Expansion of supraglacial lake area, volume and extent on the Greenland ice sheet from 1985 to 2023

Yubin Fan[1,2], Chang-Qing Ke[2] (iD), Lanhua Luo[2], Xiaoyi Shen[3], Stephen John Livingstone[4] (iD) and James M. Lea[5]

[1]School of Geography and Planning, Chizhou University, Chizhou, Anhui, China; [2]Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Key Laboratory for Land Satellite Remote Sensing Applications of Ministry of Natural Resources, School of Geography and Ocean Science, Nanjing University, Nanjing, Jiangsu, China; [3]School of Earth Sciences and Engineering, Hohai University, Nanjing, Jiangsu, China; [4]School of Geography and Planning, University of Sheffield, Sheffield S10 2TN, UK and [5]Department of Geography and Planning, School of Environmental Sciences, University of Liverpool, Liverpool, UK

Abstract

Supraglacial lakes (SGLs) are widespread across the Greenland ice sheet and cause transient changes in ice flow. Here, we produce the first annual ice-sheet wide database of maximum summer SGL extents spanning 1985 to 2023 using all July and August Landsat images. Lake visibility percentages were calculated to estimate the uncertainty induced by variable image data coverage. SGLs were mainly distributed between 1000 and 1600 m elevation, with large lake area observed in northwestern, northeastern and southwestern basins. Lake area increased at a rate of 50.5 km^2 a^{-1} across the entire Greenland, and lakes advanced to higher elevations at an average rate of 10.2 m a^{-1} during 1985–2023. We leveraged spatiotemporally matched ICESat-2 and Landsat 8 reflectance data to develop a deep learning model for lake depth inversion for the period 2014–23. This model demonstrates the highest accuracy among all image-based methods, albeit with an underestimation of ~15% when compared to ICESat-2 data. A significant positive correlation between lake volume and area is used to up-scale the approach to the entire time period, indicating a lake volume increase of $221.9 \pm 63.6 \times 10^6$ m^3 a^{-1}. Increasing air/land surface temperature, surface pressure and decreasing snowfall were the most important contributing factors in driving lake variability.

1. Introduction

Supraglacial lakes (SGLs) play an important role in the response of the Greenland ice sheet (GrIS) to climate warming. The presence of SGLs reduces ice-sheet albedo, providing a positive feedback with ice-sheet ablation (Hubbard and others, 2016). Large volumes of meltwater stored in SGLs can rapidly (hours to a few days) drain into the ice sheet via moulins and crevasses, resulting in transient ice-flow perturbations (Das and others, 2008; Stevens and others, 2022). SGLs are expected to become more abundant and extend further inland towards higher elevations due to increases in surface melt extent and duration (Leeson and others, 2015; Ignéczi and others, 2016). It is, therefore, essential to have a comprehensive understanding of their spatiotemporal distribution and evolution to evaluate their potential future impacts.

Remote sensing can be used to reveal the spatial and temporal evolution of SGLs at the ice-sheet scale. Understanding of lake variability has improved due to more frequent (weekly and seasonal) and larger-scale observations (Benedek and Willis, 2021; Yang and others, 2021). However, only a few studies have detected SGLs across the entire GrIS (Selmes and others, 2011; Hu and others, 2022; Zhang and others, 2023), while regional studies have primarily focused on the southwestern and northeastern regions (e.g. Rowley and others, 2020; Lu and others, 2021; Turton and others, 2021; Yang and others, 2021). Therefore, regional differences in lake evolution (e.g. Noël and others, 2019) and underlying controls remain poorly understood. Finally, studies utilizing medium- to high-resolution imagery have mainly focused on lake changes in the recent decade (Liang and others, 2012; Miles and others, 2017; Williamson and others, 2018); long-term (multi-decadal) variations in lake evolution are less well constrained.

Although most studies typically monitor changes in lake area (e.g. Gledhill and Williamson, 2018; Turton and others, 2021; Hu and others, 2022), their depth and volume are of crucial significance in quantifying changes in stored surface meltwater, the amount of water entering the ice sheet through crevasses and moulins and the water lost through subglacial systems (Smith and others, 2015). Depth inversion from optical imagery includes empirical-based (Liang and others, 2012) and physically based methods (Pope and others, 2016; Williamson and others, 2017; Arthur and others, 2022). ICESat-2 laser altimetry can also detect the bathymetry of lakes up to 11.55 m deep using the ATL03 photon data to differentiate the double reflection of the water surface and bottom (Datta and Wouters, 2021). These methods exhibit strengths and limitations in estimating lake depths. Empirical methods typically require substantial in situ data as reference values for deriving depth–reflectance curves. Physically

based methods can offer continuous spatial coverage with high temporal resolution, but their applicability is limited to lakes <5 m deep (Pope and others, 2016). Although ICESat-2 has the capability to accurately measure lake depths, it is confined to its tracks, which are spatially distant and exhibit a coarse temporal sampling interval of 91 d under cloud-free conditions. This makes it challenging to assess lake depths on weekly or monthly timescales.

To capture the interannual variability of SGLs, we first mapped the maximum summer SGL extent across the entire GrIS using all Landsat images acquired between July and August from 1985 to 2023 with cloud cover <50%. Interannual variations in the location, area and elevation of SGLs were quantified. Spatiotemporally matched ICESat-2 and Landsat 8 band reflectance data were used to construct a deep learning model for lake depth inversion to calculate lake depth/volume for the period 2014–23, which aligns with the operational period of Landsat 8. The relationship between lake area and volume were used to construct a linear relationship to estimate lake volume from 1985 to 2013. Finally, relationships between climatic factors and lake development were investigated.

2. Methods

2.1 Identification of SGLs

Given the 16 d temporal resolution of Landsat images and frequent cloud cover, it becomes extremely challenging to observe SGLs on monthly or weekly timescales at a large scale. Consequently, we focused on lake area around the peak of the melt season (July–August) to analyse interannual variability (following Arthur and others, 2022). To fully leverage the Landsat Collection 2 Tier 1 Top-Of-Atmosphere Reflectance data in the Google Earth Engine (GEE) platform, we used all available Landsat images during July and August each year to provide continuous and comprehensive SGL observations. We filtered to only include images with cloud cover <50% and sun elevation angles larger than 20° (Moussavi and others, 2020). We used a relatively high cloud cover filter as images still provided partial cloud-free observations.

To fill data gaps caused by the Landsat 7 ETM+ Scan Line Corrector (SLC)-off data, we employed interpolation techniques using the focal statistics function in the GEE platform (Tedstone and Machguth, 2022), which utilizes a circular buffer with a radius of ∼8 pixels to calculate the mean value of non-void pixels and took this mean value as the reflectance of the void pixel. All Landsat images have a spatial resolution of 30 m.

For all Landsat images, we first used the ice-sheet boundary from Tedstone and Machguth (2022) to mask rock, land and sea regions. Additionally, the boundary data provided by them (100k_boxes.shp) represent the upper limit of meltwater areas. This shapefile was used to define the boundary for our lake extraction operations, helping to focus the analysis on regions with SGL influence. In this paper, we refer to the unmasked areas as the study area. Clouds were masked out using a shortwave infrared (SWIR) band threshold of <0.1. The normalized difference snow index (NDSI) of >0.8 (Eqn (1)) was used to mask snow (Moussavi and others, 2020):

$$NDSI = \frac{Green - SWIR}{Green + SWIR} \qquad (1)$$

where *Green* and *SWIR* bands correspond to the reflectance in the green and SWIR bands, respectively.

Overall, 1935 Landsat 5 images, 3550 Landsat 7 images and 8179 Landsat 8 images across the study area were used in this
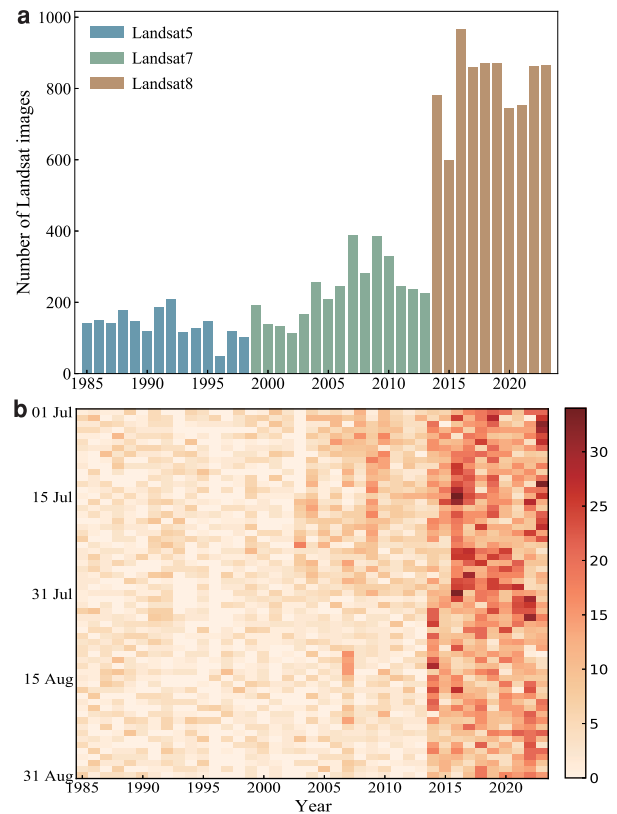

Figure 1. (a) The total number of used Landsat images taken in individual years from 1985 to 2023 in the study area of GrIS. (b) Number of available Landsat images on individual days in the months of July and August from 1985 to 2023.

study, with an average of 350 images per year (Fig. 1a). The sensing times of these images varied between years, but most images were obtained in July (Fig. 1b). A composite image of the study area was obtained each year, with each pixel taken from the image with the highest Normalized Difference Water Index ($NDWI_{ice}$, Eqn (2)) in the image collection (using the qualityMosaic function in GEE). The $NDWI_{ice}$ was then used to extract SGLs:

$$NDWI_{ice} = \frac{Blue - Red}{Blue + Red} \qquad (2)$$

where *Blue* and *Red* correspond to the reflectance in the blue and red bands, respectively. In order to capture as many shallow SGLs as possible, a threshold of 0.2 was applied to segment water (Datta and Wouters, 2021).

