Tundra Snow Depth Retrievals from Wideband Radar Observations in Trail Valley Creek

WEI WANG[1], RICHARD KELLY[1], AND JOSHUA KING[2]

ABSTRACT

Recent experiments have explored the potential of using airborne wide-band snow radar observations to estimate snow accumulation. Most if not all experiments have focused on sea ice applications with a series of campaigns conducted by Operation Ice Bridge (OIB) successfully demonstrating the applicability of wideband observations (2-8 GHz) to estimate snow on sea ice. Most applications of the wideband approach for terrestrial snow applications have been conducted using ground-based systems. However, in 2019 a wideband radar system (2-18 GHz) was flown over Trail Valley Creek in the North West Territories, Canada coinciding with ground-based field measurements of snow depth and SWE properties from a range of measurement systems. This paper describes the application of the using the interface-based pulse peakiness snow depth retrieval method to estimate snow on land. The approach was tested on the six major vegetation types present at the study area (white and black spruce trees, tall shrub, riparian shrub, dwarf shrub, tussock, and lichen). Snow depth derived from Airborne Laser Scanner (ALS) point clouds was used as the reference for snow depth retrieval. To address the differences between snow on land and on sea ice, a recalibration of the algorithm parameters was completed as well as introducing additional measures to screen abnormal observations with a low Signal to Noise Ratio. It was concluded that the principles behind the pulse peakiness approach was generally valid for tundra snow depth estimates at study sites characterized by smooth and low slope surfaces and with low stand vegetation. The presence of surface vegetation and increased surface roughness led to increased complexity of radar waveforms, which resulted in increased snow depth retrieval uncertainty.

INTRODUCTION

Seasonal snow is an essential component of the hydrologic cycle, where changes in snow cover and distribution can have significant impacts on precipitation, wildlife, and water supply. In addition, variations in snow cover can serve as an indicator for climate change (Croce *et al*., 2018). A timely accurate estimate of snow depth or snow water equivalent (SWE) can be used for hydrology and climate modeling, managing local water supply, as well as presenting a quantifiable parameter that responds to climate change.

It is because of the necessity of snow monitoring that the methods for snow depth / SWE estimates have been developed throughout the years. While field measurements or records from weather stations are known to be reliable, they are localized in representativeness and so to cover the vast areas around the globe, remote sensing technology including passive and active microwave sensors have been utilized along with numerical models.

[1] Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON, Canada
[2] Environment and Climate Change Canada, Toronto, ON, Canada

More specifically for snow depth or SWE retrieval using active sensors, there have been experimental observation methods developed utilizing radar systems at Ku or X band (Yueh *et al*., 2009; Oveisgharan *et al*., 2020; Leinss *et al*., 2015). As explained by Leinss *et al*. (2015) and Oveisgharan *et al*. (2020), it is possible to retrieve SWE with the interferometric phase differences between two measurements at specific frequencies, especially at L-band. Cui *et al*. (2016) proposed using Synthetic Aperture Radar (SAR) observations at Ku and X band coupled with a backscatter forward model to estimate the optical thickness of the snowpack. By contrast, Operation Icebridge (OIB) deployed an experimental wide band radar with a frequency range of 2-8 GHz which differs from the radar systems used by Leinss *et al*. (2015) and Cui *et al*. (2016) in that the frequency of the transmitted signal is constantly modulated. In addition, various other FMCW radar systems have also been used for SWE retrieval and snow stratigraphy measurements (Marshall and Koh, 2008). However, as those mostly ground-based FMCW radar systems are not the focus of this paper, a wide band radar here is specifically defined as an airborne FMCW radar capable of operating in a wide bandwidth similar to the ones used in OIB and AWI Icebird campaigns.

Currently, most snow depth retrieval algorithms using data collected by a wide band radar are based on obtaining two crucial interfaces, air-snow interface, and snow-ice / snow-ground interface (Kwok *et al*., 2017). As OIB which provided a large percentage of the data used for developing various retrieval algorithms is focused on offering detailed information about polar ice, most of these retrieval methods are for snow on sea ice. According to Kwok *et al*. (2017), the difference amongst these algorithms mainly lies in how the two interfaces are obtained and how they are optimized for a specific radar system. Kurtz and Farrell (2011) and Kurtz *et al*. (2013) proposed an interface identifying method that combined searching for local maximum and empirical thresholding, which was adopted by the National Snow and Ice Data Center (NSIDC). Later on, a waveform fitting method was developed by Kurtz *et al*. (2014) where a physical model was established to simulate wide band radar waveforms over sea ice and the two interfaces were obtained by fitting data to the model. In addition, a retrieval based on the gradient of radar intensity return was adapted from Koenig *et al*. (2016) and tested in Kwok *et al*. (2017). In comparison, Newman *et al*. (2014) took a slightly different approach that avoided the need for fixed sets of thresholds used by several previously mentioned methods. As demonstrated by Newman *et al*. (2014) and Kwok *et al*. (2017), a Haar-CWT was utilized to detect abrupt changes within the waveform and those identified abrupt transitions were used to determine the two interfaces under specific criteria. To cope with the gradual improvement of the radar system used in different campaigns in OIB, a simplified version of the algorithm from Kwok and Maksym (2014) was developed where interfaces were located by measuring their deviation from their neighbors. Kwok *et al*. (2017) concluded that all methods described above were capable of reproducing the high snow depth pattern expected in the study area, but the performance of each individual method when compared with field measurement differs. More specifically, according to Kwok *et al*. (2017), when comparing the retrieval snow depth from these five methods with field data from Eureka, their corresponding mean bias are -0.8 cm (Kurtz *et al*., 2013), -5.7 cm (Kurtz *et al*., 2014), 1.3 cm (Koenig *et al*., 2016), 2.0 cm (Newman *et al*., 2004), and -2.2 cm (Kwok and Maksym, 2014)

Since the start of OIB, most algorithms developed including the ones described above are for snow on sea ice because of several possible reasons. First, the difference between dielectric properties of snow and frozen soil is generally weaker compared with snow and sea ice; second, the presence of vegetation presents a sizable challenge. When there is vegetation buried in the snowpack, it would likely increase the complexity of the snowpack while if the vegetation is emerging from the snowpack, it could interfere with radar waveforms before they reach the snowpack. Thirdly, terrain features on land like steep slopes and hummocks could further increase the complexity of the waveforms. In addition, the frequency ranges of the radar systems used by these algorithms are often not wide enough to cover shallow snow. However, Icebird 2019 Winter collected data over the Trail Valley Creek research station using a wide band radar with a frequency range of 2-18 GHz and this provides an opportunity to examine waveforms for snow over various land terrain features including different types of vegetation as well as hummocks and to determine whether principles behind snow on sea ice algorithms can be used for snow on land.

STUDY AREA

The study area for this experiment is based at the Trail Valley Creek (TVC) research station, located at the northern boundary of the tundra-boreal forest ecotone, roughly 50 km north of Inuvik, Northwest Territories (Pohl *et al*., 2005; Pomeroy *et al*., 1998). The TVC research basin is within the Inuvialuit Settlement Region near the Inuvik-Tuktoyaktuk Highway. The watershed drains approximately 58 square kilometers into the Husky Lakes Estuary network. In addition, the basin itself is underlain with a continuous ice-rich permafrost.

Topographically, the area is dominated by gently rolling hills (Pohl *et al*., 2005). A major terrain feature of the study area is the hummocks, which are defined as sub-meter to meter-scale circular to oval mounds (Verret *et al*., 2019). The existence of hummocks leads to a relatively large amount of the micro terrain that is below the spatial resolution of most field measurements. More complex terrain is present in the southern margins of the study area, in proximity to forests.

According to Boike and Grünberg (2019) and Palmer *et al*. (2012), the study area contains 6 main vegetation types including trees (white and black spruce (Picea glauca and Picea mariana), tall shrubs, riparian shrubs, dwarf shrubs, tussocks, and lichens. Tussocks, lichens and dwarf shrubs are the dominant vegetation types while trees and tall shrubs have a relatively sparse distribution. For tall shrubs particularly, there are often other low height vegetation in between two tall shrubs (Boike and Grünberg, 2019). The landscape also features several lakes as shown in Figure 1 with a number of streams and rivers running through the study area.

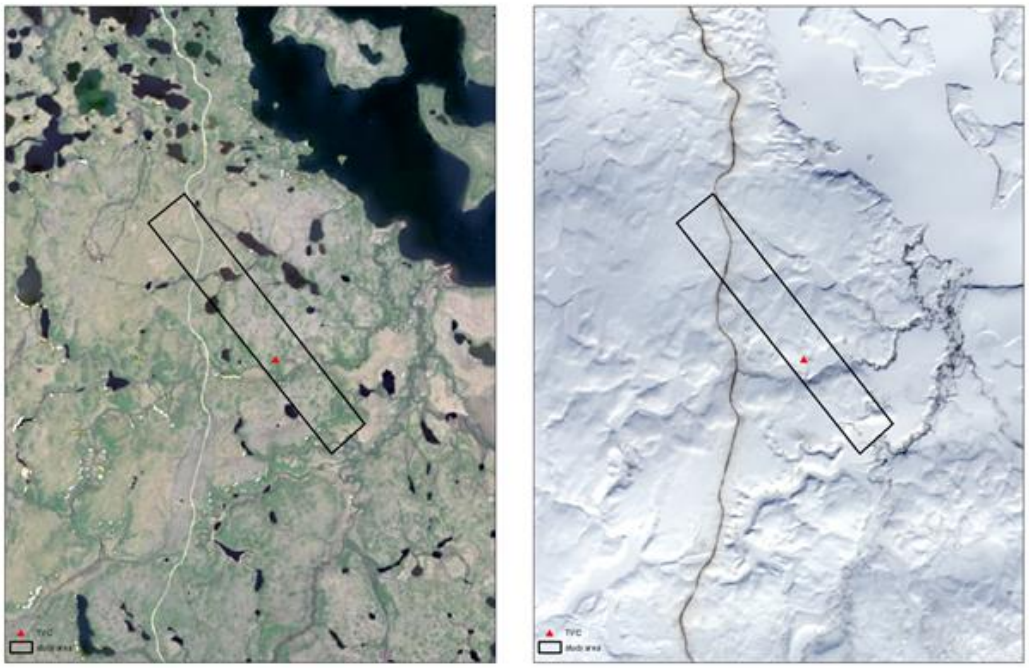

Figure 1. Landsat 8 images of study area; left is from 2019.06.18 and right is from 2019.04.15. Red triangle represents the location of TVC research station.

