

The automated retrieval of supraglacial lake depth and extent from
ICESat-2 photon clouds leveraging DBSCAN clustering

Gijs van Leeuwen

Supervised by Dr. Ir. Bert Wouters



**Utrecht
University**

FOR COMPLETION OF THE MSc. CLIMATE PHYSICS GIVEN IN COURSE NS-MO552M

INSTITUTE FOR MARINE AND ATMOSPHERIC RESEARCH UTRECHT (IMAU)
07/11/2022

Table of Contents

Abstract	ii
Introduction	1
Methodology	4
2.1 Study area	4
2.2 Data	4
2.2.1 ICESat-2	4
2.2.2 Sentinel-2	5
2.3 Processing chain	5
2.3.1 DBSCAN photon classification	5
2.3.2 Generation of the continuous surface and bottom depths	7
2.3.3 Lake classification	8
2.3.4 Validation & extrapolation	9
Results	13
3.1 Photon classification	13
3.2 Surface and bottom extraction	14
3.3 Lake classification	16
3.4 Lake depths	17
3.4.1 Assessment of the fit	17
3.4.2 Out-of-sample test	17
3.4.3 Relations applied	18
Discussion	20
4.1 Photon classification	20
4.2 Surface and bottom retrieval	21
4.2.1 Validation	22
4.2.2 Extrapolation	24
Conclusion	25
Appendix & data availability	31

Abstract

Meltwater feedbacks are a core component to the Greenland ice sheet surface mass balance but the contributions are hard to quantify, which makes predictions very uncertain. Effort has been taken to quantify the total volume of supraglacial meltwater with optical remote sensing. However, the requirement for in-situ measurements and tuning for each hydrological feature keeps the method from scaling to an automated ice-sheet wide solution. The ICESat-2 altimeter photon retrieval product provides high resolution (0.7m footprint) height estimates over the Greenland ice sheet at regular intervals (90 day revisit time). The Advanced Topographic Laser Altimeter System (ATLAS) onboard ICESat-2 returns both photons reflected from the water surface and the lake bottom which allows for depth retrieval. In this thesis a method is proposed to classify the ATL03 photon returns as surface and bottom photons leveraging an iterative Density-Based Spatial Clustering of Applications with Noise (DBSCAN) approach. The photon classifications yield estimates for the surface and lake bottom heights across the complete ICESat-2 track and a lake classification based on the slope of the surface and the estimated depth. Extrapolation of the along-track depths across the Greenland ice sheet, using Sentinel-2 optical imagery, is investigated. For extrapolation a generalized form for an empirical, physical and machine learning technique are considered. The lake classification, as compared to Sentinel-2 lake detection has an accuracy of 97.8% and provides consistent and smooth depth estimates over the complete ICESat-2 track. Extrapolation of the depth estimates using a random forest regressor shows an out-of-sample R^2 of 0.7 with an RMSE of 1.54 across the complete validation set. The results suggest that DBSCAN clustering of ICESat-2 altimeter data can yield consistent and accurate depth estimates that can be used in a generalized method to estimate lake depths across the entire Greenland ice sheet with a low error margin and small computational overhead.

Keywords: ICESat-2, supraglacial lake, Sentinel-2, DBSCAN, machine learning, Greenland

Introduction

The melt of land ice is the most important factor in global sea level rise for this decade after thermal expansion [Mouginot et al. (2019), Pörtner et al. (2022)] and the uncertainty of the future impact is high. The Greenland ice sheet (GrIS) is currently the largest contributor to the total land ice budget [Bamber et al. (2018)] and is of immense importance for climate tipping points [Pattyn et al. (2018), Pörtner et al. (2022)]. The uncertainty in current mass-balance estimates is a limiting factor for the research of land ice impact [Edwards et al. (2021), Stap et al. (2019)]. Meltwater plays an important role in the mass balance of ice sheets as surface water directly decreases albedo and surface run-off further promotes mass loss [Smith et al. (2015), Bell et al. (2018)]. Other indirect effects like water penetration into the subglacial environment [Zwally et al. (2002), Bartholomew et al. (2012), van de Wal et al. (2008)] and Antarctic ice-shelf collapse [Scambos et al. (2009)] also contribute significant uncertainty to land ice contribution estimates for the future. Meltwater dynamics have become more important to the Greenland ice sheet mass balance over time [Enderlin et al. (2014) Mouginot et al. (2019)] which can be attributed to the non-linear increase of melt with rising temperatures [Trusel et al. (2018)] that shows in the increasing meltwater production on the ice sheet [Van den Broeke et al. (2016)]. Supraglacial lakes on the GrIS act as storage for meltwater and play an important role in the positive meltwater feedback loop [Lüthje et al. (2006)]. Lakes on the GrIS have also become more abundant [Sundal et al. (2009)] thereby further increasing the uncertainty in the surface mass balance (SMB) estimates, as drainage events are chaotic and hard to predict [Williamson, Willis, Arnold & Banwell (2018)]. Quantifying meltwater impact on the surface mass balance is difficult as runoff is hard to directly measure [Smith et al. (2017)] and even meltwater entering the subglacial environment does not always directly relate to mass loss [Willis et al. (2015)]. The water volume of supraglacial lakes can be used as a proxy to improve SMB models and thereby quantify meltwater contributions [Yang & Smith (2012)].

Remote sensing has played an important role in the improvement of supraglacial lake depth and volume retrieval as gathering in-situ data is complex for remote areas like the GrIS. Box and Ski first leveraged optical images from the MODIS satellite to extrapolate sparse in-situ data to supraglacial lake volumes using an empirical relation between the MODIS reflectance and the lake depth [Box & Ski (2007)]. Optical imagery has also been used in deriving lake depths via a physical method based on the attenuation of light through the water column and the albedo of the lake substrate [Philpot (1989), Sneed & Hamilton (2007), Pope et al. (2016)]. Both solutions have seen widespread usage in the retrieval of lake depth, lake volumes and the mapping of surface hydrology on the ice sheet leveraging various optical satellites [Legleiter et al. (2014), Pope et al. (2016), Moussavi et al. (2016), Williamson, Banwell, Willis & Arnold (2018), Dell et al. (2020), Moussavi et al. (2020)]. However, the dependence of both methods on a homogeneous surface roughness, water column, bottom slope and substrate albedo means extrapolation across larger areas is difficult [Sneed

& Hamilton (2007)]. Problems also arise for lakes of greater depths as reflection no longer significantly changes [Box & Ski (2007)] with depth. On top of this, open source optical satellite imagery has relatively high resolution (Sentinel-2 has a 10m resolution) but can only be used to classify features that encompass multiple pixels ; as pixels over mixed terrain have a mixed reflectance so cant be associated with any particular class.

With the launch of ICESat-2 in 2018 a data source of non-optical imagery came available allowing for new physically based methods of lake depth retrieval which immediately saw widespread usage for bathymetry mapping [Parrish et al. (2019), Li et al. (2019), Ma et al. (2020), Hsu et al. (2021)]. The laser altimeter instrument returns photons reflected from both the water surface and the lake bottom enabling lake depth retrievals. In Fair et al. (2020) a method is proposed that uses elevation-based histograms to separate the lake surface from the bottom in along-track windows that have confirmed lakes as indicated by Landsat imagery. A more automated method that does not require optical imagery was proposed in Datta & Wouters (2021) which additionally, provides different classes for the lake bottom dependent on the strength of the bottom reflection and the surface slope. The estimates provided by the ICESat-2 methods are deemed more accurate than the physical method applied to optical imagery [Fricker et al. (2021)] and with additional automation and extrapolation estimates of the total water on the ice sheet during the melt season can be estimated [Datta & Wouters (2021)]. However, histogram-based methods are computationally expensive and reduce the horizontal resolution of the depth estimates due to the along-track windows required. This can be overcome using methods that classify individual photon returns as bottom, surface or noise photons instead of retrieving a bottom line estimate from the aggregated dataset [Xie et al. (2021)].

In this thesis I present a new automated method for supraglacial lake depth estimation from individual ICESat-2 photon returns using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm [Ester et al. (1996)] which was earlier explored for lake bathymetry in Ma et al. (2020). The depth estimates come from a physical approach eliminating the need for in-situ validation measurements and allowing extrapolation to retrieve lake volumes for the entire GrIS. As a proof of concept the cross-lake generalisability of the empirical and physical methods, to extrapolate depth retrievals beyond the ICESat-2 tracks with optical imagery from Sentinel-2, is explored in comparison to a multi-spectral machine learning approach. With the classification of individual photons the native ICESat-2 resolution is retained and automated smoothing windows and DBSCAN parameter selection eliminate track by track manual intervention. The proposed method is independent of optical imagery and its resolution limitations, which allows for the classification of significantly smaller supraglacial hydrological features like water filled crevasses, sloped streams and partially ice covered lakes.

The method is used to classify a track on the outer Western bounds of the GrIS which has a high abundance of lakes and good Sentinel-2 data coverage. The abundance of hydrological features allows for fair validation of the depth estimates and lake classification. The aim is to show a new reliable method for supraglacial lake depth retrieval with lower computational overhead, a native resolution and a method to scale this into 3d returns with optical imagery

that does not need to be tuned for every single hydrological feature. The data sources, and study area as well as the methods for depth retrieval, extrapolation and validation are laid out in the methodology section followed by results for a subset of the classified lakes. The discussion will go into depth on the effects of assumptions and choices made and provide material for future research after which the conclusion will list the most important findings of this thesis.

Methodology

2.1 Study area

The area of interest for this case study is located in the ablation region of the GrIS between 63° and 69° latitude and 46° and 50° longitude. The area has a diverse surface hydrology which is optimal for testing the proposed generalised and automated method. The same area was covered in both Fair et al. (2020) and Datta & Wouters (2021) which makes qualitative comparisons and future comparison research easier. The area was chosen due to the relevance to climate research, good data coverage and complex supraglacial hydrological setting. Only data in locations that coincide with the ICESat-2 tracks shown in figure 2.1 are processed.

