REFINING ICESAT-2 BATHYMETRIC PHOTONS USING OPTICAL SATELLITE IMAGES

KAIQI JIANG

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Master of Science in Geography

Department of Geography, Environment and Geomatics
Faculty of Arts
University of Ottawa

© Kaiqi Jiang, Ottawa, Canada, 2026

Abstract

The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2), launched by NASA in 2018, is equipped with the Advanced Topographic Laser Altimeter System (ATLAS), a photon counting laser altimeter. ATL24 is a global data product that classifies seafloor photons from the geolocated photon returns from ATLAS, and these seafloor photons can be used for bathymetry. Published evaluations report encouraging overall accuracy of ATL24. We obtained ATL24 for six Canadian coastal sites spanning turbid and Arctic waters and found much poorer quality: the RMSE of raw ATL24 depths against NONNA10, which is the Canadian Hydrographic Service's 10 m non-navigational bathymetric grid and serves as the reference throughout this study, ranged from 2.9 to 24.6 m across sites. We developed a refinement workflow that uses Sentinel-2 optical imagery as an independent check on the photons. Within each ICESat-2 track and Sentinel-2 scene, the workflow compares photon elevation with a logarithmic blue to green band ratio, removes outliers, and retains photon groups with sufficient vertical range and a Spearman correlation below an empirically chosen fixed threshold of −0.75; a photon enters the final dataset only when its track passes in at least two scenes. Refinement reduced the per site RMSE range from 2.9–24.6 m to 0.4–1.2 m. To test whether this improvement carries into satellite derived bathymetry (SDB), six SDB methods were calibrated at each of the six sites, once with raw ATL24 photons, and once with refined photons, and the resulting depth maps were evaluated against NONNA10 over the same evaluation pixels, which were capped at the 95th percentile of the calibration photons' depth. Refined calibration produced the lower RMSE in 29 of 36 site and method combinations, with a median RMSE reduction of 0.6 m. A threshold sensitivity analysis then repeated the Spearman decision over ten thresholds and five SDB methods with cross validation across the six sites and found no threshold that transferred better than this fixed value. This study provides evidence that optical imagery can be used to filter the ATL24 product and remove erroneously classified seafloor photons, which in turn improves its use as SDB calibration data in turbid and Arctic Canadian waters.

Statement on the Use of Artificial Intelligence

This thesis was prepared with assistance from generative AI, disclosed here in accordance with the Artificial Intelligence Policy for Graduate Students of the Department of Geography, Environment and Geomatics (November 2024).

Tools used

Two tools from Anthropic were used for most of the work: Claude (a conversational AI assistant, accessed through its web interface) and Claude Code (an AI coding agent that runs on the author's computer and can read, write, and execute code locally). In the early stages of the project, ChatGPT and Codex (OpenAI) were also used for code editing; the later work was carried out with the Anthropic tools.

How and why the tools were used

Code and data analysis: The author decided what every program needed to do and specified its required functions, inputs, and outputs in written task instructions; Claude Code implemented these requirements, writing, modifying, and executing most of the Python code used in this thesis, including the photon screening workflow, the SDB calibration experiments, the sensitivity analysis, and the scripts that produced the figures and tables. ChatGPT and Codex (OpenAI) assisted with code editing in the early stages of the project. All analyses ran locally on the author's real data; the AI tools were not used to generate, simulate, or substitute any data or results, and every output was reviewed by the author before use.

Writing: Chapter drafts were produced iteratively: the author provided the content requirements, results, and structure; Claude drafted text; the author reviewed, corrected, and frequently rewrote the drafts, and the supervisor's comments drove further revisions. Claude was also used for language editing and terminology consistency. The author read and approved every sentence in the final thesis.

Literature work: Claude Code was used to extract and quote passages from locally stored copies of the cited publications, so that statements in the literature review could be checked against the original sources. A citation audit was carried out in which every citation in the thesis was traced back to the cited document and its supporting passages were read; citations

that could not be supported were corrected or removed.

Verification and accountability

Since AI-generated text and code can contain errors, a verification process was used throughout: numerical values in the thesis were required to trace back to registered analysis outputs; AI-drafted text was checked against the underlying data and sources; errors found on either side (the author's or the AI's) were logged and corrected. AI tools contributed no ideas, data, or conclusions that the author has not independently verified and adopted as their own. The author takes full responsibility for the entire content of this thesis, including any remaining errors.

Table of Contents

1 Chapter 1: Introduction 1

1.1 Shallow Water Bathymetry 1

1.2 Satellite Derived Bathymetry 2

1.3 ICESat-2 Bathymetry 3

1.4 Thesis Objective 5

1.5 Thesis Format 5

2 Chapter 2: Refining ICESat-2 Bathymetric Photons Using Optical Satellite Images 7

2.1 Introduction 7

2.2 Study Areas and Data 8

2.2.1 Study areas 8

2.2.2 ATL24 photons 11

2.2.3 Sentinel-2 imagery 11

2.2.4 Vertical datum and reference data 12

2.3 Methods 14

2.3.1 Overview 14

2.3.2 Vertical referencing 14

2.3.3 Scene preparation and the band ratio 16

2.3.4 Track screening within each scene 16

2.3.5 Two-scene rule 17

2.3.6 Photon level evaluation 17

2.3.7 SDB calibration experiment 18

2.3.8 Sensitivity to the Spearman threshold and other fixed values 19

2.3.9 Visual assessment 21

2.3.10 Contribution of each stage in the screening workflow 21

2.3.11 Seasonal analysis 22
2.4 Results 23
2.4.1 Photon level vertical accuracy 23
2.4.2 SDB calibrated with raw and refined photons 27
2.4.3 Sensitivity to the Spearman threshold and other fixed values 30
2.4.4 Results of the contribution of each stage in the screening workflow 34
2.4.5 Seasonal analysis results 37
2.5 Discussion 39
2.6 Conclusion 42
3 Chapter 3: Conclusion 44
4 References 46
5 Appendix A: Parameter sensitivity tests 54
6 Appendix B: Supplementary threshold sensitivity analysis of the multiband log linear model 56
7 Appendix C: Per-site RMSE at each Spearman threshold 58
8 Appendix D: Per-site SDB RMSE at each screening stage 61

1 Chapter 1: Introduction

1.1 Shallow Water Bathymetry

Bathymetry describes the depth and shape of the seafloor. In coastal waters, this information supports both practical decisions and scientific work. Vessels need reliable depths to avoid shoals and other hazards, coastal engineers use bathymetry to plan channels, harbors, and shoreline infrastructure, while flood and wave models depend on the form of the nearshore seafloor. Habitat studies, sediment research, marine spatial planning, and assessments of shoreline change also require depth information. Better knowledge of the seafloor therefore contributes to the wider goal of understanding and managing the ocean sustainably (Wölfl et al., 2019).

Most modern hydrographic surveys use shipborne acoustic instruments or airborne bathymetric lidar. A single-beam echo sounder measures a profile beneath a vessel, whereas a multibeam system covers a wider swath. Airborne lidar can map shallow water from above when the water is sufficiently clear and weather and sea conditions are suitable. Both approaches can produce accurate measurements, but neither is easy to apply everywhere. A ship must travel through the area being surveyed, which is difficult in remote waters, close to hazards, or where the vessel may ground. Airborne surveys cover larger areas more quickly, but they require a dedicated campaign and remain expensive. Coastal bathymetry therefore involves a continuing trade off among cost, coverage, access, and safety (Cesbron et al., 2021).

These constraints are particularly important in shallow coastal zones. Such areas can change rapidly, yet they are often difficult to reach and are precisely where fine spatial detail is most useful. Large parts of the shallow coastal ocean remain poorly surveyed by modern standards (Cesbron et al., 2021). The problem is evident along northern and remote Canadian coasts, where the amount, age, and spatial coverage of available depth data differ greatly from one location to another (International Hydrographic Organization, 2026; Knudby, 2025). The consequences are concrete. Between 2000 and 2018 the Transportation Safety Board of Canada received 74 reports of vessel groundings or bottom contact in the Canadian Arctic, and in at least seven of these cases the grounding hazard was visible in satellite imagery yet absent from the chart in use (Knudby, 2025).

Satellite observations provide a useful complement to conventional surveys. They cannot replace hydrographic measurements where certified navigation products or very high vertical accuracy are required, but they can extend regional coverage, support reconnaissance, reveal spatial patterns, and help identify areas that need more detailed surveying. Optical satellite imagery has been used for bathymetric mapping for several decades, and the resulting methods range from empirical regressions to semi-analytical, physics-based, machine learning, and hybrid approaches (Ashphaq et al., 2021).

1.2 Satellite Derived Bathymetry

Satellite derived bathymetry (SDB) estimates water depth from sunlight that penetrates the water column, reflects from the seafloor, and returns to a satellite sensor. Its physical basis is that light is attenuated approximately exponentially with depth, at rates that differ among wavelengths, so the spectral composition of the water leaving signal carries depth information wherever the seafloor is visible (Philpot, 1989). The same signal is also influenced by bottom reflectance and the optical properties of the water, and in shallow water these effects can be difficult to distinguish from the effect of depth (Maritorena et al., 1994). Most SDB methods address this ambiguity in one of two ways. Empirical methods fit a statistical relation between image values and known depths, either from log transformed bands (Lyzenga, 1978) or from the ratio of two log transformed bands, which suppresses much of the bottom type variation (Stumpf et al., 2003), and they remain widely used because they are simple to implement and require only the imagery and a set of calibration depths. Physics based methods instead invert a radiative transfer model and can retrieve depth without depth calibration data (Lee et al., 1998; Klonowski et al., 2007; Jung et al., 2025), at the cost of greater model complexity and sensitivity to the optical assumptions.

Over the past decade, many studies have replaced the empirical regression with machine learning models, including random forests, support vector machines, nearest neighbor regression, and neural networks (Sagawa et al., 2019; Kibele & Shears, 2016; Mudiyanselage et al., 2022; Wu et al., 2021; Surisetty et al., 2021), and more recently with deep learning architectures (Lumban-Gaol et al., 2022; Wan & Ma, 2021; Xie et al., 2024; Z. Chen et al., 2026). However, comparisons made with the same imagery and calibration data, show that the advantage of these flexible models is conditional rather than universal. In water shallower than approximately 15 m, classical and machine learning methods perform similarly, and simple linear or ratio models are sometimes slightly better; the flexible models pull ahead mainly in the deeper part of the retrievable range (Misra et al., 2018; Zhou et al., 2023; Cheng et al., 2023), and when the maximum depth is restricted, the differences among models shrink

(Eugenio et al., 2022). Deep learning comparisons show a similar pattern: one direct assessment found only a marginal improvement over classical methods, whereas larger gains at depth came from augmenting the training data and stratifying the model by optical depth (Z. Chen et al., 2026). Transfer between sites also remains difficult. A network that reached submeter accuracy where it was trained produced errors several times larger at other sites, with no way to predict in advance which pretrained model would work (Lumban-Gaol et al., 2022).

Across these comparisons, SDB accuracy is decided less by the model than by the calibration data it is given. Overfitting to small calibration sets, the inability of nonparametric models to extrapolate beyond the calibrated depth range, and the need for calibration depths that span the full range being mapped appear repeatedly, across methods and decades (Lyzenga, 1985; Kibele & Shears, 2016; Sagawa et al., 2019; Zhou et al., 2023), and a synthesis of machine learning practice in remote sensing reached the same conclusion, that a large and accurate training set matters more than the model applied to it (Maxwell et al., 2018). The practical bottleneck of SDB is therefore not the regression: it is obtaining calibration depths that are accurate, well distributed in space and depth, and available at the site of interest.

That requirement is hardest to meet exactly where SDB is most attractive, in remote and poorly surveyed waters. This has motivated a search for calibration sources that do not require a field campaign, including transferring models between sites (Lumban-Gaol et al., 2022) and calibrating on depths measured by satellite laser altimetry. Among the alternatives that have been compared, altimetric bathymetry from ICESat-2 offers the best combination of accuracy and availability, limited chiefly by its sparse spatial coverage and by the time difference between the altimeter pass and the image (Cao et al., 2024). ICESat-2 has consequently become a widely used source of calibration depths for SDB in waters without modern surveys.

1.3 ICESat-2 Bathymetry

ICESat-2 carries the Advanced Topographic Laser Altimeter System (ATLAS), a photon counting lidar that emits green (532 nm) laser pulses along six ground tracks and records the arrival time of individual returned photons (Markus et al., 2017). ATLAS records both laser returns and photons from solar background light, so the raw observations form a geolocated point cloud in which signal must be separated from noise (Neumann et al., 2019). Green light can penetrate clear water, and some photons return from the seafloor. Under favorable

conditions, ICESat-2 has been shown to measure depths of up to tens of meters, which is what makes it usable as a source of SDB calibration data (Parrish et al., 2019).

Two processing problems stand between the photon cloud and usable depths. The first is refraction: light bends and slows at the air water interface, so uncorrected subsurface photons appear too deep. A correction assuming a flat water surface was established early, reduces the depth bias to the decimeter level, and has been adopted widely (Parrish et al., 2019). Later studies refined the same framework by accounting for sea surface slope (Ma et al., 2020), for the dependence of the refractive index on temperature and salinity (Ranndal et al., 2021), and most recently by reconstructing the instantaneous wavy surface itself instead of assuming a plane (Zhang et al., 2022). Although these refinements differ at the decimeter scale, a refraction correction is required before the photon elevations can be used as reference water depth.

The second problem is extracting seafloor photons from noise, and it has produced a large family of methods. Density based clustering adapted from DBSCAN (Density-Based Spatial Clustering of Applications with Noise) identifies signal as locally dense regions of the along-track point cloud (Ma et al., 2020; Xie et al., 2021), and later variants elongate or reorient the search neighborhood to follow the along-track geometry of the seafloor (Y. Chen et al., 2021; Leng et al., 2022; Wang et al., 2024). In a controlled comparison of elliptical against circular neighborhoods of equal area, the overall RMSE differed by only 0.02 m, but the elliptical search retained more than twice as many photons in the 15–25 m depth range, where the circular search lost the deep seafloor entirely (Leng et al., 2022). The shape of the neighborhood thus matters less for overall accuracy than for how much of the deeper seafloor photons survive the filtering. Adaptive elliptical clustering with automatically selected parameters has also reduced RMSE by more than 1 m relative to standard DBSCAN (Xie et al., 2023). Learned approaches replace the hand-built neighborhood with features trained from labeled photon clouds and generalize across more environments (Lin & Knudby, 2023), and the most recent assessment identifies the fixed geometry of the search neighborhood, whether circular, elliptical, or a preset receptive field, as the shared limitation of this whole family (Qin et al., 2026). All of these methods, hand-built or learned, rely on structure within the photon cloud itself. How well their parameters generalize across environments is largely unknown, and supervised approaches would likely require retraining at new sites (Jung et al., 2025).

A parallel line of research combines ICESat-2 with optical imagery. ICESat-2 depths have been used to calibrate Sentinel-2 SDB (Hsu et al., 2021; Ma et al., 2020), and ICESat-2 photon extraction has been combined with Sentinel-2 SDB in single workflows (Xie et al., 2021). More recent studies couple the two data sources iteratively, use optical slope information to guide photon extraction, or combine information from multiple ICESat-2 passes (Hu et al., 2025; Li et al., 2025; Cao et al., 2025). Active and passive observations thus carry partly independent information about the same shallow water, and each can be used to evaluate or refine the other.

In 2025 the extraction problem was moved upstream into a standard product. ATL24 applies a global classification to the ATL03 photon cloud and distributes refraction corrected, orthometrically referenced bathymetric photons directly (Magruder et al., 2025b). The product evaluation reported encouraging accuracy at eight validation sites, while also documenting a persistent error in which photons at or near the water surface are misclassified as seafloor (Parrish et al., 2025). However, published assessments cover a narrow range of environments: the validation sites lie on the U.S. East Coast, Gulf Coast, and Caribbean, independent evaluations so far are limited to temperate and low to mid latitude sites (Dandabathula et al., 2025; Kristoffersen, 2026), and a review of the field notes that most ICESat-2 bathymetric studies concern small sites in clear water, so that performance in other conditions is essentially unknown (Jung et al., 2025). Turbidity limits the depth to which the seafloor can be detected (Kristoffersen, 2026), and high latitude waters are strongly affected by false seafloor classifications (Parrish et al., 2025). How ATL24 performs in turbid and Arctic shallow waters, precisely the settings where alternative calibration data are scarcest, has not been tested. This thesis asks that question, and asks whether refining ATL24 improves the SDB calibrated from it.

1.4 Thesis Objective

The objective of this study is: (1) to develop a workflow that uses optical imagery to refine ATL24 bathymetric photons, (2) to apply and test this workflow at six shallow Canadian study sites spanning turbid and Arctic regions, (3) to evaluate whether the improved quality of the ATL24 photons translates into improved quality of the SDB that uses them as calibration data.

1.5 Thesis Format

This thesis follows a manuscript style format. Chapter 1 reviews shallow water bathymetry,

SDB, and ICESat-2 Bathymetry. Chapter 2 is a standalone manuscript, Refining ICESat-2 Bathymetric Photons Using Optical Satellite Images, presenting the study areas, data, methods, results, and discussion. Chapter 3 summarizes the findings and their limitations.

2 Chapter 2: Refining ICESat-2 Bathymetric Photons Using Optical Satellite Images

2.1 Introduction

Accurate nearshore bathymetry supports navigation safety, coastal engineering, habitat assessment, and hazard response, yet large parts of the shallow coastal ocean remain poorly surveyed because ship and airborne surveys are expensive and difficult to repeat in remote areas (Wölfl et al., 2019; Cesbron et al., 2021). Satellite derived bathymetry (SDB) can map shallow water at low cost over large areas. Empirical SDB methods fit a statistical relation between image values and known depths; they are widely used because they are simple to implement and require only the imagery and a set of calibration depths, and their accuracy depends on the quality, quantity, and distribution of those calibration depths more than on the model fitted to them (Ashphaq et al., 2021; Maxwell et al., 2018; Zhou et al., 2023). Physics based methods can retrieve depth from a radiative transfer model without local calibration data (Jung et al., 2025), at the cost of greater model complexity and sensitivity to their optical assumptions. In remote and poorly charted waters, where SDB would be most useful, the calibration depths that empirical methods require have commonly been missing.

