https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

Open Access Earth System Science Data Discussions

An 8-day Antarctic supraglacial lake dataset from MODIS

Shuo Wei[1,2], Lei Zheng[1,2], Qi Liang[1,2], Teng Li[1,2], Xiao Cheng[1,2]

[1]School of Geospatial Engineering and Science, Sun Yat-sen University, and Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China
[2]Key Laboratory of Comprehensive Observation of Polar Environment (Sun Yat-sen University), Ministry of Education, Zhuhai 519082, China

Correspondence to: Lei Zheng (zhenglei6@mail.sysu.edu.cn)

Abstract. Supraglacial lakes (SGLs) are widely distributed across Antarctica and play an important role in modulating surface energy balance through albedo feedback, promoting ice-shelf disintegration via hydrofracture, and influencing ice dynamics. Existing SGL monitoring studies mainly rely on narrow-swath satellite data, such as Landsat and Sentinel-2, resulting in discontinuous observations with relatively long revisit intervals and limiting the ability to capture the rapid evolution of supraglacial hydrological processes. Here, an 8-day Antarctic SGL fraction dataset spanning 2000–2023 is presented. The dataset is generated by integrating high-temporal-resolution Moderate Resolution Imaging Spectroradiometer (MODIS) imagery with high-spatial-resolution Sentinel-2 data within a machine-learning framework. The dataset reveals that Antarctic SGLs are highly dynamic and short-lived, with approximately 83% of lakes exhibiting mean persistence rates below 5%. The multi-year mean maximum SGL area is estimated at 4,103 ± 1,479 km^2. Clear spatial heterogeneity in peak timing is further revealed, with SGL extent peaking approximately one week earlier in West Antarctica than in East Antarctica and about two weeks earlier than on the Antarctic Peninsula. Spatially, approximately 65% of SGLs are located within 10 km of grounding lines and are closely associated with blue-ice and exposed rock areas. These long-term, high-temporal-resolution observations provide a valuable basis for investigating the spatiotemporal variability of Antarctic SGLs and their associated impacts. The dataset is publicly available at https://doi.org/10.5281/zenodo.19936100.

1 Introduction

Polar ice sheets are subject to accelerated mass loss (Mouginot et al., 2019; Rignot et al., 2019; Shepherd et al., 2012; van et al., 2016; Velicogna et al., 2014). Accounting for approximately 91% of Earth's ice mass, the Antarctic Ice Sheet is the primary potential contributor to global sea-level rise (The IMBIE Team, 2018). As global warming accelerates, rising air temperatures can directly enhance surface melting of the Antarctic Ice Sheet, leading to the formation of supraglacial lakes (SGLs) within topographic lows over impermeable snow and ice surfaces (Echelmeyer et al., 1991).

Once formed, SGLs can substantially influence the energy budget and mass dynamics of the Antarctic Ice Sheet via a range of processes. SGLs lower surface albedo, enhance melting, and warm the adjacent ice column (Hubbard et al., 2016; Lüthje et al., 2006; Tedesco et al., 2012). When lake water drains to the glacier bed, it enhances basal sliding and accelerates ice flow,

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

as observed in both Greenland and the Antarctic Peninsula (Bartholomew et al., 2010; Das et al., 2008; Shepherd et al., 2009; Tedesco et al., 2013; Zwally et al., 2002). Rapid lake drainage can also induce ice-shelf collapse through flexure and hydrofracture, leading to substantial mass loss (Banwell et al., 2013; Scambos et al., 2000). Antarctic-wide melt rates are projected to double by 2050, increasing both the extent and the frequency of SGLs (Arthur et al., 2020; Trusel et al., 2015). Systematic quantification of the spatiotemporal evolution of SGLs is therefore critical for assessing their role in ice-sheet stability and future sea-level rise.

Remote sensing provides an effective approach for investigating the spatial distribution and temporal evolution of SGLs in Antarctica (Niu et al., 2023). High-spatial-resolution optical satellite data, such as Landsat, Sentinel-2, and WorldView, are widely used for SGL monitoring (Jiang et al., 2022; McMillan et al., 2007; Safarov et al., 2024; Shu et al., 2023). However, despite their fine spatial detail, the effective temporal resolution of these data over Antarctica is often limited by revisit frequency, cloud cover, and illumination conditions, resulting in discontinuous observations. These limitations hinder the capture of the rapid evolution of SGLs, thereby limiting the understanding of their short-term dynamics. To improve temporal coverage, medium- to low-resolution satellite imagery, such as data from the Moderate Resolution Imaging Spectroradiometer (MODIS), has been used for SGL monitoring due to its high revisit frequency (Hubbard et al., 2016; Kingslake et al., 2015; Lenaerts et al., 2017; Selmes et al., 2011). However, these datasets are typically combined with hard classification approaches that assign each pixel to a discrete class. Under these conditions, a substantial proportion of pixels represent mixtures of surface types (e.g., water, ice, and slush). This mixed-pixel effect introduces significant bias in lake extent estimates, reduces the detectability of small or narrow water bodies, and ultimately restricts monitoring to larger lakes (Hubbard et al., 2016; Kingslake et al., 2015; Lenaerts et al., 2017). Consequently, trade-offs between revisit frequency and swath width have prevented existing monitoring products from achieving both frequent observations and applicability to lakes of different sizes (Leeson et al., 2013).

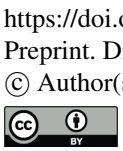
In this study, a new Antarctic SGL fraction dataset derived from MODIS imagery is presented. This dataset contains SGL fraction estimates at 8-day temporal and 1 km spatial resolutions, covering Antarctic melt seasons between 2000 and 2023. The SGL fraction retrieval is achieved by integrating moderate-resolution observations (MODIS) with high-resolution data (Sentinel-2) within a machine-learning–based soft classification framework. By combining high temporal resolution with sub-pixel characterization, the dataset enables high-frequency monitoring of SGL area variations across a wide range of lake sizes. This long-term record facilitates the investigation of the rapid spatiotemporal evolution of Antarctic SGLs and supports studies of their driving mechanisms, interactions with ice-shelf stability, and implications for ice-sheet energy and mass balance.

2 Data

2.1 MODIS surface reflectance data

Due to its wide swath, high temporal resolution, and long-term continuity, MODIS imagery from the Terra satellite is used for Antarctic SGL monitoring in this study. MOD09GA and MOD09A1 are two MODIS surface reflectance products used in this

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026



Open Access Earth System Science Data Discussions

study. Both products provide seven spectral bands spanning the visible to infrared range, include a quality assurance layer, and are atmospherically corrected (Justice et al., 2002). The MOD09 product is available at a native spatial resolution of 500 m. To improve retrieval robustness and reduce sensitivity to noise and residual cloud contamination, both MOD09GA and MOD09A1 are resampled to a common spatial resolution of 1 km. Compared with the native 500 m resolution, the new grid provides more stable spectral signals and reduces sub-pixel variability, thereby improving the reliability of SGL fraction estimates (Markham et al., 2023; Wu and Li, 2009).

The MOD09GA product, available at daily resolution, is used to construct temporally matched observations with Sentinel-2 imagery (Fig. 1) and to provide high-temporal-resolution input for developing the SGL fraction retrieval algorithm. MOD09A1 is an 8-day composite derived from MOD09GA with substantially reduced cloud contamination and improved radiometric consistency. It is employed to generate the final 8-day SGL fraction dataset across Antarctica. Both products can be obtained from the Google Earth Engine (GEE) data catalog.

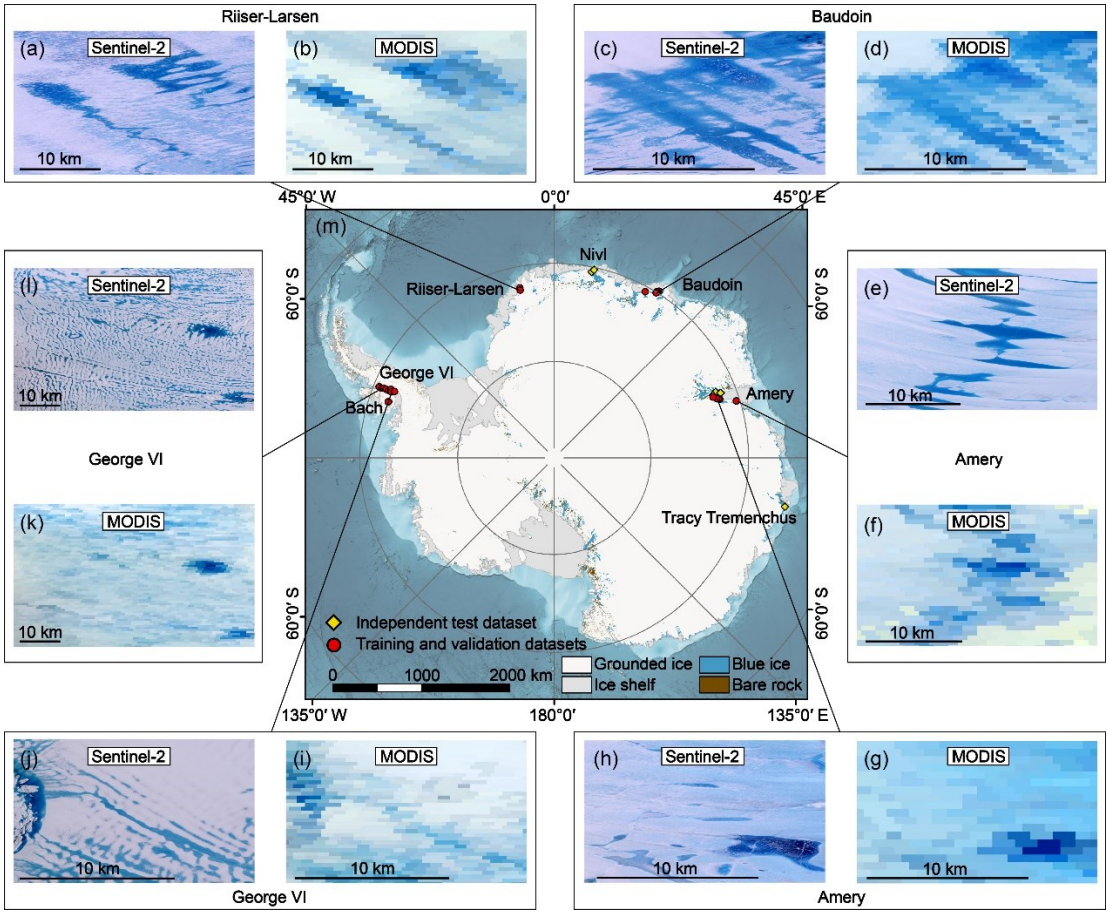


Figure 1: Representative SGL samples derived from MODIS and Sentinel-2 imagery across the Antarctic Ice Sheet. (a)–(l) True-color composite images from Sentinel-2 and MODIS. (m) Geographic distribution of sample locations across Antarctica.

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

2.2 Sentinel-2 surface reflectance data

Owing to its 10 m spatial resolution, Sentinel-2 imagery provides detailed observations of lake extents and serves as a reliable reference for sub-pixel lake characterization. Sentinel-2 data, available via the GEE data catalog, include 13 spectral bands covering the visible to shortwave-infrared wavelengths, including the red-edge region (Drusch et al., 2012). This study uses Level-1C products, with details of the selected imagery listed in Table S1. To capture the wide range of SGL morphologies and evolutionary stages across Antarctica, Sentinel-2 images are selected from multiple regions and periods throughout the melt season (Fig. 1). The selection of these regions is guided by previously reported SGL distributions and representative lake-forming zones (Dell et al., 2020; Kingslake et al., 2015, 2017; Stokes et al., 2019). Quality control criteria are applied to ensure reliable lake extraction, including cloud coverage below 20% and a solar elevation angle greater than 20° (Williamson et al., 2018).