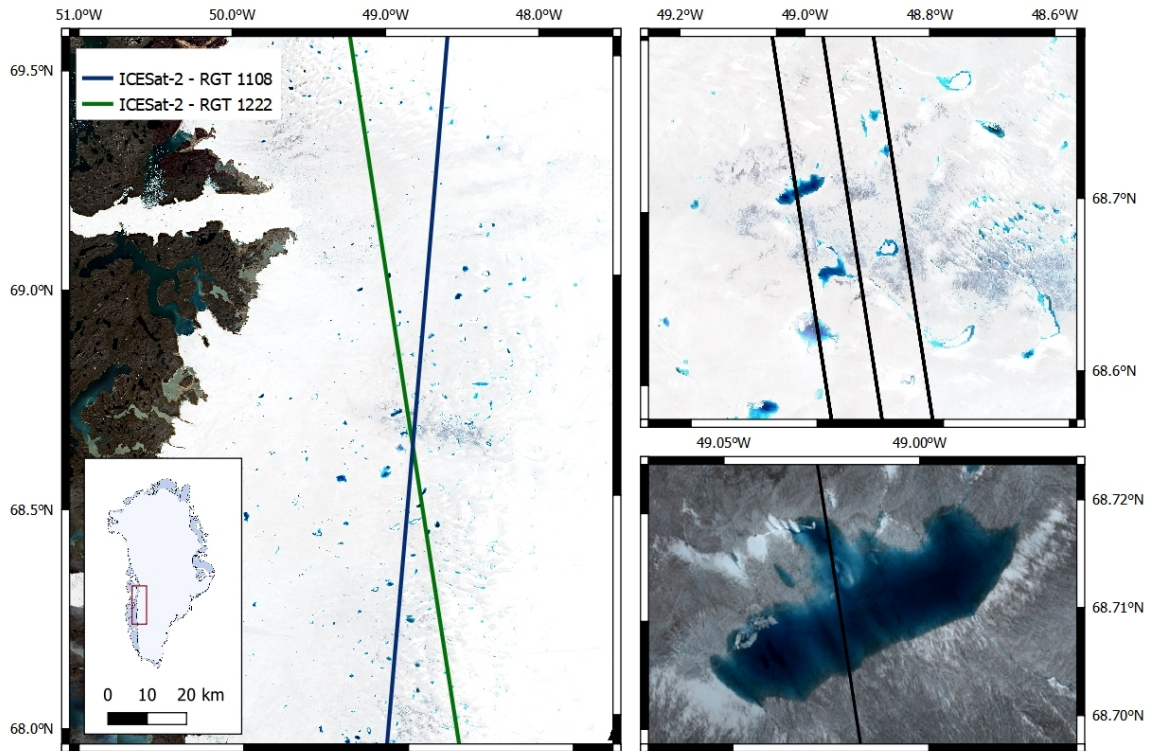


Figure 2.1: The study area in Western Greenland overlaid by the reference ground track of the studied ICESat-2 overpass

2.2 Data

2.2.1 ICESat-2

For the estimation of lake depths the method proposed in this thesis relies solely on the ICESat-2 photon-counter data as provided in the Global Geolocated Photon Data (ATL03) product version 005 [Neumann et al. (2021)]. The Advanced Topographic Laser Altimeter System (ATLAS), the lidar instrument onboard of ICESAT-2 returns 6 beams of data in 3

pairs. Each pair is separated by 3.3 km and the left (weak) and right (strong) beam are separated by 90 meter. For this thesis only the strong left beams are used as a proof of concept. The ATLAS instrument records the two-way travel time for photons at a wavelength of 532 nm send in pulses at a Pulse Repetition Frequency (PRF) of 10 khz [Neumann et al. (2021)]. The travel time translates into a height estimate for each signal photon with an estimated geolocation accuracy of 3.5 m at a footprint of 10.9 m [Magruder et al. (2021)] which is higher than mentioned in the mission requirements [Neumann et al. (2021)]. The ATL03 product also provides the along-track distance, a Digital Elevation Model (DEM) height for each segment from the GIMP-DEM [Howat et al. (2014)] and the aforementioned geolocation as Longitude and Latitude in WGS84. A list of all the used data products used from this mission is given in 2.1.

Table 2.1: ICESat-2 data

RGT	Sensing date	# of photons	Beams
1222	2019-06-17	73,232,889	GT1L, GT2L, GT3L
1108	2019-06-09	30,766,210	GT1L, GT2L, GT3L

2.2.2 Sentinel-2

The Sentinel-2 constellation consists of two passive optical satellites Sentinel-2A, and Sentinel-2B. The mission supplies a Top of Atmosphere (TOA) reflectance product with the Red, Green, Blue, (RGB), Near-InfraRed (NIR) and Short-Wave InfraRed (SWIR) bands at a 10 or 20m spatial resolution. The Sentinel-2 products are downloaded via the Copernicus API [Data availability: section 4.2.2] and used as level 1C radiance data, as TOA reflectance is required for the empirical and physical methods. All images used for this case study are from June 2019 taken 12 hours from the ICESat-2 passing so they can be best used for validation and visual exploration. No cloud masking is implemented in the proposed method so care is taken to choose dates with low cloud cover as estimated from the Sen2cor L2A product [Main-Knorn et al. (2017)]. A list of all the data products used from this mission can be found in table 6.4 in the appendix.

2.3 Processing chain

The photon classification process for retrieving the icesheet surface and potential lake bottoms is a process taken in several steps. The full workflow can be inspected in figure 2.2. The next subsections will cover steps taken for the retrieval of the photon classification, bottom and surface line and lake classification. Section 2.3.4 will go into depth on the retrieval of validation depths from the Sentinel-2 images.

2.3.1 DBSCAN photon classification

Clustering is a method widely used for classification problems in and outside of glaciology [Zhang et al. (2014), Khan et al. (2014), Rizzoli et al. (2017), Gallagher et al. (2020)]. The premise of clustering is that individual data points can be given a similarity score based on

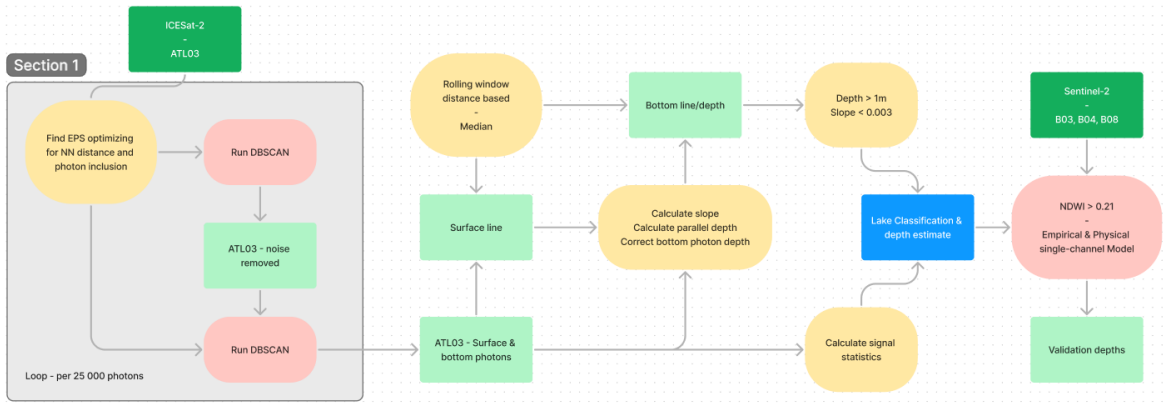


Figure 2.2: Flow diagram indicating the data flow through all sections of the method with Darkgreen boxes indicating the input data, light green boxes intermediate data returns, Yellow and red boxes indicate processing steps and the blue box being the final product.

various factors. The simplest method to measure similarity is the standardized euclidean distance across the feature set. However, one could use more complex metrics if needed. DBSCAN is a variation of the standard k-means that aims to solve the clustering problem for any arbitrary shaped cluster, something that standard distance based similarity doesn't, which makes DBSCAN an optimal algorithm for non-spherical clusters like the surface/bottom lines expected for the ICESat-2 retrievals. DBSCAN is also an optimal method for removing noise as explained later on in this section.

DBSCAN operates with two parameters, Eps and $MinPts$ so that the Eps-neighbourhood can be defined as a circle around a point with radius Eps . In DBSCAN a point is considered a core point if the number of points within it's Eps-neighbourhood is larger or equal to $MinPts$. Beyond core points there is a hierarchy implemented for the way points are connected to their nearest neighbours: (I) Points are directly density-reachable (DDR) if they are in the Eps-neighbourhood of the same core point, (II) a point is density-reachable (DR) if the point is within Eps distance of another core point which is DDR from the current core point, (III)

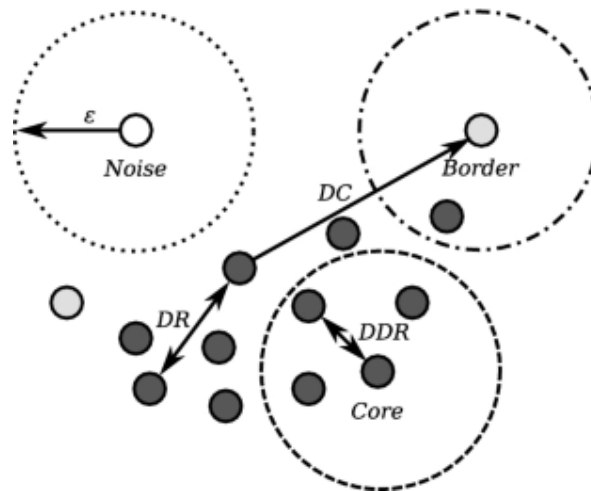


Figure 2.3: Reachability in the DBSCAN algorithm, figure from Götze et al. (2019)

a point is Density Connected (DC) if the point is within the Eps -neighbourhood from a common DR core point. Lastly a point will be classified as noise if the point is neither DDR, DR or DC and can therefore not be part of a cluster. A cluster can now be defined as a collection of points which are either DDR, DR or DC to one of the other points in the cluster like portrayed in figure 2.3.

Drawing inspiration from Ma et al. (2020) DBSCAN is used to cluster the ATL03 photon returns in signal and noise photons as a first step. For this part of the processing the dataset split into the 6 beams is grouped in batches of 25000 photons in order of the photon event occurrence. The dataset has only two dimensions, along-track distance and photon height. First the euclidean distance from each photon return to its *MinPts* neighbour is calculated. Calculating this for every point gives an indication of the density distribution in the window and will be used to give a best estimate for the Eps parameter. The data is compiled for all 25000 photons and plotted on a normalized graph as nearest neighbour distance vs photons included. An example of such a graph is included in the result section figure 3.4. The distance from the line to the origin is calculated for the entire graph and the closest point is considered the optimal cost-benefit of increasing Eps versus including more photons. The non-normalised y coordinate of this point on the line is taken as the Eps for this window with *MinPts* set to 6. DBSCAN is ran on the two dimensional dataset with these parameters classifying individual clusters in the dataset. The result is then reclassified as "signal", anything that was classified in a cluster, or "Noise", anything which was not assigned a cluster, after which all noise data is discarded.

The data classified as signal is now used to redo the Eps optimization step with *MinPts* = 3 so to again differentiate two classes within the signal photons, surface or bottom. An additional parameter is used in this step dubbed *Strict* which influences the scaling of the x-axis in the normalized plot. Increasing the *Strict* parameter (default = 1) will make the automated Eps optimization choose lower values meaning less photons are classified as signal. The resulting DBSCAN cluster classification is again reclassified into a binary system with signal and noise. All photons classified as "Noise" by DBSCAN in this second iteration will be used for the bottom height retrieval laid out in section 2.3.2, all photons classified as "signal" are considered surface returns for the surface height retrieval.

The classification result from the two DBSCAN for every loop consisting of photons classified as: Noise, bottom or surface is combined into a single dataset after the loop commences. This dataset is used to extract the surface and bottom height at the initial ATL03 sampling rate.