Following Pope and others (2016) and Williamson and others (2018), regions with areas <5 pixels (4500 m^2) and widths <2 pixels were removed to avoid potential misclassifications caused by mixed slush and supraglacial rivers. However, slush zones often comprise large connected areas that would not be removed by these thresholds. Hill and cloud shadows also caused misclassification. Therefore, as a final check, the results were manually inspected to remove or correct misclassifications. Misclassified objects caused by shadows tended to be identified in the NO basin, whereas misclassifications due to slush were found in the SW basin (Fig. S1). To quantify the accuracy of using $NDWI_{ice}$ to identify SGLs, we compared the automatic classification and manually corrected results, resulting in a kappa coefficient of 0.84, a mean intersection over union of 0.82 and F1-score of 0.88. These results suggest the automatic $NDWI_{ice}$ classification has an accuracy of ∼84% highlighting the importance of manual checking to improve identification accuracy.

2.2 Lake visibility percentage calculation

As the coverage of the composite image varied from year to year, particularly prior to the Landsat 8 mission, we calculated the coverage of the Landsat 5 and 7 data compared to the year of 2014 (complete coverage) to provide an indication of data gaps. For these gaps, mapped lakes represent minimum estimates of the true lake area. To account for uncertainty in lake area due to visibility issues, we followed the method of Tuckett and others (2021), which uses the 'lake visibility percentage' (LVP) to evaluate the impact of variable data coverage on lake area.

The calculation of the LVP involves the determination of image visibility scores (IVSs) and lake pixel contribution scores (LPCSs). IVSs represent the percentage of visible ice cover (i.e. cloud free) within each individual image. Subsequently, the frequency of lake pixels contributed by each image within the mosaicked imagery was calculated using the ee.Reducer.frequencyHistogram function in GEE. The LPCS for each image was obtained by multiplying its lake pixel frequency by the corresponding IVS. The LVP for a given year is then calculated by summing the LPCSs of all images. For Landsat 7 images, we did not perform gap-filling before calculating the LVP to avoid introducing additional uncertainty. A comprehensive description of the methodology can be found in Tuckett and others (2021). By performing this assessment of lake coverage, we could scale mapped lake area results up to 100%, thereby attaching an upper uncertainty bound to the minimum mapped lake areas. Years with no image coverage in certain basins were excluded from further analysis. We used the scaled maximum summer SGL extents to analyse the lake variability and climatic factors controlling lake development in this paper.

2.3 Lake depth data from ATL03 data

ICESat-2 ATL03 data can estimate lake bathymetry of SGLs based on the distinct photon returns received from the lake surface (air–water) and bottom (water–ice) interfaces (Fair and others, 2020; Datta and Wouters, 2021; Xiao and others, 2023). The Watta automated depth detection algorithm (Datta and Wouters, 2021) was employed to estimate lake depth as it exhibits good agreement with the original photon data, meeting the requirements for large-scale lake depth extraction (Fricker and others, 2021).

To acquire an extensive set of lake depth training samples, ICESat-2 ATL03 data that were temporally (±3 d) and spatially concurrent with Landsat 8 imagery from July to August during 2019–21 were selected. Given the impact of cloud cover on laser altimetry data and the absence of clear double reflection characteristics in all SGLs, we applied the sliderule technique (Shean and others, 2023) to pre-generate photon profiles for lake locations, which offers an open-source framework for processing the archive of low-level data products from the ICESat-2 mission in the cloud. ICESat-2 data with double reflection characteristics, indicating the ability to detect lake depth, were manually selected (Fig. 2). As the Watta algorithm operates at the photon scale, we averaged depths within the same Landsat 8 pixel to spatially match the resolution, and both the average depth and Landsat 8 band reflectance were utilized as samples for the deep learning model.

2.4 Depth inversion using deep learning

To build a depth inversion model, we initially evaluated the correlation between Landsat 8 reflectance in different bands and the depth of SGLs derived from ICESat-2 (Table 1). The findings revealed a strong correlation between lake depth and reflectance in the visible, near-infrared (NIR) and panchromatic bands, while the correlation is comparatively lower in the SWIR band. Consequently, the depth inversion model was constructed using reflectance from coastal, blue, green, red, NIR and panchromatic bands, excluding the SWIR band to enhance model precision during training.

A multi-layer perceptron was employed for data training (Fig. S2), which is an artificial neural network designed for supervised learning (Gaudart and others, 2004). It consisted of multiple layers of nodes or neurons, organized in an input layer, one or more hidden layers and an output layer. Each connection between nodes was associated with a weight, and the network learned to adjust these weights during training to optimize its performance on data classification and regression. The depth dataset was partitioned, with 70% allocated to the training set, 20% to the validation set and 10% to the test set. The Adam optimization algorithm was selected as the optimizer (Zhang, 2018). Through iterative experimentation involving different configurations of hidden layers and neurons per layer, we identified the configuration that yielded the smallest RMSE in the validation set.
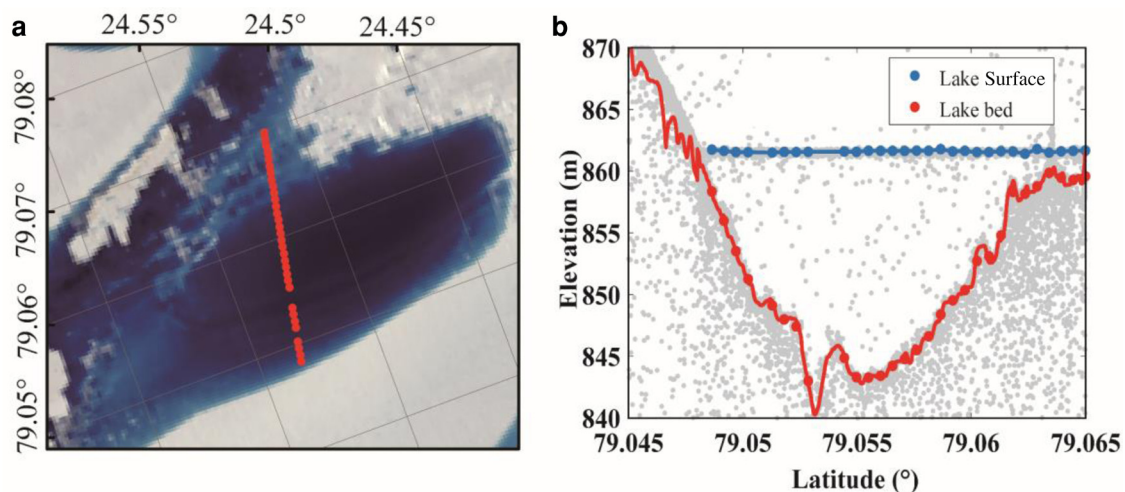

Figure 2. Example of the Watta algorithm for SGL depth detection. (a) ICESat-2 ATL06 track overlaid on a Landsat 8 image acquired on 3 August 2020, showing an SGL. (b) Original ICESat-2 ATL03 photon data collected over the lake on 2 August 2020. The top (blue line) and bottom (red line) of the double reflection correspond to the lake surface and bed derived from Watta.

Table 1. Correlation between SGL depth from ICESat-2 and Landsat 8 reflectance in each band

Band	Coastal	Blue	Green	Red	NIR	SWIR1	SWIR2	Pan	Cirrus
Correlation	−0.61	−0.66	−0.68	−0.53	−0.39	−0.24	−0.23	−0.62	−0.16

All significance levels <0.01.

Specifically, a model comprising three hidden layers with respective neuron counts of 12, 25 and 12 is chosen. The rectified linear unit was adopted as the activation function for feedforward. Parameters such as batch size (16), iteration count (100) and learning rate (0.001) were set to establish the depth inversion model. We treated RMSE between lake depths derived from deep learning and ICESat-2 in the validation dataset as the inversion uncertainty.

Three other typical approaches for determining lake depth were used for comparison with the deep learning approach. The physically based method was approximated as follows (Pope and others, 2016):

$$\text{depth} = \frac{[\ln(A_d - R_\infty) - \ln(R_w - R_\infty)]}{g} \quad (3)$$

where A_d represents the lake bottom albedo, estimated through the average reflectance in a circular buffer of three pixels around the lake. R_∞ is the reflectance of deep water (>40 m), which is calculated as the average value of the dark ocean areas in four Landsat 8 images from 2019 to 2020, R_w is the reflectance of the lake water and g is the bidirectional attenuation coefficient, $g = 0.7507$ in the red band, and $g = 0.3817$ in the panchromatic band. The lake depth is equal to the mean value of the red band and panchromatic band.

The logarithm ratio of blue and green band reflectance (Moussavi and others, 2016) was calculated using the below equation:

$$\text{depth} = a\left[\ln\left(\frac{R_b}{R_g}\right)\right]^2 + b\ln\left(\frac{R_b}{R_g}\right) + c \quad (4)$$

where R_b and R_g represent the reflectance of blue and green bands, a, b and c are the coefficients of its quadratic polynomial fit of $\ln(R_b/R_g)$.

Finally, depth was calculated using multi-variable linear regression based on reflectance bands 1–5 and band 8 for comparison with the results of deep learning.

