Airborne Laser Scanning (ALS) Point Clouds of Trail Valley Creek, NWT, Canada (2018)

Metadata Documentation – 2021-06-08

Stephan Lange, Inge Grünberg, Katharina Anders,
Jörg Hartmann, Veit Helm, Julia Boike

GENERAL INFORMATION

This data set and documentation is a repetition of our previous work of ALS data and GNSS data aquired in 2016 (Anders et al. 2018a&b). Therefore, we keep the setup and the processing chain, we used the same software and code and we updated this document, the figures and the statistics of the actual data set.
Airborne Laser Scanning of the Trail Valley Creek (TVC) research watershed in Northwest Territories, Canada (68°44'25" N 133°29'36" W, Marsh et al. 2008) took place in August 2018 (2018-08-22). Point cloud data were acquired with a Riegl LMS-Q680i airborne laser scanner on board the Alfred Wegener Institute's POLAR-5 science aircraft. The location and extent of the data set are displayed in Fig. 1.

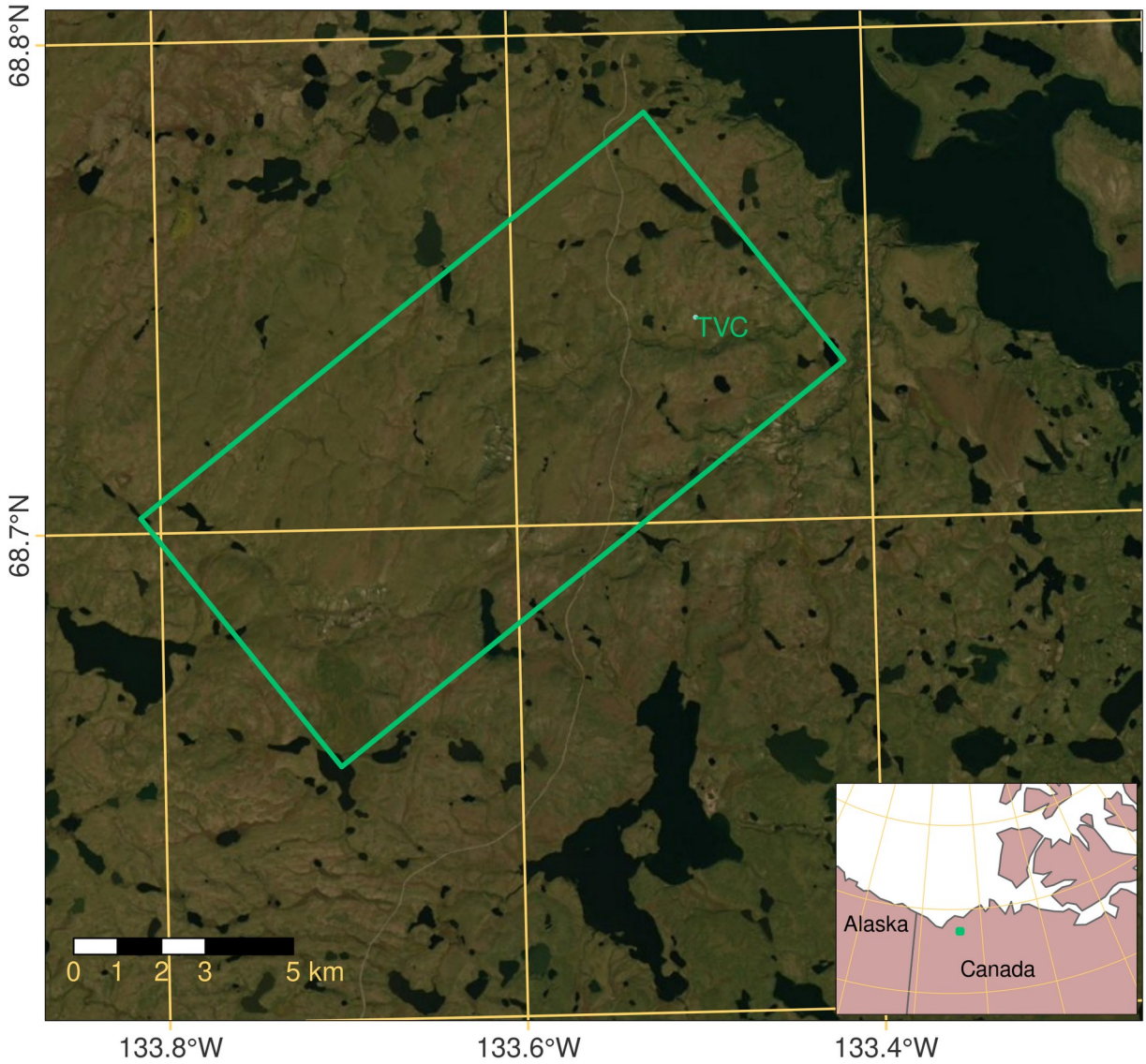


Figure 1: Location of the Trail Valley Creek (TVC) study area in Northwest Territories, Canada, and extent of the ALS data acquired in 2018 (approx. 7 x 15 km). Source of imagery: Esri, Maxar, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community.

Table 1: Instrument specifications for Riegl LMS-Q680i ALS.

Parameter	Specification
Accuracy	20 mm at 250 m scanning range
Precision	20 mm at 250 m scanning range
Laser beam divergence	≤ 0.5 mrad

The ALS instrument nominal specifications are listed in Tab. 1. For more information from the manufacturer see Riegl LMS (2012). Fig. 2 contains a map with the flight strips of the ALS data acquisition and the extent of the resulting dataset, which is provided in this publication. The ALS survey parameters are listed in Tab. 2.