2.3.2 Generation of the continuous surface and bottom depths

The photon height of all photons classified as surface is used in a distance based rolling median to generate a smooth line that follows the most consistent photon returns. The window size is set to 20m sampled at every 5m. It is important to have a smooth and consistent surface line as the slope directly impacts the lake classification and the depth calculation. A 20m window size showed to still represent complex geometries, like crevasses, while securing a consistent surface height estimate for regions with sparse photon returns.

The sample rate has significant impact on the computation time, as every further halving would double the computation time for this step. Setting the sample rate at 5m has the best cost-benefit as it provides a smooth window without significantly slowing down the overall computation. The window size and sample rate are set and are not necessarily influenced by the photon statistics so should generalise well. To create a complete dataset the surface line is linearly interpolated to the initial ATL03 photon sample rate.

After generating the surface height estimates the absolute slope is calculated as $rise/run$ in which $rise$ is the height change over a 1 photon step size and run the along track distance covered in that step. The slope is used to calculate the slope-adjusted depth between the surface line and every single photon classified as a bottom return. The formula used, $D_{adj} = D\sqrt{1+m^2}$ with D_{adj} the slope adjusted depth, D the distance between the surface line and the bottom photo and m the surface slope, assumes the slope and eventual bottom line to be parallel. This slope adjusted depth calculation is used to correctly estimate the depth for potential lakes and streams on slanted surfaces. Any photon classified as bottom but within 0.2meter of the surface line is removed from the bottom photon class so to eliminate wrongly classified surface photons. Removing photons further below the surface would make the bottom classification to erratic and complicate its process in finding the correct lake boundary.

The depth of all photons classified as a bottom are corrected for the refraction difference between air and water, for light at 532nm, by dividing the depth by 1.33 [Parrish et al. (2019)]. This correction shifts all photons slightly upwards lowering the previous depth estimates. The constant used has shown great accuracy for solving for the physical implications of light travel time through both air and water in other research as well [Datta & Wouters (2021), Fricker et al. (2021)]. The constant is created for specifically the wavelength used by the ATLAS instrument aboard ICESat-2.

A distance based rolling median with a window size of 25m along-track distance and sample rate of 5m is calculated to extract a bottom line for the entire beam. The window size is chosen slightly higher than the surface so to correct better for the noisy bottom returns. To create a complete dataset the bottom line is linearly interpolated to the initial ATL03 photon sample rate.

2.3.3 Lake classification

Before the bottom line is interpreted as either a lake return or not, several statistics are calculated in a 25000 photon window. The statistics help to dismiss ICESat-2 returns under cloudy conditions, regions with insufficient signal in comparison to noise and regions with an inconsistent surface return in comparison to the bottom. For each window the total amount of surface, bottom and noise photons are counted. Two different ratios are calculated from the photon counts: Surface to Noise (SNR) shows how robust the surface detection is and Surface to Bottom (SBR) indicates whether or not enough bottom photons were detected to make a judgement on the depth. All photons in a window where the SNR is above 0.5 and $2.5 < SBR < 10$ are considered valid from a statistics standpoint. These parameters were chosen based on visual exploration and would need to be automatically parameterized

Table 2.2: Classes for the lake classification and their requirements

Class	D-adj	slope	statistics
lake	> 1	< 0.003	X
sloped lake	> 1	0.003	X
flat surface	1	< 0.003	X

to better scale this method to more ICESat-2 tracks. Further classification is done based on a combination of the surface slope and the slope adjusted depth measured as the difference between the surface and bottom depth estimates. All sections with a slope < 0.003 and a depth > 1 are considered lakes. Two extra classes exist for suspiciously flat areas and sloped regions with high depth as displayed in table 2.2. The per-photon classification is then transformed into a smooth classification with a distance based rolling mode at a window size of 10m and a sample rate of 10m. The 10m window is chosen so that all features larger than 5m can be classified. The sampling rate is the same as the window size to limit the computation time as the computational requirements for mode calculation are significantly larger than those for the median or mean. To create a complete dataset the lake classification result is propagated to the initial ATL03 photon sample rate.

2.3.4 Validation & extrapolation

To quantify the accuracy of the ICESat-2 depth estimates all results along the track were compared with both a physical and empirical method for depth retrieval using optical imagery [Box & Ski (2007), Sneed & Hamilton (2007)]. The empirical method can be used for extrapolation of the ICESat-2 results as shown in Datta & Wouters (2021) while the physical method is only used for validation. An extra effort was made to compare the physical and empirical method, which are both reliant on a single channel and lake by lake based parameterisations, to a multi-spectral machine learning approach.

For all 3 validation and extrapolation methods we use the Sentinel-2 data described in section 2.2.2 from the 2019-6-16 : 2019-6-17 overpass by Sentinel-2A. The B11 and B12 bands used in the machine learning approach have been resampled to a 10m resolution using a nearest neighbour algorithm. Additionally the enhanced Normalized Difference Water Index (NDWI) is calculated for every Sentinel-2 scene at a 10 meter resolution with equation 2.1. The enhanced NDWI does not use the Near Infra Red (NIR) and Short Wave Infra-Red (SWIR) band like the normal NDWI and has been used before to classify lakes in Greenland to great success[Pope et al. (2016), Moussavi et al. (2016), Moussavi et al. (2020)]. A threshold of 0.21 is used for the NDWI to identify lakes. This values provided the best visual classification results and is similar to the value used in the three aforementioned papers.

$$NDWI_{enhanced} = \frac{Blue - Red}{Blue + Red} \quad (2.1)$$

The Sentinel-2 data is then matched with the ICESat-2 dataset by extracting the Sentinel-2 pixel value at the longitude/latitude combination of every 25th photon in the ATL03

dataset. The computation time for this method is significantly lower than extracting data for every single photon and the resolution is not impacted as we can expect 0.6-3.9 photons per 0.7 m pulse Neumann et al. (2019) and Sentinel has a new pixel value for roughly every 10m of along-track distance. The gaps in the data set are then filled by forward filling the latest value.

Only a subset of the data is valid to be used for validation. First off the same statistics as leveraged in section 2.3.3 are used to eliminate parts of the track with insufficient or too noisy data. An additional set of statistics is used to eliminate geographical complex regions which can not be properly classified by the algorithm, this will be further dissected in the discussion. The statistics are calculated over the same window as the DBSCAN classification, which is 25000 photons. The first criteria is that the maximum height difference between any photon in the window and the GIMP DEM is smaller than 400 m, this removes any clear data inconsistencies due to clouds or geolocation errors. Second the range, the height difference between the highest and lowest photon, needs to be lower than 400 m. Lastly the mean slope in the region has to be lower than 0.1 eliminating very geographically intense regions like crevasses. All threshold values were chosen after visual detection of anomalies and extensive exploration of the statistics across the entire track. This statistical filtering decreases the size of the validation dataset by 6%. Additionally the validation dataset consists only of regions of the track for which Sentinel-2 data was available and lakes are present indicated by both a $NDWI > 0.21$ and a classification following table 2.2 as "lake".

Confusion matrix

The first method for validation is a confusion matrix. This sets out the truth, here considered to be the Sentinel-2 lake classification based on the NDWI threshold, against the expected. The expected is the lake classification as described in section 2.3.3 with only class "lake" as given in table 2.2 counting as a correct estimate. The only four options for a classification are a True positive (TP), a True Negative (TN), a False Positive (FP) and a False Negative (FN) which are all gathered on a per-photon basis. The confusion matrix is accompanied by the accuracy, sensitivity and precision metrics which are calculated from the photon counts per category. The accuracy is calculated as $(TP + TN) / (TP + FP + TN + FN)$ and gives an overall estimate for the classification success. To compensate for the skew in the categories two more metrics are calculated. The precision shows the percentage of classified lakes that are truly lakes, as seen by Sentinel-2, and is calculated as $TP / (TP + FP)$. The sensitivity indicates the reliability of the test for "no Lake" classifications and thereby gives an insight into how many hydrological features are missed. The sensitivity is calculated as $TN / (TN + FN)$. The four classes in the confusion matrix will be expressed as percentages of the total number of photons. The bottom depth generation in this thesis is a continuous estimation and not a classification problem. However, the confusion matrix provides valuable insight into, not only the classification of this bottom line, but also how it compares to the use of Sentinel-2 imagery for lake classification.

Empirical

The validation dataset is used to fit the empirical relation (Eq. 2.2) first described in Box & Ski (2007). Here D is the depth (m), R the TOA reflectance (unitless) in either the green or red spectrum and $\theta_0, \theta_1, \theta_2$ are the parameters to be tuned. The function is optimized for the Root Mean Square Error (RMSE) using the depth estimates and related reflectance from the validation dataset as training data. The same formula and method is used for both the green and red reflectance. The optimal parameters can be used to calculate lake depth from reflectance and thereby extrapolate the depth estimates over the entire Sentinel-2 image Datta & Wouters (2021). In this work no extrapolation beyond the ICESat track is done but the empirical relation is used as a benchmark when comparing to the machine learning approach in the results section. In contrast to common usage the method is not fitted on a per-lake basis but over all lakes in the entire track. This method is chosen as to show the generalisability of the method at scale.

$$D = \frac{\theta_0}{(R + \theta_1)} + \theta_2 \quad (2.2)$$

Physical

The single channel physical method (Eq. 2.3) as proposed in Philpot (1989) and later adapted for supraglacial lakes in Sneed & Hamilton (2007) and Pope et al. (2016) can provide results without in-situ measurements or the need for optimization. The method relies on the attenuation of light through water so that the reflectance varies with depth. The relation is non-linear and heavily dependent on multiple factors like the substrate albedo and pollution of the lake water. This physical approach requires one to extract parameters for each individual lake to scale and tune the method. In this thesis we aim to show that one can extrapolate lake depths using a generalised method hence why a similar approach to the empirical method is used (as described in section 2.3.4 to fit the physical method for the varying parameters instead. In below equation D is the lake depth in m, A_d is the substrate albedo, R_w is the TOA reflectance of the water and R_{inf} is the TOA reflectance for optically deep water ($> 40m$). Values for g are taken from Williamson, Banwell, Willis & Arnold (2018) as $g_{green} = 0.1413$ and $g_{red} = 0.8304$ while the value for $R_{inf} = 0.04$ is taken from Moussavi et al. (2020). These parameters were set to these referenced values as they were found as a general suitable value for large datasets over the Greenland coast and should apply well to the study area of this thesis. For A_d the value was estimated between 2 bounds from Moussavi et al. (2016) and then found via RMSE optimization for the validation data. The search bounds for the parameter are set as: Green 0.3 – 0.7 and red 0.15 – 0.6 so to make sure the optimal value is within. A_d is computed this way to remove the need for tuning over individual lakes and show the generalisability of the physical method as a benchmark. The physical method is not intended to be generalised over different lakes but should still provide a constant comparison for the ICESat-2 retrievals and the machine learning approach.

$$D = \frac{\ln(A_d - R_{inf}) - \ln(R_w - R_{inf})}{g} \quad (2.3)$$

Multispectral machine learning

ICESat-2 depth estimates can be used to extrapolate lake depths to lake volumes using optical imagery and the empirical method as shown by Datta & Wouters (2021). However, this method requires tuning on a lake by lake basis making it unusable for smaller features or widespread extrapolation. The empirical method relies on only one band of optical imagery which removes the possibility for it to generalise for the lake parameters like substrate albedo, surface roughness and water contents. By using machine learning and a multispectral dataset (Red, Green, Blue, NIR, SWIR [-2 bands] and the NDWI) parameterisation for lake differences may be internalized and no longer needed to be tuned for. A random forest regression algorithm is used with 250 estimators. The validation dataset is slightly altered so to only include unique combinations of the reflectance over all bands. This is done by grouping the data per unique reflectance combination and taking the median depth for this set. This is done to prevent overfitting of the model as there are many photons geolocated to each single sentinel-2 pixel. The Sentinel-2 bands B02, B03, B04, B08, B11, B12 and the NDWI at 10 meter resolution are used as training data for the model. To estimate the performance of the model a cross-validation process is followed with 25 folds and a test dataset size of 15%. The cross-validation is important to show the accuracy for out of sample results, how would the model react to new data. The chosen fold size matches the requirements for this dataset. A smaller fold size would have too much variance in the sample test datasets and a larger value would exclude too much data so that the model can't train. The same out of sample test is applied to the empirical method for which the results are given in figure 3.7a.

