Article

Leaf Area Index (LAI) Estimation in Boreal Mixedwood Forest of Ontario, Canada Using Light Detection and Ranging (LiDAR) and WorldView-2 Imagery

Graham Pope * and Paul Treitz

Department of Geography, Queen's University, Kingston, ON K7L 3N6, Canada;
E-Mail: paul.treitz@queensu.ca

* Author to whom correspondence should be addressed; E-Mail: gpope@outlook.com;
Tel.: +1-416-657-2367 (ext. 225); Fax: +1-613-533-6122.

Abstract: Leaf Area Index (LAI) is an important input variable for forest ecosystem modeling as it is a factor in predicting productivity and biomass, two key aspects of forest health. Current *in situ* methods of determining LAI are sometimes destructive and generally very time consuming. Other LAI derivation methods, mainly satellite-based in nature, do not provide sufficient spatial resolution or the precision required by forest managers for tactical planning. This paper focuses on estimating LAI from: (i) height and density metrics derived from Light Detection and Ranging (LiDAR); (ii) spectral vegetation indices (SVIs), in particular the Normalized Difference Vegetation Index (NDVI); and (iii) a combination of these methods. For the Hearst Forest of Northern Ontario, *in situ* measurements of LAI were derived from digital hemispherical photographs (DHPs) while remote sensing variables were derived from low density LiDAR (*i.e.*, 1 m^{-2}) and high spatial resolution WorldView-2 data (2 m). Multiple Linear Regression (MLR) models were generated using these variables. Results from these analyses demonstrate: (i) moderate explanatory power (*i.e.*, $R^2 = 0.53$) for LiDAR height and density metrics that have proven to be related to canopy structure; (ii) no relationship when using SVIs; and (iii) no significant improvement of LiDAR models when combining them with SVI variables. The results suggest that LiDAR models in boreal forest environments provide satisfactory estimations of LAI, even with narrow ranges of LAI for model calibration. Models derived from low point density LiDAR in a mixedwood boreal environment seem to offer a reliable method of estimating LAI at high spatial resolution for decision makers in the forestry community. This method can be easily incorporated into simultaneous modeling efforts for forest inventory variables using LiDAR.

1. Introduction

The Boreal Forest of Canada covers over 300 million hectares, stretching more than 1,000 km from the Atlantic to Pacific coasts [1]. Ontario alone is covered by more than 71 million hectares of forest, resulting in almost two-thirds of the land base being forested, over half of which is in the Boreal region [2]. As one of the dominant features of this country's landscape, the careful monitoring, management and protection of the Boreal Forest is crucial for sustaining balance in this biome. However, the vast extent requires cooperation between federal, provincial and territorial governments, commercial foresters and conservationists in order to "*...manage Crown forests to meet social, economic and environmental needs of present and future generations*" [3].

In order to monitor vegetation health and sustainability at strategic (*i.e.*, future planning) and tactical (*i.e.*, short term) scales it is important to properly select vegetation metrics that are robust and easy to measure. The *State of the Forest Report* [2] lists, among other things, primary productivity as a key indicator of forest sustainability. Net primary productivity (NPP), the rate at which an ecosystem accumulates biomass, is a good long term gauge of ecosystem health due to its reliance on a combination of basic ecosystem drivers: water, nutrient availability and sunlight [4]. Tracking NPP over time can provide managers with a relative estimate of forest health and long-term growth.

When focusing on the need to monitor forest health for commercial or conservation applications, we are interested in assessing and monitoring primary productivity or biomass over large areas. Leaf Area Index (LAI), a key input to productivity models, can be estimated using a variety of remote sensing techniques (e.g., [5–7]). The focus of this research is on the ability of passive (*i.e.*, high resolution multispectral satellite data) and active (*i.e.*, Light Detection and Ranging (LiDAR)) sensors to estimate LAI. The results of this research will provide insight into the accuracy and precision of these two remote sensing data types for estimating LAI over large tracts of boreal mixedwood forest in Ontario as well as the utility of each for providing suitable inputs to productivity models.

LAI has been used in many studies to derive, or correlate to, primary productivity or biomass (e.g., [8–11]). Current productivity and biomass estimates from provincial government organizations use only growth estimate statistics and intensive ground sampling to derive biomass estimates from models [2]. Allometric measurements (*i.e.*, using known relationships between parts of an organism and its whole) of tree height or diameter at breast height (DBH) rely on statistical distributions derived from prior destructive sampling. This technique is one of the fastest *in situ* methods of providing forest mensuration data such as biomass, but has problems relating to phenology and variation in environmental conditions from the original curve-defining measurements [12].

LiDAR LAI models have the potential to provide fast, repeat assessments of a forest, returning variables that can provide accurate estimations of biomass and productivity. Studies have been carried out around the world that assess LAI using LiDAR (e.g., [13–15]). Building on those studies, this research will examine techniques for a Canadian boreal mixedwood setting using relatively low point density LiDAR, a requirement for cost effective, large area forest resource inventories.

The other aspect of this research, *i.e.*, examining spectral vegetation indices (SVI) and their correlations to LAI, has been well documented, but primarily for lower resolution satellite sensors (*i.e.*, Landsat, MODIS and AVHRR) (e.g., [16–19]). Here, we will examine high spatial resolution satellite imagery (*i.e.*, WorldView-2) to determine the relationship of SVIs to plot level estimates of LAI. These relationships can then be compared to estimates modeled from LiDAR data. Forest managers are eager to have the ability to create products such as these at improved spatial and temporal resolutions [20].

It is hypothesized that there will be a statistically significant relationship between *in situ* measurements of LAI and low point density LiDAR data and/or high resolution Normalized Difference Vegetation Index (NDVI) data. Models of LAI combining these two data sources will also be examined to determine the utility of integrating spectral and structural information for modeling LAI.

2. Materials and Methods

2.1. Study Area

The field campaign to collect *in situ* digital hemispherical photographs (DHPs) for model calibration/validation took place in June and July 2011 in the Hearst Forest, centered roughly on Hearst, Ontario, Canada (49.7°N, 83.7°W) (Figure 1). This forest falls within the boreal mixedwood region and covers approximately 1.23 million ha; 1.00 million ha of which is productive forest [21]. It is an actively managed, commercial forest, with approximately 60,000–70,000 ha harvested annually [22].

