mainly small or weak-signal targets, with an average area of only 0.126-0.164 km^2, thus having a limited impact on the total iceberg area. In low-resolution imagery, the radar signal of small icebergs is often weak or their boundaries become blurred due to noise and complex sea conditions, making it challenging to accurately identify all icebergs even after manual correction.

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

Open Access Earth System Science Data Discussions

Table 3. Performance evaluation of the incremental random forest classifier.

Year	Iceberg samples	Non-iceberg samples	ACC	Precision	Recall	F1
2018	188	209	0.950	0.922	0.973	0.948
2019	188	208	0.965	0.939	0.989	0.964
2020	182	215	0.900	0.905	0.995	0.948
2021	187	213	0.943	0.906	0.979	0.941
2022	190	209	0.937	0.910	0.963	0.936
2023	185	211	0.957	0.924	0.989	0.956

Table 4. Results of the time series Cross-Validation method with rolling window validation.

Iteration	Iceberg samples	Non-iceberg samples	ACC	Precision	Recall	F1
1	915	1056	0.950	0.915	0.985	0.948
2	917	1063	0.948	0.909	0.987	0.946
3	983	1008	0.950	0.941	0.959	0.950
4	967	1020	0.943	0.926	0.959	0.942
5	933	1036	0.963	0.938	0.988	0.962

Furthermore, in the SLIC algorithm, the low contrast between icebergs and sea ice or open water in low-resolution images leads to blurred iceberg edges, making the boundaries between adjacent icebergs indistinct and causing nearby icebergs to be erroneously merged into a single object or to exhibit contour deviations. Given that false detections are negligible after manual correction, the maximum area uncertainty due to iceberg detection errors in the tile sample areas is 3.15%. Therefore, we adopt 4% as a conservative and reasonable estimate.

We assessed the uncertainty in iceberg area and mass attributes using Equation (2) and (3). The maximum uncertainty in the area of a single iceberg was 22.4 km^2. From 2018 to 2023, the total area uncertainty for each year was as follows: 4,549 km^2, 5,007 km^2, 5,177 km^2, 5,102 km^2, 4,371 km^2, and 4,591 km^2 respectively. The uncertainty in iceberg area primarily stems from the uncertainty in the iceberg perimeter, indicating that, for icebergs of equal area, rectangular icebergs have greater area uncertainty compared to elliptical ones. The uncertainty in iceberg mass is mainly influenced by the uncertainty in iceberg thickness. The average uncertainty in iceberg mass over the six years was 3,029 Gt, with annual error fluctuations ranging from 34.07% to 34.92%. This result aligns with the 37% error margin suggested by Jacobs (Jacobs et al., 1992).

4.3 Consistency of a multisource iceberg database

4.3.1 Compare with BYU/NIC iceberg database

The BYU/NIC iceberg database provides detection dates and geolocation information for icebergs with a major axis exceeding 5 km. To match the time range of this study's iceberg dataset, we filtered the BYU/NIC iceberg database to include only the

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

October data from 2018 to 2023. To ensure comparability, our database retains only records of icebergs with a major axis greater than 5 km. During the matching process, if an iceberg's record within the same month exhibits consistent interannual trajectories and its geographic location falls within a predetermined spatial threshold, it is considered a successful match.

In our dataset, the number of icebergs with a major axis greater than 5 km (ranging from 288 to 475 per year during 2018-2023) is significantly higher than the records in the BYU/NIC iceberg database (46 to 54 per year). Our dataset achieves a recall rate of approximately 96% to 98%, indicating that most of the icebergs recorded in the BYU/NIC iceberg database have been successfully detected. The geographic locations of the matched icebergs show high consistency between the two databases, with 92% of the BYU/NIC iceberg coordinates falling within the iceberg polygon vectors of our study, and the remaining positional errors being within a few kilometers.

Three icebergs recorded in the BYU/NIC iceberg database were not detected in our dataset, primarily due to incomplete satellite image coverage or complex sea ice conditions leading to missed detections. A few matched icebergs exhibit positional errors of up to several tens of kilometers, likely due to substantial differences in observation times and high iceberg drift speeds, reflecting differences in data sources and detection methods. Moreover, our dataset detects a large number of icebergs not recorded in the BYU/NIC database, owing to the use of higher resolution imagery and a more sensitive detection algorithm that identifies smaller or transient icebergs.

4.3.2 Compare with Altiberg database

The Altiberg database provides a merged grid product of iceberg detection from multiple satellite missions, incorporating quality control and calibration procedures to yield spatiotemporal information on iceberg volume, area, and other variables. To evaluate both the overall consistency and local differences between our dataset and Altiberg's, we generated our iceberg volume data using the same 100 km × 100 km grid. Specifically, for each grid cell, we multiplied the total iceberg area in our dataset by a fixed thickness of 232 m (see Sect.3.5.2), thereby obtaining the gridded average iceberg volume for 2018-2023. We then performed a visualization and difference analysis to compare this dataset with Altiberg's across both regional and global domains.

In October, the extent of Antarctic sea ice remains substantial. Consequently, Altiberg's data show missing or low-value cells in high-latitude and coastal regions with dense sea ice, primarily due to its reliance on altimeter signals, which are easily weakened or disrupted by ice cover (Tournadre et al., 2015). This limitation makes it difficult for altimeters to distinguish or detect icebergs in regions of high sea-ice concentration. In contrast, our approach utilizes high-resolution SAR imagery that can capture iceberg outlines even beneath sea ice, leading to higher iceberg volume estimates in these regions. The difference map indicates a marked positive bias (our dataset > Altiberg) in sea ice-dominated areas. Meanwhile, the histogram reveals that, in open-water or lower sea ice concentration zones, most grid-cell volume differences fall below 0.692 km^3, indicating good overall consistency. Altiberg's detection model was initially designed for medium- to small-scale icebergs (0.01-9 km^2), whereas our method imposes no upper limit on iceberg size. Consequently, if a grid cell contains extremely large or multiple large icebergs, the total iceberg volume can become substantially higher than Altiberg's, resulting in significant differences. This phenomenon is reflected in the histogram, where a small number of grid cells exhibit differences exceeding 100 km^3,

