Charles University

Faculty of Science

Study programme: Geography

Branch of study: Cartography and geoinformatics

Bc. Samo Rusnák

SUPRAGLACIAL LAKES DETECTION AND VOLUME ESTIMATION FROM REMOTE SENSING DATA

DETEKCE SUPRAGLACIÁLNÍCH JEZER A KVANTITATIVNÍ ODHAD JEJICH OBJEMU Z DAT DPZ

Diploma thesis

Supervisor:

Ing. Lukáš Brodský, Ph.D.

Prague, 2023

Vysoká škola: Univerzita Karlova **Fakulta:** Přírodovědecká

Katedra: Aplikované geoinformatiky a kartografie **Akademický rok:** 2022/2023

Zadání diplomové práce

pro Bc. Samo Rusnák

obor Kartografie a geoinformatika

Název tématu: Detekce supraglaciálních jezer a kvantitativní odhad jejich objemu z dat DPZ

Zásady pro vypracování

Práce provede rešerši problematiky detekce hloubek supraglaciálnich jezer z optických snímků dálkového průzkumu Země. Na základě rešerše problematiky navrhne řešitel práce model pro detekci hloubek supraglaciálnich jezer se zaměřením na problematiku parametrizace hodnoty albeda dna jezera. Výsledky modelu budou demonstrovány na vybraném území v rámci Grónska. Součástí práce bude validace výsledků modelu.

Rozsah grafických prací: dle potřeby

Rozsah průvodní zprávy: cca. 50 stran

Seznam odborné literatury:

PHILPOT, W. D. (1989): Bathymetric mapping with passive multispectral imagery. Applied Optics, 8, 28, 1569.

POPE, A., SCAMBOS, T. et al. (2016): Estimating supraglacial lake depth in West Greenland using Landsat 8 and comparison with other multispectral methods. The Cryosphere, 10, 1, 15-27.

TEDESCO, M., STEINER, N., POPE, A. (2015): In situ spectral reflectance and depth of a supraglacial lake in Greenland, Arctic Data Center, https://doi.org/10.5065/D6FQ9TN2

Vedoucí diplomové práce: Ing. Lukáš Brodský, Ph.D.

Konzultant diplomové práce:

Datum zadání diplomové práce: 15.9.2022

Termín odevzdání diplomové práce: červenec 2023

Platnost tohoto zadání je po dobu jednoho akademického roku.

…… .....……

Ing. Lukáš Brodský, Ph.D. doc. RNDr. Přemysl Štych, Ph.D.

Vedoucí diplomové práce Garant studijního oboru

V Praze dne 20.7.2023

Declaration:

I declare that I have prepared the thesis independently and that I have listed all sources of information and literature used.. Neither this thesis nor any substantial part of it has been submitted for another or the same academic degree.

Prague, 28.7.2023

.....

Samo Rusnák

Acknowledgements

I want to express my gratitude to Ing. Lukáš Brodský, Ph.D., for valuable advice, constructive feedback, and support during the consultation of my thesis. I would also like to thank my family and friends for their encouragement and support throughout the process of working on my thesis.

Supraglacial lakes detection and volume estimation from remote sensing data

Abstract

Supraglacial lakes play an important role in understanding glacier dynamics, including their response to climate change. This thesis explores the problematics of estimating lake depth and volume using a physical model. This brings challenges in considering the influence of various factors, such as cryoconite on glacier surface and suspended particular matter, which influences physical model, which is in research mostly neglected. Regression analysis of the g parameter of a physical model, representing light attenuation coefficient, and supervised classification of supraglacial lakes is applied in this thesis. The results reveal the variability of parameter Ad, representing lake bottom albedo reflectance, and its impact on predicted supraglacial lakes depth and volume. The results highlight the problem of global parameterisation of the physical model of supraglacial lakes and the need for further research to improve its accuracy and explore future possibilities in this field.

Keywords: supraglacial lake, remote sensing, machine learning, physical model, depth estimation, regression analysis

Detekce supraglaciálních jezer a kvantitativní odhad jejich objemu z dat DPZ

Abstrakt

Supraglaciální jezera představují důležitou roli v pochopení dynamiky ledovců, včetně jejich reakce na klimatické změny. Tato práce se zabývá problematikou odhadu hloubky a objemu jezer pomocí fyzikálního modelu. To přináší výzvu z hlediska zohlednění vlivu více faktorů, jakými jsou kryokonit na povrchu ledovce a nerozpuštěné látky ve vodě jezera, jejichž vliv na fyzikální model je ve výzkumu většinou opomíjen. V této práci je použita regresní analýza parametru *g* fyzikálního modelu, který představuje koeficient útlumu světla, a řízená klasifikace supraglaciálních jezer. Výsledky odhalují variabilitu parametru *Ad*, který představuje odrazivost dna jezera, a jeho vliv na modelovanou hloubku a objem supraglaciálních jezer. Výsledky poukazují na problematiku globální parametrizace fyzikálního modelu supraglaciálních jezer a upozorňují na potřebu dalšího výzkumu s cílem zlepšit jeho přesnost a prozkoumat budoucí možnosti v této oblasti.

Klíčová slova: supraglaciální jezero, dálkový průzkum Země, strojové učení, fyzikální model, odhad hloubky, regresní analýza

1. Content

2. Used abbreviations

a – absorption coefficient

A – area

A_d – lake bottom albedo

D_u – upwelling light distribution function

K_d – diffuse attenuation coefficient for downwelling light

MAE – Mean Absolute Error

NDWI – Normalised Difference Water Index

R_∞– reflectance of optically deep water

R_{nir} – reflectance in the near-infrared band

R_r –reflectance in the red band

R_w – reflectance of lake water

R_x – reflectance of the particular band (r – red, g – green, b – blue, nir – near infrared)

RMSE – root-mean-square error

SGL – Supraglacial lake

TOA – Top of the atmosphere

V – volume

z – depth

3. List of Tables

4. List of Figures

5. Introduction

Supraglacial lakes (SGLs) are a particular type of lake which form during the summer melt season on the surface of glaciers. Their presence on the glacier's surface indicates its dynamic nature and their behaviour raises interest in glaciology, hydrology, and climate research. These lakes play an essential role in the mass balance of glaciers, and their accurate mapping and volume estimation are necessary for understanding the glacier processes.

5.1. Research questions

The research question of this master thesis is to investigate the current literature and conduct research in the study area of SGLs with the aim of SGLs mapping, depth, and volume estimation. The main research question is how to quantify the inaccuracies of SGL modelling and their influence on lake depth and volume modelling. More specifically, the thesis focuses on the impact of the lake bottom albedo parameter (A_d) in the physical model on the overall accuracy of SGL modelling. Comparison between A_d parameter obtained as a single value versus variable determination of this parameter is an important part of this study. This is achieved by following associated goals:

- Detection of SGLs.
- Obtaining reference data.
- Investigation of lake bottom albedo variability.
- Impact of A_d on the overall modelling of the lake.

SGLs detection in the thesis is carried out using a supervised classification, which is expected to enhance the classification accuracy compared to the thresholding methods used in previous studies.

Acquiring accurate reference data for SGLs involves utilising remote sensing data and in-situ measurements. These include various sources, such as ICESat-2, Landsat imagery and in-situ data.

Unlike many previous studies that treat A_d as a constant, this thesis delves into its spatial variability, which is investigated through the study area.

Lastly, the thesis aims to quantify the role of A_d on SGL depth modelling by determining the accuracy of lake depth and volume estimation. The study compares the results obtained using a single, constant value of A_d against those achieved through variable determination to assess its implications on the overall accuracy of SGL modelling.

6. Theoretical background

6.1. Supraglacial lakes in general

SGLs are lakes that occur on the glaciers' surface. Their occurrence is documented in various locations within glaciers. These lakes belong to the group of lakes which water is in touch with the glacier, from which at least part of the lake is in contact with the glacier (Ashley 2002). These lakes can be located under the glacier (subglacial lakes), inside the glacier (englacial), on the top of the glacier (supraglacial), or at the edge of the glacier. SGLs form on the surface of the glaciers in various shapes, from isolated depressions on the inactive parts of the glacier to complex connected structures. The lifespan of SGLs is relatively short; their dimensions change together with the melting of the glacier, and their depth, in general, is shallow.

Due to the inaccessibility of the places where SGLs occur, there is only a limited number of conducted research on SGLs, especially in-situ. Interestingly, the first person ever to attempt to swim in a SGL did so in January 2020 (Guinness World Records Limited 2023). Therefore, using remote sensing methods to research SGLs is beneficial.

SGLs are generally located on Antarctic Ice Sheet, Greenland Ice Sheet, and some mountain glaciers. A global study of SGLs estimated that the volume of SGL water worldwide is 156,5 km^3 (Shugar et al. 2020). Some regional inventories of SGLs are also provided, for example, in the West Antarctic Ice Sheet (Corr et al. 2022), Greenland (Hu et al. 2022), Cordillera Blanca mountain range in Peru (Baťka 2015), Khan Tengri-Tumor Mountains in Central Asia (Qiao, Mayer, Liu 2015), and in China–Pakistan Economic Corridor (Li, Shangguan, Anjum 2020).

6.2. Remote sensing of supraglacial lakes

Observing several characteristics of these lakes is possible when studying SGLs using remote sensing techniques. Current studies focus primarily on the issues of detecting SGLs, their number and position, and determining their depth and volume. Satellite data, such as Landsat or MODIS, are mainly used for these purposes.

Box and Ski (2007) conducted a study in which they estimated the depth of the SGLs in Greenland from MODIS imagery from April to September of 2000 to 2005. They utilised for study MODIS bands 1, 3 and 4. Additionally, as reference data, they measured the depth of two SGLs using the Garmin Fish Finder 100 sonar device in August 2005. Lake classification was achieved using the thresholding of the reflectance ratio of MODIS 1 and 3 bands, with the value of 1.2 as well as a threshold for the red band chosen on the lake-by-lake basis in the interval 0.5 – 0.6 to prevent the influence of differing concentrations of ice impurities ***(Figure 1 left).*** Consequently, for lake depth estimation, the parametrised model was obtained from a combination of two best-fit functions over measured lakes ***(Figure 1 right).***

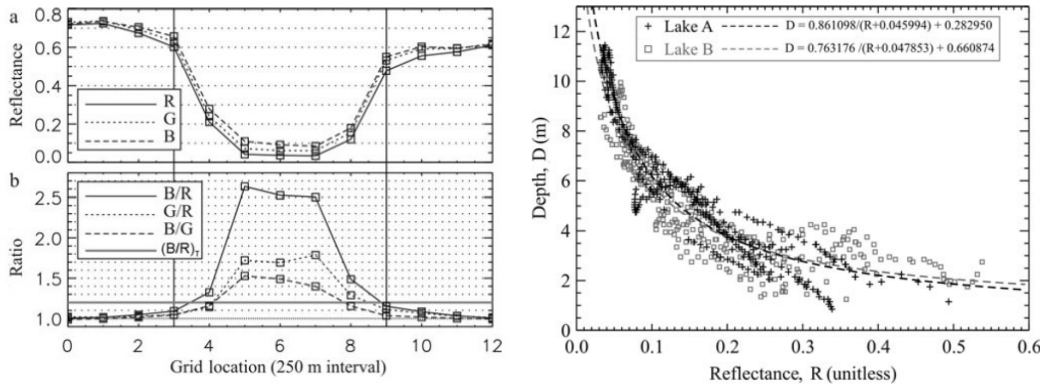


Figure 1: ***Left:*** *Lake transects: (a) reflectance and (b) reflectance ratios. R, B, and G indicate red, blue, and green, respectively, and $(B/R)_T$ is the blue/red reflectance threshold 1,2 used in the study.* ***Right:*** *Lakes MODIS band 1 reflectance vs depth with best-fit functions. Source: Box, Ski (2007).*

Murray and James (2011) were in their study watching rapid drainage events in Greenland from MODIS data in the period of 2005 – 2009. A single threshold value for SGL classification was used instead of variable threshold values. Around each input pixel was defined a 21 x 21 pixel window. If the reflectance of the central pixel is less than 0.65 of the mean reflectance inside a defined window, the pixel is classified as an SGL.

Fitzpatrick et al. (2014) investigated the evolution of lake volumes on Russell Glacier in western Greenland, covering an area of approximately 6500 km2, between 2002 and 2012. In the study, they worked with 502 MODIS images between May and September. They used a modified normalised difference water index (NDWI) for the SGLs detection:

$$NDWI = \frac{R_{MODIS\,3} - R_{MODIS\,1}}{R_{MODIS\,3} + R_{MODIS\,1}}.$$

To determine the depth of SGLs, they used an empirically derived model based on two lakes bathymetric data and reflectance in MODIS band 1 from July 2010:

$$z = \frac{0.65}{0.035 + R_{MODIS\,1}} - 0.4\,.$$

Sundal et al. (2009) investigated the evolution of SGLs areas in three different regions of Greenland during the melt season from May to October in the years 2003 and 2005 to 2007. The detection of SGLs was based on bands 1, 3 and 4 of MODIS imagery, corresponding with the visible spectrum. Definiens eCognition Professional 4.0 software environment's fuzzy logic membership functions were used for the classification of SGLs, based on MODIS band 1 reflectance and the ratio between MODIS band 3 and MODIS band 1. MODIS band 4 was in this study only for visual analysis.

The research conducted by Stokes et al. (2019) focused on the distribution of SGLs in coastal regions of East Antarctica in January 2017. They utilised satellite imagery from Sentinel 2 and Landsat 8 missions. SGLs identification was provided by the utilised NDWI index as follows:

$$NDWI = \frac{R_g - R_{nir}}{R_g + R_{nir}},$$

in which they used a threshold value of 0.3 and higher to detect SGLs. The authors also monitored the volume of the lakes as a secondary objective of the study. Due to the complications associated with using multiple sensors, the authors decided to estimate the lake volume based on its surface area. They utilised the data from two studies conducted in the Antarctic region, which provided information about SGLs volumes and areas. Through regression analysis, they established a relationship between the volume and area of the lake as follows:

$$V = 7.16e^{-4}A.$$

6.3. Depth estimation based on Bouguer-Lambert-Beer law

Several studies work in estimating the depth of SGLs based on the Bouguer-Lambert-Beer law, which was used in the study conducted by Philpot (1989) to create a model for optically shallow water bodies from multispectral images. Based on this model, Sneed and Hamilton (2007) developed an algorithm for estimating the depth of SGLs using multispectral ASTER images using the VNIR1 (520 – 600 nm) and VNIR3N (780 – 860 nm) bands in the area of west Greenland. They used the following relationships:

$$z = \frac{\ln(A_d - R_\infty) - \ln(R_w - R_\infty)}{-g},$$

$$g \approx K_d + aD_u,$$

where z represents depth, A_d represents the bottom of substance albedo, R_∞ represents the reflectance of optically deep water, R_w represents the reflectance at some depth, K_d represents the diffuse attenuation coefficient for downwelling light, a represents beam absorption coefficient and D_u represents distribution function of the reciprocal of the upwelling average cosine. To solve the equation, they used the VNIR1 band. VNIR3N band was used to obtain A_d value, where they searched for adjacent pixels that show a rapid decrease in the reflectance. Consequently, they used the resulting pixels in the VNIR3N band to find the same pixels in the VNIR1 image for the equation's A_d value. The result was not validated with field measurements, but the authors stated that obtained depths correspond with other published studies.

Georgiou et al. (2009) in their study used the method mentioned above by Sneed and Hamilton while they were following the development of the area and depth of one SGL on the Jakobshavn Isbræ glacier in western Greenland between 17th June 2002 and 12th August 2005. They used the ASTER data, validated by lidar measurements from May 1995. The value of g was determined as $0.206 \pm 0.010\ \mathrm{m}^{-1}$, R_∞ values were obtained from seawater, reaching values 0.07 – 0.14 and A_d values were obtained from locations where, based on lidar data, were estimated locations with shallow water together with method using the VNIR3N band. The results of estimated depths were compared with the lidar data, which were recorded at a time with the lake was not filled

with water. The resulting root mean square deviation between the modelled and observed depths of the SGL reached 0.3 m.

In their research, Morriss et al. (2013) monitored the evolution of SGLs in West Greenland in an area of approximately 1600 km^2 over a summer period of 10 years between 2002 and 2011. Their study used the MODIS data bands 1 and 2 and from Landsat 7 ETM+ data bands 3 and 4. The Normalized Difference Lake Index (NDLI) was used to determine the lake areas:

$$NDLI = \frac{R_r - R_{nir}}{R_r + R_{nir}},$$

where R_r represents the red band reflectance and R_{nir} represents the reflectance in the near-infrared band. The threshold for pixel-to-pixel classification of the lake water was determined to be 0.21. Lake depth analysis used a modified Sneed and Hamilton method, in which the lake depth can be derived as follows:

$$z = \frac{\ln\left(\frac{A_d - R_\infty}{R_r - R_\infty}\right)}{-g}.$$

The value for R_∞ was obtained from the centre of the bay on 3rd July 2011, which was subsequently used for all images of both sensors. The value of the g parameter was determined to be $0.868\ \mathrm{m}^{-1}$.

Pope et al. (2016) presented a study that applied Landsat 8 data, where they compared estimates of SGL depth using empirical, as well as Philpot's model for optically shallow water with in-situ measurements, WorldView DEM and various multispectral methods. The empirical model for the study utilised a combination of two bands based on the following relationships:

$$z = a + bX + cX^2,$$

$$X = \ln\left(\frac{R_1}{R_2}\right),$$

where R_1 and R_2 represents the reflectance of two different bands. The model for optically shallow water utilised the following relationship:

$$z = \frac{\ln(A_d - R_\infty) - \ln(R_w - R_\infty)}{g}.$$

In the research, they used in-situ measurements from West Greenland taken on the 2nd and 3rd of July 2010, which collected data on lake depth and spectral characteristics containing a total of 2226 points (Tedesco, Steiner 2011). They convolved these points to the relative spectral response of multiple multispectral sensors to emulate the reflectance at the sensors in question (Tedesco, Steiner, Pope 2015). Pope et al. (2016) subsequently used these data to determine the relationships between the reflectance of the individual bands, as well as combinations of the OLI and ETM+ products and the lake depth (Figure 2) to identify the most appropriate bands for deriving the depth for each model, as well as the coefficients R_∞, A_d and g. They found that the estimated depths predicted OLI as well as or better than ETM+, and based on the correlation coefficient and the lowest root mean square errors, they determined for further work on the study for the optically shallow water model the OLI bands 3 (green), 4 (red), 8 (panchromatic), and a combination of bands 4+8 and 2+3, and for the empirical model a combination of bands 1+3 and 1+8. However, the information obtained this way may not correspond for lakes deeper than 5 m since in-situ measurements did not exceed this value. The R_∞ coefficient value used in the analysis was obtained from the dark ocean or lake in the scene. If the scene did not contain such an option, the following scene was used, assuming the same atmospheric conditions occurred. The A_d value was obtained using the reflectance of pixels next to the identified lake areas. The researchers, after performing all methods, recommended the average of depths calculated using bands 4 (red) and 8 (panchromatic) as the best method for OLI. This method resulted in an average error of 0.0±1.6 m, and they suggester its implementation for future research.

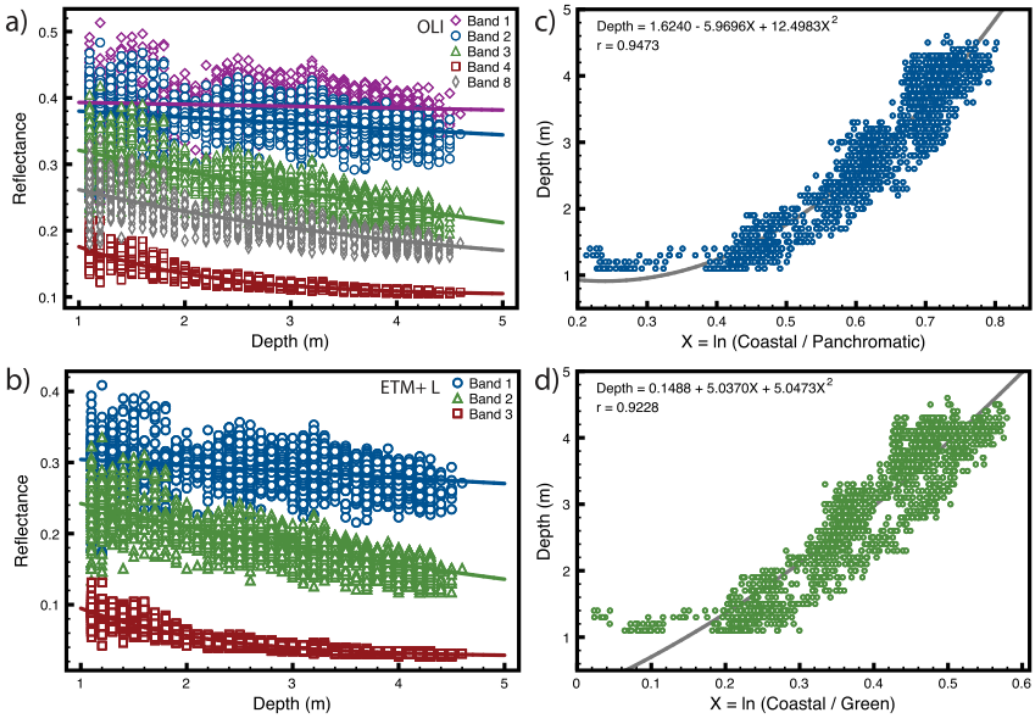


Figure 2: Regression plots for in-situ measured reflectance spectra used to emulate OLI and ETM+ reflectance and sonar-measured depths, including OLI single band (a), ETM+ low gain single band (b), OLI coastal and panchromatic (c), and OLI coastal and green (d). Source: Pope et al. (2016).