Results

3.1 Photon classification

Figure 3.4 shows the output for the surface classification alongside the parameterisation for the Eps parameter. The classification result does not discriminate lakes already but highlights photons part of the surface (grey) and photons part of the bottom signal (blue). The return is constant meaning every photon could be either noise, bottom or surface ; no other restrictions are set. However, this does also mean that the bottom and surface classifications are never truly noise free and there is a trade-off in eliminating noise and retaining signal. In the top right of 3.4 it is visible that the first Eps parameterisation is done in such a way that 85% of the photons have at least 6 other photons around them in a circle with radius $r = Eps$. With an Eps of 1.838 m this step is eliminating all noise from the dataset as can be seen in the top left image of figure 3.4.

The second classification step is taken more strictly so to extract just the surface photons which exhibit a stronger density than the bottom photons. Roughly 80% of the photons share enough neighbours to be a core point. The bottom classification shown in the bottom left of figure 3.4 indicates 2 gaps between the surface photon cluster and the bottom photon cluster. We might identify these features as lakes as they have a distinct shape and disconnect with the surface; this is something the method further highlights in the next processing step. Figure 6.11 in the Appendix shows an additional three lakes including their parameterisation.

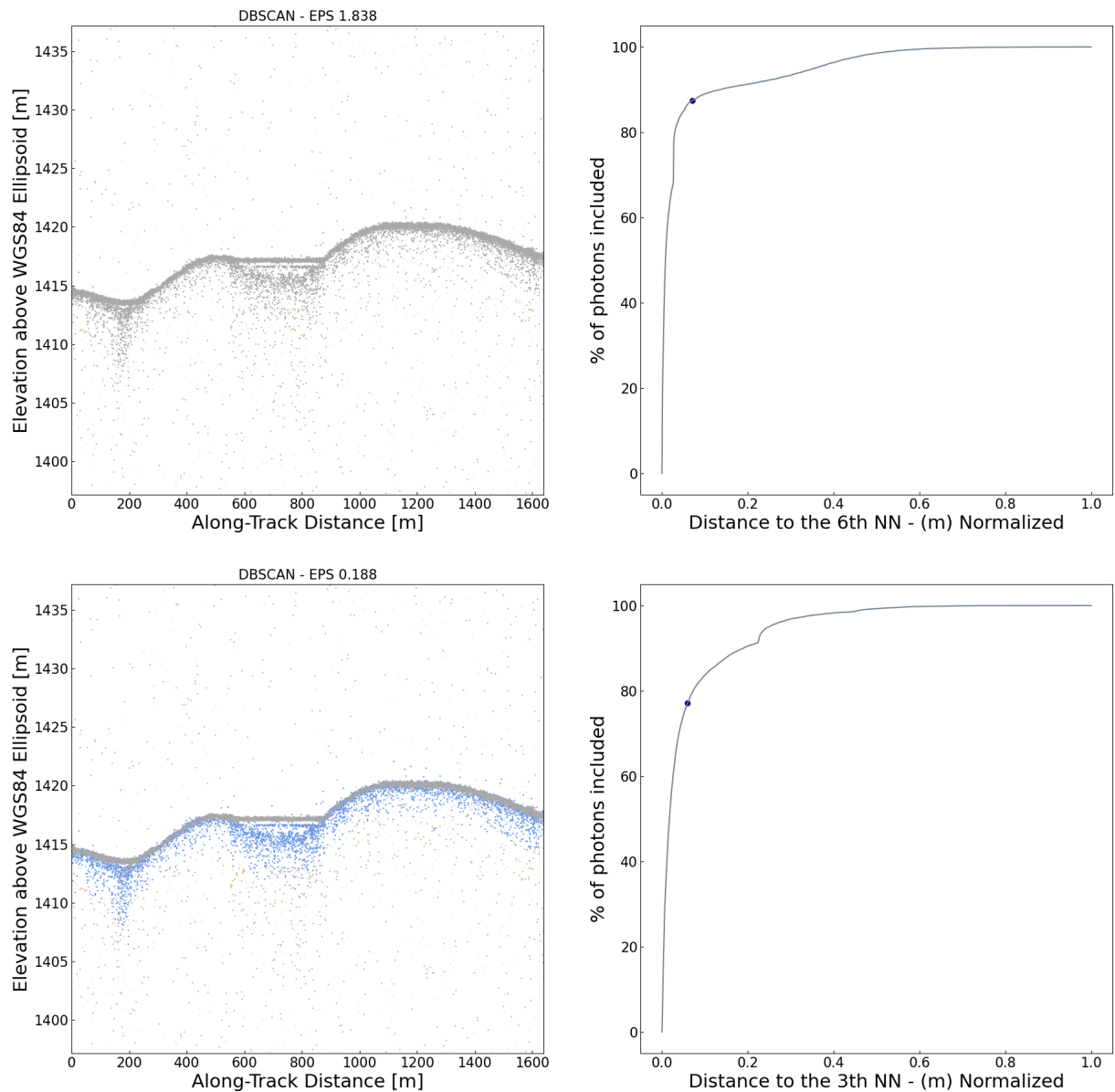


Figure 3.4: Left – Classification of Surface photons (grey) and bottom photons (blue) by DBSCAN for step 1 (top) and step 2 (bottom) plotted with the height on the y-axis and along track distance on the x-axis; both in meters. Right – Graph showing the increase of core points for DBSCAN (x-axis) as the EPS would increase (y-axis), the dot shows the automatically chosen value.

3.2 Surface and bottom extraction

The surface and bottom lines as extracted from the photon classification, using the method displayed in section 3.1, are detailed in Figure 3.5 together with a scaled image of the lake from Sentinel-2 imagery. The method classifies lakes on various scales from just a few meters wide and deep (bottom right) to kilometer scale lakes deeper than 15 m (top right). The method shows very consistent results in a native resolution and the classification visually aligns with the classification by Sentinel-2. The method is not able to identify the water

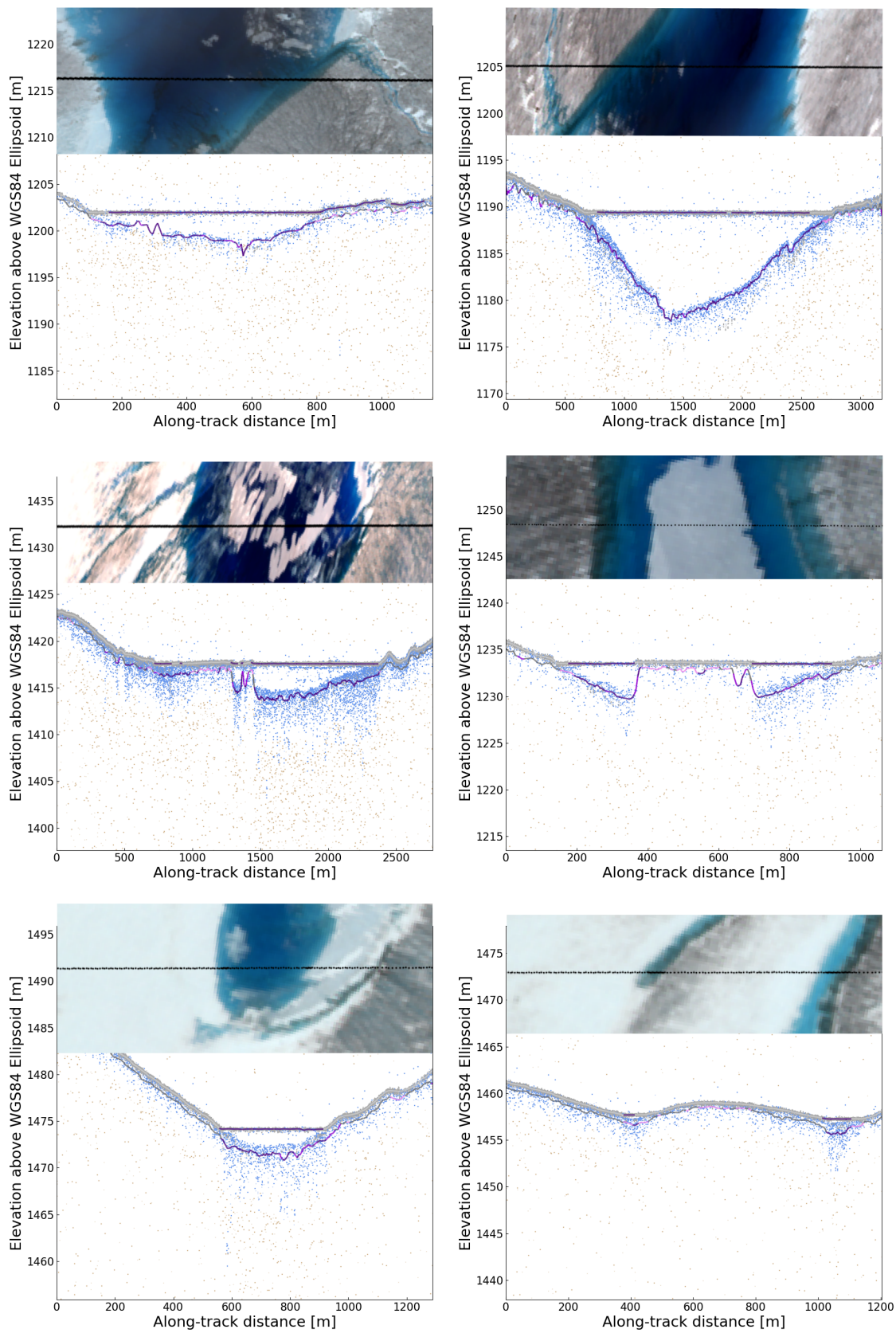


Figure 3.5: Figures showing the individual photon returns classified as surface (grey) or bottom (blue), the surface line (grey) which shows purple when Sentinel-2 $NDWI > 0.21$ and the bottom line which is purple for lake, violet for sloped lake and violet for flat surface. RGB images from sentinel-2 at the top of each image show the lakes at scale.