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

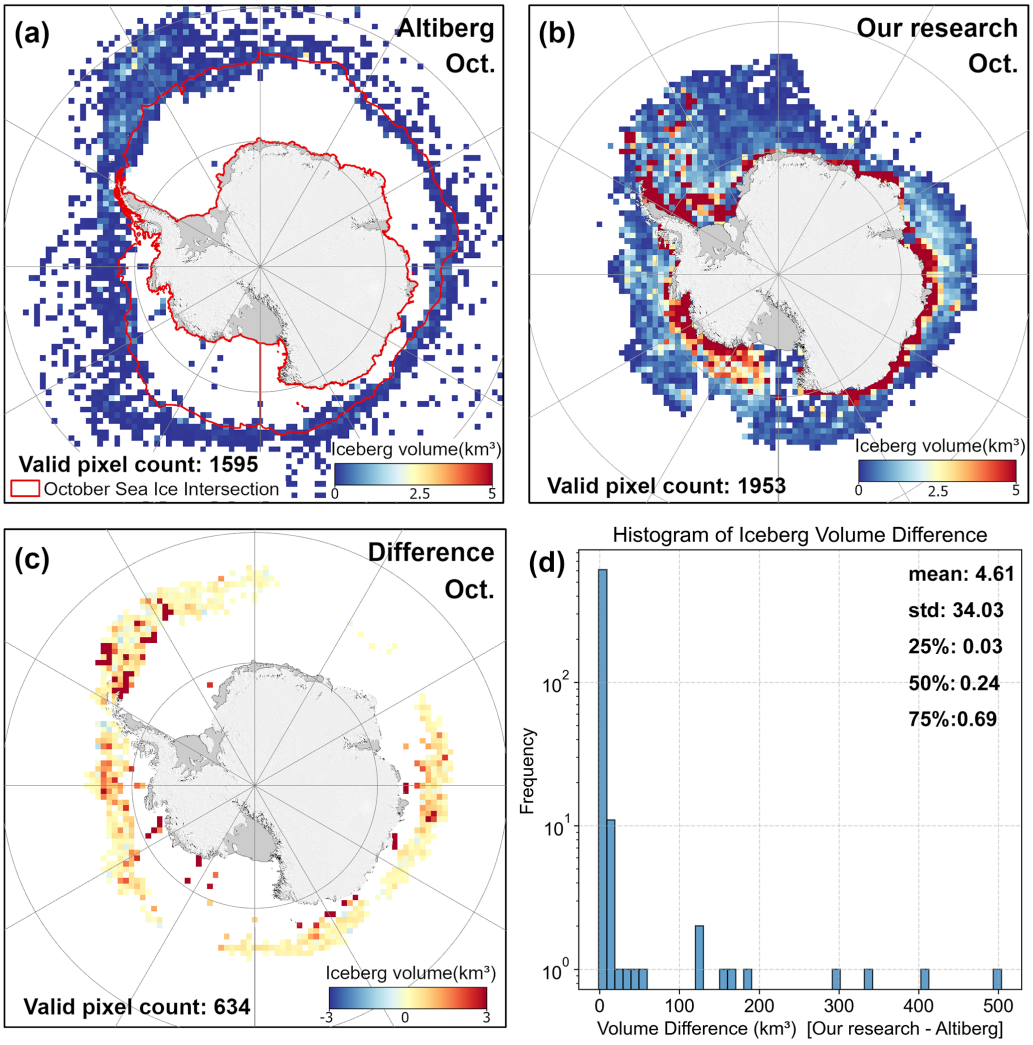

Figure 6. Panel (a) shows the six-year average iceberg volume from the Altiberg database for each October from 2018 to 2023. Panel (b) displays the six-year average iceberg volume from our dataset over the same time period and grid. Panel (c) presents the volume differences (our dataset minus the Altiberg database), and panel (d) summarizes the statistical distribution of these differences.

raising the overall standard deviation to 34 km^3. These findings suggest that while Altiberg provides a continuous, long-term record suitable for open-water regions, our dataset more comprehensively identifies and quantifies icebergs within sea ice-covered areas.

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

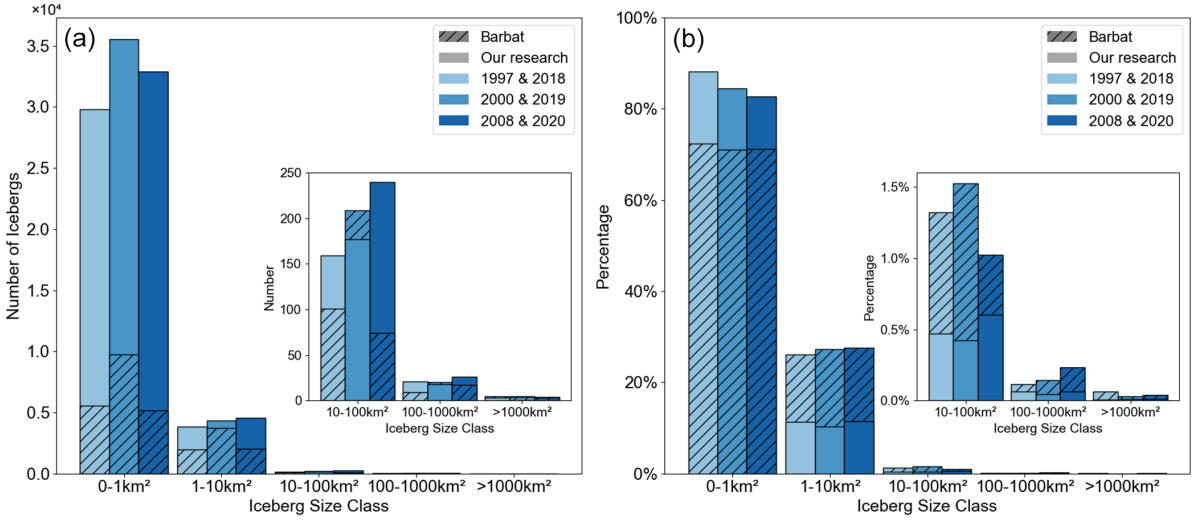

Figure 7. Comparison with the results of Barbat et al. (2019a): (a) Number of Antarctic icebergs and (b) Proportion of different categories.

4.3.3 Compare with other research

Compared with the Antarctic coastal icebergs larger than 0.1 km^2 identified by Barbet using RAMP data (Barbat et al., 2019b), our dataset covers a broader area and employs a lower minimum detection threshold, thereby capturing a larger number of icebergs with smaller scales and resulting in certain differences in the overall findings. Relying solely on coastal data tends to underestimate the actual number of small icebergs, because these smaller icebergs are often rapidly transported by coastal currents to the open ocean shortly after formation. Coastal regions mainly record the icebergs released during the initial stages of ice shelf and glacier calving, and due to their small size, small icebergs are more strongly influenced by ocean currents; as a result, their proportion in coastal observations is significantly lower. Despite the significant differences in total iceberg numbers between the two studies, as shown in Fig. 7(b), the relative proportions of icebergs by size category are generally consistent and exhibit minimal interannual variation, indicating that the size structure of Antarctic icebergs has maintained a certain degree of temporal stability.