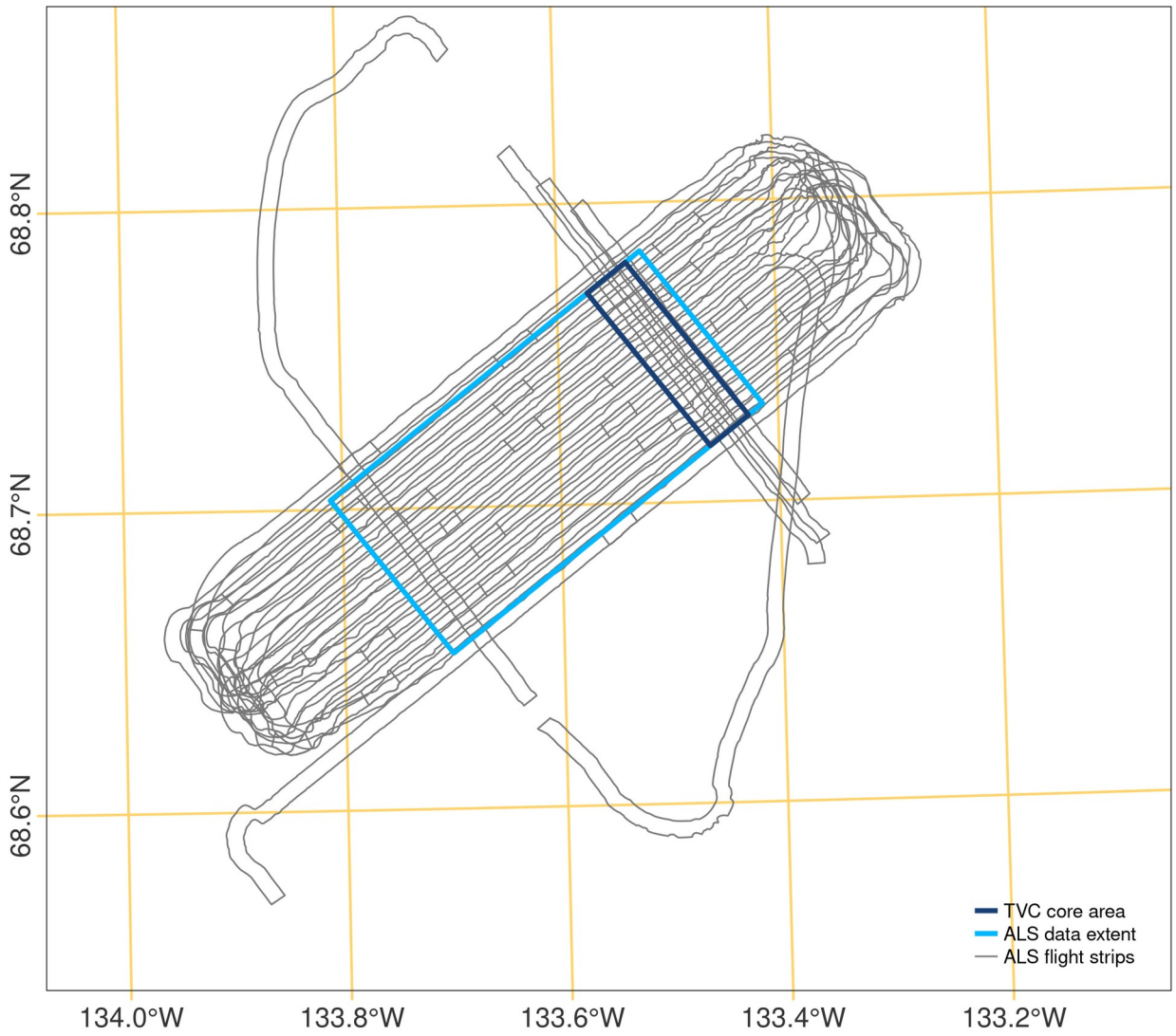


Figure 2: Flight strips of the ALS acquisition in August 2018 from which the final ALS point cloud of the displayed extent is generated. Perpendicular ALS flight strips were recorded over the core area of TVC (approx. 2 x 7 km).

Table 2: ALS survey parameters.

Instrument	Riegl LMS-Q680i
Pulse repetition rate	200 kHz
Flying altitude	2900 ft a.g.l.
Flying speed	~120 kn
Swath width	900 m
Attributes	Amplitude, Echo Width, Echo Information (echo number and count)

ALS POINT CLOUD PROCESSING

BASIC GNSS PROCESSING

The GNSS data were acquired in 20 Hz resolution with a Novatel OEM4 receiver on board the POLAR-5 aircraft. The GNSS trajectory was post-processed using precise ephemerides and the commercial software package Waypoint 8.6 (PPP [precise point positioning] processing).

BASIC ALS DATA PROCESSING

The raw data (sdf format) acquired with the Riegl LMS-Q680i is recorded as proprietary binary full-waveform format. The software RiAnalyze (version 6.2.1, Riegl LMS 2015) was used to extract a 3D point for each laser return with the attributes:

X, Y, Z, UTC, Amplitude, Echo Width, Target Type, Target, Target Count

UTC is the time stamp of the respective point record. Echo width is the width of the returned laser pulse. Target type, target, and target count are internal counters of the returned signal, for example when multiple echoes are returned per laser pulse in vegetated areas. The preprocessed data was output as binary data files in sdc format.
After this preprocessing, a software developed at AWI was used to combine the post-processed GNSS flight trajectory, inertial navigation system (INS) data and preprocessed ALS data to determine the georeferenced point cloud. Within this post-processing, the lever arm, attitude and squint angle corrections were applied. Squint angles were determined using a semi-automatic cross calibration approach. To run the calibration step, data sets of cross flights along runways, hangar buildings or rough topography are essential. For each campaign a set of squint angles are determined and used in the final post-processing. The output of the post-processing was a georeferenced point cloud in WGS84. Coordinates were transformed to UTM (Zone North8, EPSG 32608) and converted to ASCII files separated by the flight lines.
We used the georeferenced point cloud strips of each flight line with the attributes amplitude, echo width, and echo information of the internal counters (number of echo within laser pulse return and total count of echoes). The single point cloud strips were merged into one point cloud of the final extent (cf. Fig. 2). A point source ID for each ALS strip was added as attribute to the points in the merged point cloud dataset, so that the points can be distinguished by the single strips from which they originate.

CLASSIFICATION OF ECHO TYPES

Using the echo information of the acquired ALS dataset, we classified all points into four echo types regarding the respective laser returns from which they derived:

0: Single – the echo is the only echo returned for the laser pulse
1: First – the echo is the first of multiple laser returns

2: Intermediate – the echo is one of multiple laser returns (neither the first nor the last)
3: Last – the echo is the last of multiple laser returns

This attribute can be useful for preliminary filtering of the point cloud. We used this attribute in the derivation of terrain information, which is described in a subsequent section. Ground surface points would not contain first or intermediate echoes (echo types 1 and 2) as these imply that the laser pulse traveled on from this target (e.g. penetrating vegetation) so that the return is unlikely to derive from a solid surface.