DATASETS

Snow Radar Data and Airborne Laser Scanning (ALS) Data from Icebird 2019 Winter

The Icebird 2019 Winter campaign was aimed at conducting sea ice surveys over different ice regimes to observe, analyze, and estimate sea ice thickness in addition to examining regional snow distribution (Hendricks *et al*., 2019). An airborne wide band snow radar system observed snow on

sea ice and an ALS system observed surface roughness features (Hendricks *et al*., 2019). Five field observation periods were planned for the 2019 winter campaign, in which the fourth period, conducted between April 6th and April 11th, 2019, was when the snow radar and ALS data were collected and are used in this study. The particular study date that covered the TVC region was April 10th, 2019.

The onboard ALS system was a Riegl VQ-580 (Hendricks *et al*., 2019). The ALS point cloud data used in this study was level 1b with the included metadata consisting of geolocation and positioning information. With AWI toolbox (Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, 2019), it was possible to filter the large point clouds by time period and narrowed it down to point clouds that were within the TVC study area.

The snow radar system, mounted on the Polar-6 Basler BT-67, aircraft was set to operate in Frequency-Modulated and Continuous Wave (FMCW) mode for low-altitude missions. The aircraft flew at an elevation of approximately 61 m (200 ft) and 488 m (1600 ft) when using the radar for observation (Hendricks *et al*., 2019). The wide band radar operated at a frequency between 2 and 18 GHz with 2 receiving channels (horizontal and vertical). Four combinations of polarization observations were achieved, VV, VH, HV, HH. The theoretical range resolution within the snowpack, assuming a snow density of 300 kg m^3, was 1.14 cm. At 61 m (200 ft) height, the radar had an across-track footprint of 2.1 m and an along-track footprint of 2.0 m while at 488 m (1600 ft) height, the two respective footprints were 6.0 m and 10.8 m. According to Hendricks *et al*. (2019), initial processing of the raw wide band radar was done using CRESIS Matlab Toolbox, followed by coherent noise removal and deconvolution of echograms (Julita *et al*., 2021).

After the pre-processing, the data was separated in to 24 continuous segments of the flight path in a Matlab file format. Within the TVC study area, elements from 6 different flight paths were used (path numbers 1, 2, 6, 7, 11, 12, at VV polarization). The reason for using VV polarization instead of HH is that VV waveforms are generally of higher quality, for example, less noise and interference, and are more suited for identifying peaks. Using PySnowRadar (King *et al*., 2020), the files were read into python and each waveform was presented as a 1x14050 numpy array, with each element of the array being an intensity return. Each waveform data record included the impulse data, longitude, latitude, and the elevation of the aircraft. Figure 2 shows the flight line locations. Waveforms over water bodies (lakes and rivers or streams) were removed from the analysis leaving a total of 3128 radar data locations.

ALS Point Clouds From 2016

A snow-free ALS data set for the study region was acquired on September 13th, 2016, using a Riegl LMS-Q680i. The reason for needing this flight was to provide a snow-off DEM for the development of a lidar snow depth product when combined with a snow-on DEM from 2019. Details of data acquisition can be found in Anders *et al*. (2016). Three products from Anders *et al*. (2016) were used in this study including a digital terrain model (DTM), maximum vegetation height map and mean vegetation height map, as is shown in Figure 3. All three data have a spatial resolution of 1 m and were later clipped to the study area. The 2016 DTM was used as a snow-off reference for the 2019 snow-on DTM to generate a snow depth map. Both vegetation height maps were used in conjunction with the TVC vegetation map to provide a guideline for typical heights of different vegetation types.

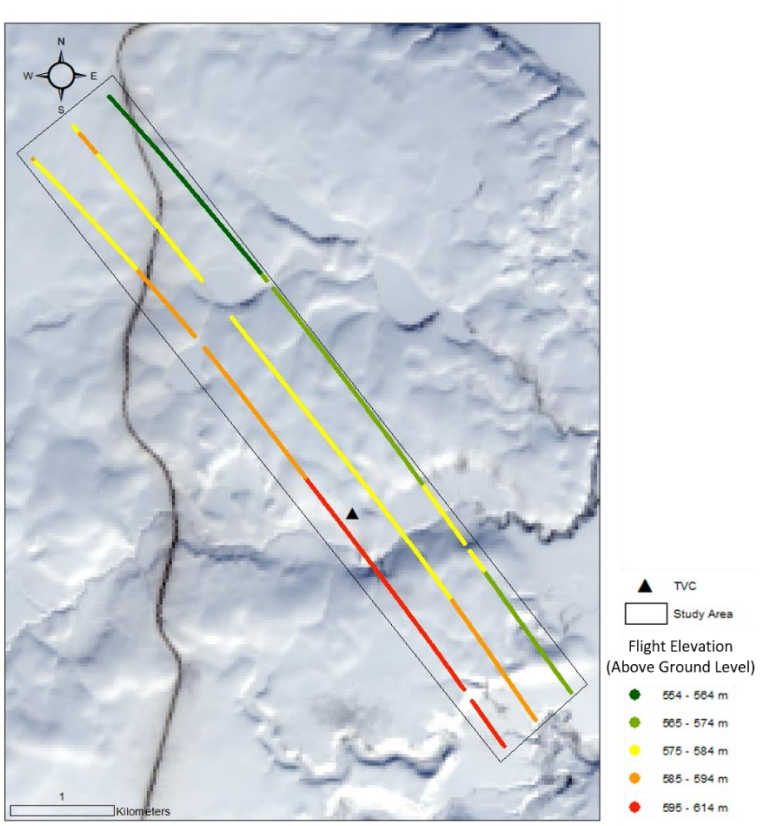

Figure 2. Location of the used radar data points colored by flight elevation (Landsat 8 from 2019.04.15 as background).

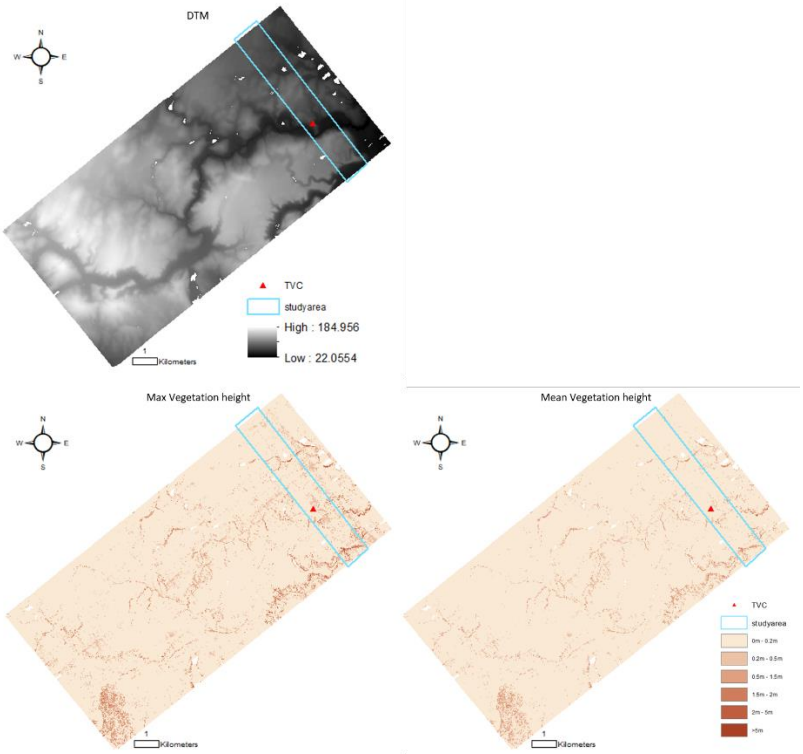

Figure 3. 2016 snow off DTM, max vegetation height and mean vegetation height maps.

Field Snow Depth Probe Measurements

Field snow depth data were collected during three field campaigns in the winter of 2018-2019. Snow measurements focused on the open tundra area around the TVC research station. Snow depth measurements were made using a GPS-enable depth measuring probe, called a Magnaprobe (SnowHydro, n.d.). Table 1 shows the dates of the three campaigns as well as the number of depths recorded. In this paper, only the data from the last campaign were used because this campaign was temporally closest to when the radar flights took place on April 10th. Due to the limited coverage of the 2019 ALS point clouds, only approximately half of the measurements from the third Campaign were used (4334 points), whose locations and snow depth distribution are shown in Figure 4 and Figure 5. Most points have a snow depth between 0.3 and 0.5 m. The mean snow depth for all the used measurements is 0.42 m with a standard deviation of 0.13 m. As the field measurements do not cover gulley where deep snow is most likely to accumulate, there are few points with snow depth higher than 0.9 m.

Table 1. Details of three field campaigns.

