Cost-Effective Estimation of Vegetation Regrowth on Anthropogenic Features in NE BC (REMB-2017-02)

FINAL PROJECT REPORT

October 31, 2018

Project [1475-1]

Prepared for:

Jacqueline Howard
Resource Development Analyst
BC Oil & Gas Commission
Victoria, BC
Jacqueline.Howard@BCOGC.ca
(250) 419-4496

BC OGRIS
BC Oil and Gas
Research and Innovation Society

Prepared by:

Forsite Consultants Ltd.
Salmon Arm, BC V1E 4R1
Canada
250-832-3366

&

Rezatec Ltd.
Oxfordshire OX11 0RA
United Kingdom
+44 (0)1865 817545

FORSITE
Forest Management Specialists

Rezatec

Acknowledgements

This project was recommended by the Research and Effectiveness Monitoring Board (REMB) of the British Columbia Governments Boreal Caribou Implementation Plan (BCIP) initiative, and funded by BC Oil and Gas Research and Innovation Society (BCOGRIS). Special thanks to Brian Thomson and Steve Wilson with REMB and BC OGRIS for project funding and coordination.

Technical support and advice was provided by Steve Wilson (BCOGRIS), Jacqueline Howard, Janice Traynor, Aron Bird (BCOGC) and Alexandre Bevington (FLNRD). Data for this project was made available by multiple vendors and agencies. Special thanks to Andre Boyer and Laurent Giugni at Radarsat International for their project support and data sponsorship and to Sam Leif at Planet Labs for access to SPOT image data for northeast British Columbia. We gratefully acknowledge MacDonald Dettwiler and Associates Ltd and its corporate affiliates including MDA GIS for provision of RADARSAT-2 data under the RADARSAT-2 SOAR-CPT LOAN AGREEMENT.

Overview

This project was undertaken jointly by Forsite Ltd (BC, Canada) and Rezatec Ltd (United Kingdom), with Rezatec providing model development and primary data analysis, and Forsite providing GIS support and assessment of the modelled results.

Primarily, this project aimed to map the current state of vegetation regrowth along anthropogenic features within Northeast British Columbia using remotely sensed data from satellite sensors. The purpose is to provide a cost-effective solution to identifying treatment priority areas to ensure sufficient recovery and reduce the impact of linear features on boreal caribou migration. This project specifically aims to assess the suitability of Earth Observation (EO) sensors for providing an up-to-date estimation of vegetation height across the full area of interest.

To achieve an estimation of vegetation height, Rezatec applied a machine-learning algorithm to Earth Observation imagery, using LiDAR data as training, to model vegetation height across NE BC.

The modelled vegetation height results were assessed both quantitatively, through direct comparison of a sample of modelled heights versus LiDAR CHM heights, and qualitatively, through visual comparisons between the modelled height layer and the LiDAR CHM in terms of general trends. In general, the lower modelled heights were overestimated while the higher modelled heights were underestimated (compared to the CHM). An adjustment equation based on least squares regression was constructed and applied to the original model heights. The adjusted modelled height output more closely matched the LiDAR CHM data, but tended to generalize heights in some areas compared to the 5m CHM.

The key primary conclusion from this project is that low-cost sensors have demonstrated the ability to model vegetation heights over large areas (millions of hectares) in a robust and repeatable approach. At a landscape level, modelled heights are useful for showing canopy information, major disturbances and vegetation patterns. However many of the seismic features are below the spatial resolution of the height model (5m) and therefore the modelled height outputs do not have sufficient detail to estimate regrowth on all the seismic lines in Northeast BC. As such, the output is primarily of benefit for larger linear disturbances and clearings.

Recommendations to further improve the output modelled vegetation height include improved training data variables (spatially and / or spectral) and alternative modelling approaches. In particular, we recommend incorporating higher spatial resolution inputs to the modelling process which would enable higher spatial resolution outputs that would better characterize vegetation regrowth on the more narrow seismic line features (<5m).

Ultimately, given the resolutions of the sensors incorporated in the model, the result of this project is a very good estimation of vegetation heights over a vast area, giving valuable information for rapid assessment of the current state of vegetation regrowth along the entire anthropogenic framework, and subsequent prioritisation of further ground-based measures in identified ‘high-concern’ areas.

Contents

List of Figures

List of Tables

1 Introduction

Primarily, this project aimed to map the current state of vegetation regrowth along anthropogenic features (access roads, pipelines, seismic lines, and well sites) within Northeast British Columbia using remotely sensed data from satellite sensors. The purpose is to provide a cost-effective solution to identifying treatment priority areas to ensure sufficient recovery and reduce the impact of linear features on boreal caribou migration. Current methods of vegetation regrowth estimation and a current inventory are prohibited due to the size of the area that needs to be assessed (>8 million ha), and the expense of LiDAR acquiring LiDAR across this region. This project specifically aims to assess the suitability of Earth Observation (EO) sensors for providing an up-to-date estimation of vegetation height across the full area of interest (AOI – see Figure 1 for extent).

