Application for the ESRI Scholarship 2013

As a PhD student in the Integrated Remote Sensing Studio at UBC, I study the impact of disturbances on forest structure in the Canadian boreal using optical remote sensing and Light Detection and Ranging (Lidar) data. Specifically, I am interested in improving our understanding of vegetation response to disturbance in order to better understand how boreal forests may be altered by a changing climate. I use ArcGIS and Python extensively in my research to explore the spatial processes that drive forest structure in the Canadian boreal. Prior to coming to UBC, I received a Bachelor’s degree in Environmental Science and a Master’s degree in Remote Sensing and Geographic Information Systems (GIS) from Boston University. During my Master’s, I used remote sensing and GIS to develop improved methods of crop yield forecasting in the US and to detect the spread of insect infestations in New England.

Here, I present a study that explores the agreement between global canopy height maps and canopy height estimates derived from airborne Lidar data over Canada. This analysis utilizes a combination of ArcGIS, Python, R programming and PostgreSQL tools to efficiently process and analyze large quantities of spatial data. I hope to demonstrate and share how powerful this combination of tools can be for processing large datasets with other ESRI users at Vancouver’s Regional User Conference. This study was submitted for publication in the *Canadian Journal of Remote Sensing* and presented at the SilviLaser 2012 Conference to reach an international audience of researchers interested in estimating carbon storage in forests.

Carbon storage in forest aboveground biomass is a critical, yet difficult, component of the global carbon cycle to estimate. Canopy height, a key indicator of site carbon storage, can be estimated with Lidar data collected by the Geoscience Laser Altimeter System (GLAS) aboard the Ice, Cloud and land Elevation Satellite (ICESat). As GLAS does not provide spatially exhaustive coverage of the world’s forests, two recent studies developed wall-to-wall canopy height maps by extrapolating GLAS-derived height estimates with spatially contiguous variables. Lefsky (2010) extrapolated GLAS derived heights using segmented spectral data from the Moderate Resolution Imaging Spectroradiometer (MODIS), while Simard et al. (2011) used climatic, topographic and other globally available ancillary variables. As these products can complement field-based inventories and provide valuable information on carbon storage in forests, their accuracy and the errors they will introduce to carbon models needs to be further assessed and better understood. Therefore, I investigated the accuracy of these products over Canada by comparing them against higher resolution estimates of canopy height derived from airborne Lidar data.

In the summer of 2010, the Canadian Forest Service (CFS) acquired 34 transects of small-footprint discrete return airborne Lidar data over Canada, spanning from Newfoundland in the east to the Yukon in the west (Figure 1a, Wulder et al. 2012). The Lidar data was divided into 25- by 25-m plots and a suite of Lidar metrics describing the distribution and density of Lidar returns was calculated for each plot (Wulder et al. 2012). For this analysis, the 95th height percentile of Lidar returns within each plot was selected as a measure of canopy height to compare against the Lefsky (2010) and Simard et al. (2011) global canopy height maps. Height percentiles describe the vertical distribution of Lidar returns and correlate to forest attributes such as mean tree height and aboveground biomass. The 95th height percentile, along with other Lidar metrics (see Wulder et al. 2012 for a complete list of metrics) were stored in a PostgreSQL database.

To allow for a direct comparison between Lefsky (2010) and Simard et al. (2011) canopy height estimates, the products were first reprojected using nearest neighbor interpolation into Lamberts Conformal Conic projection in ArcGIS (Figure 1b and 1c). A 925-m grid was chosen, which closely corresponds to the 30 arc-second resolution of the Simard et al. (2011) product over the study area. While the Simard et al. (2011) height product was derived at a constant resolution of 30 arc seconds, height estimates in the Lefsky (2010) product were derived for forest “patches” that averaged 25 km^2 in size. So while the Lefsky (2010) height product was delivered at 500-m spatial resolution, a minimal amount of information was lost by scaling the product to 925-m.