	Dates	Number of Depths Recorded
First Campaign	November 12th -18th 2018	6185
Second Campaign	January 11th-20th 2019	6740
Third Campaign	March 19th-27th 2019	8541

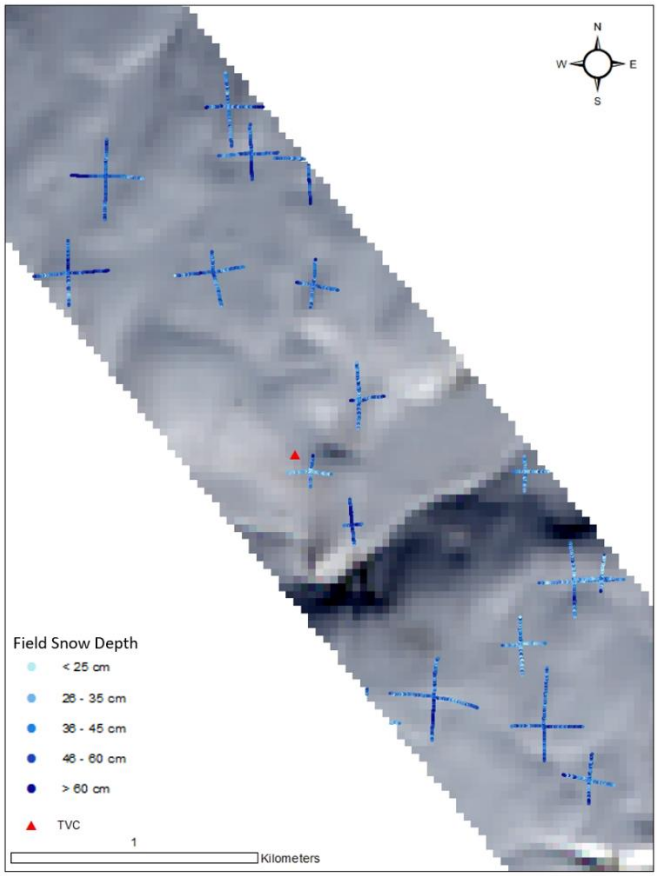

Figure 4. Locations of used field measurement (Landsat 8 from 2019.04.15 as background).

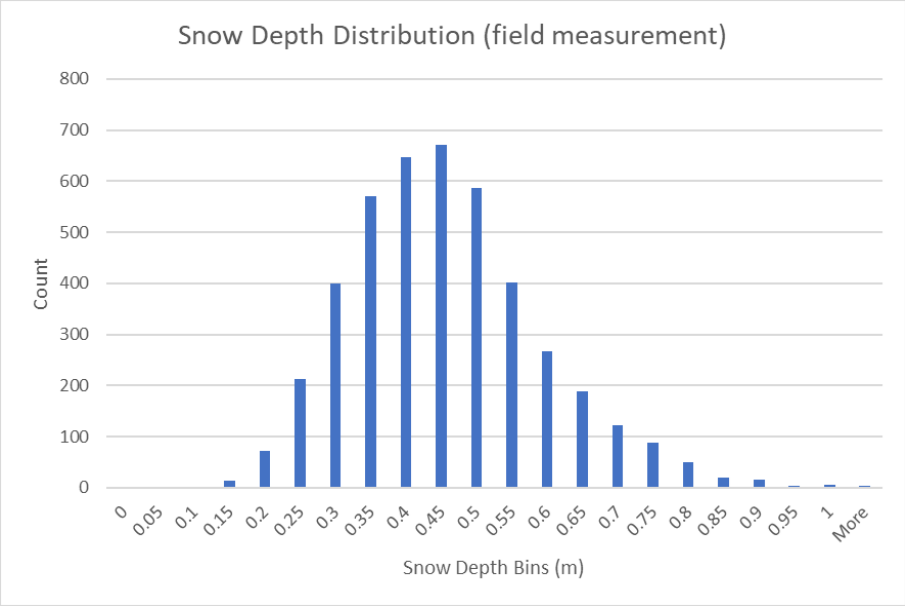

Figure 5. Histogram distribution of snow depth measurements (collected between 2019.03.21 and 2019.03.26).

In addition, Table 2 shows the mean and median snow depth and its standard deviation for each vegetation type from the field data. Among the six vegetation types, tree and tall shrub have the deeper snow but also a larger standard deviation while dwarf shrub, tussock and lichen are comparable in terms of mean, median snow depth and standard deviation.

Table 2. Mean and median snow depth and standard deviation for each vegetation type.

	Tree	Tall Shrub	Riparian Shrub	Dwarf Shrub	Tussock	Lichen
Mean (m)	0.56	0.63	0.49	0.42	0.42	0.43
Median (m)	0.56	0.65	0.47	0.41	0.41	0.43
Std (m)	0.21	0.16	0.17	0.13	0.13	0.12

TVC Vegetation Map

To account for the effect of different underlying vegetation on the radar waveform, a vegetation map was used to analyze spatial variability of vegetation type in the study region. A vegetation map of TVC was provided by Boike and Grünberg (2019), with a spatial resolution of 10 m. Vegetation is classified into 6 classes including tree, tall shrub, riparian shrub, dwarf shrub, tussock, as well as lichen as is shown in Figure 6. The map also contains a seventh class, water. The map covers the same area as the 2016 ALS point clouds data introduced above and is clipped to the study area.

According to Boike and Grünberg (2019), the vegetation map was generated through combining ALS point clouds and airborne orthophotos. Reference areas were chosen and vegetation types in these areas were classified manually, after which a set of parameters were calculated to characterize each vegetation type. Parameters include terrain attributes derived from ALS point clouds (e.g., slope, aspect, vegetation height, *etc*.) as well as spectral attributes from orthophotos (e.g., average hue, average saturation, *etc*.).

In their study, Boike and Grünberg (2019) reported a final map accuracy of 87.7% for the validation dataset and 89.5% for the calibration dataset. However, it was noted that with a 10 m spatial resolution grid, mixing of different vegetation types was likely in a single cell, specifically for tall shrub, where the actual tall shrubs are sparse and between the tall shrubs are normally low vegetation.

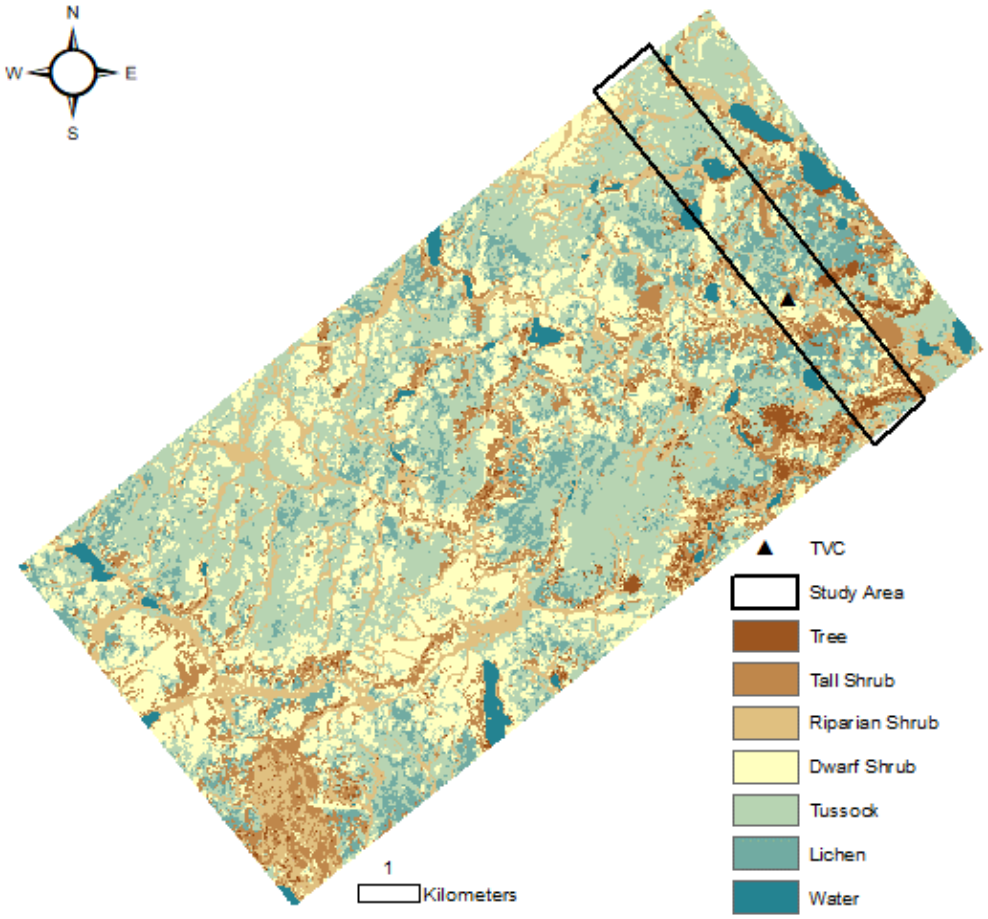

Figure 6. TVC vegetation map (Boike and Grünberg, 2019).

METHODS

Pre-processing

Data pre-processing was divided into 2 components: the generation of a snow depth product for April 2019 and radar waveform pre-processing. Because the sparse field snow depth measurements did not spatially match the flight path of the airborne wide band radar, to provide snow depth reference for each radar data location, a snow depth product was generated by differencing the snow-off DTM from September 13th, 2016 from the snow-on DTM from April 10th 2019. In addition, based on the snow depth distribution from the field data, any snow depth that is above 2.5 m or a negative value is treated as invalid (NaN).