**Figure 1.** Map depicting Hearst, Ontario and surrounding cities.
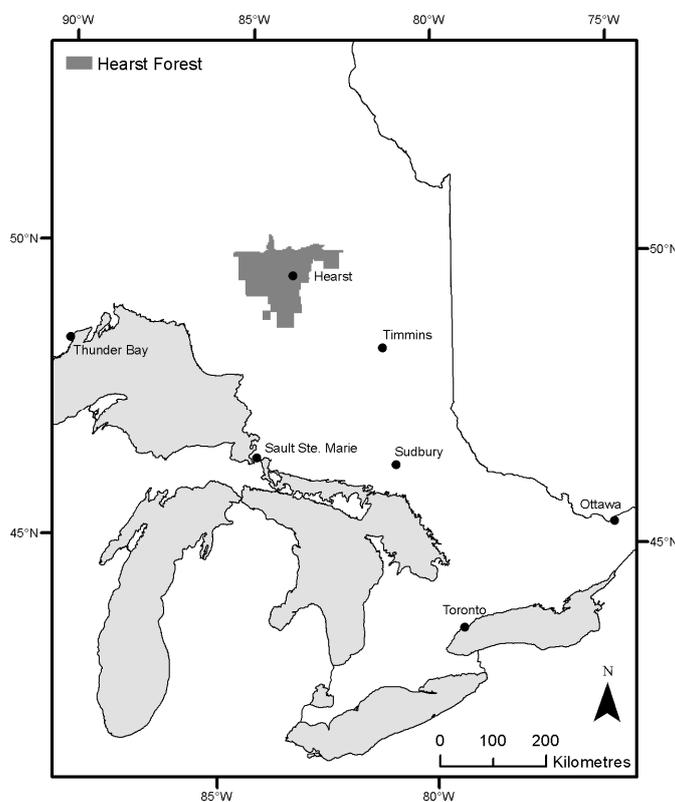
The majority of the tree species in the Hearst Forest are coniferous, with black spruce (*Picea mariana* Mill. B.S.P.) present in approximately two-thirds of the forest units. The coniferous species in

this forest also include jack pine (*Pinus banksiana* Lamb.), white spruce (*Picea glauca* Moench Voss), balsam fir (*Abies balsamea* L. Mill.) and tamarack (*Larix laricina* Du Roi K. Koch). Deciduous species in the study area include white birch (*Betula papyrifera* Marsh.), trembling aspen (*Populous tremuloides* Michx.) and balsam poplar (*Populous balsamifera* L.).

The plots selected for this project were extracted from a pre-existing pool of 446 plots. These plots were established in 2010 in support of two integrated LiDAR projects, *i.e.*, Geomatics for Informed Decisions (GEOIDE) and Advanced Forest Resource Inventory Technologies (AFRIT). Each plot was classified by forest unit and growth stage. The circular 0.04 ha plots (*i.e.*, 11.3 m radius) were sampled for multiple height and wood fiber variables in 2010 [23]. Of these plots, 249 were chosen to provide a representative range of forest unit types, growth stages, and basal area to ensure a range of canopy closures. On a secondary level, these chosen plots were examined again to ensure as even spatial distribution as possible over the forest, while still maintaining ease of access from both the town of Hearst and accessible logging roads (Figure 2).

Figure 2. Plot distribution in the Hearst Forest.
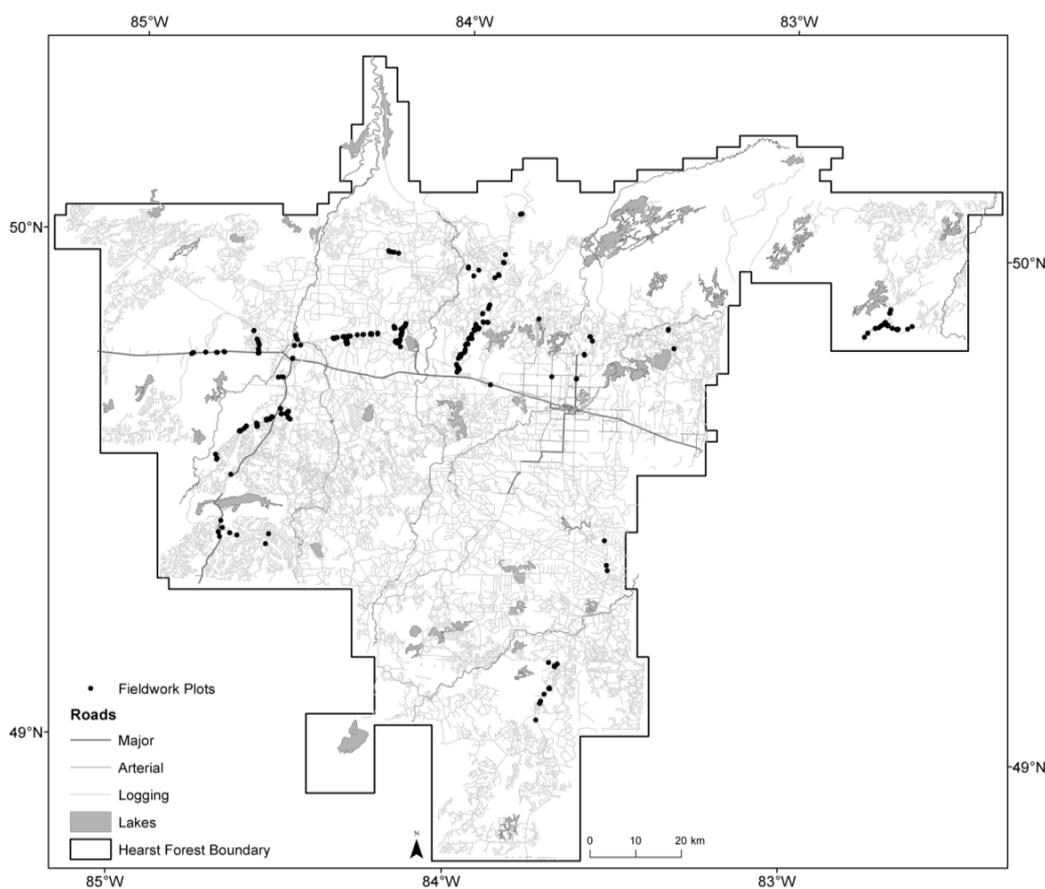
2.2. Remotely Sensed Data

2.2.1. LiDAR Data Acquisition

LiDAR data were acquired by North West Geomatics Ltd. during the period of 4 July to 4 September 2007 during leaf-on conditions. LiDAR data were collected using an Optech ALS50 sensor mounted in a Cessna 310 aircraft. These data were discrete pulse, with up to four returns measured.

The properties of the LiDAR acquisition are summarized in Table 1, with the resulting LiDAR point density being, on average, 0.81 m^{-2} across the Hearst Forest. Filtered and classified data were provided by North West Geomatics Ltd to partners at the Ontario Ministry of Natural Resources, who then generated the full suite of LiDAR predictor variables, including height and density metrics (Table A1). Variables that may be unfamiliar include D1 through 9 and cc0 through 28. DX is a canopy density metric that provides the cumulative proportion of returns found in 10 bins equally spanning the range of heights. ccX is the crown closure from X meters upwards through the canopy. Values are closest to 100% near ground level and decrease with height. These calculations were done at both the plot level and the forest level at a 20 m pixel resolution.

