

Article

Detailed Lacustrine Calving Iceberg Inventory from Very High Resolution Optical Imagery and Object-Based Image Analysis

Julian Podgórski ¹  and Michał Pełlicki ^{2,*} ¹ Institute of Geophysics, Polish Academy of Sciences, ul. Księcia Janusza 64, 01-452 Warsaw, Poland; jpodgo@igf.edu.pl² Centro de Estudios Científicos (CECs), Av. Prat 514, Valdivia 5110466, Chile

* Correspondence: michal@cecs.cl

Received: 30 April 2020; Accepted: 26 May 2020; Published: 3 June 2020



Abstract: In the field of iceberg and glacier calving studies, it is important to collect comprehensive datasets of populations of icebergs. Particularly, calving of lake-terminating glaciers has been understudied. The aim of this work is to present an object-based method of iceberg detection and to create an inventory of icebergs located in a proglacial lagoon of San Quintín glacier, Northern Patagonia Icefield, Chile. This dataset is created using high-resolution WorldView-2 imagery and a derived DEM. We use it to briefly discuss the iceberg size distribution and area–volume scaling. Segmentation of the multispectral imagery produced a map of objects, which were classified with use of Random Forest supervised classification algorithm. An intermediate classification product was corrected with a ruleset exploiting contextual information and a watershed algorithm that was used to divide multiple touching icebergs into separate objects. Common theoretical heavy-tail statistical distributions were tested as descriptors of iceberg area and volume distributions. Power law models were proposed for the area–volume relationship. The proposed method performed well over the open lake detecting correctly icebergs in all size bands except 5–15 m² where many icebergs were missed. A section of the lagoon with ice melange was not reliably mapped due to uniformity of the area in the imagery and DEM. The precision of the DEM limited the scaling effort to icebergs taller than 1.7 m and larger than 99 m², despite the inventory containing icebergs as small as 4 m². The work demonstrates viability of object-based analysis for lacustrine iceberg detection and shows that the statistical properties of iceberg population at San Quintín glacier match those of populations produced by tidewater glaciers.

Keywords: object-based image analysis; WorldView-2; glacier calving; lacustrine calving; Northern Patagonia Icefield; iceberg; power law; random forest

1. Introduction

Lacustrine calving of glaciers is mechanical detachment of icebergs when a glacier terminates in fresh water [1,2]. The phenomenon of iceberg calving in lakes differs slightly from the more widely known tidewater calving. In the case of sea- and ocean-terminating glaciers, an interplay between tidal action, thermal undercutting by subglacial plumes and buoyancy forces leads to calving processes [2]. However, in the case of glaciers terminating in lakes, buoyancy forces [1] are the primary calving factor, promoting formation of tabular icebergs, as observed in proglacial lakes of, e.g., Patagonia [1] and Alaska [3]. While freshwater calving does not contribute to the global sea level rise as strongly as the tidewater calving, it influences dynamics of proglacial lakes, which have potential to store large amounts of meltwater [4] and provide water for agriculture and even hydropower downstream [5].

Nowadays, radar imagery is widely used for detection and monitoring of large icebergs in marine contexts [6,7]. Optical imagery, both satellite and aerial, is employed when detection of small icebergs is of concern [8,9]. On these datasets, both object- and pixel-based analyses are performed to arrive at iceberg positions and inventories. The major differences, relevant for remote sensing, between lacustrine and tidewater calving are the size of icebergs and color of the surrounding water. Lacustrine calving very often produces small icebergs which are hardly detectable on medium resolution satellite data. This is the case also with small tidewater glaciers. In such cases, methods which measure the calving events instead of the floating iceberg population are used to quantify ice fluxes, such as LiDAR scanning, seismic and acoustic measurements [10] or time-lapse terrestrial photography [11,12]. These in-situ techniques do not differ significantly in application near tidewater or lake-terminating glaciers. Lacustrine calving happens in freshwater ice bodies, often with high sediment load, which changes the water color. Therefore, icebergs covered with sufficient amount of debris can be mistaken for water on multispectral images. Scarcity of lacustrine iceberg and calving datasets was pointed out as a factor hindering research into the calving's impact on lake-terminating glacier dynamics [12].

The size distribution of icebergs is a lingering topic in the glaciological literature [13–15]. Data on size distribution of calved ice bodies contribute to understanding of calving mechanisms [8] and iceberg decay processes [15,16], and they are used in modeling efforts to parameterize the freshwater influx from glaciers and icesheets [17]. Earlier studies prove that heavy-tailed distributions are best approximates of iceberg size frequencies [15,18]. Most of them, however, were focused on iceberg populations observed in front of calving tidewater glacier fronts [18], or coming from decay of icesheet margins [14]. Studies of size distributions of lacustrine icebergs are lacking; lacustrine calving literature focuses on detailed calving event descriptions [12] or studies of calving mechanisms [3]. The only size frequency-related work in Patagonia estimates iceberg volume from time-lapse imagery and surface wave measurements [11].

Object-Based Image Analysis (OBIA) is a paradigm of analysis of raster data (image analysis) that relies on segments identified within the image [19]. Instead of operating on properties of individual pixels, OBIA employs properties of groups of pixels to perform classifications. The advantage of OBIA compared to the pixel based object analysis is that the list of properties available to the classification algorithm is much wider. In the case of the pixel-based operations only the values of pixels in the available bands can be used, while OBIA can include geometrical features of segments, textural measures and statistical properties of pixel populations enclosed by segments. In the cryospheric research, OBIA finds uses in studies dealing with mapping features that would have been otherwise problematic to discern due to spectral overlap between surface types. Examples include mapping of debris-covered glaciers [20,21], glacier surface facies [22] or glacial lakes [23]. The advantages of OBIA can easily be exploited in iceberg detection as well. On multispectral imagery of floating icebergs, there are sharp boundaries between bright ice and dark water. The dark, debris-covered specimens, or ones with lines of sediment embedded in their faces stand out texturally against smooth water surface. OBIA has already been used successfully to study ocean-borne icebergs on radar [6] and high-resolution aerial optical [24] imagery, showing that the method is viable for iceberg research.

In this work we intend to expand the usefulness of OBIA method to the field of lacustrine iceberg detection. The objective of this paper is to present a method of mapping of icebergs with use of OBIA techniques on high-resolution optical satellite imagery. The novelty of our work is in proposition of a method to detect icebergs in the environment where the spectral contrast between ice and water is not always clearly visible. We develop an inventory of icebergs in a lagoon surrounding the San Quintín glacier and prove its usefulness by deriving size distributions and area–volume scaling laws. We test whether the sizes and relationship between cross-sectional surface area and total volume follow power law in lacustrine setting.

2. Study Area

San Quintín glacier is the largest outlet glacier of Northern Patagonia Icefield (NPI) [25], located in southern Chile (Figure 1). It has an area of 793 km² and has an extensive tongue that forms much of the ablation area of the icefield [26].

The glacier went through a period of oscillation in the 20th century with alternating advance and retreat episodes culminated with a very fast retreat, starting 1995 [27]. A rapid disintegration of the snout occurred between 1999 and 2013, which resulted in San Quintín's acceleration, greater than in the case of other NPI outlets [28]. The retreat, however, paused between 2013 and 2015, with a part of the snout rapidly advancing [28]. Radar measurements of ice movements showed higher velocity in the middle of the glacier's tongue compared to its margins and a diverse velocity field at the entrance to the lagoon [29] in line with reports of a variegated advance speed of the snout.

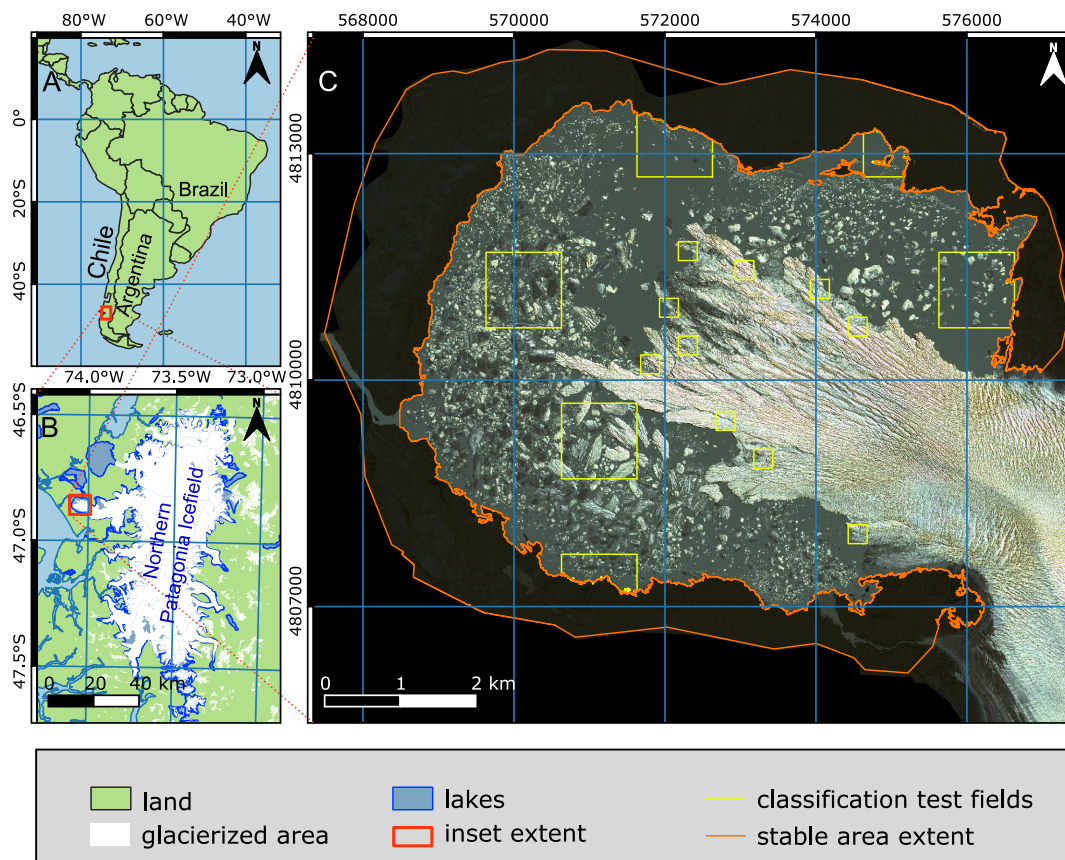


Figure 1. Map of the study area: (A) location of the Northern Patagonia Icefield within the South America; (B) location of the San Quintín terminus and lagoon in the region; and (C) WV-2 image of the study area. Coordinates of (C) in UTM 18S coordinate system.

The surface of the lakes at the fronts of San Quintín glacier has increased from 5.42 to 42.37 km² between 1979 and 2013 together with the glacier's retreat of 4.5 km [30]. The lagoon into which the glacier terminates is surrounded by push moraines created during the glacier's advance in the second half of the 20th century [27]. The lake is drained to the Pacific Ocean by a narrow, meandering stream which prevents iceberg evacuation. Consequently, icebergs calved off San Quintín's snout have no means of escaping the lagoon and remain in the lake for their entire life cycle. Melting of sediment-laden icebergs and influx of meltwater drive large amounts of sediment to the lagoon changing the water's color. The terminal lagoon of San Quintín increased in area between 1947 and 2011 and is the largest of the proglacial lagoons surrounding NPI [4]. In such environment, OBIA promises better classification results than a pixel-based study thanks to its ability to classify objects by analyzing features other than just spectral response in four bands. The dirty icebergs blend with the dirty water,

but they differ from water with their elevation, image texture or contrast between pixels, properties which can only be analyzed within groups of pixels.

3. Data and Methods

3.1. Datasets Used and Imagery Preprocessing

We used a WorldView-2 (WV-2) satellite image composed of a four-band multispectral file and a separate file with the panchromatic band (Table 1). The product was acquired as a set of OrthoReady (OR2A level) images subject exclusively to radiometric and sensor correction [31]. In addition to the multispectral and panchromatic imagery, we included a Digital Elevation Model (DEM) in the study. Our DEM was created with the WorldView-2 stereoimage pair, of which our WV-2 scene was a part. The DEM was created in ERDAS IMAGINE photogrammetry suite with a workflow that included correlation-based matching of two stereoimages [32]. For quality assessment of the DEM, we used a freely available TanDEM-X DEM at 90 m/pixel spatial resolution [33,34].

Table 1. List of the datasets used for iceberg inventory creation and DEM quality assessment.