Four pre-processing steps were applied to radar waveform to prepare them for analysis. First, for each radar waveform (14050 range bins), the first 500 and the last 500 bins were removed to eliminate residual noise from these range bins, as is shown in Figure 7. More specifically, the maximum return of each waveform is assumed to be the snow-ground interface because the difference of dielectric properties between snow and soil is theoretically the largest amongst all different interfaces. However, for some waveforms, the maximum return is located in the beginning range bins or in the final range bins and is an artefact of system processing rather than geophysical signal since both sets of range bins are located geometrically far from where the snow-ground interface is expected.

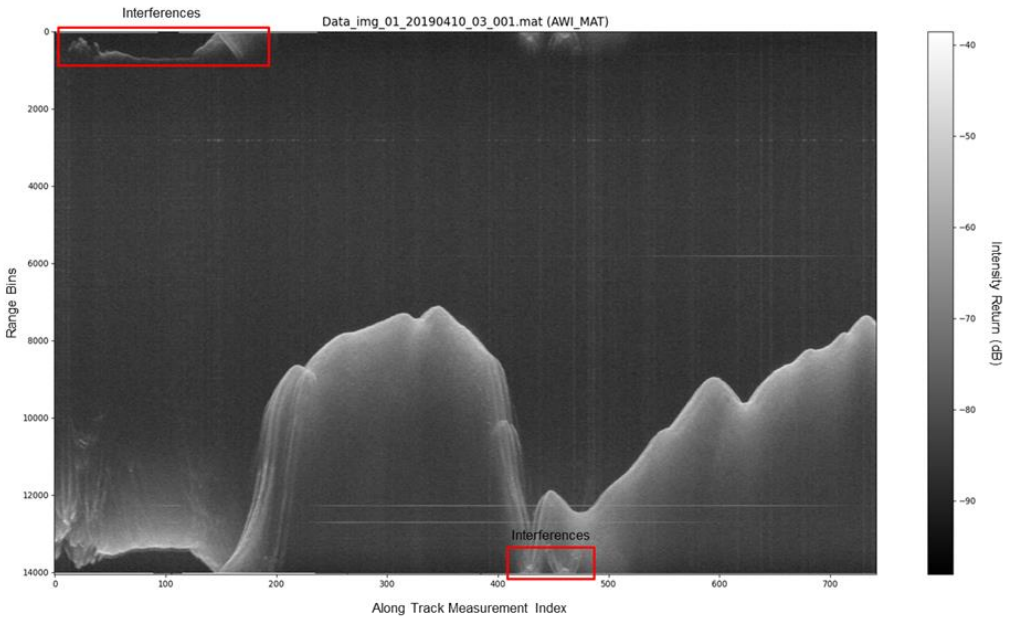

Figure 7. Interference in the beginning and final bins.

Second, a mean noise value was recorded for each waveform by averaging the first 100 returns after the first step. This was used as base reference to distinguish an interface peak in the waveform from noise. Third, the waveform was further reduced in length to 300 range bins, 150 bins before and 149 bins after the waveform maximum. The waveform maximum response is assumed to be the ground surface (snow-ground interface) which has the strongest power return. The reason for centering each waveform on its maximum was to provide a consistent reference when comparing different waveforms, regardless of ground elevation and flight elevation. However, the number of bins is arbitrary, and is largely influenced by the distribution of snow depth. Fourth, the waveforms were linearly normalized by dividing each return by its maximum response.

Evaluation of Lidar Snow Depth Product

Field snow depth data from King *et al*. (2020) collected between 2019.03.21 and 2019.03.26 were used as a ground reference to evaluate the lidar snow depth product obtained from differencing the 2016 and 2019 DTMs. The 1-m cells of the lidar product were converted into a point shapefile using the centroid of each cell. For each field snow depth measurement, its coincident lidar snow depth was obtained by selecting the closest converted lidar snow depth centroid within 1 m as is shown in Figure 8. Comparisons of field and lidar snow depths were conducted for each vegetation type. Distribution parameters for both data were calculated (mean, standard deviation, and median) in addition to obtaining mean absolute error (MAE).

Estimation of Snow Depth Using the Pulse Peakiness Approach

Julita *et al*. (2021) described the pulse peakiness approach as one of a range of peak picking methods based on the pulse peakiness of local peaks. Most of these peak picking methods have been applied for sea ice applications (Kwok *et al*., 2017; Newman *et al*., 2014; Kurtz *et al*., 2013). As a peak in the waveform can potentially indicate the boundary between two mediums with different dielectric properties, the approach was essentially searching for peaks that represent the air-snow and the snow-ice interfaces. Logarithmic scale of the impulse return was used for locating the air-snow interface and linear scale of the intensity returns was used for obtaining the snow-ice interface. The reason for using linear for the snow-ice interface is that the difference of dielectric properties between snow and sea ice is large enough that using a linear scale will decrease the number of peaks detected while preserving the interface peak.

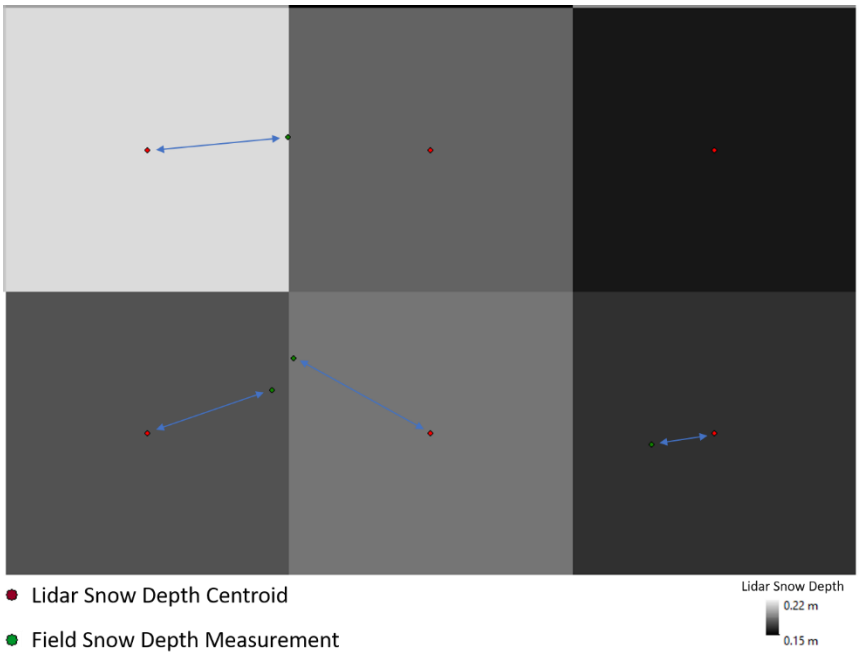

Figure 8. Matching field snow depth and lidar snow depth.

Eligible peaks were selected by applying a threshold over all detected peaks based on manually input parameters (log threshold & linear threshold) as well as the mean noise. A screening process was applied after identifying all the eligible peaks: if in linear scale, there were more than 5 eligible peaks, the waveform was considered too complicated to produce a robust result. For waveforms with fewer than 5 linear eligible peaks, two pulse peakiness values were calculated for each eligible peak. Unlike the pulse peakiness used in sea ice classification, Julita *et al*. (2021) defined these two pulse peakiness (left and right) as:

$$PP_l = \frac{S_{peak}}{\frac{1}{N}\sum_{i=1}^{N} S_{peak-i}} \times N$$

$$PP_r = \frac{S_{peak}}{\frac{1}{N}\sum_{i=1}^{N} S_{peak+i}} \times N$$

where S_{peak} is the linearly normalized intensity return, N is the bin range before and after the local peaks indicating how many bins are included into the calculation. It was recommended to use N = 10 because it was 2 times the range resolution of the radar (Julita *et al*., 2021). Finally, air-snow interface was defined to be the first eligible peak whose PP_l exceeded a set threshold while snow-ice interface was defined to be the last eligible peak whose PP_r exceeded another set threshold.

Two changes were made to the pulse peakiness method that was developed for snow on sea ice. First, the process of defining the snow-ground interface using pulse peakiness "right" was removed and instead, it was assumed that the strongest return would be the snow-ground interface. For sea ice applications, there may be circumstances where the strongest return was the air-snow interface due to high brine concentrations in the lower part of the snow. Geldsetzer *et al*. (2019) has shown that the C-band radar response from snow on first year sea ice is complicated by the brine layer. The effect of the brine layer on other frequencies as wide band radar covers from 2GHz to 18GHz is less well described and possible future work is needed. Pulse peakiness right is introduced to address the situations where air-snow interface is the strongest return. However, for terrestrial snow, this is irrelevant. Second, additional screening process were added to overcome the complexity of the

waveforms. Compared with snow on sea ice, it is expected that there are considerably more complex reflections in waveforms on land especially due to complicated vegetation interactions. Two kinds of complex interaction are expected: interactions between the wave, snow, and emergent vegetation such as trees and tall shrubs, and interactions between snow and buried vegetation such as dwarf shrubs. Therefore, in addition to the original screening where the waveform must have no more than 5 eligible linear peaks to have a valid output, 3 more restrictions were added. To begin with, the normalized linear return of the air-snow interface peak must be less than 0.6. Moreover, the air-snow interface must be ahead of the strongest return (assumed snow-ground interface) in range bin sequence. Finally, a target peak was introduced to further filter out the waveforms that were too sophisticated. A target peak was defined to be the closest eligible logarithmic peak to the reference air-snow interface, where the reference air-snow interface was obtained by converting the lidar snow depth into range bins using an assumed snow density (Fig. 9). For any waveform, if no target peaks were present within 20 range bins of the reference air-snow interface, then the waveform was considered to be ambiguous and was removed from the analysis.