5 Result and discussion

5.1 Number, area, and mass of circum-Antarctic icebergs

The statistics of circum-Antarctic icebergs from 2018 to 2023 are presented in Table 5, showing significant interannual variations in both iceberg number and area. In 2018, a total of 33,823 icebergs were observed in the circumpolar region, covering an area of 37,606 ± 4,549 km^2. In 2019, the number of icebergs increased to 40,034, and the area rose to 42,485 ± 5,007 km^2. Although the number of icebergs slightly decreased to 38,086 in 2020, the total area continued to increase, reaching 45,958 ±

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

5,177 km^2. In 2021, both the number of icebergs and their area peaked over the six-year period, with 51,332 icebergs and an area of 50,810 ± 5,103 km^2. In 2022, the number of icebergs dropped to 37,626, and the area decreased to 46,840 ± 4,372 km^2. However, in 2023, the number of icebergs went up again to 44,538, with an area of 43,409 ± 4,591 km^2. The interannual variations in the number and area of icebergs reflect the dynamic nature of the Antarctic ice sheet and its response to climate change. Furthermore, We calculated the intersection of the effective observation areas for each year (Fig. 3) and, based on this intersected area, computed the proportion of icebergs falling within it relative to the total annual iceberg number, as reported in the "percentage" column of Table 5.

In contrast to the variations in iceberg numbers, the total mass of Antarctic icebergs showed an increasing trend from 2018 to 2021, rising from 7,416 ± 2,590 Gt in 2018 to 10,020 ± 3,434 Gt in 2021, before decreasing to 9,237 ± 3,147 Gt in 2022, and 8,560 ± 2,947 Gt in 2023.

Table 5. Total number, area, mass of icebergs and percentage of icebergs in the intersection area in the circum-Antarctic region from 2018 to 2023.

Year	Total number	Total area (km^2)	Total mass (Gt)	Percentage
2018	33,823	37,606 ± 4,549	7,416 ± 2,590	96.08%
2019	40,034	42,485 ± 5,007	8,378 ± 2,917	94.81%
2020	38,086	45,958 ± 5,177	9,063 ± 3,140	93.76%
2021	51,332	50,810 ± 5,103	10,020 ± 3,434	91.19%
2022	37,626	46,840 ± 4,372	9,237 ± 3,147	97.61%
2023	44,538	43,409 ± 4,591	8,560 ± 2,947	97.47%

5.2 Spatial distribution of icebergs

Fig. 8 shows the distribution of icebergs in October for each year from 2018 to 2023. Overall, the density of icebergs is high in the West Antarctic region (e.g., near the Thwaites and Dotson ice shelves) and in the East Antarctic region (e.g., around the Holmes and Mertz ice shelves), indicating that calving activity in these areas is both frequent and intense. In contrast, in large ice shelf regions such as the Ross Sea and Weddell Sea, although calving events occur less frequently from year to year, when a large-scale fracture does occur, it typically leads to the rapid formation of a high-density iceberg zone in a short period. Fig. 9 further illustrates the distribution of icebergs by size, showing that medium-to-large icebergs tend to be concentrated in near-coastal waters and are spatially more scattered, whereas small icebergs are widely distributed throughout the Southern Ocean.

Following Wesche and Dierking (2015)'s rule, all detected icebergs are classified into five size categories, as shown in Fig. 10: A1 (<1 km^2), A2 (1-10 km^2), A3 (10-100 km^2), A4 (100-1,000 km^2), and A5 (≥1,000 km^2). From 2018 to 2023, the number of the smallest icebergs (A1) shows significant fluctuations, alternating between increases and decreases and consistently accounting for over 85% of the total iceberg count, thus driving the overall variability in iceberg numbers. In contrast, the number of medium-sized icebergs (A2 and A3) generally increases, reaching a peak in 2020 before slightly

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

declining; their fluctuations are much smaller compared to those of the A1 category, comprising roughly 10% of the total. Large icebergs (A4 and A5) are relatively rare, and their occurrence is closely associated with major ice shelf calving events—years such as 2017/18 (A68a), 2019 (D28), 2020(A69) and 2021 (A74 and A76a) see a surge in this size (Braakmann-Folgmann et al., 2022; Deakin et al., 2024). Moreover, small icebergs not only result from continuous small-scale calving but can also originate from the further breakup of large icebergs during their drift. Based on this, although the annual iceberg count is predominantly driven by small icebergs, following a large ice shelf fracture the rapid increase in large icebergs is typically accompanied by their subsequent fragmentation, which in turn leads to an additional rise in the number of small icebergs.

To assess the spatial distribution of icebergs, the circumpolar ocean region was divided into five sectors based on longitude: Ross Sea Sector (160°E to 130°W), Amundsen and Bellingshausen Seas Sector (130-60°W), Weddell Sea Sector (60°W to 20°E), Indian Ocean Sector (20-90°E), and Western Pacific Ocean Sector (90-160°E) (Parkinson and Cavalieri, 2012). Fig. 11(a) and (b) present the number of icebergs and their relative percentages in each sector. The results show that over these six years, the Western Pacific Ocean Sector contributed the highest number of icebergs, while the Weddell Sea Sector recorded the fewest from 2018 to 2021, but in 2022 its iceberg count surpassed that of the Ross Sea. In the Ross Sea Sector, the iceberg proportion remained stable at around 16% in 2018 and 2019, increased to 21.7% in 2020, and then rapidly declined to 14% and 9.8% in 2021 and 2022, respectively. The proportions in the Indian Ocean and Amundsen and Bellingshausen Seas sectors remained relatively stable at approximately 20% over the six-year period.

5.3 Distinctive spatial characteristics and formation mchanisms of Small-Scale icebergs in the Southern Ocean

This study's dataset is unique in both the scales and quantity of icebergs, particularly as it is the first to include small icebergs in the 0.04-0.1 km^2 size area derived from remote sensing imagery. Over the six-year period, the average number of icebergs in this size range was 8,272, accounting for 15.25% to 29.02% of the total number each year, with an average area of 559.5 km^2, contributing 0.97% to 1.93% of the total area.

To examine the spatial distribution and formation mechanisms of these small icebergs, we divided the Southern Ocean into 50 km $\times$ 50 km grids and calculated the average number of small icebergs in each grid from 2018 to 2023, as well as the average distance between these small icebergs and large icebergs (>100 km^2) (Fig. 12). The results show that small icebergs are mainly concentrated at ice shelf fronts, though their distribution is sparse at the fronts of the Ross Ice Shelf, Filchner-Ronne Ice Shelf, and Riiser-Larsen Ice Shelf. Due to their size, these icebergs have short lifespans and are more sensitive to changes in surrounding sea ice and ocean conditions.