Band	Wavelength	Original Resolution	Date	Source
Blue	450–510 nm	2 m/pixel	17 February 2017	[35]
Green	510–580 nm			
Red	630–690 nm			
NIR	770–895 nm	0.5 m/pixel	17 February 2017	
Panchromatic	450–800 nm			
WV-2 DEM	Elevation	2 m/pixel	17 February 2017	
TanDEM-X DEM	Elevation	90 m/pixel	22 January 2011–8 September 2014	[33,34]

For this study, we chose a less oblique photo of the pair (ID 17FEB04145515-M2AS-059470802010_01_P001). We subjected the imagery to a pre-processing routine before any analysis (Figure 2).

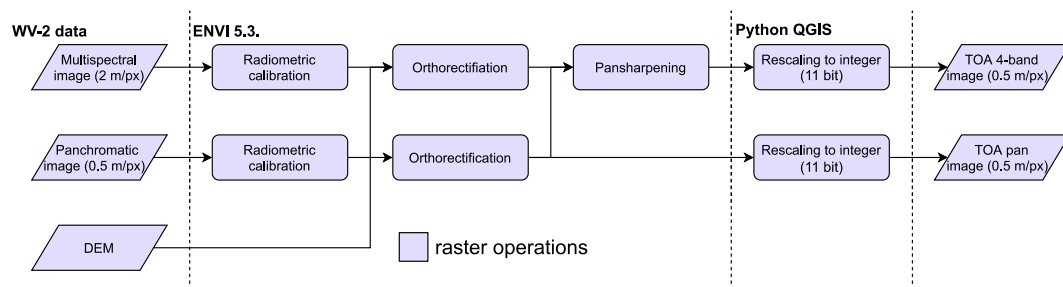


Figure 2. Procedure of pre-processing applied to the WV-2 images.

The raw images obtained from vendor were radiometrically calibrated to the Top of Atmosphere (TOA) reflectance, orthorectified and the multispectral image was pansharpened to the resolution of the panchromatic band (0.5 m/pixel). We performed these operations in ENVI 5.3. In the orthorectification step, we used the WV-2 DEM as an elevation model. The final product was a pair of orthorectified TOA reflectance images—single-band panchromatic and four-band multispectral—in 0.5 m/pixel spatial resolution. The bitrange of the WV-2 is 11 bits [31] and it was preserved along the whole preprocessing routine. Finally, we computed Normalized Difference Snow Index (NDSI) [36] with use of green and NIR bands.

3.2. DEM Preprocessing

A drawback of all feature-matching DEM generation methods is their unreliability in smooth areas devoid of distinct features [37,38]. In our case, this resulted in a “hilly” topography in areas where flat water table would be expected. We rectified this error semi-manually (Figure 3).

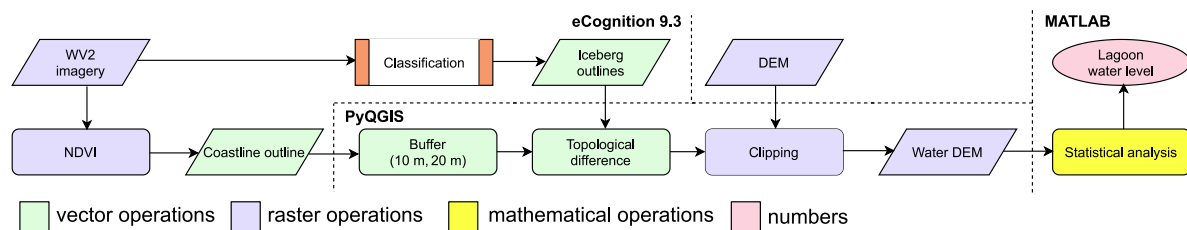


Figure 3. Procedure of determination of elevation of lagoon’s water table.

Water table elevation was measured as the mean elevation of a 10 m wide strip of pixels distanced 10 m from the coastline. This way we made sure to exclude land area from the calculations. The coastline was outlined with an NDVI mask calculated and digitized in eCognition. All pixels falling beneath this mask were assigned a unique value of 16.85 m a.s.l. To enable calculation of overwater elevation of icebergs, we created a relative elevation (Z_{rel}) map (Digital Relative Elevation Model (DREM)). The surface of reference was the level of the water table of the lagoon. Its elevation above the sea level was subtracted from all elevations in the DEM. At the end, we produced a modified DREM. We found artifacts in the original DREM that showed negative relative elevation, probably stemming from incorrect detection of parallax by the stereoinaging algorithm or from movement of icebergs in the short time between acquisition of two stereoimages. We created a DREM₀ where all $Z_{rel} < 0$ values are replaced by 0.

3.3. DEM Uncertainty Assessment

To assess precision of the DREM, we used freely available TanDEM-X DEM with 90 m/pixel spatial resolution [34] as a reference model (Figure 4). We resampled the WV-2 DEM to the 90 m/pixel of the TanDEM-X DEM and isolated an area of land surrounding the lagoon. This part of the image was assumed to not undergo elevation changes between the date of WV-2 image and TanDEM-X acquisition period and is sufficiently flat to ensure good quality of the radar DEM. We created a map of differences between the two products and filtered it to 10th–90th percentile bounds.

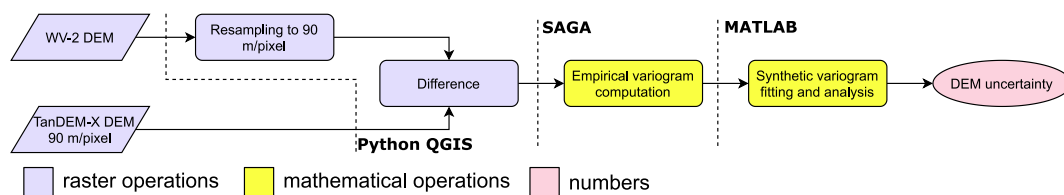


Figure 4. Procedure of the WV-2 DEM accuracy determination.

We assumed that elevation differences are spatially correlated over a limited distance. Such spatial correlation impacts measurement of averaged difference and its variance [39]. We used a semivariogram-based method of Rolstad et al. [39] to estimate the range of correlation and the variance of the difference between DEM and TanDEM-X DEM (Equation (2)). We computed the empirical semivariogram of the TanDEM-X-WV-2DEM difference map in SAGA, then fitted a synthetic variogram and found its properties in MATLAB. The σ_u —standard deviation of the elevation differences—and σ_p —a partially-correlated error of the mean difference—served as candidate measures of DREM precision. The method is identical to one used in [38] for assessment of TanDEM-X DEM precision over a small area.

$$\sigma_u = std(\Delta Z) \quad (1)$$

$$\sigma_p^2 = \begin{cases} 0 & L \leq \Delta h \\ n(\frac{\Delta h^2}{L^2}) + s(1 - \frac{L}{h} + \frac{1}{5} \cdot (\frac{L}{h})^3) & \Delta h < L < r \\ n\frac{\Delta h^2}{L^2} + \frac{1}{5}s\frac{r^2}{L^2} & L > r \end{cases} \quad (2)$$

$$\Delta h = \frac{\Delta x}{\sqrt{\pi}} \quad (3)$$

In these equations, σ_u is the standard deviation of a spatially non-correlated DEM difference map, σ_p is the standard deviation of a partially correlated DEM difference map, Δx is the spatial resolution of our dataset, L is the radius of a circular region whose S equals S of the ΔZ averaging region and N is the number of pixels included in averaging.

3.4. Iceberg Map Creation

Our method for iceberg map creation and analysis uses several tools for different tasks (Figure 5 and Table A1 in Appendix A).

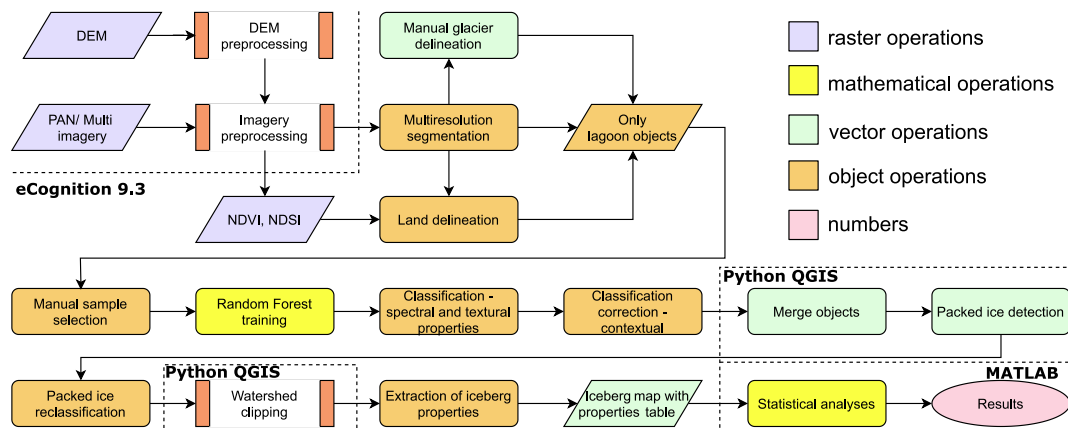


Figure 5. Entire procedure of iceberg inventory creation. Steps enclosed in white-brown rectangular boxes are shown in detail on other figures.

We used eCognition Developer 9.3. [40] to extract iceberg contours from the imagery. In the first step, we used the *Multiresolution Segmentation* algorithm to divide the satellite image into a collection of objects. This algorithm works by clumping neighboring pixels according to criteria of similarity of raster values and compactness of the resulting shape [40]. All image bands and DREM were used for segmentation. The abstract parameter of the algorithm governing the size of the resulting segments was set at 100.

As we intended to detect icebergs in the lake, we excluded two land cover categories (Land and Glacier) from the classification. We discriminated land area from the lagoon using the Normalized Difference Vegetation Index (NDVI) [41]. Objects whose mean NDVI was above 0.1 were classified as “Land”, while remaining land cover types (water, glacier, iceberg and sediment-covered ice) remained unclassified. Only these unclassified objects were included in further classifications. In addition, any sections of the image located within the land area where $NDVI < 0$ were reclassified to “Land” to avoid contamination of the final results by water puddles or rock outcrops outside the lagoon. We manually added areas of sandy and rocky beaches to the “Land” class after classification, as they were only discovered as members of “Dirty Ice” class when classification was finished.

We outlined the extent of the San Quintín glacier snout manually. The reason for a non-automated discrimination of the glacier was the presence of dense ice melange concentrated in front of the glacier

tongue. This area of the image is too chaotic to allow for reliable determination of glacier margin by automated means. We sketched a line along the margin of the glacier, creating a loop of objects enclosing the whole ice body. Thanks to the division of the image into objects, the interpreter did not have to precisely trace the margin pixel-by-pixel, but rather object-to-object, which significantly sped up the digitization effort. All objects inside the hand-drawn boundary were classified as “Glacier” and excluded from the iceberg detection procedure.

We used Random Forest (RF) algorithm implemented in eCognition to divide the set of objects into classes. Random Forest is a supervised classification method, which uses a collection of decision-tree predictors (a forest) to sort objects into classes [42,43]. The eCognition implementation is based on the OpenCV version of the algorithm [44]. We selected manually samples of classes from objects located on the lake. Table 2 shows the classes and their brief characteristics, while Table 3 lists all independent parameters used for training of the algorithm and their importance as reported by the eCognition classifier. At this stage, we performed Principal Component Analysis (PCA) to identify the object properties most important for the classification. We exported the table of properties of the classified objects and ran PCA in MATLAB on the dataset to obtain Principal Components (PC) and the loadings of particular properties on each of the PCs.

The setting of the San Quintín lagoon—isolated from the ocean—leads to accumulation of sediments brought into the lake with melting icebergs. As a result, the water of the lagoon is spectrally similar to surfaces covered in the local sediment—such as sky-facing surfaces of debris-covered icebergs. Moreover, the setting causes the areas of brash ice—clusters of icebergs too small to resolve on the available images—to look similarly to debris covered iceberg with a mixture of bright and dark objects producing a grainy texture. Therefore, corrections of the initial classification were needed to rectify the misclassified objects. Just as initial classification used statistical properties of the data within the polygons, the correction step looked at the contextual information available on the classified map of objects. We used properties such as neighborhood of particular classes, object shape and number of “holes” inside an object to move objects between classes. In this step, we assigned the objects classified initially as “Shadow” to one of the other classes. While discrimination of shadows cast by icebergs was necessary in the spectral classification step, the shadows themselves do not present value as a separate class of objects, being either sections of icebergs (e.g., in crevasses) or surface of water shadowed by a tall piece of ice. The section of ice melange in the vicinity of the glacier’s margin was also reclassified with use of contextual and spectral rules. The exact extent of this area was selected as a sum of all “Ice”, “Dirty Ice” and “Brash Ice” polygons which touch the glacier margin. These polygons were dissolved in QGIS 3 and used as an area-of-interest template to select objects in eCognition warranting the reclassification. A set of rules, different from the classification correction rules mentioned before, was developed to divide the ice melange area into four classes: “Large Icebergs”, “Small Icebergs”, “Water” and “Brash Ice”. It is worth noting that in this case we did not use any classification algorithm, just a set of rules dividing the population of objects into classes basing on their geometric, spectral and topographic properties.