2.5 Climatic factors controlling SGL development

Fifth generation of European Centre for Medium-Range Weather Forecasts atmospheric reanalysis (ERA5) offers high-resolution, continuous data from 1950 to present, with higher accuracy in simulating downward solar and infrared radiation fluxes compared to ERA-Interim and Modèle Atmosphérique Régional (Delhasse and others, 2020). Therefore, climatic factors from ERA5 reanalysis data, comprising 2 m air temperature, skin temperature, snowfall, surface net thermal radiation, surface net solar radiation and wind speed, were utilized to analyse potential controls governing lake development. These factors have been detected as key controls governing SGL reoccurrence at the GrIS and the Antarctic ice sheet (Rowley and others, 2020; Turton and others, 2021; Arthur and others, 2022). The July–August mean of each basin was calculated using the cumulative lake areas over the entire study period, and Pearson's correlation analysis was performed between lake areas and mean climatic factors to investigate relationships.

3. Results

3.1 Impact of variable data coverage on SGL area

LVPs ranged from 33.13 to 86.27%, with an average of 72.08% and a median of 73.73% across the entire GrIS from 1985 to 2013 (Fig. 3b). However, notable differences were observed between LVPs from Landsat 5 (66.56%) and Landsat 7 (77.23%) images, with the latter exhibiting higher mean values. The discrepancies in Landsat 7 data can be mainly attributed to gaps caused by the SLC failure despite gap filling. The ice-sheet basins with the lowest LVPs were primarily located in the southwest, particularly in the SO and SW basins from 1985 to 1991 (Fig. 3b). Additionally, the NO basin exhibited low LVPs in 1986, 1996,
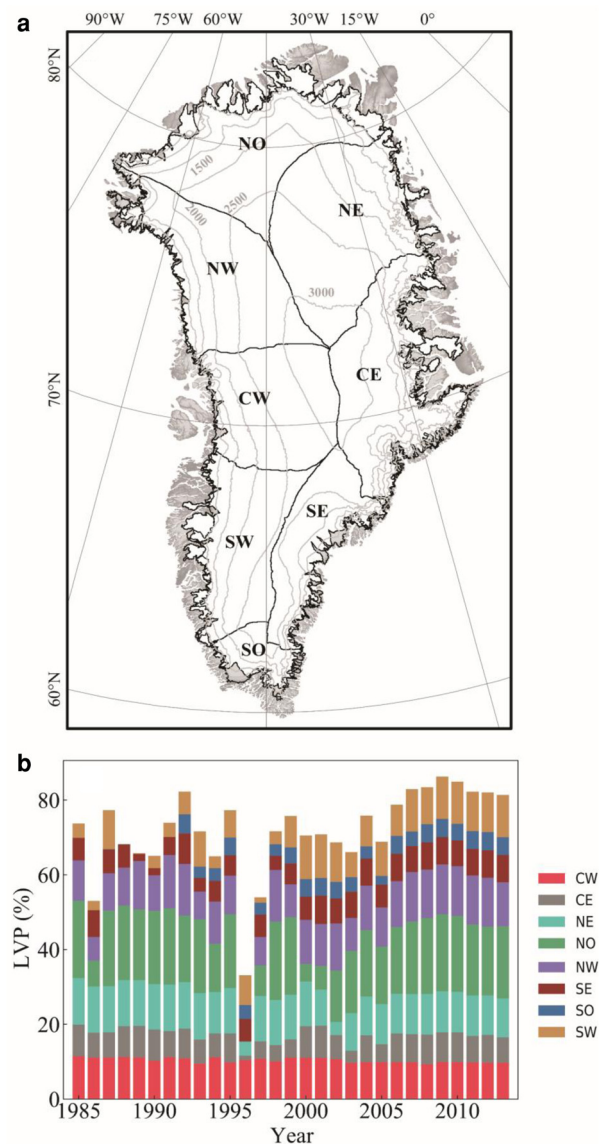

Figure 3. (a) Location of Greenland and the division of eight basins delineated by Zwally and others (2012). Grey lines represent contours with a contour interval of 500 m, produced from the 1 km ArcticDEM. (b) LVP for Greenland basins from 1985 to 2013.

1997, 2000 and 2001, while the CE basin experienced significant data loss only in 1996. By using LVPs to estimate maximum lake areas, we were able to address lake area underestimations due to data gaps in early images. On average, incorporating LVPs into lake area estimates resulted in a 50.29% increase in lake area for Landsat 5 and a 29.76% increase for Landsat 7 images.

3.2 Distribution of Greenland SGLs

Using 14 064 Landsat images, we mapped for the first-time ice-sheet-wide distribution of maximum summer SGL extents during the melt season and evolution peak from 1985 to 2023. Over this time period, an average of 8791 ± 2388 SGLs were mapped annually across the GrIS. Generally, SGLs were found to be widespread within 150 km of the ice-sheet margin, with large lake area observed in NW, NE and SW basins.

Approximately 90% of the total number of SGLs had relatively small areas, ranging from 0.045 to 0.5 km² (Fig. S3a), consistent with the findings of Hu and others (2022). Despite their abundance, these smaller SGLs represent ∼30% of the total lake area. In contrast, lakes of larger area (0.5–20 km²) account for only ∼10% of the total number but are a substantial proportion (∼65%) of the total lake area. Maximum lake area exhibited a significant increase of 0.4 km² a⁻¹ ($p < 0.01$) from 1985 to 2023 (Fig. S3b). The distribution of SGLs on the GrIS is predominantly concentrated within the elevation bin of 1000–1600 m (Fig. 4a). Larger lakes were mainly observed in the SW and NE basins of Greenland within the elevation range of 900–1100 m (Figs 4b, S4), while smaller lakes are prevalent across the ablation zone, consistent with Ignéczi and others (2018).

3.3 Variability and trends in SGL area development

SGLs on the GrIS exhibit significant interannual variation between 1985 and 2023, but with an overall positive trend in lake area of 50.5 km² a⁻¹ ($R^2 = 0.81$, $p < 0.01$) (Fig. 5e). The northern basins exhibited increases in lake area of 7.4–7.6 km² a⁻¹, which is a relatively higher rate of increase compared to the central and southern basins. The SW basins exhibited the highest increase in lake area (15.8 km² a⁻¹, $R^2 = 0.41$), whereas the SE and SO basins had relatively lower rates of 0.02 and 0.9 km² a⁻¹, respectively (Figs 5a, c). Regions with higher rates of change also exhibited larger interannual variations. For example, the NE basin had a total lake area exceeding 645.44 km² in 2019, compared to only 228.85 km² in 2013. The rate at which lake area increased was not uniform across the different elevation bins, with the greatest increase between 1000 and 1800 m, particularly after 2004 (Fig. 4a). Lake elevation increased at a rate of 10.2 m a⁻¹ over the 39 year period (Fig. 5f), with the magnitude of different basins ranging from 4.1 to 14.3 m a⁻¹ (Figs 5b, d). Overall, after calculating the LVP, the trend decreased from 60.8 to 50.4 km² a⁻¹ for the entire GrIS (Fig. S5). Most basins show a decreased slope after the LVP calculation, with the SW basin, exhibiting the most significant decrease of 7.2 km² a⁻¹. This indicates that not accounting for data gaps can lead to an overestimation of lake growth rates.

SGL reoccurrence refers to the number of times lakes occur at each pixel. At the ice-sheet scale, SGLs reoccur more often at low ice-surface elevations (600–1800 m) near the ice-sheet margin where surface meltwater is prevalent each melt season enabling basins to be re-filled (Fig. S6). In the north, both the average reoccurrence and maximum reoccurrence rate of SGLs is higher compared to the south suggesting a more stable surface hydrological system. Conversely, the SE basin exhibits lower reoccurrence rates, indicating a higher likelihood of new lakes forming in these areas, and greater sensitivity to variations in melt (Fig. 6). The average reoccurrence is ∼5 years, with ∼70% of

pixels having relatively low reoccurrence values of <5 years. Regions with high reoccurrence rates (>10 years) were concentrated at a surface elevation of 600–1500 m, with a mean surface elevation of ∼1068 m. This elevation spectrum aligns with zones characterized by the presence of deep and large lakes (see also Zhu and others, 2022). Differences in reoccurrence rate is likely related to variations in lake filling due to the amount of melt, variations in the extent of melting and the timing of lake draining.

3.4 Temporal evolution of lake depth and volume

Deep learning demonstrates excellent performance in lake depth inversion (Fig. 7), with a correlation (r value) of 0.92, a mean absolute deviation is 0.85 m and an RMSE of 1.26 m (depth inversion uncertainty) for the validation sets. The RMSE of depth inversion for lakes with very shallow depths (0–1 m) is 1.17 m, which is relatively high, as also noted by Xiao and others (2023). The errors may be partly attributed to the limited number of samples available for shallow water, as indicated by the data sparsity at these shallow depths (0–1 m), observed in Lv and others (2024). The accuracy of depth inversion remains consistent for lakes with depths from 1 to 10 m, yielding an RMSE of 1.02 m and a mean absolute deviation of 0.70 m. However, for deeper lakes, the accuracy experiences a notable decline, resulting in an RMSE of 2.26 m and a mean absolute deviation of 1.79 m. However, by performing the method over large scales, this uncertainty is minimized.

The average depths of lakes remained consistent from 2014 to 2023, ranging from a minimum of 2.06 m in 2019 to a maximum of 2.58 m in 2016. Overall, 65% of lakes have an average depth below 2 m, 24% fall within the 2–4 m depth range, and only 2% exceed depths of 7 m. This distribution is consistent with findings from Fitzpatrick and others (2014), although it likely underestimates the deepest lakes due to saturation of the signal of Landsat 8 or ICESat-2 in our deep learning method. Despite minor interannual variations among different basins, distinct spatial characteristics emerge, with northern basins generally exhibiting higher lake depths than southern basins (Fig. 8).

The NO and NE basin stand out with the highest average depths, reaching 2.4–2.7 m, while SO basins typically have lower average depths ranging from 1.5 to 2 m (Fig. 8). This is likely due to a high basal slip ratio in the NO and NE basins facilitating the transfer of longer basal wavelengths to the surface at higher elevations, favouring the formation of larger and deeper SGLs (Ignéczi and others, 2016). Lakes >50 m deep have been observed in NE Greenland (Neckel and others, 2020), which both supports our findings, while also highlighting how the depth inversions based on Landsat images or ICESat-2 have an upper limit on depth estimation, probably leading to an underestimation of the true depth/volume of lakes here.