ICESat-2 changed this situation by providing a satellite source of calibration depths. Its photon counting laser altimeter records individual photon returns, some of which come from the seafloor in sufficiently clear and shallow water (Markus et al., 2017; Parrish et al., 2019). A series of methods has been developed to separate these seafloor photons from noise, based on local photon density, adaptive search neighborhoods, or learned point cloud features, and refraction corrections at the decimeter level are well established (Ma et al., 2020; Leng et al., 2022; Lin & Knudby, 2023; Jung et al., 2025). ICESat-2 depths have been used directly to calibrate Sentinel-2 SDB (Hsu et al., 2021). ICESat-2 photon extraction has been combined with Sentinel-2 imagery in single workflows (Xie et al., 2021), and recent work couples the two more closely, through iterative fusion, optical slope guidance, and multiple pass calibration (Hu et al., 2025; Li et al., 2025; Cao et al., 2025).

The ATL24 product, released in 2025, applies a global classification to the ICESat-2 photon cloud and distributes refraction-corrected bathymetric photons as a standard dataset (Magruder et al., 2025b), removing the need for each user to implement photon extraction. The product evaluation reported encouraging accuracy while documenting a persistent error in which photons at or near the water surface are misclassified as seafloor (Parrish et al.,

2025).

Assessments of ATL24 published to date cover a narrow range of environments. The reference data of the product evaluation were drawn from the U.S. East Coast, Gulf Coast, and Caribbean (Parrish et al., 2025), and independent evaluations have so far been limited to temperate and low to mid latitude sites (Dandabathula et al., 2025; Kristoffersen, 2026). ICESat-2 bathymetric studies more generally have concentrated on clear waters (Jung et al., 2025); turbidity restricts the depth to which the seafloor can be detected (Kristoffersen, 2026), and high latitude waters are strongly affected by false seafloor classifications (Parrish et al., 2025). The performance of ATL24 in turbid and Arctic shallow waters therefore remains largely untested, although these are precisely the settings where alternative calibration data are scarcest.

The objective of this study is threefold: first, to develop a workflow that uses optical imagery to refine ATL24 bathymetric photons; second, to apply and test this workflow at six shallow Canadian study sites spanning turbid and Arctic regions; and third, to evaluate whether the improved quality of the ATL24 photons translates into improved quality of the SDB that uses them as calibration data.

2.2 Study Areas and Data

2.2.1 Study areas

The study focused on six Canadian coastal sites: Royal Geographical Society Islands, Sandspit, Montague, Cambridge Bay, Havre aux Maisons, and Igloolik. Six additional sites — Kitimat, Eureka, Resolute, Sanirajak, Kugluktuk, and Grise Fiord — were also investigated, but provided too few usable ATL24 photons and too little NONNA10 reference coverage for meaningful analysis; the outcome at these sites is reported in Section 2.4.1. The sites are distributed across western, eastern, and Arctic coastal regions. These sites were selected to cover, as broadly as possible, the range of shallow water environments along the Canadian coast, including different water conditions, photon densities, image availability, and reference coverage.



Figure 1. Locations of the six study areas in Canada.

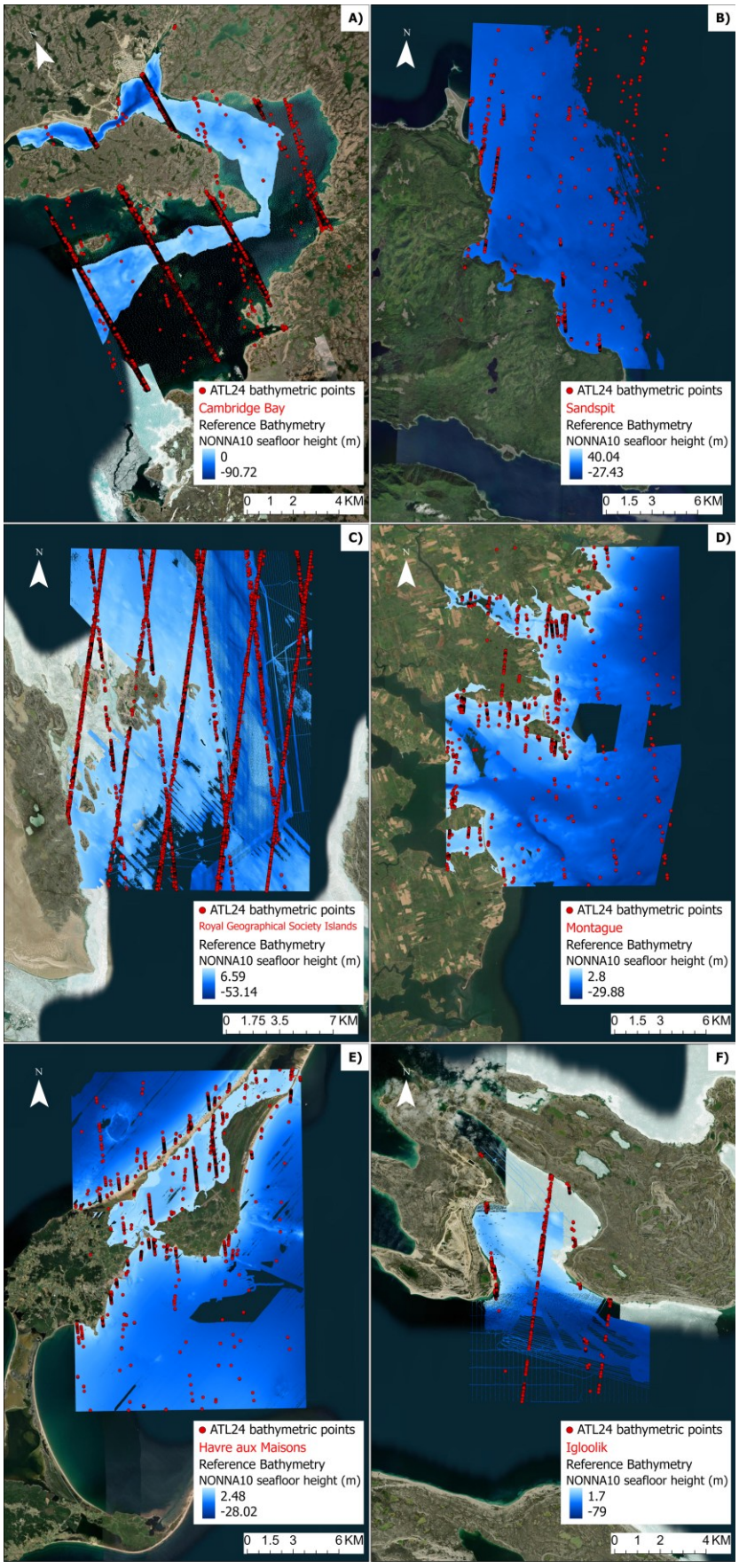
A)
ATL24 bathymetric points
Cambridge Bay
Reference Bathymetry
NONNA10 seafloor height (m)
0
-90.72
0 1 2 4 KM
B)
ATL24 bathymetric points
Sandspit
Reference Bathymetry
NONNA10 seafloor height (m)
40.04
-27.43
0 1.5 3 6 KM
C)
ATL24 bathymetric points
Royal Geographical Society Islands
Reference Bathymetry
NONNA10 seafloor height (m)
6.59
-53.14
0 1.75 3.5 7 KM
D)
ATL24 bathymetric points
Montague
Reference Bathymetry
NONNA10 seafloor height (m)
2.8
-29.88
0 1.5 3 6 KM
E)
ATL24 bathymetric points
Havre aux Maisons
Reference Bathymetry
NONNA10 seafloor height (m)
2.48
-28.02
0 1.5 3 6 KM
F)
ATL24 bathymetric points
Igloolik
Reference Bathymetry
NONNA10 seafloor height (m)
1.7
-79
0 1 2 4 KM

Figure 2. Detailed maps of the six study areas. Red points show ATL24 bathymetric photons, and blue raster show the available NONNA10 reference bathymetry.

2.2.2 ATL24 photons

ATL24 photons were obtained through the SlideRule framework's atl24x workflow (Shean et al., 2023). The request covered 1 January 2018 to 1 January 2024, retained the bathymetry class, excluded low confidence output, and used the full, non-compact output. The ATL24 data used in this thesis are Version 1 (Magruder et al., 2025a), accessed in June and July 2026; Version 1 has since been superseded by Version 2. Two height representations were preserved for each photon: orthometric height was used for optical filtering, and ellipsoidal height was retained for the later conversion of ATL24 heights to the chart datum used by the NONNA10 reference data. The analysis therefore began with classified bathymetric photons rather than repeating the full ATL03 extraction process. Table 1 summarizes, for each of the six study sites, the number of tracks, raw ATL24 photons, refined photons (hereafter refined photons: the photons retained by the optical screening workflow described in Section 2.3), and refined photons matched with NONNA10 reference depths.

Table 1. The number of ICESat-2 tracks and ATL24 photons at the six study sites.

Site	Tracks	Raw ATL24 photons	Refined photons	Refined photons matched with NONNA10
Royal Geographical Society Islands	93	26,250	7,511	5,512
Sandspit	56	1,890	414	414
Montague	80	5,377	945	945
Cambridge Bay	77	43,748	6,073	290
Havre aux Maisons	63	4,818	1,083	1,083
Igloolik	28	3,338	1,503	829

2.2.3 Sentinel-2 imagery

Sentinel-2A and Sentinel-2B imagery (Drusch et al., 2012) was searched for July to September of 2018–2023, with reported cloud cover limited to 10%. Where local availability allowed, at least five scenes were downloaded for each required Military Grid Reference

System (MGRS) tile and each satellite. When a study area crossed more than one MGRS tile, a scene was used only when all required tiles were available; the tiles were then mosaicked into a single scene, with all bands aligned to the 10 m grid of Band 2. The number of usable scenes therefore varied among sites (Table 2).

Table 2. Sentinel-2 acquisitions downloaded and scenes used in the analysis at the six study sites.

Study site	Required MGRS tile(s)	Downloaded acquisitions	Scenes used in the analysis
Royal Geographical Society Islands	14WMB	13	13
Sandspit	09UUU, 09UUV	18	6
Montague	20TNS	12	12
Cambridge Bay	13WDS, 13WES	32	2
Havre aux Maisons	20TNT	16	16
Igloolik	17WMS	14	14
Total		105	63

Table 2 lists the Sentinel-2 acquisitions downloaded and the scenes used at each site. A downloaded acquisition is one satellite pass on one day. At the four sites covered by a single MGRS tile, every downloaded acquisition entered the analysis. At the two sites that span two tiles, Sandspit and Cambridge Bay, tiles were mosaicked into one scene only when both tiles had a valid product from the same satellite pass, matched by sensor, date, and acquisition time. Most passes at these two sites produced a product for only one of the two tiles, because a single swath does not always cover both tiles and cloud screening was applied per tile. At Sandspit, 18 acquisitions therefore yielded 6 usable scenes, and at Cambridge Bay, 32 acquisitions yielded only 2.

2.2.4 Vertical datum and reference data

The CANEAST, CANNORTH, and CANWEST products used in this study were obtained directly from the Canadian Hydrographic Service (CHS) upon request. These products provide “Chart Datum height relative to ITRF2014(2010)” (Canadian Hydrographic Service, 2026b), describing the spatially varying vertical relationship between the local Chart Datum surface and the ITRF2014(2010) reference surface (International Terrestrial Reference Frame

2014, epoch 2010.0). They do not provide bathymetric depths. Instead, they provide the vertical conversion information required to transform ATL24 ellipsoidal heights to heights relative to the local Chart Datum. The appropriate CANEAST, CANNORTH, or CANWEST product was assigned to each study site according to its location, including a site specific subfile where required.

After the vertical conversion, NONNA10 was used as the reference bathymetry. NONNA10 is the Canadian Hydrographic Service's non-navigational bathymetric data product, distributed as gridded tiles at a nominal resolution of 10 m. It consolidates digital bathymetric sources managed by the CHS (Canadian Hydrographic Service, 2026a); the underlying surveys differ in age and quality, and the product is not intended for navigation. In this study it served as the reference against which both raw and refined ATL24 photons were evaluated, and the six study sites were covered by 20 NONNA10 tiles. The horizontal coordinates of each photon were transformed to the coordinate reference system of the corresponding NONNA10 raster, and the value of the raster cell containing the photon location was sampled directly. If the cell contained nodata or a non-finite value, the photon remained unmatched; photons outside valid NONNA10 coverage were not assigned reference values from nearby locations.

2.3 Methods

2.3.1 Overview

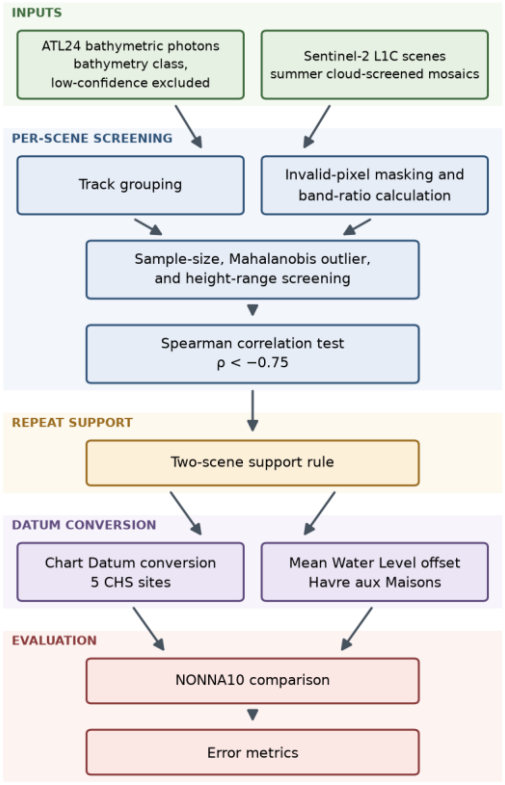

Figure 3. Overview of the workflow

Figure 3 outlines the workflow. The same rules and parameter values were applied at every site. ATL24 photons were grouped by track, Sentinel-2 scenes were prepared and masked, and each track was tested separately in each scene by comparing its photon depths with the band ratio along the track. A track could therefore pass in some scenes and fail in others. Photons were kept only from tracks that passed in at least two scenes, and a photon supported by several scenes was counted once. The retained photons were converted to each site's vertical datum and compared with NONNA10 as described in Sections 2.3.2 and 2.2.4.

2.3.2 Vertical referencing

At the five study sites covered by CHS conversion data, the photon heights were converted to Chart Datum for comparison with NONNA10, whose depths are given relative to Chart Datum. The CHS conversion values used in this study are the Chart Datum height relative to the ITRF2014(2010) ellipsoid, so the conversion began from the ATL24 ellipsoidal height. The conversion was applied after the optical screening had been completed. It therefore did

not affect the screening decisions and influenced only the reference matching and error calculations.

Because the CHS conversion values vary spatially, the conversion value at the location of photon i, denoted by s_i, was estimated by inverse distance weighting from the four nearest CHS data points, with a power of two:

$$s_i = \frac{\sum_{j=1}^{4} w_{ij}\ s_j}{\sum_{j=1}^{4} w_{ij}}$$

$$w_{ij} = \frac{1}{d_{ij}^2}$$

Here, s_j is the conversion value provided at CHS point j, and d_{ij} is the distance between ATL24 photon i and CHS point j. If the nearest CHS point was more than 10 km from the photon, the conversion value was recorded as invalid, and the photon did not enter the NONNA10 error calculation.

The three CHS products were applied with a conversion operation:

$$h_i^{\mathrm{CD}} = h_i^{\mathrm{ITRF2014}} - s_i$$

In this equation, h_i^{ITRF2014} is the ellipsoidal height of ATL24 photon i; s_i is the CHS conversion value interpolated at the photon location; and h_i^{CD} is the height of photon relative to the local Chart Datum. The sign convention follows the CHS metadata. For CANNORTH and CANWEST, the delivered value is the Chart Datum height relative to ITRF2014(2010), and the equation applies directly. The CANEAST file gives the same quantity with the opposite sign, the height of ITRF2014(2010) relative to Chart Datum, so its values were applied with the sign reversed and the same equation holds in all three regions.

Havre aux Maisons lies outside the valid coverage of the CANEAST, CANNORTH, and CANWEST products used in this study, so the conversion described above could not be applied there. A separate vertical correction was used instead: the Mean Water Level (MWL) at the CHS Havre-aux-Maisons tidal station (station 01976) is 1.02 m above Chart Datum (Canadian Hydrographic Service, 2026c). Because the MWL is a good approximation of the geoid at the station, and the ATL24 orthometric height is a height above the geoid, adding this value to the orthometric height places the photons on Chart Datum:

$$h_i^{\mathrm{Maison}} = h_i^{\mathrm{ortho}} + 1.02\ \mathrm{m}$$

Havre aux Maisons was treated with a constant Mean Water Level offset rather than a CHS separation surface, so its vertical correction carries a different systematic error. Its absolute errors are therefore reported separately. Combining them with the errors of the five CHS corrected sites would mix two different error sources.

2.3.3 Scene preparation and the band ratio

Each scene was processed separately. At the sites that span two tiles, tiles were mosaicked into one scene only when both came from the same satellite pass, matched by date and acquisition time; scenes from different dates were never mosaicked together. A screening mask removed invalid and nodata pixels, together with any pixel whose Band 11 value exceeded 1,500, Band 8A exceeded 1,700, or Band 10 exceeded 1,025. These bands were chosen because clear water is dark in the shortwave infrared and cirrus bands, so pixels bright in Band 11, Band 8A, or Band 10 are almost always cloud, ice, or land rather than water. The thresholds were set well above typical water values to remove such pixels while keeping the mask permissive, since photons on poor pixels can still be rejected by the later screening steps. Photons located outside the valid scene area or on masked pixels were excluded from that scene.