STATISTICAL OUTLIER FILTER

Spatially isolated points were removed from the point cloud via a statistical outlier filter (Rusu and Cousins 2011), for which the number of neighbors was set to 12 and the standard deviation multiplier threshold to 2.0 (in comparism to 2016 with 6 numbers of neighbors and a standard deviation threshold of 1.0). Via this filter, 0.995% of points were removed from the point cloud as outliers.

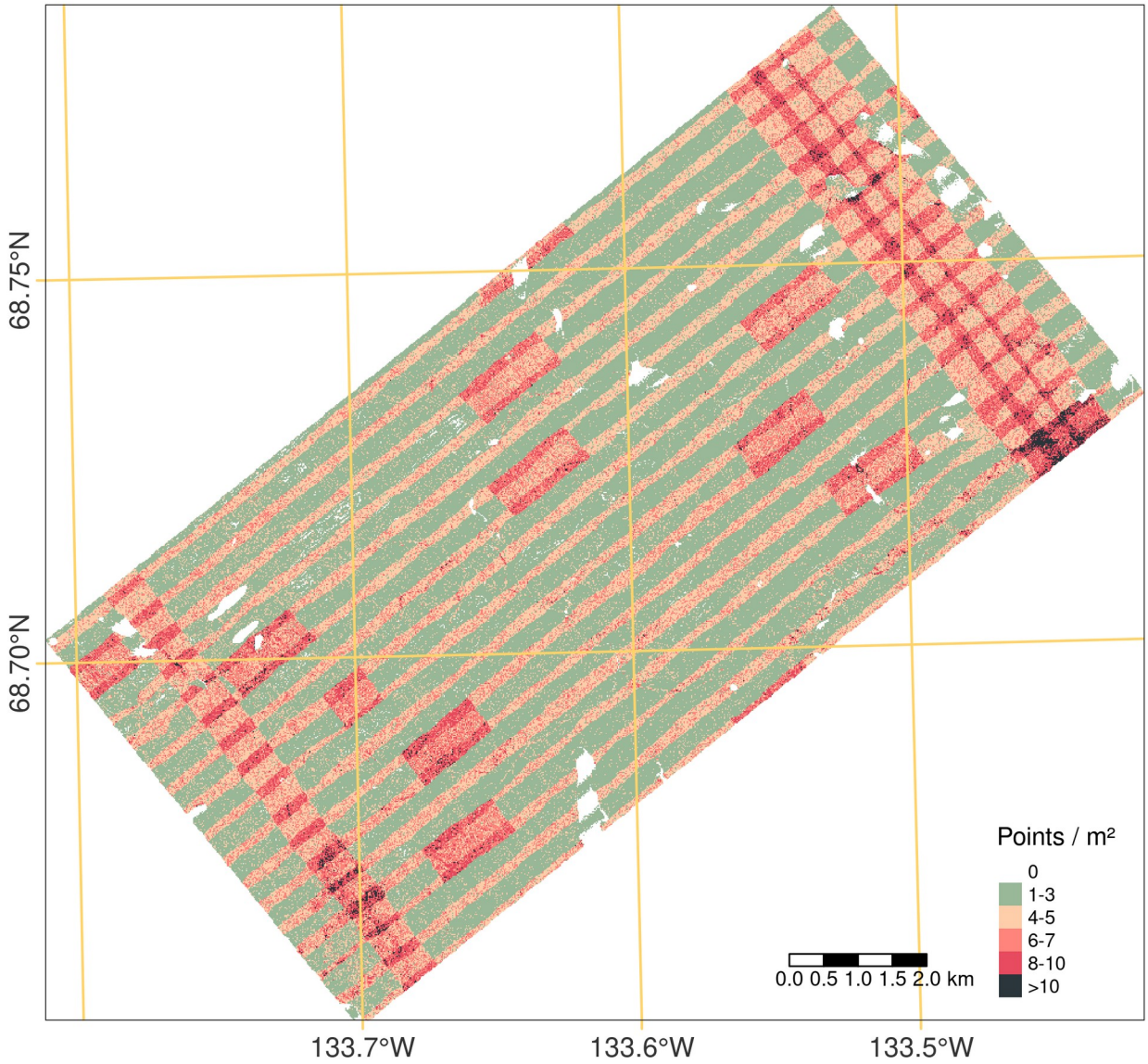


Figure 3: Point density of the processed ALS point cloud. Areas mostly have 0 points where water occurs (i.e. lakes), as the emitted laser pulse is absorbed and no signal is returned.

After these preprocessing steps, the ALS point cloud contains 420,329,915 points. The average point density is 8 points per square meter in the core area of TVC (approx. 2 x 7 km, acquired with perpendicular ALS flight strips, cf. Fig. 2) and 5 points per square meter in the remaining area (Fig. 3).

DERIVATION OF TERRAIN INFORMATION AND DTM GENERATION

The processed 3D ALS point cloud served as basis to derive the terrain elevation and, subsequently, vegetation heights in the study area. The terrain is defined as the boundary surface between the solid ground and the air. Terrain elevation is commonly represented in a Digital Terrain Model (DTM) and can be reconstructed from ALS point measurements of the terrain (Pfeifer & Mandlburger 2008).

GROUND SURFACE CLASSIFICATION AND INTERPOLATION

The vegetation cover of the tundra landscape in the study area influences the geometric representation of the terrain in the ALS point cloud in two ways: 1) Where vegetation is low, the laser return might not be separated correctly into vegetation returns and the (final) ground surface return. 2) Where vegetation is dense, the laser beam might not (fully) penetrate so that no echo is returned from the ground surface and the lowest point in the local area would be a vegetation echo (Mücke et al. 2010). Given these two aspects, the geometric properties of points are not sufficient indicators of a point representing the terrain or vegetation. Therefore, Mücke et al. (2010) developed the approach of terrain echo probability assignment, which assigns probabilities to each 3D data point, indicating if the point is likely to represent the terrain. These probabilities are determined using full-waveform observables, by modelling the distribution of echo widths dependent on the corresponding amplitude values. The method is described in detail in Mücke et al. (2010). We implemented the terrain echo probability assignment in a Python script and all other related files at AWI gitlab (see Git section).