In analyzing the distances between small and large icebergs, we further validated the small iceberg formation mechanism proposed by Tournadre (Tournadre et al., 2016). The results indicate that small icebergs in the Southern Ocean follow two main patterns: one where small icebergs are found near large icebergs, suggesting they may originate from fragmentation, share a common source, or drift along similar paths; and another where small icebergs exhibit "free drift," unrelated to any large icebergs, drifting far from their calving sources, such as in the Ross Sea, Bellingshausen Sea, and eastern Weddell Sea. In these regions, the drift of small icebergs plays a key role in transporting ice shelf and large iceberg material, significantly

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

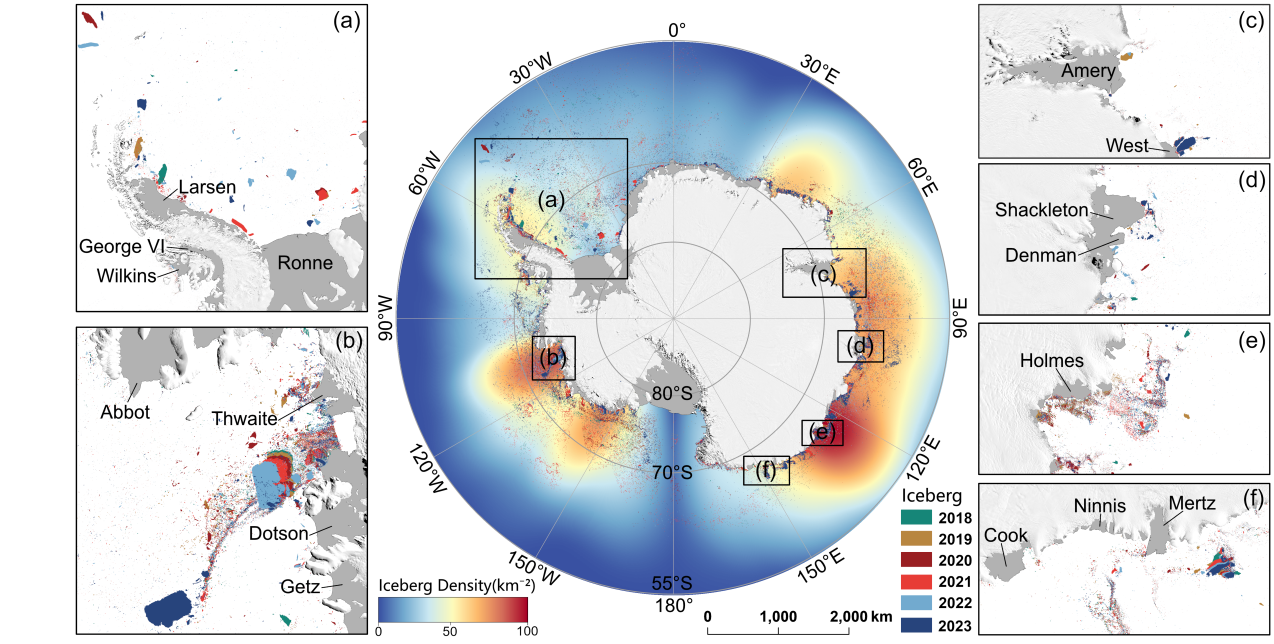

Figure 8. Distribution of Icebergs in the Circum-Antarctic Region from October 2018 to October 2023. The central map represents the distribution of icebergs over the six years, with different colors indicating different years. The base map shows the iceberg density. Panels (a)-(f) display the distribution of icebergs at the front of ice shelves that are prone to calving.

influencing regional ice flow and freshwater flux. The drift paths can extend thousands of kilometers, forming independent "drifting alley".

6 Conclusions

This study successfully identified circum-Antarctic icebergs from 2018 to 2023 using Sentinel-1 SAR mosaic data obtained from the Google Earth Engine (GEE) platform, combined with an incremental random forest algorithm and manual corrections. The smallest identifiable iceberg had an area of 0.04 km^2. This is the first high-precision dataset covering the entire Southern Ocean, including small icebergs. Small icebergs dominate in terms of quantity, and their distribution is critical for initializing coupled ocean-iceberg models, aiding in more accurate simulations of iceberg melting effects on ocean circulation and global climate.

Although this study primarily used data from October each year, when the difference in backscatter characteristics between icebergs and other oceanic features is most pronounced, and the identification results are optimal, the method is not limited to this period. In the future, images from other months can be obtained via the GEE platform, enabling the study of seasonal variations and year-round iceberg dynamics. This approach compensates for the limitations of snapshot data, providing a more comprehensive understanding of iceberg formation, drift, and melting processes.