Bands 2 and 3 were sampled at each remaining photon position. Reflectance was obtained from the digital numbers, following the Sentinel-2 Level-1C convention with the radiometric offset (European Space Agency, 2026), as

$$\rho_b = \frac{DN_b - 1000}{10000}$$

and the optical variable was the logarithmic blue to green ratio of Stumpf et al. (2003):

$$R = \frac{\ln(1000\,\rho_{B02})}{\ln(1000\,\rho_{B03})}$$

The ratio was calculated only when both digital numbers exceeded 1,000, so that both reflectances were positive and the logarithms were defined.

2.3.4 Track screening within each scene

Within each scene, the photons on unmasked pixels with a valid band ratio were grouped by track. A group entered screening only if it contained at least ten photons. Each photon in a group has two values, its band ratio and its orthometric height, and photons that lie far from the rest of the group in these two values are likely noise. Such outliers were removed with the

Mahalanobis distance, which measures how far each photon lies from the center of the group while accounting for the spread and correlation of the two values; the group covariance matrix was inverted with a pseudo inverse so that groups with almost no variation could still be handled. Photons whose squared distance exceeded the 99th percentile of the chi-square distribution with two degrees of freedom were discarded. A group failed the screening if fewer than ten photons remained after the Mahalanobis outlier removal. Panel (c) of Figure 4 in Section 2.3.9 shows an example, with the removed outliers marked.

Within each group, the heights of the photons that survived the Mahalanobis outlier removal then had to span more than 3 m:

$$\max(h) - \min(h) > 3 \text{ m}$$

This condition excluded nearly level segments, where the height ordering mostly reflects noise and a rank correlation means little. Finally, the Spearman rank correlation between the band ratio and the photon height was computed, and the group was accepted only when the correlation was −0.75 or stronger. Spearman correlation was chosen because the expected optical relation is monotonic but not necessarily linear. The expected relation is negative because the band ratio increases with depth while the photon height decreases.

2.3.5 Two-scene rule

Photons from all groups that passed the screening in Section 2.3.4 were combined across scenes, and a photon was retained in the final dataset only when its track passed in at least two scenes. This rule tests whether the relation between the photon heights and the band ratio repeats in independent images. A photon whose track passed in several scenes was counted once. The heights of the retained photons were converted to the local chart datum and matched with NONNA10 following Sections 2.3.2 and 2.2.4.

2.3.6 Photon level evaluation

Photons entered the error analysis when their heights had been converted to Chart Datum and they were matched with a valid NONNA10 value at their location. The error for matched photon i was defined as:

$$e_i = h_i^{\text{ATL24, corrected}} - z_i^{\text{NONNA10}}$$

where $h_i^{\mathrm{ATL24,\ corrected}}$ is the ATL24 height after the applicable vertical correction and z_i^{NONNA10} is the NONNA10 reference value at the same location. The error keeps its sign: it is positive when the photon is shallower than the reference and negative when it is deeper.

Three metrics were then computed over all matched photons of a site:

bias,

$$\mathrm{Bias} = \frac{1}{n}\sum_{i=1}^{n} e_i$$

mean absolute error,

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |e_i|$$

and root mean squared error,

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}$$

all in meters. Raw metrics used every matched ATL24 photon, before any screening. Refined metrics used the matched photons retained by the optical screening and the two-scene rule. Photon level comparisons use all matched photons, with no depth limit.

2.3.7 SDB calibration experiment

To test whether photon refinement improves the bathymetry calibrated from the photons, six SDB methods were calibrated twice at each of the six sites: once with the raw ATL24 photons and once with the refined photons. The raw calibration started from all ATL24 photons at the site, without any optical screening; the refined calibration started from the photons retained by the optical screening and the two-scene rule. Both photon sets then went through the same preparation for calibration, which removed photons outside the scene window, photons on masked pixels, and photons without a valid value in each of the four bands used as SDB inputs (Bands 2, 3, 4, and 8). Calibration at each site used one Sentinel-2 scene fixed in advance, chosen as the scene containing the most refined photons. Depths were expressed as positive downward values on the local chart datum.

The six methods were as follows. The band ratio model of Stumpf et al. (2003) was fitted by ordinary least squares; the blue to green or blue to red ratio was selected separately for each site and each calibration set, taking the ratio with the higher calibration R^2. The multiband log linear model after Lyzenga (1978) used Bands 2, 3, and 4; its deep water reference reflectance was set, per band, to the mean minus two standard deviations of the reflectance over the deepest 10% of the valid NONNA10 area, chosen dark enough that shallower pixels stay above it. Random forest regression used 500 trees with a minimum leaf size of 5. Histogram gradient boosting used 500 iterations with a learning rate of 0.05. K nearest neighbor regression used k = 5 with equal weights. Support vector regression used a radial basis function kernel, tuned over a 36 combination grid of kernel width, tolerance, and regularization on a random 80/20 split, then refitted on the full calibration set. The machine learning methods used the reflectance of Bands 2, 3, 4, and 8 as features, without standardization. All random steps, namely the random forest and gradient boosting training and the 80/20 split for the support vector regression, used fixed seeds, so the results are repeatable.

The resulting depth maps were evaluated against NONNA10 over a common set of pixels: valid NONNA10 cells within the area where the SDB model produced predictions, limited to a per-site depth cap. The cap was the 95th percentile of the calibration photon depths, computed separately for the raw and the refined calibration, and the smaller of the two caps was applied to both, so that the two maps were compared over exactly the same pixels. This depth cap applies only to the SDB map evaluation; the photon level evaluation in Section 2.3.6 uses all matched photons, with no depth limit.

2.3.8 Sensitivity to the Spearman threshold and other fixed values

The sensitivity of the results to the Spearman threshold was evaluated on the SDB maps: changing the threshold changes which photons are retained, and the question is whether the bathymetry calibrated from those photons becomes better or worse. Ten thresholds from −0.95 to −0.50, at intervals of 0.05, were tested with all other filtering rules unchanged, at the six sites of the SDB calibration experiment in Section 2.3.7. For each site, all ten thresholds used the same Sentinel-2 scene as in Section 2.3.7, and the resulting depth maps were compared with NONNA10 over the evaluation pixels of Section 2.3.7, including its per-site depth cap. These pixels were taken from Section 2.3.7 as they were and used unchanged for all ten thresholds, so that differences in RMSE reflect the threshold and not a change of scene or evaluation area.

The threshold sensitivity analysis used five SDB methods: the Stumpf band ratio, random forest, gradient boosting, nearest neighbor regression, and support vector regression. The multiband log linear model (Lyzenga, 1978) was included in the SDB calibration experiment of Section 2.3.7 but excluded from this analysis. Unlike the other five methods, it predicts depth by linear extrapolation and is not bounded by the calibrated range: pixels brighter than any calibration photon are extrapolated past zero to negative depths, and such pixels are excluded before scoring. In the calibration experiment this does not affect the result, because raw and refined calibration are compared on the same pixels within each site and method. In this analysis, however, the excluded area changes with the threshold. The model was run at the same sites and thresholds as the five methods, and it produced a depth map in every case. But at Sandspit, 28–39% of the map came out as negative depths, and at Havre aux Maisons 6–13%, with the percentage changing from one threshold to the next. A change in its RMSE between two thresholds could therefore reflect a change in the evaluation area rather than the effect of the threshold itself, so the model was left out (Appendix B). A site and threshold combination was considered trainable only when at least 30 calibration photons from at least two tracks remained after filtering.

Whether a threshold selected at some sites works as well at others was tested with a cross validation. Each site was excluded in turn, and the threshold with the lowest mean RMSE across the five remaining sites was selected. When several thresholds had mean RMSEs within 0.05 m of the best, they were treated as tied, and the tie went to the threshold that retained more photons. The selected threshold was then evaluated at the excluded site against the fixed value of −0.75. Two rules were set in advance for replacing the fixed threshold, one about stability and one about benefit. First, the six selections had to point to a common value: at least four of the six selected thresholds had to fall within 0.05 of their median. Second, the change had to be worth making: on the excluded sites, which took no part in the selection, the selected thresholds had to reduce the mean RMSE by at least 0.05 m compared with −0.75. Only if both held would the fixed threshold be replaced. The procedure was specified with support vector regression as the deciding method, and the other four SDB methods repeated the same procedure to check that the conclusion did not depend on the method.

For the other fixed values, the same kind of sensitivity analysis was run at three sites, Montague, Igloolik, and Cambridge Bay. Four kinds of changes were tested one at a time, with all other parameter values kept at their original values: removing the two-scene rule, changing the mask thresholds, changing the minimum number of photons required in a group, and changing the elevation range used in the optical screening. After each change, the RMSE of the retained photons against the NONNA10 reference, as defined in Section 2.3.6, was recomputed.

2.3.9 Visual assessment

Tracks were inspected with a three panel figure showing, for one track, the along-track ATL24 and NONNA10 profiles, the comparison of the ATL24 depths with NONNA10, and the relation between the band ratio and the photon heights in the scene with the most negative Spearman correlation. Seeing these side by side shows what a track looks like along its path, how well its depths agree with the reference, and how the band ratio and the photon heights relate in the scene used for the screening. Visual inspection did not change any screening decisions.

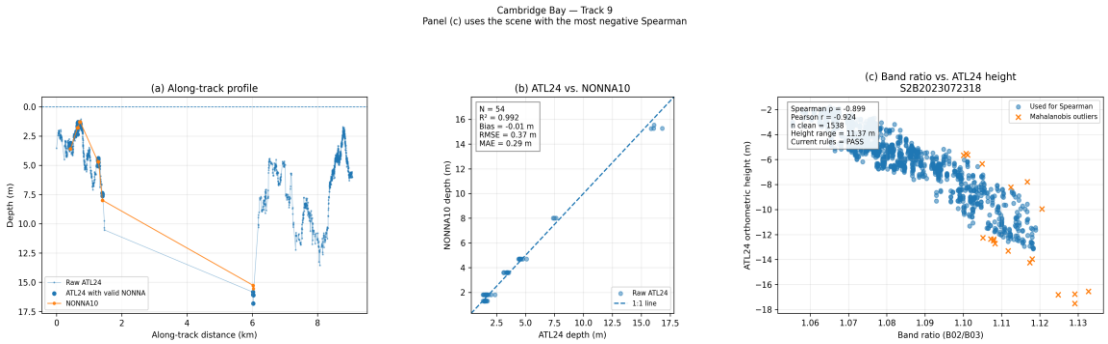

Figure 4. Example of the visual assessment for one track (Cambridge Bay, Track 9): the along-track ATL24 and NONNA10 profiles (a), the ATL24 to NONNA10 comparison (b), and the band ratio relation in the scene with the most negative Spearman correlation (c).

2.3.10 Contribution of each stage in the screening workflow

The screening workflow follows a fixed sequence of rules, and the accuracy reported in Section 2.4.1 reflects the combined effect of all these rules. To isolate the contribution of each individual rule, we decomposed the process into seven cumulative "stages". Stage 0 consists of all raw ATL24 photons for a given site. Each following stage keeps the photons that survive one additional rule, applied in the order of the screening workflow: Stage 1 keeps the photons that fall on a valid, unmasked pixel in at least one scene; Stage 2 additionally requires the photon's track to have at least ten such photons in that scene; Stage 3 keeps the photons that survive the Mahalanobis outlier removal; Stage 4 requires the track to span more than 3 m in height; Stage 5 requires the track to pass the Spearman threshold of −0.75; and Stage 6 requires all the above conditions to be met in at least two scenes. Stage 6 is therefore exactly the refined photon set of Section 2.3.5, and each stage is a subset of the one before it.

For each stage at each site, we calculated three metrics. The first metric is the number of

retained photons. The second metric is the RMSE of the retained photons relative to the NONNA10 data (see Section 2.3.6 for the calculation method). The third metric is the RMSE of the SDB maps calibrated with the photons of each stage. The five SDB methods of Section 2.3.8 were recalibrated with each stage's photons, reusing the scenes, hyperparameters, random seeds, and evaluation pixels of the SDB experiment, so that the only difference between the stages is the calibration photons.

In addition, the photons removed at each step and matched with NONNA10 were split into accurate (within 0.5 m of the reference, the value that the maximum allowable total vertical uncertainty of an IHO Order 1 survey approaches in shallow water; International Hydrographic Organization, 2024) and inaccurate, to count how many accurate photons each rule removed.

2.3.11 Seasonal analysis

ATL24 photon data for the study sites were collected throughout the year, including months when the sites were covered by sea ice. To test whether the photon errors depend on the season of acquisition, each photon was assigned the acquisition date of its corresponding ICESat-2 track. Based on the sea ice climatology for each site, the months of the year were categorized into three types: open water months, ice months, and breakup and freeze-up months. At the three Arctic sites, sea ice typically breaks up in July and freezes up in October, as reported for the Canadian Arctic Archipelago as a whole, including Victoria Strait at Royal Geographical Society Islands (Howell et al., 2013; Canadian Ice Service, 2021b), and confirmed locally for Foxe Basin at Igloolik (Gough, 2025) and for Dease Strait at Cambridge Bay (Braybrook et al., 2026 ; Butterworth et al., 2025), so the open water months are August and September, July and October are the breakup and freeze-up months, and the remaining eight months are ice months. At Montague and Havre aux Maisons, in the Gulf of St. Lawrence, the open water months are May to November, December and April are the breakup and freeze-up months, and January to March are ice months (Canadian Ice Service, 2021a). Sandspit is ice free in all months, as sea ice does not occur on the Pacific coast of Canada (Derksen et al., 2019).

The seasonal analysis was designed to answer three questions. The first question is whether the accuracy of the raw photons depends on the season. To answer it, the raw ATL24 photons of each site were compared with the raw photons from open water months only, using the metrics of Section 2.3.6, and the errors of the photons from the breakup and freeze-up months

and from the ice months were examined separately to show where the large errors come from. The second question is whether selecting photons by month could replace the optical screening. To answer it, the open water selection was compared with the refined photons of Section 2.3.5, and a third selection that also includes the breakup and freeze-up months was added as a sensitivity variant, because the result of a month-based selection depends on how these boundary months are categorized. The third question is which seasons the refined photons come from. To answer it, the refined photons from each month type were counted and their errors computed.

2.4 Results

2.4.1 Photon level vertical accuracy

Among the 3,699 track and scene evaluations, where one evaluation is one track tested in one scene, 209 passed all filtering conditions. Fifty-nine tracks passed in at least one scene, and 42 of them passed in at least two, so their photons entered the refined dataset. The refined dataset contained 17,529 photons, of which 9,073 were matched with a valid NONNA10 value at their location.

At all six sites the refined RMSE was lower than the raw RMSE (Figure 5). At Royal Geographical Society Islands the RMSE fell from 14.78 m in the raw set to 0.71 m in the refined set, at Sandspit from 5.36 m to 0.35 m, at Montague from 2.91 m to 0.38 m, at Igloolik from 19.39 m to 0.98 m, and at Cambridge Bay from 24.64 m to 0.65 m. At Havre aux Maisons, evaluated separately under its Mean Water Level correction, the RMSE fell from 3.35 m to 1.20 m.

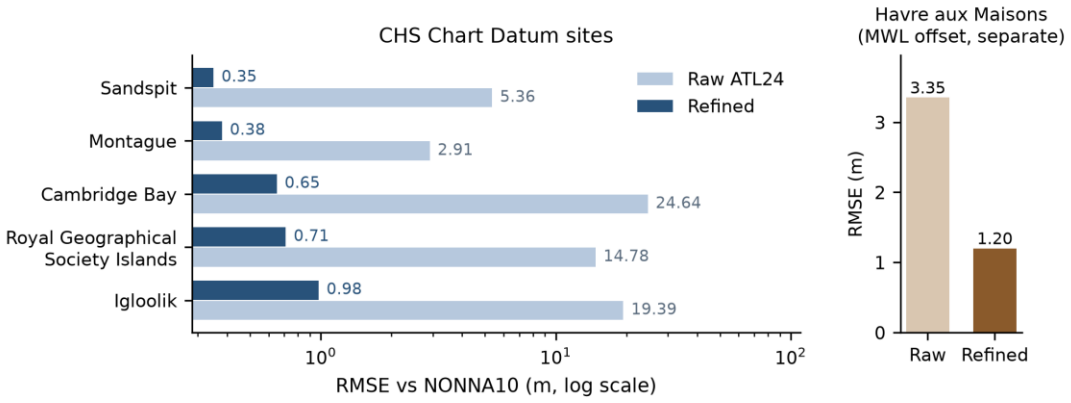

Figure 5. Site level RMSE against NONNA10 for the raw and refined ATL24 photon sets.

The scatter of raw and refined photons against NONNA10 shows where this improvement comes from (Figure 6). At the four sites shown, the raw sets contain a dense vertical cluster of photons whose heights approach zero while the reference is much deeper. These photons are sea surface returns incorrectly classified as seafloor in the ATL24 product. The refined sets lie along the one-to-one line. At Cambridge Bay, where this cluster holds most of the raw photons, refinement reduced the RMSE to about one fortieth while retaining 6,073 photons.