Williamson et al. (2017) developed and applied a fully automated algorithm for tracking the area and volume of SGLs using MODIS imagery in West Greenland. The algorithm uses only MODIS MOD09 imagery and the area of interest as input parameters. Comparison of the algorithm results was with the derived values obtained from Landsat 8 imagery using the algorithm of Popa et al. (2016) and the empirical model of Fitzpatrick et al. (2014). To select an appropriate algorithm for determining lake areas, Williamson et al. (2017) tested three different approaches; NDWI, a composite of the blue and red band, a threshold for combinations of red and blue band data, and a dynamic threshold for the red band using a 25x25 pixel resolution neighbourhood pixel. The dynamic threshold of 0.64 was the most successful result and was later incorporated into the framework of the developed automatic algorithm. To calculate the depth of a given lake, they followed the method of Sneed and Hamilton (2007) applied to the MODIS red band, which achieved accurate depths and volumes when compared with the derived values from Landsat 8 data (the root mean square error for depth reaches 1.27 m and for volume $5.9 \times 10^7 m^3$) and are comparable to the empirical model.

6.4. Supraglacial lakes depth estimation

Evidently, two main approaches to modelling the depth of SGLs are based on the above-mentioned information in the previous chapter.

One approach is empirical modelling, which involves utilising a model based on in-situ bathymetric or lidar measurements and remote sensing data in the form of reflectance values from satellite imagery. These models establish a quantitative relationship between lake depth and another variable to estimate other lake depths.

The second approach is physical modelling, based on parameters derived from the Bouguer-Lambert-Beer law. This law describes the absorption of light moving through a medium, such as water (Rafferty 2023). In the context of SGLs, the law was utilised for physical modelling by Philpot (1989) with parameters such as lake bottom albedo A_d, the reflectance of optically deep water R_∞ and losses in upward and downward travel through the water column g used for calculation of lake depth. The variability of physically based model parameters is high, as shown in Table 1. The differences in coefficients vary; for example, the g parameter from 0.118 (Sneed, Hamilton 2011) to 0.868 (Morriss et al. 2013), and A_d varies from 0.19 (Moussavi et al. 2016) to 0.5639 (Sneed, Hamilton 2011). R_∞ parameter values were relatively consistent among studies.

Table 1: Physically based model parameters. Source: Brodský et al. (2022)

Table 1. Physically based model parameters.

Attenuation Coefficient (g) (m^{-1})	Bottom Albedo (A_d)	Deep Water Reflectance ($R\infty$)	Method	Reference Publication
0.118	0.5639	0.038	g: NA [1], A_d: lake ice [3], R_∞: ocean [4]	Sneed and Hamilton [42]
0.65	0.34 + 0.062	NA [1]	$g = 2$ K [2]; A_d: lake ice [3]	Tedesco and Steiner [26]
0.868	0.5	0.02	Fitting (g); A_d: NA [4]; R∞: ocean [4]	Morriss et al. [47]
0.69	0.19	0.03	Fitting (g, A_d, and R_∞)	Moussavi et al. [24]
0.73	0.28	0.03	Fitting (g, A_d, and R_∞)	Moussavi et al. [24]
0.80	0.228	0.0375	Fitting (g)	Pope et al. [14]
0.83	0.212	0.047		

[1] Not Available, [2] Kd value from Smith and Baker [46], [3] lake adjacent ice average from image, [4] ocean reflectance average.

6.5. Problems of physical modelling

The physical model has limitations, as it assumes minimal levels of dissolved, or suspended particles in the water, neglects wave formation on the lake surface due to the wind, assumes a homogenous lake bed and disregards non-elastic scattering (Sneed, Hamilton 2011).

Sneed and Hamilton (2007) stated that the volume results for lakes are most sensitive to the A_d parameter, in which a 1% decrease in the bottom albedo results in a 16% decrease in calculated meltwater volume. They stated that accumulation of organic and inorganic matter at the bottom of the lake would decrease the bottom albedo, while sun glint caused by surface waves would increase the bottom albedo. They find that model is less sensitive to other values in physical modelling. A 10% decrease in g value results in a 10% volume increase, while a 2% increase in R_∞ value results in a 0.5% volume increase.

Pope et al. (2016) stated that physically based depth retrievals rely on accurate bottom albedos A_d and water absorption coefficient g. A comparison of laboratory-measured g with in-situ derived data shows that when laboratory-measured g is higher, lake depths are overestimated, and in reverse, when it is smaller, lake depths are underestimated, compared to WorldView DEM data. Water absorption properties vary with different wavelengths, influencing the physical model. Red wavelength with relatively higher values of g, compared to green, make it less sensitive to errors.

Tedesco and Steiner (2011) observed that the spectral behaviour of A_d is similar to glacier ice albedo. They also observed that the A_d obtained from in-situ measurements and the mean reflectance from Landsat data was approximately 10%, which led to an average error in depth retrieval of -11.8%, in the case of band 1, or -15.9% in the case of band 2.

Brodský et al. (2022) demonstrated the negative effects of suspended particulate matter in the SGL water on the performance of the physically based model on the example of an SGL on the Glacier de la Plaine Morte in Switzerland. The presence of the suspended particulate matter in the lake affected the reliability of the results, limiting usability to only up to 3 meters of depth in that specific lake.

From the problems in the research is evident that the properties of lakes in the real world are different from those modelled in the physically based optical model, in which one of the assumptions was that there is no presence of suspended particulate matter in the lake. Various authors approached the parameterisation of the physical model differently. Authors approach the parameter A_d differently; some choose a single value for the entire study area, while others use a technique that determines the value based on the surrounding pixels of the lake. The variability of the A_d parameter due to changing amount of suspended particulate matter in the water and cryoconite on the glacier surface area is acknowledged. However, no research was yet conducted on this type of problem. Considering another model parameter, R_∞, most studies use the value obtained from the image where the sea occurs. However, the performance can be influenced, for example, by the presence of ice floes in the water, which is also a problem not recognised in most studies using the physical model for SGL depth estimation.

7. Methodology

In this chapter, the methodology of this thesis is introduced. The applied methodology is divided into six sections, shown in the flowchart (figure 3), as follows:

1) Imagery data and preparation.
2) Obtaining reference data from ICESat-2.
3) Regression analysis of g parameter for the physical model.
4) Classification of image with the primary objective of finding SGLs.
5) Estimation of A_d parameter for SGLs.
6) Estimation of lake depth and volume.

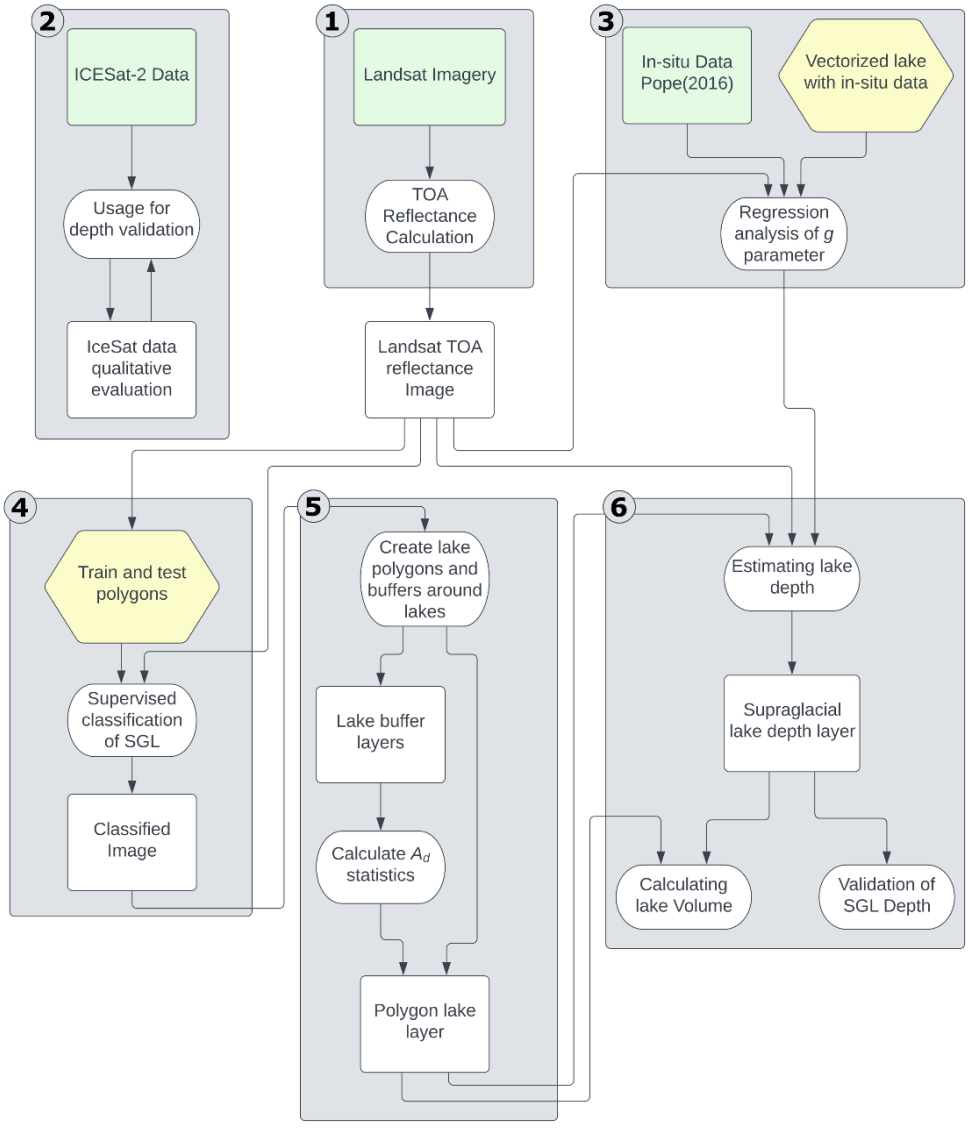


Figure 3: Methodology flowchart. Inputs: green – input data, yellow – manually edited data

Firstly, the imagery data, in this case Landsat data, is prepared for further work by calculating Top of Atmosphere (TOA) reflectance. Secondly, the possibility of usage of ICESat-2 data for validation is conducted. This involves searching for the intersection between Landsat and ICESat-2 data. Thirdly, from reference in-situ data, the parameter g for further use in the physical model is parametrised from in-situ data. This is accomplished by the method of least squares from regression analysis. Fourthly, the classification of TOA image to identify SGLs is conducted. The resulting classified image is used in the fifth step for creating SGLs polygons and buffers around each lake at various distances to estimate the A_d parameter. Lastly, all available work is combined into calculating lake depth and consequently their volume with subsequent validation.

7.1. Data preparation

The imagery used for SGLs detection was from Landsat 7 and 8 collections. The choice of image was in the specified area of Greenland island, in the west part of the island near Disco Island. The date of the selected image corresponds with the summer melt season, in which most of the SGLs occur. Images have, within its extent, visible areas with optically deep water, ideally in the form of a sea or deep lake. Used images are without or with a small amount of cloud cover.

Based on available reference data, mainly ICESat-2 data set ATL03 (Neumann, T. A., et. al. 2021) and In-Situ data from Tedesco, Steiner, Pope (2015), the chosen scenes were as follows (LANDSAT_PRODUCT_ID):

- LE07_L1TP_008011_20100709_20200911_02_T1
- LE07_L1TP_010011_20100707_20200911_02_T1
- LC08_L1TP_007012_20210708_20210713_02_T1
- LC08_L1TP_007013_20210708_20210713_02_T1

TOA reflectance was calculated for the selected images. This was accomplished by using coefficients provided in image metadata (USGS 2022). This provided a composite multi-band TOA Landsat image for further analysis.

7.2. Obtaining references from ICESat-2

The researchers analysed the ICESat-2 data set product ATL03 as a potentially promising reference data source. ATL03 data set provides geolocated photons data containing height above the WGS 84 ellipsoid (Neumann, T. A., A. Brenner, D. Hancock, J. Robbins, J. Saba, K. Harbeck, A. Gibbons, J. Lee, S. B. Luthcke, T. Rebold 2021). Further information about data is provided in Chapter 8.1. The algorithm *Watta* was used to calculate SGL bathymetry (Datta, Wouters 2021). The algorithm is provided in the *Matlab* software environment as a series of scripts, which from initial ATL03 data finds lakes on the glacier surface and provide lake depth, together with other features occurring in SGLs, such as surface or sub-surface ice. However, if this data aren't able to use, another method of reference will be used by cross-validation methods.

7.3. Approach to g parameter estimation

A few key parameters are necessary for using a physical model for SGL depth estimation. One of them is the parameter g, which figures in the approximated physical model (Sneed, Hamilton 2007):

$$z = \frac{\ln(A_d - R_\infty) - \ln(R_w - R_\infty)}{g},$$

where z represents lake depth, A_d represents lake bottom albedo, R_∞ represents the reflectance of optically deep water, R_w represents the reflectance of lake pixel and g represents losses in upward and downward travel through the water column. Assuming that the other parameters of the model are known, the value of the parameter g can be derived using parametrisation from in-situ data, obtained by regression analysis with the least square error method. In this regression, parameter g was regressed based on z obtained from in-situ data by Tedesco, Steiner, Pope (2015), A_d obtained from calculated albedo of surrounding ice, R_∞ was used with value 0,048 as in (Pope et al. 2016) and R_w was obtained from the red band of Landsat 7 TOA imagery.

The regression analysis of the parameter g was conducted using published in-situ data by Tedesco, Steiner, Pope (2015). The dataset contains depth data collected from a

lake situated on the Greenland Ice sheet between the 2nd and 5th of July 2010. These data include information about the geographical position in the form of x and y coordinates and depth at the location in meters rounded at a tenth of a meter. The regression analysis included all in-situ data from all available dates, except the data from 4th July, which contains only two in-situ measured depths, as well as only a subset of the data based on the dates or by outlier measurements. The used data for regression were also used differently based on used Landsat reflectance image into two main categories:

- In-situ data points contained within both Landsat 7 TOA images (dates 7th and 9th July 2010). Here the regression of g parameters was conducted for both images with the following data subsets:
    - In-situ data for 2nd July 2010, containing all data points, except for data points with a depth around value 300.
    - In-situ data for 3rd July 2010, containing all data points.
    - In-situ data for 5th July 2010, containing all data points.
    - In-situ data for 5th July 2010, containing all data points, except for the pixel with outlying value (see Figure 11 in Chapter 8.2.).
    - In-situ data for 2nd, 3rd and 5th July 2010, containing all data points, except for data points with a depth around value 300 from 2nd July 2010.
- In-situ data points which are located within Landsat 7 TAO image from 9th July 2010, where regression of g parameter was conducted with the following data subsets:
    - In-situ data for 5th July 2010, containing all data points.
    - In-situ data for 5th July 2010, containing all data points, except for the pixel with outlying value (see Figure 11 in Chapter 8.2.).
    - In-situ data for 2nd, 3rd and 5th July 2010, containing all data points, except for data points with a depth around value 300 from 2nd July 2010.

Based on the results (presented in Chapter 8.2.), a combination of in-situ data from 5th July 2010 and Landsat 7 imagery from 9th July 2010 was used for subsequent steps. The lake outline was manually obtained over the Landsat 7 image. Later around

the lake edge was created a buffer with a width of 90 meters, from which the average reflectance in the red band was calculated, and a histogram of all reflectance values in the red band was obtained. The buffer width of 90 meters corresponds to the 3 Landsat 7 pixels.

Depth in-situ data with a corresponding value of Landsat 7 TOA pixel reflectance were split into training and validation datasets with a ratio of 7:3, following the methodology conducted by Brodský et al. (2022). The regression analysis was performed by Python script using the *curve_fit* function from the *scipy* library. Subsequently, various statistics parameters were calculated, including RMSE, MSE and R^2 Score, to evaluate the regression model performance.

To analyse how A_d uncertainty propagates through the SGL depth modelling, a Monte Carlo simulation of the fitted g parameter was conducted. The input data for the simulation included a histogram of all reflectance values within the buffer around the SGL edge, which was used to calculate parameter A_d, the regressed value g and previously used datasets with A_d and R_w values. The calculation was later also provided for all values of A_d obtained from average reflectance in 90 meters surrounding of lake edge. Using the mean and standard deviation of the A_d data, 100 values for 100 simulations were calculated. Results were reproduced in the table, and summary statics of RMSE and R^2 scores were calculated.

7.4. SGL outline mapping

Searching for the areas with SGLs is necessary for further analysis. The knowledge of the extent of SGLs is crucial for accurately estimating the SGL edge, which is further essential for obtaining the parameter A_d for the physical model. Although most of the studies used for the detection of SGLs thresholding of various bands, in this thesis, to achieve higher accuracy of results, mainly due to inaccuracies caused by changing threshold value in space, the supervised image classification will be provided.

Training samples were manually edited on chosen images in the adequate number

for both classification categories: ice and SGLs. Other land use categories for this classification which occur in the image, weren't classified. The total number of training polygons was 84, with a total area of roughly 67 km^2, corresponding roughly to 74 400 pixels. Subsequently, the model for supervised classification was trained on half of the training samples, while another half was used to validate the resulting classification. The samples were distributed between the training and validation datasets randomly.

Supervised classification of the SGLs was conducted with available training samples on multi-band raster containing bands 1-8 with the Random Forest method. The classification was performed in the software environment ArcGIS Pro 2.5, with chosen supervised method *Random Trees* (ESRI 2023b) with parameters as follows:

- Maximum Number of Trees: 50
- Maximum Tree Depth: 30
- Maximum Number of Samples per Class: 1000

After runs, the area was checked, and new train polygons were added in the places of inconsistencies. This increased the dataset used for training compared to the validation dataset.

The final classification was validated with a reference dataset, from which 1000 equalised stratified random points were created for each class, provided by the ArcGIS Pro 2.5 classification wizard. The control points were also provided from the previously created validation dataset, created by ArcGIS Pro function *Create Random Points* (ESRI 2023a), with parameters:

- Number of points: 100
- Minimum allowed distance: 30 meters

which provided per each polygon a maximum number of 100 control points, with a distance between each other of a minimum 30 meters.

For both provided control points datasets were calculated following indicators for SGL water classification: Precision, Recall, F1 Score and Accuracy were calculated for this data with equations shown below:

$$Precision = \frac{True\ positive}{True\ positive + False\ positive}$$

$$Recall = \frac{True\ positive}{True\ positive + False\ negative}$$

$$F1\ Score = \frac{Precision * Recall}{(\frac{Precision + Recall}{2})}$$

$$Accuracy = \frac{True\ positive + True\ negative}{True\ positive + True\ negative + False\ positive + False\ negative}$$

Provided values are shown in the confusion matrix template (Table 2), where the lines represent the value in the classified image, and the columns represent values in the reference dataset. Cohen's Kappa index was also calculated with the provided equation:

$$\kappa = \frac{P_o - P_e}{1 - P_e},$$

where κ represents the Kappa index, P_o represents the proportion of units the judges agree on, and P_e represents the proportion of units for which agreement is expected by chance (Cohen 1960). The calculation of the parameters with the previous parameters can be provided as follows:

$$P_o = Accuracy$$

$$P_e = \frac{Expected\ positive}{Total} * \frac{Total\ positive}{Total} + \frac{Expected\ negative}{Total} * \frac{Total\ negative}{Total}$$

Table 2: Confusion matrix example with provided coefficients for calculation.

Training -> Result	SGL Water	Iceberg	Total
SGL Water	$True\ positive$	$False\ positive$	$Total\ positive$
Iceberg	$False\ negative$	$True\ negative$	$Total\ negative$
Total	$Expected\ positive$	$Expected\ negative$	$Total$

7.5. A_d estimation for supraglacial lakes

Estimating the lake depth using the physical model requires obtaining parameter A_d, representing the lake bottom albedo in the equation for the depth estimation process. This parameter was calculated locally from the average reflectance in each bounding box for the red band of Landsat 7 imagery (Band 3). The calculation was also provided for global values for the A_d parameter. The global parameter was obtained as a median, average, maximum and minimum values of all lakes in 0 – 90 meters bounding box, as well as a value obtained from a bounding box 0 – 90 meters for the lake used in the study conducted by Pope et al. (2016).

Validated Random Forest classification identified SGLs, which were transformed into polygon layer. The analysis in the study considered only SGLs with an area larger than 20 000 m^2, or in terms of pixels, larger than 22. To ensure that only water areas within the glacier's region were considered and to avoid using areas outside of the glacier, the iceberg mask of the Greenland ice sheet obtained from OpenStreetMap data via the Geofabrik database was used to identify only the water areas within the glacier's extent (Geofabrik GmbH, OpenStreetMap Contributors 2022).

This polygon layer was identified for each lake's bounding box at various distances. These distances were derived from the pixel size resolution of Landsat 7 optical sensors. Distances are as follows: 0 – 30 m, 0 – 60 m, 0 – 90 m, 30 – 60 m, 30 – 90 m and 60 – 90 m. In the analysis, pixels falling within the gap mask of the Landsat 7 Scan Line Corrector (SLC) failure and pixels previously classified as water were excluded from the bounding box. Landsat 7 SLC failure produces data gaps, which on the image appear as black lines with no data, which scenes only have 78 % of pixels remaining (USGS 2023). By excluding water pixels and the SLC failure data gap, the analysis avoids the influence of neighbouring water bodies on the lake surroundings (Figure 4).