We exported the shapefile map of the “Iceberg” class (a joint set of “Dirty Ice” and “Ice” classes). The post-processing of shapefile map of icebergs has been done within QGIS 3.8 environment. The polygons in the “Iceberg” class have been merged into a set reflecting the contours of icebergs, regardless of their color. In many cases, the icebergs were touching each another or were in other cases way too close to be correctly separated during the classification, resulting in single polygons covering several icebergs. We used the watershed algorithm to separate the icebergs in such cases. Watershed algorithm is an image analysis method for separation of parts of images which overlap or touch, first proposed by Beucher [45]. We used the OpenCV implementation of the algorithm in the Python 3 language [46].

Table 2. Description of image classes used to classify the image objects.

Class	Interpretation	Visible Brightness	Relative Elevation	Texture	Comments
Clear Ice	Ice without debris cover	Very bright	Positive	Uniform to moderately grainy	—
Dirty Ice	Icebergs or their sections covered with debris	Low, moderately variable	Positive	Non-uniform due to variable amount of debris	Easily mistaken for brash ice
Brash Ice	Fields of icebergs too small to be resolved individually	Moderately bright	Close to 0	Very highly textured	Easily mistaken for dirty ice due to similar brightness
Shadow	Areas in deep shadow cast by an iceberg or glacier	Extremely low	Variable, depends on location	Very uniform	“Technical” class, merged with other classes in postprocessing
Water	Open water area, without ice	Narrow band of brightness	0	Very uniform	Brightness values within range of Dirty and Brash Ice
Big Iceberg	Tabular icebergs trapped in tightly packed ice melange	Variable, from clear ice to very dark	Very high	Variable, mostly uniform	Only used to reclassify the ice melange region
Small Iceberg	Small icebergs trapped in tightly packed melange	Moderately bright, with very bright pieces of clear ice	Variable, positive, close to 0	Uniform within small objects	Only used to reclassify the ice melange region

Table 3. List of properties of image objects used as variables during the Random Forest classification and results of PCA analysis of their importance. “RF imp.” stands for “Random Forest Importance”, as returned by eCognition. Columns 3–6 list loadings of properties in the four Principal Components with largest explained variance in the classified objects dataset. Absolute loadings higher than 0.25 and properties carrying them are shown in bold. All minima, maxima, means and standard deviations are computed from pixels within a single object.

Image Object Property	RF imp.	Parameter Loadings			
		PC1	PC2	PC3	PC4
Border contrast Panchromatic band	0.024	0.08	0.07	0.68	0.17
GLCM Contrast	0.030	0.06	−0.27	0.22	0.77
GLCM Entropy	0.032	0.00	0.00	0.00	0.00
GLCM Homogeneity	0.029	0.00	0.00	0.00	0.00
GLCM Mean	0.047	0.03	0.01	0.01	−0.02
Max. relative elevation	0.031	0.00	0.00	0.00	−0.01
Max. Green band reflectance	0.053	0.29	−0.29	−0.20	−0.04
Max. Panchromatic band reflectance	0.052	0.42	−0.38	−0.11	0.00
Max. Red band reflectance	0.045	0.28	−0.27	−0.14	−0.02
Max. NDSI value	0.025	0.00	0.00	0.00	0.00
Max. Blue band reflectance	0.052	0.28	−0.27	−0.21	−0.03
Mean Blue band reflectance	0.069	0.26	0.08	0.12	−0.13
Mean relative elevation	0.044	0.00	0.00	0.00	−0.01
Mean Green band reflectance	0.066	0.28	0.07	0.13	−0.14
Mean NDSI value	0.027	0.00	0.00	0.00	0.00
Mean NIR band reflectance	0.027	0.17	0.09	0.20	−0.17
Mean Panchromatic band reflectance	0.023	0.40	0.14	0.29	−0.20
Mean Red band reflectance	0.035	0.27	0.11	0.21	−0.13
Min. NDSI value	0.024	0.00	0.00	0.00	0.00
Min. Panchromatic band reflectance	0.028	0.27	0.46	−0.17	0.22
Min. relative elevation	0.024	0.00	0.00	0.00	0.00
Min. Blue band reflectance	0.032	0.16	0.27	−0.23	0.20
Min. Red band reflectance	0.027	0.16	0.31	−0.16	0.25
Min. Green band reflectance	0.037	0.18	0.28	−0.23	0.22
Std. NIR band reflectance	0.030	0.02	−0.07	0.04	0.04
Std. of Panchromatic band reflectance	0.029	0.04	−0.18	0.11	0.22
Std. of Relative elevation	0.034	0.00	0.00	0.00	0.00
Std. of NDSI value	0.024	0.00	0.00	0.00	0.00
% of explained variance	—	83.84	10.42	1.89	1.19

Our workflow progressed on a “per-berg” basis to avoid memory-intensive graphic processing of the entire high-resolution image (Figure 6). Only icebergs larger than 100 m² were cut. The division workflow starts with a conversion of the iceberg vector polygon to a binary raster image where the value of “1” corresponds to iceberg presence and “0” to iceberg absence. With morphological erosion, thin “isthmuses” between icebergs were deleted, leading to disconnection of iceberg kernels. The watershed algorithm filled the space between the kernels, leaving pixel-lines in places where

contours of recreated iceberg polygons met. These lines were vectorized and used to cut the initial vector iceberg polygon, dividing a shape of connected icebergs into several smaller polygons.

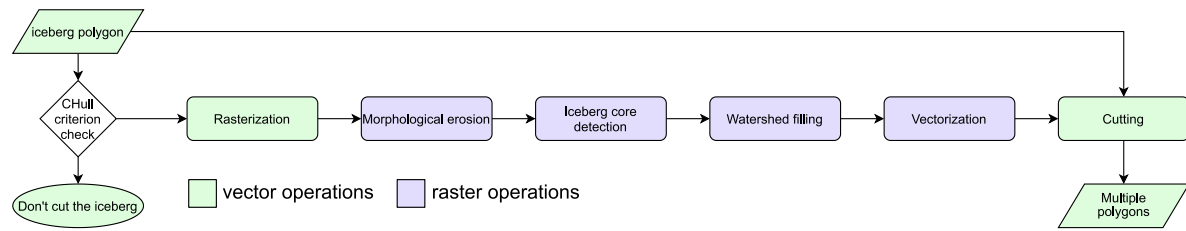


Figure 6. Procedure of watershed cutting of a single iceberg.

Finally, we imported the divided iceberg map to eCognition and extracted statistical properties (Table 3) of pixel populations enclosed by the iceberg contours. Additionally, we computed the volume of each iceberg with use of the Equation (4):

$$V = \mu Z_{rel} \cdot S \cdot \frac{\rho_w}{\rho_w - \rho_i} \quad (4)$$

where V (m^3) is the iceberg's entire (under- and overwater) volume, μZ_{rel} (m) is the mean elevation above the water table and S (m^2)—the surface area of the waterline cross-section. We assumed ice density $\rho_i = 900 \text{ kg m}^{-3}$ and water density $\rho_w = 1000 \text{ kg m}^{-3}$, as the lagoon is filled with fresh water. This amounts to an assumption that 90% of iceberg volume is submerged.

The mean relative elevation of an iceberg was multiplied by 9 to obtain iceberg's total height. The lower multiplier in this case stems from an assumption of irregular shape of an iceberg. The draft of an iceberg computed by multiplying its freeboard by the ice density ratio was shown to overestimate the iceberg's depth [47].

3.5. Iceberg Map Uncertainty Assessment

We provide two measures of uncertainty of the iceberg map created with the method described above: the error of the number of icebergs detected by the algorithm and the error of the surface area of an extracted iceberg. To obtain these values, we compared the classified iceberg map to a manually created map. Such approach has been previously used [24], albeit our procedure is not identical to it. We divided the area of the lagoon without the proximal ice melange field into a grid of squares with 500 m long sides. We randomly chose 10% of the squares as verification fields (VF, see Figure 1) and manually digitized all icebergs located within them on the map of segments produced in eCognition. A manual map of icebergs was exported as a shapefile and compared to a subset of the iceberg inventory overlapping with the VF. We calculated area error by overlying the two shapefiles and subtracting the areas of the overlapping polygons. We also counted false positives and false negatives. The former are icebergs from the automatically classified dataset which do not overlap with any icebergs from the manual dataset. The latter are those of the manually digitized polygons that do not have counterparts in the classified dataset. The same procedure was repeated on the ice melange area but with a grid of 250 m wide squares. We calculated area error by comparing areas of polygons of icebergs produced by classification (S_{cla}) with area of manually digitized iceberg polygons (S_{man}). To ensure comparability between area errors between many size scales, we computed unitless relative error of area (ΔS_{rel}) with Equation (5).

$$\Delta S_{rel} = \frac{S_{man} - S_{cla}}{S_{man}} \quad (5)$$

Histograms of frequency of area were produced, as well as 2D histograms of ΔS_{rel} to present the classification quality assessment results.

3.6. Data Filtering

Before fitting area–volume scaling model, we applied filters to the obtained iceberg inventory to discard cases, which were unreliable due to the DREM imperfection:

- We used k-nearest neighbors (knn) clustering algorithm to discard outlying data points. All items which had fewer than four neighbors within radius of 0.2 in the log area–log volume space were considered outliers.
- All icebergs with $\mu Z_{rel} < \sigma_{DEM}$ were discarded. We assumed objects to be unreliable when their elevation measurements are below the precision of the DREM.
- We filtered out icebergs which were unrealistically high with respect to their width. We used a formula proposed in [48] (Equation (6)) for the critical width/height ratio at which an iceberg capsizes. We assumed that any iceberg with width/height below the critical ratio should capsize and thus the elevation measurement is not realistic.

$$\epsilon_c = \sqrt{6 \cdot \frac{\rho_i}{\rho_w} \left(1 - \frac{\rho_i}{\rho_w}\right)} \quad (6)$$

where ϵ_c is the critical ratio, equal to 0.73 under our assumed ice and water densities.

In addition, we divided the iceberg population into sub-populations of small and large icebergs. To determine the threshold for the division, we created a volume variance signal by sorting the dataset by area and computing mean volume variance within a 100-wide running window. We found a change point (V_{var}) in this signal by dividing the signal into two parts at many candidate points and choosing the one at which the sum of deviations from the mean within each section is minimal.

3.7. Area and Volume Distribution

We investigated statistical distributions of volume and area of icebergs with use of *powerlaw* package implemented in Python 3 programming language [49]. This suite of functions uses fitting methods described by Clauset et al. [50] to accurately fit power law, exponential and log-normal distributions to datasets. We fitted power law, exponential, log-normal and exponentially truncated power law to the vectors of surface area and volume of icebergs. Both datasets were filtered to remove icebergs smaller than 4 m² (corresponding to 1 pixel on the 2 m/pixel DEM, smallest available size). We assessed the quality of fit of the curves with Kolmogorov–Smirnov (*K-S*) measure provided by the *power law* [49]. In addition, normal distribution curves were fitted to histograms of logarithm of area and volume of filtered dataset with the least squares method in MATLAB. Only icebergs larger than 1 m² were used for the fitting effort.

3.8. Area–Volume Scaling

We searched for the area–volume scaling model with use of MATLAB. We fitted models to the area–volume data points in two cases: with volumes computed with DREM and DREM₀. We fitted power law model to the entire unfiltered inventory, to the unfiltered sections on the two sides of the V_{var} threshold and to the fully filtered dataset. In addition, a linear model was fitted to the log area–log volume filtered data. The fits were performed with non-linear least squares method in case of power law fitting and linear least squares for log-log model fitting. We used two measures to assess the quality of fitted models:

- R^2 of the fit; and
- CV_{10f} —the product of k-fold cross-validation (CV) of a model. We divided the area–volume dataset into 10 subsets (folds) and performed the fitting with 9 subsets leaving one of them out for verification. R^2 of the volumes computed with the fitted line vs. the measured volumes was computed at each of the repetitions and the mean of these R^2 s is reported as CV_{10f} .