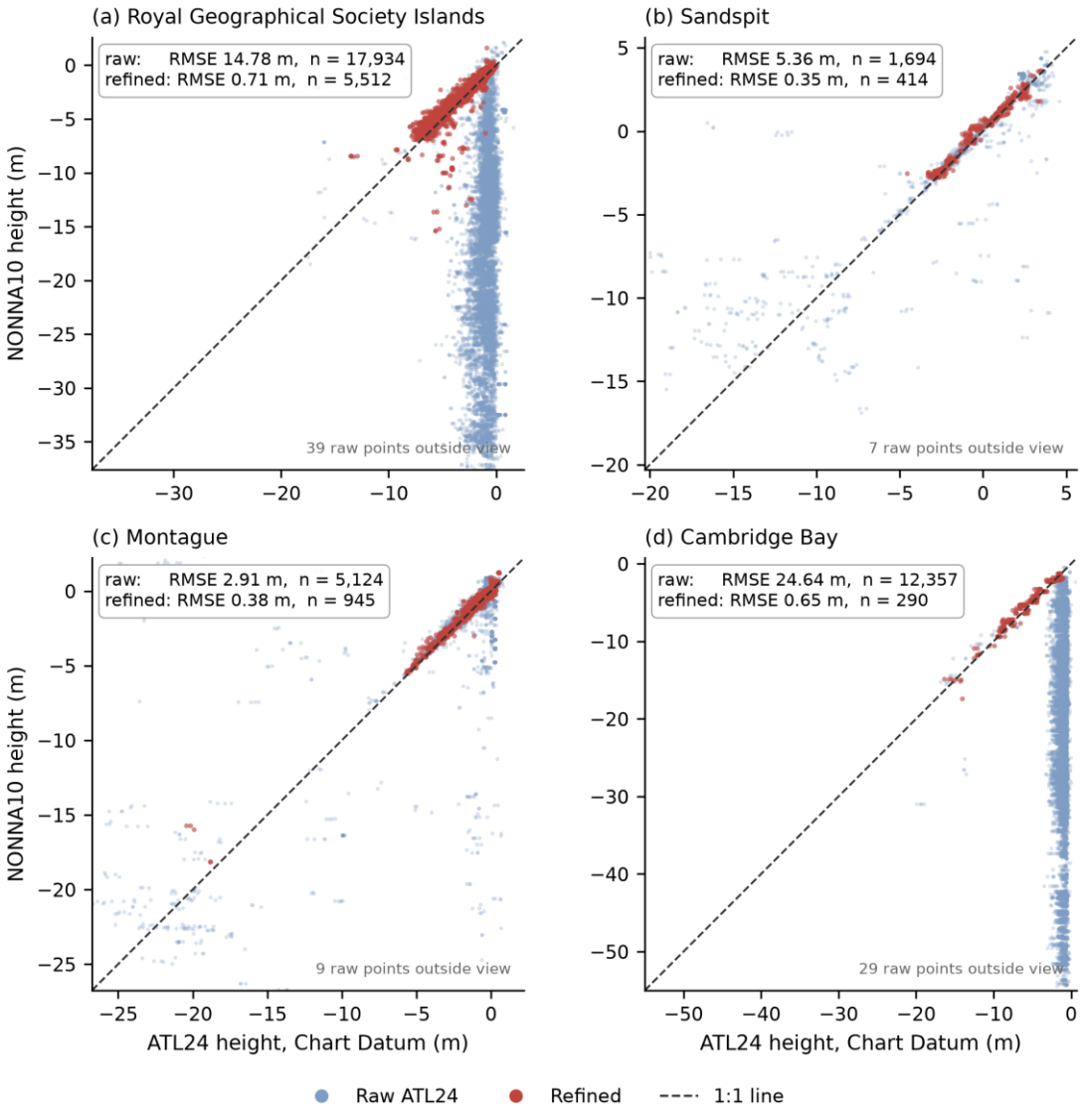

Figure 6. Raw and refined ATL24 photons against NONNA10 at four sites: (a) Royal Geographical Society Islands, (b) Sandspit, (c) Montague, (d) Cambridge Bay.

The number of retained photons and the number of retained photons matched with a reference value were two different things. All 414 refined photons at Sandspit and all 945 at Montague had a valid NONNA10 value, whereas at Cambridge Bay only 290 of the 6,073 retained photons, or 4.8%, were matched with a reference value, so the refined error there describes a small part of the retained set.

The same workflow was applied at six further sites. At Kugluktuk and Grise Fiord, which began with 3,057 and 299 raw photons, no photon passed the screening. At Sanirajak, 122 photons passed, but none lay within the NONNA10 coverage. At Kitimat, Eureka, and Resolute, 217, 69, and 202 photons passed, of which only 23, 5, and 8 lay within the NONNA10 coverage; these few matches disagreed with the reference by 112.44, 33.42, and 36.21 m RMSE. For comparison, the smallest number of matched refined photons among the six analyzed sites is 290. Therefore, these sites were included only as initial sites and were not included in the following tests and analyses.

Whether a track passed could also differ among scenes of the same site. At Havre aux Maisons, 15 of 63 tracks passed in some scenes and failed in others, at Montague 17 of 80, and at Igloolik 5 of 28. At Cambridge Bay every evaluated track passed in every scene. Figure 7 shows four tracks, each a different case. One passed with close reference agreement. One agreed with the reference but failed, because its correlation never reached the threshold in any scene. One passed on the twelve photons that remained after the outlier removal, yet disagreed with the reference by almost ten meters. One passed in eight of its thirteen scene evaluations.

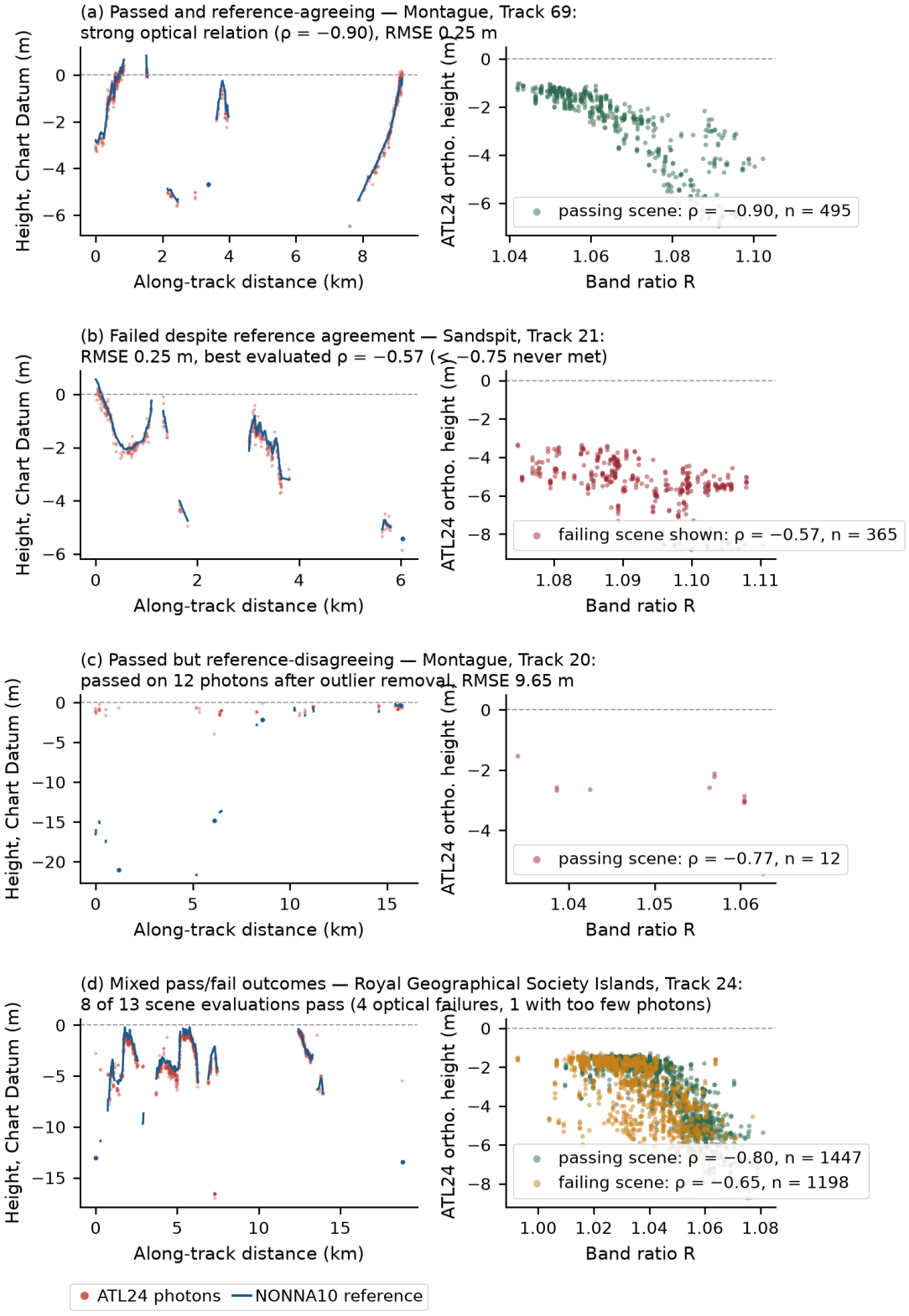

Figure 7. Four tracks, one per row, each a different case: (a) passed with close reference agreement (Montague, Track 69), (b) failed although it agreed with the reference (Sandspit, Track 21), (c) passed on the twelve photons that remained after the outlier removal, yet disagreed with the reference

(Montague, Track 20), (d) passed in eight of thirteen scene evaluations (Royal Geographical Society Islands, Track 24). Left: along-track profile. Right: band ratio against photon height in one scene.

2.4.2 SDB calibrated with raw and refined photons

The SDB calibration experiment described in Section 2.3.7 tested whether refinement improves the bathymetry calibrated from the photons. Refined calibration produced the lower RMSE in 29 of the 36 site and method combinations, with a median RMSE reduction of 0.6 m across the 36 combinations (Figure 8).

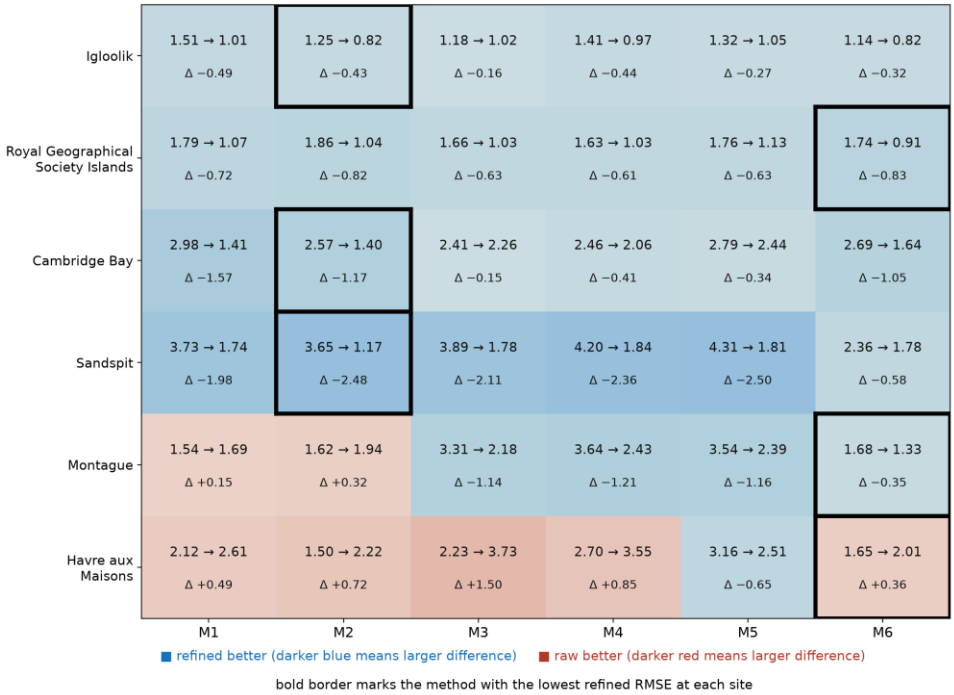

Figure 8. RMSE of the SDB maps calibrated with raw and with refined photons for every site and method combination. Each cell gives the raw and refined RMSE in meters and their difference. Blue cells are combinations where the refined RMSE is lower, red cells where the raw RMSE is lower, and the bold border marks the method with the lowest refined RMSE at each site. Methods: M1 Stumpf band ratio, M2 multiband log linear, M3 random forest, M4 gradient boosting, M5 nearest neighbor regression, M6 support vector regression.

The exceptions were concentrated at Havre aux Maisons. Across the six methods its mean RMSE rose from 2.23 m with raw calibration to 2.77 m with refined calibration, and only the nearest neighbor regression performed better with refined photons. The remaining two

combinations in which raw calibration was better were at Montague, with the band ratio and multiband log linear models.

Figure 9 shows the depth maps at the six sites, each shown with the SDB method that produced the lowest refined RMSE there. The maps show all pixels where the model produced predictions, while the RMSE values reported in Figure 8 are computed on the evaluation pixels of Section 2.3.7 only. The per-site depth cap of the evaluation pixels is the smaller of the 95th percentile calibration depths of the raw and refined sets.

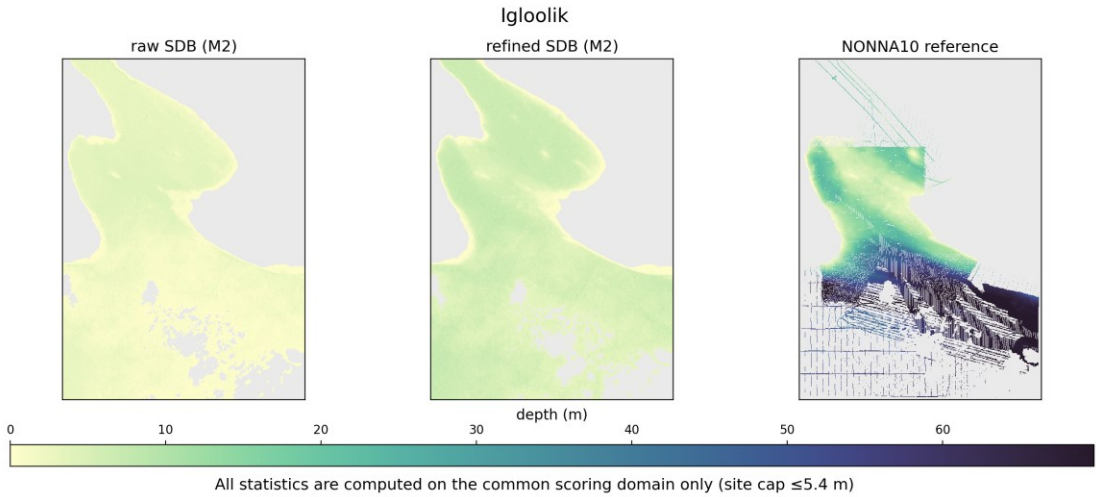

Figure 9. Raw calibrated SDB, refined calibrated SDB, and the NONNA10 reference at the six sites: Igloolik, Royal Geographical Society Islands, Cambridge Bay, Sandspit, Montague, Havre aux Maisons. The method is above each map, and the per-site depth cap is beneath each map.

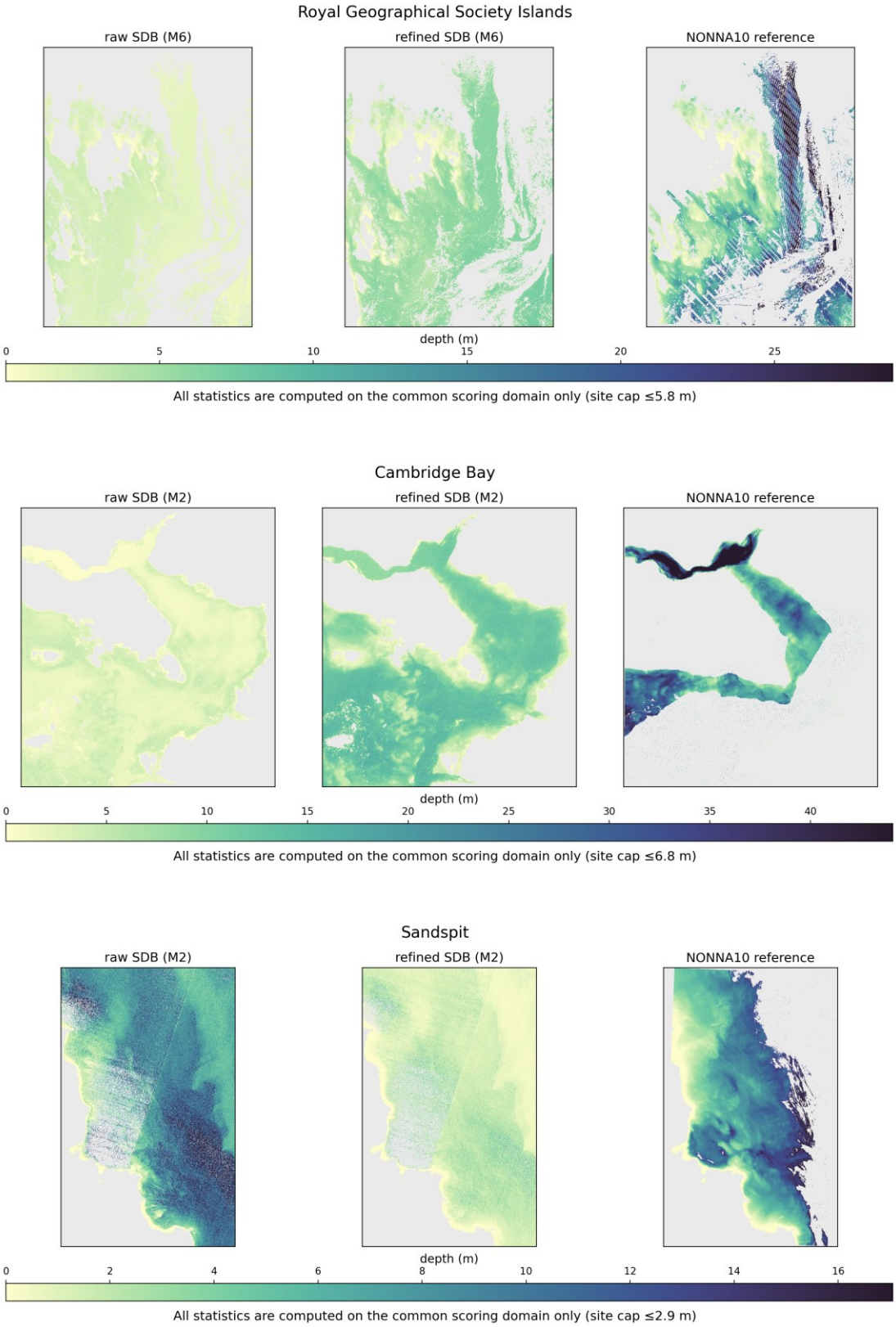

Figure 9 (continued).