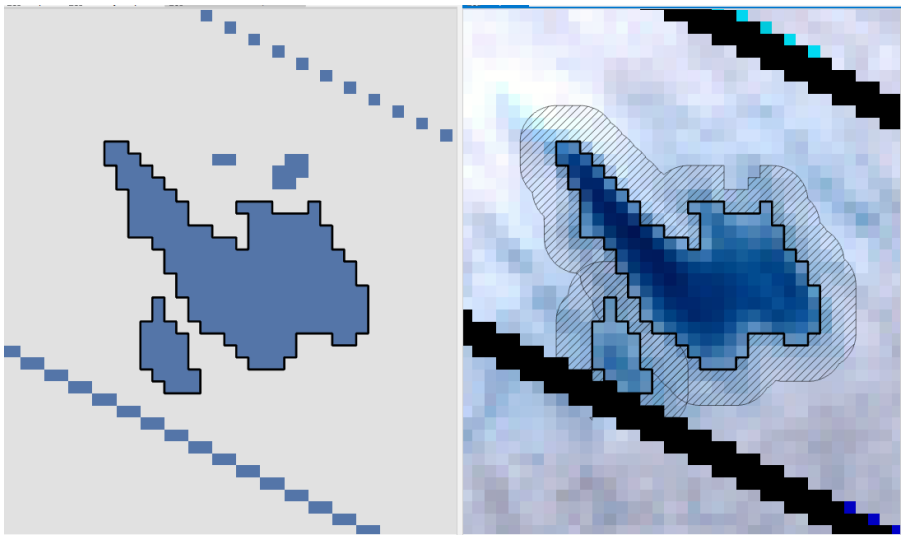


Figure 4: Example of 0-90 m buffer around lake edge interrupted by another classified lake, smaller lakes and Landsat 7 gap. Left: Classified Landsat 7 image, where grey represents ice and blue water. Selected SGLs for study with black frame. Right: Landsat 7 TOA RGB composite image with selected lakes. The black part of the image is caused by Landsat 7 SCL failure. Source: USGS (2023).

7.6. Estimating lake depth and volume

Lake depth estimation was accomplished by using a model, which is an approximation of the Bougher-Lambert-Beer law, introduced by Philpot (1989), for estimating the depth of optically shallow water bodies from multispectral imagery. It is accomplished by the following equation (Sneed, Hamilton 2007):

$$z = \frac{\ln(A_d - R_\infty) - \ln(R_w - R_\infty)}{g},$$

where z represents lake depth, A_d represents the lake bottom albedo, R_∞ represents the reflectance of optically deep water, R_w represents the reflectance of lake pixel, and g represents losses in upward and downward travel through the water column. A_d was obtained from the average ice reflectance values in the various distances from the edge of the iceberg (as written later) as well as it was obtained as a global parameter of average, median, maximum, and minimum value of lake's average value of the A_d parameter in the distance 0 – 90 meters from the lake edge. An average of ice surrounding the lake with available in-situ data was also used at the same distance. Overall, these methods for obtaining parameter A_d were used:

- Local parameter A_d
    - Distances from the lake edge: 0-30 m, 0-60 m, 0-90 m, 30-60 m,

30-90 m, 60–90 m.

- Global parameter A_d
    - Absolute lowest local A_d parameter in 0-90 m distance.
    - Median value of local A_d parameter in 0-90 m distance.
    - Average value of local A_d parameter in 0-90 m distance.
    - Absolute highest local A_d parameter in 0-90 m distance.
    - Value of local A_d from lake used for obtaining g parameter in 0-90 m distance.

R_∞ was established as a global parameter for the image, with the value of 0,048, obtained from Pope et al. (2016), and g was used from previously conducted regression with the value 0.893. The estimation was saved as a raster file with calculated lake water depth values.

The previously created raster file was used for lake volume calculations for each SGL in the polygon layer. A few of the lakes had missing values inside them because the reflectance in the pixel was lower than the R_∞ parameter. Lake volume calculation was assigned a depth value of 7 meters to these pixels, based on the closest approximation of nearly R_∞ (shown in Table 5 in chapter 8.4.1.). However, the actual depth of the lake can be higher. Similarly, the pixels with negative depth values were not considered for volume calculation. The resulting volumes of a lake, calculated for each value of the A_d parameter, were saved into the lake polygon layer.

Resulted lake depths were evaluated based on the multiple factors based on the difference in the equation´s A_d parameter, for example, in differences in the bounding box size for calculation of the parameter and overall variability in this parameter. This was performed for each SGL.

Lastly, validation of resulted depth pixels was calculated from provided lake depths. This is provided by comparing in-situ measured geolocated depths with provided lake depth calculations for each method of obtaining A_d parameter. The comparison was provided for every in-situ geolocated depth value, as well as for the average and most frequent value in each Landsat 7 TOA pixel containing at least one measurement. The calculation of parameters RMSE, R^2, and MAE were calculated.

8. Results & Discussion

8.1. Calibration data from IceSAT-2

ICESat-2 dataset offers a collection of data products derived from the Advanced Topographic Laser Altimeter System (ATLAS) instrument located on the Ice, Cloud and land Elevation Satellite-2 (ICESat-2) (National Snow and Ice Data Center 2023). The ATLAS instrument utilises a laser divided into six beams, organised into three pairs, to measure the slope of the Earth's surface accurately. These pairs are located roughly 3.2 km apart, while the distance inside the pairs is approximately 90 meters (see Figure 5). The intended use for this data was to use them as calibration and validation data, either only for determining lake edges or also, as Datta and Wouters (2021) showed in their study, for depth estimations. However, due to the limitation of the data itself, this data can be used only for SGLs located within these tracks. Therefore, the probability of overlap between SGLs and measured data is relatively low. Furthermore, the intersections with lakes can occur only within regions with similar lakes, given that flight paths mostly align with the north-south direction.

Another limitation of data is possible to see when the data is plotted in a scatter plot (Figure 6). Although the tracks occur in the area with SGLs, IceSAT-2 data often have gaps in them, in which the available photon data bounces between various heights and often shows only noise, sometimes even though data doesn't bounce, the visually identified surface of the iceberg starts to fade away. This nature of data further reduces the possibility of usage for the purpose of reference.

The Watta algorithm (Datta, Wouters 2021), after necessary adjustments, was used for lake identification and providing depth measurements. However, the outcomes generated by this algorithm exhibit on provided extent limited effectiveness and are unusable for further analysis. Specifically, the identification of the lakes by this algorithm identifies the presence of numerous lakes in an area where no SGLs are visually identifiable from Landsat 8 imagery (Figure 7). These areas typically correspond to small valleys within the ice sheet, showing a minimal amount of water, while the algorithm does not consider many bigger SGLs.

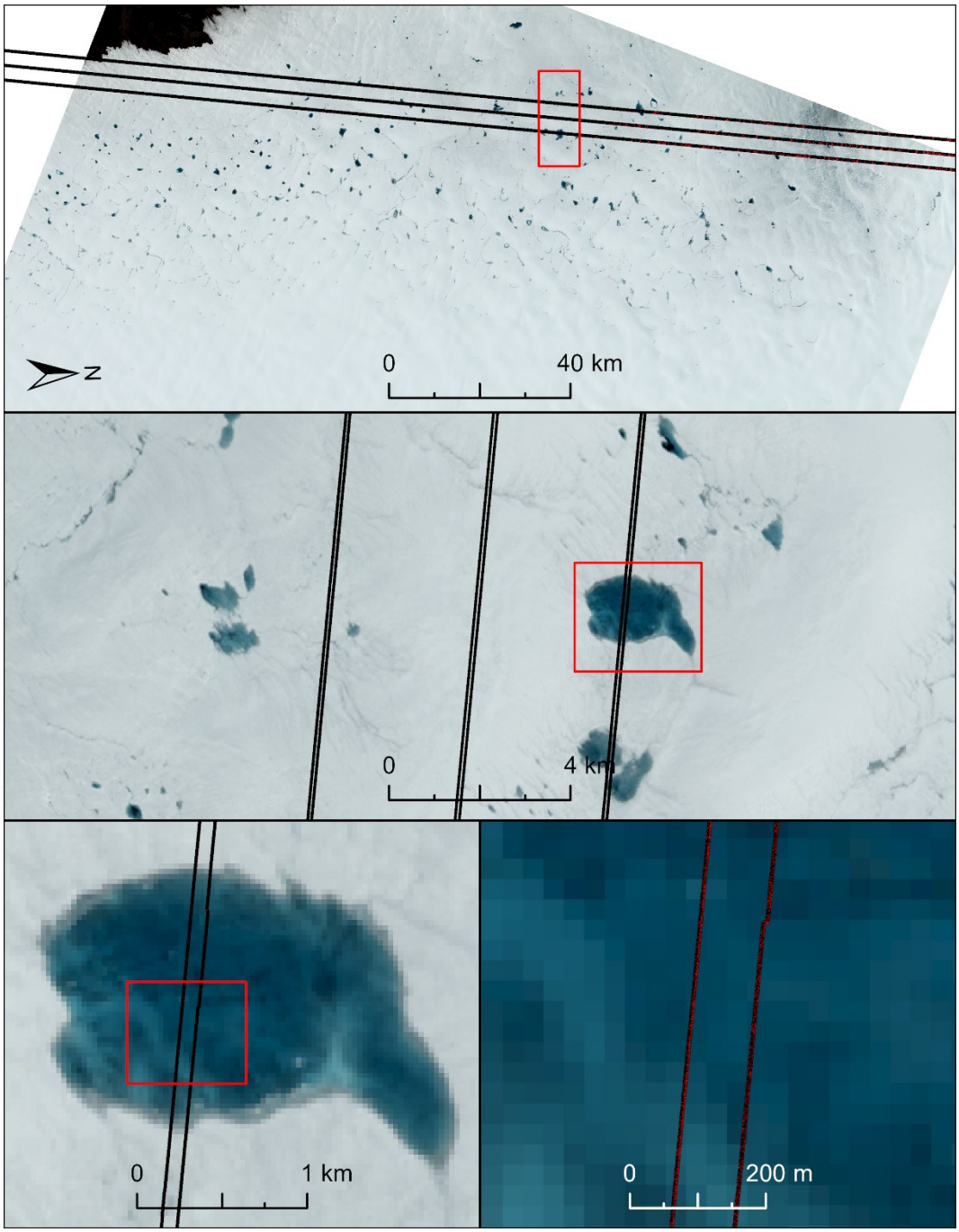


Figure 5: Example of ATL03 data (red point layer) in the area with occurred SGLs on 8th July 2021. Top image with different orientation, others are oriented to the north. Source: USGS (2023), Neumann, T. A., A. Brenner, et al. (2021).

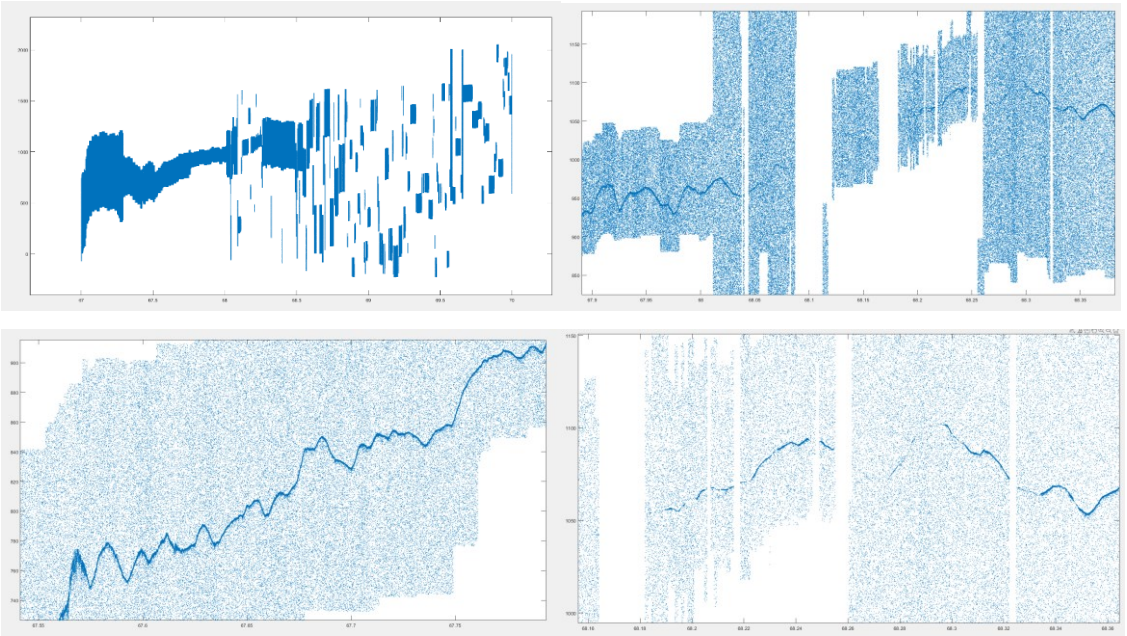


Figure 6: Examples of plotted IceSAT-2 ATL03 data in the scatter plot. Source: Neumann, T. A., A. Brenner, et al. (2021).

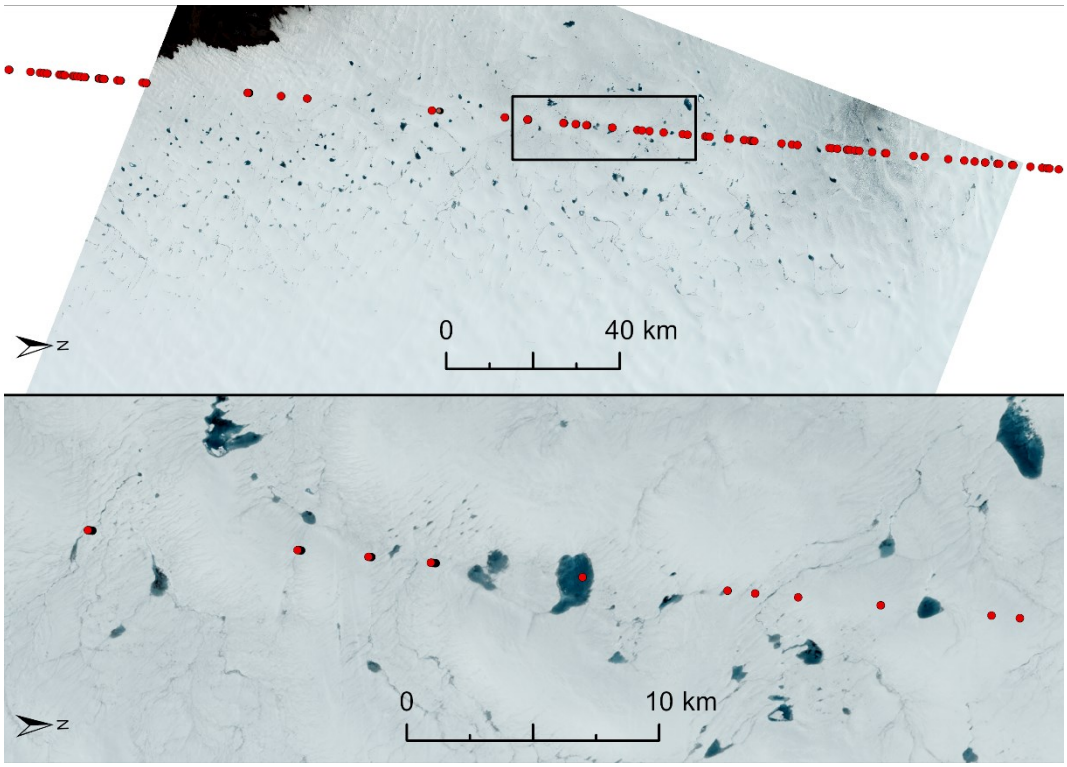


Figure 7: Identified SGL data (red dots) from ATL03 IceSAT-2 dataset by Watta algorithm. Source: USGS (2023), Neumann, T. A., A. Brenner, et al. (2021).

In one of the few instances of an accurately identified SGL (Figure 8), the algorithm identified the depths reaching approximately 2 meters. Although this SGL depth corresponds with the expected value, it is worth noting that the scatterplot reveals

only a limited number of data points upon which the calculations are based. Therefore, the output from the algorithm is regarded as overly optimistic due to substantial noise within the product ATL03. Consequently, the algorithm's performance is excluded from further calculations of SGLs, as their inclusions may introduce additional potential inaccuracies.

Datta and Wouters (2021) demonstrate the successful automatic detection of a wide range of SGL types using their algorithm in their study. However, they acknowledge the challenge of false positive SGL identification in sloping surfaces, which needs visual verification. These results correspond with observations found in this study. In general, the usage of IceSAT-2 data for verifying the depths of SGLs holds promise, especially with provided results (Figure 9); however, further studies on the identification of SGLs and obtaining accurate depth measurements need to be done.

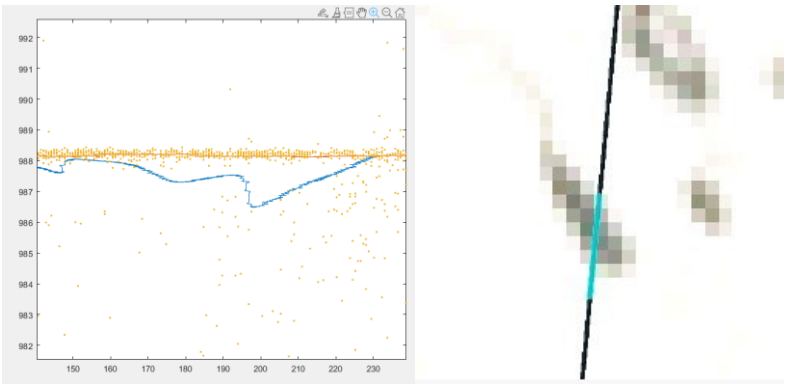


Figure 8: Small SGL identified by Watta algorithm (right) showing in blue used part of data for depth analysis and scatter plot of depth analysis of SGL (left) showing photon data (orange), surface elevation (red) and lakebed (blue). Source: USGS (2023), Neumann, T. A., A. Brenner, et al. (2021).

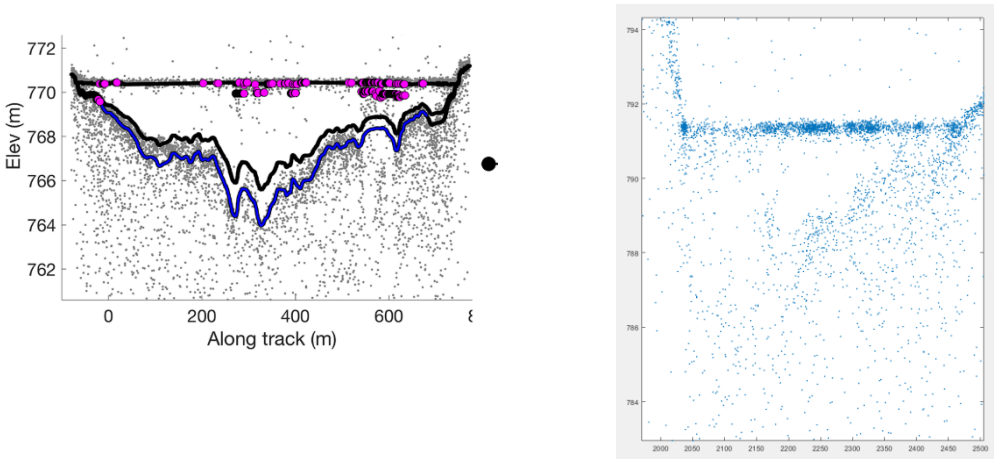


Figure 9 Raw ICESat-2 product and output of Watta script. Left: Output of Watta script. Right: Scatterplot of ATL03 data over identified SGL. Source: Datta (2021), USGS (2023), Neumann, T. A., A. Brenner, et al. (2021).

8.2. Parameter g estimation

Regression analysis of the g parameter for model calibration on an SGL was conducted on a lake used in the study by Pope et al. (2016). In-situ data, providing lake depth measurements for four days between 2nd and 5th July 2010, and two Landsat 7 TOA corrected images in the red band for 7th and 9th July 2010 (Figure 10) were used as input into the regression analysis. Landsat 7 images from multiple dates were used for analysis due to discontinuity in the data, also presented in the image's metadata gap mask. This gap passes through the measured lake and also part of the in-situ measurements in the case of the image from 7th July 2010 and for a small portion at the edge of the lake and the southwest part of the lake surroundings.