Table 1. Light Detection and Ranging (LiDAR) acquisition properties [24].

| Property	Value
Pulse Rate	119 KHz
Scan Rate	32 Hz
Field of View	30 Degrees
Flying Height	2,400 m
Line Spacing	1,000 m
Overlap	20%
Point Density	0.81 m^{-2}
Vertical Accuracy	<30 cm
Horizontal Accuracy	

2.2.2. WorldView-2 Data Acquisition

One hundred km^2 tiles of WorldView-2 data were acquired on 26 June, 2011 between 13:02 and 13:03 local time (*i.e.*, each tile was acquired on the same orbital path). These areas included eight multispectral channels at 2.0 m spatial resolution along with a panchromatic band at 0.5 m. These areas were selected to maximize the number of plots sampled, *i.e.*, 122 of 249 plots fall within the WorldView-2 coverage (Figure 3).

WorldView-2 images were delivered in five segments spanning the three areas of interest. Image calibration was first done to convert relative radiance to absolute radiance, then conversion of the raw. TIFF image file to .BIL (binary interleaved by line) to prepare the image for atmospheric correction. Atmospheric correction was performed using the FLAASH module in ENVI© *5.0* (Boulder, CO, USA) with the atmospheric model set to "sub-arctic summer" (applicable for the 49°N latitude) and the aerosol model to "rural." Ground elevations were extracted from the LiDAR derived digital elevation model (DEM) of the Hearst Forest. Initial visibility, set to 24.1 km, was taken from the Environment Canada climate archives [25]. The tiles were then mosaicked, with the *mean* blending operator used for overlapping areas, to make further processing more efficient. An NDVI was calculated using the red and near infrared-1 bands as seen in Equation (1). Mean NDVI for each 0.4 ha plot was extracted and added to the table of LiDAR predictor variables for each plot.

$$NDVI = \frac{NIR1 - Red}{NIR1 + Red} \tag{1}$$

Figure 3. WorldView-2 satellite image coverage in the Hearst Forest.
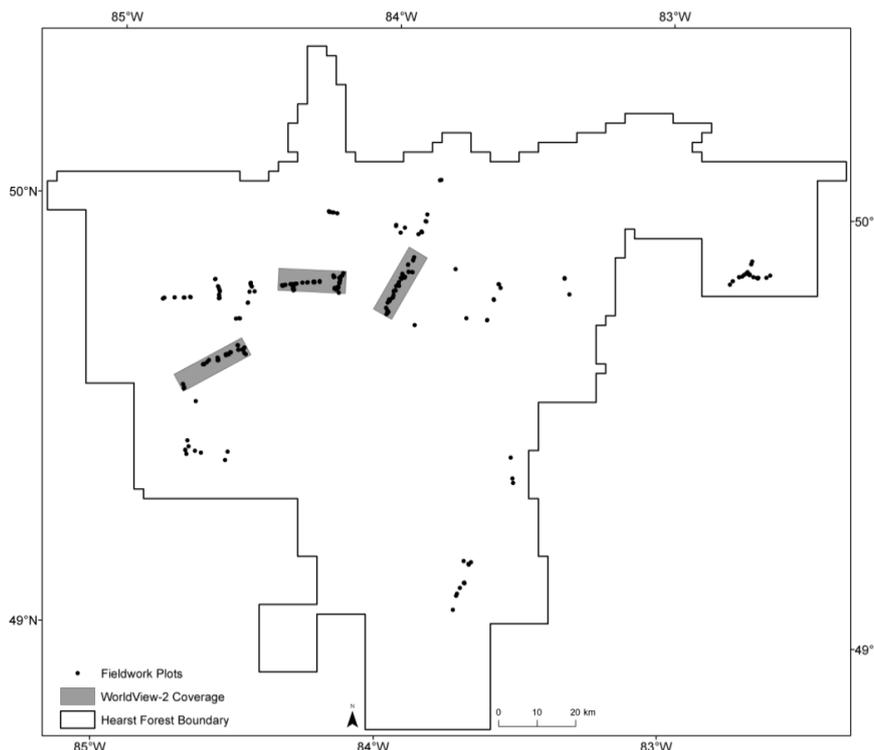
2.3. Digital Hemispherical Photography Collection

At each plot, DHPs were taken in order to estimate LAI using methods described in the DHP/TRACwin software manual [26] and by Leblanc *et al.* [6]. The sampling design used was a modification of several different patterns found in the literature, all designed specifically for the particular area or research question (e.g., [14,26–28]). Due to our circular plots with radii of 11.3 m, a gridded, 3-by-3, north-aligned design was selected with the central image being taken at plot center (Figure 4). Testing was completed afterwards to assess the necessity of capturing nine images and will be discussed later. The camera was oriented so that the top of the image always faced north. The 10 m spacing was measured using a Haglof Vertex© Hypsometer (Långsele, Sweden), calibrated daily.

Figure 4. Digital hemispherical photographs (DHP) sampling design.
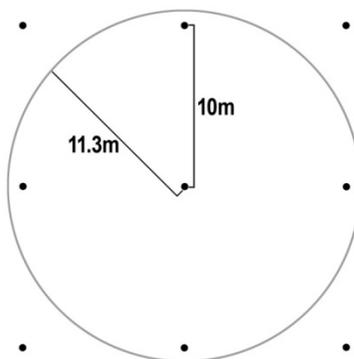
The images were taken with a Nikon© D700 (Tokyo, Japan) and Fisheye-Nikkor© 8 mm f/2.8 lens (Tokyo, Japan) (Figure 5). The D700 featured a 12.1 megapixel, FX full frame sensor required for use

with the specialized lens. The lens itself provided a 180° field of view projected as a circular image. The camera was set to take medium quality images (*i.e.*, resolution of 3,184 by 2,120 pixels) as this was greater than the image resolution recommended in the literature [6], and optimized storage requirements. Using the automatic light metering on the camera, each scene was captured attempting to under-expose the image by one or two stops of shutter speed. This exposure produced the best relationship of dark foliage/woody biomass to slightly overexposed background sky, although varying lighting and canopy conditions required constant manual adjustments. Vegetation within 1 m of the lens was moved out of frame view.

Figure 5. Setup of camera and lens system in a plot.

Images were taken with the focal point at standardized breast height (*i.e.*, 1.3 m) (Figure 5) [26]. Before taking an image, the camera was leveled using a spirit level. Most plots were topographically flat, but in plots with varied elevation or a consistent slope the camera was oriented to be level with the local ground surface, or artificial horizon, not the true horizon [26]. Ideal sampling conditions consisted of overcast, uniformly illuminated skies (*i.e.*, diffuse light). The lens was cleaned regularly to remove collected dust and fingerprints. Figure 6 provides an example of a correctly exposed DHP; foliage/biomass pixels are underexposed, *i.e.*, verging on black, and sky pixels are close to saturated, *i.e.*, near white.

Using a combination of *DHP 4.7s* and *TRACWin 5.1.0* software (Ottawa, Canada), 2,241 DHPs were processed to retrieve true LAI values. Images were first processed by "thresholding" [26]. In this process, thresholding refers to a manual process consisting of two stages. First, the logarithm of the histogram of digital numbers in the blue channel of the image was examined (Figure 7). The blue channel was used as it provides better delineation between vegetation, sky and mixed pixels [6]. The goal of this initial inspection was to place the threshold bounding the linear portion of the histogram. Then, an iterative process was used to refine the two thresholds to best represent the unique scene and zenith conditions for each of several image rings. Figure 8 shows a full resolution zoom of a comparison between a raw blue channel image and the resulting thresholded image.