2.3 Auxiliary data

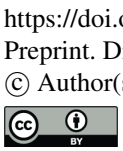
Several auxiliary datasets are employed to improve the reliability of SGL mapping across Antarctica. The ice-covered domain of Antarctica is delineated using high-resolution coastline vector polygons from the Antarctic Digital Database (ADD), provided by the Scientific Committee on Antarctic Research (SCAR). To minimize misclassification between SGLs and other surface features, exposed bedrock polygons from the ADD are applied as a spatial mask. Surface slope is calculated from the CryoSat-2 digital elevation model (DEM) at 1 km spatial resolution (Slater et al., 2018). This dataset is available from the Centre for Polar Observation and Modelling (CPOM) at the University of Leeds and can be accessed via the GEE data catalog. In addition, two independent datasets are incorporated to further constrain the SGL mapping results: (1) a daily Antarctic melt-flux dataset (Zheng et al., 2025), which provides information on surface melting conditions, and (2) an existing Antarctic SGL dataset (Tuckett et al., 2025).

The presence of mixed pixels makes it essential to estimate the fraction of SGLs within each pixel to accurately characterize lake extent across Antarctica. To address this issue, the dataset is generated using a two-stage machine-learning framework, in which a Random Forest classification (RFC) model and a Random Forest regression (RFR) model are applied to lake detection and fraction retrieval, respectively. The RF algorithm is an ensemble learning approach that builds on multiple decision trees and is well known for its strong generalization ability (Breiman, 2001). Surface reflectance from MODIS imagery is used as the primary input to the RF models, while ground-truth data are derived from Sentinel-2 imagery. Sentinel-2 data are processed using supervised classification to obtain lake class labels and lake fractions (Fig. 2).

3 Methods

The presence of mixed pixels makes it essential to estimate the fraction of SGLs within each pixel to accurately characterize lake extent across Antarctica. To address this issue, the dataset is generated using a two-stage machine-learning framework, in which a Random Forest classification (RFC) model and a Random Forest regression (RFR) model are applied to lake

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026



Open Access Earth System Science Data Discussions

detection and fraction retrieval, respectively. The RF algorithm is an ensemble learning approach that builds on multiple decision trees and is well known for its strong generalization ability (Breiman, 2001). Surface reflectance from MODIS imagery is used as the primary input to the RF models, while ground-truth data are derived from Sentinel-2 imagery. Sentinel-2 data are processed using supervised classification to obtain lake class labels and lake fractions (Fig. 2).

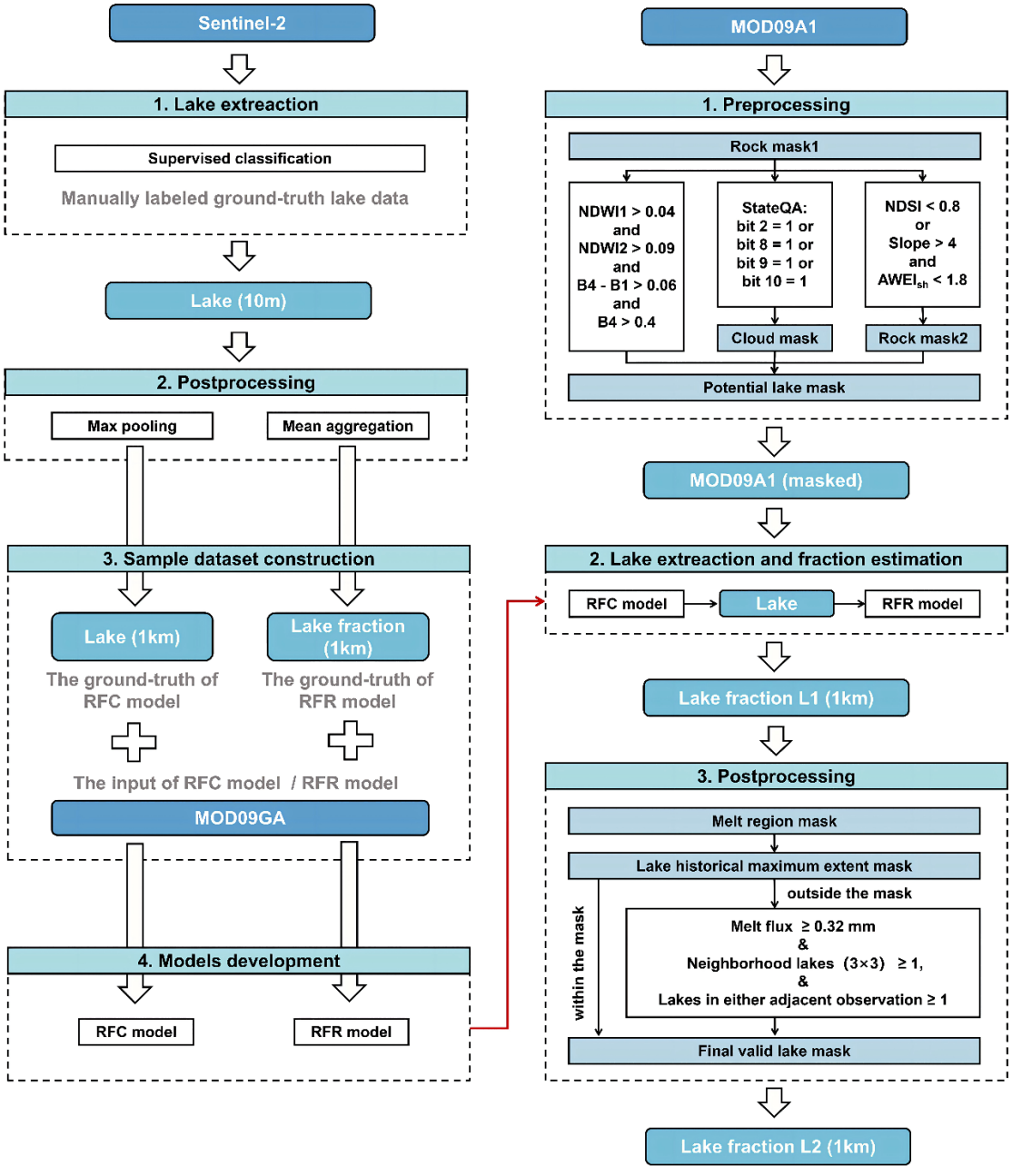


Figure 2: Workflow of the Antarctic SGL fraction dataset generation from Sentinel-2 and MODIS imagery.

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

The production workflow, which is divided into several stages, is shown in Fig. 2. First, masks are applied to exclude bare rock and cloud-covered areas, and candidate lake pixels are identified from MOD09GA surface reflectance product using a threshold-based approach. Subsequently, a two-stage machine-learning soft classification framework (RFC and RFR) is implemented, consisting of lake classification followed by fraction estimation. Finally, auxiliary datasets, including melt flux data and existing Antarctic SGL datasets, are incorporated to constrain the retrieval results and improve their physical consistency.

3.1 Extraction of SGLs from Sentinel-2 imagery

High-resolution Sentinel-2 data are used to construct a reference SGL dataset (Fig. 2). Sentinel-2 image selection is guided by known lake locations and manual visual interpretation. This selection ensures broad spatial coverage across Antarctica and representation of multiple melt seasons (Kingslake et al., 2015, 2017; Stokes et al., 2019). Each selected image is processed using a supervised classification algorithm to derive a single-band surface class map, where water is labeled as 1 and non-water as 0 (Niu et al., 2023). Only images meeting the classification accuracy threshold (> 95%) are selected for the final dataset.

Classification results at 10 m resolution are resampled to 1 km using two different strategies. First, max pooling is applied to generate binary SGL labels at MODIS pixel scale. Second, mean aggregation is used to derive lake fraction values at 1 km resolution (Liang and Liu, 2021; McKenna et al., 2025). These two datasets are used as ground-truth data for the MODIS-based RFC and RFR models. Following quality assessment, a total of 26 regional sample areas are retained, comprising approximately 5,000 lake samples and 10,000 non-lake samples. Among these regions, 19 regions are used for model training and parameterization, whereas 7 spatially independent regions are reserved for validation, allowing an unbiased assessment of model performance (Fig. 1 and Table S1).

3.2 Extraction of SGLs from MODIS imagery

Reliable estimation of SGL fractions depends on accurate identification of MODIS SGL pixels. In this study, any MODIS pixel with a lake fraction greater than 0 is defined as an SGL pixel. While this definition facilitates the detection of small lakes, it also increases the risk of misclassification due to mixed surface signals within individual pixels. To address this challenge, a two-stage hierarchical classification approach is adopted (Fig. 2).

In the first stage, potential lake pixels are identified using a multi-threshold approach that exploits spectral contrasts between water, snow, and ice. Following Corr et al. (2022), two normalized difference water indices (NDWI and $NDWI_{ice}$) are used to identify candidate SGL pixels in MODIS data. NDWI enhances open water signals by contrasting water with surrounding land or ice surfaces (McFeeters, 1996), while $NDWI_{ice}$ is specifically designed to improve the discrimination of meltwater from high-reflectance snow and ice backgrounds (Yang et al., 2013). By combining these two indices, lake pixel detection is enhanced, as their complementary sensitivities help reduce spectral ambiguity and misclassification under varying surface conditions.

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

Open Access Earth System Science Data Discussions

$$NDWI = \frac{B3 - B1}{B3 + B1} \quad (1)$$

$$NDWI_{ice} = \frac{B4 - B2}{B4 + B2} \quad (2)$$

Because MODIS pixels often contain mixed surface types, threshold values are empirically calibrated using reference lake samples derived from high-resolution Sentinel-2 imagery, as described in Text S1. The thresholds for NDWI (0.04) and $NDWI_{ice}$ (0.08) are shown in Fig. S1.

Additional thresholds based on visible bands ($B4 - B1 > 0.06$ and $B3 > 0.3$) are applied to further refine the candidate pixels (Moussavi et al., 2020). These thresholds increase the spectral separability of water from surrounding ice or snow surfaces, thereby improving SGL pixel identification. However, NDWI-based classification may still lead to the misclassification of rock shadows and cloud-contaminated pixels as water. To address this issue, the Automated Water Extraction Index for shadow areas (AWEIsh (< 1.8), Eq. (3)) is introduced in combination with surface slope (> 4°) to better distinguish SGL from shadowed surfaces (Feyisa et al., 2014).

$$AWEI_{sh} = B3 + 2.5 \times B4 - 1.5 \times (B2 + B6) - 0.25 \times B7 \quad (3)$$

In addition, cloud and cloud-shadow contamination are mitigated using the quality assurance bands provided with MODIS surface reflectance products, which have been shown to be effective in Antarctic environments (Ackerman et al., 1998, 2008).

In the second stage, candidate pixels are refined using machine learning to determine the final set of SGL pixels. The RFC model uses seven MODIS surface reflectance bands (B1–B7) as input features (Fig. 3). Binary SGL labels are based on Sentinel-2 imagery (Sect. 3.1). Model performance, including robustness and generalization, is assessed via five-fold cross-validation (Wong, 2015). Key RFC hyperparameters are optimized using a grid search strategy to balance model accuracy and the risk of overfitting (Kuhn and Johnson, 2013).