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

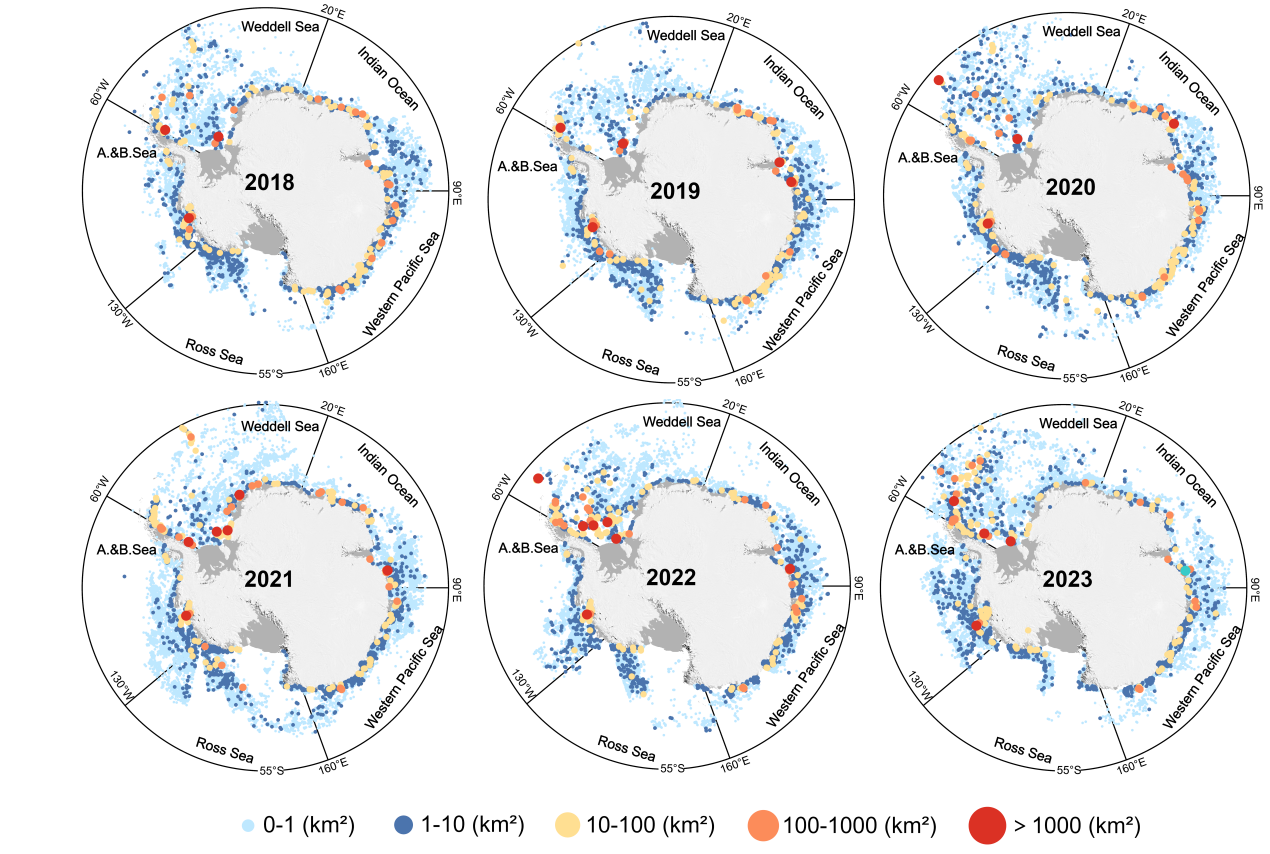

Figure 9. Iceberg counts for different size classes in various sea sectors from 2018 to 2023. Each point represents an individual iceberg, point sizes represent five size categories(A1-A5)

Despite the extensive coverage of Sentinel-1 SAR data, data gaps existed in certain years and regions, such as in parts of the Indian Ocean in 2018, which may have led to an underestimation of iceberg numbers in these areas. Additionally, in estimating iceberg mass, fixed average thickness and density values (232 m and 850 kg/m^3) were assumed, but the actual thickness and density may vary depending on iceberg size, shape, and melting status, introducing uncertainty into mass estimates. Furthermore, Although we employed a high-precision iceberg identification model supplemented by manual corrections within a semi-automated workflow, in complex marine and terrestrial environments (e.g., regions with dense sea ice and iceberg calving zones), the radar signals of icebergs are often weak and their boundaries blurred due to noise and adverse sea conditions, potentially resulting in varying degrees of omissions, erroneous merging, and contour deviations. Future research could consider integrating multi-source remote sensing data and incorporating more advanced deep learning algorithms to further improve iceberg identification accuracy.

Overall, this study provides the first high-precision iceberg distribution dataset for the Southern Ocean, including small icebergs. It lays the foundation for a deeper understanding of the impact of icebergs on the marine environment and global

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

Open Access Earth System Science Data Discussions

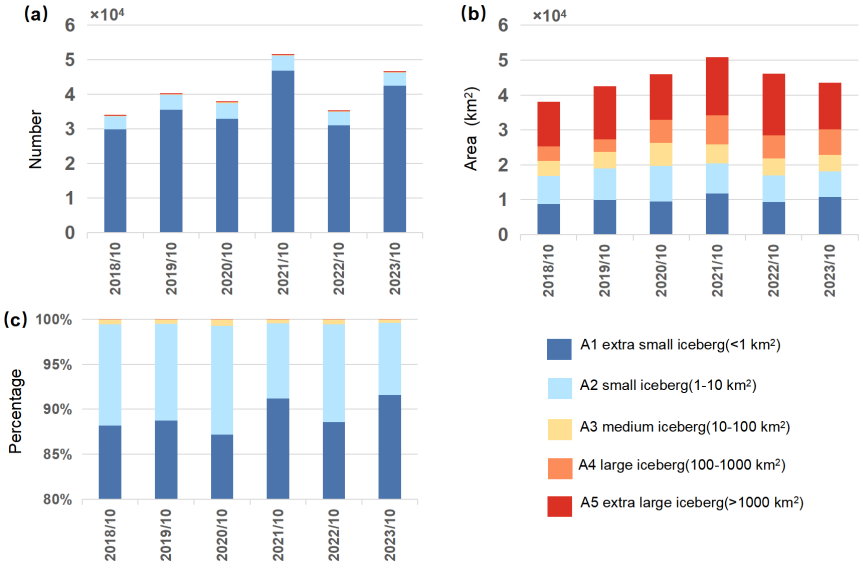

Figure 10. Annual distribution characteristics of Antarctic icebergs of five categories from October 2018 to October 2023. Panels (a)-(c) present the number, area, and number percentage of icebergs of five categories, respectively.

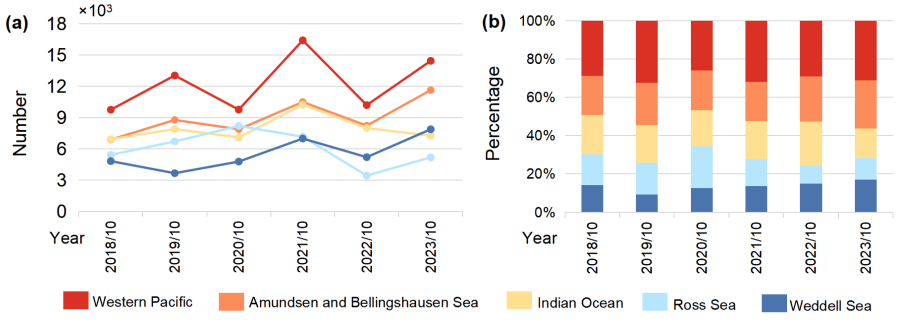

Figure 11. Annual variation trends of icebergs in five major Southern Ocean sectors from Oct. 2018 to Oct. 2023. Panels (a) and (b) present the number and percentage of icebergs of five categories in different sea sectors.

climate and offers valuable data support for future research. Moving forward, we plan to use imagery from additional months to study seasonal and interannual variations in iceberg distribution and their long-term impacts on marine ecosystems and climate systems. Besides, we attempt to backtrack and update this product as a "living" dataset, meaning it will be continuously updated and expanded as new input observations available, such as Sentinel-1A/B before 2018 and Sentinel-1C after 2024.