Figure 6. DHP image example (Jack Pine LAI = 1.5).

Figure 7. Logarithm of digital number histogram used for thresholding and initial threshold placement (relative y-axis scaling of image digital numbers).
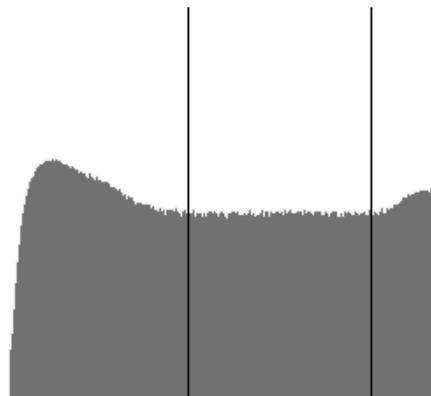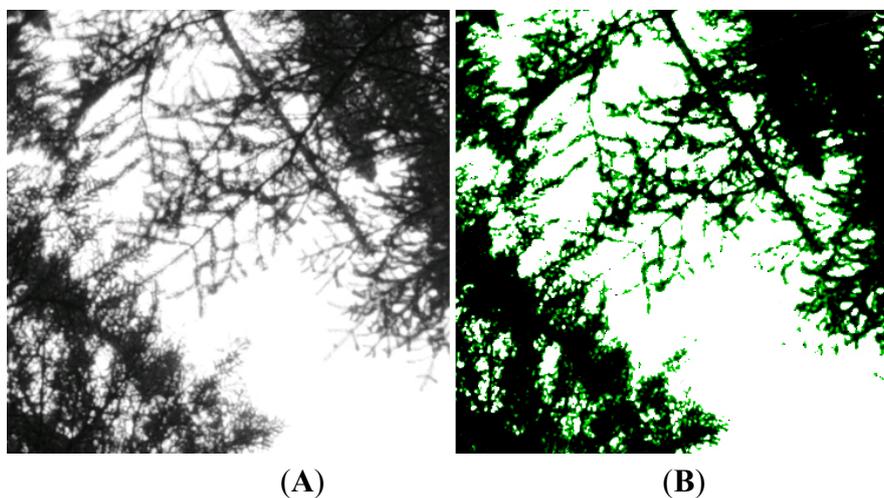

Figure 8. DHP thresholding example (1:1 resolution image (**A**); threshold image (**B**) black = vegetation, green = mixed, white = sky).

(**A**)   (**B**)

DHP provided a measure of Effective LAI (LAI$_e$). To derive LAI, data were exported from *DHP* to *TRACWin*. In this software package, two ratios for each tree species were applied to account for clumping (*i.e.*, needle:shoot ratio for coniferous species) and the expected amount of woody biomass in an image (*i.e.*, woody:total ratio) (Table 2). Species that had published values included black spruce, jack pine, balsam fir and trembling aspen [29]. The application of these ratios for other species (e.g., black spruce for tamarack and which spruce) was performed after personal communication with Drs. Jing Chen, Sylvain Leblanc and Valerie Thomas [30]. Through this consultation it was concluded that these additional species had not been verifiably measured, and that using pre-existing, similar species parameters was appropriate.

Table 2. Species TRACWin ratios.

Species	Needle:Shoot Ratio	Woody:Total Ratio
Black Spruce	1.35	0.14
Tamarack	1.35	0.14
White Spruce	1.35	0.14
Balsam Fir	1.77	0.08
Jack Pine	1.30	0.03–0.34
White Birch		0.21
Trembling Aspen		0.21
Baslam Poplar		0.21

After all images were processed, the mean of each set of nine individual LAI values for each plot was calculated to provide an estimate of plot LAI. Any plots with a mixed species composition where the minority species had greater than 20 percent coverage by basal area of trees ≥ 10 cm DBH were linearly weighted. That is to say, a plot with 85% black spruce and 15% jack pine would be treated entirely as black spruce. This methodology was consistent with LAI processing work conducted by Karin van Ewijk [31].

There is a temporal gap between LiDAR collection (2007) and DHP/WorldView-2 collection (2011), but it has been shown that annual variability in maximum LAI is relatively stable [32]. Intra-annual variability has been shown to occur based mainly on differences in cumulative degree days impacting the timing of leaf emergence. DHP collection was timed with the WorldView-2 acquisition, but LiDAR data were collected before this projects inception and therefore had to be used in its original form.

2.4. Statistical Methods

Multiple Linear Regression (MLR) was the statistical method used for modeling LAI using the LiDAR variables. A preliminary goodness-of-fit test of LAI normality resulted in identifying a normal distribution, though with distinct outliers. An assessment of outliers determined that they corresponded to plots that included balsam fir, a species whose *TRACWin* ratios were in question due to apparent over-estimation of LAI values. Removing these ten plots resulted in a statistically verified normal distribution (Shapiro-Wilk W test: Prob < W = 0.37). From the extensive suite of predictor variables (Table A1) reduction was performed using a scatterplot matrix, visually and with R^2, to avoid

multicollinearity due to the high similarity between some variables. Most notably, the percentile (*i.e.*, P10–P100 [maximum]) and density (*i.e.*, D1–D9) variables were highly correlated and were reduced to three or four variables within each set. If two or more seemingly unrelated variables were highly correlated (*i.e.*, high R^2) the more straightforward (*i.e.*, simple statistical variables over derived indices) was usually retained.

From the much smaller subset of variables, several automated techniques were used to further investigate the relationships for modeling. Forward and backward stepwise selection of variables was run using the minimum Akaike Information Criterion (AIC) as the criteria for termination. Also, an automated decision tree approach provided insight into some of the variables with the most explanatory power. Variable reduction was completed beforehand even though automated stepwise methods in *JMP©* could have been run with the entire set of predictors. This pre-screening was due to the tendency of model accuracy overestimation with large pools of predictor variables [33]. The detrimental influence of plots with exceedingly little canopy cover was noted, and a single plot with a measured LAI value < 0.1 was removed. The remaining set of plots, a total of 225, was randomly divided into two-thirds calibration and one-third validation. From the most significant variables discovered through automated forward and backward exploratory step-wise methods, manual forward stepwise selection was performed using live-updating sum of squares. Variable addition terminated when an increasing R^2 was offset by an overabundance of predictor variables and a minimized AIC [34]. Model residuals were graphed and were seen to be uniformly distributed. Three separate models were created: (i) LiDAR only predictor variables; (ii) vegetation indices; and (iii) a combination of LiDAR predictor variables with the vegetation indices. Model validation was performed using a matched pairs t-test.