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

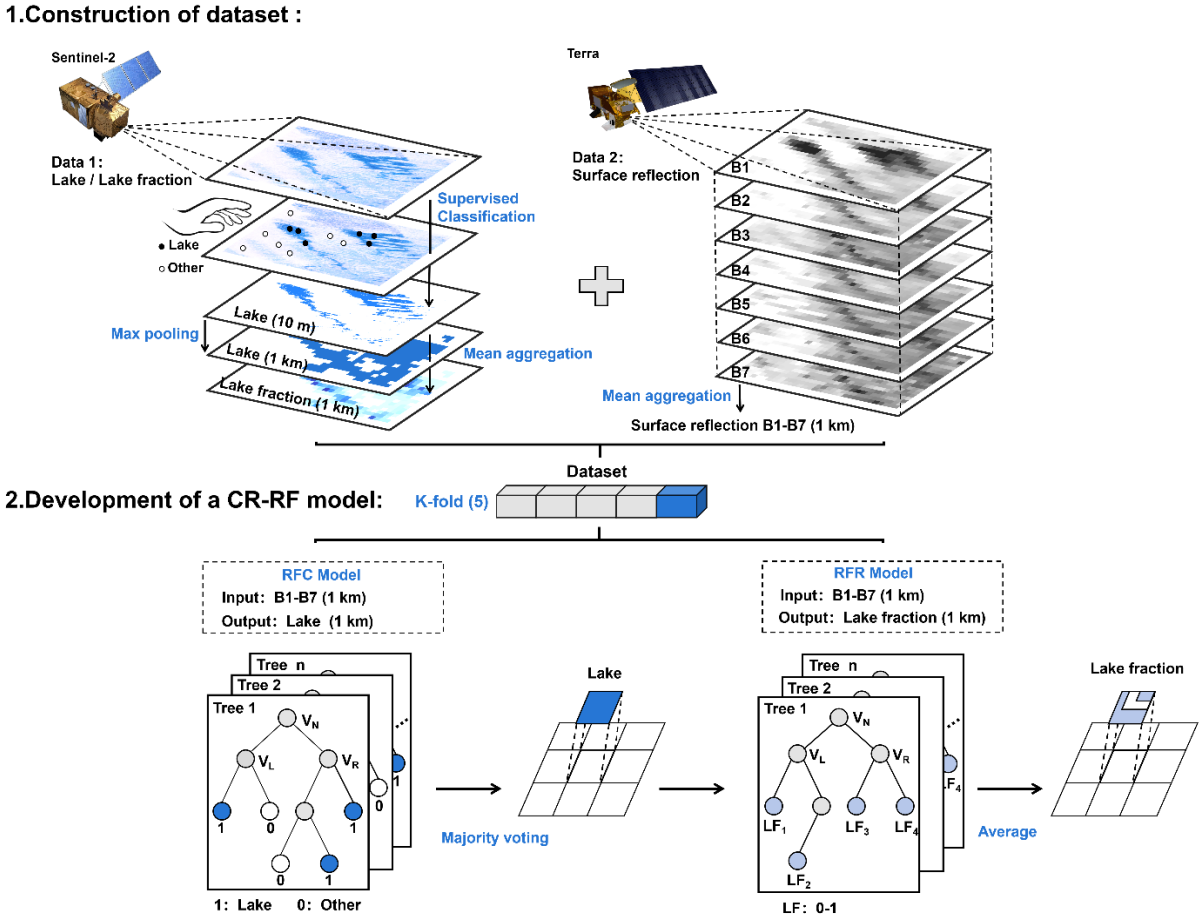


Figure 3: Architecture of the two-stage (RFC and RFR) model for MPF estimation.

3.3 Estimation of SGL fraction from MODIS imagery

Lake fraction retrieval is conducted using an RFR model. Its model architecture is similar to that used in lake pixel identification, with seven MODIS surface reflectance bands serving as input features (Fig. 3). Model parameters are also optimized using a grid search strategy, and performance is evaluated using five-fold cross-validation (Kuhn and Johnson, 2013; Wong, 2015). Unlike the RFC model, the RFR model produces continuous lake fraction estimates.

The regression model is trained using the daily MOD09GA surface reflectance product to maximize spatial and temporal consistency in the training samples. For operational product generation, the 8-day composite MOD09A1 product is adopted, as it effectively reduces cloud contamination and improves the stability of the retrieved lake fraction estimates. The trained RFR model is applied to retrieve lake fraction values for all MODIS pixels identified as containing SGLs, forming the basis of the Antarctic SGL fraction dataset

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

Open Access Earth System Science Data Discussions

3.4 Post-processing

Surface melt is widely recognized as a prerequisite for supraglacial meltwater accumulation (Leeson et al., 2012; Stokes et al., 2019). An independent Antarctic melt flux dataset derived from SSM/I observations is used as an auxiliary constraint to improve the physical consistency of the retrieved SGL dataset (Zheng et al., 2025). Melt masks are constructed for each 8-day SGL fraction image when the melt criterion exceeds a threshold corresponding to a volumetric liquid water content of approximately 0.04 (Mätzler, 1996; Ulaby et al., 1981; Tedesco, 2007; Fig. 2).

In addition to the physical constraint imposed by surface melt, a historical spatial constraint is applied to reduce misclassification caused by noise in MOD09A1 observations (Gignac et al., 2021). A historical SGL distribution dataset derived from Landsat imagery for 2006–2021 (Tuckett et al., 2025) is used to generate a maximum historical lake extent mask. Given the high temporal sampling frequency of MODIS observations, lake pixels outside the historical maximum extent are not automatically removed. Instead, they are conditionally retained when both spatial and temporal consistency criteria are satisfied (Fig. 2).

4 Results and discussion

4.1 Accuracy assessment

The performance of SGL classification and fraction retrieval is evaluated using common metrics, including overall accuracy (OA), precision, recall, F1-score (F1), the correlation coefficient (r), and root mean square error (RMSE) (Foody, 2002; Legates and McCabe, 1999; Mountrakis et al., 2011; Willmott, 1981).

The performance of RFC and RFR models is initially evaluated through five-fold cross-validation. The samples are divided into five equal subsets. In each iteration, four subsets are used for training and the remaining one is used for evaluation. For SGL classification, all metrics are around 0.90, with F1-scores consistently exceeding 0.93 across all folds (Table 1). The small variation among folds indicates stable model behavior and demonstrates that lake pixels can be reliably identified even under relatively inclusive lake definitions.

Table 1. Accuracy of SGL extraction and fraction estimation.

Fold	OA	Precision	Recall	F1	r	RMSE
1	0.90	0.92	0.94	0.93	0.85	0.12
2	0.91	0.93	0.95	0.94	0.86	0.12
3	0.90	0.91	0.95	0.93	0.84	0.13
4	0.90	0.92	0.93	0.93	0.85	0.12
5	0.89	0.92	0.93	0.93	0.85	0.12
All	0.90	0.92	0.94	0.93	0.85	0.12

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

Open Access Earth System Science Data Discussions

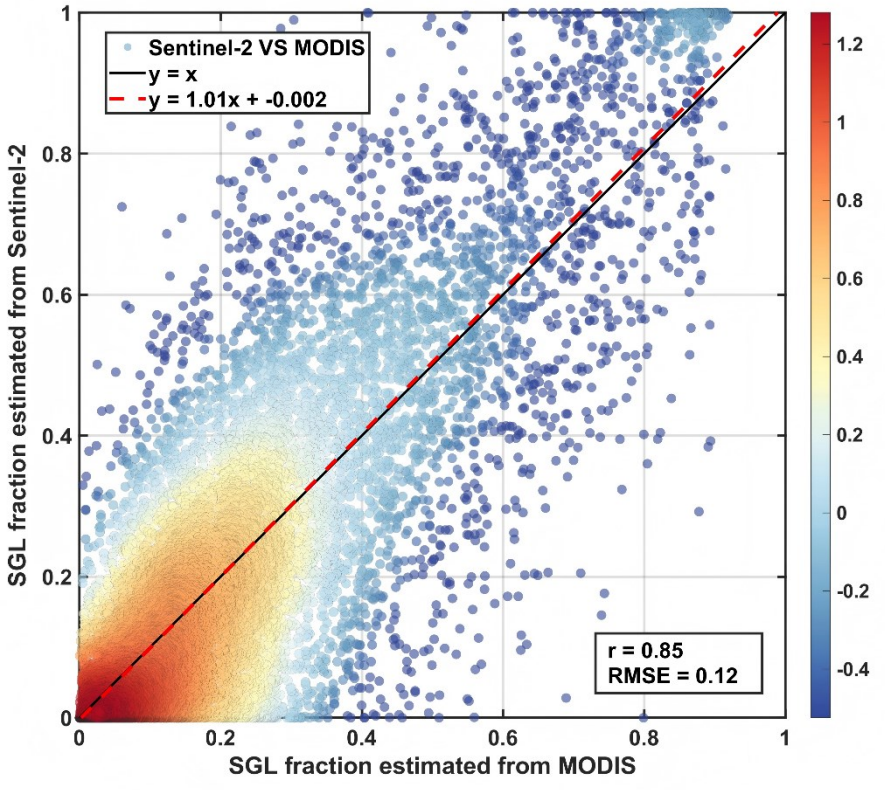


Figure 4: Validation of MODIS-derived SGL fractions against Sentinel-2 reference fractions. The black line indicates the 1:1 relationship, whereas the fitted regression is indicated by the red dashed line.

To further assess model transferability, independent validation is conducted using samples from seven spatially distinct regions covering multiple melt seasons from 2017 to 2024 (Fig. 1). Model performance varies due to differences in surface conditions, illumination geometry, and atmospheric variability among the validation samples (Table 2). Nevertheless, the independent tests consistently yield OA > 0.82 and r > 0.73, indicating stable performance across diverse Antarctic environments. Figure 5 illustrates representative validation examples, covering the full workflow from lake detection to fraction retrieval. Comparison with 10 m Sentinel-2 imagery shows that the retrieved SGL fractions remain spatially consistent under varying temporal and environmental conditions. This supports the applicability of the dataset for Antarctic analyses.

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

Open Access Earth System Science Data Discussions

Table 2. Classification and regression accuracy of independent test samples.