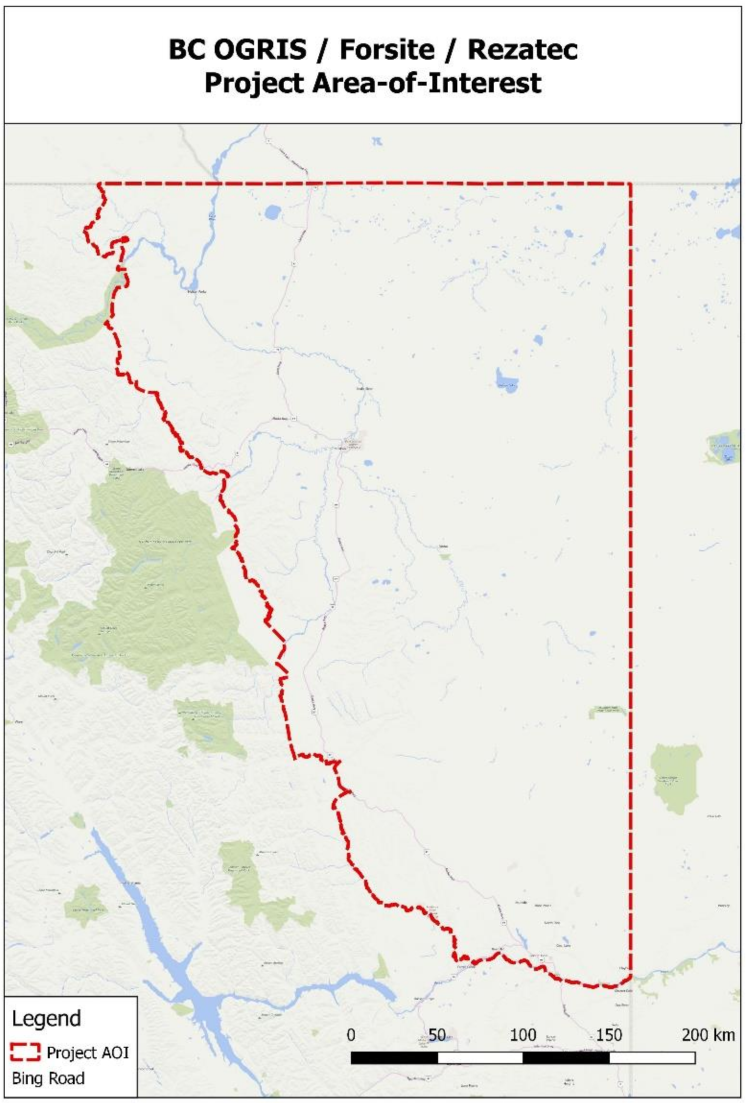

Figure 1. Project AOI

2 Methods Description

To achieve an estimation of vegetation height, Rezatec applied a machine-learning algorithm to Earth Observation imagery, using LiDAR data as training, to model vegetation height across NE BC. The process can be divided into two parts as described below; training data preparation, and model parameterisation.

2.1 DATA SELECTION AND PREPARATION

Approximately 74,000 ha of archived LiDAR data, distributed across three regions of the AOI, was obtained for training and validation (see Figure 2). Acquisition dates ranged between 2013 and 2015, and point densities between 1.3 and 4.7 pules / m2.

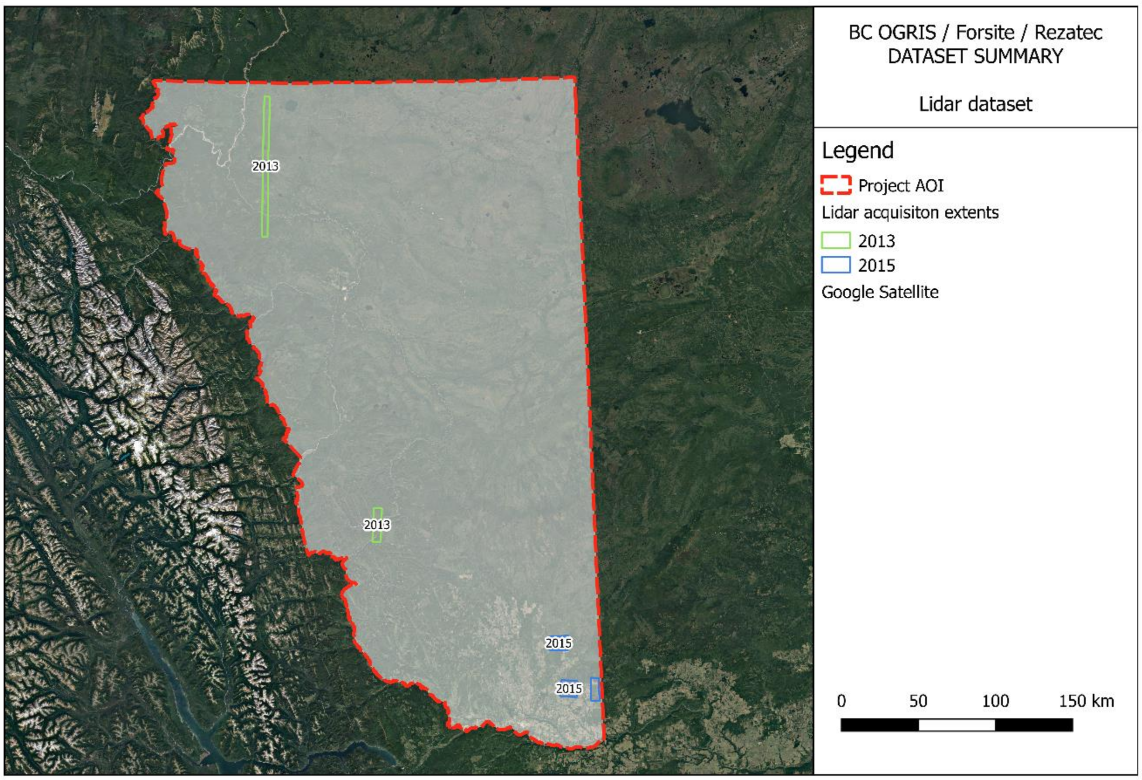

Figure 2. LiDAR data acquisition pattern

A 5m spatial resolution canopy height model (CHM) was constructed from the original LiDAR point data (maximum height within each pixel extent). The 5m spatial resolution was chosen based on the relatively coarse LiDAR point density, and also because it matched the agreed upon spatial resolution of the modelled outputs for this project.

Rezatec utilised several EO input datasets to parameterise the vegetation height model. The inputs were selected for the model implementation based on their unique sensing capabilities (see Table 1 for a summary of EO inputs). The selection criteria were spectral wavelength (optical versus radar), spatial resolution, and affordability.