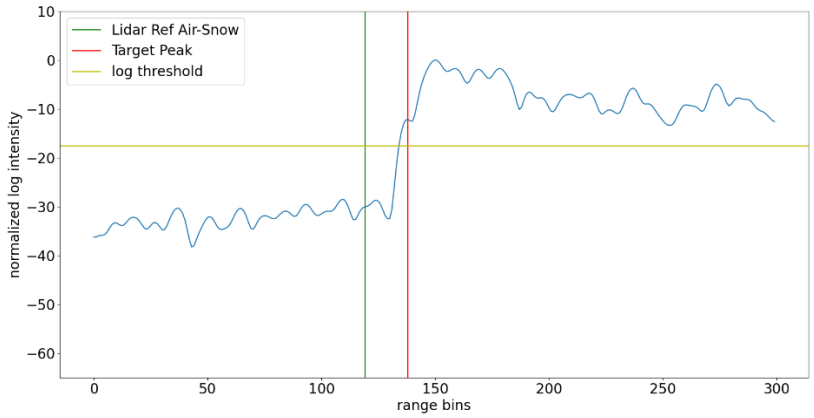

Figure 9. Example target peak defined by the closet eligible log peak to the lidar reference air-snow interface (waveform shown is from tussock).

In summary, the snow-ground interface was assumed to be the strongest return in the waveform. Therefore, the algorithm required 3 parameters instead of the 4 originally defined by Julita *et al.* (2021): 1) logarithmic threshold for determining eligible peaks in logarithmic scale; 2) linear threshold for determining eligible peaks in linear scale, and 3) the left pulse peakiness threshold for choosing the air-snow interface. All three parameters were optimized by using lidar snow depth data as reference. For both logarithmic and linear threshold, the range was set to be between 0.1 and 0.9; for the left pulse peakiness threshold, the range was from 10 to 35. Optimization scoring for each set of parameters is shown below:

$$score = 0.75 \times MAE - 0.25 \times number\ of\ valid\ outputs$$

Both MAE and the number of valid outputs were normalized to ensure they were similarly scaled (from 0 to 1). In addition, to ensure there would be sufficient points for evaluation, a requirement for the number of valid outputs under any set of parameters must be above 10% of the total points was added. Optimization was performed for each vegetation type separately. Finally, the best set of parameters were defined as the ones with the lowest score.

The optimization algorithm was run using 3 different divisions of the dataset. First, it was run on the entire dataset, and then the optimal parameters were applied to the entire dataset, which represented a best-case scenario. Second, it was run on an 80-20 split of the dataset; 80% of the data were used for optimization/training and 20% of the dataset for testing. Third, a 50-50 split of the dataset, 50% optimization/training and 50% testing.

RESULTS

Evaluation of Lidar Snow Depth Product

Figure 10 shows the 1-m spatial resolution lidar snow depth product from the 2016 snow-off DEM and 2019 snow-on DEM. The snow depths were around 30-40 cm in relatively flatter grounds mostly covered by tussocks, lichen, and dwarf shrubs. This is similar to the distribution of field snow depth measurements despite there being a two-week gap between when the field measurements and wideband radar measurements were made. In areas near streams or some valleys in the bottom part of the study area, the snow depth was greater than 60 cm, as indicated by the bright white in Figure 10. This is due to the collection of snow in these gullies and depressions.

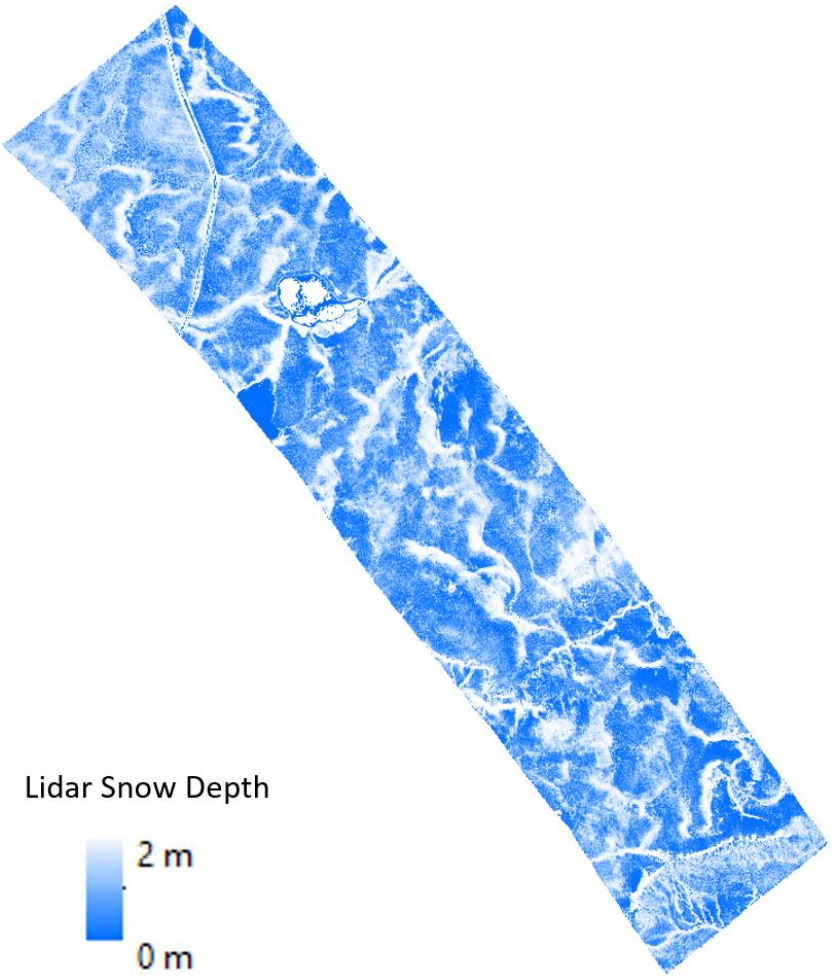

Figure 10. Lidar snow depth product for the study area.

There were 4334 field snow depth measurements used for the evaluation of lidar snow depth product. Table 3 shows the results of comparing the lidar snow depth with field measurement. Treating field snow depth as the reference, the lidar estimates underestimate snow depth for all six vegetation types, especially for tree, riparian shrub, and dwarf shrub classes as is shown in Figure 11. It is necessary to note that there was a 14-day difference between the time when these two data were collected. According to local weather station data, the temperature remained below zero with no significant precipitation event, however, some snowmelt might have been present in the field which could explain part of the reason for the underestimation. Further work to verify this possibility is ongoing. In addition, it is possible that the underestimation was the product of mismatch between snow-off and snow-on dates (the snow-off DEM used was from 2016); the underlying terrain or

vegetation may have changed, leading to an inaccurate lidar snow depth. Equally, there could be an over-probe of up to 7.6 cm for the magnaprobe, as is mentioned in King *et al*. (2020), which would contribute to field snow depth being relatively greater.

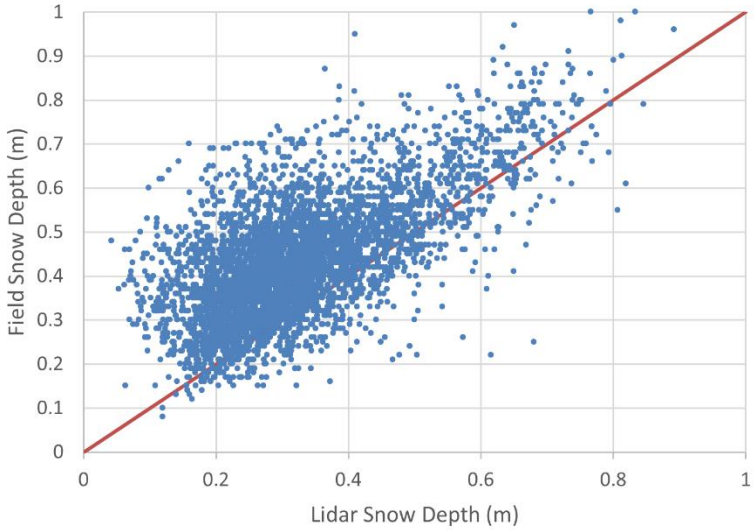

Figure 11. Scatterplot comparing lidar and field snow depth.

Table 3. Evaluation of lidar snow depth.

	Tree		Tall Shrub		Riparian Shrub		Dwarf Shrub		Tussock		Lichen	
	Field	Lidar	Field	Lidar	Field	Lidar	Field	Lidar	Field	Lidar	Field	Lidar
Mean (m)	0.56	0.45	0.63	0.53	0.49	0.38	0.42	0.28	0.42	0.33	0.43	0.35
Std (m)	0.21	0.20	0.16	0.17	0.17	0.16	0.13	0.11	0.13	0.11	0.12	0.12
Median (m)	0.56	0.39	0.65	0.57	0.47	0.34	0.41	0.26	0.41	0.31	0.43	0.33
RMSE (m)	0.19		0.15		0.17		0.19		0.13		0.13	
Bias (m)	-0.11		-0.10		-0.11		-0.15		-0.09		-0.08	
Std_bias (m)	0.16		0.11		0.13		0.12		0.09		0.10	
Count	13		151		231		1316		1456		1167	

Pulse Peakiness Approach Results

Table 4 shows the optimal pulse peakiness parameters for each vegetation type using a 100-0 split, and Figure 12 shows the retrieved snow depth variation for each vegetation classes. The pulse peakiness estimates used are from the optimal parameters obtained for a 100-0 split. The reason for using a 100-0 split is that the randomness in splitting the data has a significant effect on the optimization results, especially for tree, tall shrubs, and riparian shrubs. Moreover, a 50-50 split may result in insufficient amount of the waveforms for the algorithms, which will enhance the effect of the randomness in splitting. It is necessary to note that this is the first attempt of developing the method and a stratified random split could be tested in the future.