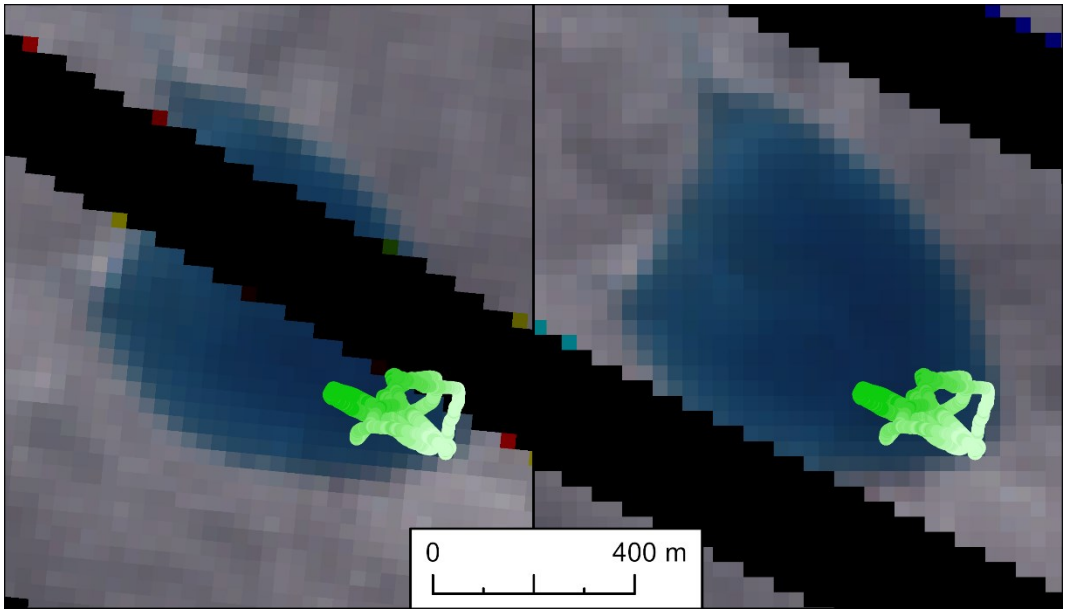


Figure 10: Landsat 7 TOA RGB composite image showing boat data for all 4 measurement days (green) for 7th July 2010 (left) and 9th July 2010 (right). Source: USGS (2023), Tedesco, Steiner, Pope (2015).

The outlier values of the in-situ data from 5th July 2010 and the image from 9th July 2010 (Figure 11) were from some regressions excluded. These outlier data were probably caused due to the pixel of the lake edge consisting of mixed lake and ice land cover. Other excluded data was part of the data from 2nd July 2010, showing the depth of the SGL over 300 meters, which is an obvious measurement error.

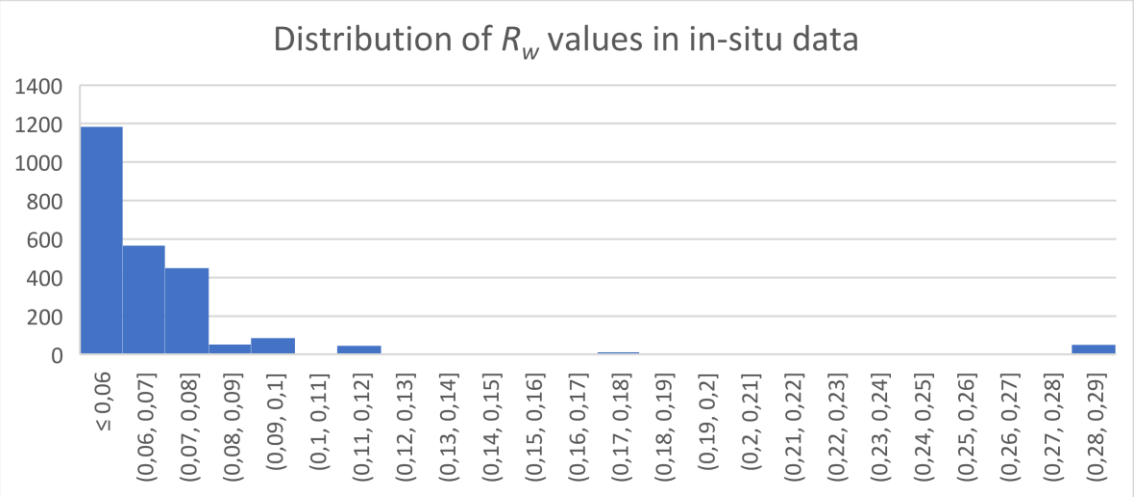


Figure 11: Histogram of in-situ measurement values from 5th July (y-axis) compared to R_w Landsat B3 TOA reflectance for the image from 9th July 2010. Source: USGS (2023), Tedesco, Steiner, Pope (2015).

Regression of parameter g based on multiple input parameters (Attachment 1) brings the value of parameter g into values between 0.866 – 0.927. Overall, the regressed values of the g parameter tend to be higher if the image from 9th July 2010 was used, while this image's RMSE values are lower. The lowest RMSE value 0,3375 was found in the dataset of in-situ data from 5th July 2010 with matching data points on both images, calculated on the image from 9th July 2010. Only by a small fraction, the second lowest RMSE value belongs to the combination of the dataset from 5th July 2010 (which excludes one edge value) and the image from 9th July 2010. Based on the smallest RMSE values and highest R^2 values, for further Monte Carlo simulations were chosen regressed g parameters from the 9th July image with all in-situ points (last three rows of Attachment 3).

Results of Monte Carlo simulation for the selected data show that when comparing the results of regressed g value, they fit best to the buffer of 0 – 30 meters around the lake used for regression analysis of mentioned parameter, based on the lowest RMSE value (Attachment 4). However, the differences between other buffer values are minimal. Similarly, the value differences are negligible in the case of the R^2 value.

However, in the simulation for A_d obtained from the average from all SGLs surroundings on Landsat 7 image (obtained from supervised classification, see chapter 8.3.), the RMSE show a significant increase in its values of around 0.17 m compared to

the values of lake buffers of the regressed lake.

For further analysis in this study, the regressed g parameter obtained from the combination of in-situ data for 5th July 2010 and Landsat 7 TOA image band 3 from 9th July 2010, with the value 0.893, will be used. The fitting curve of this parameter g shows good fitting with values in the range of 2 to 4 meters; however, for depths less than 2 meters, the curve starts to overestimate the actual depth (Figure 12). This discrepancy could be caused by the in-situ depth measurements, which were mainly taken at depths larger than 2 meters, while depths less than 1 meter were not measured. Additionally, the shape of the curve is governed by the nature of the physical model. A better fit could potentially be achieved by using an empirical model.

Compared to the other studies, used parameter g, with a value of 0.893, shows its value as one of the highest. A comparable value of this parameter, 0.868, was presented by the study of Morriss et al. (2013), which was also obtained from fitting the g parameter. A study by Moussavi et al. (2016) established the g parameter for ETM+ Band 3 as 0.69.

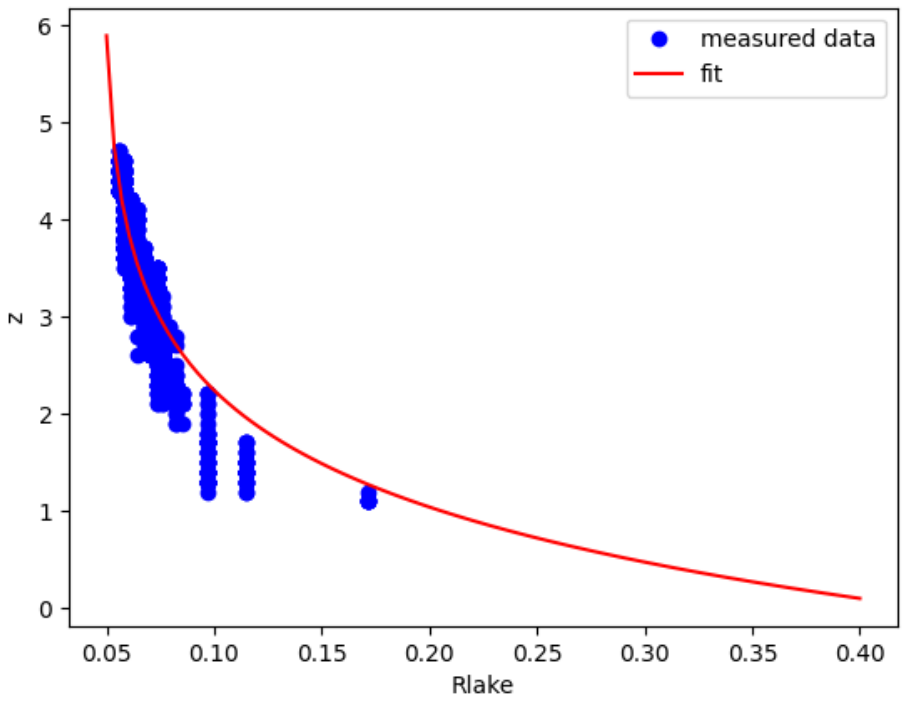


Figure 12: Regression analysis of the g parameter for in-situ SGL depth data from 5th July 2010 (y-axis [m]) and TOA corrected red band reflectance of Landsat 7 image from 9th July 2010. Source: USGS (2023), Tedesco, Steiner, Pope (2015).

8.3. Identifying SGL in the Landsat scene

This chapter presents the classification results and accuracy assessment of SGLs in the Landsat 7 TOA scene. The classification resulted in 726 identified SGLs in the Landsat scene, with most of them shown on maps in Attachments 1 and 2. The SGLs are provided as a polygon layer attached to this thesis, with each lake labelled by a unique identifier called *LAKE_ID*. This identifier allows for easy reference to individual lakes in subsequent text sections. Notice that some lakes, in the area of Landsat 7 SLC failure, may be split into multiple polygons due to the gap in the image data. The gaps in data are visible in the attached maps as black lines.

The accuracy of classification compared to the equalised stratified random points (Table 3) is 0.9765, and the F1 score for SGL water classification is 0.976. At the same time, the Kappa index has a value of 0.953. The precision of the SGL classification is 0.957, and the recall is 0.996.

Table 3: Confusion matrix of classified image for 2000 equalised stratified random points.

	SGL Water	Iceberg	Total
SGL Water	957	4	961
Iceberg	43	996	1039
Total	1000	1000	2000

The accuracy of image classification compared to the random points created from the validation dataset (Table 4) is 0.984, and the F1 score for SGL water classification is 0.981. At the same time, the Kappa index has a value of 0.967. The precision of the SGL classification is 0.998, and the recall is 0.963.

Table 4: Confusion matrix of classified image for 3079 random points created from validation dataset.

	SGL Water	Iceberg	Total
SGL Water	1232	47	1279
Iceberg	2	1798	1800
Total	1234	1845	3079

Above mentioned indicators of the classification show the high accuracy of all parameters. Although the accuracy indicators achieve high accuracy values, there is still possible to see some problems in classification. Most wrongly classified pixels among created control points resulted only at the lake's edge (Figure 13). This may be due to the consequences of manual vectorisation, where the edges of the lakes were estimated from the image. Future studies could benefit from employing more precise methods to define lake edges, such as the ICESat-2 or LIDAR measurements.

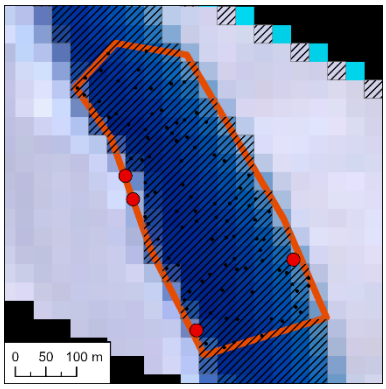


Figure 13: Example of wrongly classified lake control points near the lake's edge. Control points – black (true positive), red (false negative), Outline of train reference polygon (red) and Classified SGL water (hatched area) over Landsat 7 TOA image. Source: USGS (2023).

Another classification inaccuracy, mostly not noted from control points, but is visible when manually inspecting the image, is the problem of classification in wet ice areas (Figure 14). In these places, the classified image resembles noise. Most of these problems were solved by placing new training polygons inside these areas; however, some smaller areas with this problem still occur. Therefore, to create polygons of SGL lakes minimum area was established as a threshold for considering the area of classified water as SGL.

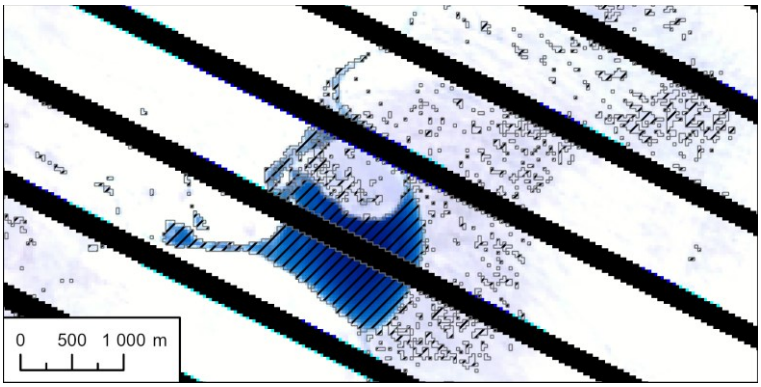


Figure 14: Example of a problem with classification in the wet ice area. Classified SGL water (hatched area) over Landsat 7 TOA image. Source: USGS (2023).

8.4. A_d variability

8.4.1 Impact of A_d variability on lake volume

Global parameters of A_d were firstly calculated based on previously determined parameters for further calculations. The results for global A_d parameters are as follows:

- Absolute lowest local A_d parameter in 0-90m distance – 0.3356.
- Average value of local A_d parameter in 0-90m distance – 0.6506.
- Median value of local A_d parameter in 0-90m distance – 0.69256.
- Absolute highest local A_d parameter in 0-90m distance – 0.7368.
- Value of local A_d from lake used for obtaining g parameter in 0-90 m distance – 0.4115.

After obtaining all necessary parameters, all lake depths within the study area were calculated. The problem with some lake pixels within SGL where the reflectance of Landsat 7 image was lower than R_∞ occurred. This is caused due to a negative logarithm from the physical model. The solution for this problem is that all SGL pixels impacted by this problem are considered as 7 meters deep. This was determined as the closest approximation of resulted depth from multiple A_d values based on defined g value 0.893, R_∞ value 0.048 and R_w value 0.049 (Table 5).

Table 5: Relationship between various A_d values on resulted z values [m] in conditions of R_w value being close to R_∞ value.

A_d	0.3356	0.4115	0.6506	0.69256	0.7368
z	6.33994	6.60190	7.16825	7.24363	7.31797

Variability of A_d parameter strongly influences the resulting lake depth and volume estimates. As expected, the most prominent contrast is between the usage of the absolute lowest and highest value of A_d as a global parameter. With the lowest value of A_d parameter, approximately 240 lakes from the total dataset did not exhibit any pixels with a positive value after the calculation of depths. This led to the conclusion that these lakes do not contain any water, and as a result, they were not considered as water bodies in the model analysis. The lakes most affected by this were mostly smaller shallow lakes located in areas with the highest reflectance of the surrounding glacier. If the used

value was the value of A_d from the lake used for calibration and with in-situ data, the number of not classified lakes is around 170. All other values from this study produced at least some pixels for lake volume calculations. An example of such a lake is lake 268 (Attachment 6), in which volume is hugely overestimated, in the case of global A_d with value 0.4115 (in-situ lake), reaching only 2% of volume from other methods. The global maximum A_d value achieves comparable results to more distant local A_d values, and this is mainly because in the more distant surroundings of this lake, the reflectivity of the glacier reaches high values.

Another way around if the maximum global A_d parameter is used in an area with very dirty ice, for example, as in the case of Lake 458 (Attachment 6), it is notable that (as well as global average and median A_d values) overestimate the lake volume by approximately 200 %, compared to the locally obtained A_d values. The lake's buffer 0-90 meters corresponds with the global minimum A_d value, which brings resulted lake volume, together with other local methods, very close to each other.

Lake 343 (Attachment 6) is the lake, which was used to calibrate the g parameter with A_d obtained from 0-90 meters buffer around the lake. Consequently, the parameters and results of local A_d derived from 0-90 meters and global A_d for the in-situ lake are nearly identical. The most significant difference in lake volumes is observed when comparing mentioned A_d parameter with the maximum global A_d value, resulting in a 140% increase in lake volume.

The most significant differences in dependence of local A_d value on lake volume can be seen in areas where parts of the surrounding iceberg exhibit lower reflectance values, while other parts show higher reflectance values. An example of such variation can be seen in Lake 98 (Attachment 6). This is also enhanced by the relatively small size of the lake. If the buffer of 1 pixel around the lake edge is compared to 3 pixels around the lake edge, the difference in the resulting lake volume results in around 30% change.

The sensitivity to changes in A_d value can be observed more in smaller and shallower lakes. Compared to smaller Lake 558 (Attachment 6) and larger Lake 538 (Attachment 6), globally obtained A_d values were, compared to locally obtained A_d values, making a much more significant difference in lake volume. However, in this

example, the locally obtained A_d value provides smaller variability in resulted lake volumes due to more consistent surroundings around the smaller lake.

Another factor influencing the SGL volume estimates is the presence of pixels with a R_w value lower than the R_∞ value. Due to the influence of the specified depth value of 7 metres for these pixels (chapter 7.6.), lakes in which the proportion of these pixels is high show a lower rate of change in lake volume estimates. This can be demonstrated on Lakes 604 and 610 (Attachment 6), where the smaller lake (604) shows less change in lake volume, with the most significant difference between predicted volumes reaching around 14%. Lake 610, the largest classified lake, also offers only a small difference in predicted volume; however, with 18%, it is slightly higher.

Comparison of the average and median difference in predicted lake volume for all lakes (Attachment 5), especially for globally obtained A_d parameter, is heavily influenced by smaller lakes, which disproportionately distort the values of lake volume difference. However, for locally obtained A_d parameter for the average difference in volume between the first pixel (0-30 m) versus the second (30-60 m) and third pixel (60-90 m) was approximately 120.4 and 123.3 %, respectively 114.8 and 115.4 % for median. This suggests distinct characteristics within the buffer surrounding the classified lake edge. A possible explanation could be an inaccurately defined outline of the lake, leading to the inclusion of additional features like wet ice or disturbances in the glacier structure. Further investigation, incorporating higher-resolution data or field surveys, is essential to investigate the reasons behind these variations and enhance the precision of A_d parameter calculations for estimating SGL depth and volume.

The summary of lake volume estimates for the entire Landsat 7 scene reveals an expected increase of SGLs volumes with similar patterns as in the previously shown examples of lakes (Table 6). In the case of local A_d values, the total volumes show a rise from approximately 176 mil. m^3 to 194 mil. m^3, which corresponds to an overall 10 % difference. However, when using global parameters, the volumes rise from approximately 124 mil. m^3 to 207 mil. m^3, which represents a difference of nearly 66 % from the lowest value.

Table 6: Summary of SGL volume across Landsat 7 scene based on different obtained A_d values.

Local A_d	Method	Volume [m³]	Global A_d	Method	Volume [m³]
	0-30	175 816 193		Minimum	124 285 255
	0-60	182 577 242		In-situ	142 197 756
	0-90	186 443 711		Average	191 261 856
	30-60	192 262 784		Median	199 355 752
	30-90	193 108 151		Maximum	207 392 719
	60-90	193 792 567			

In conclusion, the A_d parameter variability strongly influences the resulting lake depth and volume estimates. Different global and local A_d values in the physical model lead to significant variations in lake volume calculation (Figure 15), especially for smaller and shallower lakes.

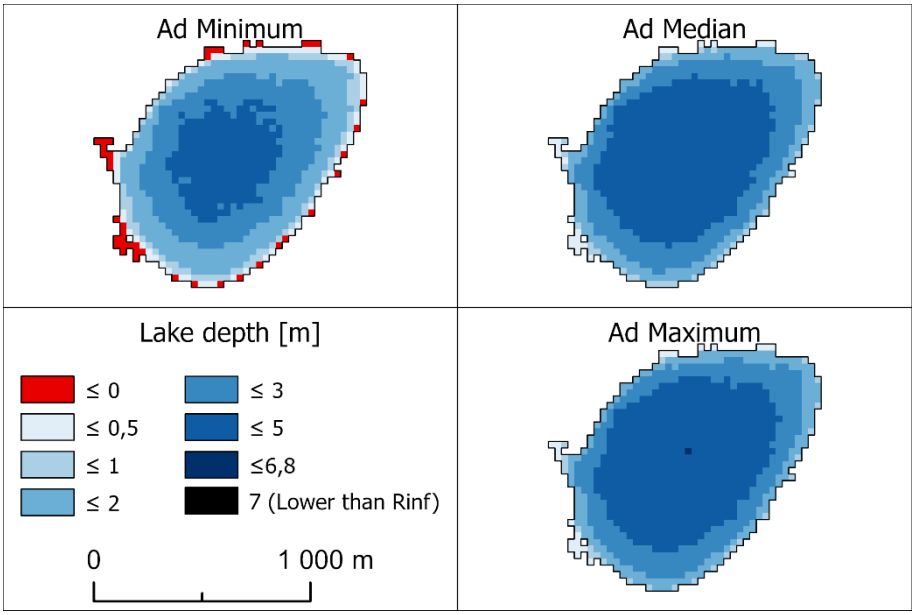


Figure 15: Multiple global A_d parameter variability influence on the SGL depths. Global A_d minimum (top left), median (top right) and maximum (bottom right) methods presented.