We multiplied the lake depth of each pixel by the area of that pixel and summed them to get lake volume. Lake volume on the GrIS demonstrates notable interannual variations from 2014 to 2023. On average, the lake volume is 8.7 km³, reaching a maximum of 10.6 km³ in 2015, and a minimum of 5.8 km³ in 2018 (Fig. 9). The SW, NW and NE basins generally exhibit larger lake volumes, while the SO basin displays a smaller volume, consistent with findings from Ignéczi and others (2016). In 2018, all basins experienced a negative volume anomaly, with a substantial volume decrease in the SW, NW and NE basins, ranging up to −60%. All basins except for the SW basin exhibited positive anomalies in 2019. There is a robust positive correlation between lake volume and area (Fig. 10), characterized by the equation (RMSE = 1.299×10^6 m³, $R^2 = 0.83$, $p = 0.000$):

$$\text{SGL Volume } (10^6 \text{ m}^3) = 4.396 \times \text{SGL area } (\text{km}^2) \quad (5)$$
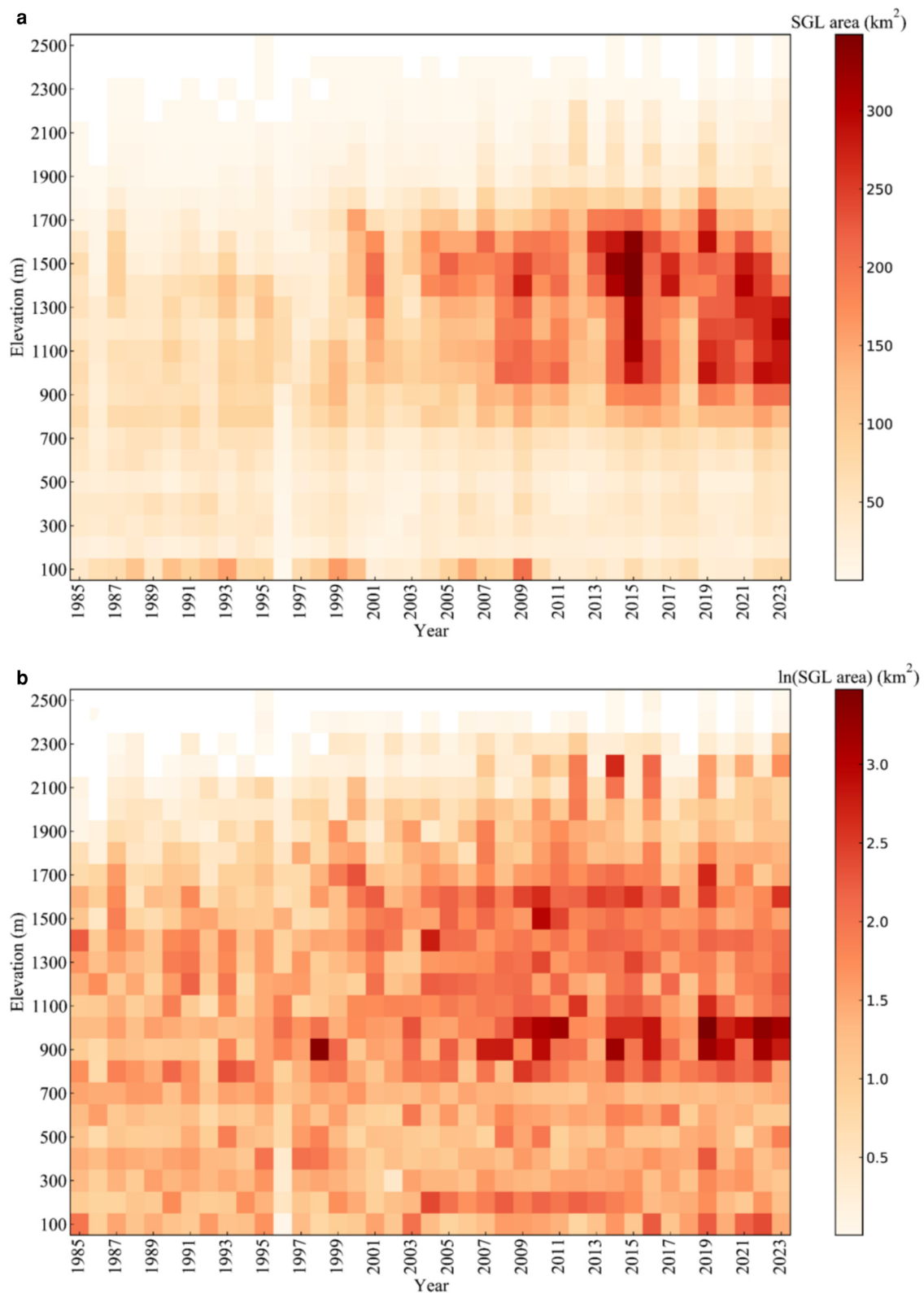
Figure 4. Distribution and variation of total lake area (a) and maximum lake area with a log scale (b) with elevation in the study area of the GrIS from 1985 to 2023.

This relationship aligns with the findings of Gledhill and Williamson (2018), who reported that SGL volume (10^6 m³) is 4.71 times the SGL area (km²). We therefore scaled up lake area to estimate volume over our entire time period, revealing an average estimated lake volume of 4.8 km³ during 1985–89 increasing to 11.7 km³ over 2019–23. Overall, lake volume increased at a rate of $221.9 \pm 63.6 \times 10^6$ m³ a^{-1} from 1985 to 2023.

4. Discussion

4.1 Comparison with other SGL area studies

The distribution of SGLs is consistent with previous studies (e.g. Selmes and others, 2011; Hu and others, 2022; Zhang and others, 2023), with largest total lake area in NW, NE and SW basins. These regions exhibit significant negative surface mass balance
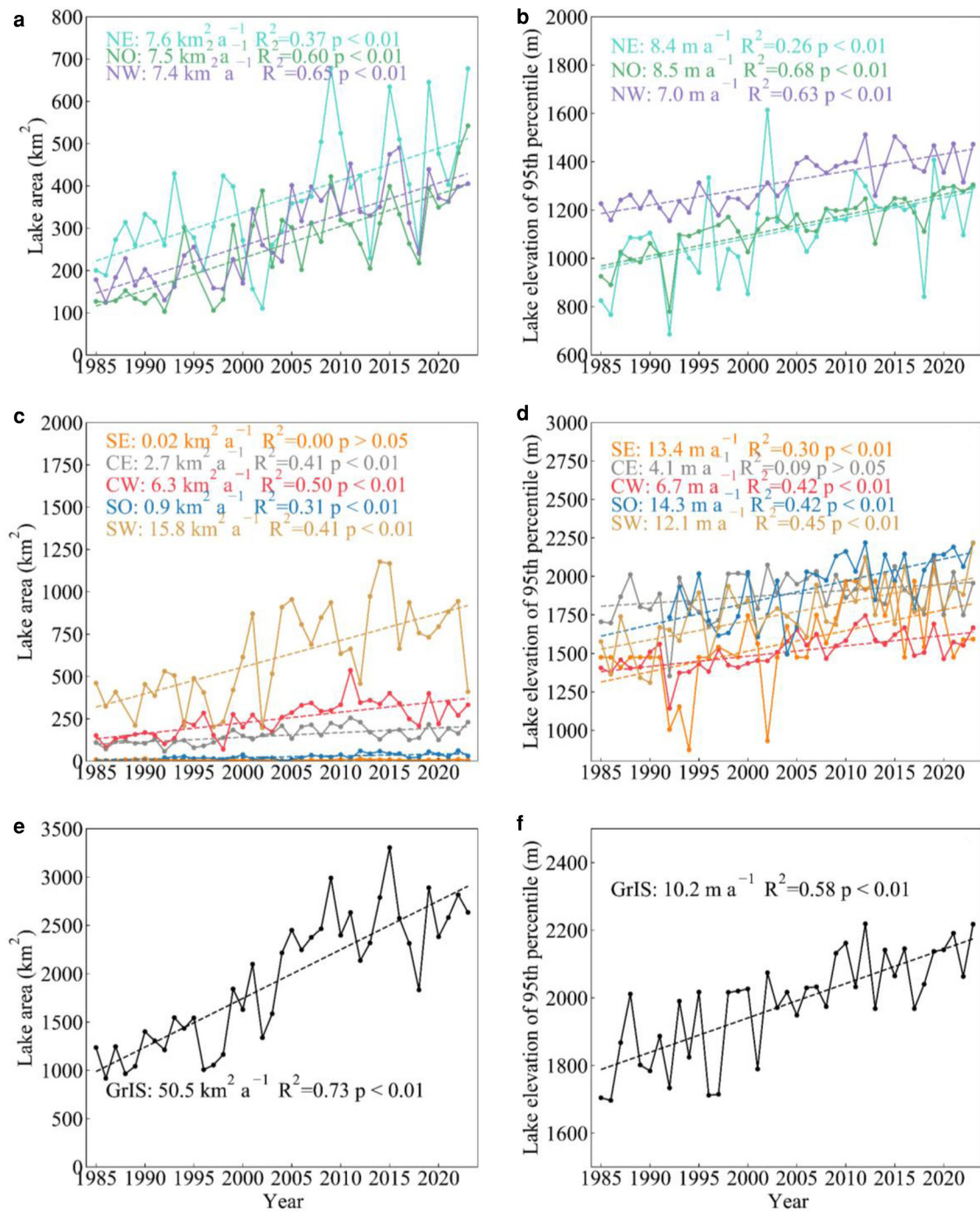
Figure 5. Interannual variability in maximum summer SGL extents (a, c, e) and lake elevation of 95th percentile (b, d, f) of the GrIS from 1985 to 2023. The first row represents the northern region, the second row represents the central and southern regions and the third row represents the entire GrIS. The trend of mapped lake area results can be found in Figure S5.