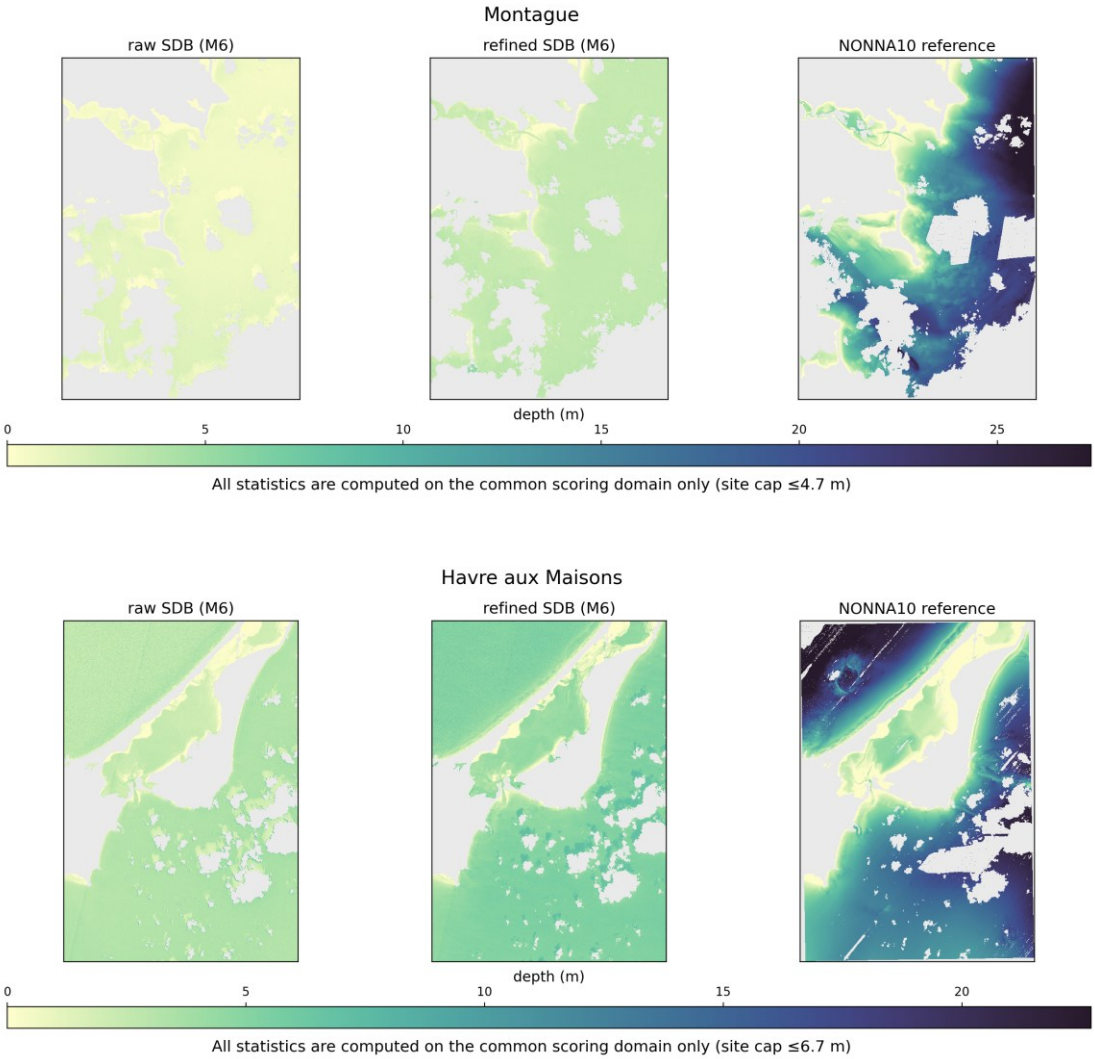

Figure 9 (continued).

2.4.3 Sensitivity to the Spearman threshold and other fixed values

Among the 60 site and threshold combinations, 48 met the training criterion described in Section 2.3.8 and 12 did not. All 12 occurred at thresholds of −0.85 and stricter, where fewer than two tracks retained photons in every case. The RMSE of every method at every site and threshold is listed in Appendix C. Figure 10 shows the six site mean RMSE of the five methods over the seven thresholds from −0.80 to −0.50, at which every site could be trained. Support vector regression produced the lowest mean RMSE at every threshold, stayed nearly level across the seven thresholds, and was 0.074 m lower at −0.50 than at −0.80. For the Stumpf band ratio, random forest, gradient boosting, and nearest neighbor regression, the mean RMSE was 0.32 to 0.53 m higher at −0.50 than at −0.80, and most of the increase occurred between −0.70 and −0.65.

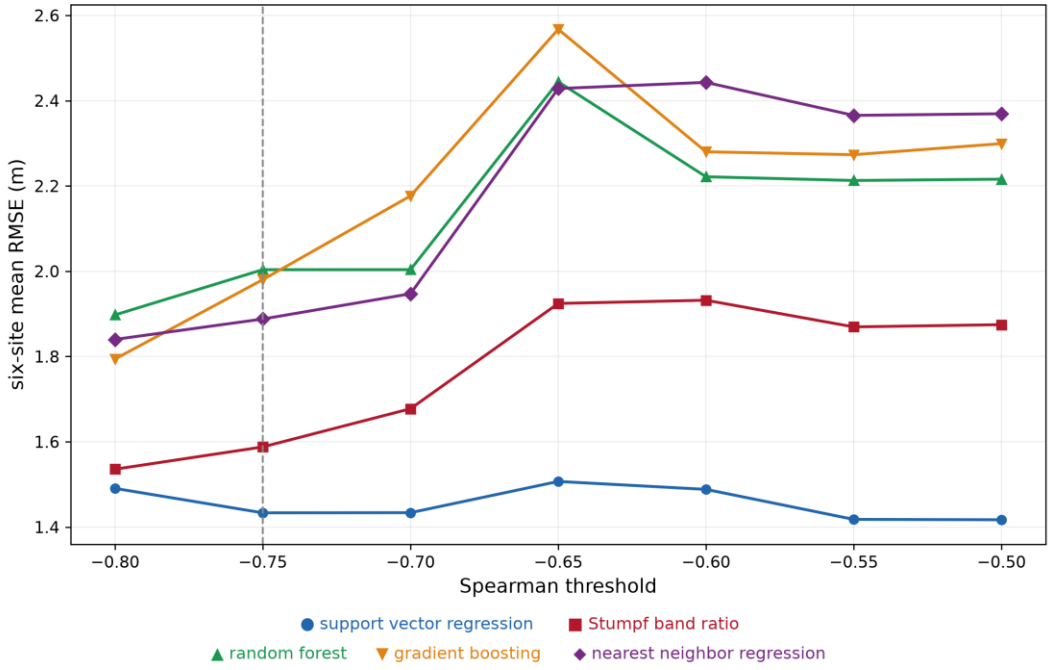

Figure 10. Mean RMSE of the five SDB methods across the six sites, at the seven Spearman thresholds at which all six sites could be trained. The dashed line marks the threshold of −0.75.

For the Stumpf band ratio, random forest, gradient boosting, and nearest neighbor regression, the six site mean RMSE was lowest at −0.80, the strictest of the seven thresholds, as shown in Figure 10. However, a threshold that gives the lowest mean RMSE does not necessarily give the lowest RMSE at each site. Figure 11 therefore shows the RMSE of the five methods at each site separately. Among the seven thresholds at which all six sites could be trained, no threshold gave the lowest RMSE at all six sites for any method. At Montague the RMSE of the four methods was 1.65 to 3.31 m higher at the threshold of −0.50 than at the threshold of −0.80, while at Igloolik, Royal Geographical Society Islands, and Sandspit the RMSE of every method changed by less than 0.15 m over the same threshold range. Cambridge Bay moved in the opposite direction, and all five methods had a lower RMSE at the threshold of −0.50 than at −0.80. The magnitude and direction of the change therefore differed from site to site. Since the six sites were averaged with equal weight, the mean RMSE in Figure 10 was influenced more by sites with larger changes.

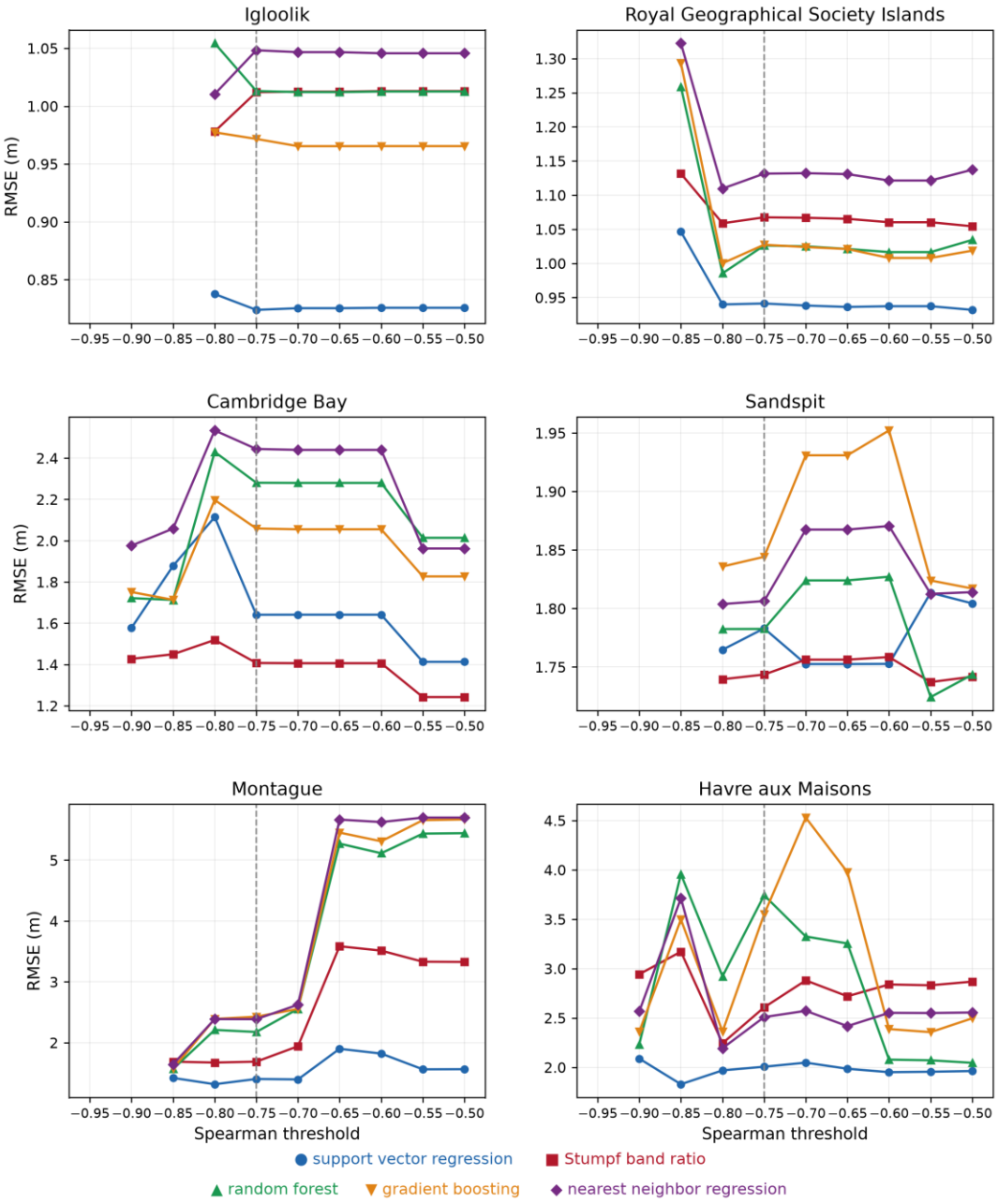

Figure 11. RMSE of the five SDB methods at each of the six sites. In each panel, points are shown only at the Spearman thresholds at which that site could be trained, and the dashed line marks the threshold of −0.75.

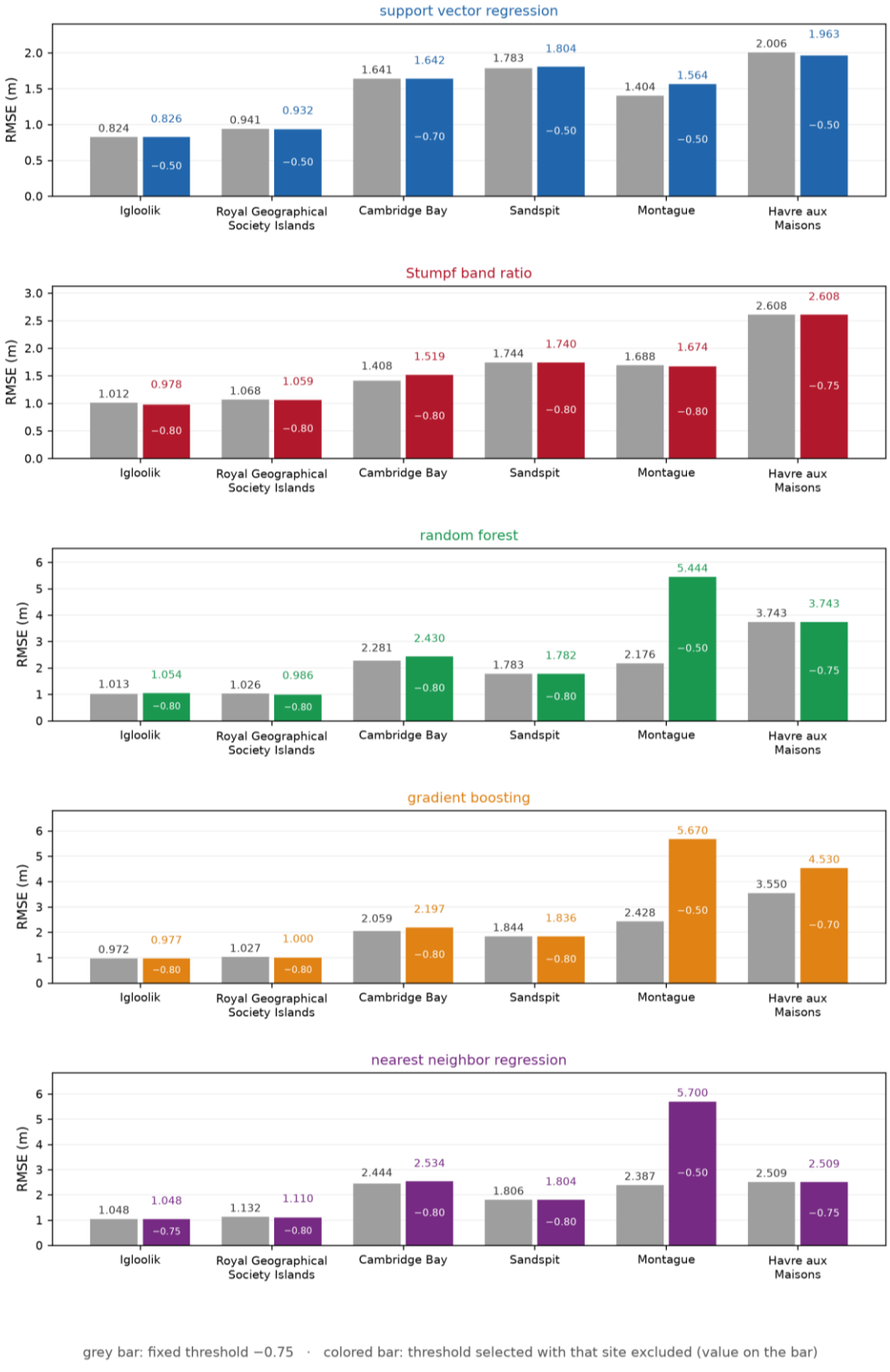

Figure 12. Cross validation of the Spearman threshold for five SDB methods, with each site excluded in turn. Grey bars: RMSE at the excluded site with the fixed threshold of −0.75; colored bars: RMSE at

the same excluded site with the threshold selected from the other five sites, with the selected value written on the bar.

Figure 12 shows the result of the cross validation described in Section 2.3.8. The six sites were excluded in turn. Each time a site was excluded, each of the five methods selected the threshold with the lowest mean RMSE across the five remaining sites, and this selected threshold was then applied at the excluded site. The six exclusions and five methods gave 30 selections in total, of which −0.80 was selected 16 times, −0.50 eight times, −0.75 four times, and −0.70 twice. For each method, every excluded site therefore received a threshold selected without it, and the SDB map calibrated at that site with that threshold gave one RMSE. The six exclusions gave six such RMSE values, one per excluded site, and these six values were averaged. This average was higher than the corresponding average with the fixed threshold of −0.75, by 0.008 m for the Stumpf band ratio, 0.570 m for random forest, 0.722 m for gradient boosting, 0.563 m for nearest neighbor regression, and 0.022 m for support vector regression. No method met the requirement described in Section 2.3.8 of reducing the RMSE at the excluded site by at least 0.05 m. The fixed threshold of −0.75 was therefore retained for all sites.

For the other fixed values, the sensitivity analysis described in Section 2.3.8 was run at Montague, Igloolik, and Cambridge Bay. Four kinds of changes were tested one at a time: removing the two-scene rule, changing the mask thresholds, changing the minimum number of photons required in a group, and changing the elevation range used in the optical screening. After each change, the RMSE of the retained photons against the NONNA10 reference, as defined in Section 2.3.6, was recomputed. Removing the two-scene rule raised this RMSE from 0.38 m to 2.24 m at Montague and by 0.21 m at Igloolik. At Montague, relaxing the mask thresholds raised the RMSE by 1.41 m, and lowering the minimum number of photons to eight raised it by 0.23 m. Tightening the mask thresholds left no retained photons at any of the three sites, so no RMSE could be computed for that change. Every other change, including every change to the elevation range, changed the RMSE at all three sites by less than 0.02 m. The full results are given in Appendix A.

2.4.4 Results of the contribution of each stage in the screening workflow

Table 3. Number of photons retained at each stage of the screening workflow.

Site	S0	S1	S2	S3	S4	S5	S6
Royal Geographical Society Islands	26,250	24,858	24,766	24,548	12,245	7,647	7,511

Sandspit	1,890	1,862	1,703	1,685	1,354	526	414
Montague	5,377	5,314	5,184	5,155	3,380	1,345	945
Cambridge Bay	43,748	43,328	43,225	42,574	7,959	6,147	6,073
Havre aux Maisons	4,818	4,788	4,669	4,628	3,995	1,476	1,083
Igloolik	3,338	3,293	3,247	3,202	2,428	1,572	1,503

Table 4. RMSE (m) of the matched photons against NONNA10 at each stage.