Keywords of all lakes which are presented in Attachment 6:

- 98 – Small lake, dirty ice, crescent shape
- 268 – Medium lake, shallow, very clear ice, elongated shape
- 343 – Lake with in-situ data, used for estimation of g parameter, medium lake, dirty ice, round shape
- 458 – Small lake, very dirty ice
- 538 – Medium lake, clear ice, round shape

- 558 – Medium lake, very clear ice, shallow, crescent shape (“dolphin”)
- 604 – Big lake, mostly dirty ice, deep lake, round shape
- 610 – Biggest classified lake, dirty ice, deep lake, crescent shape in the end and oval shape on another

8.4.2 Variability of A_d with distance from iceberg edge

Based on the analysis presented in the previous chapter, it is evident that the A_d parameter exhibits variability among different lakes. To enhance geographic comprehension of this phenomenon, the relationship between distance and observed A_d value was projected into a scatter plot (Figure 16). The plot shows an evident relationship between the distance from the glacier's edge and the value of A_d obtained from the vicinity of the SGL. This corresponds with the expectation that, at least in the case of Greenland, the westerly side of the glacier will be influenced by wind-borne particles, such as cryoconite. This effect is evident until approximately 40-45 km from the iceberg edge, until which the A_d value increases, and later the value remains balanced.

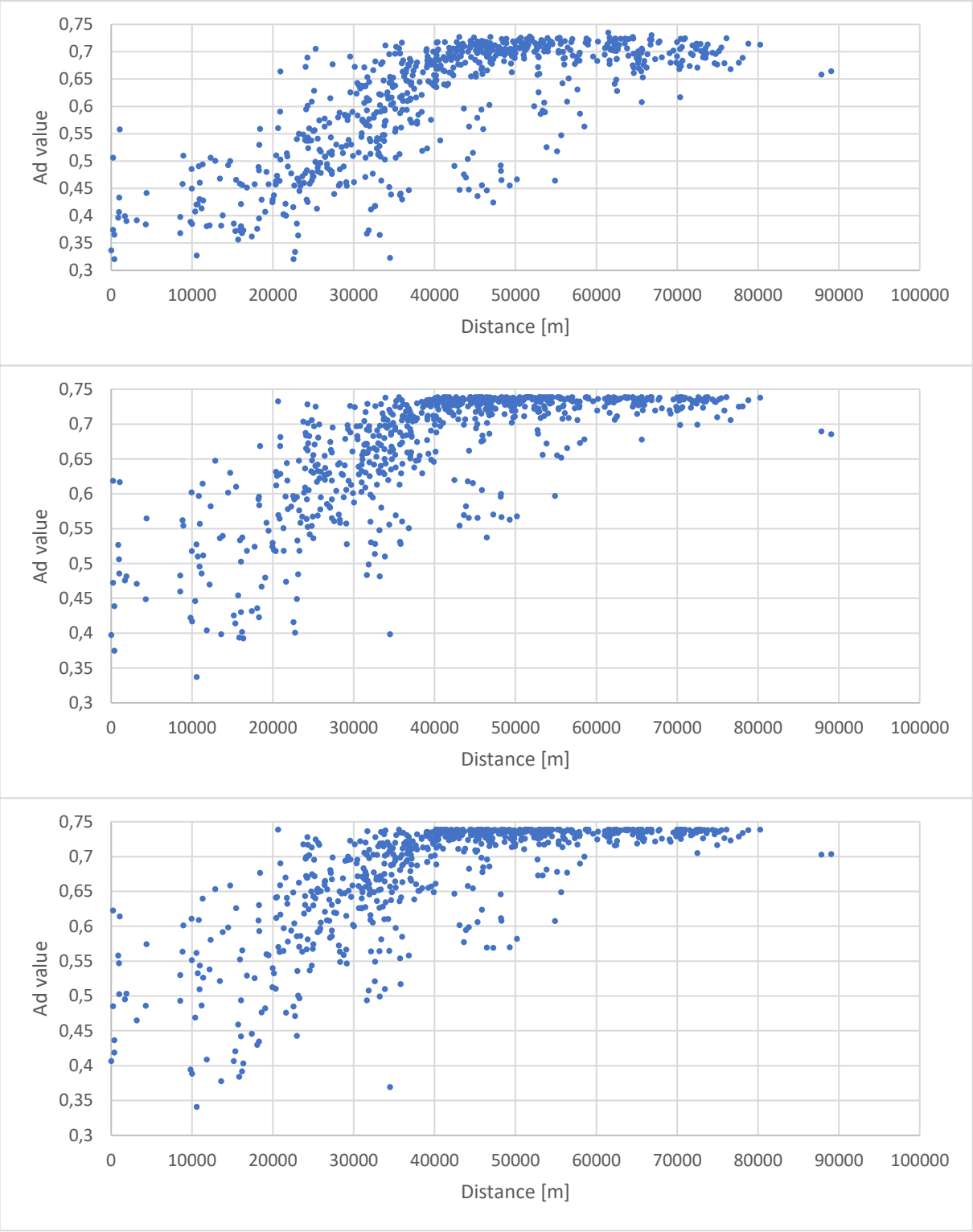


Figure 16: Relationship between distance from the edge of the iceberg and A_d value for various buffers around lake edge. Upper: 1px (0-30 m), Middle: 2px (30-60 m), Bottom: 3px (60-90 m).

8.5. Validation of the SGL depth retrieval

The resulting classification of the SGL was compared to 4675 individual in-situ measurements of depth, as well as aggregations of these measurements over 39 Landsat 7 pixels areas. The comparison was provided for all three measured days combined and each day individually.

The aggregated values over pixels by methods of averaging values of points inside the pixel and searching for the most frequent values showed similar results (Figure 17). Note that in-situ measurements in their original form are rounded to the nearest tenth of a metre. This nature of the data influences the resulting scatter plots of comparison shown later.

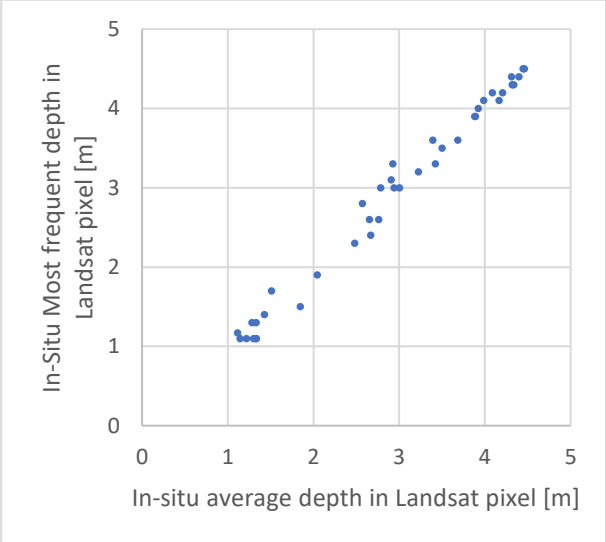


Figure 17: Comparison of two methods for aggregation multiple in-situ depth measurements into Landsat 7 pixel area.

The statistical indicators, in the case of aggregated values, show that the locally obtained A_d parameter for the physical model is obviously better for predicting SGL depth than global parameters (Attachment 7). Note that the global parameter "Lake" corresponds with the locally obtained parameters in distances 0 – 90 m; therefore, this global parameter will not be mentioned in the text below in the cases of the globally applied A_d parameter. In the area of RMSE values, the local parameters showed values around 0.33 – 0.35 m, while global parameters showed these values in the 0.42 – 0.79 m range R^2 values with local A_d parameters show values around 0.9, while global parameters show values in the range 0.49 – 0.86. MAE is also smaller in the locally

obtained A_d parameter, with values between 0.26 – 0.27 m. The global parameters show this value between 0.35 – 0.72 m. Considering the aggregation method of in-situ values into Landsat pixels, the average seems to bring lower MAE and RMSE values, while R^2 values are similar. Overall, the locally obtained A_d parameter 0-90 m shows, by the indicators, the most accurate results. Compared to the in-situ data (Figure 18), there is possible to see a good approximation of lake depths from 2 m depth up; however, under this value, the performance results are deteriorating.

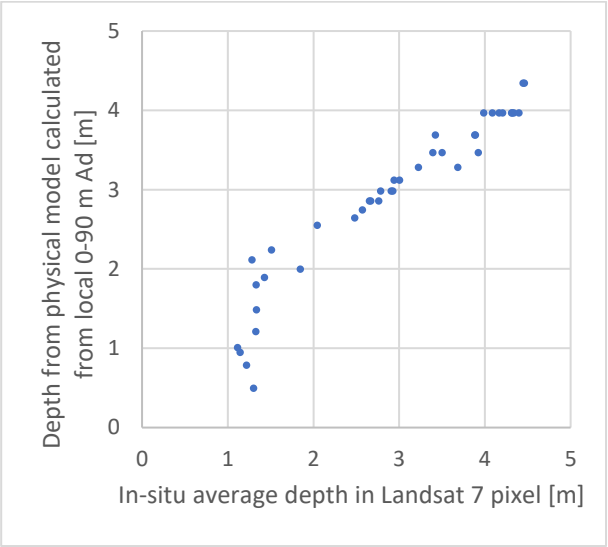


Figure 18: Comparison between in-situ aggregated SGL depth measurements into Landsat 7 pixel area and depths resulting from the physical model with A_d obtained as local reflectance in the distance 0-90 m from SGL edge.

The performance of the resulting depths from the physical model was also compared to in-situ point measurements from all dates. The overall results show a similar pattern regarding the indicators, as mentioned above (Attachment 8). The locally obtained A_d parameter for physical model values for RMSE varies from 0.37 – 0.4 m, R^2 around 0.85, and MAE between 0.3 – 0.33 m. The global A_d parameter for physical model values for RMSE varies from 0.47 – 0.79 m, R^2 between 0.36 – 0.77, and MAE is between 0.4 – 0.7 m. Compared to aggregated depths, the lowest MAE is here achieved from locally obtained A_d parameter from a distance 60-90 m around the lake edge; however, this difference is not significant. The lowest RMSE and highest R^2 value still reach the A_d parameter from a distance 0-90 m around the lake edge. Compared to the in-situ data (Figure 19), there is possible to see, likewise in the previous example, a good approximation of lake depths from 2 m depth up, followed by a worse approximation under these values.

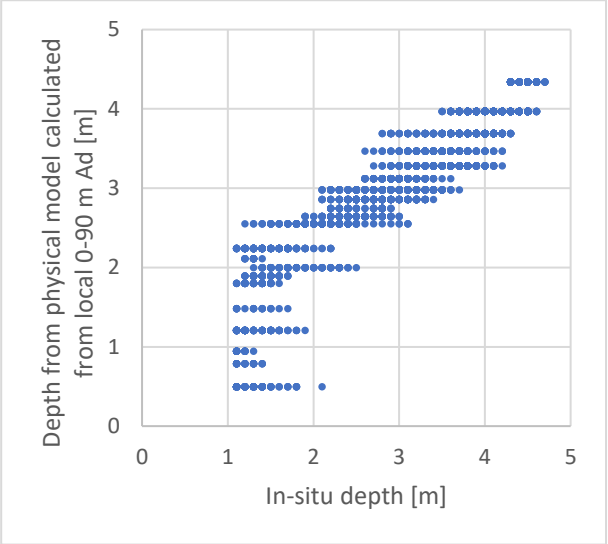


Figure 19: Comparison between in-situ all SGL depth measurements and depths resulting from the physical model with A_d obtained as local reflectance in the distance 0-90 m from the SGL edge.

Minor differences could be seen when the in-situ dataset was split into three dates. The approximated depths fit best the depths measured on 5th July. In the case of locally obtained A_d parameter, the other two dates show, in general, the higher values of RMSE and MAE and lower R^2 values accordingly to their time distance. This is the expected result due to the Landsat 7 TOA image being obtained on 9th July 2010, from which the closest day is 5th July and due to data from 5th July being used for parameter g calibration.

Overall, a physical model using locally derived A_d parameter from 0-90 m distance from the lake edge best approximates the in-situ data. This is especially the case of in-situ data from 5th July. This result can be caused mainly due to bias caused due to the g parameter of the model obtained from precisely the same conditions. However, looking at other locally derived A_d parameters shows their approximation is also very good. Some even outperform the previously mentioned 0-90 m distance parameter in certain indicators. The results also correspond well with other dates; however, other independent data would be beneficial for better validation of depths approximated by the physical model.

9. Conclusions and future development

The objective was to quantify the inaccuracies of SGL modelling and their influence on SGLs depth and volume estimation, specifically focusing on the A_d parameter influence on the physical model.

The results highlight the spatial variability of the A_d parameter, making it unsuitable to be represented by a single value for the entire scene. Using a global A_d parameter can lead to up to 66% inaccuracies. Notably, in the studied area of West Greenland, there is possible to see an increase in the reflectance of the glacier surrounding SGLs from the west to the east. The results show that the immediate surroundings of the SGLs have different reflectance compared to the surrounding of SGL further away, which results in approximately 10 % different volume estimates across the Landsat scene.

Therefore, estimating the A_d parameter is essential for understanding SGL modelling and significantly affects the depth and volume estimation accuracy. The spatial variability of the A_d parameter in the study area highlights the need for acquiring in-situ reference data from SGLs with varying albedo values. Only relevant field study was conducted in 2010 on one specific SGL (Tedesco, Steiner, Pope 2015), which is also shown in this thesis. ICESat-2 data brings many inaccuracies; however, there is still room for further research to improve the accuracy and applicability of these data.

Moving forward, developing a method that respects natural conditions with variable albedo values is crucial. This approach will improve the accuracy of SGL modelling and enhance the understanding of these glacial features. By accounting for the spatial variability of the A_d parameter with the integration of possibly newly obtained field data, a more reliable method can be obtained, advancing knowledge in SGL dynamics and their influence on glacial systems.

10. References

ASHLEY, G. M. (2002): Glaciolacustrine Environments. In: Modern and Past Glacial Environments: Revised Student Edition. Oxford, 335–359.

BAŤKA, J. (2015): Analýza supraglaciálních jezer: případová studie z pohoří Cordillera Blanca, Peru. Univerzita Karlova v Praze.

BOX, J. E., SKI, K. (2007): Remote sounding of Greenland supraglacial melt lakes: Implications for subglacial hydraulics. Journal of Glaciology, 181, 53, 257–265.

BRODSKÝ, L., VILÍMEK, V., ŠOBR, M., KROCZEK, T. (2022): The Effect of Suspended Particulate Matter on the Supraglacial Lake Depth Retrieval from Optical Data. Remote Sensing, 23, 14.

COHEN, J. (1960): A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 1, 20, 37–46.

CORR, D., LEESON, A., MCMILLAN, M., ZHANG, C., BARNES, T. (2022): An inventory of supraglacial lakes and channels across the West Antarctic Ice Sheet. Earth System Science Data, 1, 14, 209–228.

DATTA, R. T., WOUTERS, B. (2021): Supraglacial lake bathymetry automatically derived from ICESat-2 constraining lake depth estimates from multi-source satellite imagery. Cryosphere, 11, 15, 5115–5132.

ESRI (2023a): Create Random Points (Data Management), https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/create-random-points.htm (20. 7. 2023).

ESRI (2023b): Train Random Trees Classifier (Image Analyst), https://pro.arcgis.com/en/pro-app/latest/tool-reference/image-analyst/train-random-trees-classifier.htm (20. 7. 2023).

FITZPATRICK, A. A. W., HUBBARD, A. L., BOX, J. E., QUINCEY, D. J., VAN AS, D., MIKKELSEN, A. P. B., DOYLE, S. H., DOW, C. F., HASHOLT, B., JONES, G. A. (2014): A decade (2002-2012) of supraglacial lake volume estimates across Russell Glacier, West Greenland. Cryosphere, 1, 8, 107–121.

GEOFABRIK GMBH, OPENSTREETMAP CONTRIBUTORS (2022): Greenland - Geofabrik Download Server, http://download.geofabrik.de/north-america/greenland.html (19. 10. 2022).

GEORGIOU, S., SHEPHERD, A., MCMILLAN, M., NIENOW, P. (2009): Seasonal evolution of supraglacial lake volume from aster imagery. Annals of Glaciology, 52, 50, 95–100.

GUINNESS WORLD RECORDS LIMITED (2023): First person to swim beneath the Antarctic Ice Sheet, https://www.guinnessworldrecords.com/world-records/613580-first-person-to-swim-beneath-the-antarctic-ice-sheet (30. 6. 2023).

HU, J., HUANG, H., CHI, Z., CHENG, X., WEI, Z., CHEN, P., XU, X., QI, S., XU, Y., ZHENG, Y. (2022): Distribution and Evolution of Supraglacial Lakes in Greenland during the 2016 – 2018 Melt Seasons.

LI, D., SHANGGUAN, D., ANJUM, M. N. (2020): Glacial lake inventory derived from landsat 8 OLI in 2016-2018 in China-Pakistan economic corridor. ISPRS International Journal of Geo-Information, 5, 9.

MORRISS, B. F., HAWLEY, R. L., CHIPMAN, J. W., ANDREWS, L. C., CATANIA, G. A., HOFFMAN, M. J., LÜTHI, M. P., NEUMANN, T. A. (2013): A ten-year record of supraglacial lake evolution and rapid drainage in West Greenland using an automated processing algorithm for multispectral imagery. Cryosphere, 6, 7, 1869–1877.

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.

NATIONAL SNOW AND ICE DATA CENTER (2023): ICESat-2, https://nsidc.org/data/icesat-2 (10. 7. 2023).

NEUMANN, T. A., A. BRENNER, D. HANCOCK, J. ROBBINS, J. SABA, K. HARBECK, A. GIBBONS, J. LEE, S. B. LUTHCKE, T. REBOLD, et al. (2021): ATLAS/ICESat-2 L2A Global Geolocated Photon Data, Version 5., https://doi.org/https://doi.org/10.5067/ATLAS/ATL03.005.

PHILPOT, W. D. (1989): Bathymetric mapping with passive multispectral imagery. Applied Optics, 8, 28, 1569.

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, 1, 10, 15–27.

QIAO, L., MAYER, C., LIU, S. (2015): Distribution and interannual variability of supraglacial lakes on debris-covered glaciers in the Khan Tengri-Tumor Mountains, Central Asia. Environmental Research Letters, 1, 10.

RAFFERTY, J. P. (2023): Beer's law, Encyclopædia Britannica, https://www.britannica.com/science/Beers-law (20. 7. 2023).

SELMES, N., MURRAY, T., JAMES, T. D. (2011): Fast draining lakes on the Greenland Ice Sheet. Geophysical Research Letters, 15, 38, 1–5.

SHUGAR, D. H., BURR, A., HARITASHYA, U. K., KARGEL, J. S., WATSON, C. S., KENNEDY, M. C., BEVINGTON, A. R., BETTS, R. A., HARRISON, S., STRATTMAN, K. (2020): Rapid worldwide growth of glacial lakes since 1990. Nature Climate Change, 10, 10, 939–945.

SNEED, W. A., HAMILTON, G. S. (2007): Evolution of melt pond volume on the surface of the Greenland Ice Sheet. Geophysical Research Letters, 3, 34.

SNEED, W. A., HAMILTON, G. S. (2011): Validation of a method for determining the depth of glacial melt ponds using satellite imagery. Annals of Glaciology, 59, 52, 15–22.

STOKES, C. R., SANDERSON, J. E., MILES, B. W. J., JAMIESON, S. S. R., LEESON, A. A. (2019): Widespread distribution of supraglacial lakes around the margin of the East Antarctic Ice Sheet. Scientific Reports, 1, 9, 1–14.

SUNDAL, A. V., 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, 10, 113, 2164–2171.

TEDESCO, M., STEINER, N. (2011): In-situ multispectral and bathymetric

measurements over a supraglacial lake in western Greenland using a remotely controlled watercraft. The Cryosphere, 2, 5, 445–452.

TEDESCO, M., STEINER, N., POPE, A. (2015): In situ spectral reflectance and depth of a supraglacial lake in Greenland, Arctic Data Center, https://doi.org/10.5065/D6FQ9TN2 (10. 4. 2020).

USGS (2022): Using the USGS Landsat Level-1 Data Product, https://www.usgs.gov/landsat-missions/using-usgs-landsat-level-1-data-product.

USGS (2023): Landsat 7 | U.S. Geological Survey, https://www.usgs.gov/landsat-missions/landsat-7?qt-science_support_page_related_con=0#qt-science_support_page_related_con (30. 6. 2023).

WILLIAMSON, A. G., ARNOLD, N. S., BANWELL, A. F., WILLIS, I. C. (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.

11. Attachments

Attachment 1: Classified SGLs over the west part of Landsat 7 TOA Image from 9.7.2010 with enhanced lakes used as an attachment. Source of image: USGS (2023).

Attachment 2: Classified SGLs over the south part of Landsat 7 TOA Image from 9.7.2010 with enhanced lakes used as an attachment. Source of image: USGS (2023).

Attachment 3: Results of Regression of g parameter based on input in-situ data subset and image, with provided number of in-situ points used for regression.

Attachment 4: Results of Monte Carlo simulations for provided regressed g parameter.

Attachment 5: Average and median values of differences in estimated lake volume [%] by used parameter A_d in the physical model for SGL depth estimation.

Attachment 6: Differences in estimated lake volume [%] by used parameter A_d in the physical model for SGL depth estimation for particular lake (lake number written in the corner of the table) and figures of lake; From top left to the right: Lake outline over Landsat 7 TOA image from 9th July 2010, lake depth estimations for global parameter of Ad, with the name of the parameter on the top of the figure; bottom from left to right: legend of figures, lake depth estimations for locally obtained parameter of Ad, with the distance from the edge of the lake on the top of the figure. Together eight pages in the following order:

- Lake 98
- Lake 268
- Lake 343
- Lake 458
- Lake 538
- Lake 558
- Lake 604
- Lake 610

Source of image: USGS (2023).