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

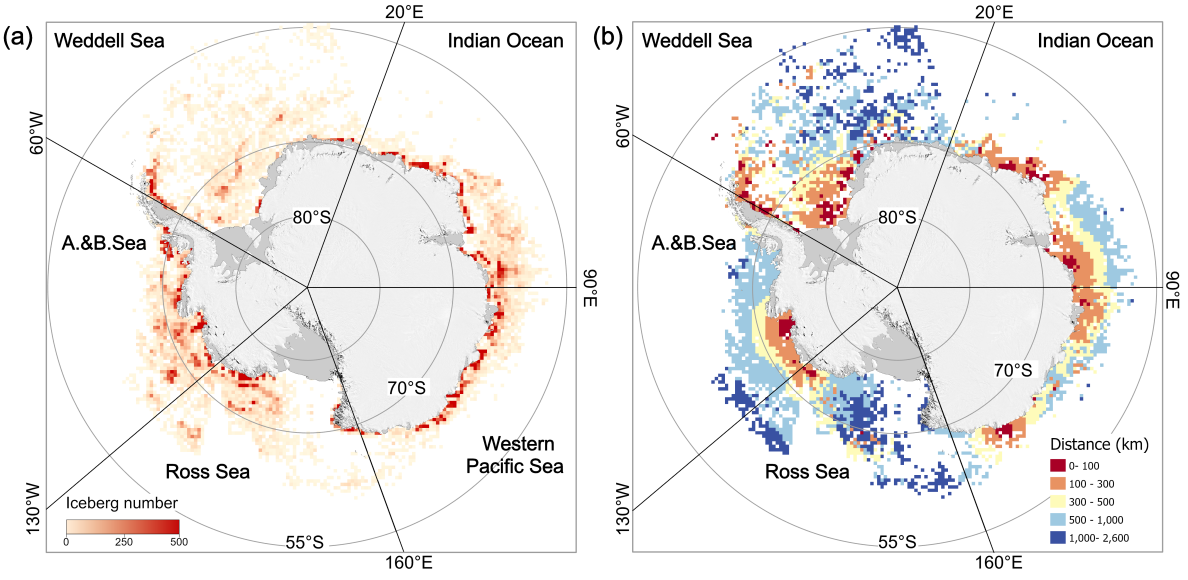

Figure 12. The spatial distribution characteristics of icebergs with sizes between 0.04 and 0.1 km^2 in 50 km × 50 km grids in the Southern Ocean. Panel (a) represents the average number of icebergs in each grid cell from 2018 to 2023; Panel (b) shows the average distance from the icebergs in each grid to the nearest large iceberg (area greater than 100 km^2).

7 Code and data availability

The GEE code for data acquisition, the MATLAB code for image segmentation, feature extraction, and the dataset of icebergs outlines in shapefile format along with their latitude and longitude, area, perimeter, and other attribute information, are all available at https://doi.org/10.5281/zenodo.15332566 (Liu and Chen, 2025), last access: 3 May 2025.

Appendix A: Feature description

(1) Statistical features: Calculated from the pixel backscatter values of each segment

1. CenterBackscatter: The grayscale value at the center position of the superpixel object. A superpixel is defined as a small, contiguous cluster of adjacent pixels that share similar backscatter characteristics, effectively representing a meaningful image region rather than individual pixels.

2. CenterStd: The standard deviation within a 3×3 range near the center of the superpixel. If there are fewer than 3×3 pixels around the center, then CenterStd = 0.

3. WeightedMean: Obtained from Equation A1:

$$\text{WeightedMean} = \sum_{ij} \frac{1}{D_{ij}} x_{ij} \tag{A1}$$

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

Open Access Earth System Science Data Discussions

Table A1. Feature Categories and Descriptions

Category	Feature	Note
Statistical features	CenterBackscatter	Calculated from the pixel backscatter values of each segment
	CenterStd	
	WeightedMean	
	Energy	
Histogram-based features	Mean	Calculated from the histogram of each segment
	Variance	
	Skewness	
	Kurtosis	
	Mode	
	Median	
	Slope	
Texture features	Entropy	Calculated from the Grey Level Co-occurrence Matrix (GLCM) of each segment
	$\text{Contrast}_{0/45/90/135^\circ}$	
	$\text{Correlation}_{0/45/90/135^\circ}$	
	$\text{Homogeneity}_{0/45/90/135^\circ}$	

where x_{ij} is the grayscale of the pixel at position (i, j), and D_{ij} is the distance from that pixel to the centroid of the superpixel.

4. Energy: Obtained from Equation A2:

$$\text{Energy} = \frac{1}{N}\sum_{ij} x_{ij}^2 \tag{A2}$$

where N is the total number of pixels within the superpixel.

(2) Histogram-based features (bin=0.1): Calculated from the histogram of each segment.

1. Mean: The average of all pixel grayscale values within the superpixel.

2. Variance: The variance of all pixel grayscale values within the superpixel.

3. Skewness: Used to measure the asymmetry of the histogram distribution of grayscale values of all pixels within a superpixel.It can derived from the equation A3:

$$\text{Skewness} = E\left[\left(\frac{x-\mu}{\sigma}\right)^3\right] \tag{A3}$$

4. Kurtosis: Characterizes the height of the peak at the mean of the probability distribution curve, that is, the shape of the curve's peak. The larger the kurtosis, the sharper the peak.

$$\text{Kurtosis} = E\left[\left(\frac{x-\mu}{\sigma}\right)^4\right] \tag{A4}$$

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

5. Mode: The most frequent value in the grayscale values of the superpixel. If multiple values occur with the same frequency, the Mode is the smallest of these values.

6. Median: The median of the grayscale values of all pixels within the superpixel.

7. Slope: The one-sided slope of the probability distribution curve.

$$\text{Slope} = \tan^{-1}\left(\frac{P(M)}{\max(x) - M}\right) \tag{A5}$$

Where M is the median of the grayscale values, and $P(M)$ is the probability density corresponding to the median.

(3) Texture features: Calculated from the Grey Level Co-occurrence Matrix(GLCM) of each segment.

1. Entropy: It characterizes the overall distribution of grayscale values in the image.

$$\text{Entropy} = -\sum_{n} P(i) \cdot \log_2 P(i) \tag{A6}$$

where n is the number of grayscale levels obtained by binning the histogram of all pixel grayscale values within a superpixel with bin = 0.1, and $P(i)$ is the probability density value corresponding to the i-th grayscale level.

2. $\text{Contrast}_{0/45/90/135^\circ}$

3. $\text{Correlation}_{0/45/90/135^\circ}$

4. $\text{Homogeneity}_{0/45/90/135^\circ}$

In our research, the Gray-Level Co-Occurrence Matrix (GLCM) is used to calculate the texture features of superpixels. The GLCM characterizes the texture of an image by calculating the frequency of occurrence of pixel pairs with specific values and spatial relationships in the image (Haralick et al., 1973). The elements of the Gray-Level Co-Occurrence Matrix are calculated using the Equation A7:

$$P(i,j) = \frac{P(i,j,d,\theta)}{\sum_i \sum_j P(i,j,d,\theta)} \tag{A7}$$

The element $P(i,j)$ in the matrix represents the probability of the occurrence of pixel pairs at a distance d in the direction θ. In this study, we consider the GLCM for the cases when $d = 0$ and $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$. For non-rectangular superpixels, missing pixels are filled with 0. After calculating the GLCM for each superpixel in these four directions, we can further compute metrics that describe contrast, correlation, and homogeneity. The equation is as follows:

$$\text{Contrast} = \sum_{i,j} (i-j)^2 P(i,j) \tag{A8}$$

$$\text{Correlation} = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i \sigma_j} \tag{A9}$$

$$\text{Homogeneity} = \sum_{i,j} \frac{P(i,j)}{1+(i-j)^2} \tag{A10}$$

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

Author contributions. ZC: Writing original draft, Visualization, Software, Formal analysis, Methodology. XL: Formal analysis, Conceptualization, Methodology. ZG: Methodology, review and editing. TL: review and editing, Supervision, Methodology. XC: Conceptualization, Supervision, Funding acquisition. TL: Formal analysis, review and editing. YL, QL, LZ and JL: review and editing. All authors participated in result interpretation.