In order to compare the airborne Lidar height estimates (i.e., the 95th height percentile) against the Lefsky (2010) and Simard et al. (2011) height maps, the airborne Lidar data needed to be scaled from the 25-m plot to the 925-m pixel scale. As the Lidar dataset consisted of over 17 million 25-m plots and plot coordinates were stored in 192 point files, a strategic approach was needed to perform this analysis. First, a raster was created with the same extent and resolution (925-m) as the reprojected GLAS products. This raster contained a unique value for each pixel, which would serve as a pixel ID. This raster was converted into a polygon file, and iteratively intersected with each of the 192 point files in Python using the arcpy library. This assigned each of the 17 million Lidar plots with a pixel ID that corresponded to the 925-m pixel to which it belonged. Next, the pixels that contained Lidar plots were identified, and an empty text file was produced for each pixel (39,525 text files). By utilizing the arcpy and PostgreSQL libraries in Python, Lidar metrics were queried from the PostgreSQL database for each of the 17 million plots and stored in the corresponding text file (i.e., plot data were stored according to pixel ID). The text files were then read into the R statistical package, where the Lidar metrics were averaged from the 25-m plot to the 925-m pixel scale. To be included in the 925-m pixel average, a plot needed to satisfy several rules. The plot needed to be forested according to the Earth Observation for Sustainable Development of Forests (EOSD) land cover classification (this information was stored for each plot in the PostgreSQL database). The plot needed to have a vegetation cover of at least 10%, which was determined as the percentage of airborne Lidar first returns that were above 2 m. The plot needed to have a 95^{th} height percentile below 50 m, as values above 50 m were assumed to be erroneous – too tall – in the study area. Finally, a 'spatial uniqueness' test was performed on the Lidar plots to insure that no areas were double counted in locations where flight lines crossed.

To determine which 925-m pixels were used in the comparison, several conditions needed to be met. First, the pixel needed to contain both a Lefsky (2010) and Simard et al. (2011) height estimate above 2 m. Second, the pixel needed to contain at least 100 25-m Lidar plots. Finally, the pixel needed to be at least 75% forested according to the EOSD, to ensure the analysis was restricted to forested areas. The percentage of each 925-m cell that was forested according the 25-m EOSD classification was calculated using focal statistics in the arcpy library. This resulted in a total of 8,656 pixels suitable for comparison across the three data sources. The root mean square error (RMSE) between the canopy height maps and the 95^{th} height percentile from airborne Lidar was then calculated for all ecodistricts that contained more than 50 925-m pixels. Ecodistricts represent the finest scale division of geomorphology, soil, vegetation and climate characteristics defined by the Ecological Stratification Working Group (1995).

As terrain slope has been shown to adversely affect canopy height estimates from GLAS data, we compared ecodistrict RMSE against terrain roughness (i.e., the standard deviation of elevation within each ecodistrict according to the Canadian Digital Elevation Dataset, CDED). As the CDED is produced at a high spatial resolution (0.75 arc seconds) and delivered for individual map sheets, mosaics were iteratively created for each ecodistrict using the arcpy library and the standard deviation of elevation was calculated for each individually (Figure 2).

The 95^{th} height percentile from airborne Lidar agreed more closely to the Simard et al. (2011) height estimates than the Lefsky (2010) estimates over the sampled Canadian ecodistricts. The RMSEs for the Lefsky (2010) product (Figure 1d) were large (between 9-15 m) in many ecodistricts, which could translate into potentially large errors in carbon storage estimates for carbon modeling activities. The RMSEs for the Simard et al. (2011) product (Figure 1e) were highest in ecodistricts above 60° N, most likely because GLAS data above 60° N were not used in the Simard et al. (2011) product. Figure 3a reveals that many of the ecodistricts with high RMSEs for the Lefsky (2010) product occurred in areas with high terrain roughness, while Figure 3b reveals that Simard et al. (2011) product performs better in these areas of high slope. The large discrepancies between the Lefsky (2010) and Simard et al. (2011) datasets over Canada highlight how differences in data processing and approaches can lead to large differences in canopy height estimates. The outputs of this analysis are important for determining which approaches are most successful in various ecosystems. These results will better inform the current users of these global products as well provide insights for the generation of future large area height products with Lidar, models, and remotely sensed data.

Ecological Stratification Working Group 1995. A national ecological framework for Canada. Centre for Land and Biological Resources Research, Research Branch, Agriculture and Agri-Food Canada. Avaliable from http://sis.agr.gc.ca/cansis/publications/ecostrat/cad_report.pdf

Lefsky, M.A. 2010. A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. *Geophysical Research Letters*, Vol. 37, L15401, pp. 1–5.

Simard, M., Pinto, N., Fisher, J.B., and Baccini, A. 2011. Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research*, Vol. 116. doi: 10.1029/2011JG001708.

Wulder et al. 2012. Lidar plots — a new large-area data collection option: context, concepts, and case study. *Canadian Journal of Remote Sensing*, Vol. 38, No. 5, pp. 600–618.