3. Results

3.1. Leaf Area Index Estimation

Descriptive statistics for the final set of plots can be seen in Table 3, with the final LAI distribution in Figure 9. Over 90 percent of the LAI values fall between 1.0 and 3.5. This narrow range is not ideal for regression analyses, in spite of the fact that sampling was specifically designed to sample areas of expected high and low LAI to acquire data from within some of the anticipated "tails" of the distribution. The distribution of values is statistically normal, a prerequisite for multiple linear regression model generation.

Table 3. Leaf Area Index (LAI) statistics for final plots.

	Final Plots
Mean	2.26
Standard Deviation	0.83
Range	4.64
Minimum	0.37
Maximum	5.01
Total Count (n)	225

Figure 9. Frequency Histogram of LAI values from 225 final model plots.
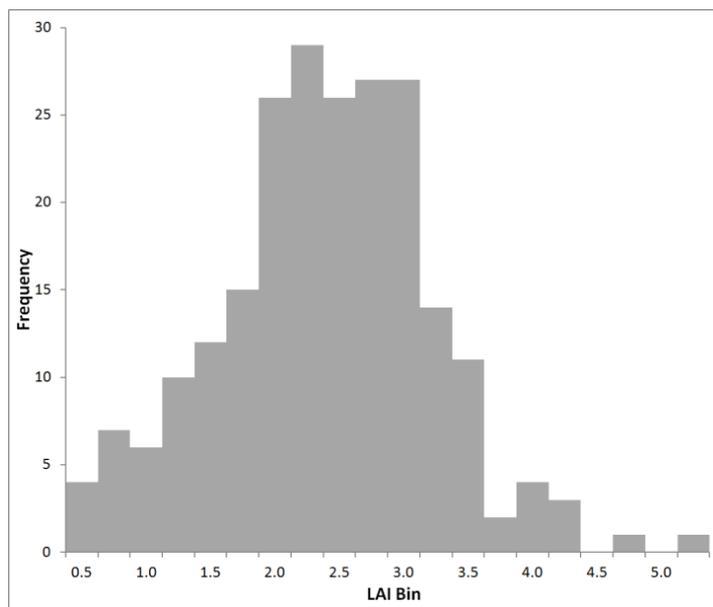
The statistics presented in Table 3 are based on the mean of all nine DHPs collected for each plot. Testing was performed to investigate the effects of the positioning of the DHPs collected at each plot (Figure 4). The same statistics using: (i) the five DHPs within the plot boundary (*i.e.*, center and four cardinal directions within the 11.3 m radius); and (ii) the center DHP alone are presented in Table 4. Over the three scenarios (*i.e.*, nine DHPs, five DHPs and one DHP), the LAI mean remains constant, likely due to the large number of samples and small range of LAI values. As the number of photographs used in the calculations at each plot decreases, the standard deviation and range of values increases. Conversely, at the intra-plot level five DHPs in the cross pattern have less variation than the nine DHPs. This trend, which seems unexpected, is likely caused by: (i) the tighter spacing of the center and immediately adjacent DHPs; (ii) the probable overlap of fields-of-view within adjacent DHPs; and (iii) the fact that these DHPs are more likely to fall within the densest central portion of potentially smaller forest stands than in differing fringe areas.

Table 4. Descriptive statistics of three trials using different combinations of DHPs to estimate LAI.

LAI Statistics	All DHPs (9)	Cross Pattern (5) (within Plot)	Center (1)
Mean	2.26	2.26	2.26
Standard Deviation	0.83	0.86	0.97
Range	4.64	4.68	6.00
Minimum	0.37	0.32	0.12
Maximum	5.01	5.00	6.12
Intra-plot Standard Deviation	0.38	0.36	--

While the statistics below demonstrate that there are only small statistical variations in LAI between sampling designs, Figure 10 illustrates in more detail the plot-by-plot differences observed between full and subset sampling. With a 1:1 line shown, there appears to be strong correlation ($R^2 = 0.84$) at the single DHP level, albeit with a moderate spread and general overestimation at higher LAI values.

The variability around the 1:1 line results from the lack of any averaging, and the overestimation is likely because of the nature of the original plot selection. A small number of plots represent smaller forest stands close to the size of the DHP grid. These few small stands were originally visited to ensure adequate species representation in the initial studies undertaken in Hearst. This size discrepancy means that the central photograph was usually placed in the densest, central part of the stand. Examining some WorldView-2 scenes in both the multispectral and panchromatic confirms that several of the overestimated plots indeed fall in these smaller pockets. When using five DHPs within the plot a near-perfect correlation ($R^2 = 0.98$) to the full nine photograph sample is observed, with minor variability around the 1:1 line and only slight overestimation at higher LAI values. From this figure it could be argued that future research in this area could be completed using a five photograph cross pattern with almost no loss of accuracy.

Figure 10. Comparison of plot LAI between the nine sample DHP average (x-axis); the five sample DHP cross pattern average (y-axis) and the single sample DHP estimate (y-axis).
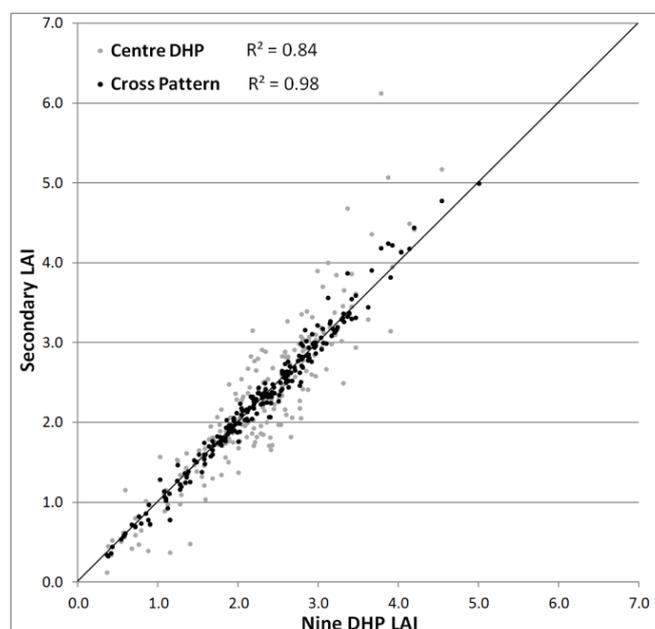
3.2. LiDAR Model Generation

From the 53 original LiDAR predictor variables, variable selection was performed using scatterplot matrices and correlation analysis. Some of the most simple exclusions were variables from different sets that had similar calculations (*e.g.*, median and P50, standard and absolute deviation). As median and P50 are actually the same value, it was simple to remove P50 from the potential list. Other variables like the Shannon Weaver Index (H) and the VCI were so closely related that it was easy to eliminate the more complex variable, H. The most difficult variables to exclude were height and density metrics, as they provided a large amount of information, but were also severely inter-correlated. Percentile variables correlated strongly with their adjacent variables (*i.e.*, P20 to both P10 and P30) ranging from R^2's of 0.80 to 0.98, and adjacent density variables were even more inter-correlated, ranging from R^2's of 0.89 to 0.97. Based on inter-correlations and correlations to LAI itself, 22 predictor variables were selected for model development (Table 5).