(Khan and others, 2022), indicating a strong link between lake presence and ice-sheet melting processes. There was a general paucity of lakes in the SE basin of Greenland, which is characterized by steep ice-surface slopes and thick firn making it difficult for meltwater to accumulate and form SGLs (MacFerrin and others, 2019).

Existing research on lake area change at the ice-sheet scale has generally focused on a limited 5 year period (Lu and others, 2021; Hu and others, 2022; Zhang and others, 2023), making it difficult to directly compare with our study. Nevertheless, our results suggest that during this 5 year period (2016–20) lake area was relatively consistent in 2016, 2017 and 2020, smaller in 2018 (∼1965 km^2) but expanded significantly to ∼2997 km^2 in 2019.

This finding aligns with previous studies, and directly reflects the impact of surface melt intensity on lake area. The distribution of lake elevations (1000–1600 m) is consistent with the findings of Ignéczi and others (2018) who demonstrated that lakes tend to form in regions of moderate ice-surface relief where depressions that can hold lakes are most abundant. The elevation of lakes also varies with latitude, with SGLs on the northern GrIS ∼300–400 m lower than in the SW.

Regional mapping of lake evolution over longer periods are of the same magnitude as previous studies. For example, Gledhill and Williamson (2018) found that in the northwestern GrIS (74.1–74.7° N), and lakes advanced to higher elevations at a rate of ∼13.5 m a^{-1} and area increased at a rate of 1.4 km^2 a^{-1}
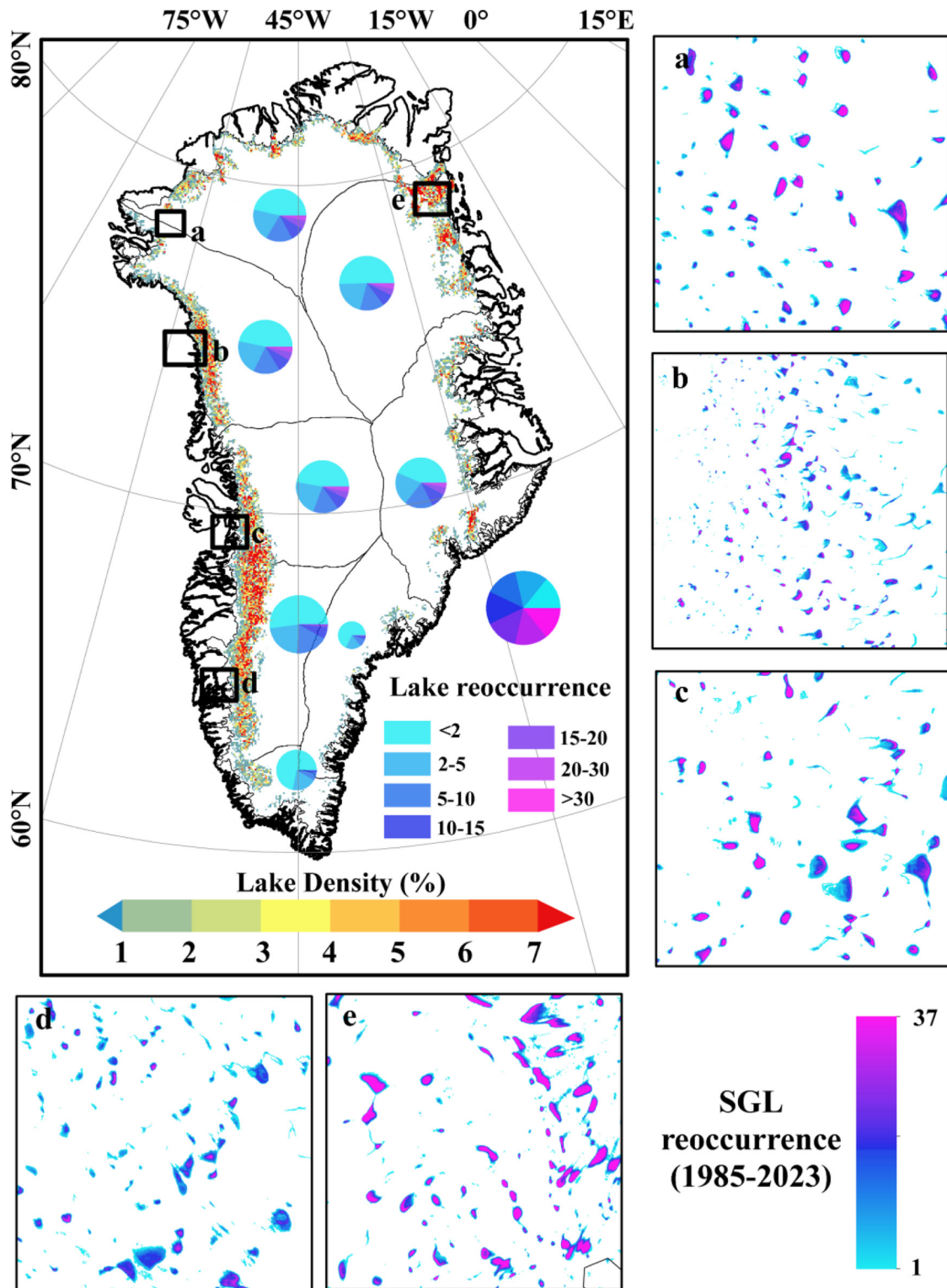
Figure 6. SGL reoccurrence on the GrIS and selected basins, that is: (a) NO, (b) NW, (c) CW, (d) SW and (e) NE basins. Reoccurrence was calculated by summing the number of times lakes occur at each pixel. The Greenland panel shows the spatial density of the lake with reoccurrence >2 over 5 km grids, indicating the proportion of this grid covered by lakes from 1985 to 2023. The pie chart indicates reoccurrence class distribution in the eight basins in GrIS, and the circle size is scaled according to the average lake area from 1985 to 2023.

from 1985 to 2016, with a noticeable acceleration after 2000. Using the same study area as Gledhill and Williamson (2018), our approach approximately reproduces the pattern and magnitude of change, with an increase in lake elevation of $8.5 \, \text{m a}^{-1}$, and an increase in lake area of $1.2 \, \text{km}^2 \, \text{a}^{-1}$. In the southwestern GrIS, Zhu and others (2022) found that lake elevation increased at a rate of $\sim 12.5 \, \text{m a}^{-1}$ from 2000 to 2020. Our data for the same period revealed a comparable elevation increase of 11.8 m

a^{-1}. The similarity of our results with more regional-scale analyses gives us confidence in our results.

4.2 Comparison with other depth inversion methods

We first analysed the relationship between band reflectance and lake depth. Blue, green and red band reflectance all decrease with increasing lake depth (Fig. 11). This absorption of light is
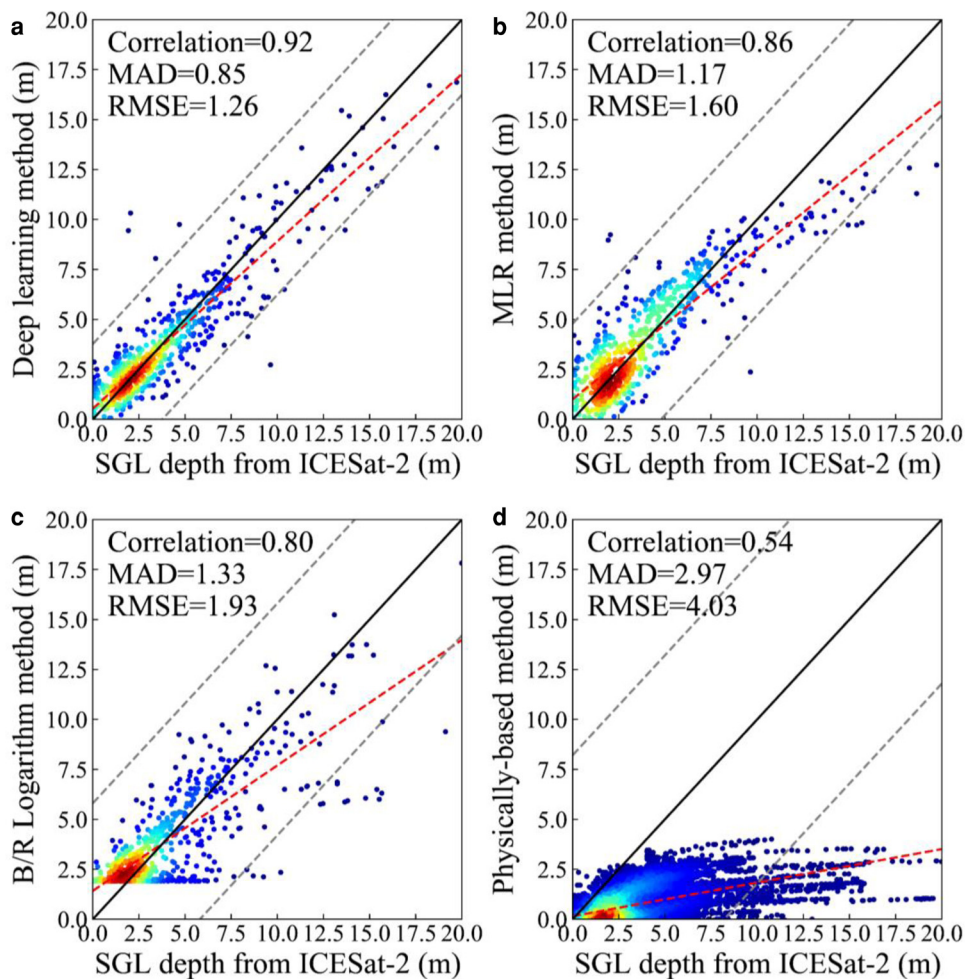
Figure 7. Comparison of SGL depths obtained by ICESat-2 and (a) deep learning, (b) multiple linear regression, (c) logarithm ratio of blue and green band reflectance and (d) physically based method. The red line represents its linear regression line, and grey dotted lines denote the range within three SDs of the mean, the text at the top left of each panel gives different statistical metrics for difference and MAD indicates the mean absolute deviation.

apparent for lake depths >7.5 m for the blue and green channels, but most pronounced for the red band. The reflectance is similar when the lake depth exceeds 5 m due to the signal saturation, which is consistent with Pope and others (2016) and Moussavi and others (2016). Shallow lakes have quite large differences of 0.4 in reflectance values, while medium-depth lakes typically exhibit similar reflectance values.