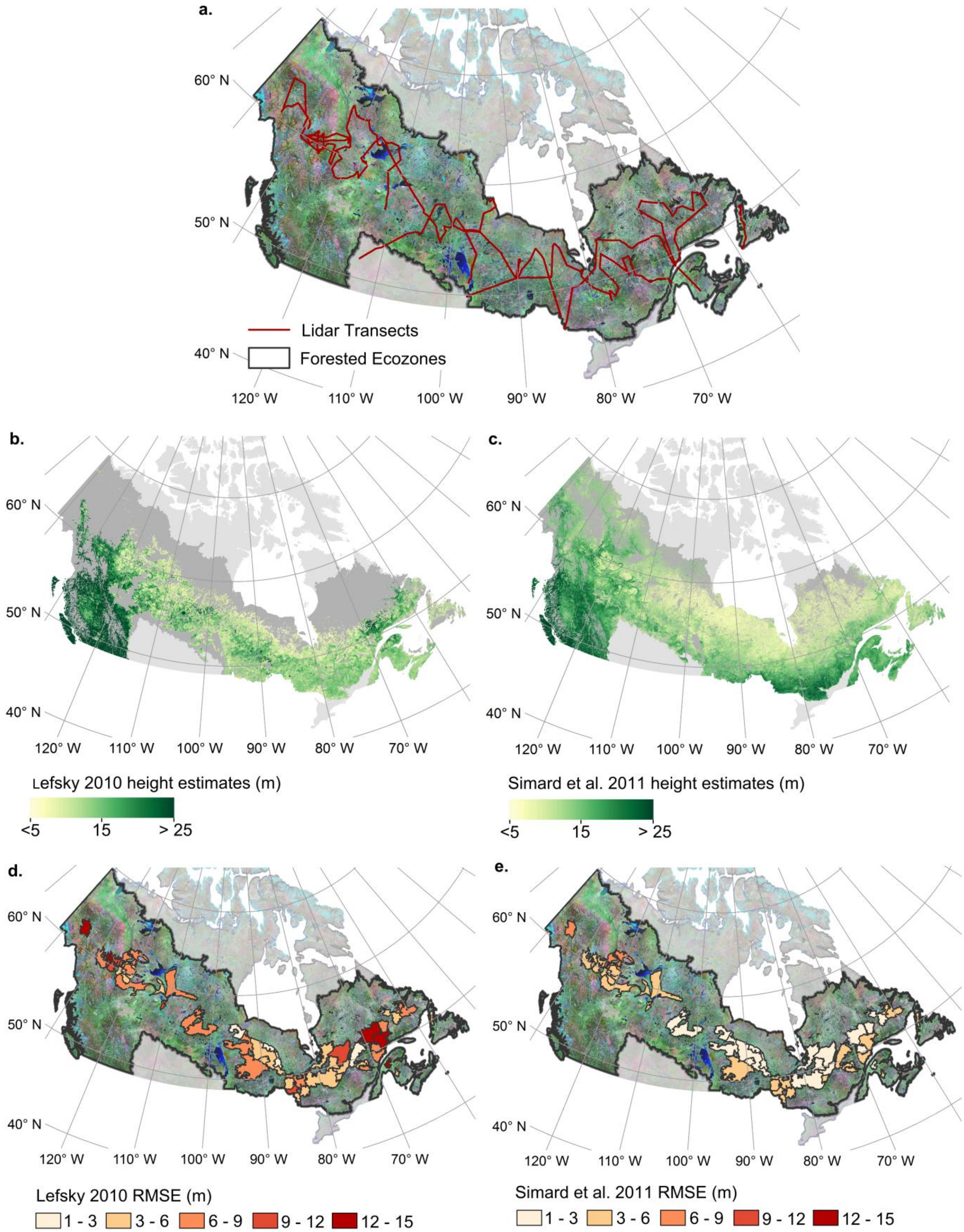

Figure 1. a) Lidar transects collected by the CFS in 2010 **b)** Lefsky 2010 height estimates over Canada **c)** Simard et al. 2011 height estimates over Canada **d)** Ecodistrict-level RMSE between Lefsky 2010 and airborne Lidar height estimates **e)** Ecodistrict-level RMSE between Simard et al. 2011 and airborne Lidar height estimates