Region	data	OA	r	RMSE
Amery	01/27/2017	0.90	0.79	0.13
Nivl	01/26/2017	0.88	0.74	0.14
Amery	01/02/2019	0.85	0.74	0.13
Amery	01/03/2019	0.86	0.76	0.14
Tracy Tremenchus	01/29/2019	0.82	0.75	0.14
Voyeykov	12/28/2019	0.82	0.73	0.15
Nivl	01/26/2020	0.85	0.75	0.14

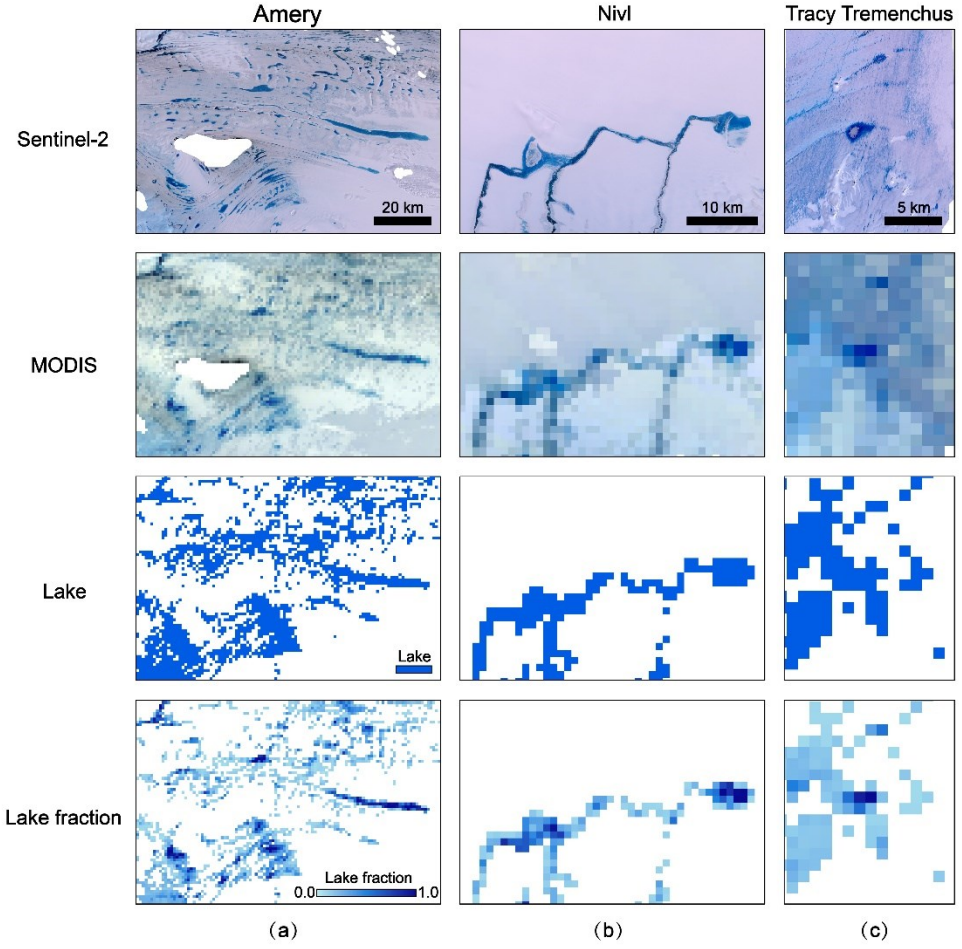


Figure 5: Lake extraction and fraction retrieval for representative independent test samples. (a) Amery Ice Shelf, 27 January 2017; (b) Nivl Ice Shelf, 26 January 2020; (c) Tracy Tremenchus Ice Shelf, 29 January 2019.

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

4.2 Occurrences of Antarctic SGLs

Time-series analyses show the evolution of SGL fractions on the Amery Ice Shelf from early December to early March across three periods (2008–2009, 2013–2014, and 2018–2019; Fig. 6). Peak SGL extent and the timing of peak development vary among the three periods, highlighting pronounced interannual variability in SGLs. Despite this variability, all periods exhibit a consistent seasonal pattern. In addition, rapid decreases can occur even during overall growth phases, as illustrated by a decline of 119 km^2 in SGL area on the Amery Ice Shelf during 9–16 January 2019 compared with the previous 8-day period. This rapid fluctuation demonstrates the dataset's ability to capture strong temporal variability in SGLs. Such transient dynamics are difficult to resolve using low-frequency optical observations but are clearly captured by this dataset, demonstrating the value of the 8-day temporal resolution for characterizing short-term variability in Antarctic SGLs.

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

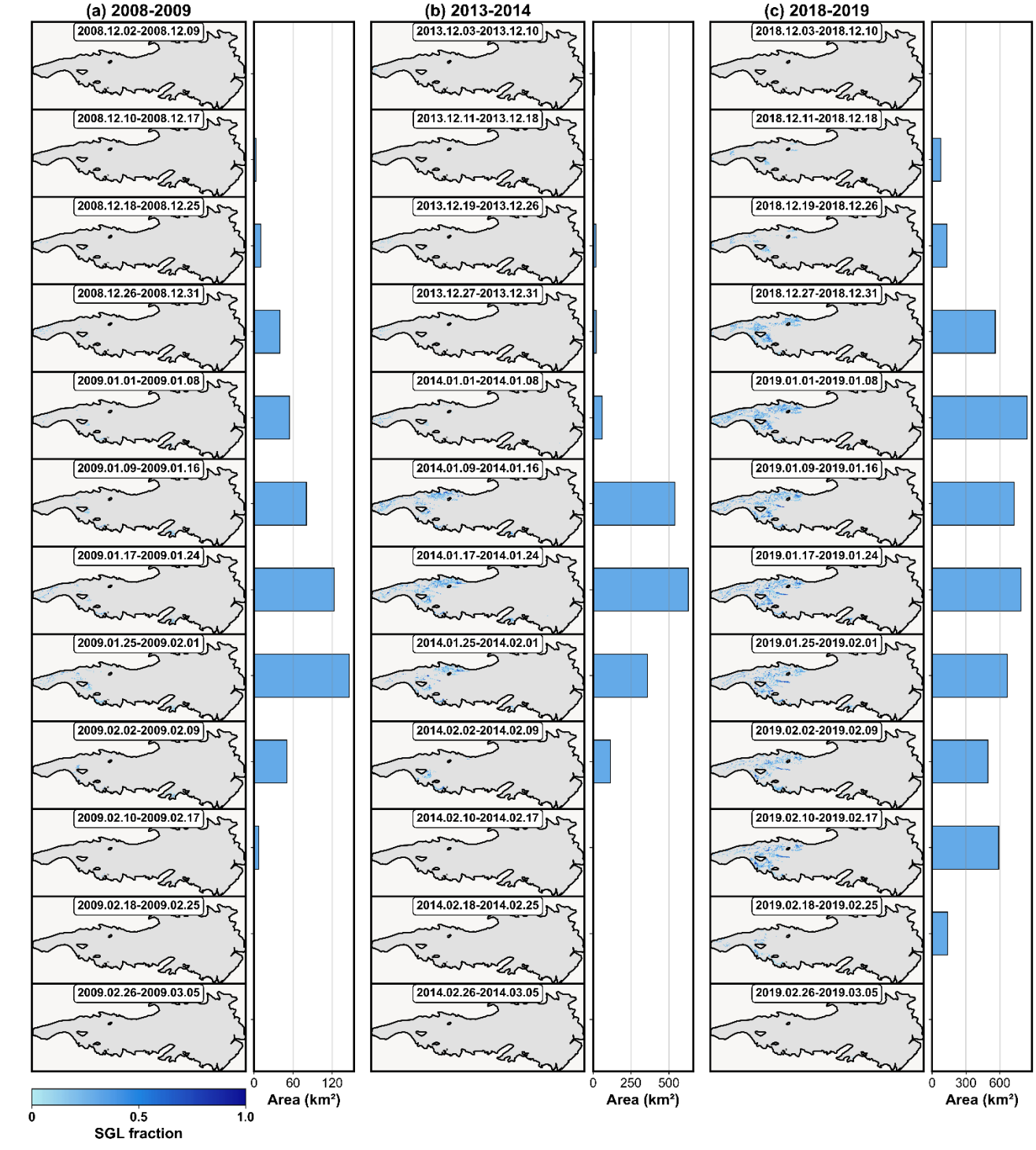


Figure 6: Evolution of SGLs on the Amery Ice Shelf from early December to early March: (a) 2008–2009; (b) 2013–2014; (c) 2018–2019.

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

Based on the 23-year 8-day SGL fraction dataset, SGL persistence is defined as the ratio of lake-detected observations to the total number of observations within each melt season (early November to early April, approximately 20 observations). The multi-year mean persistence is calculated by averaging these seasonal values over the 23-year study period (Fig. 7). Overall, Antarctic SGLs exhibit low persistence (Fig. 7). The mean persistence rate across all lakes is 2.5%. Although the maximum persistence reaches 30%, lakes with such high persistence are extremely rare and account for less than 0.01% of all lakes. In contrast, lakes with persistence rates below 5% dominate the distribution, representing approximately 83% of all lakes. These persistence statistics demonstrate that Antarctic SGLs are predominantly characterized by intermittent occurrence and low persistence.

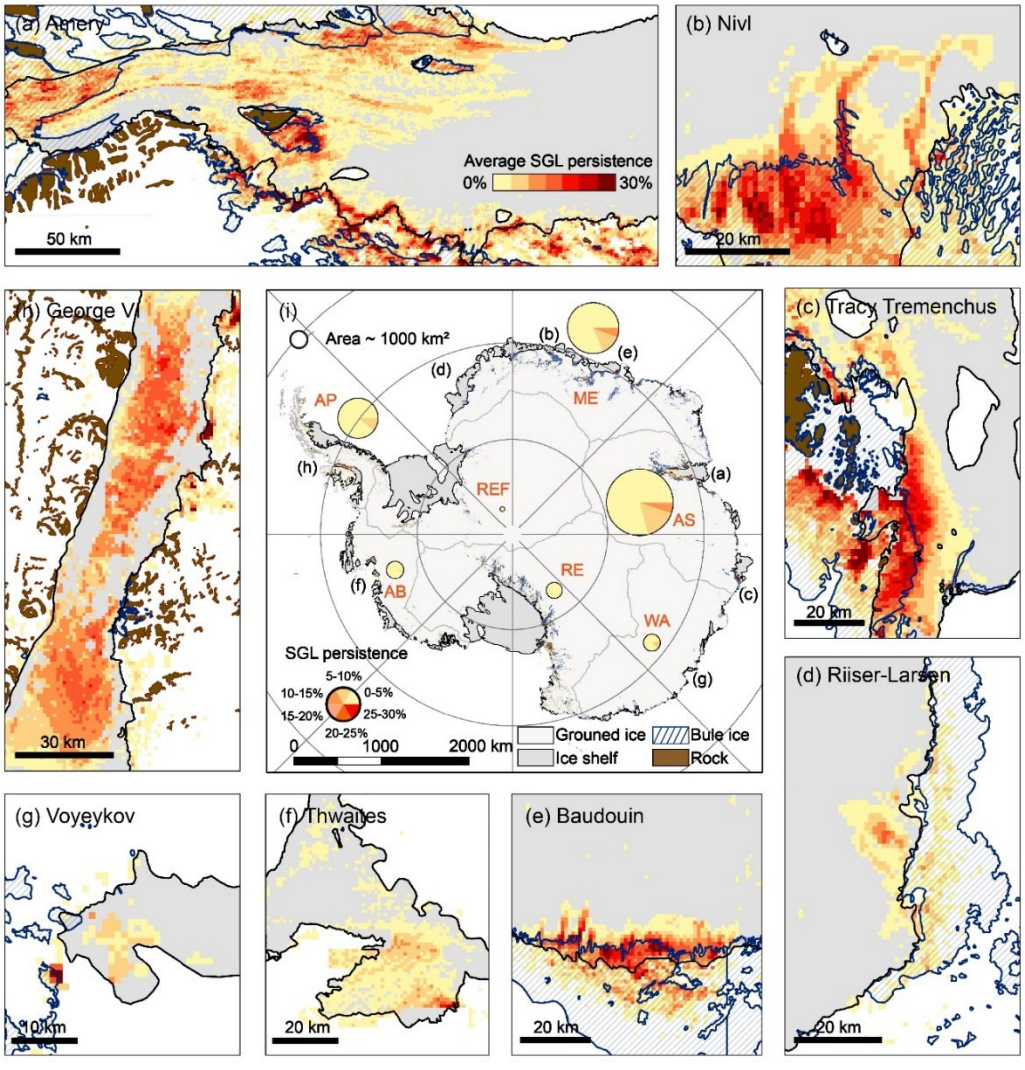


Figure 7: Multi-year mean persistence of SGLs across Antarctica (2000–2023). (a) Amery Ice Shelf; (b) Nivl Ice Shelf; (c) Tracy Tremenchus Ice Shelf; (d) Riiser-Larsen Ice Shelf; (e) Baudouin Ice Shelf; (f) Thwaites Ice Shelf; (g) Voyeykov

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

Ice Shelf; (h) George VI Ice Shelf; (i) Regional summary of multi-year mean SGL persistence. The radius of each pie chart represents the multi-year mean maximum SGL area of the corresponding region.