Table 5. Final suite of LiDAR predictor variables.

LiDAR Predictor	Description
ABS_DEV	Absolute Standard Deviation
KURTOSIS	Kurtosis
P10	First Decile LiDAR Height (m)
P40	Fourth Decile LiDAR Height (m)
P60	Sixth Decile LiDAR Height (m)
MAX	Maximum height (m)
D1	Cumulative percentage of the number of returns found in Bin 1 of 10
D5	Cumulative percentage of the number of returns found in Bin 5 of 10
D9	Cumulative percentage of the number of returns found in Bin 9 of 10
DA	First returns/All Returns
DV	First Vegetation Returns/All Returns
MEDIAN	Median Height (m)
VDR	Vertical Distribution Ratio = VDR = [Max−Median]/Max
COVAR	Coefficient of variation (STD/Mean)
CanCOVAR	Coefficient of variation (STD/Mean) of first returns only
VCI	Vertical Complexity Index
FIRSTVEG	Number of First Vegetation Returns only
ALLGROUND	Number of Ground Returns
cc2	Crown closure ≥ 2 m
cc6	Crown closure ≥ 6 m
cc12	Crown closure ≥ 12 m
cc20	Crown closure ≥ 20 m

Using automated decision tree and forward and backward, step-wise, automated variable selection methods within the model generation framework of JMP©, several key variables began to emerge. The coefficient of variation (COVAR) (*i.e.*, standard deviation divided by mean) and VCI were highly influential components of automatically derived models. As well, several other variables relating to vegetation complexity were flagged for later inspection: *i.e.*, absolute standard deviation, DA (*i.e.*, first return divided by all returns) and DV (*i.e.*, first vegetation return divided by all returns). Interestingly, height and density metrics were not selected earlier than 5th or 6th in automatic variable selection, and did not provide much additional explanatory power to the model.

Using manual entry and expert knowledge, in conjunction with monitoring the sum of squares and F-ratios, these remaining key variables were passed together in various combinations. This analysis generated the final model from the randomly selected calibration dataset of 150 plots (Equation (2)):

$$LAI = 0.03\,DA - 1.84\,COVAR + 1.10\,VCI + 0.01\,cc6 \tag{2}$$

This model has an adjusted R^2 of 0.53 and RMSE of 0.57 (RMSE% = 24.7). All parameters are statistically significant at the <0.0001 level, with the exception of VCI at a slightly higher value (*i.e.*, p = 0.009). The Durban-Watson test for residual autocorrelation gave a value of 1.07 and autocorrelation probability of 0.44, upholding the null hypothesis of a lack of autocorrelation. A plot of model residuals lacked heteroscedasticity and residuals statistically passed a Shapiro-Wilk W test for normality (Prob < W = 0.56).

Using the remaining 75 plots as validation, the model was run with the same predictor variables. This dataset results in an adjusted R^2 of 0.58 ($p < 0.0001$) and RMSE of 0.55 (RMSE% = 25.1). This R^2 corresponds well to the explanatory power of both the calibration model ($R^2 = 0.53$). A matched pairs test of the predicted and *in situ* estimated values shows no difference, statistically, with a null hypothesis-refuting (H_0 = the data sets means are statistically different) *p*-value = 0.13.

3.3. Spectral Vegetation Index Model Generation

Using the atmospherically corrected WorldView-2 data, an NDVI surface with 2 m spatial resolution was created and mean values were extracted for each plot. These NDVI values were compared to plot LAI values in an attempt to determine whether there was a relationship between NDVI and LAI for the Hearst Forest using these high resolution data (*i.e.*, DHP and WorldView-2). Upon plotting NDVI *versus* LAI at the plot level, it was apparent both visually (Figure 11) and statistically (*i.e.*, $R^2 = 0.01$) that the expected relationship between NDVI, as derived from WorldView-2 data, and LAI was not present.

Figure 11. WorldView-2 Normalized Difference Vegetation Index (NDVI) *versus* LAI for $n = 122$ plots.
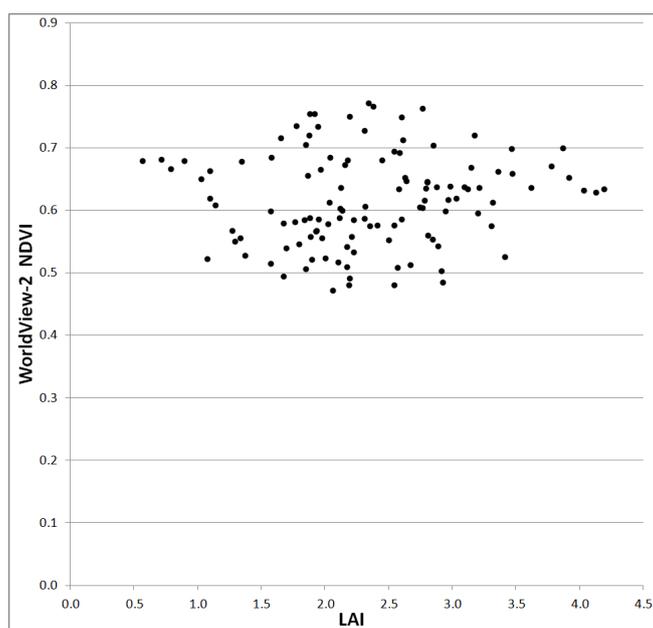
3.4. Combination Model

While optical instruments alone proved insufficient to model LAI, further testing was done to investigate whether WorldView-2 data could improve the LiDAR model. MLR models may be able to extract information from a combination of predictor variables that may not have been apparent in a simple regression model [35]. As the WorldView-2 scene only covers a subset of the plots (*i.e.*, 122), preliminary testing to evaluate model effectiveness was done using the full subset, with no calibration/validation subdivision. The first model to be tested was the initial LiDAR model shown in Equation (2) with the simple addition of the NDVI predictor variable. This new model has an adjusted R^2 of 0.52 and RMSE of 0.52 (RMSE% = 22.2); coefficients can be seen in Equation (3):

$$LAI = 0.03\,DA - 1.84\,COVAR + 1.10\,VCI + 0.01\,cc6 - 0.98\,NDVI \tag{3}$$

Although it uses a different, reduced data set, it is apparent that the addition of the new NDVI variable does not improve the predictive ability of the existing LiDAR model. In fact, the NDVI predictor coefficient has a statistically non-significant p-value of 0.64.

The second model used the forward stepwise technique with the NDVI predictor variable pre-inserted into the selection. An additional four variables were automatically added to generate a comparable model to Equations (2) and (3). As seen in Equation (4), three of the previously included variables were added again, with COVAR being replaced by D1, the proportion of the number of returns found in the lowest ten percent of the vertical canopy structure. The adjusted R^2 of this model is 0.55 with an RMSE of 0.50 (RMSE% = 21.6).

$$LAI = 0.04\,DA - 3.23\,D1 + 0.76\,VCI + 0.01\,cc6 - 1.23\,NDVI \tag{4}$$

Unlike the previous models, only three variables are statistically significant in this model, with NDVI and VCI having statistically non-significant *p*-values, *i.e.*, 0.35 and 0.17, respectively.