Table 4. Optimal pulse peakiness parameters for each vegetation type

	Log	Lin	PP_L	MAE (m)	Bias (m)	RMSE (m)	Pearson	Valid/Total Points
Tree	0.6	0.5	25	0.12	0.03	0.15	0.65	11/84
Tall Shrub	0.3	0.8	20	0.27	0.13	0.34	0.20*	226/441
Riparian Shrub	0.2	0.7	15	0.29	0.21	0.38	0.35*	209/321
Dwarf Shrub	0.4	0.5	25	0.13	0.04	0.19	0.29*	135/500
Tussock	0.5	0.5	30	0.09	0.04	0.12	0.55*	141/1224
Lichen	0.3	0.5	30	0.10	0.01	0.15	0.46*	74/558

*indicates significant P-value at $\alpha<0.01$

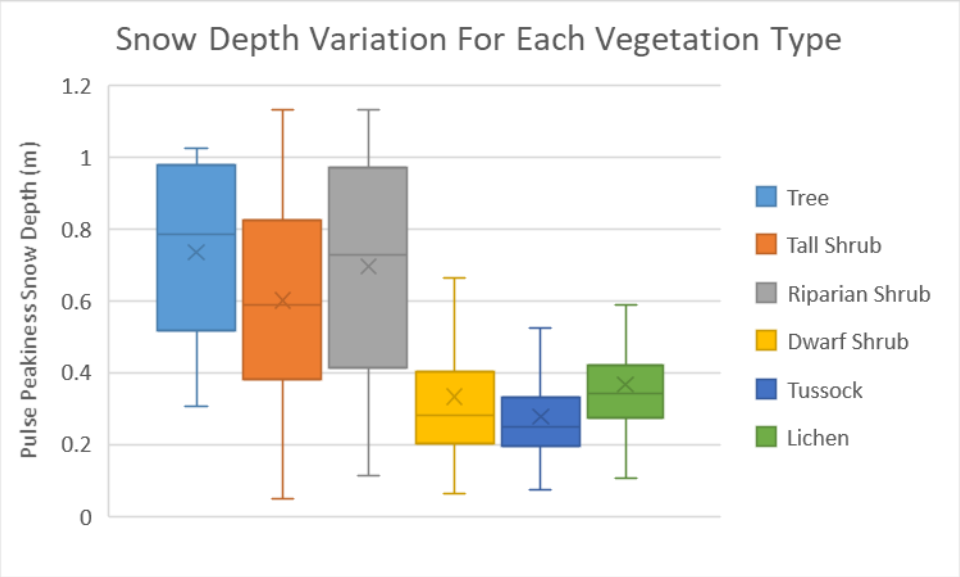

Figure 12: Pulse peakiness snow depth variation for each vegetation type

Overall, from Figure 12, snow depth variation is considerably higher for trees, tall shrubs and riparian shrubs in addition to them having generally thicker snow, possibly due to the higher vegetation.

For trees, despite a relatively high Pearson correlation coefficient and a low MAE, the *p*-value is not significant because there are few points. The scatterplot (Figure 13a) for trees further proves the high correlation and low MAE where most points fall closely around the 1:1 line. While this is promising, due to the sparse distribution of trees in the study area, the valid points shown here are likely to be returns from spaces between trees instead of actual tree returns. This is further reinforced by the similarity of parameters between tree and three other low height vegetation (dwarf shrub, tussock, & lichen). The low linear threshold at 0.5 filtered out all waveforms from actual tree returns since these waveforms are highly complex with a large number of peaks. Both tall shrub and riparian shrub have low Pearson correlation coefficients (and high MAE values) of 0.2 (0.27m) and 0.35 (0.29m), respectively. The scatterplots (Figure 13b and c) show the characteristics of relationships between radar and reference estimates with no discernible relationship between tall shrub and lidar and a general overestimation of the radar estimates for tall shrubs. An additional observation here is the high linear threshold of these two vegetation types compared with low height vegetation like tussock. As a high linear threshold leads to complex waveforms (ones often challenging for this retrieval approach) being more likely to be fed into the retrieval, the optimal linear threshold should be as low as possible provided that there will be sufficient valid output. For tall shrub and riparian shrub, a high optimal linear threshold indicates that fewer waveforms from these two vegetation types resembles the ideal two-peak shape, and the optimization was forced to include more complex waveforms to satisfy the criteria of having enough valid outputs.

For dwarf shrub, the MAE is 0.13 m with a Pearson correlation coefficient of 0.29. Due to the lower linear threshold, the increase in total points does not lead to an increase in valid points. From Figure 13d, the reason for a low Pearson correlation coefficient is the overestimation of snow depth by the pulse peakiness algorithm. More specifically, there are a series of points with increasing pulse peakiness snow depth, whose Lidar snow depths remain constant at around 0.2 and 0.3 m. This is most likely to be caused by the waveform having multiple peaks with close to maximum returns and will be further discussed later. In comparison, the underestimation of the retrieval retained the observable pattern from riparian shrub as most of the points above the 1:1 line still falls into a linear pattern.

According to Boike and Grünberg (2019) tussock and lichen classes are characteristically similar in classification terms. However, they are functionally different biophysically. They behave similarly in terms of their radar waveform. Compared with other vegetation types, they have the best retrieval results with MAE of 0.09 m for tussock and 0.1 m for lichen and Pearson correlation coefficients of 0.55 and 0.46, respectively. This result is to some degree achieved by severely constraining the waveform selection for use with the algorithm, as indicated by the low valid / total points ratio. In other words, for tussock and lichen, the two thresholds are relatively more effective at filtering out waveforms that are difficult for the algorithm to accurately obtain interfaces. Both the decrease in MAE and the increase in Pearson correlation coefficient are illustrated in the scatterplot of Figure 13e and f, where the overestimation that results in the low correlation for dwarf shrubs has mostly disappeared. It is noted that the pattern for underestimation of the method persists in both of these vegetation type, indicating that this could possibly be addressed with empirical bias correction.

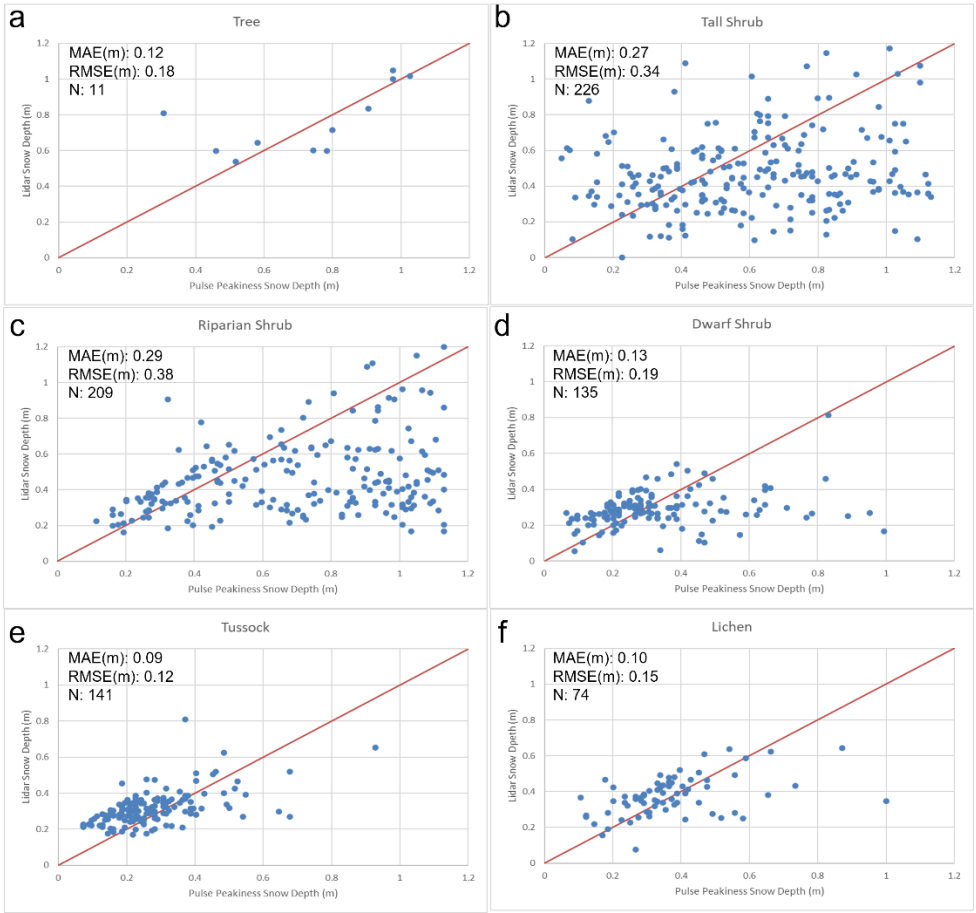

Figure 13. Scatterplots between pulse peakiness and lidar snow depth estimates for the following classes a) tree, b) tall shrub, c) riparian shrub, d) dwarf shrub, e) tussock and f) lichen.

DISCUSSION AND CONCLUSION

Underestimation and Overestimation

Both underestimation and overestimation exist but they result from different sources. With respect to underestimation, it is likely to not be from waveform / radar measurement and it follows a clear pattern, meaning that it can be mitigated to a degree with a bias correction. On the contrary, the overestimation is mainly resulted from the waveforms and needs other more delicate solutions.