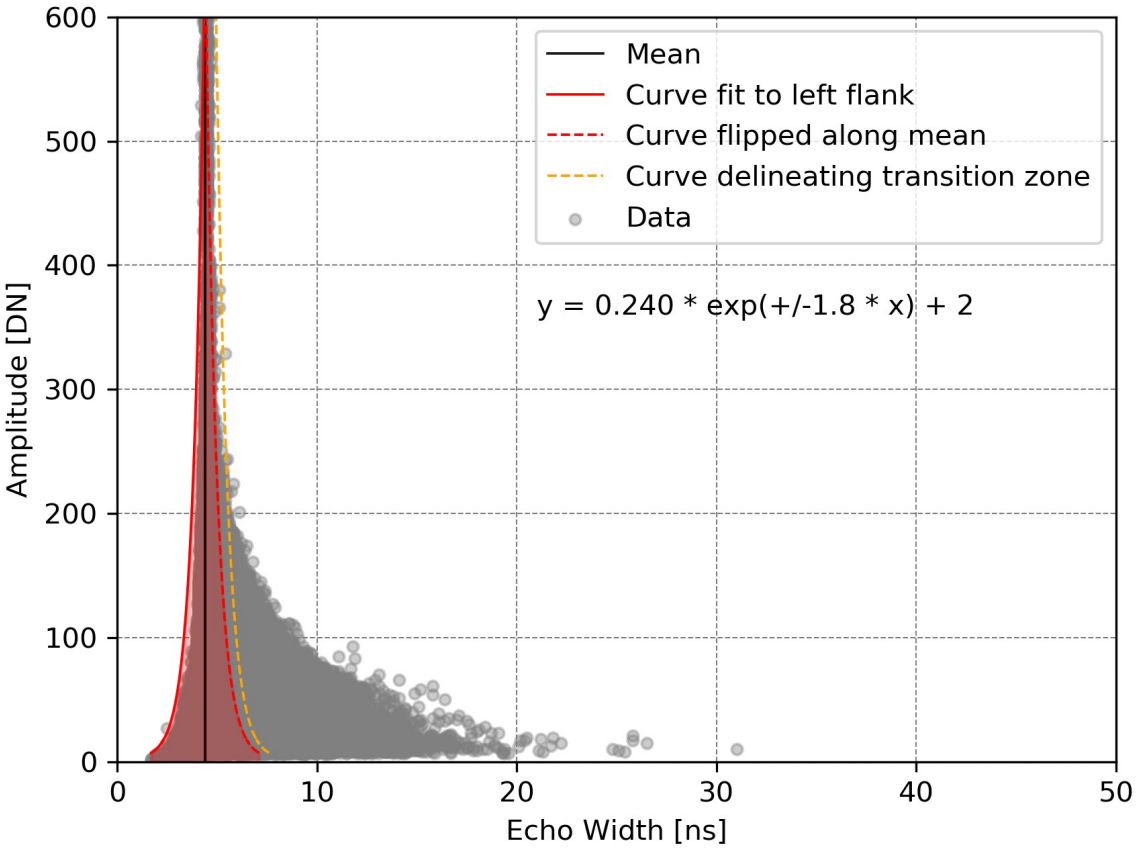
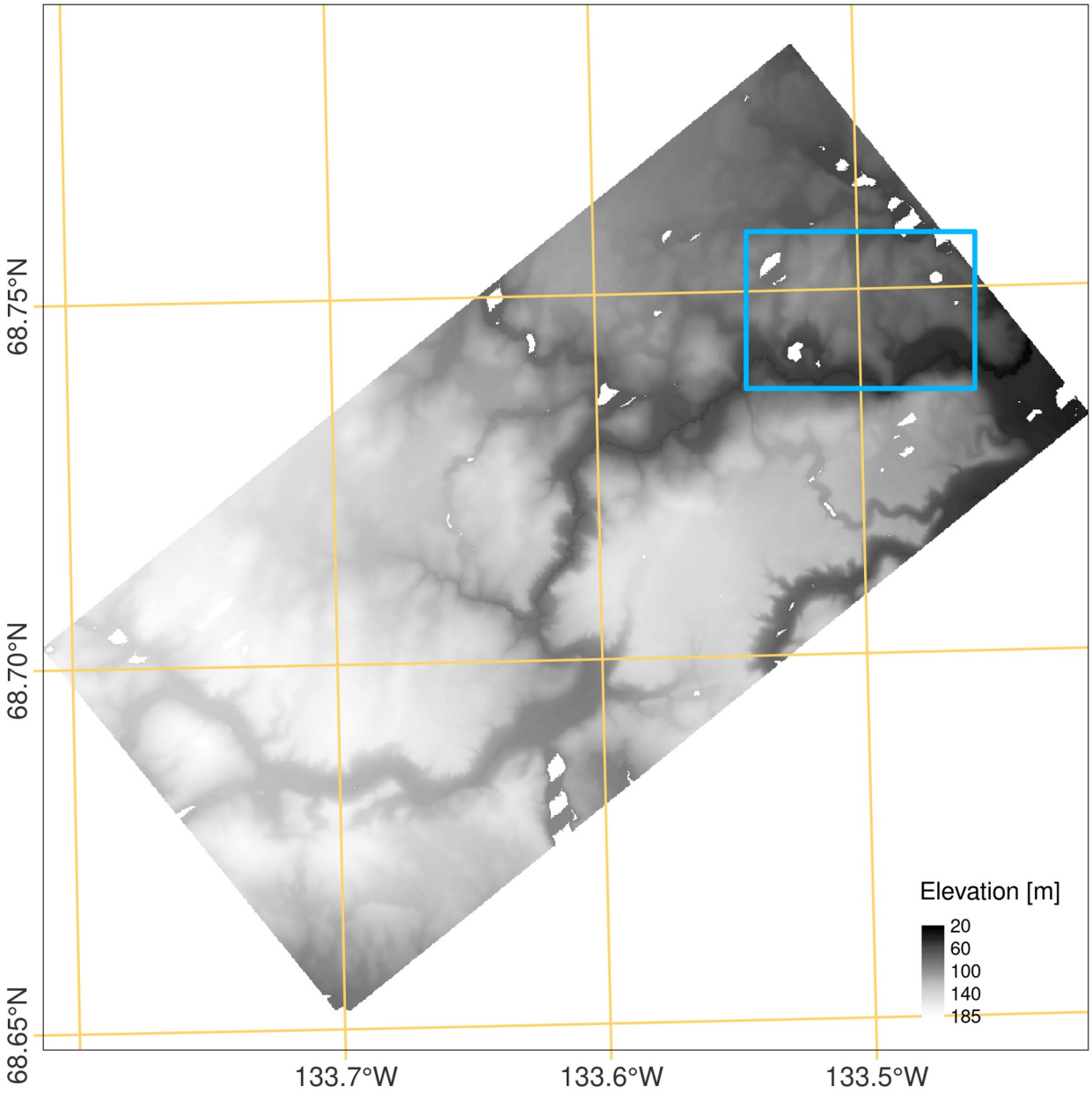


Figure 4: Echo width plotted against amplitude values with black line indicating the mean of echo width. Beneath the legend we provide the values and formula of the fitted curves.