Regionally, central and southwestern Antarctica including the Ross Embayment (RE), Ronne Embayment (REF), and the Amundsen–Bellingshausen (AB) sectors, exhibit relatively sparse SGL occurrence, with nearly all lakes showing persistence rates below 5% (Fig. 7i). In contrast, the Antarctic Peninsula shows higher persistence, with more than 10% of lakes exceeding 5%. In East Antarctica, including the Amery–Shackleton (AS), Maud–Enderby (ME), and Wilkes–Adélie (WA) sectors, nearly 20% of lakes have persistence rates above 5%, and about 5% exceed 10% in the ME and AS sectors (Fig. 7i). Nevertheless, low-persistence lakes (< 5%) account for 81%–86% of the total and remain the dominant persistence class.

Spatial patterns of SGL persistence across representative ice shelves (Fig. 7a–h) reveal strong heterogeneity in both magnitude and spatial organization. On the Amery, Nivl, Tracy Tremenchus, and Baudouin Ice Shelves, lakes with persistence rates exceeding 10% form narrow, elongated bands aligned with grounding lines and adjacent blue-ice regions. In contrast, ice shelves such as Riiser-Larsen and Voyeykov exhibit low persistence rates (1%) and sparse spatial distributions. On the Antarctic Peninsula, lakes on the George VI Ice Shelf show relatively high persistence (4%) and extensive spatial coverage, whereas the Thwaites Ice Shelf in West Antarctica exhibits lower persistence (1.5%) and reduced spatial continuity.

Overall, this long-term dataset, generated at an 8-day temporal resolution, provides a more detailed characterization of Antarctic SGL persistence and highlights pronounced regional differences that are less apparent in coarser monthly-scale observations. East Antarctica and the Antarctic Peninsula exhibit more coherent persistence patterns, whereas West Antarctica shows greater spatial variability.

4.3 Spatial distribution of Antarctic SGLs

Based on the 23-year dataset, the multi-year mean maximum SGL fraction across Antarctica is presented (Fig. 8). It is defined as the mean of the annual maximum SGL fraction over the 23-year period. Estimates of the average maximum SGL area for the continent and its subregions are also provided. Overall, SGLs are primarily distributed near grounding lines of ice shelves, particularly in regions associated with blue-ice and bare rock (Fig. 8). Quantitatively, within 10 km of the grounding line, SGLs occupy an area of 2,683 km^2, accounting for 65% of the multi-year mean maximum Antarctic SGL area (4,103 km^2, Fig. S2). Most SGLs occur on ice shelves, which account for 60% of the total SGL area, approximately 1.5 times that on grounded ice (Fig. S2). In addition, 85% of Antarctic SGLs are located within 10 km of blue-ice and bare-rock regions, and nearly all SGLs (99%) occur at elevations below 1 km (Fig. S2).

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

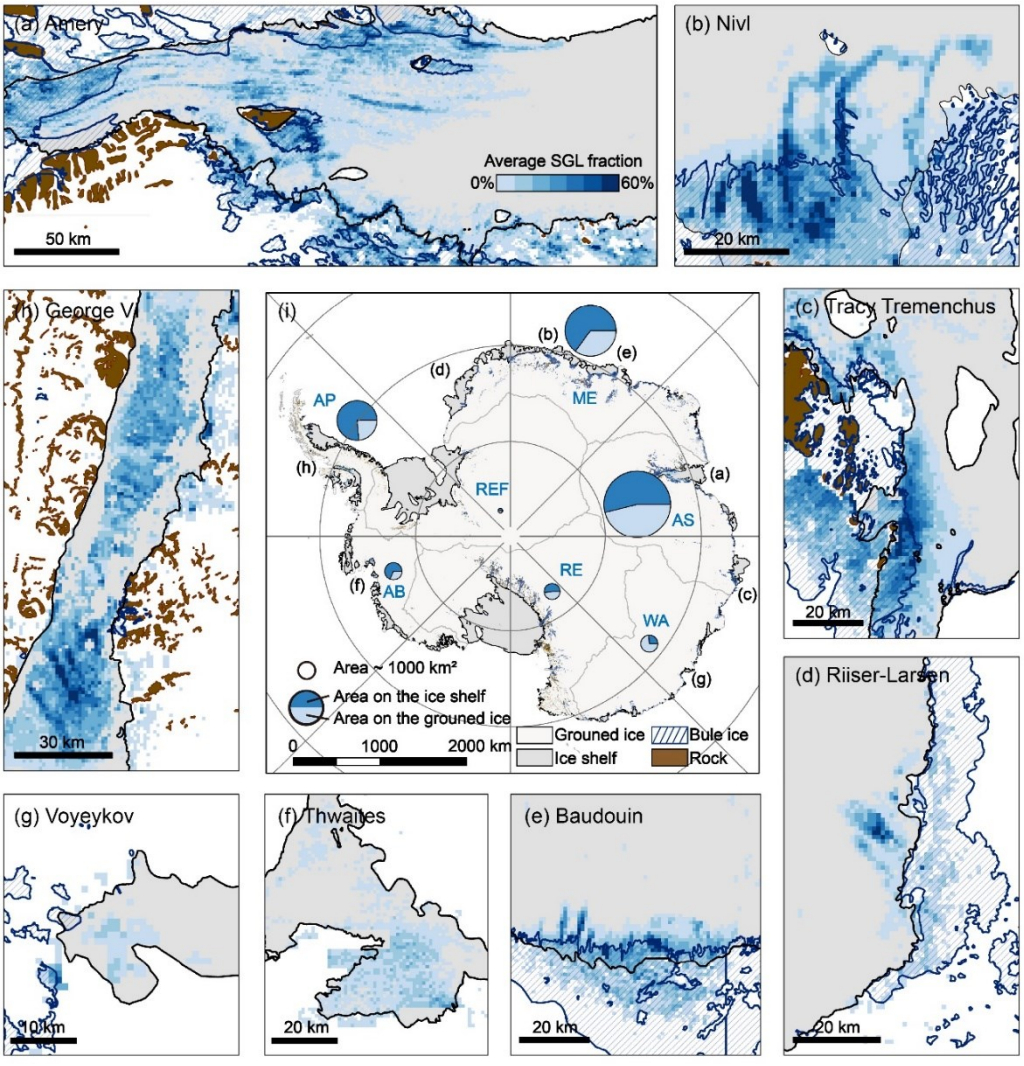


Figure 8: Mean annual SGL fraction across Antarctica (2000–2023). (a) Amery Ice Shelf; (b) Nivl Ice Shelf; (c) Tracy Tremenchus Ice Shelf; (d) Riiser-Larsen Ice Shelf; (e) Baudouin Ice Shelf; (f) Thwaites Ice Shelf; (g) Voyeykov Ice Shelf; (h) George VI Ice Shelf; (i) Regional statistics of multi-year mean maximum SGL area. Each pie chart radius represents the multi-year mean maximum SGL area.

Regionally, East Antarctica contributes the majority of SGL coverage, with an area of approximately 3,285 km^2, far exceeding the values for West Antarctica (130 km^2) as well as the Antarctic Peninsula (688 km^2). Within East Antarctica, the AS, ME, and WA sectors together account for more than 77% of the total Antarctic lake area, contributing 47%, 27%, and 3%, respectively. In addition, the Antarctic Peninsula accounts for 17% of the total Antarctic SGL area. In contrast, the AB sector in West Antarctica, together with the RE and REF regions spanning East and West Antarctica, collectively contribute less than10% of the Antarctic SGL area.

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

Spatial patterns of SGL fraction across representative ice shelves reveal pronounced variability in spatial extent (Fig. 8). On the Antarctic Peninsula, the George VI Ice Shelf exhibits relatively high lake fractions with extensive coverage, forming a dense and heterogeneous spatial distribution of SGLs across the central ice shelf. In West Antarctica, the Thwaites Ice Shelf is characterized by low lake fractions and a fragmented spatial pattern, with SGLs occurring mainly as small and isolated patches. In contrast, several ice shelves in East Antarctica exhibit substantially higher lake fractions and more coherent spatial organization. High-fraction regions on the Amery, Nivl, Tracy Tremenchus, and Baudouin Ice Shelves form broad and continuous zones with diverse spatial structures.

Overall, the dataset provides a continent-wide characterization of Antarctic SGL distribution and demonstrates the capability to resolve complex spatial patterns of SGLs. East Antarctica contains the most extensive and spatially coherent SGL distributions, while the Antarctic Peninsula exhibits locally enhanced lake coverage, and West Antarctica is characterized by limited and fragmented SGL occurrence.

4.4 Interannual and seasonal changes in Antarctic SGL area

The 23-year SGL dataset shows pronounced interannual variability, clear regional contrasts, and consistent seasonal cycles in the Antarctic SGLs area. Figure 9 illustrates the temporal evolution of the SGL area across Antarctica and its major subregions from 2000 to 2023. Antarctic SGL area exhibits a distinct annual peak during the austral summer, with a mean peak area of approximately 1,286 km^2 (Fig. 9a). Substantial interannual variability is also observed. Peak lake areas in high-area years (mean = 1,774 km^2) are more than twice those in low-area years (mean = 798 km^2). This variability corresponds to a coefficient of variation (CV) of 0.48, defined as the ratio of the standard deviation to the mean.