The tussock land cover class is used to illustrate the underestimation problem as this landcover class has the most radar points. When running the pulse peakiness approach on the 100-0 split scenario, out of 141 valid outputs, 57 had an absolute error of 10 cm or above. And out of these 57, 39 underestimated snow depth in a similar way as is shown in Figure 14. The Lidar reference air-snow interface falls into areas which could be considered noise in the logarithmic scale intensity return while the algorithm was successful in identifying the first peak that satisfied the logarithmic threshold.

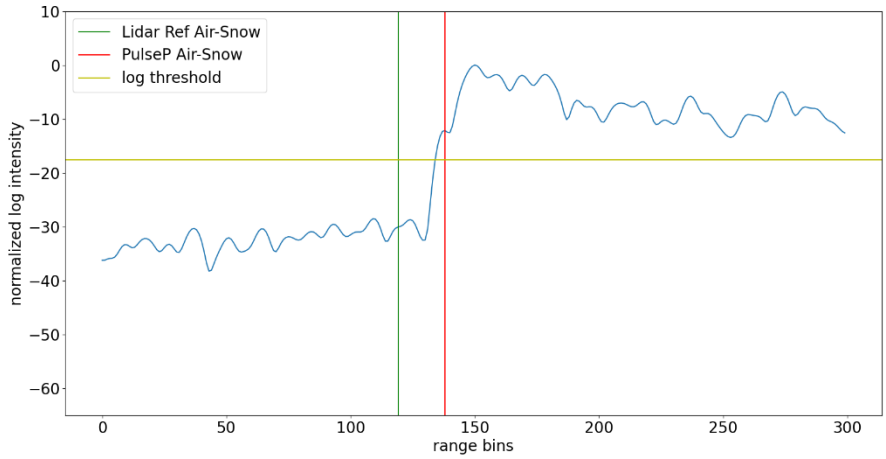

Figure 14. Sample underestimating waveforms for pulse peakiness where the Lidar reference air-snow interface falls below the logarithmic threshold.

As pulse peakiness left value alone normally cannot uniquely identify a peak as the air-snow interface, it is likely that before the interface peak, there would be numerous peaks with similar pulse peakiness left values. Consequently, it is extremely important to exclude as many non-interface peaks before the interface peak as possible to increase the chance of it being the first that exceeds the pulse peakiness left value threshold. This is achieved through the application of a logarithmic threshold. By lowering the threshold to accommodate the type of underestimation shown in Figure 14 the result would most likely lead to an overestimation of snow depth. An alternative is to apply a post algorithm bias correction based on this dataset.

To achieve this, the data were split in half into training and testing subsets, for which an average of underestimation was calculated for all the waveforms in the training subset. This average was used as the bias correction for all the waveforms in the testing set. Figure 15 shows the comparison between the un-corrected scatterplot and the corrected scatterplot. The average bias for the training set in this case is approximately -4 cm and the MAE after the correction decreased approximately 1 cm. The reason for such a small improvement is that the bias correction is applied to all the waveforms in the testing sets including those that are grossly overestimating. More intricate solutions to this bias correction, such as the identification and character generalization of underestimating waveforms should further improve the results.

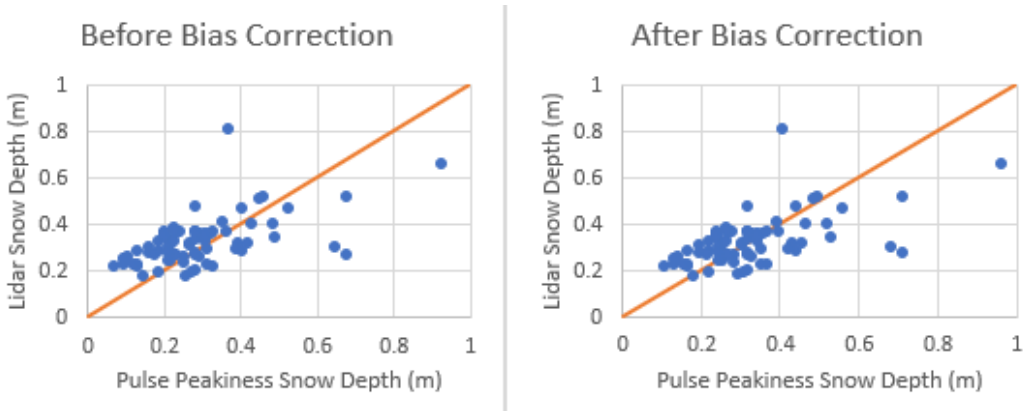

Figure 15. Before and after bias correction.

With respect to the overestimation problem, the main reason is the complexity of the waveform. A complex waveform here is defined as a waveform that has more than one peak that is close to the maximum return, in other words, a waveform that has no clear single maximum return. There can be a number of reasons why a waveform may have multiple "maximum" returns, but one plausible possibility is that a feature in the snowpack has a similar dielectric response to the ground surface.

This kind of waveform is inconsistent with the baseline assumption of a clear single maximum return peak being the snow-ground interface, and therefore disrupts the snow depth calculation which is the distance between air-snow peak and the maximum return. Figures 16 and 17 show examples of overestimation in dwarf shrubs and lichen.

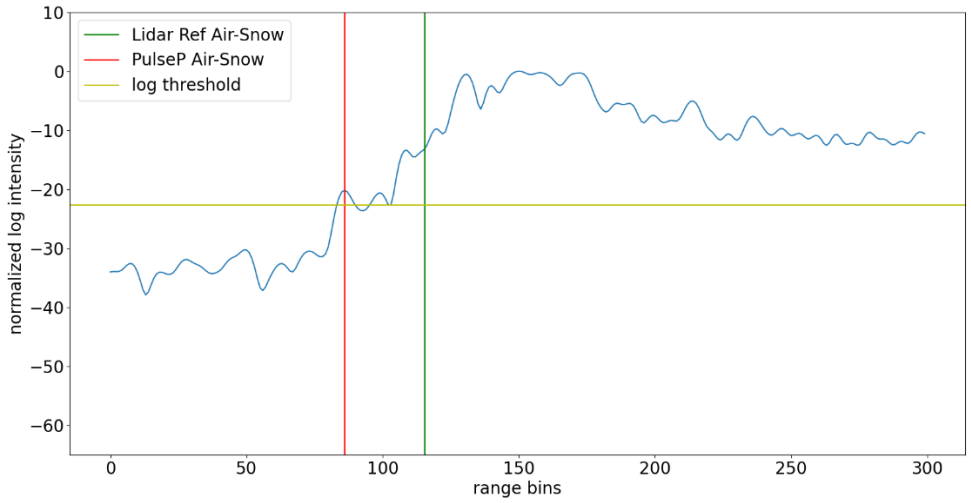

Figure 16. Sample overestimation for dwarf shrubs where there are multiple peaks close to the maximum peak.

To counteract this problem, one possible way is to apply a signal-to-noise ratio (SNR) threshold that excludes waveforms like the one shown in Figure 16. For such waveforms, adding another screening criterion that requires the waveform to have only 1 peak whose return is above a certain value should help filter them out. The challenge would be to set the correct threshold so that it does not filter out too many of the waveforms leaving the algorithm with insufficient valid outputs.

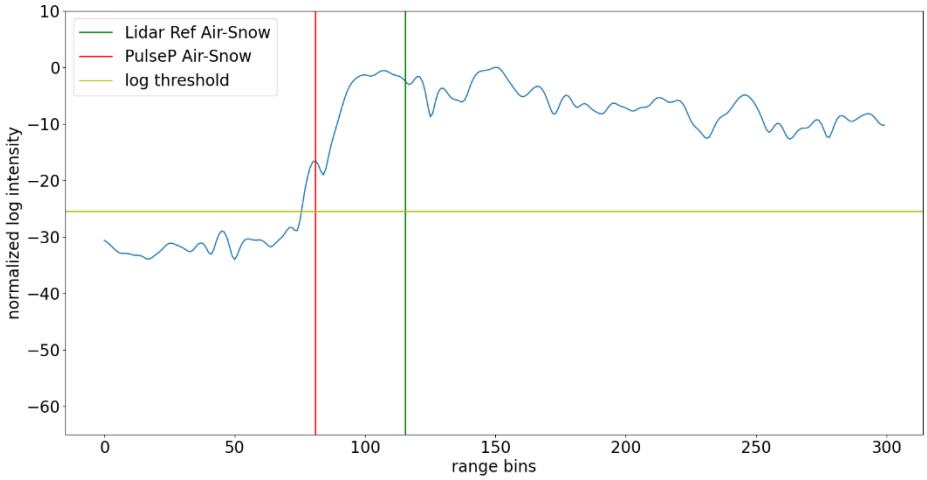

Figure 17. Sample overestimation for lichen where there are multiple peaks close to the maximum peak.

Error of Omission for the Pulse Peakiness Approach

Due to the limited number of valid outputs of the pulse peakiness approach, an analysis was done on the error of omission for dwarf shrub, tussock, and lichen. The reason for excluding trees, tall shrub and riparian shrub is that these landcover types have more complex waveform responses and therefore are omitted from the algorithm.

Table 5 shows the distribution statistics for Lidar points belonging to dwarf shrub (500), tussock (1224), and lichen (559) as well as those with a valid pulse peakiness output (135, 141, & 74). It demonstrates that for dwarf shrub and lichen, even though pulse peakiness filters excluded parts of the data, the distributions are still similar. However, for tussock, the pulse peakinesss distribution has shifted slightly to lower snow depth, which is further reinforced by the histogram shown in Figure 18. The pulse peakiness output is lacking representation of snow depth that is above 0.4m.