To evaluate the performance of deep learning method, comparisons were conducted with three alternative methods. The multi-variable linear regression model displayed a notable underestimation of ~25% compared to ICESat-2 measurements (Fig. 7b). The difference between the multiple linear regression and deep learning methods demonstrates that the accuracy of the deep learning method surpasses that of multiple linear regression in lakes with depths exceeding 5 m, while multiple linear regression performs well for lakes with depth <5 m. Our approach demonstrated a marked improvement in accuracy compared to the logarithmic ratio of blue and green band reflectance (Moussavi and others, 2016), with an RMSE of 1.26 m for the deep learning method compared to 1.93 m for the logarithmic method (Fig. 7c). In comparison to physically based methods, deep learning exhibited a considerable improvement. Physically based methods were limited to lake depths not exceeding 5 m (Fig. 7d), indicating a significant deviation from lake depths detected by ICESat-2. In summary, deep learning proves to be well suited for long time-series studies focusing on the inversion of lake depth and volume.

4.3 Factors affecting SGL variability

We analysed the influence of climatic factors on the interannual variability of lake area (Figs 12, S7). Lake areas of northern basins have a significant positive correlation with 2 m air temperature (0.62–0.66), and a slight positive correlation in the SE and SO basins (0.04–0.15), consistent with regions with enhanced negative mass balance (Medley and others, 2022), as reported by other studies (e.g. Sundal and others, 2009; Turton and others, 2021). Snowfall generally shows a negative correlation across all basins (−0.08 to −0.59), implying that decreased snowfall is associated with increased lake areas. Surface pressure shows a range of correlations with lake area, with statistically significant positive correlations in the central and northern basins. Positive correlations in the SO and SW basins are also significant but at a lower level, while the correlation in the SE is not significant. Wind speed is only significantly positively correlated in the NW basin. Surface net solar radiation and surface net thermal radiation show no significant correlation, while land surface temperature shows positive correlations (0.06–0.65). The negative correlation with snowfall in the SE and SO basins is likely linked to the thick permeable firn and relatively steep surface slopes in these regions that favour meltwater infiltration and aquifer formation (Miller and others, 2022). Reduced snowfall and increased melt in this setting would act to increase firn saturation, with refreezing of melt also creating impermeable ice lenses, enabling meltwater to more easily form lakes.
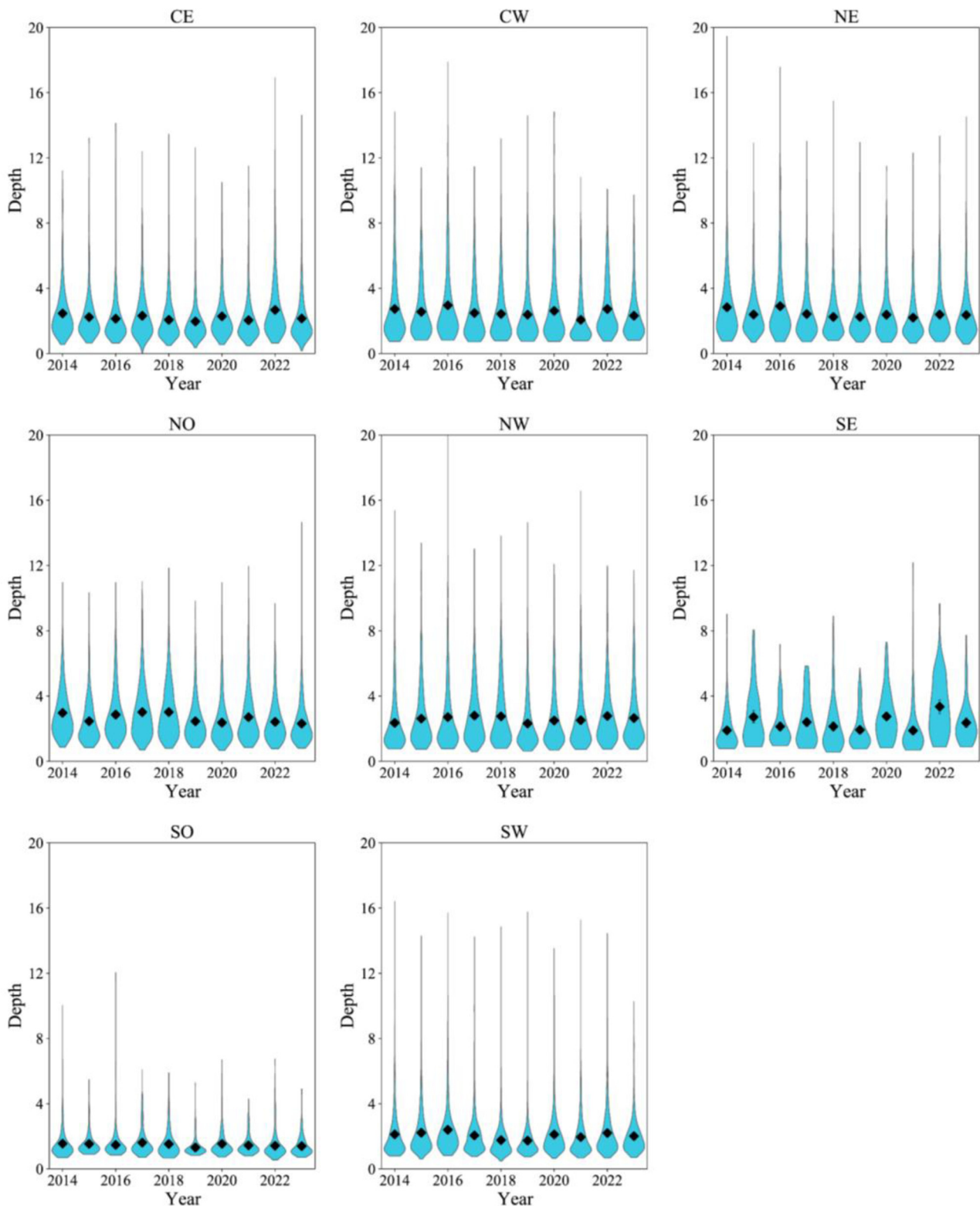
Figure 8. Violin plots showing the depth distribution of lakes in the eight Greenland basins from 2014 to 2023. The shape of each violin plot represents the kernel density estimation of depth data for each year. The black diamond markers in the centre are the mean depths.

The SW basin, with the largest lake area, showed no obvious link with climatic factors. This lack of correlation might reflect the limited number of depressions above the equilibrium line altitude compared to other basins (Ignéczi and others, 2016). As most depressions in the SW basin are likely already filled, this leaves limited capacity for further lake expansion with increased warming. In contrast, regions with more available depressions, such as the central and northern basins, show a strong correlation with temperature. Ice slabs up to several metres thick can be also found in the SW basin, which reduce vertical percolation pathways, and encourage further ice aggregation at their horizon (MacFerrin and others, 2019; Jullien and others, 2023). Recent increases in the area drained by surface rivers align closely with the extent of the ice slabs on the ice sheet (Jullien and others, 2023), and therefore likely play a significant role in controlling lake area in this basin.
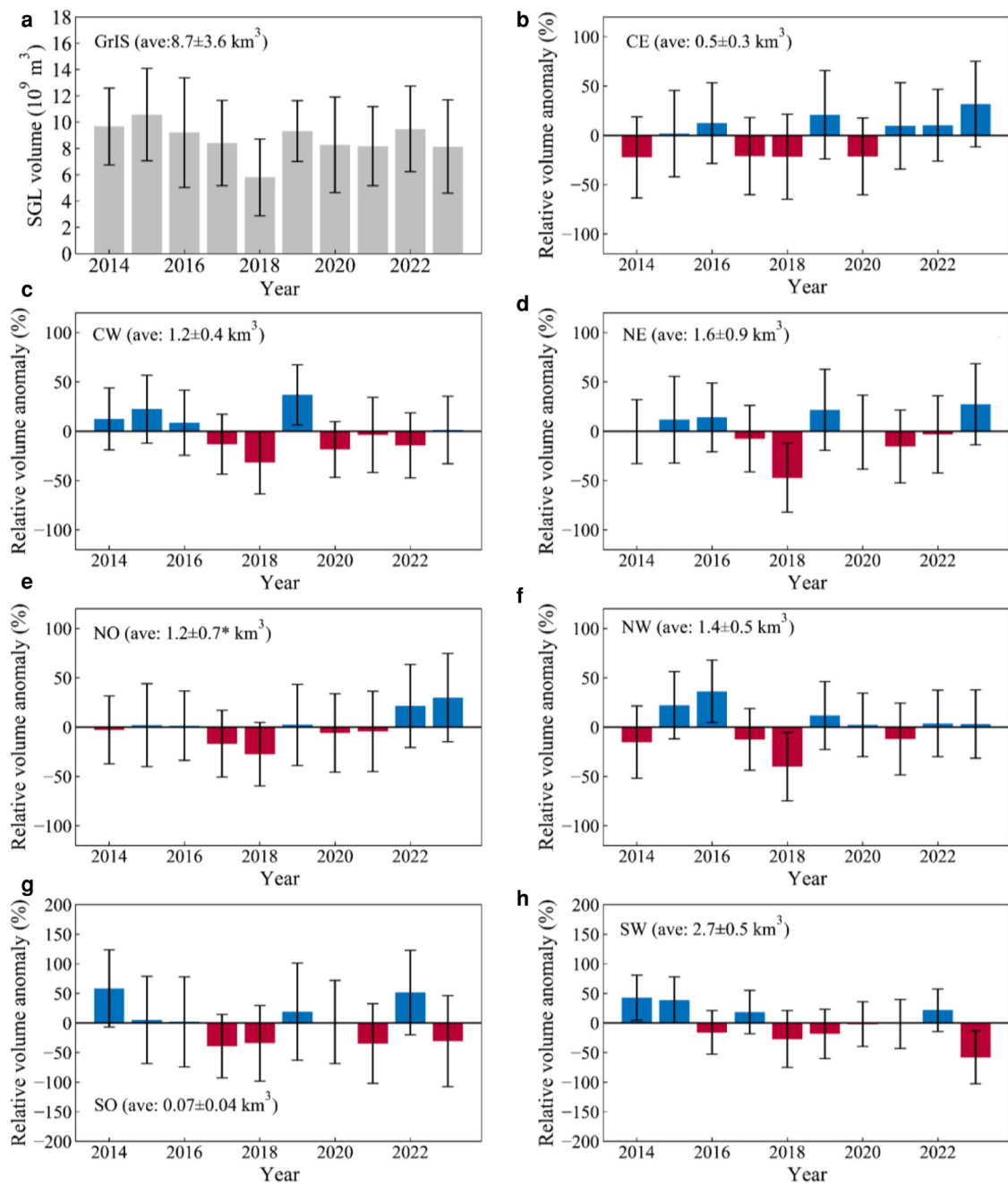
Figure 9. Interannual variations in SGL volume on the GrIS: (a) absolute lake volume and its uncertainty and (b)–(h) represent the anomalies (lake volume relative to the average from 2014 to 2023) in lake volume for each basin. The average values and uncertainty are indicated in the text, and positive and negative anomalies are distinguished by blue and red, respectively.