Site	S0	S1	S2	S3	S4	S5	S6
Royal Geographical Society Islands	14.78	14.98	14.99	14.99	11.95	0.71	0.71
Sandspit	5.36	2.36	1.84	1.84	1.93	1.29	0.35
Montague	2.91	2.80	2.30	2.27	2.11	2.24	0.38
Cambridge Bay	24.64	24.65	24.62	24.66	27.72	0.65	0.65
Havre aux Maisons	3.35	3.34	3.04	3.00	2.99	1.38	1.20
Igloolik	19.39	19.39	18.92	18.82	2.32	1.19	0.98

Table 3 lists the number of photons retained at each stage, and Table 4 lists the RMSE of the matched photons against NONNA10. The first three rules, the requirement of a valid, unmasked pixel, the requirement of at least ten photons in the track, and the Mahalanobis outlier removal, removed few photons and left the RMSE nearly unchanged, as every site retained 89% to 97% of its raw photons after Stage 3. The large improvements came from different rules at different sites. At Igloolik the height range rule reduced the RMSE from 18.82 m to 2.32 m. At Royal Geographical Society Islands and Cambridge Bay, the Spearman rule reduced RMSE from 11.95 m to 0.71 m and from 27.72 m to 0.65 m. At Sandspit the valid pixel rule alone reduced RMSE from 5.36 m to 2.36 m, and the two-scene rule reduced RMSE from 1.29 m to 0.35 m. At Montague the two-scene rule reduced RMSE from 2.24 m to 0.38 m, and at Havre aux Maisons the largest single reduction came from the Spearman rule, which reduced the RMSE from 2.99 m to 1.38 m. The RMSE did not fall at every step. The largest increase occurred at Cambridge Bay, where the height range rule raised the RMSE from 24.66 m to 27.72 m, before the Spearman rule removed most of the remaining error. At every site the refined set of Stage 6 had the lowest photon RMSE of all seven stages.

Table 5. Accurate photons removed by each rule.

Site	Accurate in S0	S1	S2	S3	S4	S5	S6	Total removed
Royal Geographical Society Islands	5,085	12 (0.2%)	0 (0.0%)	9 (0.2%)	328 (6.5%)	207 (4.1%)	72 (1.4%)	628 (12.4%)
Sandspit	1,004	0 (0.0%)	38 (3.8%)	5 (0.5%)	54 (5.4%)	493 (49.1%)	49 (4.9%)	639 (63.6%)
Montague	3,563	0 (0.0%)	34 (1.0%)	21 (0.6%)	890 (25.0%)	1,503 (42.2%)	248 (7.0%)	2,696 (75.7%)
Cambridge Bay	175	0 (0.0%)	0 (0.0%)	1 (0.6%)	19 (10.9%)	0 (0.0%)	3 (1.7%)	23 (13.1%)
Havre aux Maisons	974	0 (0.0%)	27 (2.8%)	13 (1.3%)	86 (8.8%)	528 (54.2%)	66 (6.8%)	720 (73.9%)
Igloolik	661	0 (0.0%)	5 (0.8%)	13 (2.0%)	18 (2.7%)	30 (4.5%)	8 (1.2%)	74 (11.2%)

The accurate photons removed by the screening workflow were counted as described in Section 2.3.10 and are listed in Table 5. The cost of the screening workflow was small at the Arctic sites and large at the Atlantic and Pacific coastal sites. Among the matched photons of Stage 0 that lay within 0.5 m of the reference, the screening workflow removed 11.2% at Igloolik, 12.4% at Royal Geographical Society Islands, and 13.1% at Cambridge Bay, but 63.6% at Sandspit, 75.7% at Montague, and 73.9% at Havre aux Maisons. At the Atlantic and Pacific coastal sites most of this loss came from the Spearman rule, which alone removed 49.1% of the accurate Stage 0 photons at Sandspit, 42.2% at Montague, and 54.2% at Havre aux Maisons, and at Montague the height range rule removed a further 25.0%. At the Arctic sites the removed photons were overwhelmingly inaccurate. At Cambridge Bay, 12,067 matched photons were removed in total, and only 23 of them were accurate.

The SDB RMSE of the five methods at each stage is listed in Appendix D. Stage 0 and Stage 1 produced identical depth maps in all 30 site and method combinations, because the calibration preparation of Section 2.3.7 already removes the photons that Stage 1 removes. Between the later stages the direction of change varied. Compared with the stage before it, Stage 2 lowered the SDB RMSE in 20 of the 30 site and method combinations, Stage 3 in 13, Stage 4 in 18, Stage 5 in 15, and Stage 6 in 22. The refined calibration of Stage 6 gave a lower RMSE than the raw calibration of Stage 0 in 25 of the 30 combinations. Four of the five exceptions occurred at Havre aux Maisons, possibly related to the systematic error of its

vertical correction described in Section 2.3.2. The stage with the lowest SDB RMSE, however, was not always the last one. Across the 30 site and method combinations, the lowest RMSE of the seven stages occurred 11 times at Stage 6, 7 times each at Stage 5 and Stage 4, 4 times at Stage 3, and once at Stage 2. At Cambridge Bay all five methods reached their lowest SDB RMSE at Stage 4, although the photon RMSE at that stage was the highest of the seven stages at that site.

2.4.5 Seasonal analysis results

Table 6. Errors of the raw photons by month type. Each cell gives the number of photons matched with NONNA10, their bias, and their RMSE in meters.

Site	All months	Open water months	Breakup and freeze-up months	Ice months
Royal Geographical Society Islands	17,934 / +10.49 / 14.78	5,429 / +0.01 / 2.42	4,914 / +15.56 / 18.66	7,591 / +14.71 / 16.93
Sandspit	1,694 / −0.54 / 5.36	1,694 / −0.54 / 5.36	—	—
Montague	5,124 / +0.25 / 2.91	4,250 / +0.14 / 2.55	111 / +4.60 / 8.24	763 / +0.24 / 3.30
Cambridge Bay	12,357 / +21.20 / 24.64	9 / +5.32 / 7.76	610 / +18.59 / 27.33	11,738 / +21.35 / 24.50
Havre aux Maisons	4,530 / −0.30 / 3.35	3,750 / −0.74 / 2.08	259 / −0.48 / 1.95	521 / +2.94 / 8.01
Igloolik	1,253 / +7.67 / 19.39	313 / +1.85 / 7.14	678 / +0.71 / 4.14	262 / +32.66 / 41.13

Table 6 lists the errors of the raw photons by month type. At the Arctic sites the accuracy of the raw photons depended strongly on the season. The raw photons from the ice months were almost entirely wrong. Their bias ranged from +14.71 m at Royal Geographical Society Islands to +32.66 m at Igloolik, which means the photons were much shallower than the reference depths, and the share with errors within 1 m was 3.4% at Igloolik, 1.3% at Royal Geographical Society Islands, and 0.5% at Cambridge Bay. Keeping only the open water months reduced the RMSE from 19.39 m to 7.14 m at Igloolik and from 14.78 m to 2.42 m at Royal Geographical Society Islands. At Cambridge Bay only 200 of the 43,748 raw photons came from the open water months and only 9 of them were matched with NONNA10, which is too few for a meaningful value. At Montague and Havre aux Maisons the seasonal

dependence was weaker. The open water month selection reduced the RMSE from 2.91 m to 2.55 m and from 3.35 m to 2.08 m, and the ice month photons at these two sites did not show the strong positive bias of the Arctic sites. At Sandspit all months are open water months.

Table 7. The open water selection, the sensitivity variant that also includes the breakup and freeze-up months, and the refined photons of Section 2.3.5, compared against NONNA10. Each cell gives the number of matched photons and their RMSE in meters.

Site	Open water months	Open water, breakup, and freeze-up months	Refined
Royal Geographical Society Islands	5,429 / 2.42	10,343 / 12.98	5,512 / 0.71
Sandspit	1,694 / 5.36	1,694 / 5.36	414 / 0.35
Montague	4,250 / 2.55	4,361 / 2.84	945 / 0.38
Cambridge Bay	9 / 7.76	619 / 27.15	290 / 0.65
Havre aux Maisons	3,750 / 2.08	4,009 / 2.08	1,083 / 1.20
Igloolik	313 / 7.14	991 / 5.28	829 / 0.98

Table 7 compares the two kinds of month selections with the refined photons. At every site with a meaningful value the open water selection remained far less accurate than the refined photons, with 7.14 m against 0.98 m at Igloolik, 2.42 m against 0.71 m at Royal Geographical Society Islands, 2.55 m against 0.38 m at Montague, and 2.08 m against 1.20 m at Havre aux Maisons. The sensitivity variant, which included open water months and breakup and freeze-up months, showed that the result of selecting photons by month also depends on whether the breakup and freeze-up months are included. Including them raised the RMSE from 2.42 m to 12.98 m at Royal Geographical Society Islands and produced 27.15 m at Cambridge Bay, higher than the raw set itself, but lowered the RMSE from 7.14 m to 5.28 m at Igloolik. Neither including nor excluding these months worked at all sites.

Table 8. The refined photons by month type. Each cell gives the number of refined photons from that month type, the number of them matched with NONNA10, and the RMSE in meters of the matched photons.

Site	Open water	Breakup and freeze-up	Ice
Royal Geographical Society Islands	6,619 / 4,923 / 0.72	892 / 589 / 0.59	0 / 0 / —
Sandspit	414 / 414 / 0.35	—	—
Montague	916 / 916 / 0.29	0 / 0 / —	29 / 29 / 1.43
Cambridge Bay	146 / 5 / 0.39	5,927 / 285 / 0.65	0 / 0 / —
Havre aux Maisons	1,052 / 1,052 / 1.21	31 / 31 / 0.62	0 / 0 / —
Igloolik	718 / 172 / 0.53	785 / 657 / 1.06	0 / 0 / —

Table 8 lists the refined photons by month type. At the three Arctic sites not a single refined photon came from the ice months. At Cambridge Bay 5,927 of the 6,073 refined photons came from the breakup and freeze-up months, 5,204 of them from October alone, and their RMSE was 0.65 m, so a selection by open water months alone would have discarded nearly all of the photons that the optical screening found usable at this site. The only refined photons from ice months occurred at Montague, 29 photons from February and March with an RMSE of 1.43 m.

2.5 Discussion

This study asked whether optical imagery can be used to refine ATL24 bathymetric photons. The answer was yes: the workflow removed large errors present in the raw photon sets.

Optical imagery helps because it provides information that the photon cloud itself cannot: it comes from a different type of sensor than the ICESat-2 lidar, so it offers an independent view of the same shallow water against which the photons can be checked. A sequence of misclassified photons can look like a real seafloor within the photon cloud itself: the points are dense and form a smooth line. Along the same ground path, however, the imagery records how the water attenuates light, and the blue to green band ratio changes with depth. The depths of real seafloor photons vary together with this ratio, while the depths of near surface photons misclassified as seafloor do not. This explains the large error reductions at Royal Geographical Society Islands, Sandspit, Montague, Cambridge Bay, and Igloolik, where the raw sets contained large numbers of such misclassified photons, whose depths bore no relation to the band ratio along the track. A strong optical correlation did not guarantee a small error against the reference, nor did a weak one guarantee a large error, which is expected:

reflectance responds to water and bottom properties as well as depth. A similar conclusion was reached when one band ratio workflow was applied to both a clear coast and a turbid lagoon, with results that had to be assessed case by case (Bernardis et al., 2023).

The optical rule has a blind spot: the Spearman correlation describes rank order, not absolute depth. A track can pass while tracking something other than the seafloor, as long as that surface happens to rise and fall with the water depth.

Optical screening complements the methods already used for photon extraction. Density filters, neighborhood searches, histograms, grids, and learned point cloud features all use structure within the photon data (Jung et al., 2025), and a PointNet++ model has been shown to learn this structure across many environments (Lin & Knudby, 2023). The present workflow adds a source of information from outside the photon cloud, applied after classification. The ATL24 product evaluation reported strong overall performance while still documenting persistent misclassification of near surface photons as seafloor (Parrish et al., 2025), and the results at the six Canadian sites in this study show that ATL24 photons still need to be checked before they are used.

The cost of the filter is coverage. The refined dataset contains 17,529 of the 85,421 raw photons. Whether this trade is worthwhile depends on the application: a strict filter protects an SDB model from false calibration depths, but where ATL24 offers few photons to begin with, refinement can leave nothing at all.

Whether a track passed also depended on the scene. The same track could pass in one image and fail in another. This study did not investigate why; future work could record which scenes supported each photon, together with measures of scene quality, and identify what makes a track pass in one image and fail in another.

The threshold sensitivity analysis showed that the fixed threshold is hard to improve on. On the strict side of the tested range, so few calibration photons remained that most sites could no longer train an SDB model. In the cross validation, each round picked the threshold that performed best on its five training sites, most often −0.80 and sometimes −0.50, yet these picks brought no practically meaningful improvement at the excluded sites, and for three of

the five methods they made the excluded sites worse by more than half a meter on average. In other words, a threshold that looked better on the training sites was better mainly for those particular sites and tracks, and did not carry over. The fixed threshold therefore provides the most useful main comparison: it applies the same rule everywhere, and it was never adjusted to fit the NONNA10 data used to report the errors. Among all thresholds tested, −0.75 performed as well as or better than any alternative at the excluded sites, so it was retained.

The stage-by-stage analysis of Section 2.4.4 showed what each rule contributes. The rules that reduce the photon error differ from site to site. Most of the accurate photons removed at the Atlantic and Pacific coastal sites were removed by the height range condition and the Spearman test, and at Sandspit and Montague these were not the rules that carried the improvement. The SDB maps calibrated at each stage showed that the full workflow was not always the best stopping point, as the lowest SDB RMSE occurred at an intermediate stage in 19 of the 30 site and method combinations. Future research could test whether a simpler combination of rules, or rules with site dependent settings, preserves the improvement while discarding fewer accurate photons.

The seasonal analysis of Section 2.4.5 located the source of the large errors at the Arctic sites. The raw photons from the ice months lay 15 to 33 m above the reference depths on average, the signature of surface returns classified as seafloor, and not one of them survived the screening, although the screening uses no acquisition dates. Selecting photons by month was no substitute for the optical test, because usable photons also occur outside the open water months, as at Cambridge Bay, where nearly all refined photons came from the breakup and freeze-up months.

The error statistics of this thesis treat each matched photon and each evaluation pixel as one observation, but neither is an independent sample. Neighboring NONNA10 cells share the same water and bottom conditions, photons along a track are consecutive measurements of the same ground path, and misclassified photons occur in contiguous stretches of track rather than at random. The reported RMSE values therefore describe the magnitude of the errors over the evaluated area, while the effective number of independent observations is smaller than the photon and pixel counts suggest. The screening itself acts on whole track and scene groups, so it removes or keeps errors in spatially connected blocks, and how it changes the spatial structure of the remaining errors is not quantified in this thesis. Future research could quantify the spatial autocorrelation of the photon and SDB errors and its interaction with the

screening.

The workflow developed in this study runs on the released ATL24 product, after classification. Unlike studies that use imagery to guide the extraction of individual photons (Hu et al., 2025; Li et al., 2025), it needs no access to the extraction step, so any ATL24 user with Sentinel-2 imagery can apply it. Its output has clear limits: a track that passes is consistent with the imagery but not guaranteed correct in absolute depth, and the number of retained photons varies strongly from site to site. Within these limits, the workflow makes ATL24 a more reliable source of SDB calibration data in waters where few alternatives exist.

2.6 Conclusion

We developed a workflow that uses Sentinel-2 imagery to refine ATL24 bathymetric photons and tested it at six Canadian coastal sites. Each track was tested separately in each scene: its photon depths were compared with the blue to green band ratio along the track, and the track passed a scene only when the two agreed with a Spearman correlation of −0.75 or stronger. Photons were retained only from tracks that passed in at least two scenes.

The workflow retained 17,529 of the 85,421 raw photons and removed large errors present in the raw ATL24 sets: at the five Chart Datum sites, the refined RMSE against NONNA10 fell to between 0.35 m and 0.98 m, from raw values between 2.91 m and 24.64 m. At Havre aux Maisons, whose vertical correction carries a different systematic error, the refined RMSE was 1.20 m, the highest of the six sites.

The stage-by-stage and seasonal analyses showed how the workflow earns this improvement. Different rules do the work at different sites, and the screening as a whole removed 64% to 76% of the accurate photons at the Atlantic and Pacific coastal sites. At the Arctic sites the large raw errors came almost entirely from outside the open water months, and although the screening uses no acquisition dates, not one refined photon at these sites came from an ice month. Selecting photons by month could not replace the screening.

Calibrating six SDB methods with raw and with refined photons showed that better photons make better bathymetry: refined calibration produced the lower RMSE in 29 of 36 site and method combinations, and the exceptions were concentrated at a single site. The threshold sensitivity analysis found no Spearman threshold that worked better than the fixed value of −0.75 across sites.

The workflow developed in this thesis therefore improves ATL24 where it performs poorly, in the turbid and Arctic shallow waters of Canada, and turns the ATL24 photons into better calibration data for SDB.

3 Chapter 3: Conclusion

This thesis examined whether optical imagery can be used to refine the bathymetric photons in the ICESat-2 ATL24 product. The motivation is a continuing limitation of satellite derived bathymetry: SDB models need reliable calibration data, yet such data are scarcest in the remote coastal areas where satellite mapping is most useful. ATL24 makes bathymetric photons easy to obtain, because users no longer need to extract them from the ATL03 photon cloud themselves, but the photons it classifies as bathymetric photons can still include returns that do not represent the seafloor, and its performance in Canadian turbid and Arctic waters had not been tested.

This thesis developed a workflow that refines ATL24 photons with Sentinel-2 imagery and applied it at six Canadian coastal sites spanning turbid and Arctic waters, evaluating the retained photons against NONNA10 reference data. The raw ATL24 data at these sites were far poorer than the accuracy reported by published evaluations elsewhere, with site level RMSEs from 2.9 to 24.6 m, and at several sites large numbers of sea surface returns had been classified as seafloor. The workflow greatly reduced these large errors: at the five Chart Datum sites, the refined RMSE fell to between 0.35 m and 0.98 m, from raw values between 2.91 m and 24.64 m. The improvement was not universal: at Havre aux Maisons, whose vertical correction carries a different systematic error, the refined RMSE, 1.20 m, was the highest of the six sites.