Table 5. Distribution statistics for Lidar snow depth and pulse peakiness valid output.

	Dwarf Shrub		Tussock		Lichen	
	Lidar	Pulse Peakiness	Lidar	Pulse Peakiness	Lidar	Pulse Peakiness
Mean (m)	0.32	0.33	0.38	0.28	0.44	0.37
Median (m)	0.30	0.28	0.36	0.25	0.42	0.34
Std (m)	0.14	0.19	0.14	0.13	0.16	0.16

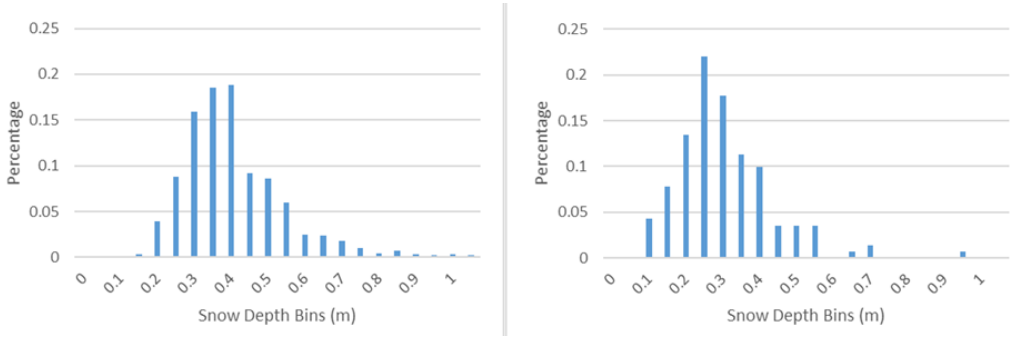

Figure 13. Distribution statistics for Lidar snow depth (left) and pulse peakiness (right) valid output.

CONCLUSION

The overall objective of this thesis is to test the applicability of a standard interface-based snow depth retrieval algorithms for snow on sea ice to snow on land. In principle, the pulse peakiness approach from Julita *et al*. (2021) should be valid for applicability to snow on land because the snow-ground interface provides a strong dielectric contrast along with the snow-air interface; these contrasts should be somewhat similar to snow on sea ice. However, the complex ground characteristics narrowed the locations where the pulse peakiness method could yield accurate results. More specifically, the presence of tall surface vegetation combined with the hummocky terrain meant that a considerably large percentage of the waveforms were discarded, and to some degree refocused the effort into how to filter out complex waveforms that are too ambiguous for the method. In conclusion, for snow on land, if the ground is generally a flat surface with minimal vegetation, it is expected that the interface-based pulse peakiness algorithm will have a good performance.

REFERENCES

Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research. 2019. *Awi als toolbox.* https://github.com/awi-als-toolbox/awi-als-toolbox. GitHub.

Anders K, Antonova S, Wegener A, Höfle B, Marsh P. 2018. Airborne laser scanning (als) point clouds of Trail Valley Creek, NWT, Canada (2016) vulnerability of tundra lakes to climate change view project 4d near real-time environmental monitoring view project. Retrieved from *https://www.uni-heidelberg.de/permasar*. *doi: 10.1594/PANGAEA.894884*.

Boike J, Grünberg I. 2019. Vegetation map of Trail Valley Creek, Northwest Territories, Canada. *doi: 10.1594/PANGAEA.904270*.

Croce P, Formichi P, Landi F, Mercogliano P, Bucchignani E, Dosio A, Dimova S. 2018. The snow load in Europe and the climate change. *Climate Risk Management*, **20**: 138-154. *doi: 10.1016/j.crm.2018.03.001*.

Cui Y, Xiong C, Shi J, Jiang L, Peng B, Ji D, Zhao T. 2016. Estimating snow water equivalent with backscattering at X and Ku bands. In: *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS):* 3454-3457. *doi: 10.1109/IGARSS.2016.7729893*.

Geldsetzer T, Langlois A, Yackel J. 2009. Dielectric properties of brine-wetted snow on first-year sea ice. *Cold Regions Science and Technology*, **58**: 47-56. *doi: 10.1016/j.coldregions. 2009.03.009*.

Hendricks S, Ricker R, Jutila A. 2019. Icebird 2019 winter icesat-2 validation data acquisition report.

Jutila A, King J, Paden J, Ricker R, Hendricks S, Polashenski C, Helm V, Binder T, Haas C. 2021. High-resolution snow depth on arctic sea ice from low-altitude airborne microwave radar data. *IEEE Transactions on Geoscience and Remote Sensing*, **60**. *doi: 10.1109/TGRS.2021.3063756*.

King J, Toose P, Silis A, Derksen C. 2020. TVCSnow 2018-2019 tundra snow depth probe measurements. Retrieved from *https://doi.org/10.5281/zenodo.4021401*.

King J, Brady M, Newman T. 2020. kingjml/pysnowradar: Updated IEEE TGRS submission. Retrieved from *https://doi.org/10.5281/zenodo.4071947 doi: 10.5281/zenodo.4071947.*

Koenig LS, Ivanoff A, Alexander PM, MacGregor JA, Fettweis X, Panzer B, Paden JD, Forster RR, Das I, McConnell JR, Tedesco M, Leuschen C, Gogineni P. 2016. Annual Greenland accumulation rates (2009-2012) from airborne snow radar. *Cryosphere*, **10**: 1739-1752. *doi: 10.5194/tc-10-1739-2016*.

Kurtz NT, Farrell SL. 2011. Large-scale surveys of snow depth on arctic sea ice from operation icebridge. *Geophysical Research Letters*, **38**(20). *doi: 10.1029/2011GL049216*.

Kurtz NT, Farrell SL, Studinger M, Galin N, Harbeck JP, Lindsay R, Onana VD, Panzer B, Sonntag JG. 2013. Sea ice thickness, freeboard, and snow depth products from operation icebridge airborne data. *Cryosphere*, **7**: 1035-1056. *doi: 10.5194/tc-7-1035-2013*.

Kurtz NT, Galin N, Studinger M. 2014. An improved Cryosat-2 sea ice freeboard retrieval algorithm through the use of waveform fitting. *Cryosphere*, **8**: 1217-1237. *doi: 10.5194/tc-8-1217-2014*.

Kwok R, Maksym T. 2014. Snow depth of the Weddell and Bellingshausen sea ice covers from ice bridge surveys in 2010 and 2011: An examination. *Journal of Geophysical Research: Oceans,* **119**(7): 4141-4167. *doi: 10.1002/2014JC009943*.

Kwok R, Kurtz NT, Brucker L, Ivanoff A, Newman T, Farrell SL, King J, Howell S, Webster MA, Paden J, Leuschen C, MacGregor JA, Richter-Menge J, Harbeck J, Tschudi M. 2017. Intercomparison of snow depth retrievals over Arctic sea ice from radar data acquired by Operation IceBridge. *Cryosphere,* **11**: 2571-2593. *doi: 10.5194/tc-11-2571-2017.*

Leinss S, Wiesmann A, Lemmetyinen J, Hajnsek I. 2015. Snow water equivalent of dry snow measured by differential interferometry. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* **8**: 3773-3790. *doi: 10.1109/JSTARS.2015.2432031*.

Marshall HP, Koh G. 2008. FMCW radars for snow research. *Cold Regions Science and Technology,* **52**: 118-131. *doi: 10.1016/j.coldregions.2007.04.008*.

Newman T, Farrell SL, Richter-Menge J, Connor LN, Kurtz NT, Elder BC, McAdoo D. 2014. Assessment of radar derived snow depth over Arctic sea ice. *Journal of Geophysical Research: Oceans,* **119**(12): 8578-8602. *doi: 10.1002/2014JC010284*.

Oveisgharan S, Esteban-Fernandez D, Waliser D, Friedl R, Nghiem S, Zeng X. 2020. Evaluating the preconditions of two remote sensing SWE retrieval algorithms over the US. *Remote Sensing,* **12**(12): 2021. *doi: 10.3390/rs12122021*.

Palmer MJ, Burn CR, Kokelj SV. 2012. Factors influencing permafrost temperatures across tree line in the uplands east of the Mackenzie delta, 2004-2010. *Canadian Journal of Earth Sciences,* **49**: 877-894. *doi: 10.1139/E2012-002*.

Pohl S, Davison B, Marsh P, Pietroniro A. 2005. Modelling spatially distributed snowmelt and meltwater runoff in a small Arctic catchment with a hydrology land-surface scheme (WATCLASS). *Atmosphere – Ocean,* **43**: 193-211. *doi: 10.3137/ao.430301*.

Pomeroy JW, Gray DM, Shook KR, Toth B, Essery RLH, Pietroniro A, Hedstrom N. 1998. An evaluation of snow accumulation and ablation processes for land surface modelling. *Hydrological Processes,* **12**(15): 2339-2367. *doi:10.1002/(SICI)1099-1085(199812)12:15<2339::AID-HYP800>3.0.CO;2-L*.

SnowHydro. n.d. GPS snow depth probe. Retrieved from *http://www.snowhydro.com/products/column2.html*

Verret M, Wang Y, Bjornson J, Lacelle D. 2019. Hummocks in alpine tundra, northern British Columbia, Canada: Distribution, morphology and organic carbon composition. *Arctic Science,* **5**: 127-147. *doi: 10.1139/as-2018-0021*.

Yueh SH, Dinardo SJ, Akgiray A, West R, Cline DW, Elder K. 2009. Airborne Ku band polarimetric radar remote sensing of terrestrial snow cover. *IEEE Transactions on Geoscience and Remote Sensing,* **47**(10): 3347-3364. *doi: 10.1109/TGRS.2009.2022945*.