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

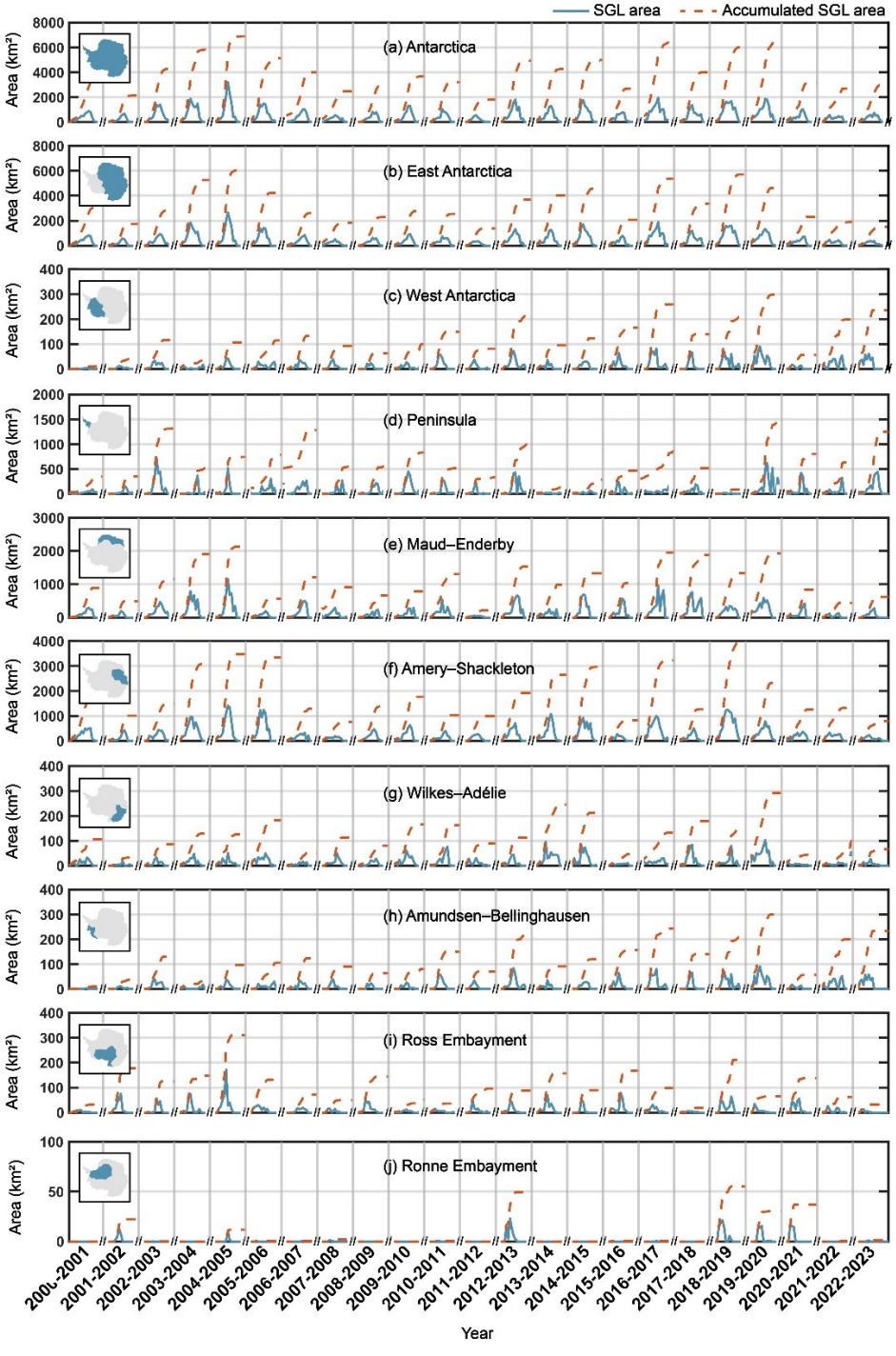


Figure 9: Changes in SGL area across Antarctica and its subregions, 2000–2023. (a) Antarctica; (b) East Antarctica; (c) West Antarctica; (d) Antarctic Peninsula; (e) Maud–Enderby region; (f) Amery–Shackleton region; (g) Wilkes–Adélie region; (h) Amundsen–Bellingshausen region; (i) Ross Embayment; (j) Ronne Embayment.

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

East Antarctica is consistently the dominant contributor to Antarctic SGL area, with mean peak area and mean annual accumulated area exceeding 1,100 km^2 and 3,200 km^2, respectively (Fig. 9b). Within East Antarctica, the AS and ME subregions exhibit the largest SGL extents, with mean peak areas of around 500 km^2, exceeding the combined total of West Antarctica and the Antarctic Peninsula (Fig. 9c, d). Compared with East Antarctica, the Antarctic Peninsula shows a mean peak area of approximately one-third of that in East Antarctica, while peak lake area in West Antarctica is generally below 100 km^2 (Fig. 9d). Subregions spanning East and West Antarctica, including RE and REF, contribute relatively small lake areas, with peak extents rarely exceeding 100 km^2 (Fig. 9i, j).

Across all regions, the dataset captures a pronounced seasonal cycle in the SGL area, with a progressive increase toward a seasonal maximum followed by a rapid decline (Fig. 10).

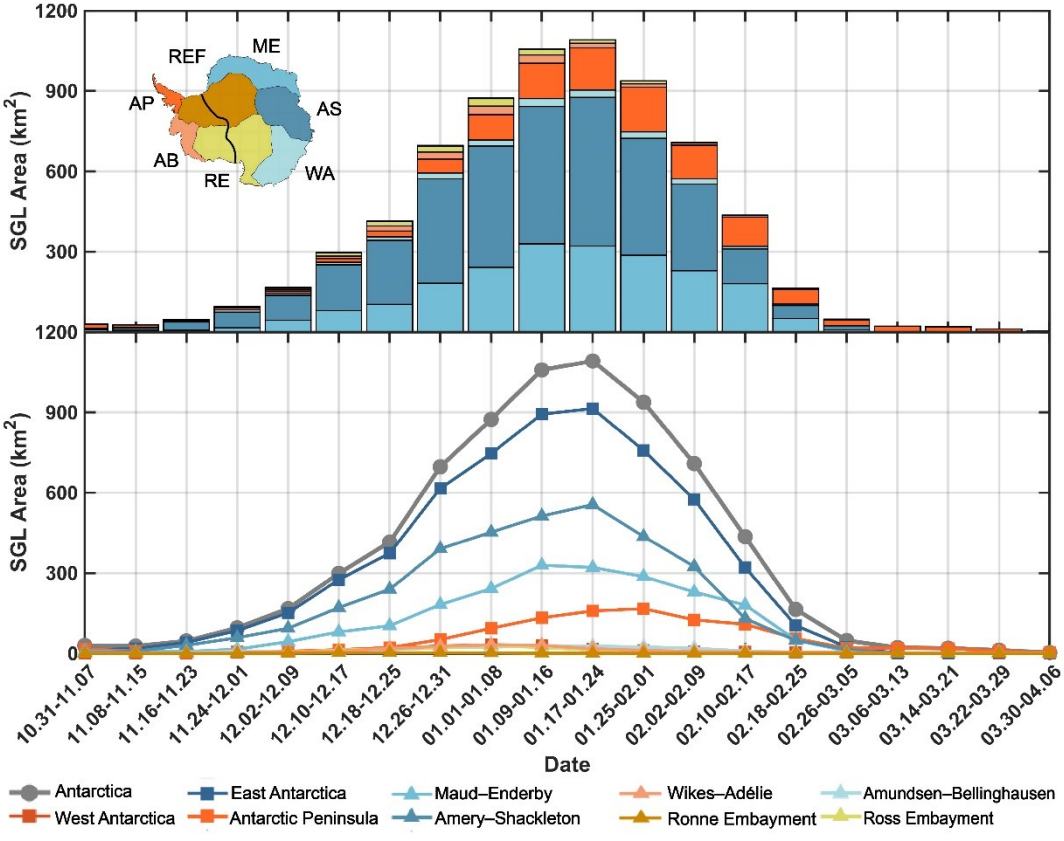


Figure 10: Seasonal dynamics of SGL area in Antarctica and regional sectors. The values represent the 23-year mean SGL area for each observation period. The bar charts show the proportional contribution of the seven subregions to the total SGL area in each period.

At the continental scale, SGL area is minimal during 31 October–1 December, with only limited SGL coverage observed during this period. A gradual increase occurs during 24 November–9 December, followed by a rapid increase in SGL area from 10 December to 16 January. The Antarctic-wide peak occurs between 17 and 24 January. After this peak, SGL area

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

declines rapidly and approaches near-zero levels by 6 March–6 April. Peak SGL extent shows strong regional variability in timing. In West Antarctica, SGL area reaches its maximum earlier, typically during 1–8 January. In contrast, East Antarctica exhibits a later peak, with the ME and WA regions reaching their maxima during 9–16 January, followed by the AS region during 17–24 January. The Antarctic Peninsula shows the latest peak timing, with maximum lake extent generally occurring during 25 January–1 February.

Regional contributions vary systematically over the season. During the early melt season (31 October–9 December), SGLs are primarily concentrated in East Antarctica, particularly in the AS and ME regions (39.80%–82.68%). During the period of rapid SGL area increase (10 December–16 January), East Antarctica remains dominant, with the AS region contributing the largest share (48.50%–57.54%) at the seasonal peak. Following the peak, the SGL area decreases across all regions. During the mid-to-late observation period (17 January–5 March), the Antarctic Peninsula maintains a relatively high proportional contribution (17.67%–41.47%), coinciding with the period when maximum SGL extent is observed in this region. In contrast, East Antarctica shows a more rapid decline after reaching its maximum extent, while West Antarctica remains characterized by consistently low lake coverage throughout the season.

Overall, the long-term, continent-wide dataset reveals pronounced interannual variability and persistent seasonal cycles in Antarctic SGL dynamics. Its 8-day temporal resolution further captures fine-scale differences in the timing and magnitude of peak SGL area, which are not readily resolved by observations with coarser temporal resolution.

4.5 Uncertainty analysis

Despite the overall good retrieval performance, uncertainties remain and should be considered when interpreting the results. These uncertainties primarily arise from limitations in optical observations, ground-truth errors, structural assumptions of the model, and the inherently complex environmental conditions of Antarctica.

First, the inherent limitations of optical remote sensing introduce observational uncertainties. Clouds and cloud shadows frequently obscure the ice-sheet surface, potentially leading to underestimation or temporal gaps in SGL fraction retrievals (Arthur et al., 2020; Husman et al., 2023; Moussavi et al., 2020). To mitigate this issue, the 8-day composite MOD09A1 product is employed, as it preferentially retains higher-quality observations (Xiang, 2013). However, residual cloud contamination may still persist even in the composited product, and its impact is difficult to quantify, introducing additional uncertainty into the retrievals.

Second, uncertainties in the reference data propagate into the modeling process. The lake fraction data derived from Sentinel-2 imagery for model training inevitably contain classification errors (Elmes et al., 2020; Pelletier et al., 2017). To reduce this impact, only samples with classification accuracies exceeding 95% are retained, and all results are further verified through manual visual inspection. Under these strict criteria, errors introduced by the ground-truth data are considered negligible at the 1 km scale.

Third, structural uncertainty arises from the two-stage framework adopted in this study, in which lake classification is followed by fraction regression. Errors introduced during the initial lake extraction stage may propagate to the subsequent fraction

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

estimation. To alleviate this effect, a hybrid strategy combining threshold-based constraints and machine learning is applied. This approach reduces misclassification under relatively relaxed lake identification criteria.

Finally, additional uncertainty arises from the complex and heterogeneous environmental conditions across Antarctica, which result in substantial variability in surface properties and observation conditions. As a result, applying a single model across the entire continent cannot fully capture all regional differences, leading to unavoidable performance limitations. This pronounced spatiotemporal heterogeneity further results in discrepancies between validation and independent testing results, with OA decreasing to 0.86 and r to 0.76. Nevertheless, the overall performance remains within an acceptable range and is sufficient to support large-scale analyses of Antarctic SGL dynamics.

Future improvements to the dataset will focus on reducing cloud-related data gaps and enhancing model generalization. This may be achieved by integrating higher-frequency observations or complementary data sources, as well as developing region-specific models. These improvements will further enhance the representation of short-lived SGL events and enable more temporally continuous estimation of lake fractions.

5 Data and code availability

The Antarctic SGL fraction dataset and all source code developed in this study are publicly available through Zenodo (https://doi.org/10.5281/zenodo.19936100; Wei et al., 2026). The repository provides permanent access to both the dataset and the scripts required to reproduce the processing workflow described in this paper.