The cost of refinement is the number of photons that remain. The refined dataset contains 17,529 of the 85,421 raw photons. Strict screening protects an SDB model from false calibration depths, but where ATL24 offers few photons to begin with, refinement can leave nothing at all.

Calibrating six SDB methods with raw and with refined photons gave the decisive test: refined calibration produced the lower RMSE in 29 of 36 site and method combinations, with a median reduction of 0.6 m. Refining the photons with optical imagery therefore improved the SDB calibrated from them at most of the sites tested. The threshold sensitivity analysis found no Spearman threshold that worked better than the fixed value of −0.75 across sites.

The stage-by-stage and seasonal analyses showed where the improvement comes from and what it costs: the rules that carry the improvement differ between sites, accurate photons are the price at the Atlantic and Pacific coastal sites, and at the Arctic sites the screening removed every photon from the ice months without using any acquisition dates.

Future research could test simpler or site dependent combinations of the screening rules, explain why the same track passes in one scene and fails in another, and test the workflow at more sites and on newer ATL24 releases. Refining ATL24 photons with optical imagery produced better SDB calibration data in the turbid and Arctic shallow waters of Canada, where such data are scarcest.

4 References

Ashphaq, M., Srivastava, P. K., & Mitra, D. (2021). Review of near-shore satellite derived bathymetry: Classification and account of five decades of coastal bathymetry research. Journal of Ocean Engineering and Science, 6(4), 340–359. https://doi.org/10.1016/j.joes.2021.02.006

Bernardis, M., Nardini, R., Apicella, L., Demarte, M., Guideri, M., Federici, B., Quarati, A., & De Martino, M. (2023). Use of ICEsat-2 and Sentinel-2 open data for the derivation of bathymetry in shallow waters: Case studies in Sardinia and in the Venice Lagoon. Remote Sensing, 15(11), 2944. https://doi.org/10.3390/rs15112944

Braybrook, C. A., Else, B. G. T., Walker, Z., Yendamuri, K., Mundy, C. J., & Komarov, A. S. (2026). Identifying thermodynamic regimes of CO2 gas exchange during the melt transition on first-year sea ice using a multi-sensor synthetic aperture radar approach. Journal of Geophysical Research: Oceans, 131(3), e2025JC023228. https://doi.org/10.1029/2025JC023228

Butterworth, B. J., Else, B. G. T., Brown, K. A., Mundy, C. J., Williams, W. J., Rotermund, L. M., & de Boer, G. (2025). Annual carbon dioxide flux over seasonal sea ice in the Canadian Arctic. The Cryosphere, 19(11), 5317–5335. https://doi.org/10.5194/tc-19-5317-2025

Canadian Hydrographic Service. (2026a). Canadian Hydrographic Service non-navigational (NONNA) bathymetric data [Data set]. Government of Canada Open Government Portal. Retrieved August 12, 2026, from https://open.canada.ca/data/en/dataset/d3881c4c-650d-4070-bf9b-1e00aabf0a1d

Canadian Hydrographic Service. (2026b). Chart Datum to ITRF2014(2010) separation products CANEAST2024v2, CANNORTH_2025_V2, and CANWEST_2026_V1: Product metadata files (CANEAST2024v2_CD_ITRF2014_metadata.txt; CANNORTH_2025_V2_[ITRF2014(2010)_CD(2025)]_metadata.txt; CANWEST_2026_V1_[ITRF(2014)_CD(2026)]_metadata.txt) [Data delivery documentation, delivery packages 2026-0129-1260-UO-ATL-2, 2026-0129-UO-OPA, and 2026-0129-UO-PAC]. Fisheries and Oceans Canada. Obtained upon request under Data Use License Agreement 2026-0522-UO.

Canadian Hydrographic Service. (2026c). Havre-aux-Maisons (01976) — Station information. Government of Canada. Retrieved August 9, 2026, from

https://tides.gc.ca/en/stations/01976

Canadian Ice Service. (2021a). Ice climate normals for the Canadian East Coast waters 1991 to 2020. Environment and Climate Change Canada. Retrieved November 25, 2026, from https://www.canada.ca/en/environment-climate-change/services/ice-forecasts-observations/latest-conditions/climatology/ice-climate-normals/canadian-east-coast-waters.html

Canadian Ice Service. (2021b). Ice climate normals for the northern Canadian waters 1991 to 2020. Environment and Climate Change Canada. Retrieved November 25, 2026, from https://www.canada.ca/en/environment-climate-change/services/ice-forecasts-observations/latest-conditions/climatology/ice-climate-normals/northern-canadian-waters.html

Cao, B., Liu, H., & Cao, B. (2024). Making satellite-derived empirical bathymetry independent of high-quality in-situ depth data: An assessment of four possible model calibration data. ISPRS Journal of Photogrammetry and Remote Sensing, 211, 336–355. https://doi.org/10.1016/j.isprsjprs.2024.04.014

Cao, B., Xiong, L., Liu, H., Chen, J., Zhang, H., Wu, S., Xu, D., & Cao, B. (2025). Optimal approach to utilize multiple-pass ICESat-2 ATL03 data for satellite-derived empirical bathymetry. ISPRS Journal of Photogrammetry and Remote Sensing, 229, 303–322. https://doi.org/10.1016/j.isprsjprs.2025.08.024

Cesbron, G., Melet, A., Almar, R., Lifermann, A., Tullot, D., & Crosnier, L. (2021). Pan-European satellite-derived coastal bathymetry—review, user needs and future services. Frontiers in Marine Science, 8, 740830. https://doi.org/10.3389/fmars.2021.740830

Chen, Y., Le, Y., Zhang, D., Wang, Y., Qiu, Z., & Wang, L. (2021). A photon-counting LiDAR bathymetric method based on adaptive variable ellipse filtering. Remote Sensing of Environment, 256, 112326. https://doi.org/10.1016/j.rse.2021.112326

Chen, Z., Wang, L., Feng, W., Gu, Y., Li, J., & Wang, Y. P. (2026). Advancing coral reef bathymetry: A GAN-augmented and stratified CNN analysis of fused ICESat-2 and Sentinel-2 dataset. IEEE Transactions on Geoscience and Remote Sensing, 64, 4202311. https://doi.org/10.1109/TGRS.2026.3659873

Cheng, J., Cheng, L., Chu, S., Li, J., Hu, Q., Ye, L., Wang, Z., & Chen, H. (2023). A comprehensive evaluation of machine learning and classical approaches for spaceborne active-passive fusion bathymetry of coral reefs. ISPRS International Journal of Geo-Information, 12(9), 381. https://doi.org/10.3390/ijgi12090381

Dandabathula, G., Ghatage, O. S., Roy, S., Yadav, S., Bera, A. K., & Srivastav, S. K. (2025). Quantitative validation of ICESat-2 ATL24 in shallow marine environments using reference-grade bathymetric data. Marine Geodesy. Advance online publication. https://doi.org/10.1080/01490419.2025.2608329

Derksen, C., Burgess, D., Duguay, C., Howell, S., Mudryk, L., Smith, S., Thackeray, C., & Kirchmeier-Young, M. (2019). Changes in snow, ice, and permafrost across Canada. In E. Bush & D. S. Lemmen (Eds.), Canada's changing climate report (pp. 194–260). Government of Canada. https://changingclimate.ca/CCCR2019/chapter/5-0/

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sensing of Environment, 120, 25-36. https://doi.org/10.1016/j.rse.2011.11.026

Eugenio, F., Marcello, J., Mederos-Barrera, A., & Marqués, F. (2022). High-resolution satellite bathymetry mapping: Regression and machine learning-based approaches. IEEE Transactions on Geoscience and Remote Sensing, 60, 1–14. https://doi.org/10.1109/TGRS.2021.3135462

European Space Agency. (2026). Sentinel-2 Products Specification Document (S2-PDGS-CS-DI-PSD, Issue 15.2). https://sentiwiki.copernicus.eu/__attachments/a_79d81f1db038da666d7831a1b76a44df399e5a67ba5e5f39f1f7038d75cb1fb9/S2-Sentinel-2%20Products%20Specification%20Document-15.2.pdf

Gough, W. A. (2025). A marine season metric for Foxe Basin, Nunavut, Canada: Insights into the evolving nature of sea-ice breakup and freeze-up. Coasts, 5(1), 7. https://doi.org/10.3390/coasts5010007

Howell, S. E. L., Wohlleben, T., Dabboor, M., Derksen, C., Komarov, A., & Pizzolato, L. (2013). Recent changes in the exchange of sea ice between the Arctic Ocean and the Canadian Arctic Archipelago. Journal of Geophysical Research: Oceans, 118(7), 3595–3607. https://doi.org/10.1002/jgrc.20265

Hsu, H.-J., Huang, C.-Y., Jasinski, M., Li, Y., Gao, H., Yamanokuchi, T., Wang, C.-G., Chang, T.-M., Ren, H., Kuo, C.-Y., & Tseng, K.-H. (2021). A semi-empirical scheme for bathymetric mapping in shallow water by ICESat-2 and Sentinel-2: A case study in the South China Sea. ISPRS Journal of Photogrammetry and

Remote Sensing, 178, 1–19. https://doi.org/10.1016/j.isprsjprs.2021.05.012

Hu, Q., Li, J., & Cheng, L. (2025). Satellite-derived bathymetry based on iterative fusion of active lidar photons and passive pseudo-photons. International Journal of Digital Earth, 18(1), 2543568. https://doi.org/10.1080/17538947.2025.2543568

International Hydrographic Organization. (2024). IHO Publication S-44: Standards for Hydrographic Surveys (Edition 6.2.0). International Hydrographic Organization. https://iho.int/uploads/user/pubs/standards/s-44/S-44_Edition_6.2.0_adopted.pdf

International Hydrographic Organization. (2026). IHO Publication C-55: Status of Hydrographic Surveying and Charting Worldwide (continuously updated). International Hydrographic Organization. Snapshot downloaded July 22, 2026.

Jung, J., Parrish, C. E., Magruder, L. A., Herrmann, J., Yoo, S., & Perry, J. S. (2025). ICESat-2 bathymetry algorithms: A review of the current state-of-the-art and future outlook. ISPRS Journal of Photogrammetry and Remote Sensing, 223, 413–439. https://doi.org/10.1016/j.isprsjprs.2025.03.016

Kibele, J., & Shears, N. T. (2016). Nonparametric empirical depth regression for bathymetric mapping in coastal waters. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(11), 5130–5138. https://doi.org/10.1109/JSTARS.2016.2598152

Klonowski, W. M., Fearns, P. R. C. S., & Lynch, M. J. (2007). Retrieving key benthic cover types and bathymetry from hyperspectral imagery. Journal of Applied Remote Sensing, 1, 011505. https://doi.org/10.1117/1.2816113

Knudby, A. (2025). Satellite imagery for hazard detection: Better use of satellite imagery could help prevent groundings in the Canadian Arctic. The International Hydrographic Review, 31(2), 30–43. https://doi.org/10.58440/ihr-31-2-a13

Kristoffersen, F. R. (2026). Satellite-derived bathymetry of Danish coastal waters using machine learning with Sentinel-2 and ICESat-2 data (Master's thesis in Geographical Information Science, No. 204). Lund University.

Lee, Z., Carder, K. L., Mobley, C. D., Steward, R. G., & Patch, J. S. (1998). Hyperspectral remote sensing for shallow waters. I. A semianalytical model. Applied Optics, 37(27), 6329–6338. https://doi.org/10.1364/AO.37.006329

Leng, Z., Zhang, J., Ma, Y., Zhang, J., & Zhu, H. (2022). A novel bathymetry signal photon extraction algorithm for photon-counting LiDAR based on adaptive elliptical

neighborhood. International Journal of Applied Earth Observation and Geoinformation, 115, 103080. https://doi.org/10.1016/j.jag.2022.103080

Li, J., Chu, S., Hu, Q., Qu, Z., Zhang, J., & Cheng, L. (2025). A slope adaptive bathymetric method by integrating ICESat-2 ATL03 data with Sentinel-2 images. Remote Sensing, 17(17), 3019. https://doi.org/10.3390/rs17173019

Lin, Y., & Knudby, A. (2023). Global automated extraction of bathymetric photons from ICESat-2 data based on a PointNet++ model. International Journal of Applied Earth Observation and Geoinformation, 124, 103512. https://doi.org/10.1016/j.jag.2023.103512

Lumban-Gaol, Y., Ohori, K. A., & Peters, R. (2022). Extracting coastal water depths from multi-temporal Sentinel-2 images using convolutional neural networks. Marine Geodesy, 45(6), 615–644. https://doi.org/10.1080/01490419.2022.2091696

Lyzenga, D. R. (1978). Passive remote sensing techniques for mapping water depth and bottom features. Applied Optics, 17(3), 379–383. https://doi.org/10.1364/AO.17.000379

Lyzenga, D. R. (1985). Shallow-water bathymetry using combined lidar and passive multispectral scanner data. International Journal of Remote Sensing, 6(1), 115–125. https://doi.org/10.1080/01431168508948428

Ma, Y., Xu, N., Liu, Z., Yang, B., Yang, F., Wang, X. H., & Li, S. (2020). Satellite-derived bathymetry using the ICESat-2 lidar and Sentinel-2 imagery datasets. Remote Sensing of Environment, 250, 112047. https://doi.org/10.1016/j.rse.2020.112047

Magruder, L. A., Parrish, C., Perry, J., Swinski, J. P., Holwill, M., Kief, K., & Corcoran, F. (2025a). ATLAS/ICESat-2 L3A along track coastal and nearshore bathymetry (ATL24, Version 1) [Data set]. NASA National Snow and Ice Data Center Distributed Active Archive Center. Retrieved June–July 2026, from https://doi.org/10.5067/ATLAS/ATL24.001

Magruder, L., Parrish, C., Perry, J., Holwill, M., Swinski, J. P., & Kief, K. (2025b). ICESat-2 coastal and nearshore bathymetry product algorithm development. Earth and Space Science, 12(10), e2025EA004390. https://doi.org/10.1029/2025EA004390

Maritorena, S., Morel, A., & Gentili, B. (1994). Diffuse reflectance of oceanic shallow waters: Influence of water depth and bottom albedo. Limnology and Oceanography, 39(7), 1689–1703. https://doi.org/10.4319/lo.1994.39.7.1689

Markus, T., Neumann, T., Martino, A., Abdalati, W., Brunt, K., Csatho, B., Farrell, S.,

Fricker, H., Gardner, A., Harding, D., Jasinski, M., Kwok, R., Magruder, L., Lubin, D., Luthcke, S., Morison, J., Nelson, R., Neuenschwander, A., Palm, S., ⋯ Zwally, J. (2017). The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation. Remote Sensing of Environment, 190, 260–273. https://doi.org/10.1016/j.rse.2016.12.029

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. International Journal of Remote Sensing, 39(9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

Misra, A., Vojinovic, Z., Ramakrishnan, B., Luijendijk, A., & Ranasinghe, R. (2018). Shallow water bathymetry mapping using support vector machine (SVM) technique and multispectral imagery. International Journal of Remote Sensing, 39(13), 4431–4450. https://doi.org/10.1080/01431161.2017.1421796

Mudiyanselage, S. S. J. D., Abd-Elrahman, A., Wilkinson, B., & Lecours, V. (2022). Satellite-derived bathymetry using machine learning and optimal Sentinel-2 imagery in South-West Florida coastal waters. GIScience & Remote Sensing, 59(1), 1143–1158. https://doi.org/10.1080/15481603.2022.2100597

Neumann, T. A., Martino, A. J., Markus, T., Bae, S., Bock, M. R., Brenner, A. C., Brunt, K. M., Cavanaugh, J., Fernandes, S. T., Hancock, D. W., Harbeck, K., Lee, J., Kurtz, N. T., Luers, P. J., Luthcke, S. B., Magruder, L., Pennington, T. A., Ramos-Izquierdo, L., Rebold, T., ⋯ Thomas, T. C. (2019). The Ice, Cloud, and Land Elevation Satellite-2 mission: A global geolocated photon product derived from the Advanced Topographic Laser Altimeter System. Remote Sensing of Environment, 233, 111325. https://doi.org/10.1016/j.rse.2019.111325

Parrish, C. E., Magruder, L. A., Neuenschwander, A. L., Forfinski-Sarkozi, N., Alonzo, M., & Jasinski, M. (2019). Validation of ICESat-2 ATLAS bathymetry and analysis of ATLAS's bathymetric mapping performance. Remote Sensing, 11(14), 1634. https://doi.org/10.3390/rs11141634

Parrish, C. E., Magruder, L. A., Perry, J., Holwill, M., Swinski, J. P., & Kief, K. (2025). Analysis and accuracy assessment of a new global nearshore ICESat-2 bathymetric data product. Earth and Space Science, 12(8), e2025EA004391. https://doi.org/10.1029/2025EA004391

Philpot, W. D. (1989). Bathymetric mapping with passive multispectral imagery. Applied Optics, 28(8), 1569–1578. https://doi.org/10.1364/AO.28.001569

Qin, X., Wu, Z., Luo, X., Li, J., Chen, P., Shang, J., Zhao, D., Zhou, J., & Wan, H. (2026). A novel ICESat-2 bathymetry model guided by photon density profiles and synergy with Sentinel-2 for automated seafloor mapping without in situ data. IEEE Transactions on Geoscience and Remote Sensing, 64, 4203924. https://doi.org/10.1109/TGRS.2026.3669597

Ranndal, H., Sigaard Christiansen, P., Kliving, P., Baltazar Andersen, O., & Nielsen, K. (2021). Evaluation of a statistical approach for extracting shallow water bathymetry signals from ICESat-2 ATL03 photon data. Remote Sensing, 13(17), 3548. https://doi.org/10.3390/rs13173548

Sagawa, T., Yamashita, Y., Okumura, T., & Yamanokuchi, T. (2019). Satellite-derived bathymetry using machine learning and multi-temporal satellite images. Remote Sensing, 11(10), 1155. https://doi.org/10.3390/rs11101155