5. Conclusions

Here, for the first time, maximum summer SGL extents are mapped across the entire GrIS during July and August and from 1985 to 2023 using the Landsat catalogue. The NE and SW basins have the largest lake areas and greatest interannual variability. All eight basins experienced lake area gains from 1985 to 2023, with the southwestern GrIS experiencing the most significant increasing trend of $15.8 \, \mathrm{km^2 \, a^{-1}}$. Lakes exhibited an average rate of inland advance of $10.2 \, \mathrm{m \, a^{-1}}$, with the largest rate of $14.3 \, \mathrm{m \, a^{-1}}$ found in the SO basin. These lake expansion rates are consistent with previous regional-scale studies, helping to provide confidence in the reliability of our results. Lakes tend to re-occur near the ice-sheet margin where higher rates of surface melting at lower surface elevations enable basins to be repeatedly filled. ICESat-2 measurements and the expansive spatial coverage

of Landsat 8 images were used to employ a deep learning method to invert lake depths from 2014 to 2023. Compared with other image-based methods, this approach exhibited substantial improvement, with only a 15% underestimation compared with ICESat-2 data. Lake volume exhibited a minimum volume in 2018 during a period of weak surface melt.

The relationships between common climatic factors and lake area differ regionally, with the key contributing factors being increasing air and land surface temperature except in the southern basins. The presence of thick ice slabs in these southern regions reduces vertical percolation pathways, potentially limiting further expansion of lakes, while the number of available depressions is generally lower than the central and northern basins, further constraining the potential for lake formation. These findings imply that, as ice-sheet melting intensifies, both the area and volume
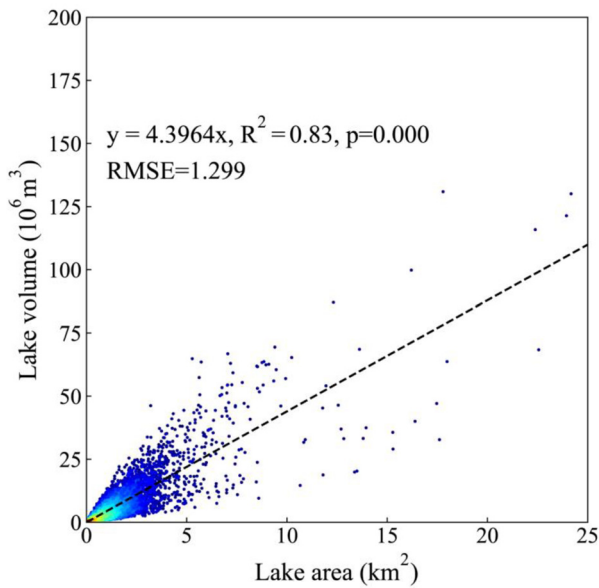
Figure 10. Density scatter plot between SGL area and volume from 2014 to 2023. The black dashed line shows an ordinary least-squares linear regression. The equation of the line and its statistical parameters are shown in the panel. Relationships between lake area and volume for different basins can be found in Table S1.
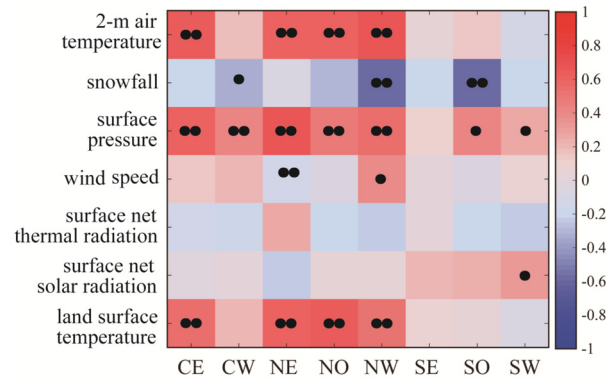

Figure 12. Pearson's correlation coefficients between SGL area and 2 m air temperature, snowfall, surface pressure, wind speed, surface net thermal radiation, surface net solar radiation and land surface temperature from the ERA5 reanalysis dataset. The *X*-axis indicates different basins, shown in Figure 2. The • and •• symbols indicate that the relationships are significant at the 95 and 99% significance levels, respectively, according to a two-tailed Student's *t* tests.
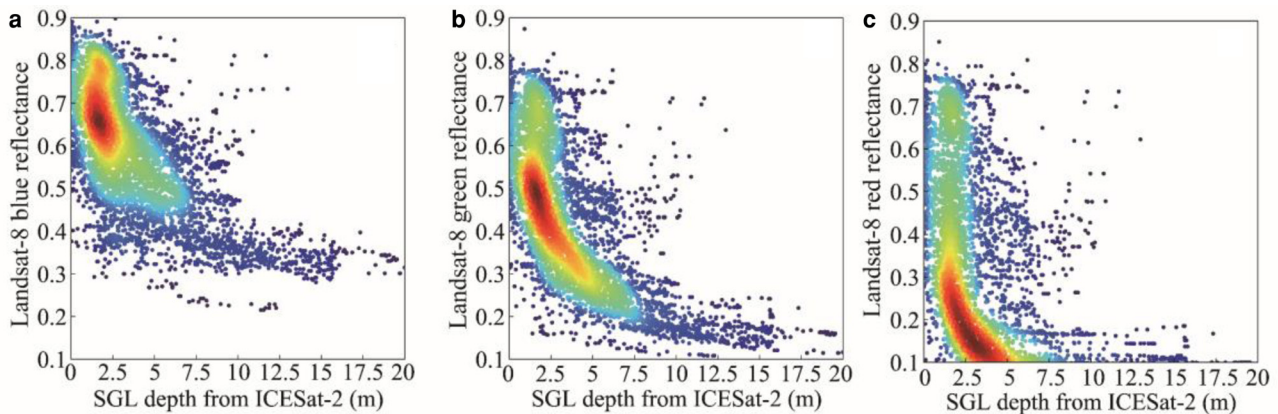
Figure 11. Density plots of depth–reflectance observations for blue (a), green (b) and red (c) bands.

of lakes will continue to expand, gradually extending their distribution towards the inland regions of Greenland but could become limited by the presence of thick ice slabs and available depressions.

More robust relationships between climatic factors and supraglacial conditions remain to be developed to constrain the interplay between climatic factors and lakes. Depth inversion accuracy can be improved by optimizing the deep learning model and integrating multi-source data for improved spatiotemporal resolution. This will contribute to a more comprehensive and continuous record of lake area/volume changes, advancing our understanding of lake evolution.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/jog.2024.87.

Data. The codes for mapping lake extents, and the lake boundary data from 1985 to 2023 can be downloaded from the National Tibetan Plateau/Third Pole Environment Data Center, Institute of Tibetan Plateau Research, Chinese Academy of Sciences at https://data.tpdc.ac.cn/en/data/77528408-ee3e-4322-88ce-0d69f68c5a63.