Spectral resolution

The key differentiator and selection criteria for earth observation data (for this project) was the spectral capability of the sensor, specifically the selection of both multi-spectral optical and synthetic aperture radar (SAR) data. Optical data is sensitive to surface feature reflectance patterns.

Spatial resolution and affordability

The features of interest in this project (anthropogenic features) vary in scale, from narrow seismic lines 3 metres in width, to large well sites in excess of 10-20 metres across. The aim of the project was to utilise EO sensors to map vegetation heights across all features if possible, and given the spatial density of linear features (e.g. seismic lines and pipelines), the focus was primarily on a high resolution capability (minimum 3 metre). However, the design of EO sensors to achieve such spatial resolutions often impacts the cost of data acquisition, and high spatial resolution sensors (sub 5 metres) are often commercial. Utilisation of such datasets would have been counter-intuitive to the low-cost requirement of the project, so the sensors used in the project were those that were either freely available (e.g. Sentinel's 1 and 2, ALOS-PALSAR) or available as a comparatively low cost to the end-user (in this instance, SPOT 5,6, and 7, and RADARSAT-2 data). This ultimately means the majority of sensors were in excess of 3 metres pixel resolution, ranging from 2.5 to 20 metres, and that the final output was modelled at a 5 metre pixel resolution to provide a relevant and functional product without exceeding the abilities of the input data.

Table 1. Earth Observation Datasets

Dataset	Sensor	Data type	Data product	Source	Spectral resolution	Spatial resolution	Temporal range	Processing level	Data description	Spatial Reference System
Sentinel-1	C-SAR	Radar	IW GRD	ESA	5.405 GHz	10m	2017	Radiometrically terrain-corrected	C-Band Synthetic Aperture Radar (SAR) data acquiring both intensity backscatter (GRD) and phase information (SLC), sensitive to vegetation canopy structure and dielectric constant.	EPSG: 32610 WGS 84 / UTM zone 10N
			IW SLC					Coherence (unitless)		
Sentinel-2	MSI	Multi-spectral	L2A	ESA	B2 (490 nm), B3 (560 nm), B4 (665 nm), B8 (842 nm)	10m	2017	L2A Surface Reflectance	Sentinel-2 is a medium-to-high resolution, multi-spectral imaging mission, supporting Copernicus Land Monitoring studies, including the monitoring of vegetation, soil and water cover.	EPSG: 32610 WGS 84 / UTM zone 10N
					B5 (705 nm), B6 (740 nm), B7 (783 nm), B8a (865 nm), B11 (1610 nm), B12 (2190 nm)	20m				
					B1 (443 nm), B9 (940 nm), B10 (1375 nm)	60m				
RADARSAT-2	C-SAR	Radar	Wide Fine	SOAR-CPT	5.405 GHz	8m	2018	Map Image (SSG)	C-Band Synthetic Aperture Radar (SAR) intensity backscatter data, sensitive to vegetation canopy structure and dielectric constant. Acquired at higher spatial granularity than Sentinel-2	EPSG: 32610 WGS 84 / UTM zone 10N
SPOT	SPOT-5	Multi-spectral	L1	Planet	Panchromatic (0.48 - 0.71 µm)	2.5/5m	2011 & 2014	Orthorectified	The SPOT constellation provides a high resolution, multi-spectral imaging mission, supporting land cover monitoring activities, including the monitoring of vegetation, soil and water cover.	EPSG: 32610 WGS 84 / UTM zone 10N
					B1 : Green (0.50 - 0.59 µm) B2 : Red (0.61 - 0.68 µm) B3 : NIR (0.78 - 0.89 µm)	10m				
					B4 : SWIR (1.58 - 1.75 µm)	20m				
	SPOT 6/7	Multi-spectral	L1		B1: Blue: 0.450-0.520 µm B2: Green: 0.530-0.590 µm B3: Red: 0.625-0.695 µm B4: NIR: 0.760-0.890 µm	6m	2017	Orthorectified		
ALOS	PALSAR	Radar	Fine Beam Dual	Alaska Satellite Facility	1270 MHz	12.5m	2010-2011	Radiometrically terrain-corrected	L-Band Synthetic Aperture Radar (SAR) intensity backscatter data, sensitive to vegetation canopy and branch structure and dielectric constant.	EPSG: 32610 WGS 84 / UTM zone 10N

All datasets were resampled to the same pixel resolution (5 metres) and grid extent, (a requirement of the random forest algorithm), thereby downscaling some inputs (SPOT 6) and upscaling others (e.g. Sentinel-1 and Sentinel-2).

2.2 MODEL PARAMETERISATION

5000 randomly selected X,Y co-ordinates were generated across the spatial extent of the training data (LiDAR CHM), capturing a range of vegetation heights, and topographical variations (elevations, aspects, slopes etc.) and intersected with the LiDAR CHM and pre-processed EO data layers. This intersected data formed the training data for the machine-learning (random forest) algorithm. 70% of the training data points were input into the model, which ascertains the relative influence of each input to predicting the model variable; vegetation height, as intersected in the LiDAR data. Having created a model, the remainder of the training data (in this instance, 30%) is evaluated against the model prediction to assess the model fit (based on correlation between actual and modelled values).

Once a good model fit is achieved, (based on user-knowledge of training data and interpretation of model outputs), the model is then applied to the EO raster datasets across the full AOI. The key elements of assessing the model fit (the ability for the model to ascertain a good correlation against the input data and the modelled variable sample) are the relative influence of the input variables (the EO data in this instance, see Figure 3), and the correlation between the predicted heights versus the withheld (independent) sample test vegetation height data (see Figure 4).