Shean, D., Swinski, J. P., Smith, B., Sutterley, T., Henderson, S., Ugarte, C., Lidwa, E., & Neumann, T. (2023). SlideRule: Enabling rapid, scalable, open science for the NASA ICESat-2 mission and beyond. Journal of Open Source Software, 8(81), 4982. https://doi.org/10.21105/joss.04982

Stumpf, R. P., Holderied, K., & Sinclair, M. (2003). Determination of water depth with high-resolution satellite imagery over variable bottom types. Limnology and Oceanography, 48(1, part 2), 547–556. https://doi.org/10.4319/lo.2003.48.1_part_2.0547

Surisetty, V. V. A. K., Venkateswarlu, Ch., Gireesh, B., Prasad, K. V. S. R., & Sharma, R. (2021). On improved nearshore bathymetry estimates from satellites using ensemble and machine learning approaches. Advances in Space Research, 68(8), 3342–3364. https://doi.org/10.1016/j.asr.2021.06.034

Wan, J., & Ma, Y. (2021). Shallow water bathymetry mapping of Xinji Island based on multispectral satellite image using deep learning. Journal of the Indian Society of Remote Sensing, 49(9), 2019–2032. https://doi.org/10.1007/s12524-020-01255-9

Wang, Z., Nie, S., Wang, C., Fu, B., Xi, X., & Yang, B. (2024). A novel bathymetric signal extraction method for photon-counting LiDAR data based on adaptive rotating ellipse and curve iterative fitting. International Journal of Applied Earth Observation and Geoinformation, 132, 104042. https://doi.org/10.1016/j.jag.2024.104042

Wölfl, A.-C., Snaith, H., Amirebrahimi, S., Devey, C. W., Dorschel, B., Ferrini, V., Huvenne,

V. A. I., Jakobsson, M., Jencks, J., Johnston, G., Lamarche, G., Mayer, L., Millar, D., Pedersen, T. H., Picard, K., Reitz, A., Schmitt, T., Visbeck, M., Weatherall, P., & Wigley, R. (2019). Seafloor mapping—the challenge of a truly global ocean bathymetry. Frontiers in Marine Science, 6, 283. https://doi.org/10.3389/fmars.2019.00283

Wu, Z., Mao, Z., & Shen, W. (2021). Integrating multiple datasets and machine learning algorithms for satellite-based bathymetry in seaports. Remote Sensing, 13(21), 4328. https://doi.org/10.3390/rs13214328

Xie, C., Chen, P., Pan, D., Zhong, C., & Zhang, Z. (2021). Improved filtering of ICESat-2 lidar data for nearshore bathymetry estimation using Sentinel-2 imagery. Remote Sensing, 13(21), 4303. https://doi.org/10.3390/rs13214303

Xie, C., Chen, P., Zhang, Z., & Pan, D. (2023). Satellite-derived bathymetry combined with Sentinel-2 and ICESat-2 datasets using machine learning. Frontiers in Earth Science, 11, 1111817. https://doi.org/10.3389/feart.2023.1111817

Xie, C., Chen, P., Zhang, S., & Huang, H. (2024). Nearshore bathymetry from ICESat-2 LiDAR and Sentinel-2 imagery datasets using physics-informed CNN. Remote Sensing, 16(3), 511. https://doi.org/10.3390/rs16030511

Zhang, D., Chen, Y., Le, Y., Dong, Y., Dai, G., & Wang, L. (2022). Refraction and coordinate correction with the JONSWAP model for ICESat-2 bathymetry. ISPRS Journal of Photogrammetry and Remote Sensing, 186, 285–300. https://doi.org/10.1016/j.isprsjprs.2022.02.020

Zhou, W., Tang, Y., Jing, W., Li, Y., Yang, J., Deng, Y., & Zhang, Y. (2023). A comparison of machine learning and empirical approaches for deriving bathymetry from multispectral imagery. Remote Sensing, 15(2), 393. https://doi.org/10.3390/rs15020393

5 Appendix A: Parameter sensitivity tests

This appendix reports single parameter sensitivity tests of the photon refinement workflow at three sites: Cambridge Bay, Montague, and Igloolik. Each test changed one parameter and kept all other parameters at the original values used in this thesis.

Table A1. Parameter sensitivity of the photon refinement workflow at the three test sites. Each column changes one parameter while all other parameters keep their original values. Change rows give the difference relative to the original values, in percent for counts and in meters for RMSE.

Metric	Original values	Two-scene rule off	Mask ×1.2	Min. photons 8	Min. photons 15	Height range 2 m	Height range 4 m
Cambridge Bay							
Passing tracks	8	8	8	8	8	8	8
Passing tracks, change	—	0%	0%	0%	0%	0%	0%
Contributing tracks	8	8	8	8	8	8	8
Contributing tracks, change	—	0%	0%	0%	0%	0%	0%
Retained photons	6073	6147	6076	6073	6073	6073	6073
Retained photons, change	—	+1.2%	+0.0%	0%	0%	0%	0%
Matched photons	290	295	291	290	290	290	290
Matched photons, change	—	+1.7%	+0.3%	0%	0%	0%	0%
RMSE (m)	0.645	0.646	0.644	0.645	0.645	0.645	0.645
RMSE (m), change	—	+0.001	−0.001	0.000	0.000	0.000	0.000
Montague							
Passing tracks	17	17	13	19	13	17	12
Passing tracks, change	—	0%	−23.5%	+11.8%	−23.5%	0%	−29.4%
Contributing tracks	9	17	10	10	8	10	7
Contributing tracks, change	—	+88.9%	+11.1%	+11.1%	−11.1%	+11.1%	−22.2%
Retained	945	1345	966	954	934	1113	901

Metric	Original values	Two-scene rule off	Mask ×1.2	Min. photons 8	Min. photons 15	Height range 2 m	Height range 4 m
photons							
Retained photons, change	—	+42.3%	+2.2%	+1.0%	−1.2%	+17.8%	−4.7%
Matched photons	945	1293	966	954	934	1083	901
Matched photons, change	—	+36.8%	+2.2%	+1.0%	−1.2%	+14.6%	−4.7%
RMSE (m)	0.376	2.244	1.791	0.602	0.378	0.362	0.379
RMSE (m), change	—	+1.868	+1.414	+0.226	+0.001	−0.014	+0.003
Igloolik							
Passing tracks	5	5	5	5	5	5	5
Passing tracks, change	—	0%	0%	0%	0%	0%	0%
Contributing tracks	5	5	5	5	5	5	4
Contributing tracks, change	—	0%	0%	0%	0%	0%	−20.0%
Retained photons	1503	1572	1504	1503	1503	1503	1470
Retained photons, change	—	+4.6%	+0.1%	0%	0%	0%	−2.2%
Matched photons	829	845	829	829	829	829	819
Matched photons, change	—	+1.9%	0%	0%	0%	0%	−1.2%
RMSE (m)	0.976	1.185	0.976	0.976	0.976	0.976	0.980
RMSE (m), change	—	+0.210	0.000	0.000	0.000	0.000	+0.004

The mask ×0.8 case is excluded because the Band 10 threshold lies only 25 DN above the radiometric offset floor, so scaling it by 0.8 removes all pixels.

6 Appendix B: Supplementary threshold sensitivity analysis of the multiband log linear model

Table B1. RMSE (m) of the SDB maps at each site and Spearman threshold, for the multiband log linear model. A dash marks a site and threshold combination that did not meet the training criterion described in Section 2.3.8.

Spearman threshold	Igloolik	Royal Geographical Society Islands	Cambridge Bay	Sandspit	Montague	Havre aux Maisons
−0.95	—	—	—	—	—	—
−0.90	—	—	1.40	—	—	3.34
−0.85	—	1.17	1.48	—	1.93	2.89
−0.80	0.78	1.03	1.54	1.17	1.94	1.88
−0.75	0.82	1.04	1.40	1.17	1.94	2.22
−0.70	0.82	1.04	1.40	1.20	2.29	2.52
−0.65	0.82	1.04	1.40	1.20	3.73	2.44
−0.60	0.82	1.03	1.40	1.20	3.65	2.42
−0.55	0.82	1.03	1.94	1.31	3.67	2.42
−0.50	0.82	1.02	1.94	1.29	3.67	2.44

Table B2. Share of pixels with negative predicted depths, as percent of the prediction domain, at each site and Spearman threshold for the multiband log linear model. A dash marks a site and threshold combination that did not meet the training criterion described in Section 2.3.8. Negative depth predictions are physically impossible and are excluded before scoring. Their share changes with the threshold, which is why the RMSE values of the model are not comparable across thresholds, as described in Section 2.3.8.

Spearman threshold	Igloolik	Royal Geographical Society Islands	Cambridge Bay	Sandspit	Montague	Havre aux Maisons
−0.95	—	—	—	—	—	—
−0.90	—	—	0.5	—	—	6.0
−0.85	—	1.6	0.6	—	11.3	6.3
−0.80	2.9	4.1	0.5	35.5	4.4	9.2
−0.75	4.1	4.1	0.3	35.6	5.0	9.5
−0.70	4.0	4.1	0.3	38.4	5.8	10.9

Spearman threshold	Igloolik	Royal Geographical Society Islands	Cambridge Bay	Sandspit	Montague	Havre aux Maisons
−0.65	4.0	4.1	0.3	38.4	8.9	10.7
−0.60	4.0	4.1	0.3	39.2	9.7	12.8
−0.55	4.0	4.1	5.7	28.2	12.5	12.8
−0.50	4.0	4.1	5.7	28.5	12.6	12.9

7 Appendix C: Per-site RMSE at each Spearman threshold

Tables C1 to C5 list, for each of the five SDB methods, the RMSE at each of the six sites and each of the ten Spearman thresholds of the threshold sensitivity analysis described in Section 2.3.8. A dash marks a site and threshold combination that did not meet the training criterion of that analysis. The six site means shown in Figure 10 and the per-site values shown in Figure 11 are computed from the values listed here.

Table C1. RMSE (m) of the SDB maps at each site and Spearman threshold, for the Stumpf band ratio model. A dash marks a site and threshold combination that did not meet the training criterion described in Section 2.3.8.

Spearman threshold	Igloolik	Royal Geographical Society Islands	Cambridge Bay	Sandspit	Montague	Havre aux Maisons
−0.95	—	—	—	—	—	—
−0.90	—	—	1.43	—	—	2.94
−0.85	—	1.13	1.45	—	1.69	3.17
−0.80	0.98	1.06	1.52	1.74	1.67	2.24
−0.75	1.01	1.07	1.41	1.74	1.69	2.61
−0.70	1.01	1.07	1.41	1.76	1.94	2.88
−0.65	1.01	1.07	1.41	1.76	3.59	2.72
−0.60	1.01	1.06	1.41	1.76	3.51	2.84
−0.55	1.01	1.06	1.24	1.74	3.33	2.83
−0.50	1.01	1.05	1.24	1.74	3.33	2.87

Table C2. As Table C1, for random forest regression.

Spearman threshold	Igloolik	Royal Geographical Society Islands	Cambridge Bay	Sandspit	Montague	Havre aux Maisons
−0.95	—	—	—	—	—	—
−0.90	—	—	1.72	—	—	2.23
−0.85	—	1.26	1.71	—	1.57	3.96
−0.80	1.05	0.99	2.43	1.78	2.21	2.92
−0.75	1.01	1.03	2.28	1.78	2.18	3.74
−0.70	1.01	1.03	2.28	1.82	2.56	3.33
−0.65	1.01	1.02	2.28	1.82	5.27	3.26

Spearman threshold	Igloolik	Royal Geographical Society Islands	Cambridge Bay	Sandspit	Montague	Havre aux Maisons
−0.60	1.01	1.02	2.28	1.83	5.11	2.08
−0.55	1.01	1.02	2.01	1.72	5.44	2.07
−0.50	1.01	1.03	2.01	1.74	5.44	2.05

Table C3. As Table C1, for gradient boosting.

Spearman threshold	Igloolik	Royal Geographical Society Islands	Cambridge Bay	Sandspit	Montague	Havre aux Maisons
−0.95	—	—	—	—	—	—
−0.90	—	—	1.75	—	—	2.36
−0.85	—	1.29	1.71	—	1.58	3.50
−0.80	0.98	1.00	2.20	1.84	2.39	2.36
−0.75	0.97	1.03	2.06	1.84	2.43	3.55
−0.70	0.97	1.02	2.06	1.93	2.55	4.53
−0.65	0.97	1.02	2.06	1.93	5.45	3.98
−0.60	0.97	1.01	2.06	1.95	5.31	2.39
−0.55	0.97	1.01	1.83	1.82	5.66	2.36
−0.50	0.97	1.02	1.83	1.82	5.67	2.50

Table C4. As Table C1, for nearest neighbor regression.

Spearman threshold	Igloolik	Royal Geographical Society Islands	Cambridge Bay	Sandspit	Montague	Havre aux Maisons
−0.95	—	—	—	—	—	—
−0.90	—	—	1.98	—	—	2.57
−0.85	—	1.32	2.06	—	1.64	3.71
−0.80	1.01	1.11	2.53	1.80	2.39	2.19
−0.75	1.05	1.13	2.44	1.81	2.39	2.51
−0.70	1.05	1.13	2.44	1.87	2.62	2.57
−0.65	1.05	1.13	2.44	1.87	5.67	2.42
−0.60	1.05	1.12	2.44	1.87	5.63	2.55

Spearman threshold	Igloolik	Royal Geographical Society Islands	Cambridge Bay	Sandspit	Montague	Havre aux Maisons
−0.55	1.05	1.12	1.96	1.81	5.70	2.55
−0.50	1.05	1.14	1.96	1.81	5.70	2.56

Table C5. As Table C1, for support vector regression.

Spearman threshold	Igloolik	Royal Geographical Society Islands	Cambridge Bay	Sandspit	Montague	Havre aux Maisons
−0.95	—	—	—	—	—	—
−0.90	—	—	1.58	—	—	2.09
−0.85	—	1.05	1.88	—	1.42	1.83
−0.80	0.84	0.94	2.11	1.76	1.32	1.97
−0.75	0.82	0.94	1.64	1.78	1.40	2.01
−0.70	0.83	0.94	1.64	1.75	1.40	2.05
−0.65	0.83	0.94	1.64	1.75	1.90	1.99
−0.60	0.83	0.94	1.64	1.75	1.82	1.95
−0.55	0.83	0.94	1.41	1.81	1.56	1.96
−0.50	0.83	0.93	1.41	1.80	1.56	1.96

8 Appendix D: Per-site SDB RMSE at each screening stage

Tables D1 to D5 list, for each of the five SDB methods, the RMSE at each of the six sites and each of the seven stages of the stage-by-stage analysis described in Section 2.3.10. Stage 0 is the raw calibration and Stage 6 is the refined calibration. The summary of these values is given in Section 2.4.4.

Table D1. RMSE (m) of the SDB maps calibrated with the photons of each stage, for the Stumpf band ratio model.

Stage	Igloolik	Royal Geographical Society Islands	Cambridge Bay	Sandspit	Montague	Havre aux Maisons
S0	1.51	1.79	2.98	3.73	1.54	2.12
S1	1.51	1.79	2.98	3.73	1.54	2.12
S2	1.50	1.78	2.98	3.57	1.48	1.71
S3	1.51	1.80	3.00	3.57	1.48	1.72
S4	1.11	1.56	1.28	3.88	2.54	1.72
S5	1.01	1.07	1.42	1.90	3.37	2.96
S6	1.01	1.07	1.41	1.74	1.69	2.61

Table D2. RMSE (m) of the SDB maps calibrated with the photons of each stage, for the random forest model.

Stage	Igloolik	Royal Geographical Society Islands	Cambridge Bay	Sandspit	Montague	Havre aux Maisons
S0	1.18	1.65	2.41	3.89	3.31	2.22
S1	1.18	1.65	2.41	3.89	3.31	2.22
S2	1.19	1.65	2.41	3.73	3.26	1.89
S3	1.20	1.65	2.42	3.73	3.27	1.86
S4	1.07	1.37	1.76	3.76	3.56	2.01
S5	1.04	1.02	2.26	1.82	4.80	4.35
S6	1.01	1.03	2.28	1.78	2.18	3.74

Table D3. RMSE (m) of the SDB maps calibrated with the photons of each stage, for the gradient boosting model.

Stage	Igloolik	Royal Geographical Society Islands	Cambridge Bay	Sandspit	Montague	Havre aux Maisons
S0	1.41	1.64	2.48	4.20	3.64	2.70
S1	1.41	1.64	2.48	4.20	3.64	2.70
S2	1.41	1.65	2.50	3.96	3.40	2.03
S3	1.39	1.66	2.46	4.01	3.40	1.97
S4	1.11	1.39	1.76	3.90	4.12	2.11
S5	1.00	1.03	2.15	1.98	4.76	4.37
S6	0.97	1.03	2.06	1.84	2.43	3.55

Table D4. RMSE (m) of the SDB maps calibrated with the photons of each stage, for the nearest neighbor regression model.

Stage	Igloolik	Royal Geographical Society Islands	Cambridge Bay	Sandspit	Montague	Havre aux Maisons
S0	1.32	1.76	2.79	4.31	3.54	3.16
S1	1.32	1.76	2.79	4.31	3.54	3.16
S2	1.33	1.76	2.78	4.29	3.60	2.36
S3	1.33	1.74	2.78	4.29	3.64	2.30
S4	1.16	1.47	1.94	4.45	3.93	2.48
S5	1.05	1.13	2.47	1.90	5.23	3.10
S6	1.05	1.13	2.44	1.81	2.39	2.51

Table D5. RMSE (m) of the SDB maps calibrated with the photons of each stage, for the support vector regression model.

Stage	Igloolik	Royal Geographical Society Islands	Cambridge Bay	Sandspit	Montague	Havre aux Maisons
S0	1.14	1.74	2.69	2.36	1.68	1.65
S1	1.14	1.74	2.69	2.36	1.68	1.65
S2	1.13	1.74	2.69	2.25	1.69	1.67
S3	1.11	1.75	2.69	2.23	1.70	1.68
S4	0.90	1.41	1.53	2.42	1.29	1.60
S5	0.82	0.91	1.64	1.76	1.80	2.04
S6	0.82	0.94	1.64	1.78	1.40	2.01