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

Disclaimer. Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. This study was supported by the National Natural Science Foundation of China (42206249, 42306256, 42276246, 42422606), the Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) under Grant (311021008).

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S.: SLIC Superpixels Compared to State-of-the-Art Superpixel Methods, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34, 2274–2282, https://doi.org/10.1109/TPAMI.2012.120, 2012.

Amani, M., Ghorbanian, A., Ahmadi, S. A., Kakooei, M., Moghimi, A., Mirmazloumi, S. M., Moghaddam, S. H. A., Mahdavi, S., Ghahremanloo, M., Parsian, S., Wu, Q., and Brisco, B.: Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 5326–5350, https://doi.org/10.1109/JSTARS.2020.3021052, 2020.

Barbat, M. M., Rackow, T., Hellmer, H. H., Wesche, C., and Mata, M. M.: Three Years of Near-Coastal Antarctic Iceberg Distribution From a Machine Learning Approach Applied to SAR Imagery, Journal of Geophysical Research: Oceans, 124, 6658–6672, https://doi.org/10.1029/2019JC015205, 2019a.

Barbat, M. M., Wesche, C., Werhli, A. V., and Mata, M. M.: An adaptive machine learning approach to improve automatic iceberg detection from SAR images, ISPRS Journal of Photogrammetry and Remote Sensing, 156, 247–259, https://doi.org/10.1016/j.isprsjprs.2019.08.015, 2019b.

Benn, D. I. and Åström, J. A.: Calving glaciers and ice shelves, Advances in Physics: X, 3, 1513 819, https://doi.org/10.1080/23746149.2018.1513819, 2018.

Bigg, G. R., Cropper, T. E., O'Neill, C. K., Arnold, A. K., Fleming, A. H., Marsh, R., Ivchenko, V., Fournier, N., Osborne, M., and Stephens, R.: A model for assessing iceberg hazard, Natural Hazards, 92, 1113–1136, https://doi.org/10.1007/s11069-018-3243-x, 2018.

Braakmann-Folgmann, A., Shepherd, A., Gerrish, L., Izzard, J., and Ridout, A.: Observing the disintegration of the A68A iceberg from space, Remote Sensing of Environment, 270, 112 855, https://doi.org/10.1016/j.rse.2021.112855, 2022.

Budge, J. S. and Long, D. G.: A Comprehensive Database for Antarctic Iceberg Tracking Using Scatterometer Data, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11, 434–442, https://doi.org/10.1109/JSTARS.2017.2784186, 2018.

Deakin, K. A., Christie, F. D. W., Boxall, K., and Willis, I. C.: Oscillatory response of Larsen C Ice Shelf flow to the calving of iceberg A-68, Journal of Glaciology, 70, e61, https://doi.org/10.1017/jog.2023.102, 2024.

Depoorter, M. A., Bamber, J. L., Griggs, J. A., Lenaerts, J. T. M., Ligtenberg, S. R. M., Van Den Broeke, M. R., and Moholdt, G.: Calving fluxes and basal melt rates of Antarctic ice shelves, Nature, 502, 89–92, https://doi.org/10.1038/nature12567, 2013.

Dowdeswell, J. and Bamber, J.: Keel depths of modern Antarctic icebergs and implications for sea-floor scouring in the geological record, Marine Geology, 243, 120–131, https://doi.org/10.1016/j.margeo.2007.04.008, 2007.

Drinkwater, M. R., Hosseinmostafa, R., and Gogineni, P.: C-band backscatter measurements of winter sea-ice in the Weddell Sea, Antarctica, International Journal of Remote Sensing, 16, 3365–3389, https://doi.org/10.1080/01431169508954635, 1995.

Duprat, L. P. A. M., Bigg, G. R., and Wilton, D. J.: Enhanced Southern Ocean marine productivity due to fertilization by giant icebergs, Nature Geoscience, 9, 219–221, https://doi.org/10.1038/ngeo2633, 2016.

Ferdous, M. S., McGuire, P., Power, D., Johnson, T., and Collins, M.: A comparison of numerically modelled iceberg backscatter signatures with sentinel-1 C-band synthetic aperture radar acquisitions, Canadian Journal of Remote Sensing, 44, 232–242, https://doi.org/10.1080/07038992.2018.1495554, 2018.

Gladstone, R. M., Bigg, G. R., and Nicholls, K. W.: Iceberg trajectory modeling and meltwater injection in the Southern Ocean, Journal of Geophysical Research: Oceans, 106, 19 903–19 915, https://doi.org/10.1029/2000JC000347, 2001.

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sensing of Environment, 202, 18–27, https://doi.org/10.1016/j.rse.2017.06.031, 2017.

Griggs, J. and Bamber, J.: Antarctic ice-shelf thickness from satellite radar altimetry, Journal of Glaciology, 57, 485–498, https://doi.org/10.3189/002214311796905659, 2011.

Hammond, M. D. and Jones, D. C.: Freshwater flux from ice sheet melting and iceberg calving in the Southern Ocean, Geoscience Data Journal, 3, 60–62, https://doi.org/10.1002/gdj3.43, 2016.

Haralick, R. M., Shanmugam, K., and Dinstein, I.: Textural Features for Image Classification, IEEE Transactions on Systems, Man, and Cybernetics, SMC-3, 610–621, https://doi.org/10.1109/TSMC.1973.4309314, 1973.

Karvonen, J., Gegiuc, A., Niskanen, T., Montonen, A., Buus-Hinkler, J., and Rinne, E.: Iceberg Detection in Dual-Polarized C-Band SAR Imagery by Segmentation and Nonparametric CFAR (SnP-CFAR), IEEE Transactions on Geoscience and Remote Sensing, 60, 1–12, https://doi.org/10.1109/TGRS.2021.3070312, 2022.

Koo, Y., Xie, H., Mahmoud, H., Iqrah, J. M., and Ackley, S. F.: Automated detection and tracking of medium-large icebergs from Sentinel-1 imagery using Google Earth Engine, Remote Sensing of Environment, 296, 113 731, https://doi.org/10.1016/j.rse.2023.113731, 2023.

Li, T., Shokr, M., Liu, Y., Cheng, X., Li, T., Wang, F., and Hui, F.: Monitoring the tabular icebergs C28A and C28B calved from the Mertz Ice Tongue using radar remote sensing data, Remote Sensing of Environment, 216, 615–625, https://doi.org/10.1016/j.rse.2018.07.028, 2018.