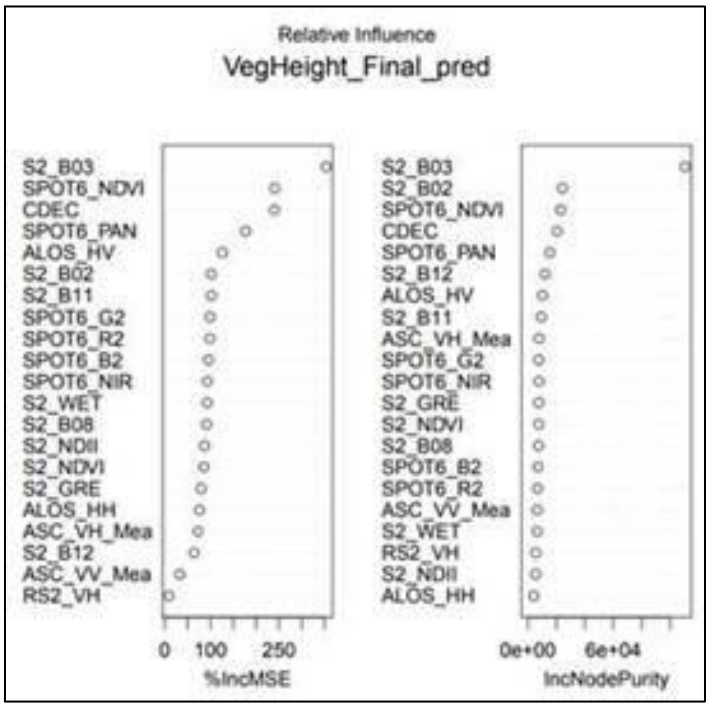

Figure 3. Relative influence of the EO variables in the final Vegetation Height model

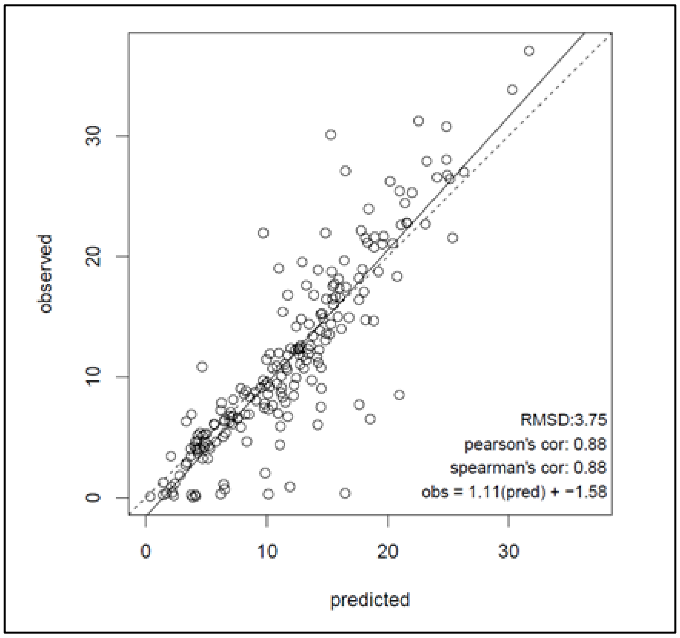

Figure 4. Correlation between predicted and observed Vegetation Height

The relative influence graph (Figure 3) reveals that the level of influence to the model for each input sensor varies. The highest contributors to the model are Sentinel-2 Band 3 (Green), SPOT-6 NDVI, Elevation (CDEC), SPOT-6 Panchromatic, and ALOS PALSAR HV radar backscatter intensity.

Ultimately, this is because the high to moderate spatial and spectral resolution of SPOT-6/7 and Sentinel-2, are sensitive to proportions of vegetation on the ground surface. Additionally, L-band radar in HV polarisation, is sensitive to volumetric scattering of emitted radar, and directly related to the varying surface texture from different vegetation types (bare ground, grass, shrubs, trees etc.) and densities.

In contrast, the RadarSat-2 VH data contributed least to the model, below the comparable C-band radar sensor, Sentinel-1. This is due to the inherent noise in radar data. The freely-available access of Sentinel-1 radar data allows for greater volumes of data (area and multi-date) to be downloaded and processed, and temporal filtering to be applies which mitigates the inherent single-scene noise. This would ultimately explain why Radarsat-2, despite having a higher spatial resolution than Sentinel-1 (8 metres, as opposed to Sentinel-1's 20 metres) contributed less to the model.

Ultimately, this chart reveals the requirement to have a multi-sensor approach to mapping vegetation height, because by combining the complimentary capabilities of each sensor type (optical and radar), the limitations of each sensor are mitigated.

The final model fit (Figure 4) selected for the final outputs reveals a good correlation between predicted and actual (observed) vegetation heights identified in the LiDAR canopy height model.