depth under floating ice but does correctly identify floating ice as surface instead, which can be seen in the middle two images in figure 3.5. The surface line is very detailed with very small scale flat surfaces and geographical features showing while also being smooth over the entire window with little erratic vertical movements. The bottom line is more erratic and very dependent on the amount of bottom photons available for its calculation. The bottom line clearly indicates the start and ending of lakes with minimal deviation from the slope break points visible in the surface line. The bottom line perfectly follows the bottom photon returns for dense returns (top right Figure 3.5) but for lakes with more inconsistent bottom photon returns (top left Figure 3.5 the bottom line can act erratic at times. Apparent multiple surface returns [Martino et al. (2020), Lu et al. (2021)] as shown in the middle left image in figure 3.5 do not materially impact the bottom line return and bottom classification. In general some photons at the surface being classified as bottom photons does not impact the visual correctness of the bottom line return. However, an inconsistent surface line over a lake, as can be seen in the bottom left image of figure 3.5, has an impact on the slope and therefore impacts the bottom classification. Misclassification due to inconsistent slope returns as "sloped lake" is the most common classification error as will be discussed in section 3.3. The bottom line is a consistent return and has a minimum depth of 0.2 m due though the line creation process as described in 2.3.2. On top of this the classification of lakes requires a minimum depth of 1 m which makes it so the all classified lakes have a minimum depth of 1 m even for the smallest lakes as can be seen in the bottom right of Figure 3.5.

3.3 Lake classification

The lake classification is performed with an accuracy of 97.8 % when compared to the Sentinel-2 NDWI threshold classification. The distribution of lakes is skewed towards no lakes which is why the precision and sensitivity are used as additional metrics. The ratio between correctly identified lakes and sections identified as lake, but not showing as lake on the Sentinel-2 images, is given by the precision at 49.6 %. Table 3.3 also gives the sensitivity at 72.2 % meaning the classification misses 27% of the total photons identified as lakes by Sentinel-2. The Sentinel-2 NDWI threshold can not be seen as the whole truth due to resolution disparities and misclassification in the NDWI method but provides a good insight into the overall performance of the lake classification. The confusion matrix shows that the lake classification does classify more photons as lakes than Sentinel-2 and misses relatively little lakes in comparison to the Sentinel-2 method. The sensitivity jumps to 96% when "sloped lake" is also counted as a positive classification, a significant increase. For this second scenario the precision stays constant at 50% indicating no loss in overall accuracy.

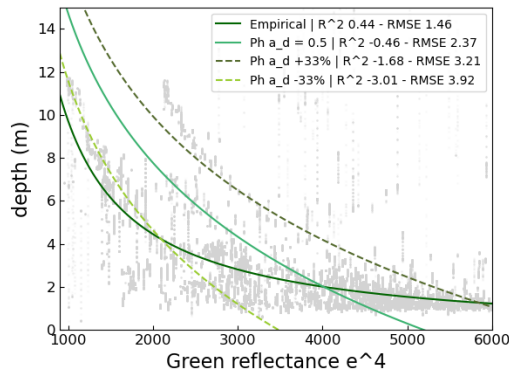
Table 3.3: Confusion matrix for the classification result

Class	Lake (ICESat)	no Lake (ICESat)	
Lake (S2)	1.61 %	0.62 %	Sensitivity: 72.2 %
no Lake (S2)	1.63 %	96.15 %	
	Precision: 49.6 %		Accuracy: 97.8 %

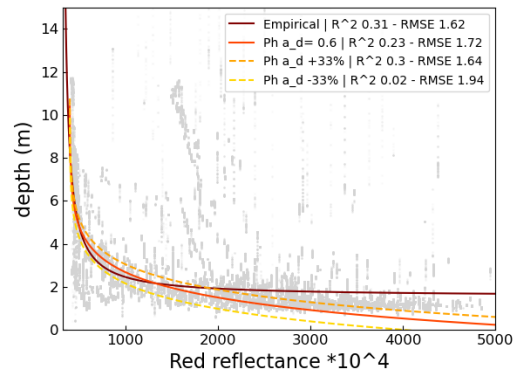
3.4 Lake depths

3.4.1 Assessment of the fit

The fitted relation for the empirical and physical method as described in 2.3.4 are shown in figure 3.6a and figure 3.6b with the background showing the validation dataset. As reflectance decreases the depth increases for both the green and red relation. The empirical relations have a significantly better fit, with Pearson correlation coefficients of 0.44 and 0.31 for the green and red band respectively, than the physical relations. The sensitivity of fit to the substrate albedo parameter A_d is shown with the dashed lines. The shape of the physical relation does not seem to fit the overall shape of the estimated ICESat-2 depths from the validation dataset. The bad fit of the physical relationship is expressed in the negative R^2 values indicating a horizontal line is a better fit to the overall dataset. The relation for the red band has a more non-linear shape than the green reflectance with depth only significantly increasing below a reflectance of 0.1. The non-linearity lowers the predictive capabilities of the model at lower reflectances. Both the green and red relation have a clear set of points which do not fit the relations as they show higher depths for a higher reflectance than expected. The physical method predicts negative depths for reflectances larger than 0.51 for both the green and red relation which is a non-physical solution.



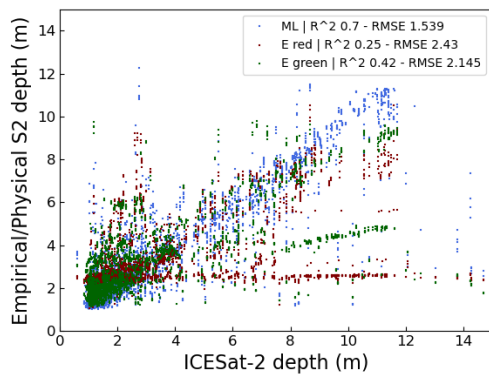
(a) Plot showing the ICESat-2 depth as grey scatters, the fitted empirical and optimal physical relation in darkgreen and seagreen respectively and two alternative physical fits as dashed lines in olive and darkolive.



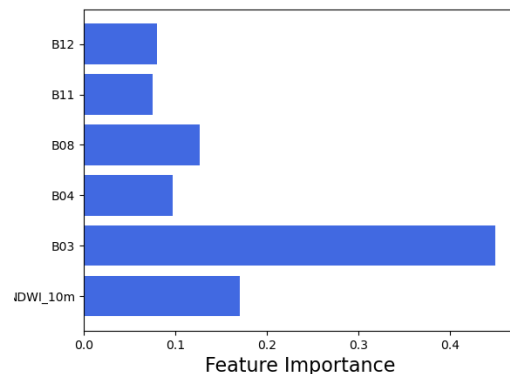
(b) Plot showing the ICESat-2 depth as grey scatters, the fitted empirical and optimal physical relation in maroon and orange respectively and two alternative physical fits as dashed lines in yellow and gold.

3.4.2 Out-of-sample test

To further assess the quality of the depth estimates the empirical and physical relation are tested out-of-sample for which the depths are given in figure 3.7a. The same test is applied to the machine learning approach as laid out in section 2.3.4. The figure shows only the empirical relations for both green and red and the machine learning approach as the physical relation gives non-physical results and does not generalise well. Extrapolation of the ICESat-2 depths across an entire track is feasible with a RMSE of 2.15, 2.43 and 1.54 meters for the green empirical, red empirical and machine learning approach respectively. The machine learning approach holds a R^2 value of 0.7 indicating a medium strong correlation between



(a) Estimated vs expected lake depth tested out of sample for the Green (Darkgreen) and Red (Maroon) Empirical method and the Machine learning (ML) approach (Blue).



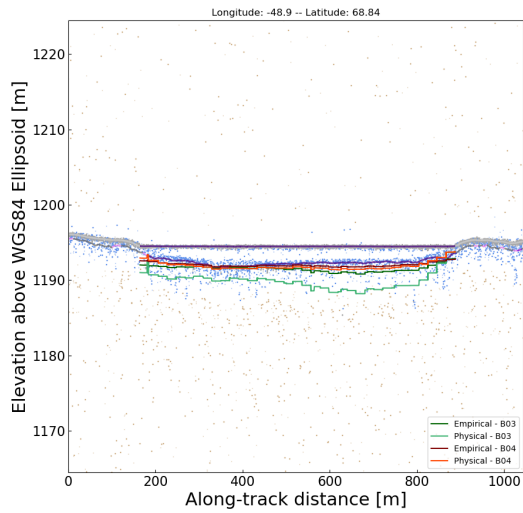
(b) The importance of the different input bands for the Machine Learning approach visualized as a bar graph.

the ICESat-2 depths and the estimated depths from Sentinel-2 imagery. The green and red empirical relations show no depth estimates deeper than 10 m and underestimate the depth for a continuous array of samples as can be seen in the bottom right quadrant of figure 3.7a. all three comparisons show a narrowing effect of their distribution towards deeper depths indicating estimating depth of shallow lakes is more ambiguous than deeper lakes.

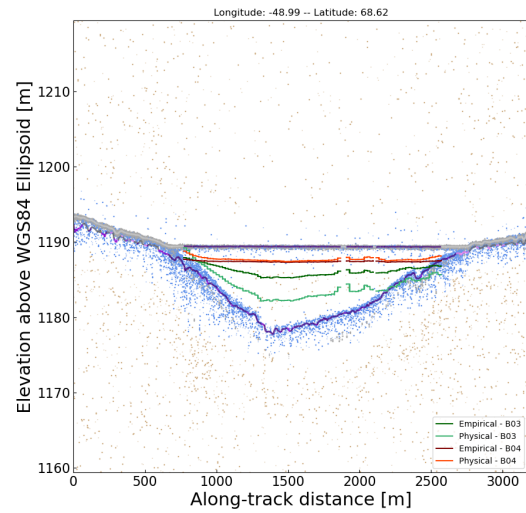
Some extra insight into the workings of the machine learning method is given in figure 3.7b. The figure shows that the green reflectance is the feature that contributes the most information to the predictive power of the machine learning method. The enhanced NDWI is the second most important feature followed by the NIR band.