4. Results

4.1. DEM Accuracy Assessment

The pixel-wise difference between WV-2 DEM and TanDEM-X 90 DEM does not follow a normal distribution (Figure 7). Thus, the typical measures of bias and precision of an elevation model, mean and standard deviation of ΔZ_{rel} , could not be used as DEM quality measures. The second reason against these values is the presence of spatial correlation within the elevation models. It invalidates the use of σ_u as precision measure, as this value is computed under assumption of lack of spatial correlation of individual pixels. We calculated the correlation distance to be 1381.15 m on the difference map. The parameters of semi-variograms led us to a value of the $\sigma_p = 1.70$ m, used as a measure of DREM precision further on.

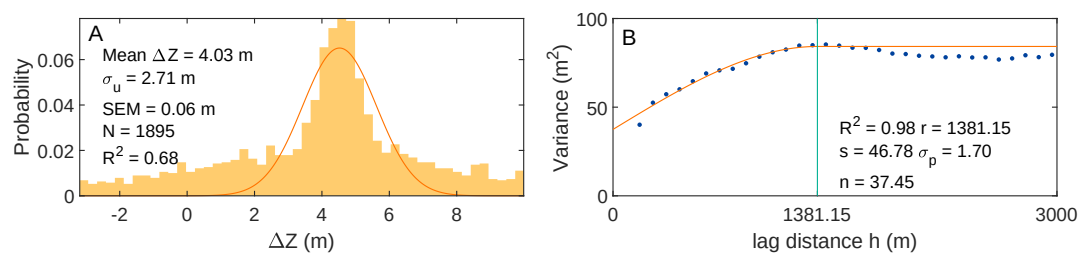


Figure 7. Results of the accuracy assessment of the WV-2 DEM as compared to the TanDEM-X 90 m/pixel DEM: (A) histogram of the values of the WV-2 DEM-TanDEM-X DEM difference map, where the orange line represents the normal distribution fitted to the histogram and (B) empirical (blue dots) and fitted (orange line) semivariogram of the same difference map with parameters of the fitted synthetic variogram, where the green line indicates the correlation range.

4.2. Dataset and Filtering

The iceberg inventory consists of over 38,000 records (Figure 8 and Table 4) each corresponding to a single detected iceberg. the majority of icebergs are between 10 and 1000 m^2 of area with just around 24% of cases outside these bounds. The largest icebergs ($S > 10^4 m^2$) contribute only under 1% of the unfiltered dataset.

Table 4. Statistical summary of the subsets of icebergs generated by different filtering methods.

Filter	Whole Dataset	Only V > 0	knn Outliers Removal	DEM Error	Iceberg Height	All Filters
N of records	38,781	20,170	20,110	6441	11,688	933
% of volume	100	100	93.06	93.66	93.56	80.99
Iceberg area percentages						
$S < 4$	18.43	15.26	15.21	20.43	5.31	0.00
$10 < S < 100$	53.36	44.25	44.34	43.81	38.07	0.00
$100 < S < 1000$	21.22	28.94	28.98	21.32	38.61	18.76
$1000 < S < 10,000$	4.93	9.03	9.01	9.39	14.50	51.34
$S > 10,000$	0.91	1.75	1.70	4.38	3.01	29.90

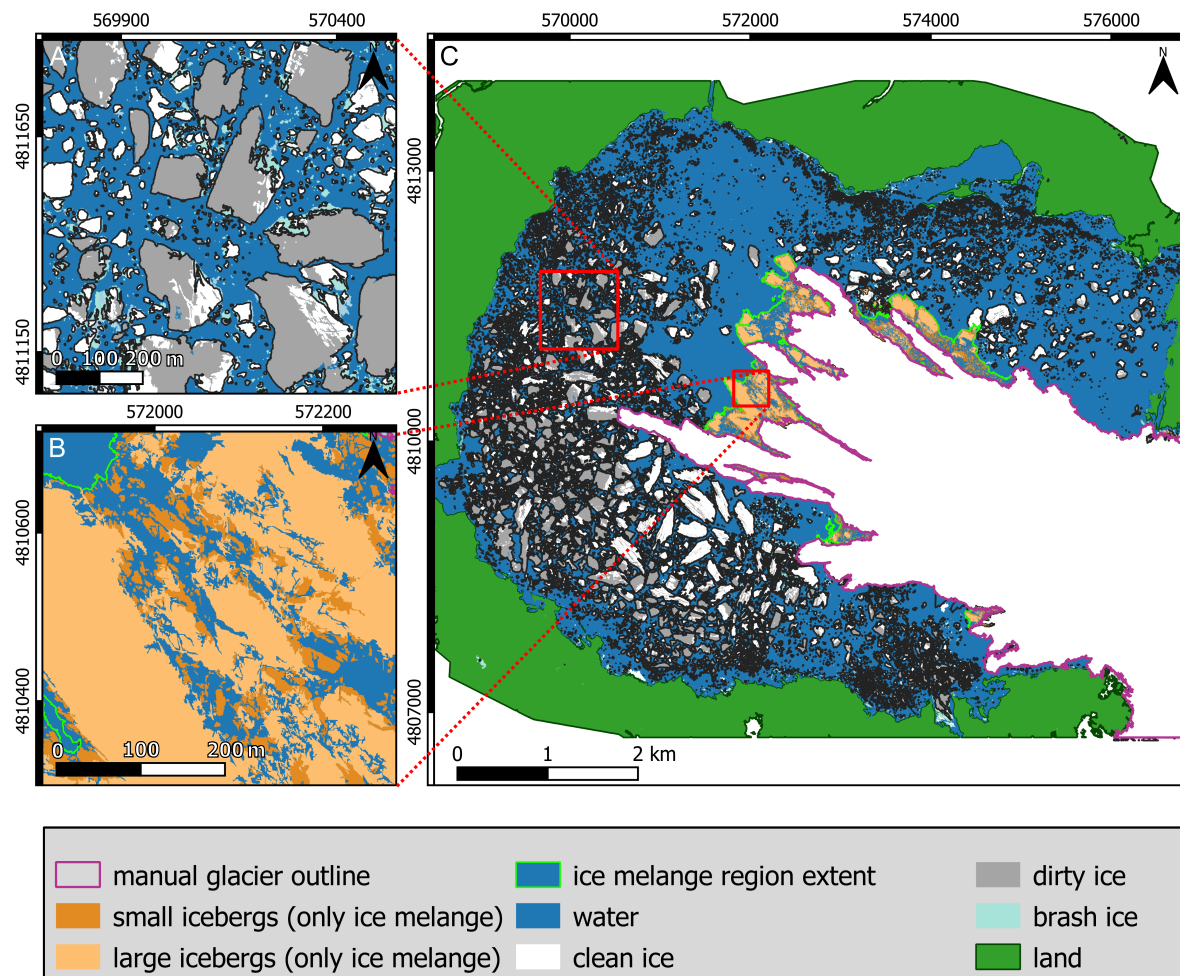


Figure 8. Final result of the iceberg map creation process: (A) example of iceberg map in the lagoon area; (B) example of iceberg map in the ice melange area; and (C) the overview of the classification over the whole lagoon and its surroundings. Coordinates in UTM 18S coordinate system.

The impact of the three filters on the dataset has been highly variable, as each filter discards a different subset of icebergs (Figure 9). The k-nearest neighbors filter eliminates only a small number of data points outlying from the clustered bulk and its impact is the lowest of the three filters. Elimination of the icebergs with mean elevation below the precision of the DREM trims the entire lower part of the cluster, corresponding to almost 70% of all data points (Figure 9).

Despite that, the icebergs remaining in the pool constitute 94% of the entire volume of icebergs found in the lagoon. The iceberg stability filter operates in an opposite manner to the previously described one. What remains after removal of the icebergs too high to be stable from the dataset, are the data points located in the lower part of the cluster and a subset of the largest icebergs (Figure 9). This filter is less restrictive than the DEM Error (over 57% of icebergs remaining) but the summary volume of the remaining icebergs is slightly lower (92% of the total). Application of all three filters to the dataset leaves only a small group of 933 icebergs, corresponding to 4.63% of the entire sample (Figure 9), which is, however, responsible for nearly 80% of the total volume of the lagoon's floating ice.

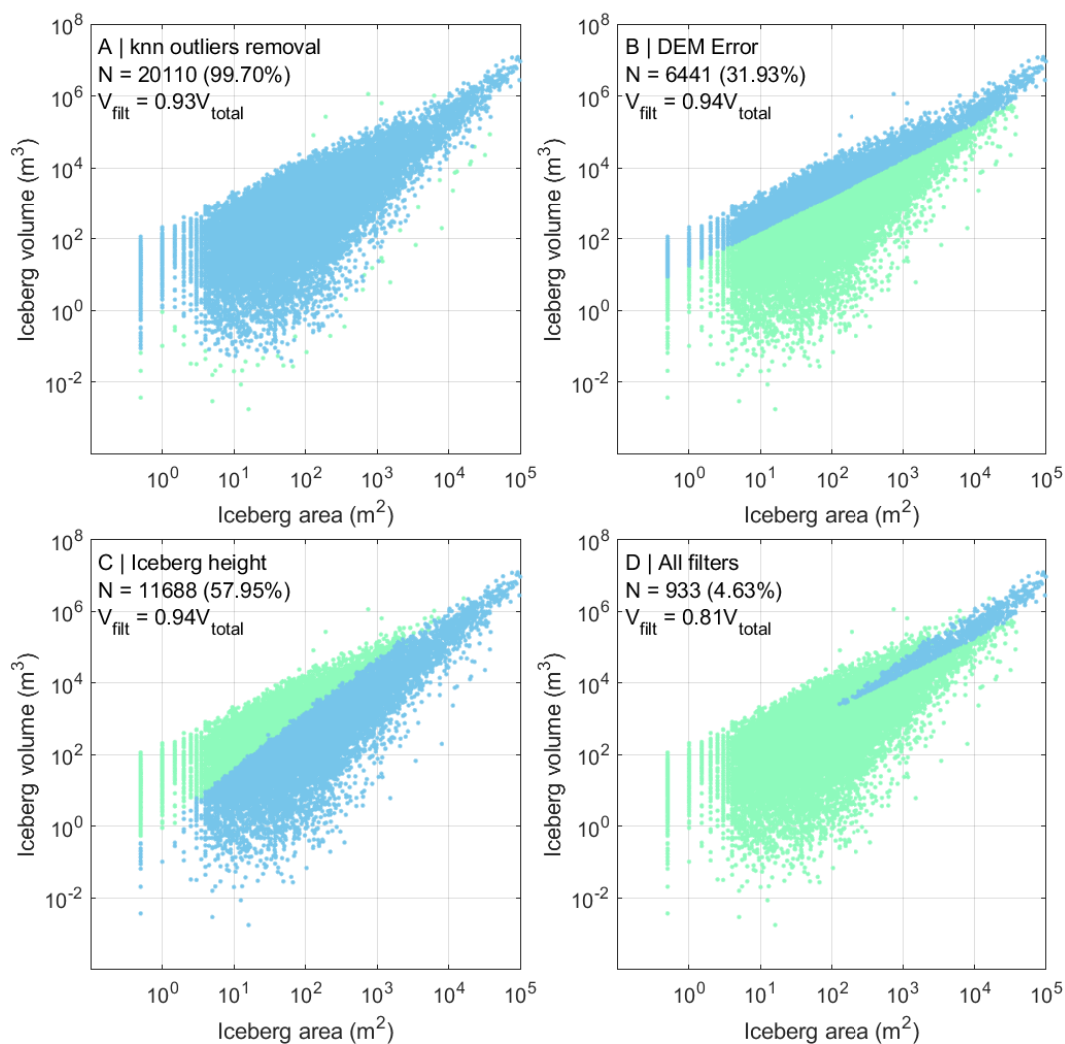


Figure 9. Subsets of the area–volume dataset produced by applying different filters. Discarded data points shown in green: (A) outliers removal with *knn clustering*; (B) icebergs with $\mu Z_{rel} < 1.7$ m removed; (C) icebergs with width/height ratio below 0.73 removed; and (D) all filters applied together. Discarded points shown in green, and the accepted ones in blue.

4.3. Classification Accuracy Assessment

The overall number of icebergs detected by the classification in the open-lagoon verification fields is very close to the number produced by manual digitization within the open-lagoon VF (3184 vs. 3213 icebergs, Table 5). The icebergs' frequency of occurrence for particular sizes is also similar between the manually and the automatically mapped icebergs (Figure 10). The only exception is underestimation of the number of icebergs with the sizes in the 5–15 m² range. This is the size band in which the majority (247) of false negatives are found. A sizeable group of undetected icebergs (151) is also present among the smallest growlers (0–5 m²). These two groups account for the underestimation of iceberg count in these bands on the histogram. The remaining 96 false positives are all below 300 m².