Lin, H., Cheng, X., Li, T., Shi, Q., Liang, Q., Meng, X., Wang, S., and Zheng, L.: Assessing the degree of impact from iceberg activities on penguin colonies of Clarence Island, Acta Oceanologica Sinica, 43, 105–109, https://doi.org/10.1007/s13131-024-2355-2, 2024.

Liu, X., Cheng, X., Liang, Q., Li, T., Peng, F., Chi, Z., and He, J.: Grounding event of iceberg D28 and its interactions with seabed topography, Remote Sensing, 14, 154, https://doi.org/10.3390/rs14010154, 2021.

Liu, X.-Y. and Chen, Z.-L.: A 6-year circum-Antarctic icebergs dataset (2018-2023) [Data Set], https://doi.org/10.5281/zenodo.15332566, 2025.

Liu, Y., Moore, J. C., Cheng, X., Gladstone, R. M., Bassis, J. N., Liu, H., Wen, J., and Hui, F.: Ocean-driven thinning enhances iceberg calving and retreat of Antarctic ice shelves, Proceedings of the National Academy of Sciences, 112, 3263–3268, https://doi.org/10.1073/pnas.1415137112, 2015.

Long, D. G., Ballantyn, J., and Bertoia, C.: Is the number of Antarctic icebergs really increasing?, Eos, Transactions American Geophysical Union, 83, 469–474, https://doi.org/10.1029/2002EO000330, 2002.

Mazur, A., Wåhlin, A., and Krężel, A.: An object-based SAR image iceberg detection algorithm applied to the Amundsen Sea, Remote Sensing of Environment, 189, 67–83, https://doi.org/10.1016/j.rse.2016.11.013, 2017.

Orheim, O., Giles, A. B., Jacka, T. H. J., and Moholdt, G.: Quantifying dissolution rates of Antarctic icebergs in open water, Annals of Glaciology, 64, 170–180, https://doi.org/10.1017/aog.2023.26, 2023.

Parkinson, C. L. and Cavalieri, D. J.: Antarctic sea ice variability and trends, 1979–2010, The Cryosphere, 6, 871–880, https://doi.org/10.5194/tc-6-871-2012, 2012.

Qadir, M., Smakhtin, V., Koo-Oshima, S., and Guenther, E., eds.: Unconventional Water Resources, Springer International Publishing, Cham, ISBN 978-3-030-90145-5 978-3-030-90146-2, https://doi.org/10.1007/978-3-030-90146-2, 2022.

Qi, M., Liu, Y., Liu, J., Cheng, X., Lin, Y., Feng, Q., Shen, Q., and Yu, Z.: A 15-year circum-Antarctic iceberg calving dataset derived from continuous satellite observations, Earth System Science Data, 13, 4583–4601, https://doi.org/10.5194/essd-13-4583-2021, 2021.

https://doi.org/10.5194/essd-2025-51
Preprint. Discussion started: 15 May 2025

Rackow, T., Wesche, C., Timmermann, R., Hellmer, H. H., Juricke, S., and Jung, T.: A simulation of small to giant Antarctic iceberg evolution: Differential impact on climatology estimates, Journal of Geophysical Research: Oceans, 122, 3170–3190, https://doi.org/10.1002/2016JC012513, 2017.

Rignot, E., Jacobs, S., Mouginot, J., and Scheuchl, B.: Ice-Shelf Melting Around Antarctica, Science, 341, 266–270, https://doi.org/10.1126/science.1235798, 2013.

Silva, T. A. M., Bigg, G. R., and Nicholls, K. W.: Contribution of giant icebergs to the Southern Ocean freshwater flux, Journal of Geophysical Research: Oceans, 111, 2004JC002 843, https://doi.org/10.1029/2004JC002843, 2006.

Smith, K. L., Robison, B. H., Helly, J. J., Kaufmann, R. S., Ruhl, H. A., Shaw, T. J., Twining, B. S., and Vernet, M.: Free-Drifting Icebergs: Hot Spots of Chemical and Biological Enrichment in the Weddell Sea, Science, 317, 478–482, https://doi.org/10.1126/science.1142834, 2007.

Stuart, K. and Long, D.: Iceberg size and orientation estimation using SeaWinds, Cold Regions Science and Technology, 69, 39–51, https://doi.org/10.1016/j.coldregions.2011.07.006, 2011a.

Stuart, K. and Long, D.: Tracking large tabular icebergs using the SeaWinds Ku-band microwave scatterometer, Deep Sea Research Part II: Topical Studies in Oceanography, 58, 1285–1300, https://doi.org/10.1016/j.dsr2.2010.11.004, 2011b.

Tournadre, J., Girard-Ardhuin, F., and Legrésy, B.: Antarctic icebergs distributions, 2002–2010, Journal of Geophysical Research: Oceans, 117, 2011JC007 441, https://doi.org/10.1029/2011JC007441, 2012.

Tournadre, J., Bouhier, N., Girard-Ardhuin, F., and Rémy, F.: Large icebergs characteristics from altimeter waveforms analysis, Journal of Geophysical Research: Oceans, 120, 1954–1974, https://doi.org/10.1002/2014JC010502, 2015.

Tournadre, J., Bouhier, N., Girard-Ardhuin, F., and Rémy, F.: Antarctic icebergs distributions 1992–2014, Journal of Geophysical Research: Oceans, 121, 327–349, https://doi.org/10.1002/2015JC011178, 2016.

Tournadre, J., Piollé, J. F., Accensi, M., and Girard-Ardhuin, F.: The ALTIBERG iceberg data base version 4.0, Antarctic and Arctic data sets, Tech. Rep. Doc. Techni.LOPS 2024- 01 version4.0, Institut Français de Recherche pour l'Exploitation de la Mer, https://doi.org/10.13140/RG.2.2.35719.71841c, 2024.

Wesche, C. and Dierking, W.: Iceberg signatures and detection in SAR images in two test regions of the Weddell Sea, Antarctica, Journal of Glaciology, 58, 325–339, https://doi.org/10.3189/2012J0G11J020, 2012.

Wesche, C. and Dierking, W.: Near-coastal circum-Antarctic iceberg size distributions determined from Synthetic Aperture Radar images, Remote Sensing of Environment, 156, 561–569, https://doi.org/10.1016/j.rse.2014.10.025, 2015.

Wu, S.-Y. and Hou, S.: Impact of icebergs on net primary productivity in the Southern Ocean, The Cryosphere, 11, 707–722, https://doi.org/10.5194/tc-11-707-2017, 2017.

Xu, L., Yan, Q., Xia, Y., and Jia, J.: Structure extraction from texture via relative total variation, ACM Transactions on Graphics, 31, 1–10, https://doi.org/10.1145/2366145.2366158, 2012.

Zhou, Z.-H.: Ensemble methods: foundations and algorithms, CRC press, https://doi.org/https://doi.org/10.1201/b12207, 2012.