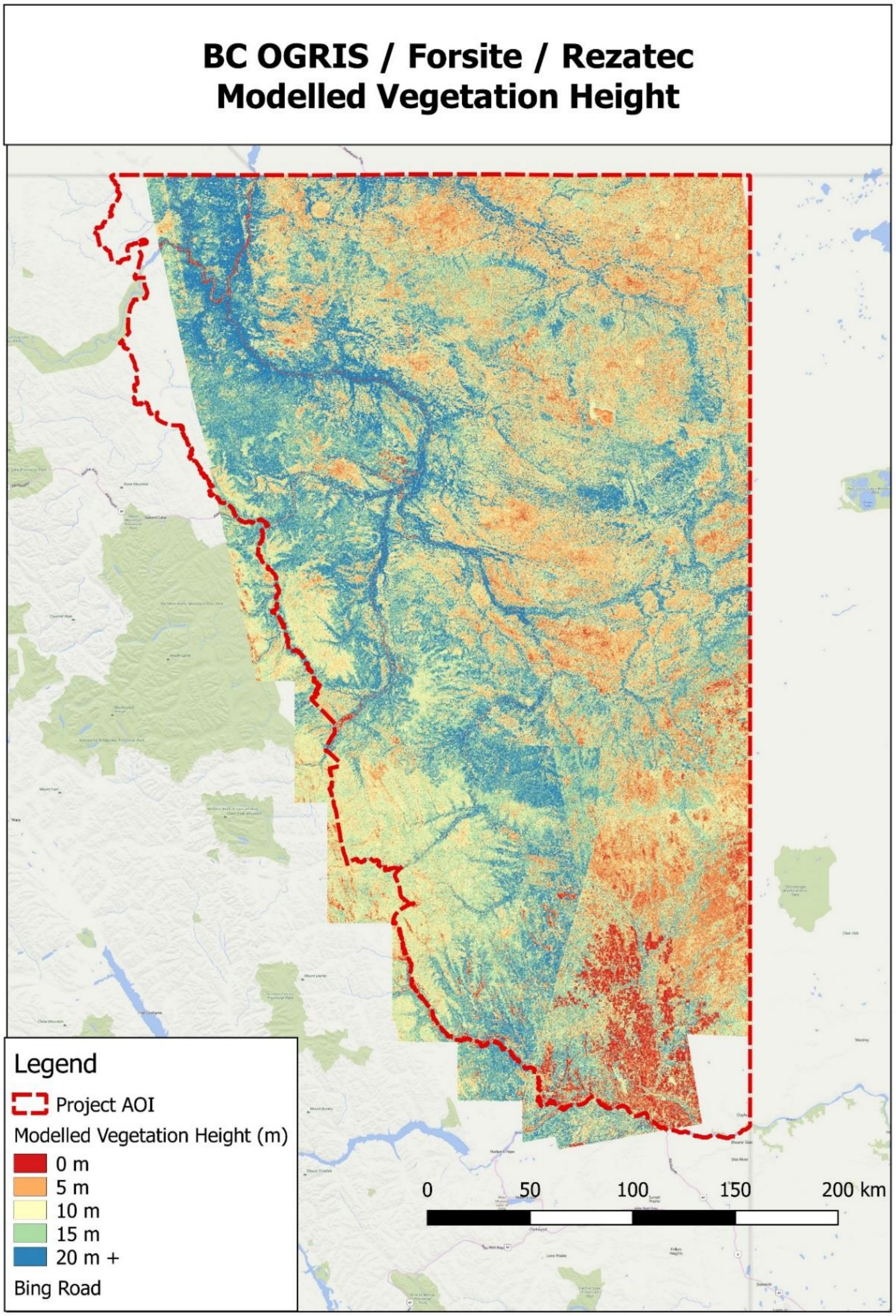

Figure 5. Project Modelled Vegetation Height raster output

3 Accuracy Assessment

The modelled vegetation height results were assessed both quantitatively, through direct comparison of a sample of modelled heights versus LiDAR CHM heights, and qualitatively, through visual comparisons between the modelled height layer and the LiDAR CHM in terms of general trends.

A set of 3,000 assessment points were generated at random across the LiDAR coverage area (independent of the model training points). Modelled heights and CHM heights were extracted at each point, plotted on a graph, and the resulting RMSE was calculated (Figure 6).

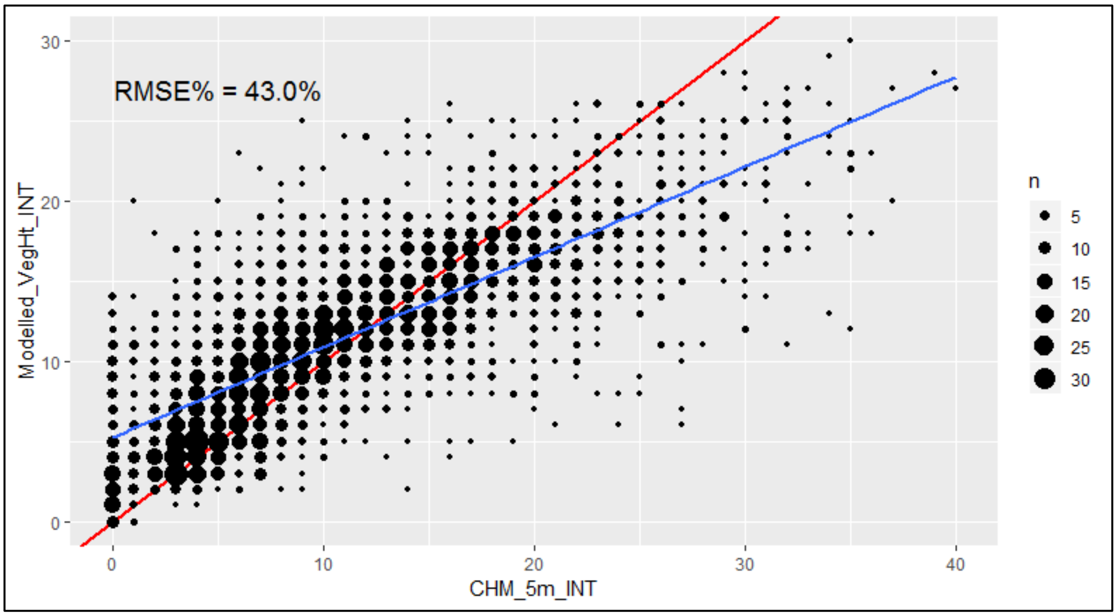

Figure 6. Correlation between modelled and CHM heights for 3,000 random points

In general, the lower modelled heights were overestimated while the higher modelled heights were underestimated (compared to the CHM).

This assessment was repeated with additional sets of stratified random assessment points generated along seismic lines and roads, and within well sites and cutblocks. The same trend was observed in all cases, and was consistent with qualitative observations between modelled heights and CHM heights across the project area (for example, see Figure 7).