Attachment 7: Statistical indices to validate the resulting physical model estimation as a function of the parameter A_d for 4675 in-situ measurements aggregated over 39 Landsat 7 pixels area.

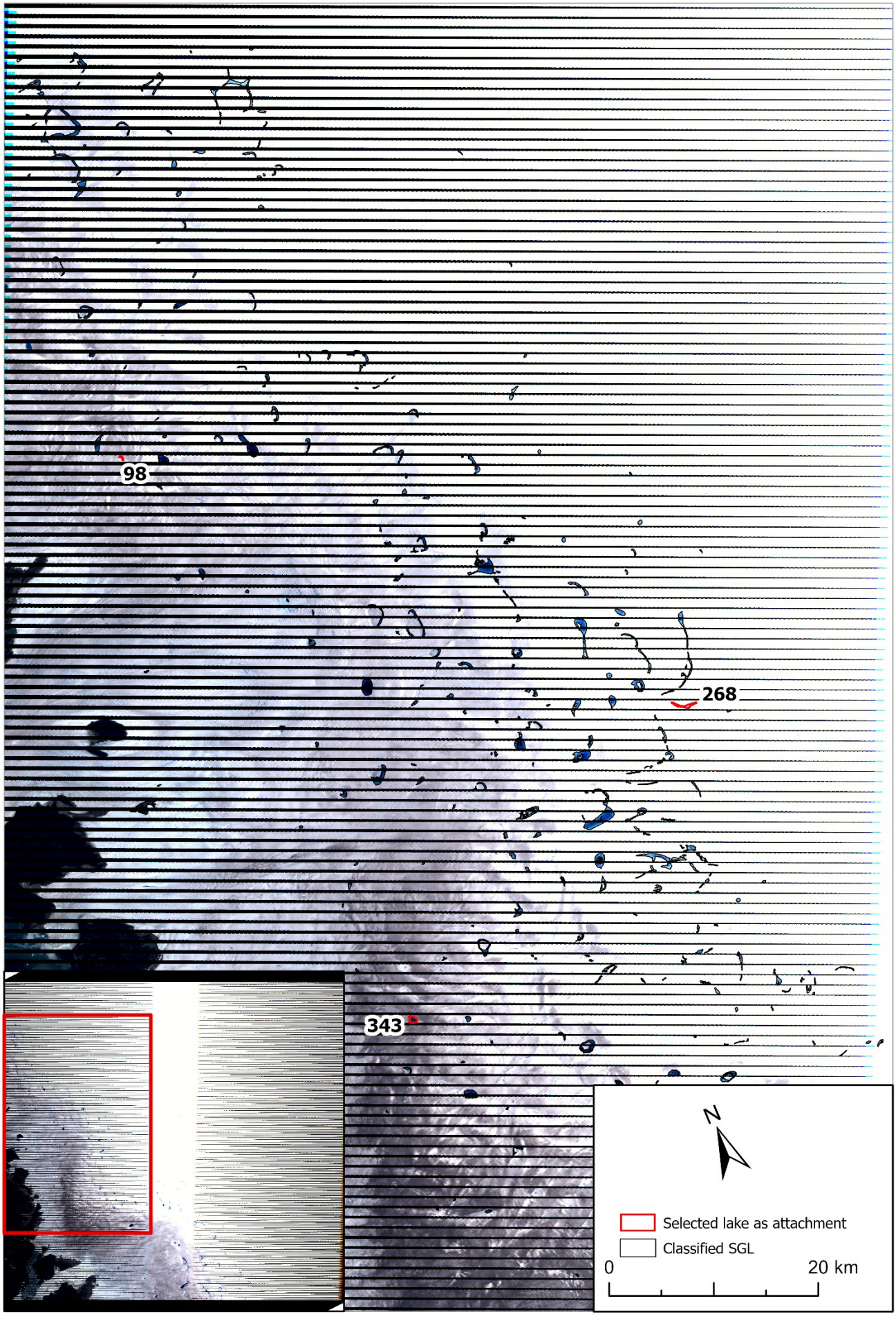
Attachment 8: Statistical indices to validate the resulting physical model estimation as a function of the parameter A_d for 4675 in-situ measurements.

11.1. Electronic attachments

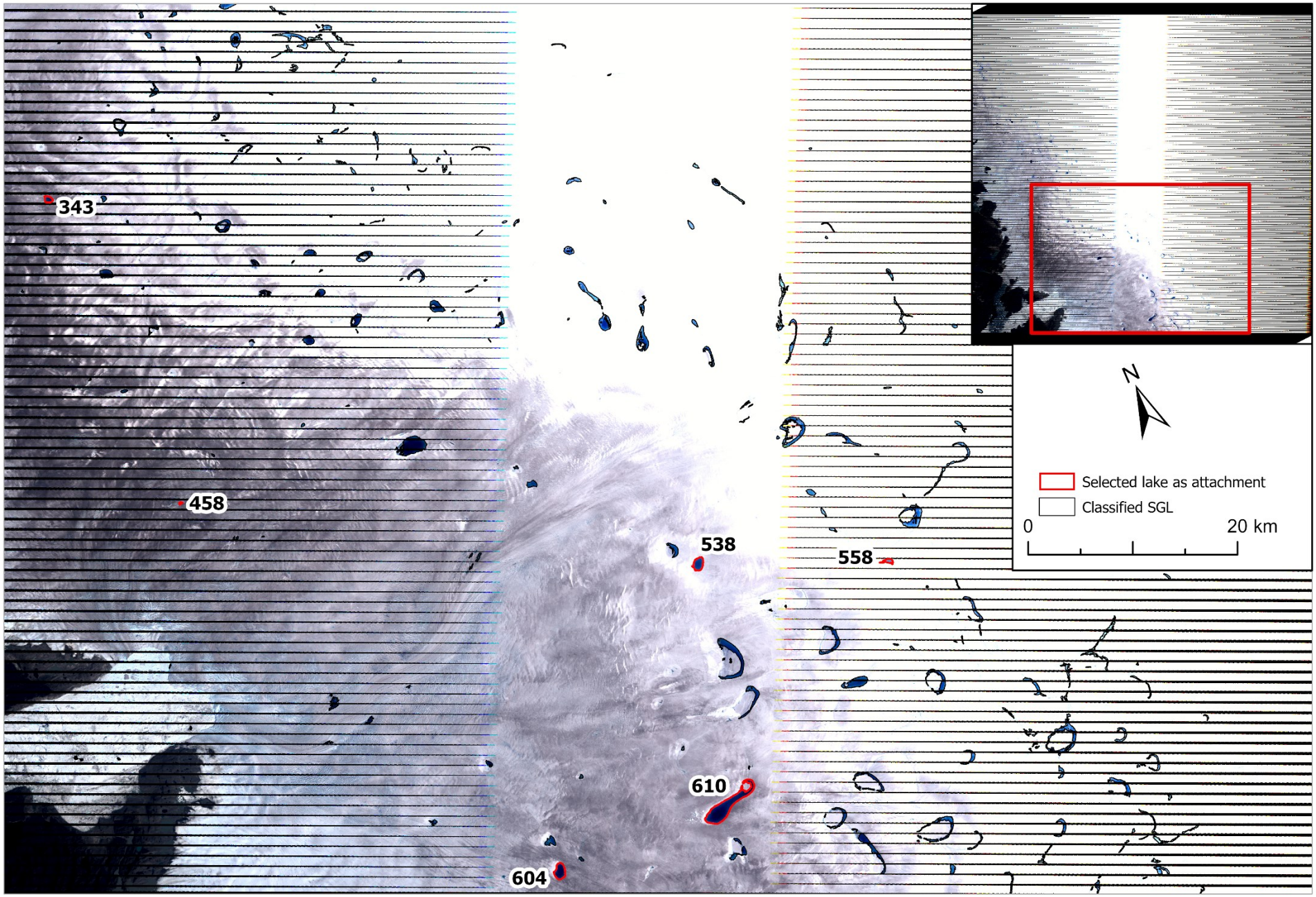
SGL.shp - Shapefile of 726 SGL lakes, resulting from supervised classification. Each lake contains information about the local Ad parameter, resulting lake volume and average depth obtained from the physical model. Metadata provided in attached text file *SGL_Meteadata.txt*.

SGL_Volume_Statistics.xlsx - Interactive Excel worksheet with lake volume comparison, used for the creation of attachment 5, as well as information about A_d values used in calculating the lake volume and distance of SGL from the iceberg edge. Metadata is included in the worksheet.

Attachment 1: Classified SGLs over the west part of Landsat 7 TOA Image from 9.7.2010 with enhanced lakes used as an attachment.



Attachment 2: Classified SGLs over the south part of Landsat 7 TOA Image from 9.7.2010 with enhanced lakes used as an attachment.



Attachment 3

Results of regression of g parameter based on input in situ data subset and image, with provided number of in situ points used for regression		Number of points used for g parameter regression			g regression for B3 image from 7th July 2010			g regression for B3 image from 9th July 2010		
In situ data subset		All	Training	Validation	regressed g	RMSE	R2	regressed g	RMSE	R2
In situ data, contained within images from 7th and 9th July 2010	5th July, all points	2247	1572	675	0.865792	0.4235	0.7737	0.888775	0.3575	0.8387
	5th July, without extreme value B3 pixel (0.281264)	2197	1537	659	0.867188	0.4131	0.7548	0.889271	0.3375	0.8362
	2nd, 3rd and 5th July, all values (expect 2nd July values 300)	4059	2841	1218	0.883691	0.4418	0.7946	0.893512	0.3812	0.8471
	2nd July, all values (expect 2nd July values 300)	602	421	181	0.894482	0.4916	0.7875	0.898770	0.4167	0.8473
	3rd July, all values	1210	847	363	0.92705	0.4098	0.8122	0.905460	0.4019	0.8187
In situ data contained within image from 9th July 2010	5th July, all points	2448	1713	735				0.893007	0.3583	0.8406
	5th July, without extreme value B3 pixel (0.281264)	2397	1677	720				0.892766	0.3376	0.8366
	2nd, 3rd and 5th July, all values (expect 2nd July values 300)	4675	3271	1403				0.896655	0.3706	0.8570

Attachment 4

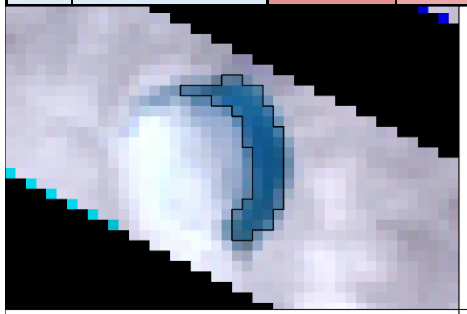
Results of Monte Carlo simulations for provided regressed g parameter	g parameter regression						Monte Carlo simulation							
							Lake with In-situ data						All lakes	
In situ data subset	Number of points for g parameter regression			Resulted g parameter and statistical indicators			0-30 Buffer		0-60 Buffer		0-90 Buffer		0-90 Buffer	
	All	Training	Valid.	g	RMSE	R2	RMSE	R2	RMSE	R2	RMSE	R2	RMSE	R2
;th July, all points	2448	1713	735	0.893007	0.3583	0.8406	0.3649	0.8306	0.3677	0.8321	0.3755	0.8249	0.5297	0.6515
5th July, without extreme value B3 pixel (0.281264)	2397	1677	720	0.892766	0.3376	0.8366	0.3493	0.8252	0.351	0.8235	0.3536	0.8209	0.5392	0.5835
2nd, 3rd and 5th July, all values (expect 2nd July values 300)	4675	3271	1403	0.896655	0.3706	0.8570	0.3812	0.8486	0.383	0.8472	0.3876	0.8435	0.5581	0.6755

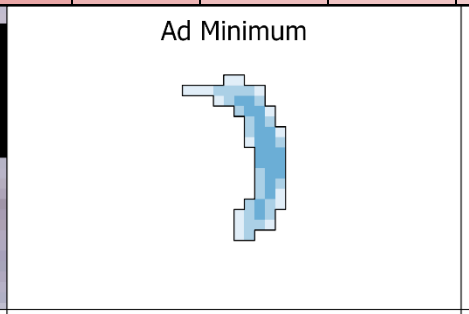
Attachment 5

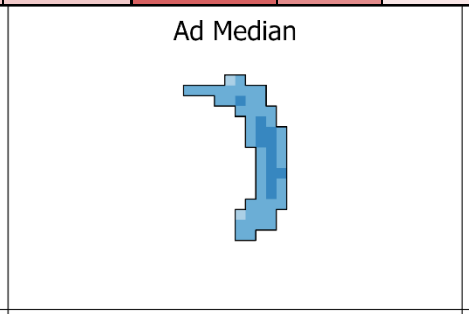
AVERAGE		Local A_d [m]						Global A_d				
		0 - 30	0 - 60	0 - 90	30 - 60	30 - 90	60 - 90	Minimum	In-Situ	Average	Median	Maximum
Local A_d [m]	0 - 30	100	107.9952	113.0454	120.4474	121.766	123.2611	29.06331	41.02111	110.15	148.4238	186.8041
	0 - 60	92.86275	100	104.3826	110.7801	111.8912	113.1573	27.32168	38.40958	99.95615	125.9317	151.9053
	0 - 90	89.40607	96.02433	100	106.0485	106.9441	108.1297	26.3832	37.00015	96.14738	123.0119	149.8596
	30 - 60	84.9423	91.04393	94.74622	100	100.912	101.8319	25.29693	35.41658	90.49397	115.1565	139.8111
	30 - 90	85.01154	90.77123	94.15176	99.38156	100	100.804	25.08049	35.0822	93.71265	130.6815	167.5341
	60 - 90	87.7037	92.321	94.75739	99.76201	99.58995	100	24.93974	34.85899	110.8915	199.3413	287.2222
Global A_d	Minimum	5474.816	5811.643	6021.448	6350.5	6397.587	6434.155	100	510.0684	4733.154	5690.593	6642.83
	In-Situ	2536.139	2681.744	2765.174	2891.513	2906.66	2927.306	53.72792	100	1992.316	2545.38	3103.293
	Average	112.5045	121.5207	126.8988	134.3053	135.5824	136.6711	24.25443	34.45175	100	128.9595	158.5889
	Median	87.87232	94.18253	97.94066	103.1567	104.0542	104.8179	23.01143	32.38153	83.50875	100	116.6489
	Maximum	76.1688	81.44031	84.5805	88.93503	89.68718	90.32589	21.93268	30.63258	74.37118	87.14281	100

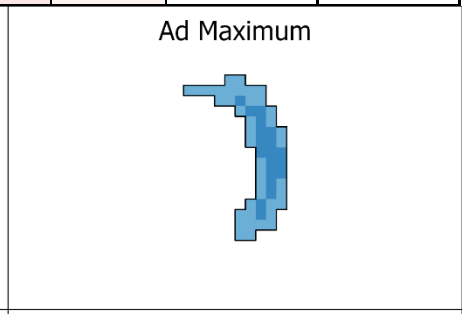
MEDIAN		Local A_d [m]						Global A_d				
		0 - 30	0 - 60	0 - 90	30 - 60	30 - 90	60 - 90	Minimum	In-Situ	Average	Median	Maximum
Local A_d [m]	0 - 30	100	106.1539	109.4901	114.817	115.25	115.443	17.36815	35.07998	97.12096	110.208	125.2469
	0 - 60	94.20074	100	103.1359	107.985	108.1954	108.6003	16.27635	33.77259	92.21771	102.9391	117.0016
	0 - 90	91.32809	96.95945	100	104.4313	104.9653	105.2571	15.62083	32.45402	90.00697	99.96559	112.7475
	30 - 60	87.0565	92.60542	95.7567	100	100.3809	100.7349	14.67619	30.6436	86.88283	96.20511	106.9656
	30 - 90	86.7676	92.4254	95.26954	99.62055	100	100.3183	14.42768	30.65905	86.39608	95.14613	106.1112
	60 - 90	86.62268	92.08078	95.00547	99.27047	99.68272	100	14.37367	29.84038	85.82869	94.58616	105.2732
Global A_d	Minimum	247.9547	264.5566	280.8747	289.3129	297.6924	300.0766	100	150.6264	289.0747	313.1041	336.3699
	In-Situ	194.1016	202.6018	211.551	222.5699	225.0719	224.6921	62.18882	100	208.2304	226.6983	243.1754
	Average	102.9644	108.3806	111.0741	115.044	115.659	116.4336	16.59682	34.18192	100	111.4232	122.9217
	Median	90.73756	97.05546	99.99774	103.8638	105.0786	105.7106	14.78557	30.42598	89.74797	100	110.2258
	Maximum	79.84231	85.45583	88.686	93.48045	94.23243	94.98683	13.29771	27.7294	81.35261	90.72288	100

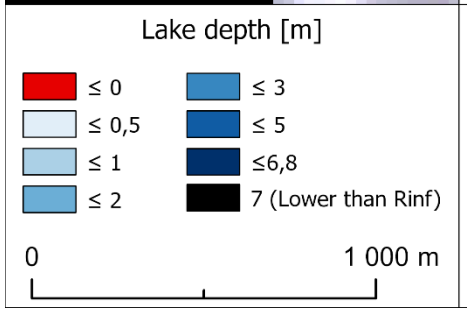
LAKE 98		Local A_d distance [m]						Global A_d					Volume [m³]
		0 - 30	0 - 60	0 - 90	30 - 60	30 - 90	60 - 90	Minimum	In-Situ	Average	Median	Maximum	
Local A_d distance [m]	0 - 30	100	112.79	121.06	126.33	130.50	133.55	72.74	97.30	150.32	157.38	164.34	56 694
	0 - 60	88.66	100	107.33	112.00	115.70	118.40	64.49	86.27	133.27	139.53	145.71	63 945
	0 - 90	82.60	93.17	100	104.35	107.79	110.31	60.08	80.38	124.17	130.00	135.75	68 635
	30 - 60	79.16	89.28	95.83	100	103.30	105.72	57.58	77.03	118.99	124.58	130.09	71 619
	30 - 90	76.63	86.43	92.77	96.80	100	102.34	55.74	74.57	115.19	120.60	125.94	73 983
	60 - 90	74.88	84.46	90.65	94.59	97.71	100	54.47	72.86	112.56	117.85	123.06	75 713
Global A_d	Minimum	137.48	155.06	166.43	173.67	179.40	183.60	100	133.77	206.65	216.36	225.93	41 239
	In-Situ	102.77	115.91	124.42	129.83	134.11	137.25	74.75	100	154.48	161.74	168.90	55 166
	Average	66.52	75.03	80.54	84.04	86.81	88.84	48.39	64.73	100	104.70	109.33	85 223
	Median	63.54	71.67	76.92	80.27	82.92	84.86	46.22	61.83	95.51	100	104.42	89 225
	Maximum	60.85	68.63	73.66	76.87	79.40	81.26	44.26	59.21	91.47	95.76	100	93 172