For the terrain probability assignment, we use only single and last returns (defined by echo types 0 and 3) of the 3D ALS point cloud. Fig. 4 shows the result of the distribution function computed for the dataset using the full-waveform observables.
Using robust surface interpolation (cf. Pfeifer & Mandlburger 2008), the 3D ALS point cloud is finally classified into ground surface and non-ground surface points. The method considers the assigned terrain probabilities as a-priori weights in the ground surface classification. In the process, the surface is iteratively re-calculated based on (updated) point weightings based on the vertical point distances from the current surface until the surface sufficiently approximates the ground surface. We use the modular program system OPALS (Orientation and Processing of Airborne Laser Scanning Data, Pfeifer et al. 2014) for this processing step (cf. TU Wien 2017). Ground points resulting from this computation are labelled according to the LAS specification (ASPRS 2011) with the attributes:

1: Unclassified
2: Ground

The DTM was saved as GeoTIFF with 1.0 m cell size using the OPALS module Grid (TU Wien 2016) based on all points classified as ground points in a robust moving plane interpolation with a search radius of 7.5 m and a maximum standard deviation of the surface interpolation of 0.5 m. Gaps in the computed DTM were filled with mean elevation values in the 3x3 m neighborhood of empty pixels. Pixels within larger gaps which were not filled in this step contain no values (NoData) in the final DTM (Fig. 5).


68.75°N
68.70°N
68.65°N
133.7°W
133.6°W
133.5°W
Elevation [m]
20
60
100
140
185

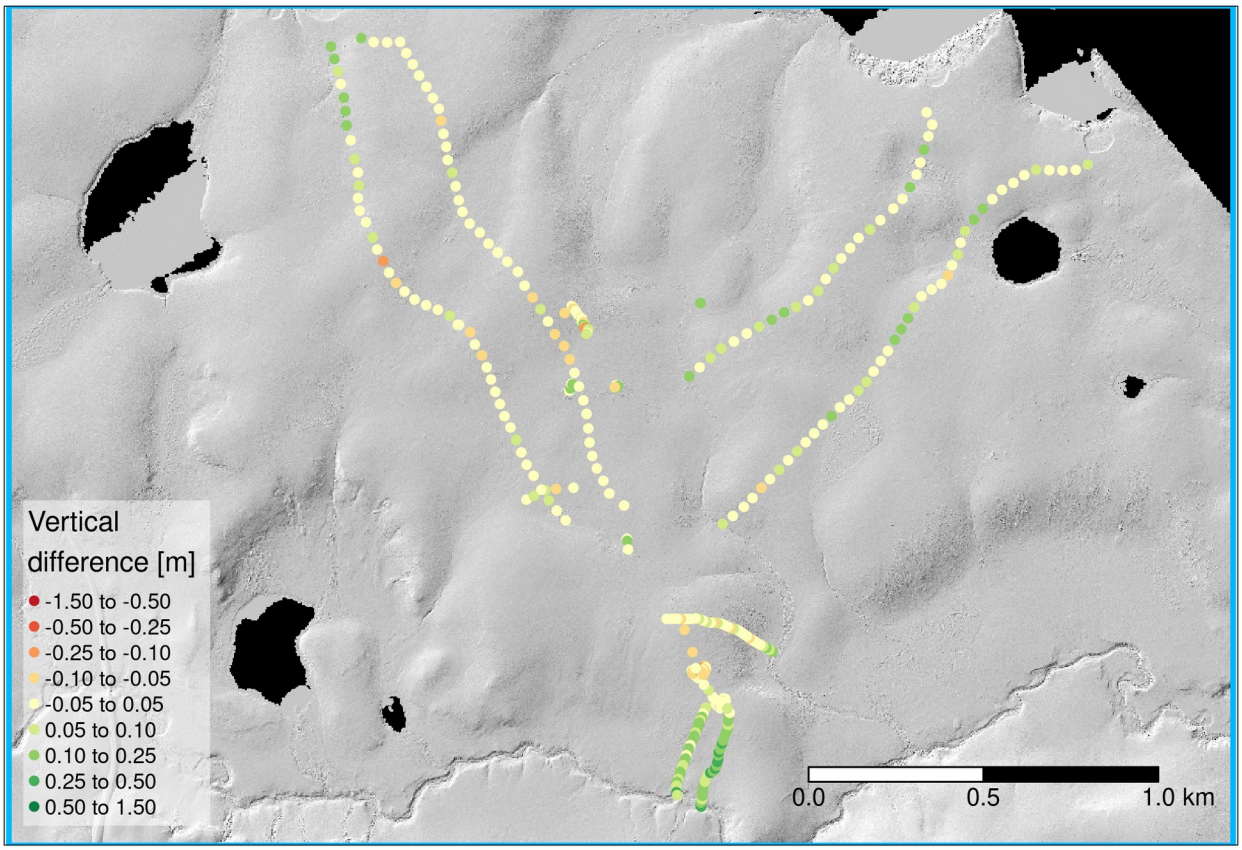

Vertical
difference [m]
-1.50 to -0.50
-0.50 to -0.25
-0.25 to -0.10
-0.10 to -0.05
-0.05 to 0.05
0.05 to 0.10
0.10 to 0.25
0.25 to 0.50
0.50 to 1.50
0.0
0.5
1.0 km

Figure 5: DTM derived from ALS ground points with close-up map extracts of GNSS point measurements in the study area (Extract background shows a hillshade of the DTM). The GNSS points are used for the DTM accuracy assessment and therefore colored by vertical difference to the corresponding DTM pixel. The extract shows measurements in the core study area, with a larger range of difference values which indicates the influence of the ground surface vegetation cover on ground point acquisition by ALS and resulting the DTM accuracy.

DTM ACCURACY ASSESSMENT

The accuracy of the DTM was tested using GNSS measurements of the ground surface acquired in the frame of a field campaign in August 2018 (cf. Fig. 5). The GNSS dataset is published in Pangaea with detailed information on the sensor system and field surveys (cf. Lange et al. 2020). To match the georeferencing of the ALS dataset, the PPP-corrected coordinates of the GNSS data were used. We exclude GNSS measurements which were not transect or repetition measurements. The higher values of vertical difference in both transects in the south in figure 5 due to very dense vegetation.