3.4.3 Relations applied

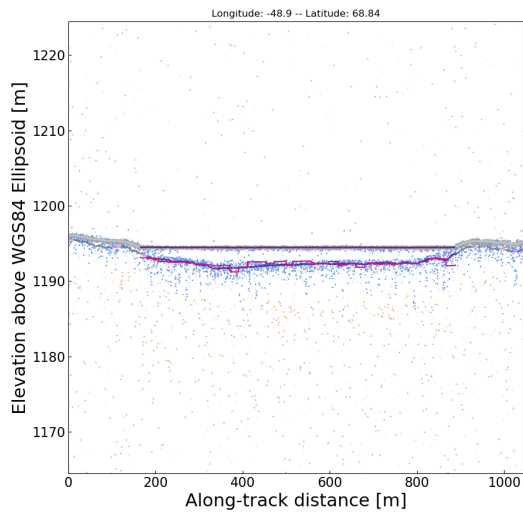
Figures 3.8a and 3.8b show an out of sample estimation of the lake depth retrieved from the sentinel 2 images using the empirical and physical method for both the green and red band. Both methods utilizing the red bands underestimate the depth of the lakes which is particularly noticeable in figure 3.8a. The methods utilizing the green band are closer to the ICESat-2 depth for the deeper lake but overestimate the depth of the shallow lake. The result for all four methods is only visible for regions indicated as lake by the Sentinel-2 images which makes the depth retrievals non-continuous. All four methods provide a bottom line which moves in steps as its dependent on the Sentinel-2 resolution and was not smoothed. Figures 3.8c and 3.8d show the same out of sample estimation but then for the machine learning method laid out in section 2.3.4. The bottom classification shows the same non-continuous structure and stepped height as bottom depth estimated from the empirical and physical relation. However, the depth estimates are much closer to the ICESat-2 depth estimates. The two example lakes have distinct differences in depth which are both captured by the trained machine learning model.



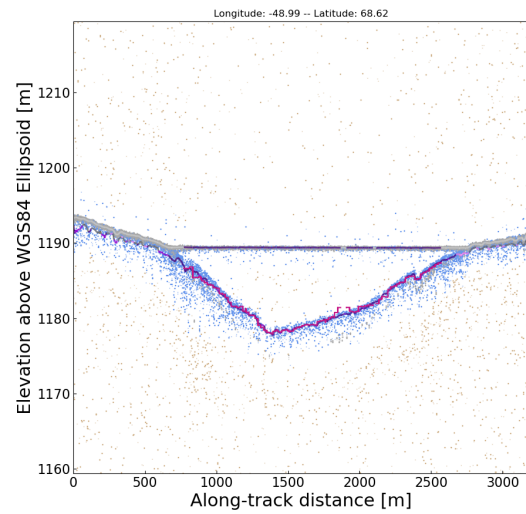
(a) Lake 1 - Empirical and physical relations



(b) Lake 2 - Empirical and physical relations



(c) Lake 1 - Machine learning



(d) Lake 2 - Machine learning

Figure 3.8: Out of sample results for the an empirical, physical and machine learning approach to the extrapolation of ICESat-2 lake depth estimates. The estimated depth to be emulated by the models is shown as the purple bottom line. The empirical and physical methods are displayed as shades of green and red in figure (a) and (b). The result from the machine learning model is visualized as a pink line in figures (c) and (d)

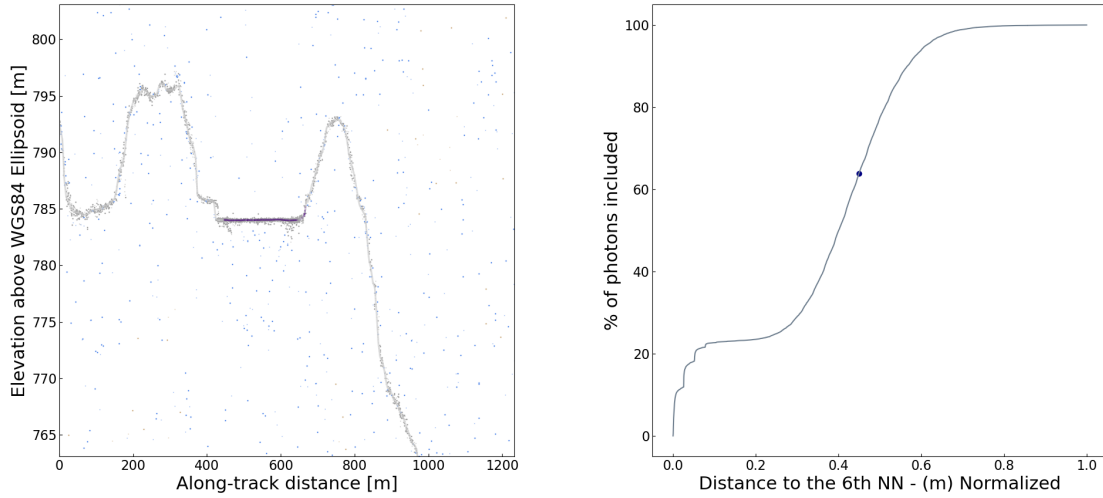
Discussion

4.1 Photon classification

The two phase clustering approach for the photon classification has delivered visually consistent and correct results. The background noise is removed consistently, leaving a very clear surface reflection and bottom separation for regions with sufficient bottom photons. However, a better screening of valid regions is needed prior to classification, besides the post classification filter, so that regions with insufficient or inconsistent photon returns can be removed or correctly parameterized for. Figure 4.9a shows an example of a non-valid region as the surface return is very light in comparison to the number background noise photons with roughly 20% of the photons being part of the displayed surface. The surface is still correctly identified but a wrongful Eps parameterisation leaves all background noise photons classified as bottom returns. Solutions could be: Identifying sparse surface returns after classification and removing all associated bottom returns, implementing a $maxEps$ parameter for step 1 of the classification, changing the automated Eps parameterisation so that it identifies breakpoints in the optimisation graph and acts on them or lastly alter the normalisation of the Nearest Neighbour (NN) distance in the Eps parameterisation to give more magnitude to the lower NN distances. Solving this parameterisation of Eps for varying statistical scenarios will also enable further generalisation of the method so that weak beams and overpasses during the night can be used.

Apparent multiple surface returns are common in many classified lakes and are often classified as either surface or bottom returns. The second implicated surface, most often, consists of less photons than the actual surface or the bottom. This means it is not picked up by the window based median and does not impact the result. This is in contrast to histogram based methods which will have to eliminate these returns from the histogram to implicate the correct surface or bottom.

Something else to take into account is that DBSCAN only operates on a two dimensional dataset with the photon height and along track distance, no weight is given to the across track distance. The across track distance is not constant meaning two photons with the same or very similar along track distances might still originate from locations several meters apart. A lake is a 3D feature so the height of these two photons might vary quite a lot, dependent on the type of hydrological feature. These 3D depth variations are represented as depth differences in 2D in the current method which distorts the 2D view of the data and impacts the classification. The best way to tackle this is introduce the across track distance as a third dimension so that the DBSCAN algorithm can take it into account when estimating the similarity. From small visual tests done early in the creation of this thesis the impact on the classification seemed to be little but the computational requirements were significantly higher. Further research on this topic would be needed quantify the impact of adding the across track distance as a third dimension to both the result and the overall computation time.



(a) 2D plot showing the estimated lake surface line (purple when classified as lake by S2), and the classified SCAN (x-axis) as the Eps would increase (y-axis), the photon scatters (grey for surface and blue for bottom). dot shows the automatically chosen value.

4.2 Surface and bottom retrieval

The photon classification is important but provides no real interpretable result as the photons returns are still chaotic and manual bottom retrieval from this would be based on arbitrary decisions. The median over a distance based window has shown as a good way to generalise the classification to a return at every sampling point. The smoothing window for both the surface and the bottom however has an impact on the resolution of the surface and bottom retrieval. On top of this it is hard to estimate the best smoothing window for the lines. Increasing the window will generate a more consistent return but will further reduce the possibility to identify intricate surface and bottom geometry. However, reducing the window size to a minimal size is not optimal either as at a certain point the window is so small that it is back on the photon level, meaning no generalisation has happened. On top of this the window size can not be smaller than the sample rate as data would be omitted further hindering an optimal estimate. Every halving of the sample rate doubles the computation time of this step in the method. This means decreasing the window size can also lead to significant increase in computation time. In this thesis a choice is made to sample the bottom at a the lowest possible window size so to catch the smallest hydrological features while the surface is smoothed over a smaller distance window but with much more photons so to make sure it is as consistent as possible. As without a good surface return the bottom return is useless because no correct depth can be calculated.

Another factor to consider is that the smoothing window for the surface and bottom requires sufficient signal photons within the size of each window or it wont return any value. In figure 4.10a a sudden depression in the surface estimate can be seen. This estimated depression is not a correct representation of the surface but a drop in surface returns made

the distance based window estimate very volatile, picking the bottom as a signal instead. If gaps are even larger and/or no bottom photons are accidentally classified as surface, then the line would have a break instead. This means there would be no way to calculate the depth at that location. A wrong depth estimate can have knock on effects in the model training for extrapolation, so having no data can be considered better than wrong data.

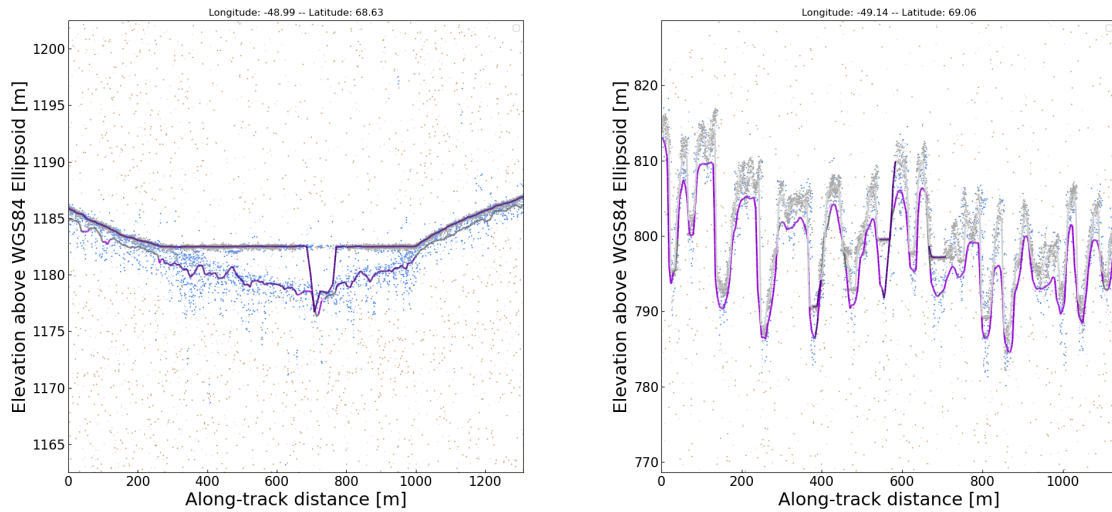
The method always calculates both a surface line and a bottom line. For most regions it can very well identify whether or not this bottom return is actual signal or noise. For intense geographical regions however, like crevasses or steep slopes, the classification has a harder time. In figure 4.10b we see a region where the surface slope is changing so quickly that the estimate is trailing the surface. This violet colour of the line indicates the two estimates are more than 1 meter apart for most of the region. However, lakes in these regions are mostly correctly identified. This is because if crevasses are water filled low slope is still detected. The classification distance window might not be perfectly aligning with the crevasse bottom though, which may make it so the next slope is classified as a lake as well. Some of these features are smaller than the bottom classification window can manage.