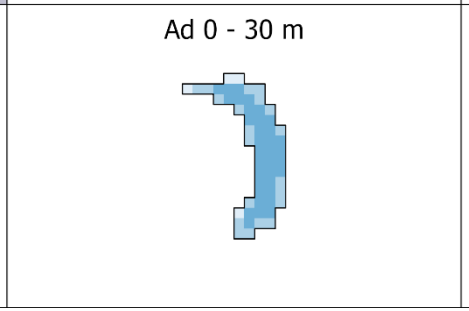


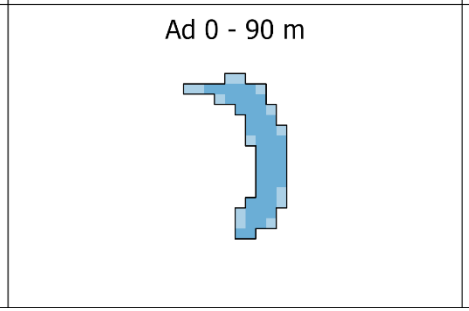


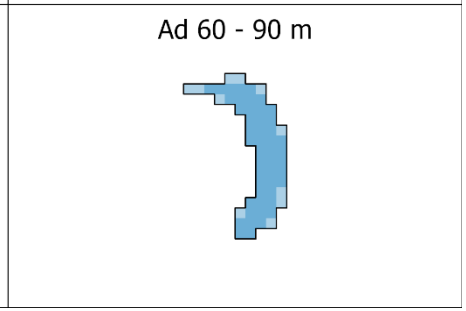




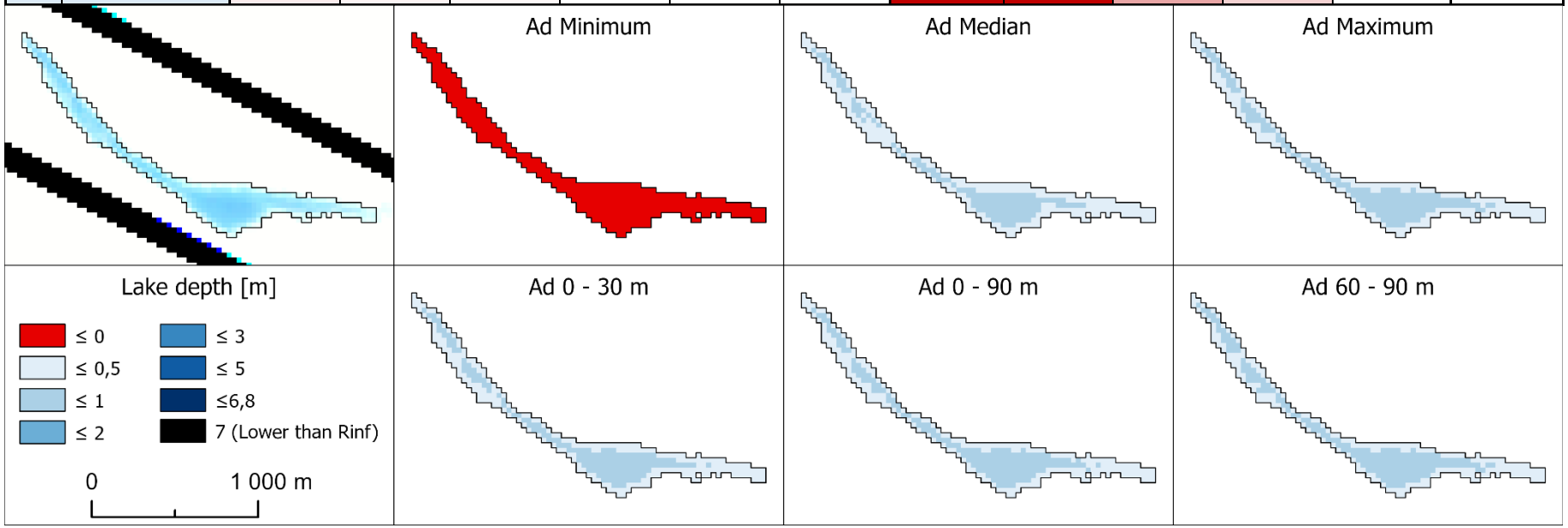




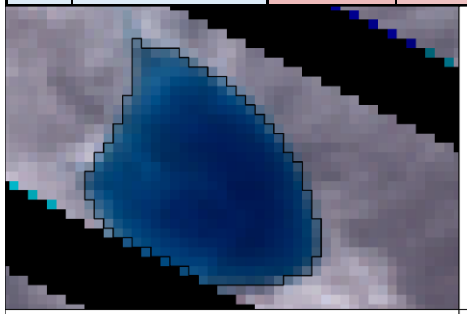


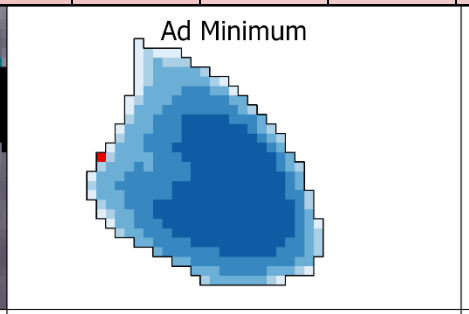


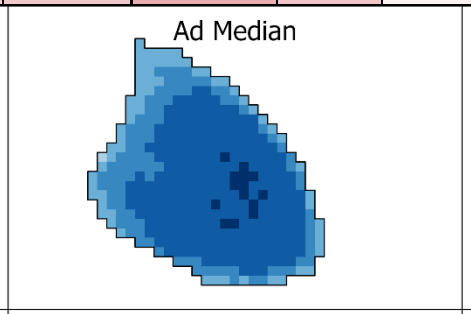
LAKE 268		Local A_d distance [m]						Global A_d					Volume [m³]
		0 - 30	0 - 60	0 - 90	30 - 60	30 - 90	60 - 90	Minimum	In-Situ	Average	Median	Maximum	
Local A_d distance [m]	0 - 30	100	103.07	104.83	107.40	107.58	107.71	0.00	2.24	73.77	90.46	107.13	148 873
	0 - 60	97.02	100	101.70	104.20	104.37	104.49	0.00	2.18	71.57	87.76	103.93	153 448
	0 - 90	95.40	98.33	100	102.46	102.63	102.75	0.00	2.14	70.38	86.29	102.20	156 058
	30 - 60	93.11	95.97	97.60	100	100.17	100.29	0.00	2.09	68.69	84.22	99.75	159 889
	30 - 90	92.95	95.81	97.44	99.83	100	100.12	0.00	2.09	68.57	84.08	99.58	160 160
	60 - 90	92.85	95.70	97.33	99.72	99.88	100	0.00	2.08	68.50	83.98	99.46	160 345
Global A_d	Minimum	X	X	X	X	X	X		X	X	X	X	0
	In-Situ	4456.54	4593.50	4671.61	4786.31	4794.42	4799.97	0.00	100	3287.75	4031.19	4774.23	3 341
	Average	135.55	139.72	142.09	145.58	145.83	146.00	0.00	3.04	100	122.61	145.21	109 829
	Median	110.55	113.95	115.89	118.73	118.93	119.07	0.00	2.48	81.56	100	118.43	134 664
	Maximum	93.35	96.21	97.85	100.25	100.42	100.54	0.00	2.09	68.86	84.44	100	159 486

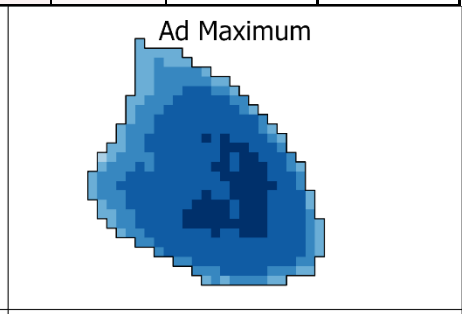


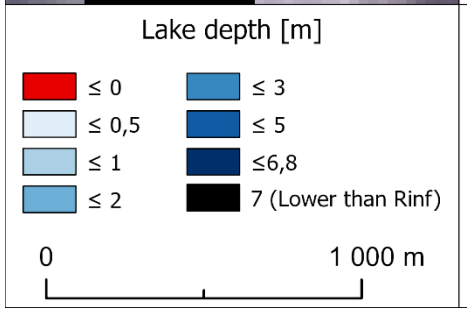
LAKE 343		Local A_d distance [m]						Global A_d					Volume [m³]
		0 - 30	0 - 60	0 - 90	30 - 60	30 - 90	60 - 90	Minimum	In-Situ	Average	Median	Maximum	
Local A_d distance [m]	0 - 30	100	103.03	104.30	107.03	106.66	106.40	94.34	104.30	125.80	128.66	131.49	985 767
	0 - 60	97.06	100	101.24	103.89	103.53	103.27	91.57	101.24	122.11	124.88	127.62	1 015 591
	0 - 90	95.88	98.78	100	102.62	102.26	102.01	90.45	100.00	120.61	123.36	126.07	1 028 149
	30 - 60	93.43	96.26	97.45	100	99.66	99.41	88.15	97.45	117.54	120.21	122.85	1 055 054
	30 - 90	93.76	96.59	97.79	100.35	100	99.75	88.45	97.79	117.94	120.63	123.28	1 051 417
	60 - 90	93.99	96.83	98.03	100.59	100.25	100	88.67	98.03	118.23	120.93	123.58	1 048 839
Global A_d	Minimum	106.00	109.20	110.55	113.45	113.06	112.78	100	110.56	133.34	136.38	139.37	930 000
	In-Situ	95.88	98.78	100.00	102.62	102.26	102.01	90.45	100	120.61	123.36	126.06	1 028 167
	Average	79.49	81.90	82.91	85.08	84.79	84.58	74.99	82.91	100	102.28	104.52	1 240 093
	Median	77.72	80.07	81.06	83.19	82.90	82.70	73.33	81.07	97.77	100	102.19	1 268 315
	Maximum	76.05	78.35	79.32	81.40	81.12	80.92	71.75	79.32	95.68	97.85	100	1 296 147

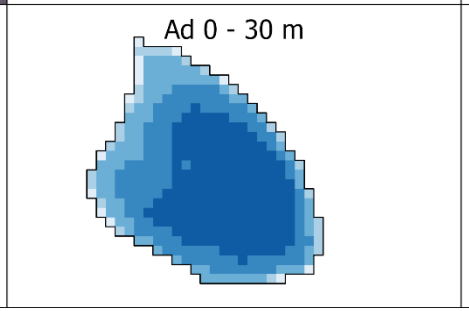


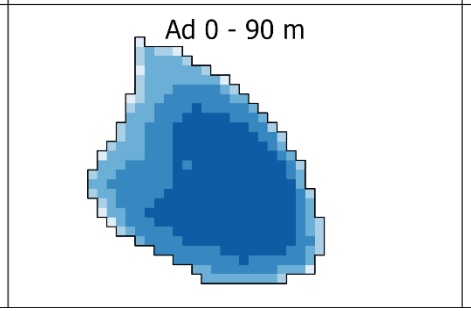


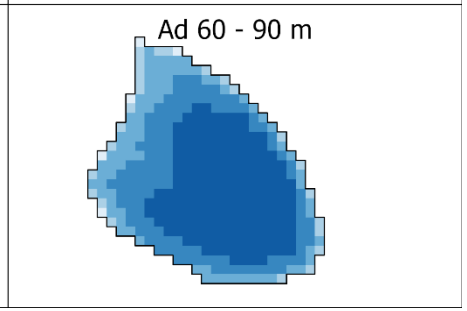




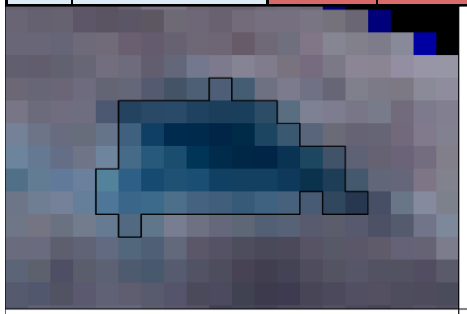


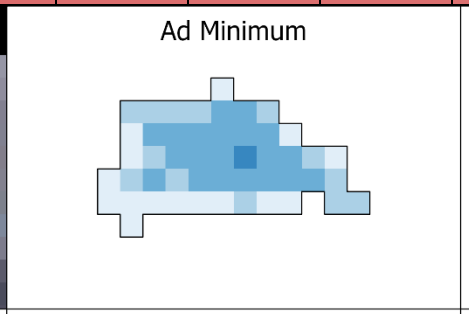


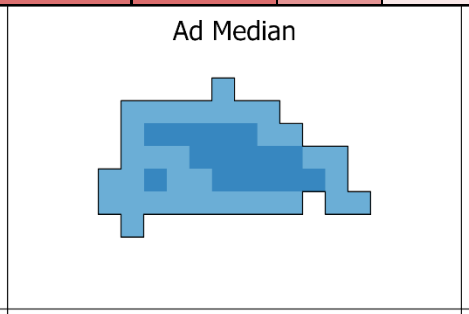


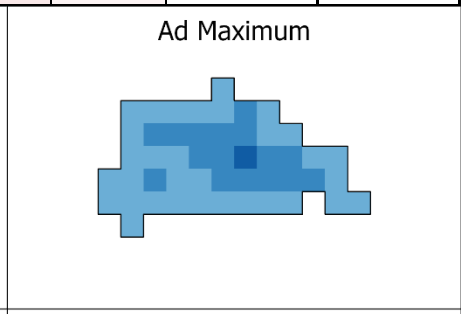


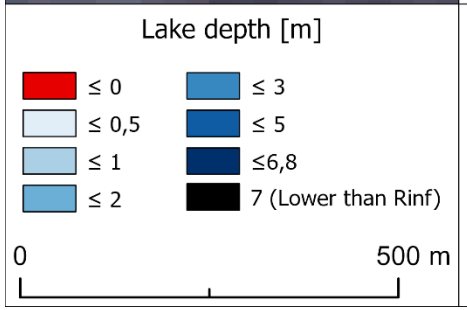
LAKE 458		Local A_d distance [m]						Global A_d					Volume [m³]
		0 - 30	0 - 60	0 - 90	30 - 60	30 - 90	60 - 90	Minimum	In-Situ	Average	Median	Maximum	
Local A_d distance [m]	0 - 30	100	101.99	103.62	104.21	105.18	105.80	103.61	132.99	196.39	204.84	213.16	39 370
	0 - 60	98.05	100	101.60	102.18	103.13	103.74	101.59	130.39	192.56	200.84	209.01	40 152
	0 - 90	96.51	98.43	100	100.57	101.51	102.11	99.99	128.35	189.54	197.69	205.72	40 793
	30 - 60	95.96	97.87	99.43	100	100.94	101.53	99.43	127.62	188.46	196.57	204.56	41 026
	30 - 90	95.07	96.96	98.51	99.07	100	100.59	98.50	126.44	186.72	194.74	202.66	41 410
	60 - 90	94.52	96.40	97.94	98.49	99.42	100	97.93	125.70	185.63	193.61	201.48	41 653
Global A_d	Minimum	96.52	98.44	100.01	100.58	101.52	102.11	100	128.36	189.55	197.70	205.74	40 790
	In-Situ	75.20	76.69	77.91	78.36	79.09	79.56	77.91	100	147.68	154.03	160.29	52 356
	Average	50.92	51.93	52.76	53.06	53.56	53.87	52.76	67.71	100	104.30	108.54	77 319
	Median	48.82	49.79	50.58	50.87	51.35	51.65	50.58	64.92	95.88	100	104.07	80 643
	Maximum	46.91	47.85	48.61	48.89	49.34	49.63	48.61	62.39	92.13	96.09	100	83 921

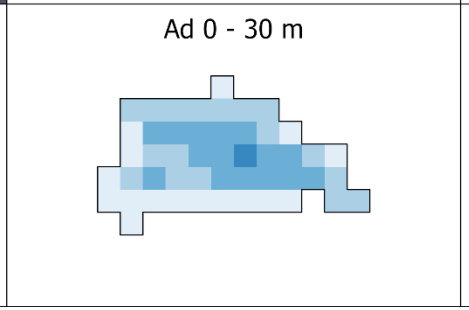


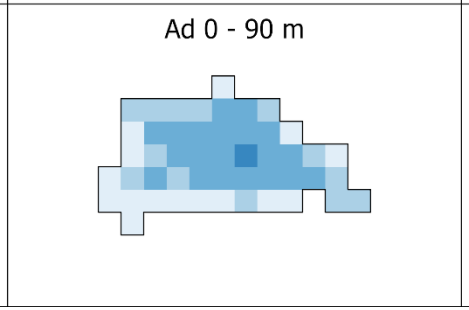


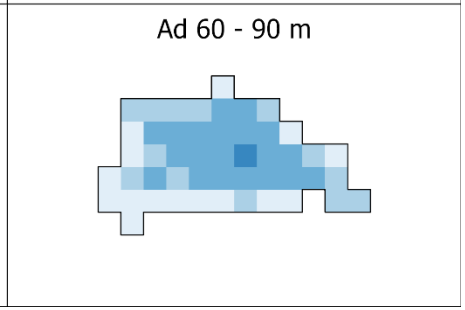




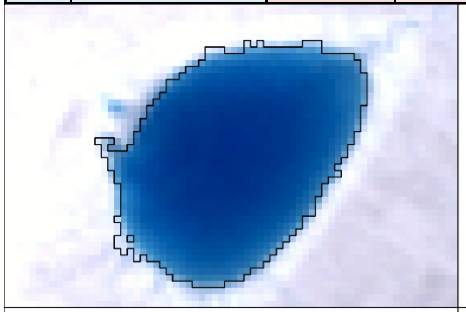


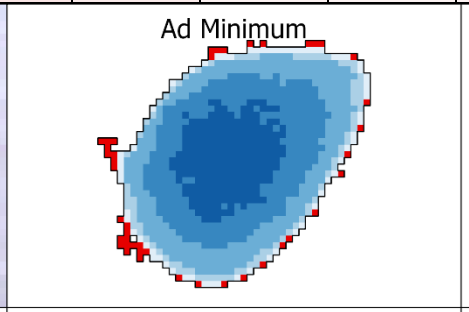


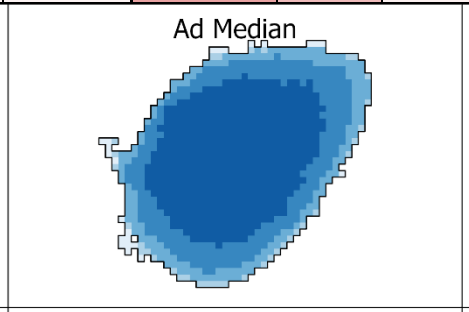


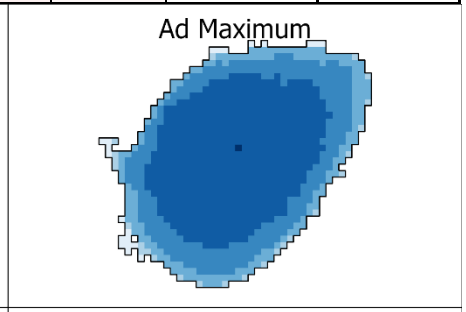


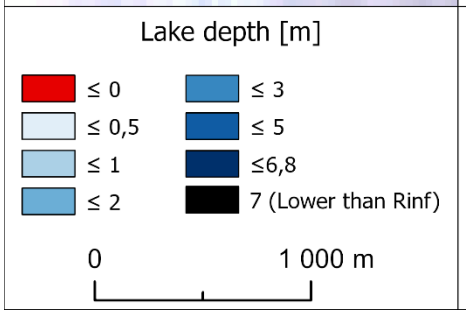
LAKE 538		Local A_d distance [m]						Global A_d					Volume [m³]
		0 - 30	0 - 60	0 - 90	30 - 60	30 - 90	60 - 90	Minimum	In-Situ	Average	Median	Maximum	
Local A_d distance [m]	0 - 30	100	104.02	106.38	109.24	109.81	110.22	77.51	87.38	109.04	111.97	114.85	2 419 619
	0 - 60	96.14	100	102.27	105.02	105.57	105.96	74.51	84.01	104.83	107.64	110.42	2 516 825
	0 - 90	94.00	97.78	100	102.69	103.23	103.61	72.86	82.14	102.50	105.25	107.97	2 573 947
	30 - 60	91.54	95.22	97.38	100	100.53	100.90	70.95	80.00	99.82	102.50	105.14	2 643 108
	30 - 90	91.06	94.72	96.87	99.48	100	100.37	70.58	79.58	99.30	101.96	104.59	2 657 025
	60 - 90	90.73	94.37	96.52	99.11	99.63	100	70.32	79.28	98.93	101.59	104.21	2 666 844
Global A_d	Minimum	129.02	134.20	137.25	140.94	141.68	142.20	100	112.74	140.68	144.46	148.18	1 875 400
	In-Situ	114.44	119.03	121.74	125.01	125.67	126.13	88.70	100	124.78	128.13	131.44	2 114 369
	Average	91.71	95.39	97.56	100.18	100.71	101.08	71.08	80.14	100	102.68	105.33	2 638 370
	Median	89.31	92.90	95.01	97.56	98.07	98.44	69.22	78.04	97.39	100	102.58	2 709 198
	Maximum	87.07	90.56	92.62	95.11	95.61	95.96	67.48	76.08	94.94	97.49	100	2 779 045