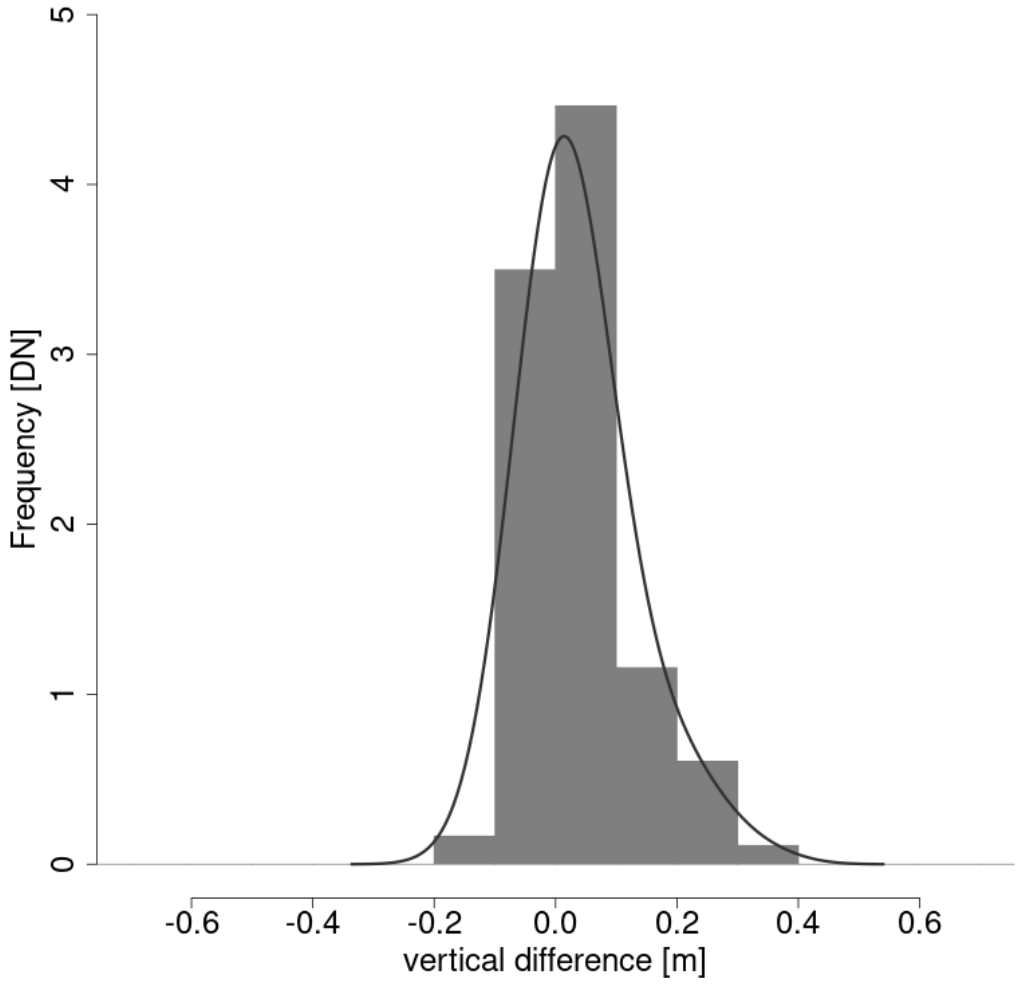


Figure 6: Distribution of vertical distances of the ALS-based DTM to GNSS point measurements in the study area.

Fig. 6 shows the distribution of difference values from the ALS-based DTM to the 363 GNSS-based reference measurements. Tab. 3 shows the statistics of differences. According to the mean value of differences, the accuracy of the DTM is 0.03 m. The precision of the DTM is 0.08 m according to the standard deviation of differences.
Please note that the accuracy of the DTM shows high spatial variability (cf. Fig. 6), depending on the characteristics of surface cover (i.e. height and density of vegetation). Errors are large where the ground surface was not recorded in the ALS campaign. The previous section on “Ground Surface Classification and Interpolation” gives some considerations on this aspect of ALS data acquisition and ground surface representation in the point cloud.

Table 3: Statistics of difference values of the ALS-based DTM to GNSS point measurements in the study area.

Minimum	-0.159
Maximum	0.364
Mean	0.038
Median	0.022
Standard deviation	0.086
Root-mean-square (RMS)	0.094

DERIVATION OF VEGETATION HEIGHT

The extent of the ALS dataset does not contain man-made objects except the bare surface of the highway. Therefore, we considered all non-ground surface points to represent vegetation. Vegetation height values in the study area were hence derived by subtracting the terrain elevation (i.e. DTM pixel values) from the non-ground surface points in the ALS point cloud. Two rasters of vegetation height were generated with a cell size of 1.0 m by using (1) the mean and (2) the maximum values of vegetation height per cell.

Fig. 7 shows the resulting vegetation height raster. We explicitly point to the sampling effect of the ALS data acquisition, which becomes visible in the raster. Vegetation height values are larger where multiple strips overlap, and lower where only two strips overlap. This effect relates to the general property of ALS surveying, that the acquisition of an area from multiple perspectives increases the likelihood of penetrating dense vegetation and recording points of the ground surface. When using the vegetation height data, this aspect needs to be considered in the analysis and interpretation of different parts of the dataset. For the purpose of easier traceability, we therefore provide a raster (GeoTIFF) with the number of flight strips covering a pixel (as shown in Fig. 8).