The area error increases gradually with the increase of area of a manual iceberg (Figure 10). A trend of positive area errors close to or at 100% of the actual area (indicating detected icebergs significantly smaller than their manual counterparts) is a trace of the false negatives. Relative area errors of icebergs in the lagoon assume diverse, mostly negative, values but are concentrated between −1 and 1.

Table 5. Quantities of icebergs used in classification quality assessment. N_{cla} stands for number of classified icebergs, N_{man} – number of manually digitized icebergs and VF – Verification Fields.

Area	N_{cla} Whole Set	N_{cla} within VF	N_{man} within VF	Detected Icebergs	Icebergs within VF	False Positives	False Negatives
Lagoon	38781	3184	3213	99.10%	8.21%	137	494
Ice melange	9275	1702	535	318.13%	18.35%	303	55

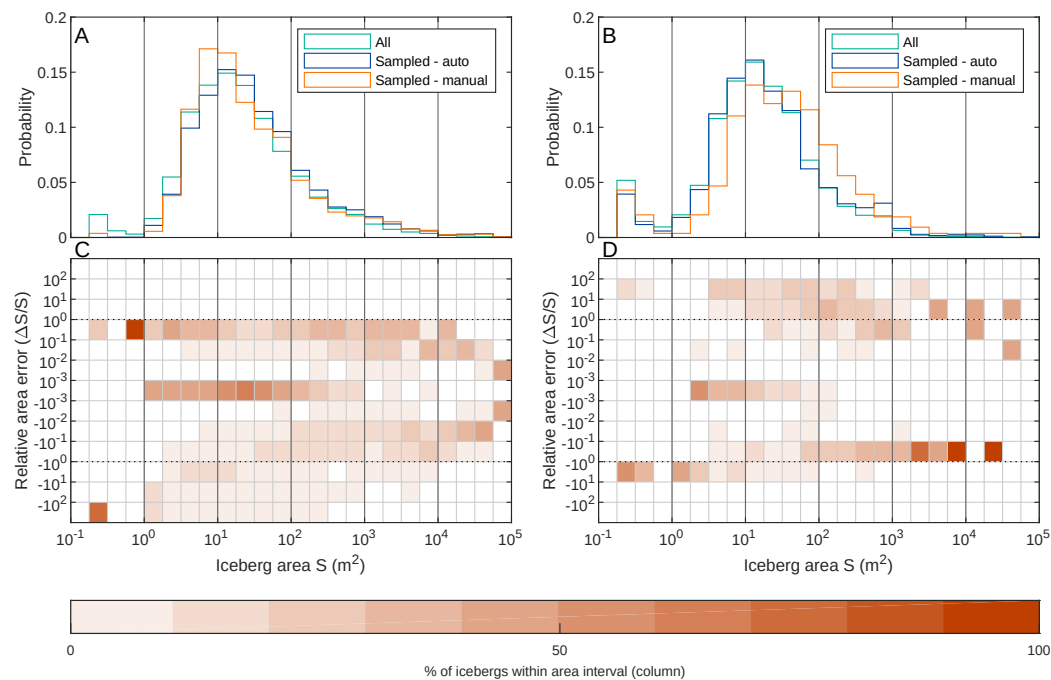


Figure 10. Results of quality assessment of iceberg map: (A,B) histograms of frequency of iceberg areas; and (C,D) 2D histograms of unsigned relative error of manually digitized icebergs compared to automatically mapped ones. Plots (A,C) relate to the lagoon, while (B,D) to the fields of ice melange. Dashed lines show the lines of 100% relative error. Values within each column add up to 100%. Relative differences were computed as $\frac{S_{man} - S_{cla}}{S_{man}}$. X-axis divisions spaced by equal $\log_{10} S$ for histogram readability.

The classification method used in the ice melange area performed visibly worse than the RF-classification of the lagoon. It has to be noted, however, that the RF-classification on the ice melange region returned an entirely unreliable map of "icebergs" consisting only of a few very large polygons with holes filled with brash ice inside and between them. The histogram of manually picked icebergs is shifted, showing that the ruleset we used underestimates the number of large icebergs and overestimates frequency of small ones (Figure 10). Contrary to the lagoon area, in this case, the errors of area compared to manually picked bergs are less diverse in the large area bands and decidedly higher for the small icebergs than for the big ones. The largest error of this case is a large overestimation of the total number of iceberg compared to the manual set (Table 5). Many of the classified polygons overlap partially with the manual polygons. This explains low number of false positives.

In both cases, the area frequency of the whole iceberg populations are nearly identical to the frequencies of the automatically classified populations from within the verification fields showing the representativeness of the verification areas for the entire dataset. The percentage of verification icebergs relative to the entire inventory (8.21% in the lagoon and 18.35% in the ice melange region) is also satisfactory.

Four principal components explain over 97% of variability between the classified objects (Table 3). Results of PCA show that reflectances in the visual bands are the most important feature of objects for the Random Forest classifier. Minima and maxima of the red, green, blue and panchromatic bands bear highest loadings on the four PCs (Table 3). Among the other properties, only *GLCM Contrast* and *Border Contrast* on the panchromatic band are important for the PC values, although only in the final two PCs, which are together responsible for only about 3% of the overall variability.

4.4. Area and Volume Distribution

The statistical distribution which fits both the area and volume histograms best is the log-normal distribution, as evidenced by the highest K-S measures of all tested models (Figures 11 and 12). In the case of area distribution, $K-S = 0.03$, while, for volume distribution, it is slightly higher ($K-S = 0.04$). In both cases, the value is lower than that of exponentially truncated power law ($K-S = 0.08$ in both cases). Goodness measures of power law and exponential distributions are several times higher than those of log-normal distributions. Exponential distribution shows the poorest fit to the area frequency among the four tested models ($K-S = 0.54$). The very poor fit of power law to volume histogram is worth noting. Its $K-S$ (0.50) is higher even than that of the exponential model (0.46). The area distribution seems to adhere to power law on a limited section, but the model fails to recreate the tail of the dataset, corresponding to the largest icebergs (Figure 11). In the case of distribution of volume, the log-normal distribution seems imperfect at the extreme ends of the distribution, near the smallest and largest icebergs (Figure 12).

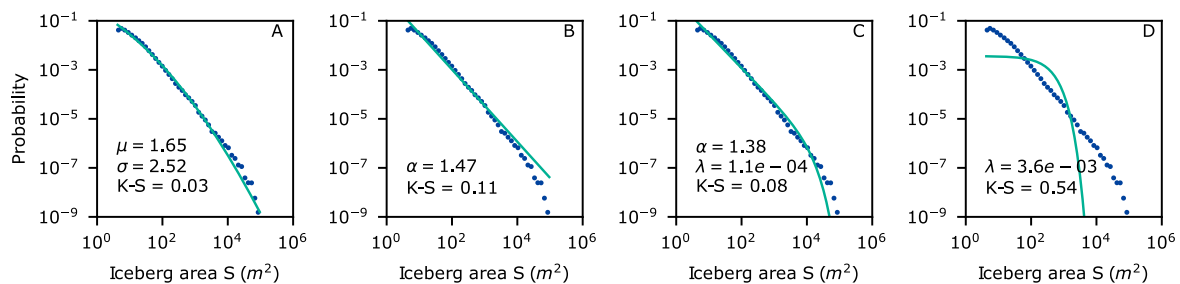


Figure 11. Fitting of heavy-tailed distributions (green lines) to logarithmically binned area data (blue dots): (A) log-normal $P(S = x) = \frac{1}{x\sigma\sqrt{2\pi}}\exp(-\frac{(\ln x - \mu)^2}{2\sigma^2})$; (B) power law $P(S = x) = x^{-\alpha}$; (C) exponentially truncated power law $P(S = x) = x^{-\alpha}e^{-\lambda x}$; and (D) exponential $P(S = x) = e^{-\lambda x}$.

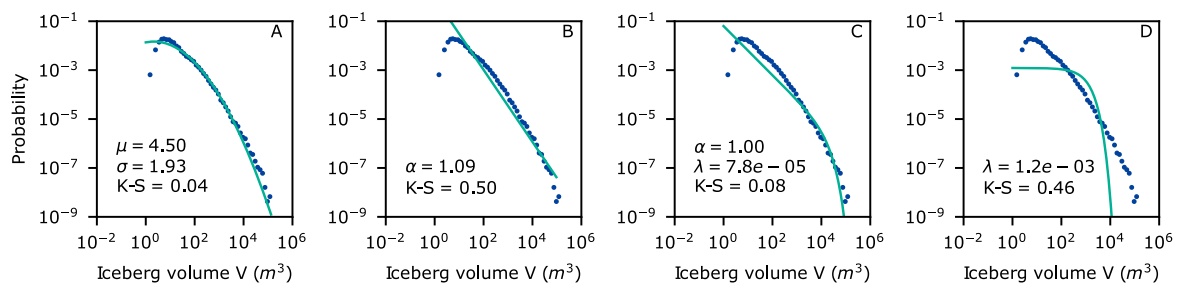


Figure 12. Fitting of heavy-tailed distributions (green lines) to logarithmically binned volume data (blue dots): (A) log-normal $P(V = x) = \frac{1}{x\sigma\sqrt{2\pi}}\exp(-\frac{(\ln x - \mu)^2}{2\sigma^2})$; (B) power law $P(V = x) = x^{-\alpha}$; (C) exponentially truncated power law $P(V = x) = x^{-\alpha}e^{-\lambda x}$; and (D) exponential $P(V = x) = e^{-\lambda x}$.

The log-normality of these distributions is confirmed by fitting the normal distribution curves to the histograms of logarithms of the respective measures (Figures 13 and 14). These achieve high R^2 values (over 0.9) regardless of whether a complete or filtered dataset is used. The exception to that is the filtered dataset of DREM₀-based data where the R^2 of the Gaussian distribution fitted to the area and volume histogram are 0.87 and 0.77.