The dataset contains Antarctic SGL fraction maps for melt seasons from 2000 to 2023 (early November to early April), covering the entire Antarctic Ice Sheet at a spatial resolution of 1 km and a temporal resolution of 8 days. Two product levels are provided: Level 0 and Level 1. Level 0 represents the direct output of the SGL fraction retrieval model. Level 1 products are generated by applying additional physical constraints based on melt flux observations and historical SGL occurrence information, thereby improving physical consistency.

All products are distributed in GeoTIFF format as floating-point raster datasets, where each pixel value represents the estimated fraction of supraglacial lake coverage within the corresponding 1 km grid cell. The data are provided in the Antarctic Polar Stereographic projection (EPSG:3031; WGS 84 datum). File names follow a standardized convention: Level 0 products are named *LF_yyyymmdd.tif*, whereas Level 1 products are named *LF_yyyymmdd_L1.tif*, where *yyyy*, *mm*, and *dd* indicate the acquisition date of the corresponding 8-day composite period.

The code package consists of two GEE scripts written in JavaScript. The first script generates the training and validation datasets used for SGL classification and fraction retrieval. For a selected observation period, the script produces an eight-band GeoTIFF image, where Bands 1–7 correspond to MODIS surface reflectance and Band 8 contains the reference SGL fraction derived from Sentinel-2 imagery. The second script implements the complete Antarctic SGL fraction retrieval workflow, including candidate lake identification, random forest classification, random forest regression, and final fraction generation.

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

Together, these scripts enable full reproduction of the dataset generation process. Users can apply the workflow to any available MODIS observation period to generate Antarctic-wide SGL fraction maps.

6 Conclusions

The dataset is generated by integrating high-temporal-resolution MODIS imagery with high-spatial-resolution Sentinel-2 data within a machine-learning framework and validated against Sentinel-2 observations, showing good agreement ($r = 0.85$). Analysis of the dataset shows that Antarctic SGLs are highly dynamic, with approximately 83% of lakes exhibiting persistence rates below 5%, indicating that most lakes are short-lived. Spatially, approximately 65% of Antarctic SGLs are located within 10 km of grounding lines and are closely associated with blue-ice and exposed rock areas, with more than 85% of lakes occurring within 10 km of these features. East Antarctica dominates total SGL coverage, contributing approximately 80% of the mean maximum SGL area, particularly in the AS and ME sectors, whereas West Antarctica shows sparse and fragmented lake development. Over the 23-year period, Antarctic SGL area shows strong interannual variability ($CV = 0.48$) while maintaining a consistent seasonal cycle characterized by rapid expansion in early austral summer, a peak in January, and subsequent decline. The dataset further reveals clear spatial heterogeneity in peak timing, with SGL extent peaking approximately one week earlier in West Antarctica than in East Antarctica and about two weeks earlier than on the Antarctic Peninsula.

Overall, this dataset provides a high-frequency, sub-pixel-scale, and continent-wide record of Antarctic SGL dynamics, substantially improving the capability to capture short-lived and rapidly evolving meltwater processes compared with existing products. By integrating long-term MODIS observations with high-resolution Sentinel-2 data within a consistent and automated framework, the dataset enables continuous monitoring and ensures temporal consistency across the full record. It bridges the gap between spatial detail and temporal continuity, offering new opportunities to investigate SGL variability across multiple timescales and regions. Resolving both the rapid changes and spatial heterogeneity of supraglacial lakes, it provides a robust foundation for advancing the understanding of Antarctic meltwater processes and their implications for ice-sheet stability, surface energy balance, and future sea-level change.

Author contributions

SW: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing.

LZ: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation.

QL: Formal analysis, Visualization.

TL: Data curation, Software.

XC: Funding acquisition, Resources, Supervision.

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

Open Access Earth System Science Data Discussions

Competing interests.

The contact author has declared that none of the authors has any competing interests.

Acknowledgements

We gratefully acknowledge NASA for providing the MODIS data and the European Space Agency (ESA) Copernicus Programme for providing the Sentinel-2 imagery used in this study. These openly accessible datasets provide essential support for Antarctic remote sensing research and long-term environmental monitoring.

References

Ackerman, S. A., Strabala, K. I., Menzel, W. P., Frey, R. A., Moeller, C. C., and Gumley, L. E.: Discriminating clear sky from clouds with MODIS, Journal of Geophysical Research: Atmospheres, 103, 32141–32157, https://doi.org/10.1029/1998JD200032, 1998.

Ackerman, S. A., Holz, R. E., Frey, R., Eloranta, E. W., Maddux, B. C., and McGill, M.: Cloud Detection with MODIS. Part II: Validation, Journal of Atmospheric and Oceanic Technology, 25, 1073–1086, https://doi.org/10.1175/2007JTECHA1053.1, 2008.

Arthur, J. F., Stokes, C., Jamieson, S. S., Carr, J. R., and Leeson, A. A.: Recent understanding of Antarctic supraglacial lakes using satellite remote sensing, Progress in Physical Geography: Earth and Environment, 44, 837–869, https://doi.org/10.1177/0309133320916114, 2020.

Banwell, A. F., MacAyeal, D. R., and Sergienko, O. V.: Breakup of the Larsen B Ice Shelf triggered by chain reaction drainage of supraglacial lakes, Geophysical Research Letters, 40, 5872–5876, https://doi.org/10.1002/2013GL057694, 2013.

Bartholomew, I., Nienow, P., Mair, D., Hubbard, A., King, M. A., and Sole, A.: Seasonal evolution of subglacial drainage and acceleration in a Greenland outlet glacier, Nature Geoscience, 3, 408–411, https://doi.org/10.1038/ngeo863, 2010.

Breiman, L.: Random forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Corr, D., Leeson, A., McMillan, M., Zhang, C., and Barnes, T.: An inventory of supraglacial lakes and channels across the West Antarctic Ice Sheet, Earth System Science Data, 14, 209–228, https://doi.org/10.5194/essd-14-209-2022, 2022.

Das, S. B., Joughin, I., Behn, M. D., Howat, I. M., King, M. A., Lizarralde, D., and Bhatia, M. P.: Fracture propagation to the base of the Greenland Ice Sheet during supraglacial lake drainage, Science, 320, 778–781, https://doi.org/10.1126/science.1153360, 2008.

Dell, R.L., Arnold, N. S., Willis, I. C., Banwell, A. F., Williamson, A., Pritchard, H., and Orr, A.: Lateral meltwater transfer across an Antarctic ice shelf, The Cryosphere, 14, 2313–2330, https://doi.org/10.5194/tc-14-2313-2020, 2020.

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., and Bargellini, P.: Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services, Remote Sensing of Environment, 120, 25–36, https://doi.org/10.1016/j.rse.2011.11.026, 2012.

Echelmeyer, K., Clarke, T., and Harrison, W. D.: Surficial glaciology of Jakobshavns Isbræ, West Greenland: Part I. Surface morphology, Journal of Glaciology, 37, 368–382, https://doi.org/10.3189/S0022143000005803, 1991.

Elmes, A., Alemohammad, H., Avery, R., Caylor, K., Eastman, J., Fishgold, L., Friedl, M., Jain, M., Kohli, D., Laso Bayas, J., Lunga, D., McCarty, J., Pontius, R., Reinmann, A., Rogan, J., Song, L., Stoynova, H., Ye, S., Yi, Z.-F., and Estes, L.: Accounting for Training Data Error in Machine Learning Applied to Earth Observations, Remote Sensing, 12, 1034, https://doi.org/10.3390/rs12061034, 2020.

Feyisa, G. L., Meilby, H., Fensholt, R., and Proud, S. R.: Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery, Remote Sensing of Environment, 140, 23–35, https://doi.org/10.1016/j.rse.2013.08.029, 2014.

Foody, G. M.: Status of land cover classification accuracy assessment, Remote Sensing of Environment, 80, 185–201, https://doi.org/10.1016/S0034-4257(01)00295-4, 2002.

Hubbard, B., Luckman, A., Ashmore, D. W., Bevan, S., Kulessa, B., Kuipers Munneke, P., Philippe, M., Jansen, D., Booth, A., Sevestre, H., Tison, J.-L., O'Leary, M., and Rutt, I.: Massive subsurface ice formed by refreezing of ice-shelf melt ponds, Nature Communications, 7, 11897, https://doi.org/10.1038/ncomms11897, 2016.

Husman, S. D. R., Hu, Z., Wouters, B., Munneke, P. K., Veldhuijsen, S., and Lhermitte, S.: Remote Sensing of Surface Melt on Antarctica: Opportunities and Challenges, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16, 2462–2480, https://doi.org/10.1109/JSTARS.2022.3216953, 2023.

Jiang, D., Li, X., Zhang, K., Marinsek, S., Hong, W., and Wu, Y.: Automatic Supraglacial Lake Extraction in Greenland Using Sentinel-1 SAR Images and Attention-Based U-Net, Remote Sensing, 14, 4998, https://doi.org/10.3390/rs14194998, 2022.

Justice, C. O., Townshend, J. R. G., Vermote, E. F., Masuoka, E., Wolfe, R. E., Saleous, N., Roy, D. P., and Morisette, J. T.: An overview of MODIS Land data processing and product status, Remote Sensing of Environment, 83, 3–15, https://doi.org/10.1016/S0034-4257(02)00084-6, 2002.

Kingslake, J., Ng, F., and Sole, A.: Modelling channelized surface drainage of supraglacial lakes, Journal of Glaciology, 61, 185–199, https://doi.org/10.3189/2015JoG14J158, 2015.

Kingslake, J., Ely, J. C., Das, I., and Bell, R. E.: Widespread movement of meltwater onto and across Antarctic ice shelves, Nature, 544, 349–352, https://doi.org/10.1038/nature22049, 2017.

Leeson, A. A., Shepherd, A., Palmer, S., Sundal, A., and Fettweis, X.: Simulating the growth of supraglacial lakes at the western margin of the Greenland ice sheet, The Cryosphere, 6, 1077–1086, https://doi.org/10.5194/tc-6-1077-2012, 2012.

Leeson, A. A., Shepherd, A., Sundal, A. V., Malin Johansson, A., Selmes, N., Briggs, K., Hogg, A. E., and Fettweis, X.: A comparison of supraglacial lake observations derived from MODIS imagery at the western margin of the Greenland ice sheet, Journal of Glaciology, 59, 1179–1188, https://doi.org/10.3189/2013JoG13J064, 2013.

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

Legates, D. R. and McCabe, G. J.: Evaluating the use of "goodness‐of‐fit" measures in hydrologic and hydroclimatic model validation, Water Resources Research, 35, 233–241, https://doi.org/10.1029/1998WR900018, 1999.

Lenaerts, J. T. M., Lhermitte, S., Drews, R., Ligtenberg, S. R. M., Berger, S., Helm, V., Smeets, C. J. P. P., van den Broeke, M. R., van de Berg, W. J., van Meijgaard, E., Eijkelboom, M., Eisen, O., and Pattyn, F.: Meltwater produced by wind–albedo interaction stored in an East Antarctic ice shelf, Nature Climate Change, 7, 58–62, https://doi.org/10.1038/nclimate3180, 2017.

Liang, J. and Liu, D.: Automated estimation of daily surface water fraction from MODIS and Landsat images using Gaussian process regression, International Journal of Remote Sensing, 42, 4261–4283, https://doi.org/10.1080/01431161.2021.1892859, 2021.