Overall the easiest way to reduce the trailing bottom line problem, is by setting the bottom height to the surface height for every section that is not classified as a lake. A small smoothing window might be needed at the edges of each lake. This solution would also remove the minimum lake depth of 1 meter as the bottom line at the edges of a lake bent better towards the surface improving the training for the extrapolation model. A closed of lake also better reflects the reality as the lake depth is considered 0 at the lake edge.

Lastly on the surface and bottom estimates, it is important to mention that there is a small assumption in the calculation of the lake depth which can cause artifacts. As described in section 2.3.2, the depth is calculated assuming the surface and bottom are parallel. This makes it so lakes that exist on a sloped surface might have their depth misrepresented. These cases are not abundant, as lakes on sloped surfaces still have a flat surface return. Shallow surface streams on a sloped surface however do not fit this criteria. Overall for future research it needs to be considered that the way depth is calculated provides a bias to the classification of hydrological features.

4.2.1 Validation

The validation used in this thesis aimed to provide both validation and generalised extrapolation at the same time. It is clear from, especially the results for the physical validation, that this was not feasible at the same time. With the current methods and lack of in-situ measurements there is no way to properly validate the accuracy of the ICESat-2 depths only from this research. Further research is needed to compare and validate the method proposed in this thesis against other unsupervised methods and in-situ measurements, like done in Fricker et al. (2021). Only the empirical and machine learning method showed any sign of generalisation. From the physical method the only conclusion that could be taken is that the ICESat-2 depths are very inaccurate which does not seem plausible given the high accuracy of ICESat-2 in other research. The proper validation method would have been to scale the physical relation for individual lakes, but even this can't be considered a ground truth to



(a) 2D plot showing the estimated lake bottom line (purple when classified as lake), lake surface line (purple when classified as lake by S2), and the classified photon scatters (grey for surface and blue for bottom). (b) 2D plot showing the estimated lake bottom line (purple when classified as lake), lake surface line (purple when classified as lake by S2), and the classified photon scatters (grey for surface and blue for bottom).

properly validate a model with. For now we assume the ICESat-2 depth estimates to be close to the potential ground truth as the ATL03 product comes from a clear physical relation and has proven to be accurate from previous research [Ma et al. (2020), Fricker et al. (2021)].

The lake classification does have strong validation data, as it was compared to a Sentinel-2 NDWI thresholding mechanism which is considered an accurate standard widely used for lake classification [Moussavi et al. (2020)]. The confusion matrix indicates that the lake classification misses only few lake pixels in comparison to Sentinel-2 but that it overestimates the amount of lake pixels compared to Sentinel-2. The confusion matrix is an aggregate statistic over the entire track consisting of more than 22 million samples. The difference noted with the Sentinel-2 classification does not have to mean the ICESat-2 method is inherently incorrect, as it has a significantly higher resolution meaning it could have identified lakes Sentinel-2 can't have seen. The Sentinel-2 and ICESat-2 images were taken 12 hours apart which can already have a material difference on highly variable environments like an ice sheet ablation zone. Features, like floating ice for example, will be different in both images impacting the confusion matrix. On top of this the Sentinel-2 data was not checked filtered for pixels with clouds. The cloud percentage on all the used images was low and no clouds were apparent over the ICESat-2 track. However this could still have an impact as Sentinel-2 misses lakes and this reflecting badly on the accuracy of the ICESat-2 classification. Besides this it is important to note that regions with denser photon returns have a higher impact on the confusion matrix as it was calculated on a per-photon basis. For more consistent validation it would have been better to sample the lake classification at a set along-track distance resolution and compare that to the Sentinel-2 classification instead. Overall the classification does seem to be relatively accurate and properly reflect floating ice, very small hydrological features and the boundaries of supraglacial lakes.

4.2.2 Extrapolation

The physical method generalises poorly overall which makes it unusable for calculating depth estimates at scale. The empirical method generalises relatively well and has some predictive capabilities. However, the relation operates at a much higher accuracy when used on a lake by lake basis as shown in Datta & Wouters (2021). The proposed machine learning method provides significantly better results than both the empirical and physical relation, which shows in the out of sample results as displayed in figure 3.8c and 3.8d. It is important to note however that the photon based dataset caused complications for the model training. Like mentioned before in 4.2.1 re-implementing some of the method to return a dataset sampled at a constant distance would mean the validation data better represents the actual distribution of the various hydrological features. If this solution is implemented, more data would be retained. This is because implementing a filter to keep only the unique values for the model training, what is done now, would no longer be needed. Relatively little data (2000 unique rows) was used in this thesis to train the machine learning model as all validation data came from the 1222 ICESat-2 track. Even less could be used to create the out of sample prediction displayed in figure 3.8c and 3.8d as the entire beam which hold that lake has to be omitted from the training. Overall I expect the generalisation capabilities of the model to significantly improve when its trained on more than one track. Figure 3.8d is also displayed in Datta & Wouters (2021) so a small comparison can be made. Both the empirical relations for the green and red band shown in the Watta algorithm paper better estimate the depth than the more generalised empirical relation found in this paper. The lake bottom found in this thesis is very similar to the Watta retrieved depth estimates and the machine learning method proposed in this thesis seem to outperform both of the empirical relations displayed in Datta & Wouters (2021).

Conclusion

This thesis presents a new method for supraglacial lake depth and surface retrieval from the ICESat-2 ATL03 product leveraging an automated parameterisation of DBSCAN clustering. The proposed method provides a consistent and visually accurate surface estimate, similar to the ATL06 product, as well as a continuous lake depth estimate with a lake classification. The method is able to detect lakes and other surface hydrological features as small as 10-20 meters at a 97.8 % accuracy, compared to the traditional NDWI lake classification, and matches very well with optical images taken of the study area. The depth estimates retrieved in this thesis are hard to validate but show a very consistent return amongst a wide variety of lake types. The depth estimates are not materially impacted by apparent multiple surface returns, indicate floating ice and show no sign of loss of predictive capability with depth. Overall the proposed method operates under relatively little computational overhead, especially the photon classification, and provides depth estimates at a native vertical resolution from only one data-source over the entirety of the Greenland Ice sheet. Beyond the depth estimates a machine learning method is proposed to extrapolate the depth estimated from the ICESat-2 imagery leveraging optical imagery by Sentinel-2. An out-of-sample test shows that the method is able to predict lake depth with a RMSE of 1.54 meters and a R^2 value of 0.7. The method can be used to estimate supraglacial lake depths, in a generalized setting for other ablation areas; and do so with a higher accuracy than the generalised version of the traditional empirical and physical relations for depth estimates. The automated parameterisation method for DBSCAN clustering laid out in this case study needs to be further advanced so that it can be used on the weak ICESat-2 beams and in situations with more noisy photon returns. Advancements are warranted to make the method more efficient and return results only for areas with classified lakes while clearly highlighting the lake boundaries. Lastly more work is needed to validate the retrieved lake depth estimates with ground truth, something that was not feasible in this research.

Bibliography

- Bamber, J. L., Westaway, R. M., Marzeion, B. & Wouters, B. (2018), ‘The land ice contribution to sea level during the satellite era’, *Environmental Research Letters* **13**(6), 063008.
- Bartholomew, I., Nienow, P., Sole, A., Mair, D., Cowton, T. & King, M. A. (2012), ‘Short-term variability in greenland ice sheet motion forced by time-varying meltwater drainage: Implications for the relationship between subglacial drainage system behavior and ice velocity’, *Journal of Geophysical Research: Earth Surface* **117**(F3).
- Bell, R. E., Banwell, A. F., Trusel, L. D. & Kingslake, J. (2018), ‘Antarctic surface hydrology and impacts on ice-sheet mass balance’, *Nature Climate Change* **8**(12), 1044–1052.
- Box, J. E. & Ski, K. (2007), ‘Remote sounding of greenland supraglacial melt lakes: implications for subglacial hydraulics’, *Journal of glaciology* **53**(181), 257–265.
- Datta, R. T. & Wouters, B. (2021), ‘Supraglacial lake bathymetry automatically derived from icesat-2 constraining lake depth estimates from multi-source satellite imagery’, *The Cryosphere* **15**(11), 5115–5132.
- Dell, R., Arnold, N., Willis, I., Banwell, A., Williamson, A., Pritchard, H. & Orr, A. (2020), ‘Fully automated supraglacial-water tracking algorithm for ice shelves (fastish).’.
- Edwards, T. L., Nowicki, S., Marzeion, B., Hock, R., Goelzer, H., Seroussi, H., Jourdain, N. C., Slater, D. A., Turner, F. E., Smith, C. J. et al. (2021), ‘Projected land ice contributions to twenty-first-century sea level rise’, *Nature* **593**(7857), 74–82.
- Enderlin, E. M., Howat, I. M., Jeong, S., Noh, M.-J., Van Angelen, J. H. & Van Den Broeke, M. R. (2014), ‘An improved mass budget for the greenland ice sheet’, *Geophysical Research Letters* **41**(3), 866–872.
- Ester, M., Kriegel, H.-P., Sander, J., Xu, X. et al. (1996), A density-based algorithm for discovering clusters in large spatial databases with noise., *in* ‘kdd’, Vol. 96, pp. 226–231.
- Fair, Z., Flanner, M., Brunt, K. M., Fricker, H. A. & Gardner, A. (2020), ‘Using icesat-2 and operation icebridge altimetry for supraglacial lake depth retrievals’, *The Cryosphere* **14**(11), 4253–4263.
- Fricker, H. A., Arndt, P., Brunt, K. M., Datta, R. T., Fair, Z., Jasinski, M. F., Kingslake, J., Magruder, L. A., Moussavi, M., Pope, A. et al. (2021), ‘Icesat-2 meltwater depth estimates: Application to surface melt on amery ice shelf, east antarctica’, *Geophysical Research Letters* **48**(8), e2020GL090550.
- Gallagher, M. R., Chepfer, H., Shupe, M. D. & Guzman, R. (2020), ‘Warm temperature extremes across greenland connected to clouds’, *Geophysical Research Letters* **47**(9), e2019GL086059.