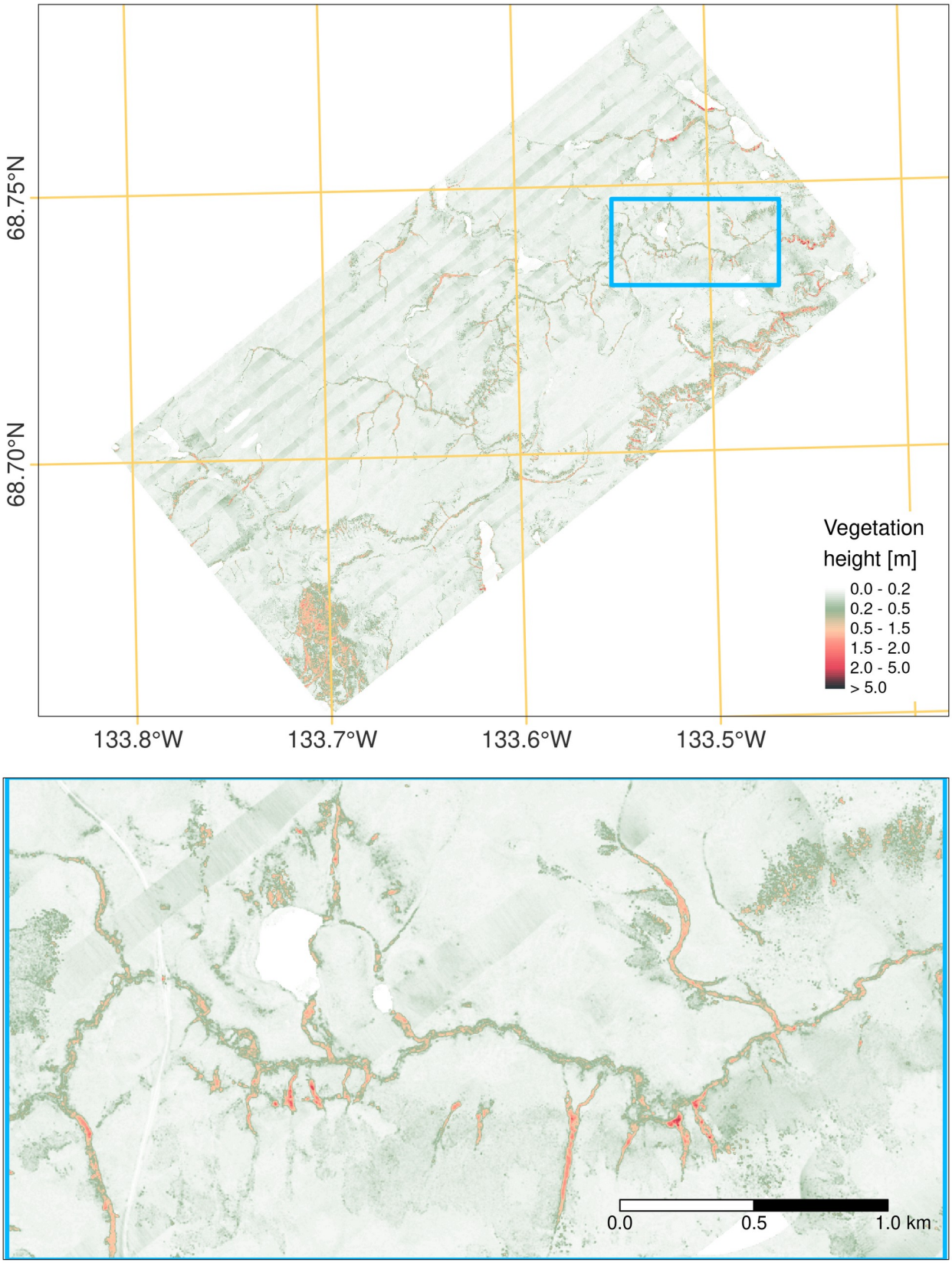


Figure 7: Maximum vegetation height raster derived from the ALS point cloud relative height of points compared to the DTM.

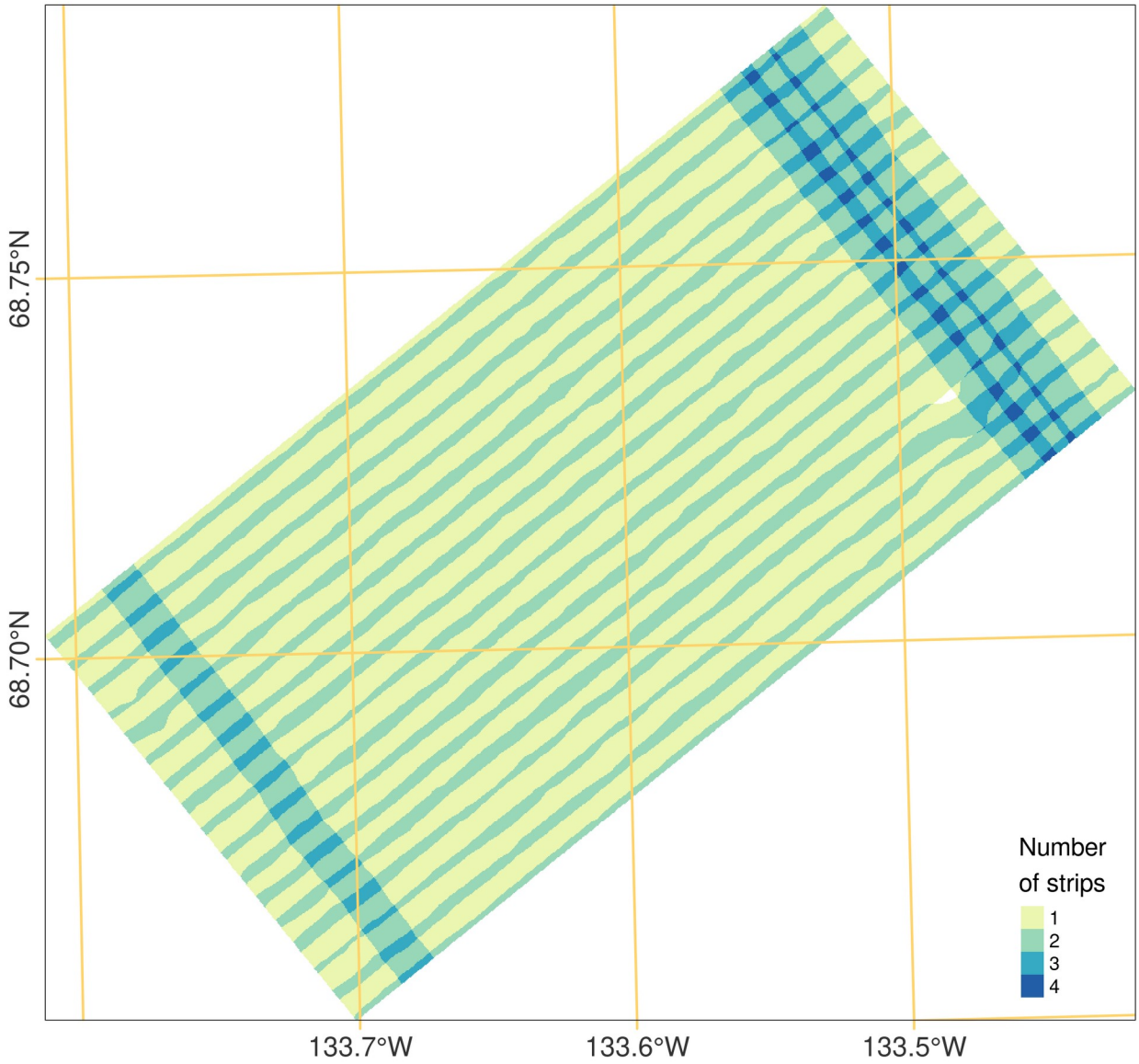


Figure 8: Number of overlapping flight strips per raster pixel (cell size of 1.0 m, corresponding to DTM and vegetation height raster).