Lüthje, M., Feltham, D. L., Taylor, P. D., and Worster, M. G.: Modeling the summertime evolution of sea‐ice melt ponds, Journal of Geophysical Research: Oceans, 111, 2004JC002818, https://doi.org/10.1029/2004JC002818, 2006.

Markham, K., Frazier, A. E., Singh, K. K., and Madden, M.: A review of methods for scaling remotely sensed data for spatial pattern analysis, Landscape Ecology, 38, 619–635, https://doi.org/10.1007/s10980-022-01449-1, 2023.

Mätzler, C.: Microwave permittivity of dry snow, IEEE Transactions on Geoscience and Remote Sensing, 34, 573–581, https://doi.org/10.1109/36.485127, 1996.

McFeeters, S. K.: The use of the normalized difference water index (NDWI) in the delineation of open water features, International Journal of Remote Sensing, 17, 1425–1432, https://doi.org/10.1080/01431169608948714, 1996.

McKenna, O. P., Lothspeich, A. C., Vacek, S., MacDonald, D., Eash, J. D., Vanderhoof, M. K., McCulloch, E. C., Ross, C. D., Sabrina, S., and Knight, J. F.: Small waterbodies of large conservation concern: Towards an integrated approach to more accurately measuring surface water dynamics, Ecological Indicators, 175, 113525, https://doi.org/10.1016/j.ecolind.2025.113525, 2025.

McMillan, M., Nienow, P., Shepherd, A., Benham, T., and Sole, A.: Seasonal evolution of supra-glacial lakes on the Greenland Ice Sheet, Earth and Planetary Science Letters, 262, 484–492, https://doi.org/10.1016/j.epsl.2007.08.002, 2007.

Mouginot, J., Rignot, E., Bjørk, A. A., Van Den Broeke, M., Millan, R., Morlighem, M., Noël, B., Scheuchl, B., and Wood, M.: Forty-six years of Greenland Ice Sheet mass balance from 1972 to 2018, Proceedings of the National Academy of Sciences of the United States of America, 116, 9239–9244, https://doi.org/10.1073/pnas.1904242116, 2019.

Mountrakis, G., Im, J., and Ogole, C.: Support vector machines in remote sensing: A review, ISPRS Journal of Photogrammetry and Remote Sensing, 66, 247–259, https://doi.org/10.1016/j.isprsjprs.2010.11.001, 2011.

Moussavi, M., Pope, A., Halberstadt, A., Trusel, L., Cioffi, L., and Abdalati, W.: Antarctic Supraglacial Lake Detection Using Landsat 8 and Sentinel-2 Imagery: Towards Continental Generation of Lake Volumes, Remote Sensing, 12, 134, https://doi.org/10.3390/rs12010134, 2020.

Niu, L., Tang, X., Yang, S., Zhang, Y., Zheng, L., and Wang, L.: Detection of Antarctic Surface Meltwater Using Sentinel-2 Remote Sensing Images via U-Net With Attention Blocks: A Case Study Over the Amery Ice Shelf, IEEE Transactions on Geoscience and Remote Sensing, 61, 1–13, https://doi.org/10.1109/TGRS.2023.3275076, 2023.

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

Pelletier, C., Valero, S., Inglada, J., Champion, N., Marais Sicre, C., and Dedieu, G.: Effect of Training Class Label Noise on Classification Performances for Land Cover Mapping with Satellite Image Time Series, Remote Sensing, 9, 173, https://doi.org/10.3390/rs9020173, 2017.

Rignot, E., Mouginot, J., Scheuchl, B., Van Den Broeke, M., Van Wessem, M. J., and Morlighem, M.: Four decades of Antarctic Ice Sheet mass balance from 1979–2017, Proceedings of the National Academy of Sciences of the United States of America, 116, 1095–1103, https://doi.org/10.1073/pnas.1812883116, 2019.

Safarov, M., Kang, S., Fazylov, A., Gulayozov, M., Banerjee, A., Navruzshoev, H., Chen, P., Xue, Y., and Murodov, M.: Estimating glacier dynamics and supraglacial lakes together with associated regional hazards using high-resolution datasets in Pamir, Journal of Mountain Science, 21, 3767–3788, https://doi.org/10.1007/s11629-024-8936-x, 2024.

Scambos, T. A., Hulbe, C., Fahnestock, M., and Bohlander, J.: The link between climate warming and break-up of ice shelves in the Antarctic Peninsula, Journal of Glaciology, 46, 516–530, https://doi.org/10.3189/172756500781833043, 2000.

Selmes, N., Murray, T., and James, T. D.: Fast draining lakes on the Greenland Ice Sheet, Geophysical Research Letters, 38, L15501, https://doi.org/10.1029/2011GL047872, 2011.

Shepherd, A., Hubbard, A., Nienow, P., King, M., McMillan, M., and Joughin, I.: Greenland ice sheet motion coupled with daily melting in late summer, Geophysical Research Letters, 36, L01501, https://doi.org/10.1029/2008GL035758, 2009.

Shepherd, A., Ivins, E. R., Gardner, A. S., Barletta, V. R., Bentley, M. J., Bettadpur, S., Briggs, K. H., Bromwich, D. H., Forsberg, R., Galin, N., Horwath, M., Jacobs, S., Joughin, I., King, M. A., Lenaerts, J. T. M., Li, J., Ligtenberg, S. R. M., Luckman, A., Luthcke, S. B., McMillan, M., Meister, R., Milne, G. A., Mouginot, J., Muir, A., Nicolas, J. P., Paden, J., Payne, A. J., Pritchard, H., Rignot, E., Rott, H., Sørensen, L. S., Scambos, T. A., Scheuchl, B., Schrama, E. J. O., Smith, B., Sundal, A. V., van Angelen, J. H., van den Broeke, M. R., Vaughan, D. G., Velicogna, I., Wahr, J., Whitehouse, P. L., Wingham, D. J., Yi, D., Young, D., and Zwally, H. J.: A reconciled estimate of ice-sheet mass balance, Science, 338, 1183–1189, https://doi.org/10.1126/science.1228102, 2012.

Shu, Q., Killick, R., Leeson, A., Nemeth, C., Fettweis, X., Hogg, A., and Leslie, D.: Characterising the ice sheet surface in Northeast Greenland using Sentinel-1 SAR data, Journal of Glaciology, 69, 1834–1845, https://doi.org/10.1017/jog.2023.64, 2023.

Stokes, C. R., Sanderson, J. E., Miles, B. W. J., Jamieson, S. S. R., and Leeson, A. A.: Widespread distribution of supraglacial lakes around the margin of the East Antarctic Ice Sheet, Scientific Reports, 9, 13823, https://doi.org/10.1038/s41598-019-50343-5, 2019.

Tedesco, M., Lüthje, M., Steffen, K., Steiner, N., Fettweis, X., Willis, I., Bayou, N., and Banwell, A.: Measurement and modeling of ablation of the bottom of supraglacial lakes in western Greenland, Geophysical Research Letters, 39, 2011GL049882, https://doi.org/10.1029/2011GL049882, 2012.

Tedesco, M., Willis, I. C., Hoffman, M. J., Banwell, A. F., Alexander, P., and Arnold, N. S.: Ice dynamic response to two modes of surface lake drainage on the Greenland ice sheet, Environ. Res. Lett., 8, L02502, https://doi.org/10.1088/1748-9326/8/3/034007, 2013.

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

The IMBIE Team: Mass balance of the Antarctic Ice Sheet from 1992 to 2017, Nature, 558, 219–222, https://doi.org/10.1038/s41586-018-0179-y, 2018.

Trusel, L. D., Frey, K. E., Das, S. B., Karnauskas, K. B., Kuipers Munneke, P., van Meijgaard, E., and van den Broeke, M. R.: Divergent trajectories of Antarctic surface melt under two twenty-first-century climate scenarios, Nature Geoscience, 8, 927–932, https://doi.org/10.1038/ngeo2563, 2015.

Tuckett, P. A., Sole, A. J., Livingstone, S. J., Jones, J. M., Lea, J. M., and Gilbert, E.: Continent-wide mapping shows increasing sensitivity of East Antarctica to meltwater ponding, Nature Climate Change, 15, 775–783, https://doi.org/10.1038/s41558-025-02363-5, 2025.

Ulaby, F. T., Moore, R. K., and Fung, A. K.: Microwave remote sensing: Active and passive, Volume 1: Microwave remote sensing fundamentals and radiometry, Artech House, Norwood, MA, 1981.

van den Broeke, M. R., Enderlin, E. M., Howat, I. M., Kuipers Munneke, P., Noël, B. P. Y., Van De Berg, W. J., Van Meijgaard, E., and Wouters, B.: On the recent contribution of the Greenland ice sheet to sea level change, The Cryosphere, 10, 1933–1946, https://doi.org/10.5194/tc-10-1933-2016, 2016.

Velicogna, I., Sutterley, T. C., and van den Broeke, M. R.: Regional acceleration in ice mass loss from Greenland and Antarctica using GRACE time-variable gravity data, Geophysical Research Letters, 41, 8130–8137, https://doi.org/10.1002/2014GL061052, 2014.

Wei, S., Zheng, L., Liang, Q., Li, T., and Cheng, X.: An 8-day Antarctic supraglacial lake dataset from MODIS (Version 1), Zenodo [data set], https://doi.org/10.5281/zenodo.19936100, 2026.

Williamson, A. G., Banwell, A. F., Willis, I. C., and Arnold, N. S.: Dual-satellite (Sentinel-2 and Landsat 8) remote sensing of supraglacial lakes in Greenland, The Cryosphere, 12, 3045–3065, https://doi.org/10.5194/tc-12-3045-2018, 2018.

Willmott, C. J.: On the validation of models, Physical Geography, Physical Geography, 2, 184–194, https://doi.org/10.1080/02723646.1981.10642213, 1981.

Wong, T.-T.: Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation, Pattern Recognition, 48, 2839–2846, https://doi.org/10.1016/j.patcog.2015.03.009, 2015.

Wu, H. and Li, Z.-L.: Scale Issues in Remote Sensing: A review on analysis, processing and modeling, Sensors, 9, 1768–1793, https://doi.org/10.3390/s90301768, 2009.

Xiang, H.: Algorithms for Moderate Resolution Imaging Spectroradiometer cloud-free image compositing, Journal of Applied Remote Sensing, 7, 073486, https://doi.org/10.1117/1.JRS.7.073486, 2013.

Yang, K. and Smith, L. C.: Supraglacial streams on the Greenland Ice Sheet delineated from combined spectral–shape information in high-resolution satellite imagery, IEEE Geoscience and Remote Sensing Letters, 10, 801–805, https://doi.org/10.1109/LGRS.2012.2224316, 2013.

Zheng, L., Shang, X., Van Den Broeke, M. R., Noël, B., Li, X., Fettweis, X., Liang, Q., Wang, K., Liu, J., and Cheng, X.: Rapid increases in satellite-observed ice sheet surface meltwater production, Nature Climate Change, 15, 769–774, https://doi.org/10.1038/s41558-025-02364-4, 2025.

https://doi.org/10.5194/essd-2026-353
Preprint. Discussion started: 16 June 2026

580 Zwally, H. J., Abdalati, W., Herring, T., Larson, K., Saba, J., and Steffen, K.: Surface Melt-Induced Acceleration of Greenland Ice-Sheet Flow, Science, 297, 218–222, https://doi.org/10.1126/science.1072708, 2002.