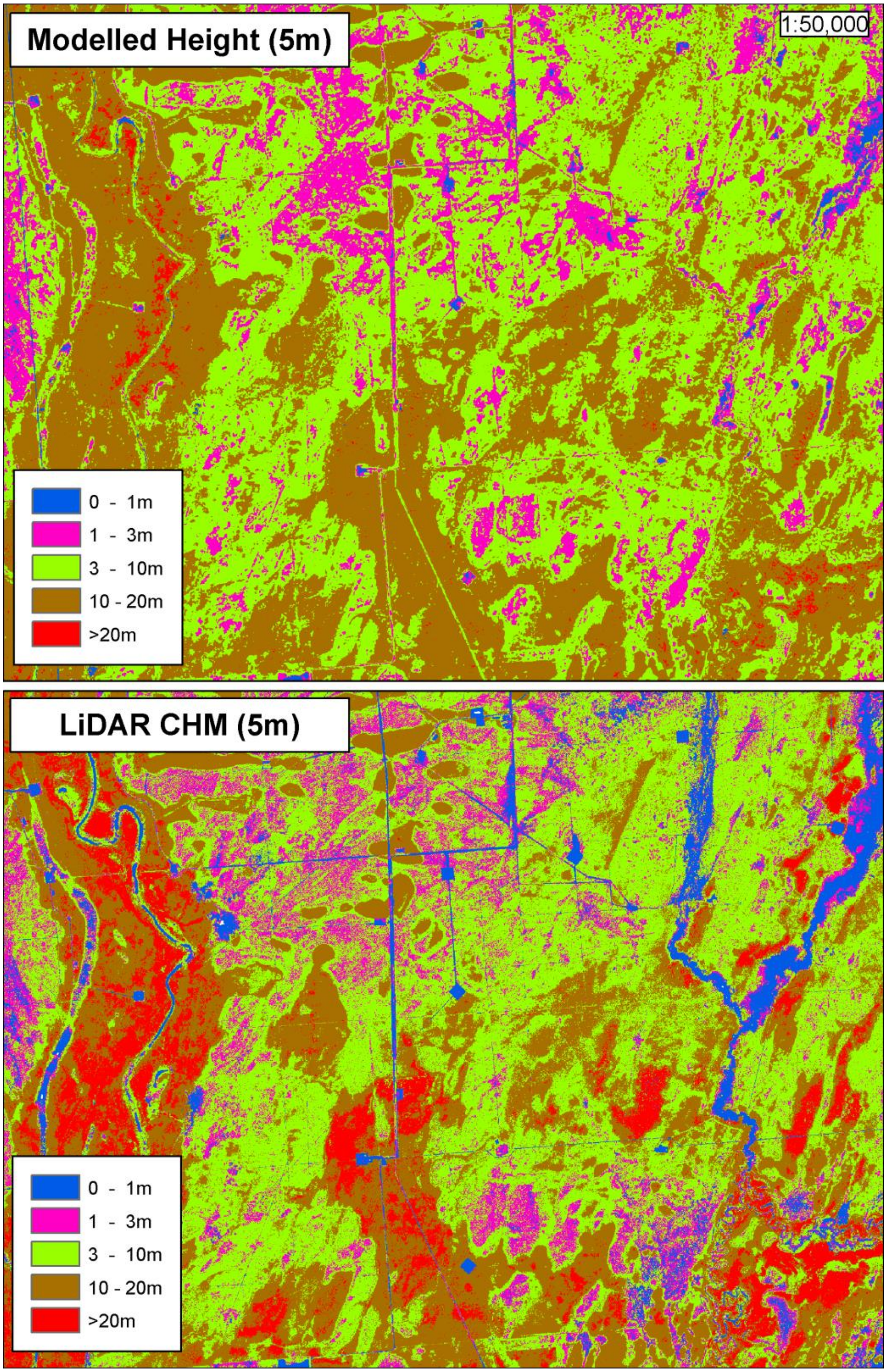

Figure 7. Comparison between modelled heights and 5m CHM heights

A best fit line using least squares regression was modelled between the modelled heights and CHM heights. An adjustment equation was then constructed and applied to the original model heights. The adjusted modelled height output more closely matched the LiDAR CHM data, but tended to generalize heights in some areas compared to the 5m CHM (Figure 8).