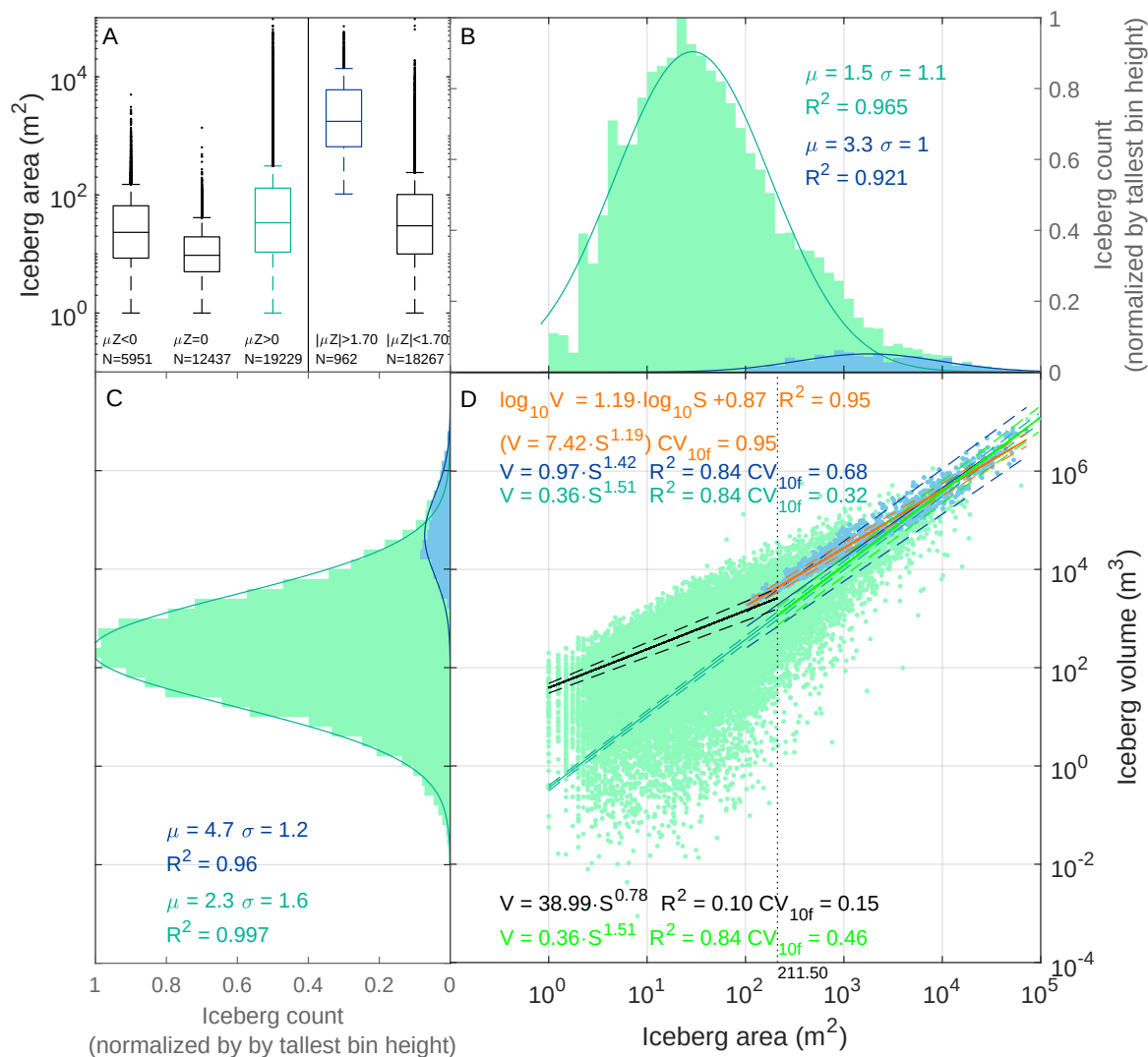


Figure 13. Results of filtering and model fitting to the area–volume data. Unmodified DREM with negative values used: (A) area distributions of iceberg classes selected by μZ_{rel} ; (B) histograms of iceberg area; (C) histograms of iceberg volume; and (D) power law relationship fitted to the area/volume data. Green fields indicate whole $\mu Z_{rel} > 0$ dataset, blue—data filtered. Green lines are models (Gaussian distributions in (B,C), power law models in (D)) fitted to unfiltered data, blue—to the filtered data. Orange line is a linear model fitted to the log-log data. Dashed lines indicate 95% confidence intervals of fitting.

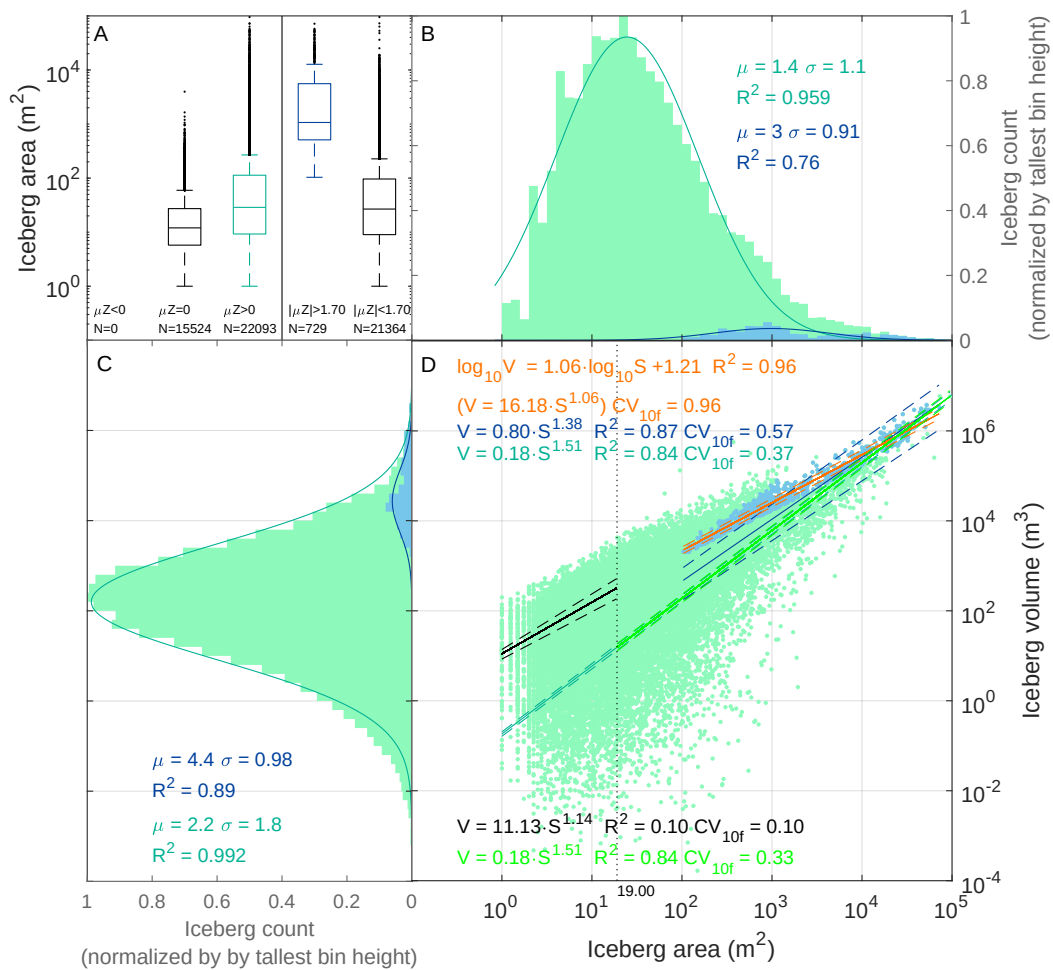


Figure 14. Results of filtering and model fitting to the area–volume data. DREM₀ with negative values replaced with 0 used: (A) area distributions of iceberg classes selected by μZ_{rel} ; (B) histograms of iceberg area; (C) histograms of iceberg volume; and (D) power law relationship fitted to the area/volume data. Green fields indicate whole $\mu Z_{rel} > 0$ dataset, blue—data filtered. Green lines are models (Gaussian distributions in (B,C), power law models in (D)) fitted to unfiltered data, blue—to the filtered data. Orange line is a linear model fitted to the log-log data. Dashed lines indicate 95% confidence intervals of fitting.

4.5. Area–Volume Scaling

Replacement of negative relative elevation ($Z_{rel} < 0$) values with 0 in DREM₀ distorts the log-normality of the volume distribution by introducing to the population a large number of small icebergs. This is manifested by a slightly lower R^2 of the Gaussian distribution fitted to the histogram of the modified volumes and over-representation of low-volume cases. (compare Figures 13 and 14). The number of icebergs with $\mu Z_{rel} = 0$ is also higher when the modified DREM is used, showing large population of icebergs which did not contain any positive DREM values within their outlines. On the other hand, the population of $\mu Z_{rel} > 0$ icebergs (green box plots in Figures 13 and 14) has also increased, hinting at presence of cases where the negative values over-weighed positives on DREM, but the positive values shifted the mean above 0 in DREM₀. Apparently, certain number of such individuals is found even in the filtered dataset, among the fairly large ($S > 99 \text{ m}^2$) icebergs, as evidenced by lowering of median iceberg area in the filtered dataset used for fitting (Figures 13 and 14). The imperfections introduced to the dataset and uncertain status of the negative Z_{rel} measurements led us to assume the unmodified DREM as the basis for further analyzes.

The choice of a fitting method of the model visibly impacts the result. The linear model fitted to the logarithms of the areas and volumes ($\log_{10} V = a \cdot \log_{10} S + b$) shows better R^2 than power law models fitted to the untransformed data and performs better in the K-fold cross-validation test (Figure 13). Depending on the subset of data and fitted model, the area–volume scaling model assumes different values and its quality varies. The best fit is achieved by a linear model fitted to the logarithms of areas and volumes. The log-linear model $\log_{10} V = 1.19 \pm 0.02 \cdot \log_{10} S + 0.87 \pm 0.06$ fitted to the filtered dataset achieves better fit ($R^2 = 0.95$) and performance (K-fold CV score $CV_{10f} = 0.92$) than other models. This model can be transformed into a power law form of $V = 7.42 \pm 1.14 \cdot S^{1.19 \pm 0.02}$. The power law scaling laws assume different forms when fitted to the different subsets of the inventory. The power law $V = 0.97 \pm 0.52 \cdot S^{1.42 \pm 0.05}$ fitted to the same filtered dataset as the log-linear model produces best quality measures ($R^2 = 0.84$, $CV_{10f} = 0.72$). This variant has the widest confidence intervals and visibly deviates from the point cloud below 1000 m² of area. A model fitted to the whole dataset with the scaling formula $V = 0.36 \pm 0.04 \cdot S^{1.51 \pm 0.01}$ sports similar $R^2 = 0.84$, but much worse $CV_{10f} = 0.49$.

We fitted two additional scaling models using the unfiltered dataset bisected with the volume-variance threshold. These yielded two very different products. The model $V = 38.99 \pm 8.71 \cdot S^{0.78 \pm 0.05}$ fitted to the subpopulation of small icebergs ($S < 211.50$ m²) performed very poorly ($R^2 = 0.10$, $CV_{10f} = 0.07$), its scores lowest of all fits. The model based on the large icebergs, in turn, has identical formula to the model fitted to the entirety of the dataset, albeit with wider fitting confidence intervals ($V = 0.36 \pm 0.10 \cdot S^{1.51 \pm 0.03}$). Its R^2 is also identical, while $CV_{10f} = 0.53$ differs in favor of the more limited model.

5. Discussion

5.1. Iceberg Classification

Results of the classification present that the Random Forest classification of objects gives very good results in iceberg detection in environments where the floating icebergs can be easily discerned from the background (water). In areas where icebergs are crowded without much space between them and in the presence of small growlers lying on top of large icebergs, neither Random Forest classification nor manually tuned classification ruleset could succeed in correctly discerning individual objects. A challenge encountered in the setting of the San Quintín lagoon was the spectral similarity of the sediment-laden water to the sediment-covered icebergs. While in most cases the elevation criterion was a sufficient determinant of whether an object belongs to the “Dirty Ice” or “Water” class, the objects located in the icebergs’ marginal zones were often flat. We have tackled this problem at the post-processing stage of classification by eliminating objects which were located between “Water” and “Dirty Ice” classes and had very elongated shape (length to width ratio greater than 3). These cases represented sections of image where an intensity gradient was present between the darker iceberg and more uniform water. The segmentation algorithm separated such group of pixels as an entity separate from the wide, uniform field of water and different from the more grainy surface of the iceberg. At the same stage we eliminated from the “Ice” class objects which represented underwater ice visible as gradient of brightened water color. Due to lagoon water’s spectral similarity to the debris, these underwater portions of clear, white ice were often classified as “Dirty Ice”. We predict that our method will perform better on imagery of fjords where reflection from deep and dark water contrasts strongly to very bright ice.

The same spectral overlap hindered the multi-scale approach to object-based classification. In this method, the image is segmented and classified several times with smaller or larger objects at each level in order to detect icebergs of different size [6]. On our image, where borders between icebergs and water are sometimes fuzzy, large scale classifications, instead of separating largest icebergs from their surroundings, produced polygons which included sections of water or brash ice within the “Dirty Ice”

class. Therefore, we had to classify the entire image at a low level—subdivided into a larger number of smaller objects—to detect subtle borders between debris-covered icebergs and their surroundings.

Our attempt at mapping icebergs tightly packed in the area proximal to the San Quintín glacier snout has not been fully successful. In large parts of this region ice melange resembles more a continuous field of bright ice than a collection of discrete icebergs or a grainy section of small ice pieces similar to brash ice. In others, the borders between particular icebergs, or even between the glacier and free-floating ice, were challenging to discern even by visual inspection, not to mention automated methods. The lack of clearly delimited shapes together with a spectral and textural similarity of image sections led to imperfect segmentation and erroneous classification. The properties of very tightly packed, bright ice melange are very similar to those of solid pieces of clear ice. Therefore, the classification algorithm has problems differentiating between these two surface types. Actual icebergs, and only large specimens, could be separated in the ice melange environment by following edges and shadows. A different method of detection would be needed for such cases, one that would unify OBIA with pixel-based methods such as edge detection and denoising.

Our workflow as presented in this work relies on many different software suites. Use of proprietary eCognition and MATLAB was motivated by their interactivity—it was possible for us to control quality of the outputs at each stage of the development of the method. As the method relies on generic algorithms (intensity-based segmentation, Random Forest classifier, Watershed cutting and property-based object reclassification), it could be theoretically implemented within a single environment, e.g. in Python programming language, in which all necessary algorithms are already available as existing libraries [43,46,51]. This article demonstrates the use of our proposed method on very high resolution image of lacustrine icebergs. There are however no technical reasons that would prevent upscaling the study to a different environment (tidewater calving), larger size or number of lakes and imagery with different spatial resolutions. Together with the aforementioned possibility of translating our method to a more unified software environment, we believe in very wide applicability of the proposed workflow.