## 4. Discussion

4.1. LiDAR Model

The inclusion of the four final variables presented above suggests some interesting aspects about the model and LiDAR prediction of LAI in general. The COVAR exhibited a strong, negative correlation to LAI (R^2 = 0.42), *i.e.*, as LAI increased, COVAR decreased (Figure 12A). High COVAR corresponds to high standard deviation (*i.e.*, more open, penetrable, variable canopies), and/or low mean height (*i.e.*, shorter, immature trees). It was anticipated that a variable of this nature would be included in the final model since it is a coarse surrogate for gap fraction (*i.e.*, greater penetration = higher gap fraction = lower LAI). In laboratory controlled experiments with high density laser scanners and artificial trees it has been shown that increased leaf count corresponds to a decrease in pulse return density at greater distances into the canopy, thereby increasing standard deviation [36].

Other predictor variables did not exhibit strong correlations to LAI. However, the nature of MLR is that the combination and interplay of trends between predictor variables can often generate more explanatory power than an individual variable. Figure 12B illustrates the relatively weak positive correlation between LAI and VCI. VCI is a variable similar to COVAR, in that it is a measure of the structural complexity of the forest canopy, but inverse to COVAR, as high VCI corresponds to high LAI (*i.e.*, multi-layered forest canopy) [37].

DA (*i.e.*, first return divided by all returns) was another variable that linked closely to the concept of gap fraction and how much penetration a canopy allows [36]. As the number of first returns was equal to the total number of laser pulses transmitted (*i.e.*, a near-constant value for each equal area plot), the only changing variable in DA was the total number of returns. If the canopy is dense, the pulse is occluded and there is a higher chance of the first return being the only return; more open canopies allow for greater pulse penetration and a higher chance of two or three returns. The expected relationship was therefore that higher DA values are indicative of higher LAI. That trend was not observed clearly with this dataset, as there were many low LAI values which actually had the highest

DA value (Figure 12C). It was observed that in non-vegetated, completely open areas the same high DA values can be seen, given first returns (*i.e.*, only returns) were ground returns. Since this dataset contained only plots that were vegetated to some degree, a non-linear relationship would not be expected, but this trend may be partially responsible for odd artifacts in clear-cut areas in the complete forest LAI estimate surface.

Figure 12. Comparison of LAI and (**A**) COVAR, (**B**) VCI, (**C**) DA, and (**D**) crown closure ($\geq$ 6 m) for *n* = 225 plots.
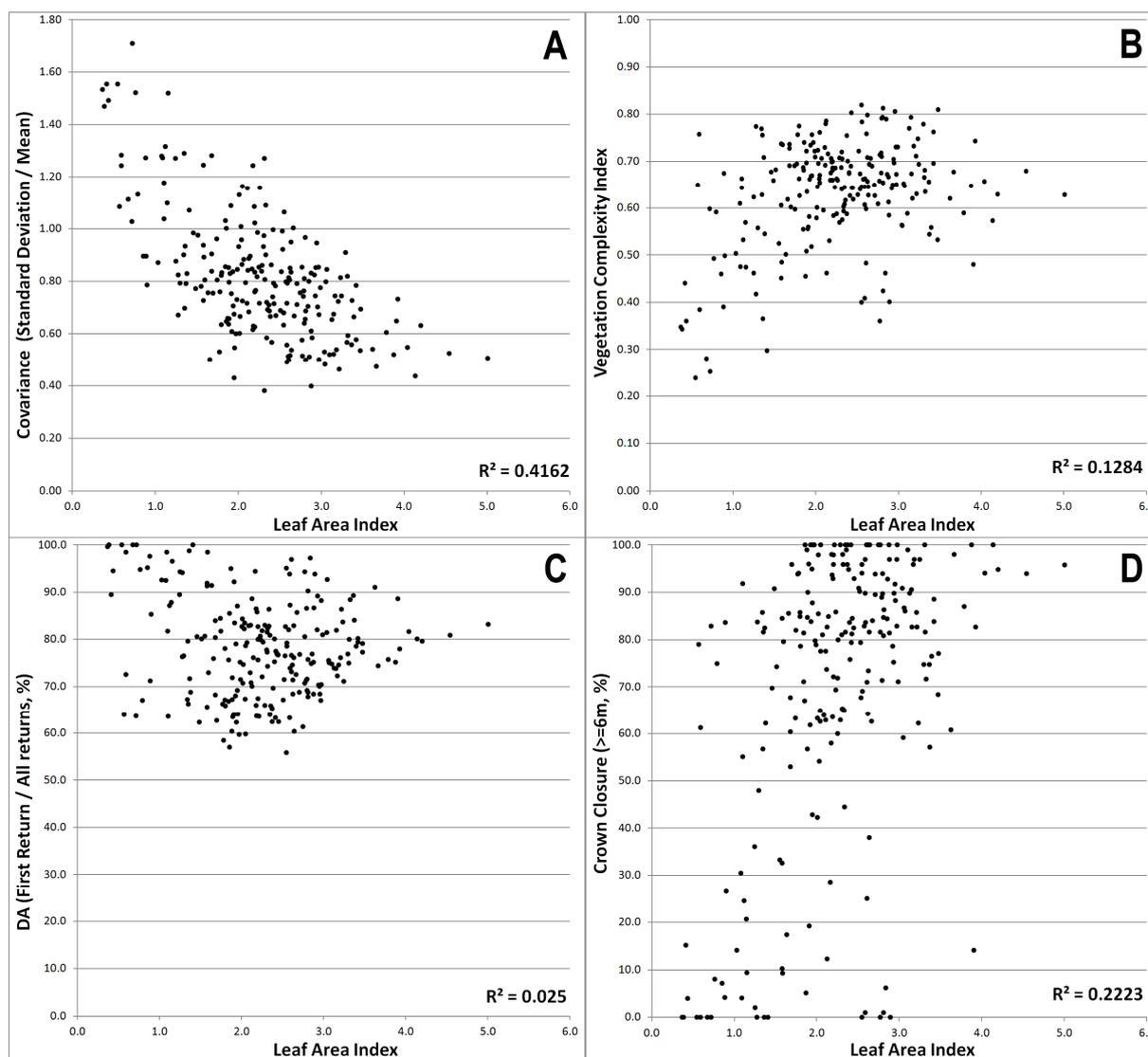

Crown closure ($\geq$6 m; cc6) was the predictor variable most unlike the others as it was not a measure of statistical spread or the direct complexity of the canopy. Crown closure at each height was calculated as a proportion of the number of 2 m sub-pixels within the 20 m pixel that matched the height criteria to the total number of pixels (*i.e.*, 100). As a measure of how closed the canopy was looking down from above to a certain height, cc6 was an interesting substitute for gap fraction. This trend shows that higher LAI corresponds to greater crown closure at six meters and above (Figure 12D). The basis of this relationship is that for the plots investigated, six meters is below a significant portion of the canopy, and is definitely above understory vegetation. Maximum tree height

in each plot is close to 15 m, so all vegetation greater than six meters is a good indication of total crown closure. This particular relationship compares especially well with the LAI values as the camera system was mounted 1.3 m from the ground and processing eliminated a conical field of view 18° above the horizon. Selecting six meters as the particular height to use from the suite of two meter intervals was done through the same step-wise methods as the primary model. Six meters was chosen as the variable from the low height ranges (*i.e.*, two–eight meters) that had the most explanatory power in the model.