Using the relative height points in the ALS point cloud, we labelled respective points according to the LAS specification (ASPRS 2011) with the attributes:

3: Low vegetation – 0.5 m < relative height ≤ 2.0 m
4: Medium vegetation – 2.0 m < relative height ≤ 5.0 m
5: High vegetation – 5.0 m < relative height

DATA FORMAT

Coordinate reference system:

Coordinates are given in UTM WGS84 Zone North8 (EPSG 32608). Elevations are given above the WGS84 ellipsoid.

Units:

Meters

File format:

Point cloud as laz-file
Rasters as GeoTIFF

Point cloud attributes:

X **Y** **Z**	3D coordinate of the laser measurement (WGS84 UTM 8N, EPSG 32608, height above WGS84 ellipsoid) [m]
Amplitude	Amplitude of the returned laser intensity (LiDAR backscatter information) [dB]
EchoWidth	Width of the returned laser signal (LiDAR backscatter information) [ns]
EchoType	Classification of echo type for the laser pulse from which the echo was returned [DN] EchoType = {0: single, 1: first, 2: intermediate, 3: last}
Terrain Probability	Terrain probability value according to the distribution function derived from the full-waveform observables amplitude and echo width (cf. section *Ground Surface Classification and Interpolation*) [Decimal Fraction]
Relative Height	Vertical height of point above ground surface elevation (based on corresponding DTM pixel value) [m]
Class	Classification according to las specification (cf. ASPRS 2011) [DN] Class = {1: Unclassified, 2: Ground, 3: Low Vegetation, 4: Medium Vegetation, 5: High Vegetation} *Note: The ASPRS classifies vegetation as low vegetation if 0.5 m < height ≤ 2.0 m, medium vegetation if 2.0 m < height ≤ 5.0 m, and high vegetation if 5.0 m < height.*
Strip ID	ID of the ALS flight strip from which the laser measurement point originates [DN]

Git

https://gitlab.awi.de/sparcs/als-data-processing/als/als_tvc_2018_b
We provide all documents and scripts we used in this processing chain. See *02_process.bat* to follow all steps. All figures created by *TVC_ALS_201809_stats_and_plots.R* with R, figure 4 is done by *terrain Probability2pi.py* with python.

Final note

No further processing (e.g., correction, classification or filtering) was applied to the data other than the steps described above.

REFERENCES

Anders, K., Antonova, S., Beck, I., Boike, J., Höfle, B., Langer, M., Marsh, P., & Marx, S. (2018 a). Multisensor ground-based measurements of the permafrost thaw subsidence in the Trail Valley Creek, NWT, Canada, 2015-2016. Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research, Bremerhaven, PANGAEA, https://doi.org/10.1594/PANGAEA.888566.

Anders, Katharina; Antonova, Sofia; Boike, Julia; Gehrmann, Martin; Hartmann, Jörg; Helm, Veit; Höfle, Bernhard; Marsh, Philip; Marx, Sabrina; Sachs, Torsten (2018 b): Airborne Laser Scanning (ALS) Point Clouds of Trail Valley Creek, NWT, Canada (2016). PANGAEA, https://doi.org/10.1594/PANGAEA.894884.

ASPRS (2011). LAS SPECIFICATION - VERSION 1.4. URL: https://www.asprs.org/wp-content/uploads/2010/12/LAS_1_4_r13.pdf (2021-06-08).

Lange, Stephan; Cable, William L; Grünberg, Inge; Boike, Julia (2020): GNSS Measurements Inuvik-Tuktoyaktuk-Highway (ITH) and Trail Valley Creek (TVC) 2018. PANGAEA, https://doi.org/10.1594/PANGAEA.918649.

Marsh, P., Pomeroy, J., Pohl, S., Quinton, W., Onclin, C., Russell, M., Neumann, N., Pietroniro, A., Davison, B., & McCartney, S. (2008). Snowmelt Processes and Runoff at the Arctic Treeline: Ten Years of MAGS Research. In M.-k. Woo (Ed.), Cold Region Atmospheric and Hydrologic Studies: The Mackenzie GEWEX Experience (Vol. 2: Hydrologic Processes). Berlin, Heidelberg: Springer. pp. 97-123. doi: 10.1007/978-3-540-75136-6_6.

Mücke, W., Briese, C. & Hollaus, M. (2010). Terrain Echo Probability Assignment based on Full-Waveform Airborne Laser Scanning Observables. ISPRS TC VII Symposium - 100 Years ISPRS, Vienna, Austria. pp. 157-162.

Pfeifer, N. & Mandlburger, G. (2008): Filtering and DTM Generation. In: Shan, J. & Toth, C. (Eds.). Topographic Laser Ranging and Scanning: Principles and Processing. CRC Press. pp. 307-333.

Pfeifer, N., Mandlburger, G., Otepka, J., & Karel, W. (2014). OPALS - A framework for Airborne Laser Scanning data analysis. Computers, Environment and Urban Systems, 45, pp. 125-136. doi:

10.1016/j.compenvurbsys.2013.11.002.

Riegl LMS (2012). Long-Range Airborne Laser Scanner for Full Waveform Analysis: RIEGL LMS-Q680i [Datasheet]. URL: http://www.riegl.com/uploads/tx_pxpriegldownloads/10_Data Sheet_LMS-Q680i_28-09-2012_01.pdf (2021-06-08).

Riegl LMS (2015). Full Waveform Analysis Software: RiANALYZE [Datasheet]. URL: http://www.rieglusa.com/pdf/rianalyze-datasheet-new.pdf (2021-06-08).

TU Wien (2017). OPALS: Module Grid. URL: http://geo.tuwien.ac.at/opals/html/ModuleGrid.html (2021-06-08).

TU Wien (2017). OPALS: Module RobFilter. URL: http://geo.tuwien.ac.at/opals/html/ModuleRobFilter.html (2021-06-08).