- Götz, M., Kononets, M., Bodenstern, C., Riedel, M., Book, M. & Palsson, O. P. (2019), ‘Automatic water mixing event identification in the koljö fjord observatory data’, *International Journal of Data Science and Analytics* **7**(1), 67–79.
- Howat, I. M., Negrete, A. & Smith, B. E. (2014), ‘The greenland ice mapping project (gimp) land classification and surface elevation data sets’, *The Cryosphere* **8**(4), 1509–1518.
- Hsu, H.-J., Huang, C.-Y., Jasinski, M., Li, Y., Gao, H., Yamanokuchi, T., Wang, C.-G., Chang, T.-M., Ren, H., Kuo, C.-Y. et al. (2021), ‘A semi-empirical scheme for bathymetric mapping in shallow water by icesat-2 and sentinel-2: A case study in the south china sea’, *ISPRS Journal of Photogrammetry and Remote Sensing* **178**, 1–19.
- Khan, K., Rehman, S. U., Aziz, K., Fong, S. & Sarasvady, S. (2014), DbSCAN: Past, present and future, *in* ‘The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014)’, IEEE, pp. 232–238.
- Legleiter, C., Tedesco, M., Smith, L., Behar, A. & Overstreet, B. (2014), ‘Mapping the bathymetry of supraglacial lakes and streams on the greenland ice sheet using field measurements and high-resolution satellite images’, *The Cryosphere* **8**(1), 215–228.
- Li, Y., Gao, H., Jasinski, M. F., Zhang, S. & Stoll, J. D. (2019), ‘Deriving high-resolution reservoir bathymetry from icesat-2 prototype photon-counting lidar and landsat imagery’, *IEEE Transactions on Geoscience and Remote Sensing* **57**(10), 7883–7893.
- Lu, X., Hu, Y., Yang, Y., Vaughan, M., Palm, S., Trepte, C., Omar, A., Lucker, P. & Baize, R. (2021), ‘Enabling value added scientific applications of icesat-2 data with effective removal of afterpulses’, *Earth and Space Science* **8**(6), e2021EA001729.
- Lüthje, M., Pedersen, L. T., Reeh, N. & Greuell, W. (2006), ‘Modelling the evolution of supraglacial lakes on the west greenland ice-sheet margin’, *Journal of Glaciology* **52**(179), 608–618.
- Ma, Y., Xu, N., Liu, Z., Yang, B., Yang, F., Wang, X. H. & Li, S. (2020), ‘Satellite-derived bathymetry using the icesat-2 lidar and sentinel-2 imagery datasets’, *Remote Sensing of Environment* **250**, 112047.
- Magruder, L., Brunt, K., Neumann, T., Klotz, B. & Alonzo, M. (2021), ‘Passive ground-based optical techniques for monitoring the on-orbit icesat-2 altimeter geolocation and footprint diameter’, *Earth and Space Science* **8**(10), e2020EA001414.
- Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U. & Gascon, F. (2017), Sen2COR for sentinel-2, *in* ‘Image and Signal Processing for Remote Sensing XXIII’, Vol. 10427, SPIE, pp. 37–48.
- Martino, A., Field, C. T. & Ramos-Izquierdo, L. (2020), ‘Icesat-2/atlas instrument linear system impulse response’.

- Mouginot, J., Rignot, E., Bjørk, A. A., Van den Broeke, M., Millan, R., Morlighem, M., Noël, B., Scheuchl, B. & Wood, M. (2019), ‘Forty-six years of greenland ice sheet mass balance from 1972 to 2018’, *Proceedings of the national academy of sciences* **116**(19), 9239–9244.
- Moussavi, M., Pope, A., Halberstadt, A. R. W., Trusel, L. D., Cioffi, L. & Abdalati, W. (2020), ‘Antarctic supraglacial lake detection using landsat 8 and sentinel-2 imagery: Towards continental generation of lake volumes’, *Remote Sensing* **12**(1), 134.
- Moussavi, M. S., Abdalati, W., Pope, A., Scambos, T., Tedesco, M., MacFerrin, M. & Grigsby, S. (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.
- Neumann, T. A., Martino, A. J., Markus, T., Bae, S., Bock, M. R., Brenner, A. C., Brunt, K. M., Cavanaugh, J., Fernandes, S. T., Hancock, D. W. et al. (2019), ‘The ice, cloud, and land elevation satellite–2 mission: A global geolocated photon product derived from the advanced topographic laser altimeter system’, *Remote Sensing of Environment* **233**, 111325.
- Neumann, T., Brenner, A., Hancock, D., Robbins, J., Saba, J., Harbeck, K., Gibbons, A., Lee, J., Luthcke, S. & Rebold, T. (2021), ‘Atlas/icesat-2 l2a global geolocated photon data, version 5. boulder, colorado usa’.
- Parrish, C. E., Magruder, L. A., Neuenschwander, A. L., Forfinski-Sarkozi, N., Alonzo, M. & Jasinski, M. (2019), ‘Validation of icesat-2 atlas bathymetry and analysis of atlas’s bathymetric mapping performance’, *Remote sensing* **11**(14), 1634.
- Pattyn, F., Ritz, C., Hanna, E., Asay-Davis, X., DeConto, R., Durand, G., Favier, L., Fettweis, X., Goelzer, H., Golleger, N. R. et al. (2018), ‘The greenland and antarctic ice sheets under 1.5 c global warming’, *Nature climate change* **8**(12), 1053–1061.
- Philpot, W. D. (1989), ‘Bathymetric mapping with passive multispectral imagery’, *Applied optics* **28**(8), 1569–1578.
- Pope, A., Scambos, T. A., Moussavi, M., Tedesco, M., Willis, M., Shean, D. & Grigsby, S. (2016), ‘Estimating supraglacial lake depth in west greenland using landsat 8 and comparison with other multispectral methods’, *The Cryosphere* **10**(1), 15–27.
- Pörtner, H.-O., Roberts, D. C., Adams, H., Adler, C., Aldunce, P., Ali, E., Begum, R. A., Betts, R., Kerr, R. B., Biesbroek, R. et al. (2022), ‘Climate change 2022: impacts, adaptation, and vulnerability. contribution of working group ii to the sixth assessment report of the intergovernmental panel on climate change’.
- Rizzoli, P., Martone, M., Rott, H. & Moreira, A. (2017), ‘Characterization of snow facies on the greenland ice sheet observed by tandem-x interferometric sar data’, *Remote Sensing* **9**(4), 315.

- Scambos, T., Fricker, H. A., Liu, C.-C., Bohlander, J., Fastook, J., Sargent, A., Massom, R. & Wu, A.-M. (2009), ‘Ice shelf disintegration by plate bending and hydro-fracture: Satellite observations and model results of the 2008 wilkins ice shelf break-ups’, *Earth and Planetary Science Letters* **280**(1-4), 51–60.
- Scheick, J. et al. (2019–), ‘icepyx: Python tools for obtaining and working with ICESat-2 data’.
URL: <https://github.com/icesat2py/icepyx>
- Smith, L. C., Chu, V. W., Yang, K., Gleason, C. J., Pitcher, L. H., Rennermalm, A. K., Legleiter, C. J., Behar, A. E., Overstreet, B. T., Moustafa, S. E. et al. (2015), ‘Efficient meltwater drainage through supraglacial streams and rivers on the southwest greenland ice sheet’, *Proceedings of the National Academy of Sciences* **112**(4), 1001–1006.
- Smith, L. C., Yang, K., Pitcher, L. H., Overstreet, B. T., Chu, V. W., Rennermalm, Å. K., Ryan, J. C., Cooper, M. G., Gleason, C. J., Tedesco, M. et al. (2017), ‘Direct measurements of meltwater runoff on the greenland ice sheet surface’, *Proceedings of the National Academy of Sciences* **114**(50), E10622–E10631.
- Sneed, W. & Hamilton, G. S. (2007), ‘Evolution of melt pond volume on the surface of the greenland ice sheet’, *Geophysical research letters* **34**(3).
- Stap, L. B., Köhler, P. & Lohmann, G. (2019), ‘Including the efficacy of land ice changes in deriving climate sensitivity from paleodata’, *Earth System Dynamics* **10**(2), 333–345.
- Sundal, A., Shepherd, A., Nienow, P., Hanna, E., Palmer, S. & Huybrechts, P. (2009), ‘Evolution of supra-glacial lakes across the greenland ice sheet’, *Remote Sensing of Environment* **113**(10), 2164–2171.
- Trusel, L. D., Das, S. B., Osman, M. B., Evans, M. J., Smith, B. E., Fettweis, X., McConnell, J. R., Noël, B. P. & van den Broeke, M. R. (2018), ‘Nonlinear rise in greenland runoff in response to post-industrial arctic warming’, *Nature* **564**(7734), 104–108.
- van de Wal, R. S., Boot, W., Van den Broeke, M., Smeets, C., Reijmer, C., Donker, J. & Oerlemans, J. (2008), ‘Large and rapid melt-induced velocity changes in the ablation zone of the greenland ice sheet’, *science* **321**(5885), 111–113.
- Van den Broeke, M. R., Enderlin, E. M., Howat, I. M., Kuipers Munneke, P., Noël, B. P., Van De Berg, W. J., Van Meijgaard, E. & Wouters, B. (2016), ‘On the recent contribution of the greenland ice sheet to sea level change’, *The Cryosphere* **10**(5), 1933–1946.
- Williamson, A. G., Banwell, A. F., Willis, I. C. & Arnold, N. S. (2018), ‘Dual-satellite (sentinel-2 and landsat 8) remote sensing of supraglacial lakes in greenland’, *The Cryosphere* **12**(9), 3045–3065.
- Williamson, A. G., Willis, I. C., Arnold, N. S. & Banwell, A. F. (2018), ‘Controls on rapid supraglacial lake drainage in west greenland: an exploratory data analysis approach’, *Journal of Glaciology* **64**(244), 208–226.

- Willis, M. J., Herried, B. G., Bevis, M. G. & Bell, R. E. (2015), 'Recharge of a subglacial lake by surface meltwater in northeast greenland', *Nature* **518**(7538), 223–227.
- Xie, H., Xu, Q., Ye, D., Jia, J., Sun, Y., Huang, P., Li, M., Liu, S., Xie, F., Hao, X. et al. (2021), 'A comparison and review of surface detection methods using mbl, mabel, and icesat-2 photon-counting laser altimetry data', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **14**, 7604–7623.
- Yang, K. & Smith, L. C. (2012), '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**(4), 801–805.
- Zhang, J., Kerekes, J., Csatho, B., Schenk, T. & Wheelwright, R. (2014), A clustering approach for detection of ground in micropulse photon-counting lidar altimeter data, *in* '2014 IEEE Geoscience and Remote Sensing Symposium', IEEE, pp. 177–180.
- Zwally, H. J., Abdalati, W., Herring, T., Larson, K., Saba, J. & Steffen, K. (2002), 'Surface melt-induced acceleration of greenland ice-sheet flow', *Science* **297**(5579), 218–222.

Appendix & data availability

Data availability

Data products:

- **ICESat-2**: NASA NSIDC DAAC (NASA National Snow and Ice Data Center Distributed Active Archive Center) via IcePyx Scheick et al. (2019–) or via <https://nsidc.org/data/atl03>
- **Sentinel-2**: Copernicus open access hub API via <https://scihub.copernicus.eu/dhus>

Processing steps and code:

- **Open source code**: https://github.com/gjvanleeuwen/Classify_lakesICESat2

Appendix

Table 6.4: Sentinel-2 data products - L1C

S2 Tile	sensing date
T22WEB	2019-06-16
T22WEC	2019-06-16
T22WEV	2019-06-16
T22WFV	2019-06-16
T22WEV	2019-06-17
T22WFS	2019-06-17
T22WFT	2019-06-17
T22WFU	2019-06-17
T22WFV	2019-06-17