Pre-eminence of reflectance as most important property during RF classification further lifts hopes for use of our method on different and larger datasets. DEMs are not always readily available, particularly in multi-temporal studies. Large loadings of maximum and minimum of reflectances in visual bands demonstrate that OBIA carries added value to the iceberg detection efforts. Division of polygons into classes can be done more effectively when statistical properties of pixels within objects can be used than just by comparison of reflectances of individual pixels. In the San Quintín lagoon where spectral differences between icebergs and water are not always unambiguous, statistical measures derived from populations of pixels provide more reliable means of differentiating between water (where difference between highest and lowest reflectances within an object is low), ice (with low reflectivity range but high minimal reflectivity) and dirty ice (where low maximum reflectivity with potentially high range of values can be expected)

5.2. Iceberg Inventory

Our iceberg inventory is a reliable source of knowledge about areas and spectral or textural properties of icebergs in the open part of the terminal lagoon of San Quintín glacier. The object-based iceberg mapping allows for collection of diverse information about icebergs that can be used for glaciological studies. The in-depth analysis of relationships between these properties is outside the scope of this work. A potential use for a high resolution iceberg inventory is as a training set for lower-resolution remote sensing exercises. A 10-30-meter-wide Sentinel or Landsat pixel can contain clear water, several very small icebergs or a part of a big one, with a possibility for clear and dirty ice in each case. Spatially detailed inventory could be used to produce datasets of iceberg density or color in grids collocated with a lower resolution image. Such basemaps could be used to improve assessment of iceberg coverage with popular medium- and low-resolution spaceborne sensors. A similar work

could be conducted to create maps of surface properties relevant for radar imaging (such as volume or angularity within a pixel), if iceberg elevation data were available.

The results of the filtering show great importance of the large icebergs for the calculation of the total volume of the floating ice. This is a consequence of the heavy-tailed distribution the set adheres to. Very wide range of size scales, ranging from less than 1 m² to several tens of thousands m², means that a single, very large iceberg will contribute more ice to the lake than even several hundred smaller ones. The relationship is even more pronounced when ice volume is considered, a property which has higher order of magnitudes, and thus higher absolute differences between largest and smallest cases. Despite the very restrictive filtering imposed upon our iceberg inventory, resulting low fraction of reliable records accounts for nearly 80% of the ice in the lagoon. This value might, however, be inaccurate due to the uncertainty of height of the icebergs. Fitting of a power law model to the subset of small icebergs showed very little correlation between their areas and volumes. This lack of relationship is an argument for their exclusion from the scaling law determination. In the realm of small areas and low freeboards, the 1.7 m of DEM uncertainty introduces noise much stronger than the actual signal.

5.3. Volume Assessment

The quality of the DEM is a constraint on the calculation of the area–volume relationship. Filtering of icebergs lower than the 1.7 m threshold of precision effectively eliminated the majority of small icebergs. The criterion of iceberg stability also favors large icebergs, filtering out the small and tall ones. The large number of unrealistically tall icebergs testifies to overestimation of elevation by our DEM. The low precision naturally impacts small, low icebergs the most. At the critical width/height ratio of 0.73, the precision limit of 1.7 m of overwater height translates to a minimum width of a stable iceberg at $W_{crit} = 9 \cdot 0.73 \cdot 1.7 \text{ m} = 11.24 \text{ m}$, leading to the minimal critical area of $\pi \cdot (\frac{W_{crit}}{2})^2 = 99.28 \text{ m}^2$ of waterline cross-section area assuming perfectly cylindrical model of iceberg. Any smaller iceberg taller than the 1.7 m would be forced to capsize. Thus, 99 m² is a theoretical lower limit of iceberg area about which meaningful elevation information can be obtained from our DREM. Despite the availability of icebergs as small as 4 m² in our inventory, the precision of the DEM (1.7 m as compared to TanDEM-X DEM) limited our fitting effort to just the icebergs taller than the precision limit and larger than 99 m². In the case of large icebergs, the very large number of pixels and high general elevation diminishes the impact of elevation uncertainty on the volume estimate. Moreover, large icebergs tend to have rich texture which lends more potential tie-points to the stereo correlation algorithm used to produce DEM. The elevations detected on highly crevassed or multi-colored icebergs can thus be considered more reliable than those computed from uniformly colored ones, particularly dark ones easily blending with surrounding water.

5.4. Area–Volume Scaling and Distributions

It was shown on an example of a tidewater glacier that the the distribution of iceberg size is different close to the calving terminus (iceberg source) than in sections of fjord farther away from a glacier [15]. Close to the source, the size of fresh icebergs follows power law distribution, which changes to log-normal as the distance from the calving front increases and icebergs have more time to decay. A pronounced log-normal distribution visible in our dataset (both areas and volumes) suggests that the old icebergs dominate the population. This is an expected result, as the lagoon serves as a trap in which icebergs decay in nearly unchanged conditions for their entire life cycle without removal of old bergs by other means than melting into the lake. It was shown that calving rates of lacustrine calving fronts are typically smaller than those of tidewater glaciers [52], albeit lake terminating glaciers calving more intensely than neighboring tidewater ones were also reported [53]. These factors working together mean that, compared to a typical marine-terminating glacier, we are looking at an increased number of old icebergs and decreased influx of fresh ice bodies. These trends increase proportion of old icebergs in the overall population and strengthen the log-normality of the distribution.

Kirkham's iceberg inventory had its lower bound at around 900 m² of waterline cross-sectional area [15], while our dataset extends the range of sizes down to 4 m² hinting at universality of the physical processes that lead to the log-normal size distribution of older icebergs. Not only does the law hold in smaller sizes, it is also possible to find it in a lacustrine setting in temperate climate zone where conditions for iceberg decay are drastically different than in an Arctic fjord.

Better R^2 and CV_{10f} scores of linear model fitted to log-log data than of power law model fitted to untransformed data hints that the power law might not be the optimal model to explain the relationship between waterline cross-sectional area of an iceberg and its total volume. The straight line on a log-log plot, often assumed as the prime indicator of the presence of a power law, can appear also when other models are actually involved [50], as evidenced by our investigation of the distribution of the iceberg sizes. It is often the case that a power law holds only over a section of the dataset, while the extreme ends deviate from this trend [50]. There is a possibility that power law is a sufficient model to explain area–volume scaling over only a portion of the iceberg population we used for fitting. In such case, below (or above) some size threshold a different mathematical relationship would take over.

Sulak et al. [18] performed a area–volume scaling exercise similar to ours on a population of icebergs derived from Landsat imagery and a WV-2 based DEM. The exponents obtained by these authors were spread between 1.26 and 1.33. Our values of 1.19 ± 0.02 and 1.42 ± 0.05 provide a similar, but wider range. Both the models of Sulak et al. [18] and ours were fitted to a multi-scale population of icebergs (size range between 10² and 10⁶ m²) with the least squares method. This fitting technique minimizes squares of residuals between the observed and modeled values. In a multi-scale study such as iceberg size analysis, the residuals in the heavy tail of the population (where absolute values of areas and volumes are very high) are much higher than in the lower orders of magnitude. Therefore, an algorithm intended to minimize these residuals will always favor fits that perform well in the region of very big icebergs and will neglect smaller ones. This approach is not necessarily bad when one tries to model the overall volume of ice because, as briefly shown above, the largest icebergs contribute the majority of ice volume in the set. This behavior of least squares is also highlighted by identical formulas of the power law model fitted to the entire population of icebergs and just the $S > 211$ m² part (Figure 13). The weight of the largest icebergs is so great that the addition of a large, uncorrelated subset to the fitting did not change the course of the line. The specifics of least-squares fitting may be the reason why fitting a linear model to log-log data performed better. In this case the least squares algorithm operates on a range of values within a single order of magnitude (from -4 to $+8$ in our dataset), which does not involve the scale problems and leads to a more robust fit.

6. Conclusions

We present a method of classification of lacustrine icebergs within object-based image analysis paradigm. We propose Random Forest classification, context-based corrections and watershed cutting as a workflow to classify open-water icebergs. Our method correctly estimates the number of icebergs in the image and provides decent measurements of their cross-sectional area. The disappointing performance of ruleset-based classification of ice melange region demonstrates the need of a specialized approach to iceberg detection in cases of densely packed icebergs. An object-based inventory of icebergs can serve as a rich source of optical and geometrical information about the studied objects. High impact of reflectance information, as compared to elevation or image texture, on the result of classification gives hope that the workflow could be used with diverse datasets. Universality of the method is realized by its reliance on generic algorithms, which allows it to be implemented within various programs or programming languages.

We used our method to map icebergs floating in the lagoon surrounding the snout of San Quintín glacier, NPI. We show that log-normal distribution is the best approximation of iceberg area and volume frequencies in a population of lacustrine icebergs. This result suggests that there are mostly old icebergs in the lake and the iceberg decay may follow universal principles regardless of whether it occurs in freshwater or in a marine setting.

We found a power law relationship between iceberg area and volume in the form of $V = 7.42 \pm 1.14 \cdot S^{1.19 \pm 0.02}$. The values of the parameters are in line with an earlier publication. Nevertheless, our model comes from conversion of a linear trend fitted to the $\log_{10}$ of the iceberg volumes and waterline cross-sectional areas. This, as well as a limited size range of the data valid for fitting (only icebergs larger than 99 m²), leaves open a question of whether the power law is the best model of the scaling. More work with more precise elevation models would be needed to accurately constrain the model and extend it to smaller icebergs and growlers.

Author Contributions: J.P. wrote all computer code, performed most of analyses and wrote the majority of the manuscript. M.P. conceived the idea of the work and designed the study, provided data and contributed to manuscript writing and interpretation of results. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: This work was funded with FONDECYT INICIACIÓN grant #11170937 from CONICYT-Chile and was partially financed as a part of the statutory activity from the subvention of the Ministry of Science and Higher Education in Poland. CECs is a non-profit research organization funded by the Basal Finance program of CONICYT, among others. JP is supported by International Environmental Doctoral School associated with the Centre for Polar Studies at the University of Silesia in Katowice. We thank Johannes Reinthaler for preparing the WV-2 DEM.

Conflicts of Interest: The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript and in the decision to publish the results.

Appendix A

Table A1. Summary of processing steps, tools used for their execution and input data at each step.

Processing step	Tool	Input data	Flowchart
WV-2 imagery preprocessing	ENVI 5	WV-2 raw images	Figure 2
DREM creation	QGIS 3.10	WV-2 DEM	Figure 3
DREM uncertainty assessment	QGIS 3.10, SAGA, MATLAB	WV-2 DEM, TanDEM-X DEM	Figure 4
Image segmentation and classification	eCognition 9.3	WV-2 processed images, DREM	Figure 5
Watershed cutting	Python 3	Shapefile map of icebergs	Figure 6
Iceberg detection quality assessment	MATLAB	Cut shapefile map of icebergs	-
Iceberg properties extraction	eCognition 9.3	Cut shapefile map of icebergs	Figure 5
Statistical analysis	MATLAB	Table of iceberg properties	-

References

- Warren, C.; Benn, D.; Winchester, V.; Harrison, S. Buoyancy-driven lacustrine calving, Glaciar Nef, Chilean Patagonia. *J. Glaciol.* **2001**, *47*, 135–146. [\[CrossRef\]](#)
- Benn, D.I.; Warren, C.R.; Mottram, R.H. Calving processes and the dynamics of calving glaciers. *Earth-Sci. Rev.* **2007**, *82*, 143–179. [\[CrossRef\]](#)
- Trüssel, B.L.; Motyka, R.J.; Truffer, M.; Larsen, C.F. Rapid thinning of lake-calving Yakutat Glacier and the collapse of the Yakutat Icefield, southeast Alaska, USA. *J. Glaciol.* **2013**, *59*, 149–161. [\[CrossRef\]](#)
- Loriaux, T.; Casassa, G. Evolution of glacial lakes from the Northern Patagonia Icefield and terrestrial water storage in a sea-level rise context. *Glob. Planet. Chang.* **2013**, *102*, 33–40. [\[CrossRef\]](#)
- Schoolmeester, T. *The Andean Glacier and Water Atlas : The Impact of Glacier Retreat on Water Resources*; UNESCO GRID-Arenal: Paris, France; Arendal, Norway, 2018.
- Mazur, A.; Wählin, A.; Krężel, A. An object-based SAR image iceberg detection algorithm applied to the Amundsen Sea. *Remote Sens. Environ.* **2017**, *189*, 67–83. [\[CrossRef\]](#)
- Soldal, I.; Dierking, W.; Korosov, A.; Marino, A. Automatic Detection of Small Icebergs in Fast Ice Using Satellite Wide-Swath SAR Images. *Remote Sens.* **2019**, *11*, 806. [\[CrossRef\]](#)
- Neuhaus, S.U.; Tulaczyk, S.M.; Begeman, C.B. Spatiotemporal distributions of icebergs in a temperate fjord: Columbia Fjord, Alaska. *Cryosphere* **2019**, *13*, 1785–1799. [\[CrossRef\]](#)