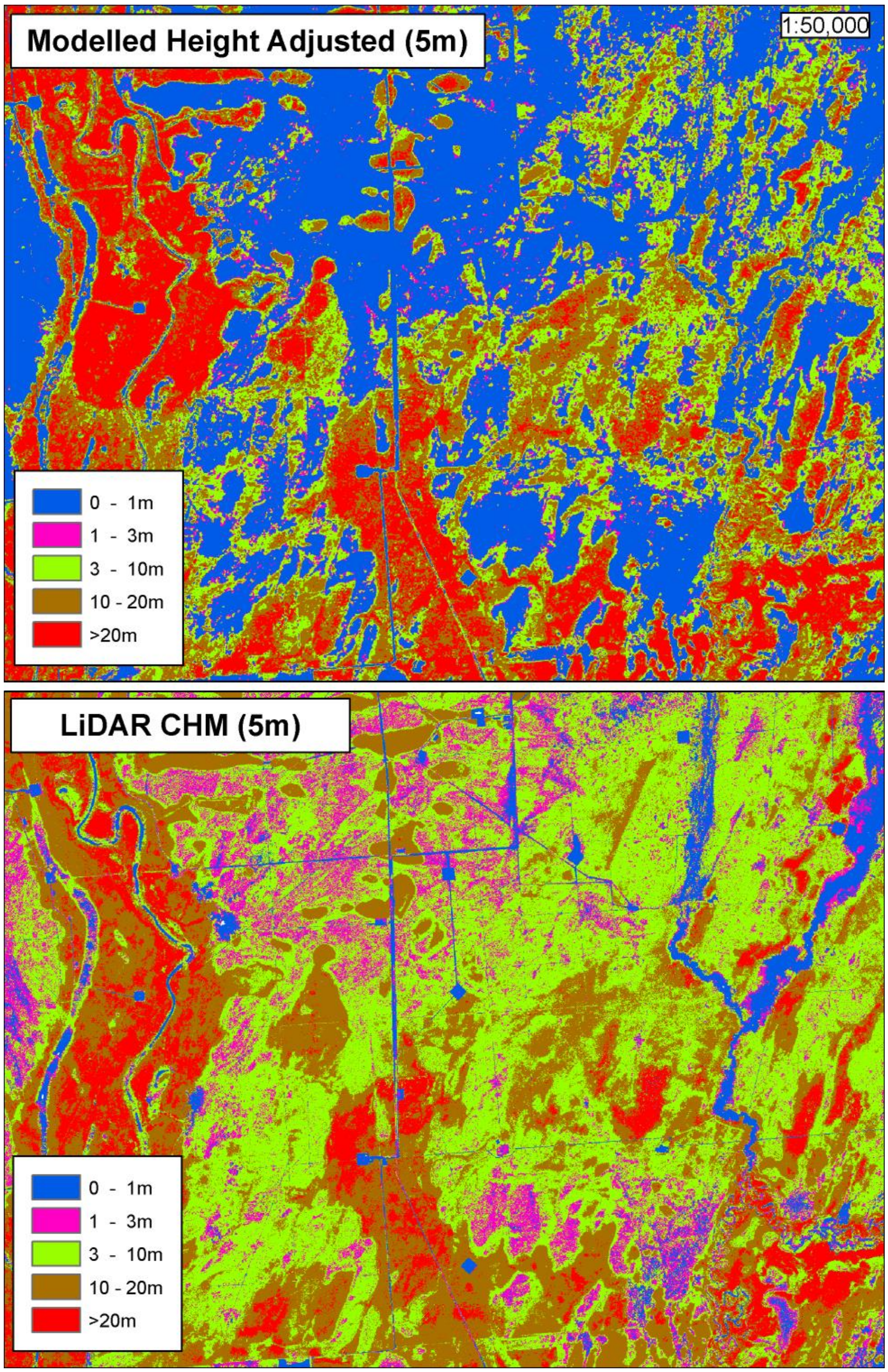

Figure 8. Comparison between adjusted modelled heights and 5m CHM heights

However, the adjusted modelled heights, when extracted from anthropogenic features only, showed reasonably close correlation to 1m CHM data (Figure 9).

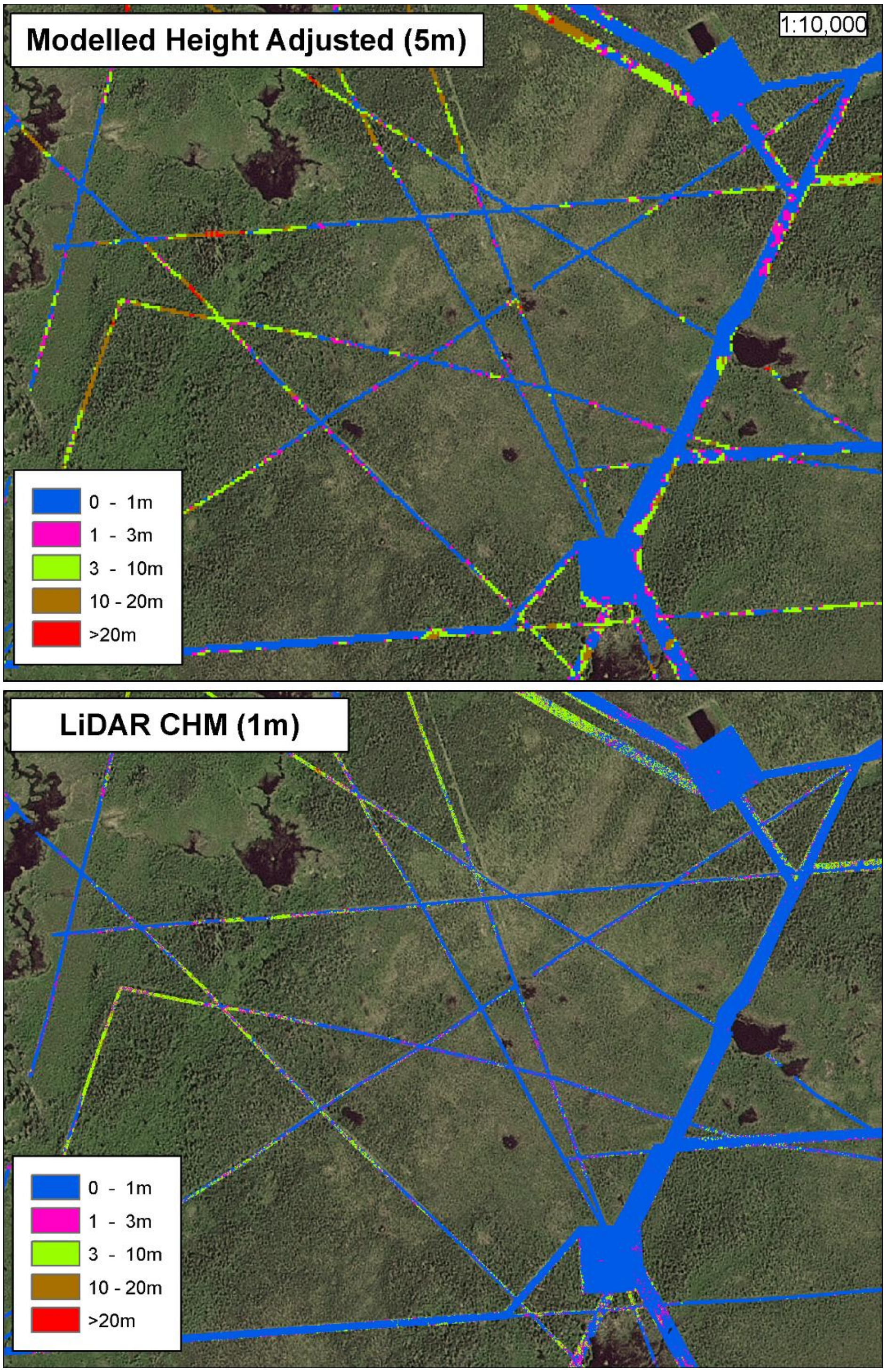

Figure 9. Comparison between modelled heights and 1m CHM heights (anthropogenic features only)

The impact of applying the adjustment equation to the original modelled heights is shown in Figure 10, which plots adjusted heights for the same point locations as those used to create Figure 6, against the LiDAR derived CHM.