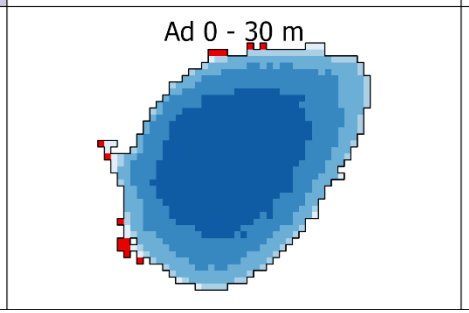


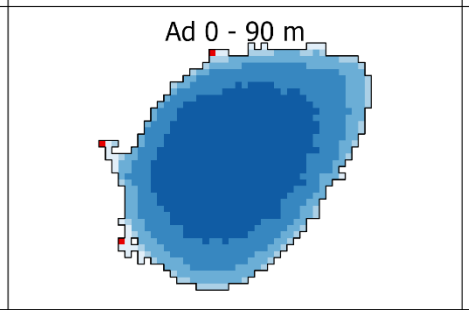


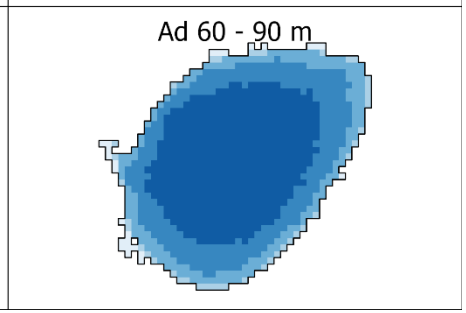




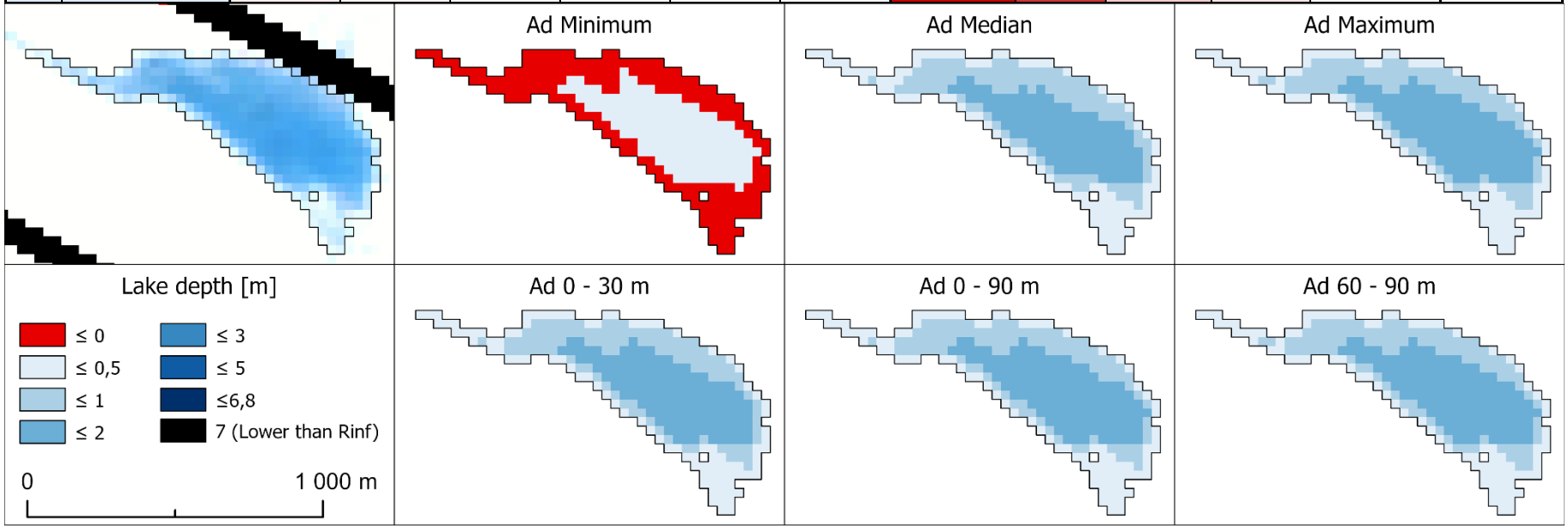




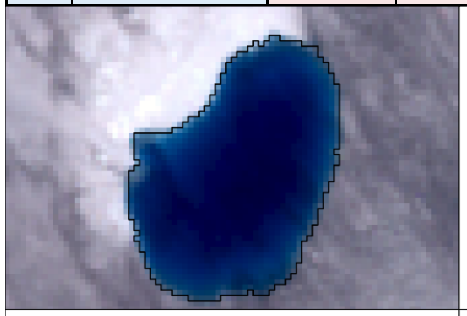


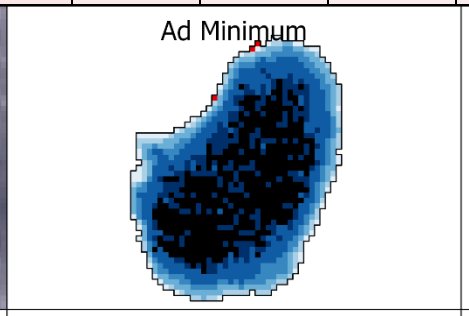


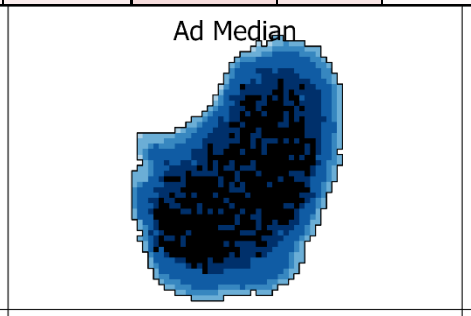
LAKE 558		Local A_d distance [m]						Global A_d					Volume [m³]
		0 - 30	0 - 60	0 - 90	30 - 60	30 - 90	60 - 90	Minimum	In-Situ	Average	Median	Maximum	
Local A_d distance [m]	0 - 30	100	102.38	103.77	105.84	106.12	106.33	13.08	30.99	87.65	97.23	106.74	231 491
	0 - 60	97.68	100	101.37	103.39	103.66	103.87	12.78	30.27	85.62	94.97	104.26	236 991
	0 - 90	96.36	98.65	100	101.99	102.26	102.47	12.60	29.87	84.47	93.69	102.86	240 228
	30 - 60	94.48	96.73	98.05	100	100.27	100.46	12.36	29.28	82.82	91.86	100.85	245 015
	30 - 90	94.23	96.47	97.79	99.73	100	100.20	12.33	29.20	82.60	91.62	100.58	245 666
	60 - 90	94.04	96.28	97.59	99.54	99.80	100	12.30	29.15	82.43	91.44	100.38	246 151
Global A_d	Minimum	764.51	782.67	793.36	809.17	811.32	812.93	100	236.95	670.12	743.34	816.03	30 280
	In-Situ	322.65	330.32	334.83	341.50	342.41	343.09	42.20	100	282.82	313.72	344.40	71 746
	Average	114.09	116.80	118.39	120.75	121.07	121.31	14.92	35.36	100	110.93	121.77	202 910
	Median	102.85	105.29	106.73	108.86	109.15	109.36	13.45	31.88	90.15	100	109.78	225 081
	Maximum	93.69	95.91	97.22	99.16	99.42	99.62	12.25	29.04	82.12	91.09	100	247 092

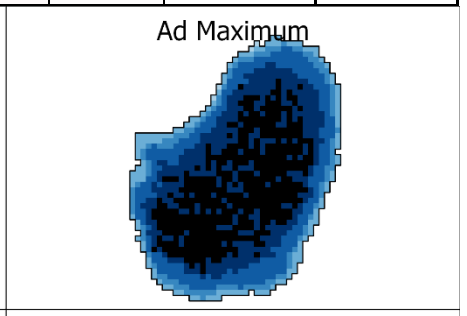


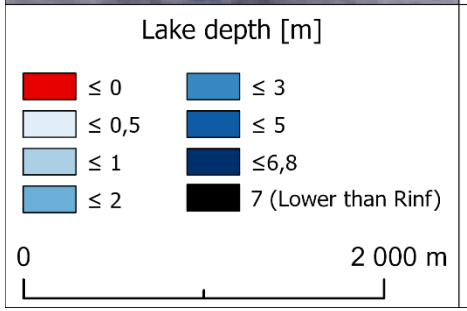
LAKE 604		Local A_d distance [m]						Global A_d					Volume [m³]
		0 - 30	0 - 60	0 - 90	30 - 60	30 - 90	60 - 90	Minimum	In-Situ	Average	Median	Maximum	
Local A_d distance [m]	0 - 30	100	101.34	101.98	102.92	102.97	103.00	96.86	100.47	108.29	109.33	110.36	5 823 662
	0 - 60	98.68	100	100.64	101.56	101.61	101.64	95.58	99.14	106.86	107.88	108.90	5 901 613
	0 - 90	98.05	99.37	100	100.91	100.96	101.00	94.97	98.51	106.18	107.20	108.21	5 939 211
	30 - 60	97.17	98.47	99.09	100	100.05	100.08	94.11	97.62	105.22	106.23	107.23	5 993 512
	30 - 90	97.12	98.42	99.05	99.95	100	100.03	94.07	97.58	105.17	106.18	107.18	5 996 338
	60 - 90	97.09	98.39	99.01	99.92	99.97	100	94.04	97.54	105.13	106.14	107.14	5 998 379
Global A_d	Minimum	103.25	104.63	105.29	106.26	106.31	106.34	100	103.73	111.80	112.88	113.94	5 640 583
	In-Situ	99.53	100.87	101.51	102.44	102.48	102.52	96.40	100	107.78	108.82	109.84	5 851 001
	Average	92.35	93.58	94.18	95.04	95.09	95.12	89.44	92.78	100	100.96	101.91	6 306 271
	Median	91.47	92.69	93.28	94.14	94.18	94.21	88.59	91.90	99.05	100	100.94	6 366 922
	Maximum	90.62	91.83	92.41	93.26	93.30	93.33	87.77	91.04	98.13	99.07	100	6 426 733

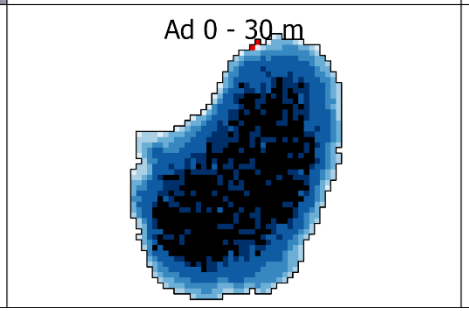


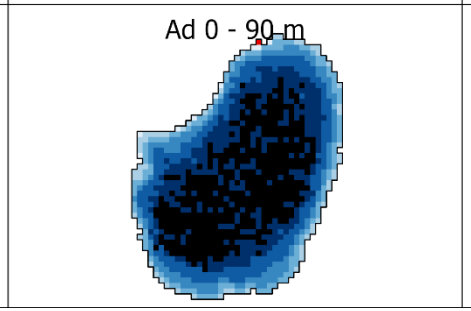


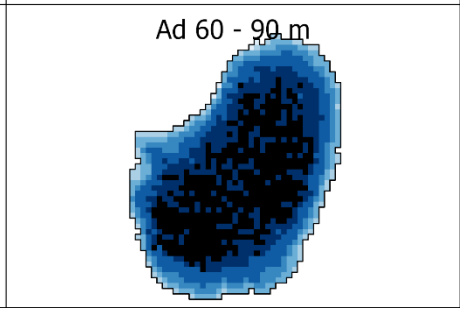




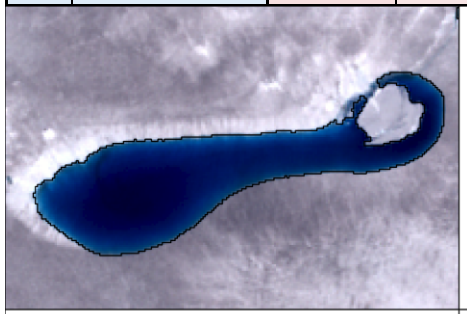


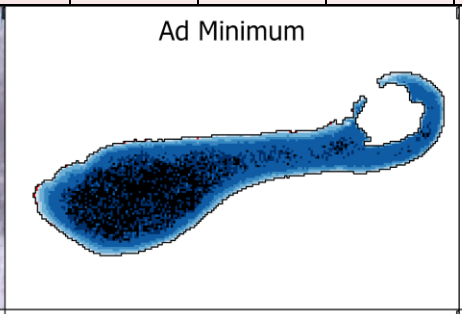


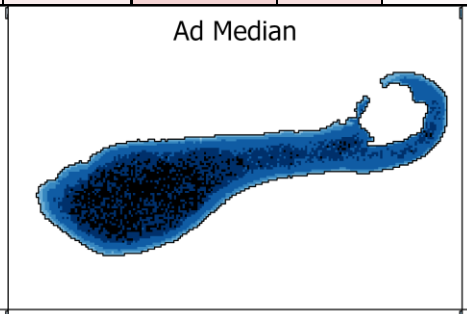


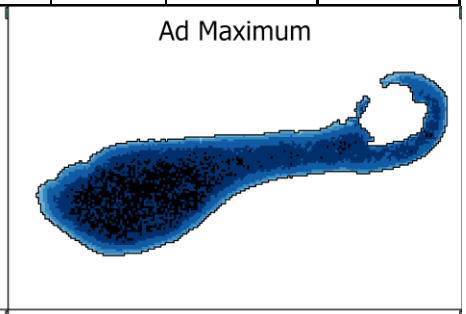


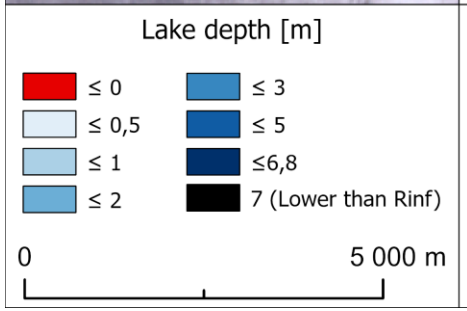
LAKE 610		Local A_d distance [m]						Global A_d					Volume [m³]
		0 - 30	0 - 60	0 - 90	30 - 60	30 - 90	60 - 90	Minimum	In-Situ	Average	Median	Maximum	
Local A_d distance [m]	0 - 30	100	102.05	103.09	104.49	104.66	104.79	95.15	99.68	109.49	110.79	112.08	24 037 797
	0 - 60	97.99	100	101.01	102.39	102.56	102.68	93.23	97.68	107.28	108.56	109.83	24 531 273
	0 - 90	97.01	99.00	100	101.36	101.53	101.65	92.30	96.70	106.21	107.48	108.73	24 779 578
	30 - 60	95.70	97.67	98.66	100	100.16	100.28	91.06	95.40	104.78	106.03	107.26	25 117 255
	30 - 90	95.55	97.51	98.50	99.84	100	100.12	90.91	95.24	104.61	105.86	107.09	25 158 093
	60 - 90	95.43	97.39	98.38	99.72	99.88	100	90.80	95.13	104.48	105.73	106.96	25 188 823
Global A_d	Minimum	105.10	107.26	108.35	109.82	110.00	110.14	100	104.77	115.07	116.45	117.80	22 870 779
	In-Situ	100.32	102.38	103.41	104.82	104.99	105.12	95.45	100	109.84	111.15	112.44	23 961 345
	Average	91.34	93.21	94.15	95.44	95.59	95.71	86.90	91.04	100	101.19	102.37	26 318 154
	Median	90.26	92.11	93.04	94.31	94.47	94.58	85.88	89.97	98.82	100	101.16	26 632 059
	Maximum	89.22	91.05	91.98	93.23	93.38	93.49	84.89	88.94	97.69	98.85	100	26 941 620

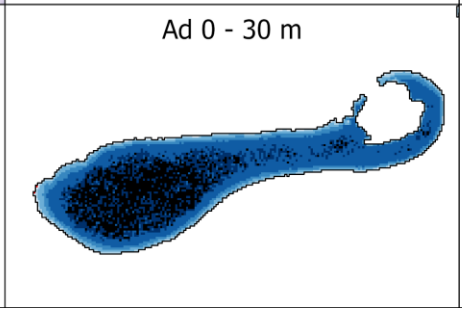


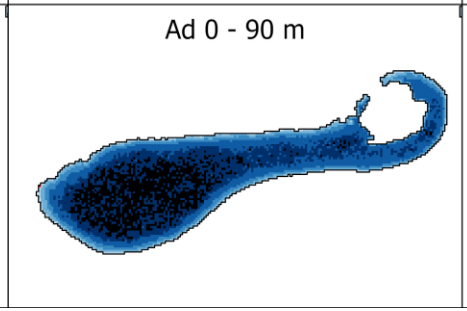


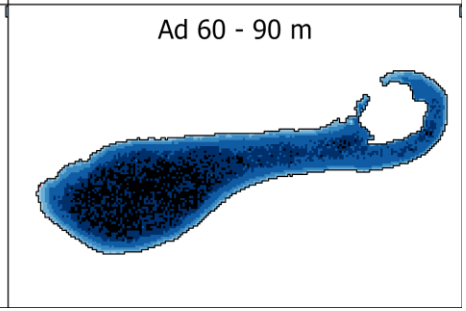












Attachment 7: Statistical indices to validate the resulting physical model estimation as a function of the parameter A_d for 4675 in-situ measurements aggregated over 39 Landsat 7 pixels area.

R Squared	Global Min	Global Lake	Global Average	Global Median	Global Max	Local 0-30m	Local 0-60m	Local 0-90m	Local 30-60m	Local 30-90m	Local 60-90m
In-situ Average	0.856533	0.910741	0.650003	0.576328	0.494706	0.900931	0.909987	0.910741	0.906255	0.907348	0.90803
In-situ Most Frequent	0.864487	0.910172	0.667205	0.599625	0.524873	0.90274	0.909917	0.910173	0.905203	0.906315	0.907019
In-situ Mean 20100702	0.797189	0.845418	0.520849	0.433422	0.337028	0.840023	0.846264	0.845421	0.836689	0.838422	0.839545
In-situ Mean 20100703	0.843544	0.889275	0.617178	0.542705	0.460461	0.882875	0.889493	0.889277	0.882823	0.884173	0.885039
In-situ Mean 20100705	0.820316	0.884965	0.542719	0.447459	0.342081	0.874391	0.884578	0.884967	0.878015	0.879575	0.880563
AVERAGE	0.836414	0.888114	0.599591	0.519908	0.43183	0.880192	0.888048	0.888116	0.881797	0.883166	0.884039

RMSE [m]	Global Min	Global Lake	Global Average	Global Median	Global Max	Local 0-30m	Local 0-60m	Local 0-90m	Local 30-60m	Local 30-90m	Local 60-90m
In-situ Average	0.421994	0.332856	0.659116	0.725178	0.791958	0.350671	0.33426	0.332856	0.341118	0.339124	0.337872
In-situ Most Frequent	0.431295	0.351148	0.675885	0.74134	0.807586	0.365385	0.351646	0.351146	0.360729	0.358608	0.357257
In-situ Mean 20100702	0.470989	0.411192	0.723938	0.787218	0.851555	0.418306	0.410065	0.411188	0.422642	0.420394	0.41893
In-situ Mean 20100703	0.444785	0.374176	0.695749	0.760417	0.825972	0.384838	0.373808	0.374174	0.384924	0.3827	0.381266
In-situ Mean 20100705	0.418171	0.33459	0.667099	0.733298	0.800174	0.349629	0.335152	0.334588	0.34455	0.34234	0.340932
AVERAGE	0.437447	0.360792	0.684357	0.74949	0.815449	0.373766	0.360986	0.36079	0.370793	0.368633	0.367252

Mean Absolute Error [m]	Global Min	Global Lake	Global Average	Global Median	Global Max	Local 0-30m	Local 0-60m	Local 0-90m	Local 30-60m	Local 30-90m	Local 60-90m
In-situ Average	0.347232	0.263804	0.58112	0.652634	0.723159	0.271324	0.262943	0.263803	0.270473	0.269228	0.268345
In-situ Most Frequent	0.376541	0.297512	0.579532	0.653004	0.727341	0.319649	0.303541	0.297521	0.290311	0.290561	0.290737
In-situ Mean 20100702	0.408919	0.324117	0.614819	0.676494	0.746325	0.345811	0.327171	0.324122	0.328846	0.327963	0.327337
In-situ Mean 20100703	0.377369	0.291326	0.600821	0.669919	0.738062	0.303615	0.291326	0.291326	0.29724	0.295621	0.294474
In-situ Mean 20100705	0.333162	0.266941	0.59958	0.669935	0.739316	0.269539	0.262526	0.266931	0.281304	0.279361	0.277984
AVERAGE	0.368645	0.28874	0.595174	0.664397	0.734841	0.301987	0.289501	0.288741	0.293635	0.292547	0.291775

Attachment 8: Statistical indices to validate the resulting physical model estimation as a function of the parameter A_d for 4675 in-situ measurements.

R Squared	Global Min	Global Lake	Global Average	Global Median	Global Max	Local 0-30m	Local 0-60m	Local 0-90m	Local 30-60m	Local 30-90m	Local 60-90m
In-situ ALL	0.77363	0.855178	0.547792	0.457011	0.356013	0.837366	0.852652	0.855175	0.852771	0.853718	0.854272
In-situ 20100702	0.778406	0.844036	0.5634	0.482485	0.392662	0.830897	0.842548	0.844034	0.840398	0.841433	0.842064
In-situ 20100703	0.780631	0.846612	0.493579	0.395476	0.286972	0.835955	0.846277	0.846612	0.839326	0.840949	0.841979
In-situ 20100705	0.719452	0.835834	0.491812	0.38462	0.264787	0.807006	0.830678	0.835828	0.837243	0.837817	0.838081
AVERAGE	0.76303	0.845415	0.524146	0.429898	0.325109	0.827806	0.843039	0.845412	0.842435	0.843479	0.844099

RMSE [m]	Global Min	Global Lake	Global Average	Global Median	Global Max	Local 0-30m	Local 0-60m	Local 0-90m	Local 30-60m	Local 30-90m	Local 60-90m
In-situ ALL	0.468583	0.374796	0.662288	0.725727	0.790344	0.397176	0.378051	0.3748	0.377898	0.376681	0.375966
In-situ 20100702	0.496024	0.416136	0.69625	0.758027	0.82118	0.43331	0.418117	0.416139	0.420962	0.419595	0.418759
In-situ 20100703	0.455635	0.381	0.692286	0.756373	0.821453	0.394013	0.381415	0.381	0.389943	0.387969	0.386711
In-situ 20100705	0.470106	0.359612	0.63271	0.696248	0.761025	0.38991	0.365216	0.359618	0.358065	0.357433	0.357142
AVERAGE	0.472587	0.382886	0.670883	0.734094	0.7985	0.403602	0.3857	0.382889	0.386717	0.38542	0.384645

Mean Absolute Error [m]	Global Min	Global Lake	Global Average	Global Median	Global Max	Local 0-30m	Local 0-60m	Local 0-90m	Local 30-60m	Local 30-90m	Local 60-90m
In-situ ALL	0.395807	0.301208	0.558911	0.630024	0.701635	0.328206	0.306973	0.301217	0.296922	0.296814	0.296774
In-situ 20100702	0.419553	0.32381	0.571366	0.638916	0.709839	0.350594	0.328687	0.323817	0.321102	0.320942	0.320827
In-situ 20100703	0.389463	0.305255	0.587133	0.658404	0.730254	0.330486	0.311039	0.305265	0.302651	0.302256	0.301976
In-situ 20100705	0.394176	0.292966	0.537126	0.60901	0.680631	0.321188	0.298935	0.292974	0.287174	0.287271	0.287407
AVERAGE	0.39975	0.30581	0.563634	0.634089	0.70559	0.332619	0.311409	0.305818	0.301962	0.301821	0.301746