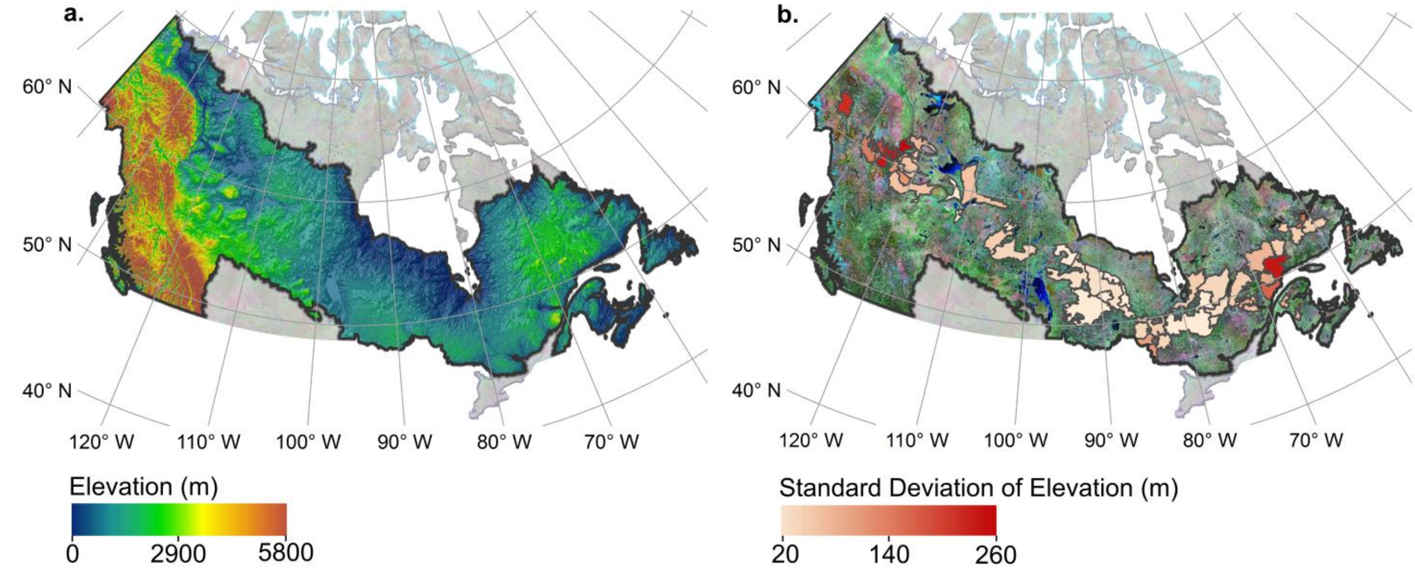

Figure 2. a) Elevation according to the GTOPO30 global digital elevation model **b)** The standard deviation of elevation within each sampled ecodistrict according to high resolution CDED elevation data

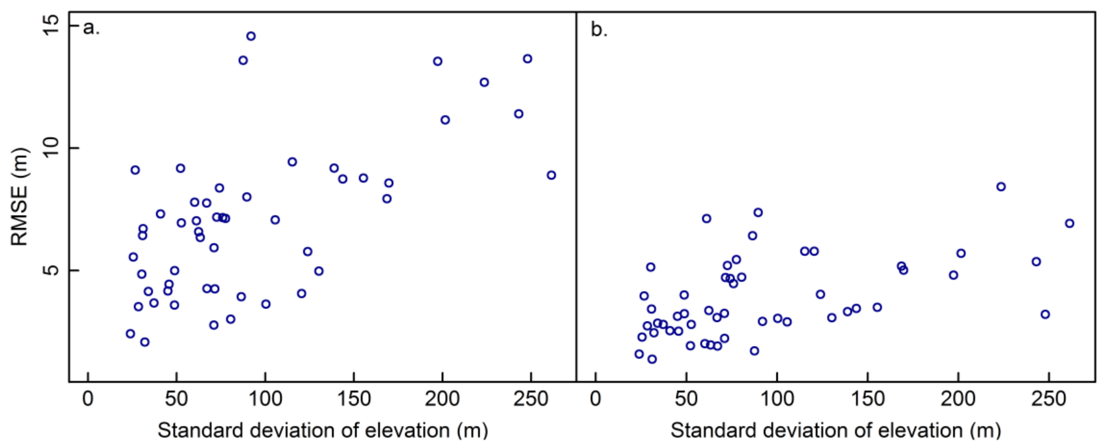

Figure 3. Relationship between ecodistrict-level RMSE and the standard deviation of elevation for **a)** Lefsky 2010 and **b)** Simard et al. 2011