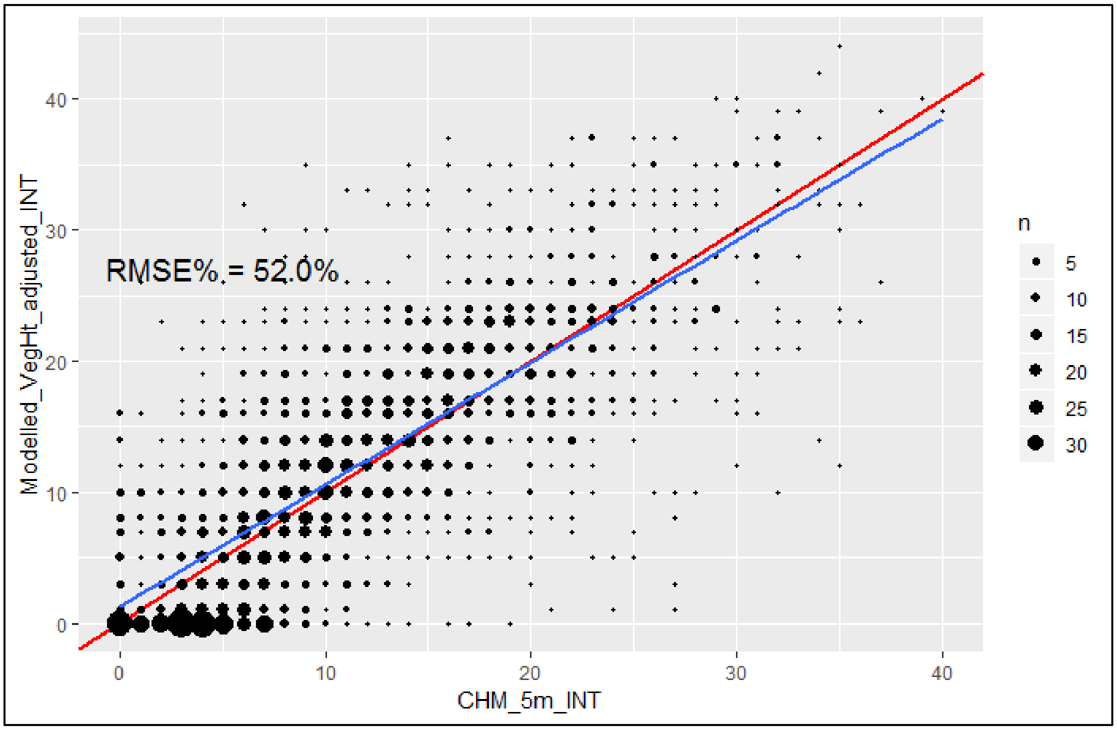

Figure 10. Correlation between adjusted modelled and CHM heights for 3,000 random points

The adjustment equation was as follows:

-9.37 + 1.78*Modelled_VegHt

4 Conclusions and Recommendations

The key primary conclusion from this project is that low-cost sensors have demonstrated the ability to model vegetation heights over large areas (millions of hectares) in a robust and repeatable approach. The model is able to capture the relative variations in all vegetation types at a landscape level and predict their height to a good level of accuracy (RMSD < 4m) in relation to the range of vegetation heights (0-40m). Compared to the relatively high cost of LiDAR, specifically the cost of acquisition and processing for very large areas, and the cost impact of high temporal revisits, the approach applied in this project provides a low-cost alternative, with the ability to monitor changes in vegetation height more frequently.

It is noted however, that the model reduces in accuracy at lower vegetation heights (less than 5 metres) and tall vegetation heights (greater than 20 metres). This is a result of the multi-sensor approach applied, and the varied range of spatial and spectral resolutions, as well as the applied method (regression based random forest machine learning algorithm). The relatively coarse inputs which influenced the model were in excess of 10 metres, thereby encapsulating a variety of vegetation heights within their spatial extent (known as the mixed pixel effect).

At a landscape level, modelled heights are useful for showing canopy information, major disturbances and vegetation patterns. However many of the seismic features are below the spatial resolution of the height model (5m) and therefore the modelled height outputs do not have sufficient detail to estimate regrowth on all the seismic lines in Northeast BC. As such, the output is primarily of benefit for larger linear disturbances and clearings.

Recommendations to further improve the output modelled vegetation height include improved training data variables (spatially and / or spectral) and alternative modelling approaches. One option for improving the spatial and spectral definition within the model, based on the outputs of this project, would be the inclusion of ALOS-2 L-Band SAR data at higher spatial resolution (3 metres as opposed to 12.5 of ALOS PALSAR). L-Band SAR played a significant factor in the prediction capability of the model, due to its interaction with vegetation structure (branches and stems), and therefore, improved spatial resolution of this data would improve the relative spatial accuracy of the model output. In addition to this, multi-temporal RADARSAT-2 at higher spatial resolution (Ultra-Fine at 3m) would account for single-scene noise and improved spatial definition.

Ultimately, given the resolutions of the sensors incorporated in the model, the result of this project is a very good estimation of vegetation heights over a vast area, giving valuable information for rapid assessment of the current state of vegetation regrowth along the entire anthropogenic framework, and subsequent prioritisation of further ground-based measures in identified 'high-concern' areas. Whereas the alterative would be labour-intensive field-based inspection of all linear features, or lidar acquisitions over the whole network, both of which are cost prohibitive. In addition, the benefits of secondary benefits (e.g. alternative uses such as cut-clock assessment) and the ability to readily scale the product increase its current value to caribou habitat restoration monitoring.