References

Arthur JF and 5 others (2022) Large interannual variability in supraglacial lakes around East Antarctica. *Nature Communications* **13**, 1711. doi: 10.1038/s41467-022-29385-3

Benedek CL and Willis IC (2021) Winter drainage of surface lakes on the Greenland ice sheet from Sentinel-1 SAR imagery. *The Cryosphere* **15**, 1587–1606. doi: 10.5194/tc-15-1587-2021

Das SB and 7 others (2008) Fracture propagation to the base of the Greenland ice sheet during supraglacial lake drainage. *Science* **320**, 778–781. doi: 10.1126/science.1153360

Datta RT and Wouters B (2021) Supraglacial lake bathymetry automatically derived from ICESat-2 constraining lake depth estimates from multi-source

satellite imagery. *The Cryosphere* **15**, 5115–5132. doi: 10.5194/tc-15-5115-2021

Delhasse A and 6 others (2020): Brief communication: Evaluation of the near-surface climate in ERA5 over the Greenland ice sheet, *The Cryosphere* **14**, 957–965. doi: 10.5194/tc-14-957-2020

Fair Z, Flanner M, Brunt KM, Fricker HA and Gardner A (2020) Using ICESat-2 and operation IceBridge altimetry for supraglacial lake depth retrievals. *The Cryosphere* **14**, 4253–4263. doi: 10.5194/tc-14-4253-2020

Fitzpatrick AAW and 9 others (2014) A decade (2002–2012) of supraglacial lake volume estimates across Russell Glacier, West Greenland. *The Cryosphere* **8**, 107–121. doi: 10.5194/tc-8-107-2014

Fricker HA and 12 others (2021) ICESat-2 meltwater depth estimates: application to surface melt on Amery Ice Shelf, East Antarctica. *Geophysical Research Letters* **48**, e2020GL090550. doi: 10.1029/2020GL090550

Gaudart J, Giusiano B and Huiart L (2004) Comparison of the performance of multi-layer perceptron and linear regression for epidemiological data. *Computational Statistics & Data Analysis* **44**, 547–570. doi: 10.1016/S0167-9473(02)00257-8

Gledhill LA and Williamson AG (2018) Inland advance of supraglacial lakes in north-west Greenland under recent climatic warming. *Annals of Glaciology* **59**, 66–82. doi: 10.1017/aog.2017.31

Hu J and 9 others (2022) Distribution and evolution of supraglacial lakes in Greenland during the 2016–2018 melt seasons. *Remote Sensing* **14**, 55. doi: 10.3390/rs14010055

Hubbard B and 12 others (2016) Massive subsurface ice formed by refreezing of ice-shelf melt ponds. *Nature Communications* **7**, 11897. doi: 10.1038/ncomms11897

Ignéczi Á and 7 others (2016) Northeast sector of the Greenland ice sheet to undergo the greatest inland expansion of supraglacial lakes during the 21st century. *Geophysical Research Letters* **43**, 9729–9738. doi: 10.1002/2016GL070338

Ignéczi Á, Sole AJ, Livingstone SJ, Ng FS and Yang K (2018) Greenland ice sheet surface topography and drainage structure controlled by the transfer of basal variability. *Frontiers in Earth Science* **6**, 101. doi: 10.3389/feart.2018.00101

Jullien N, Tedstone AJ, Machguth H, Karlsson NB and Helm V (2023) Greenland ice sheet ice slab expansion and thickening. *Geophysical Research Letters* **50**, e2022GL100911. doi: 10.1029/2022GL100911

Khan SA and 13 others (2022) Greenland mass trends from airborne and satellite altimetry during 2011–2020. *Journal of Geophysical Research: Earth Surface* **127**, e2021JF006505. doi: 10.1029/2021JF006505

Leeson A and 6 others (2015) Supraglacial lakes on the Greenland ice sheet advance inland under warming climate. *Nature Climate Change* **5**, 51–55. doi: 10.1038/NCLIMATE2463

Liang YL and 7 others (2012) A decadal investigation of supraglacial lakes in West Greenland using a fully automatic detection and tracking algorithm. *Remote Sensing of Environment* **123**, 127–138. doi: 10.1016/j.rse.2012.03.020

Lu Y and 6 others (2021) Response of supraglacial rivers and lakes to ice flow and surface melt on the northeast Greenland ice sheet during the 2017 melt season. *Journal of Hydrology* **602**, 126750. doi: 10.1016/j.jhydrol.2021.126750

Lv J, Li S, Wang X, Qi C and Zhang M (2024) Long-term satellite-derived bathymetry of Arctic supraglacial lake from ICESat-2 and Sentinel-2. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **XLVIII-1-2024**, 469–477. doi: 10.5194/isprs-archives-XLVIII-1-2024-469-2024

MacFerrin M and 13 others (2019) Rapid expansion of Greenland's low-permeability ice slabs. *Nature* **573**, 403–407. doi: 10.1038/s41586-019-1550-3

Medley B, Neumann TA, Zwally HJ, Smith BE and Stevens CM (2022) Simulations of firn processes over the Greenland and Antarctic ice sheets: 1980–2021. *The Cryosphere*, **16**, 3971–4011. doi: 10.5194/tc-16-3971-2022, 2022.

Miles KE, Willis IC, Benedek CL, Williamson AG and Tedesco M (2017) Toward monitoring surface and subsurface lakes on the Greenland ice sheet using Sentinel-1 SAR and Landsat-8 OLI imagery. *Frontiers in Earth Science* **5**, 58. doi: 10.3389/feart.2017.00058

Miller JZ and 5 others (2022) Mapping firn saturation over Greenland using NASA's soil moisture active passive satellite. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **15**, 3714–3729. doi: 10.1109/JSTARS.2022.3154968

Moussavi MS and 6 others (2016) Derivation and validation of supraglacial lake volumes on the Greenland ice sheet from high-resolution satellite imagery. *Remote Sensing of Environment* **183**, 294–303. doi: 10.1016/j.rse.2016.05.024

Moussavi M and 5 others (2020) Antarctic supraglacial lake detection using Landsat 8 and Sentinel-2 imagery: towards continental generation of lake volumes. *Remote Sensing* **12**, 134. doi: 10.3390/rs12010134

Neckel N, Zeising O, Steinhage D, Helm V and Humbert A (2020) Seasonal observations at 79 N glacier (Greenland) from remote sensing and in situ measurements. *Frontiers in Earth Science* **8**, 142. doi: 10.3389/feart.2020.00142

Noël B, van de Berg WJ, Lhermitte S and van den Broeke MR (2019) Rapid ablation zone expansion amplifies north Greenland mass loss. *Science Advances* **5**, eaaw0123. doi: 10.1126/sciadv.aaw0123

Pope A and 6 others (2016) Estimating supraglacial lake depth in West Greenland using Landsat 8 and comparison with other multispectral methods. *The Cryosphere* **10**, 15–27. doi: 10.5194/tc-10-15-2016

Rowley NA, Carleton AM and Fegyveresi J (2020) Relationships of West Greenland supraglacial melt lakes with local climate and regional atmospheric circulation. *International Journal of Climatology* **40**, 1164–1177. doi: 10.1002/joc.6262

Selmes N, Murray T and James T (2011) Fast draining lakes on the Greenland ice sheet. *Geophysical Research Letters* **38**, 1–5. doi: 10.1029/2011GL047872

Shean D and 7 others (2023) Sliderule: enabling rapid, scalable, open science for the NASA ICESat-2 mission and beyond. *The Journal of Open Source Software* **8**, 4982. doi: 10.21105/joss.04982

Smith LC and 15 others (2015) Efficient meltwater drainage through supraglacial streams and rivers on the southwest Greenland ice sheet. *Proceedings of the National Academy of Sciences* **112**, 1001–1006. doi: 10.1073/pnas.1413024112

Stevens LA and 6 others (2022) Tidewater-glacier response to supraglacial lake drainage. *Nature Communications* **13**(1), 6065. doi: 10.1038/s41467-022-33763-2

Sundal A and 5 others (2009) Evolution of supra-glacial lakes across the Greenland ice sheet. *Remote Sensing of Environment* **113**, 2164–2171. doi: 10.1016/j.rse.2009.05.018

Tedstone AJ and Machguth H (2022) Increasing surface runoff from Greenland's firn areas. *Nature Climate Change* **12**, 672–676. doi: 10.1038/s41558-022-01371-z

Tuckett PA and 6 others (2021) Automated mapping of the seasonal evolution of surface meltwater and its links to climate on the Amery Ice Shelf, Antarctica. *The Cryosphere* **15**, 5785–5804. doi: 10.5194/tc-15-5785-2021

Turton JV, Hochreuther P, Reimann N and Blau MT (2021) The distribution and evolution of supraglacial lakes on 79 N Glacier (north-eastern Greenland) and interannual climatic controls. *The Cryosphere* **15**, 3877–3896. doi: 10.5194/tc-15-3877-2021

Williamson AG, Arnold NS, Banwell AF and Willis IC (2017) A fully automated supraglacial lake area and volume tracking ('FAST') algorithm: development and application using MODIS imagery of West Greenland. *Remote Sensing of Environment* **196**, 113–133. doi: 10.1016/j.rse.2017.04.032

Williamson AG, Banwell AF, Willis IC and Arnold NS (2018) Dual-satellite (Sentinel-2 and Landsat 8) remote sensing of supraglacial lakes in Greenland. *The Cryosphere* **12**, 3045–3065. doi: 10.5194/tc-12-3045-2018

Xiao W, Hui F, Cheng X and Liang Q (2023) An automated algorithm to retrieve the location and depth of supraglacial lakes from ICESat-2 ATL03 data. *Remote Sensing of Environment* **298**, 113730. doi: 10.1016/j.rse.2023.113730

Yang K and 7 others (2021) Seasonal evolution of supraglacial lakes and rivers on the southwest Greenland ice sheet. *Journal of Glaciology* **67**, 592–602. doi: 10.1017/jog.2021.10

Zhang Z (2018) Improved Adam optimizer for deep neural networks. In *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*. IEEE, pp. 1–2. doi: 10.1109/IWQoS.2018.8624183

Zhang W and 7 others (2023) Pan-Greenland mapping of supraglacial rivers, lakes, and water-filled crevasses in a cool summer (2018) and a warm summer (2019). *Remote Sensing of Environment* **297**, 113781. doi: 10.1016/j.rse.2023.113781

Zhu D, Zhou C, Zhu Y and Peng B (2022) Evolution of supraglacial lakes on Sermeq Avannarleq Glacier, Greenland using Google Earth Engine. *Journal of Hydrology: Regional Studies* **44**, 101246. doi: 10.1016/j.ejrh.2022.101246

Zwally HJ, Giovinetto MB and Beckley MA (2012) *Antarctic and Greenland Drainage Systems*. Greenbelt, MD, USA: NASA GSFC Cryospheric Sciences Laboratory.