9. Scheick, J.; Enderlin, E.M.; Hamilton, G. Semi-automated open water iceberg detection from Landsat applied to Disko Bay, West Greenland. *J. Glaciol.* **2019**, *65*, 468–480. [[CrossRef](#)]
10. Köhler, A.; Petlicki, M.; Lefeuvre, P.M.; Buscaino, G.; Nuth, C.; Weidle, C. Contribution of calving to frontal ablation quantified from seismic and hydroacoustic observations calibrated with lidar volume measurements. *Cryosphere* **2019**, *13*, 3117–3137. [[CrossRef](#)]
11. Minowa, M.; Podolskiy, E.A.; Sugiyama, S.; Sakakibara, D.; Skvarca, P. Glacier calving observed with time-lapse imagery and tsunami waves at Glaciar Perito Moreno, Patagonia. *J. Glaciol.* **2018**, *64*, 362–376. [[CrossRef](#)]
12. Mallalieu, J.; Carrivick, J.L.; Quincey, D.J.; Smith, M.W. Calving Seasonality Associated with Melt-Undercutting and Lake Ice Cover. *Geophys. Res. Lett.* **2020**, *47*. [[CrossRef](#)]
13. Dowdeswell, J.A.; Forsberg, C.F. The size and frequency of icebergs and bergy bits derived from tidewater glaciers in Kongsfjorden, northwest Spitsbergen. *Polar Res.* **1992**, *11*, 81–91. [[CrossRef](#)]
14. Tournadre, J.; Bouhier, N.; Girard-Ardhuin, F.; Rémy, F. Antarctic icebergs distributions 1992–2014. *J. Geophys. Res. Ocean.* **2016**, *121*, 327–349. [[CrossRef](#)]
15. Kirkham, J.D.; Rosser, N.J.; Wainwright, J.; Jones, E.C.V.; Dunning, S.A.; Lane, V.S.; Hawthorn, D.E.; Strzelecki, M.C.; Szczuciński, W. Drift-dependent changes in iceberg size-frequency distributions. *Sci. Rep.* **2017**, *7*, 1–10. [[CrossRef](#)]
16. Crocker, G. Size distributions of bergy bits and growlers calved from deteriorating icebergs. *Cold Reg. Sci. Technol.* **1993**, *22*, 113–119. [[CrossRef](#)]
17. Stern, A.A.; Adcroft, A.; Sergienko, O. The effects of Antarctic iceberg calving-size distribution in a global climate model. *J. Geophys. Res. Ocean.* **2016**, *121*, 5773–5788. [[CrossRef](#)]
18. Sulak, D.J.; Sutherland, D.A.; Enderlin, E.M.; Stearns, L.A.; Hamilton, G.S. Iceberg properties and distributions in three Greenlandic fjords using satellite imagery. *Ann. Glaciol.* **2017**, *58*, 92–106. [[CrossRef](#)]
19. Blaschke, T.; Hay, G.J.; Kelly, M.; Lang, S.; Hofmann, P.; Addink, E.; Feitosa, R.Q.; van der Meer, F.; van der Werff, H.; van Coillie, F.; et al. Geographic Object-Based Image Analysis—Towards a new paradigm. *ISPRS J. Photogramm. Remote Sens.* **2014**, *87*, 180–191. [[CrossRef](#)]
20. Rastner, P.; Bolch, T.; Notarnicola, C.; Paul, F. A Comparison of Pixel- and Object-Based Glacier Classification With Optical Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 853–862. [[CrossRef](#)]
21. Robson, B.A.; Nuth, C.; Dahl, S.O.; Hölbling, D.; Strozzi, T.; Nielsen, P.R. Automated classification of debris-covered glaciers combining optical, SAR and topographic data in an object-based environment. *Remote Sens. Environ.* **2015**, *170*, 372–387. [[CrossRef](#)]
22. Jawak, S.D.; Wankhede, S.F.; Luis, A.J. Exploration of Glacier Surface Facies Mapping Techniques Using Very High Resolution Worldview-2 Satellite Data. *Proceedings* **2018**, *2*, 339. [[CrossRef](#)]
23. Mitkari, K.V.; Arora, M.K.; Tiwari, R.K. Extraction of Glacial Lakes in Gangotri Glacier Using Object-Based Image Analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 5275–5283. [[CrossRef](#)]
24. McNabb, R.W.; Womble, J.N.; Prakash, A.; Gens, R.; Haselwimmer, C.E. Quantification and Analysis of Icebergs in a Tidewater Glacier Fjord Using an Object-Based Approach. *PLoS ONE* **2016**, *11*, e0164444. [[CrossRef](#)] [[PubMed](#)]
25. Aniya, M.; Enomoto, H. Glacier Variations and Their Causes in the Northern Patagonia Icefield, Chile, Since 1944. *Arct. Alp. Res.* **1986**, *18*, 307. [[CrossRef](#)]
26. Le Bris, R.; Paul, F.; Rastner, P.; Schaub, Y. *GLIMS Glacier Database*; National Snow and Ice Data Center: Boulder, CO, USA, 2018.
27. Harrison, S.; Warren, C.R.; Winchester, V.; Aniya, M. Onset of rapid calving and retreat of Glaciar San Quintin, Hielo Patagónico Norte, southern Chile. *Polar Geogr.* **2001**, *25*, 54–61. [[CrossRef](#)]
28. Aniya, M. Glacier variations of Hielo Patagónico Norte, Chile, over 70 years from 1945 to 2015. *Bull. Glaciol. Res.* **2017**, *35*, 19–38. [[CrossRef](#)]
29. Di Tullio, M.; Nocchi, F.; Camplani, A.; Emanuelli, N.; Nascetti, A.; Crespi, M. Copernicus Big Data And Google Earth Engine For Glacier Surface Velocity Field Monitoring: Feasibility Demonstration on San Rafael and San Quintin Glaciers. *ISPRS-Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *XLII-3*, 289–294. [[CrossRef](#)]
30. Dixon, L.; Ambinakudige, S. Remote Sensing Study of Glacial Change in the Northern Patagonian Icefield. *Adv. Remote Sens.* **2015**, *4*, 270–279. [[CrossRef](#)]

31. DigitalGlobe. Base Product Series FAQ. 2016. Available online: <https://www.c-agg.org/wp-content/uploads/DigitalGlobe-Base-Product-FAQ.pdf> (accessed on 9 April 2020).
32. Eisank, C.; Rieg, L.; Klug, C.; Kleindienst, H.; Sailer, R. Semi-Global Matching of Pléiades Tri-Stereo Imagery to Generate Detailed Digital Topography for High-Alpine Regions. *GL_Forum* **2015**, *1*, 168–177. [[CrossRef](#)]
33. Rizzoli, P.; Martone, M.; Gonzalez, C.; Wecklich, C.; Tridon, D.B.; Bräutigam, B.; Bachmann, M.; Schulze, D.; Fritz, T.; Huber, M.; et al. Generation and performance assessment of the global TanDEM-X digital elevation model. *ISPRS J. Photogramm. Remote Sens.* **2017**, *132*, 119–139. [[CrossRef](#)]
34. Wessel, B.; Huber, M.; Wohlfart, C.; Marschalk, U.; Kosmann, D.; Roth, A. Accuracy assessment of the global TanDEM-X Digital Elevation Model with GPS data. *ISPRS J. Photogramm. Remote Sens.* **2018**, *139*, 171–182. [[CrossRef](#)]
35. DigitalGlobe. WorldView-2 Data Sheet and Core Imagery Product Guide. 2016. Available online: <https://dg-cms-uploads-production.s3.amazonaws.com/uploads/document/file/98/WorldView2-DS-WV2-rev2.pdf> (accessed on 9 April 2020).
36. Hall, D.K.; Riggs, G.A. Normalized-Difference Snow Index (NDSI). In *Encyclopedia of Snow, Ice and Glaciers*; Singh, V.P., Singh, P., Haritashya, U.K., Eds.; Springer: Amsterdam, The Netherlands, 2011; pp. 779–780.
37. Haala, N.; Rothermel, M. Dense Multi-Stereo Matching for High Quality Digital Elevation Models. *Photogramm. Fernerkund. Geoinf.* **2012**, *2012*, 331–343. [[CrossRef](#)]
38. Podgórski, J.; Kinnard, C.; Petlicki, M.; Urrutia, R. Performance Assessment of TanDEM-X DEM for Mountain Glacier Elevation Change Detection. *Remote Sens.* **2019**, *11*, 187. [[CrossRef](#)]
39. Rolstad, C.; Haug, T.; Denby, B. Spatially integrated geodetic glacier mass balance and its uncertainty based on geostatistical analysis: Application to the western Svartisen ice cap, Norway. *J. Glaciol.* **2009**, *55*, 666–680. [[CrossRef](#)]
40. Trimble Geospatial. Trimble eCognition Suite for Windows Operating System. 2017. Available online: http://www.infoserve.co.jp/doc/Trimble/v9_3/ReleaseNotes.pdf (accessed on 29 May 2020).
41. Rouse, J. W., J.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the Great Plains with ERTS. In *Goddard Space Flight Center 3d ERTS-1 Symp. Sect. A*; Freden, S.C., Mercanti, E.P., Becker, M.A., Eds.; NASA: Washington, DC, USA, 1974; Volume 1, pp. 309–317.
42. Bradski, G. The OpenCV Library. *Dr. Dobbs's J. Softw. Tools* **2000**. Available online: <https://www.drdobbs.com/open-source/the-opencv-library/184404319> (accessed on 12 May 2020).
43. OpenCV. OpenCV 2.4.13.7 Documentation. Random Trees. 2019. Available online: https://docs.opencv.org/2.4/modules/ml/doc/random_trees.html (accessed on 24 April 2020).
44. Trimble Geospatial. *Trimble Documentation eCognition Developer 9.3 Reference Book*; Trimble Germany GmbH: Munich, Germany, 2018.
45. Beucher, S. Use of watersheds in contour detection. In *Proceedings of the International Workshop on Image Processing (CCETT)*, Marne-la-Vallée, France, 26–28 March 1979.
46. Mordvintsev, A.; Abid, K. Image Segmentation with Watershed Algorithm. OpenCV 2.4.13.7 Documentation. 2013. Available online: https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_imgproc/py_watershed/py_watershed.html (accessed on 03 December 2019).
47. Scheick, J.; Enderlin, E.; Miller, E.; Hamilton, G. First-Order Estimates of Coastal Bathymetry in Ilulissat and Naajarsuit Fjords, Greenland, from Remotely Sensed Iceberg Observations. *Remote Sens.* **2019**, *11*, 935. [[CrossRef](#)] [[PubMed](#)]
48. Wagner, T.J.; Stern, A.A.; Dell, R.W.; Eisenman, I. On the representation of capsizing in iceberg models. *Ocean Model.* **2017**, *117*, 88–96. [[CrossRef](#)]
49. Alstott, J.; Bullmore, E.; Plenz, D. powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions. *PLoS ONE* **2014**, *9*, e85777. [[CrossRef](#)]
50. Clauset, A.; Shalizi, C.R.; Newman, M.E.J. Power-Law Distributions in Empirical Data. *SIAM Rev.* **2009**, *51*, 661–703. [[CrossRef](#)]
51. Bunting, P.; Clewley, D.; Lucas, R.M.; Gillingham, S. The Remote Sensing and GIS Software Library (RSGISLib). *Comput. Geosci.* **2014**, *62*, 216–226. [[CrossRef](#)]

52. Truffer, M.; Motyka, R.J. Where glaciers meet water: Subaqueous melt and its relevance to glaciers in various settings. *Rev. Geophys.* **2016**, *54*, 220–239. [[CrossRef](#)]
53. Larsen, C.F.; Motyka, R.J.; Arendt, A.A.; Echelmeyer, K.A.; Geissler, P.E. Glacier changes in southeast Alaska and northwest British Columbia and contribution to sea level rise. *J. Geophys. Res.* **2007**, *112*. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).