The lack of height (P'XX') and density (D'X') metrics captured in the model was noteworthy as these tend to be predominant in modeling many other forest inventory variables (e.g., biomass, height, density) (e.g., [23,38,39]). This absence likely stems from the fact that although the P'XX' and D'X' variables do give a general measure of canopy penetration and complexity, there were other variables in the LiDAR suite that were better surrogates for canopy gap fraction. Rough trials using only basic height and density metrics as potential inputs to a stepwise model proved that satisfactory models could be created, albeit at lower R^2's (e.g., 0.40–0.45). The lower accuracy may be an acceptable trade-off for users hoping to generate more basic, easily interpretable models, where simplicity could outweigh a small level of error. Overall, this portion of the study provides interesting insight into the relationships between these height and density metrics and variables more related to overall canopy characteristics and crown closure. LAI seems to be more reliably estimated with the latter.

Using the regression model created with the calibration dataset to estimate values of LAI for the validation dataset resulted in the trend of predicted *versus in situ* estimated values presented in Figure 13. Scatter around the 1:1 line may be partially due to the time lag between the acquisition of the LiDAR data (*i.e.*, summer 2007) and *in situ* data collection (*i.e.*, summer 2011). This phenomenon may have most affected younger plots, which had potential to gain most biomass and leaf area.

Figure 13. Predicted *versus in situ* estimated LAI values for *n* = 75 validation plots.
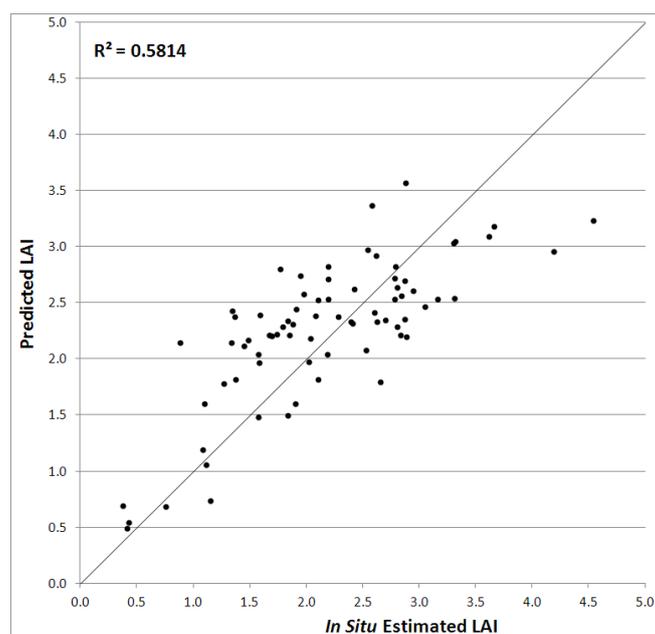
Existing work using LiDAR data alone to estimate LAI tends to show marginally better results than what was demonstrated here, but for different forest environments. A temperate coniferous forest study

by Jensen *et al*. [15] was able to obtain an adjusted R^2 of 0.65 for their true LAI model and a slightly higher value (R^2 = 0.68) when examining LAI values without clumping index processing. Their final model also used four predictors: *i.e.*, crown closure above breast height, COVAR above breast height and two simple percentile variables. As their study had an even larger suite of predictor variables to draw from than this study it is significant that they also found utility in a low height crown closure and COVAR variable. The stronger explanatory power of the Jensen *et al*. their model may be related to: (i) the different species in the temperate coniferous forest of Idaho, and/or (ii) the use of an LAI-2000 instead of DHP for *in situ* LAI data collection.

Another coniferous study in the eastern United States showed even better results [40]. This study compared *in situ* LAI derived from the LAI-2000 to LiDAR metrics derived from high density LiDAR (*i.e.*, 5 m^{-2}) for intensively managed loblolly pine (*Pinus taeda* L.) plantations. Results were presented for five models with an incrementally greater number of predictors; adjusted R^2 ranged from 0.61 with two predictors to 0.82 with six predictors. The comparable, four variable model to our study obtained an adjusted R^2 of 0.78 and used very different predictors, *i.e.*, mean (>1 m), P20 (>1 m), LPI (Laser penetration index − ground returns/(ground returns + all returns)) and mean intensity. This was one of the only studies we found that incorporated LiDAR intensity values alongside height and density metrics. High model explanatory power may have been linked to the single species (*i.e.*, plantation) nature of the study.

4.2. Spectral Vegetation Index Model

The results of this portion of the study were not entirely unexpected. A previous study by [41] examined the relationships between DHP and optically (*i.e.*, MODIS) derived LAI across several different forest types. They found poor correlations for sites that had open canopies, branches extending to the ground surface, and relatively low LAI values, particularly in black and white spruce stands. Studies with conflicting results include Stenberg *et al*. [42] and Chen and Cihlar [16], which use Landsat TM and ETM data, respectively. Stenberg *et al*. [42] found the NDVI/LAI correlation coefficient to be 0.55 in managed pine and spruce stands in Finland with LAI ranging from 0.36 to 3.72. One of the first to use NDVI to estimate LAI in boreal Canada, Chen and Cihlar [16] obtained R^2's of 0.50 and 0.42 for their two campaigns. LAI values ranged from 0.92 to 4.17.

A Landsat-5 scene, acquired 14 August, 2008, was obtained for the same region as the WorldView-2 coverage and an NDVI surface was generated. The positive relationship between the two NDVI datasets was R^2 = 0.61. A matched-pairs test of the WorldView-2 and Landsat values shows no statistical difference, with a null hypothesis-refuting p-value of 0.37. This relationship shows that the WorldView-2 NDVI values are comparable to the Landsat NDVI values.

It was originally surmised that the poor correlation was primarily due to a combination of the high spatial resolution of the sensor and the low spatial density of individual trees in the Hearst Forest (*i.e.*, open canopy), resulting in substantial shadowing and significant reflectance from the ground surface. Landsat-5 has significant, inherent averaging of reflectance, while Worldview-2 shows distinct tree crowns and gaps with associated shadows. While this phenomenon impacts measurements at the pixel level with WorldView-2 data, it is assumed that this is not the problem in this study. As was discussed earlier, for each plot the mean NDVI value was calculated, so the values being used for the

WorldView-2 calculation include both crown and gap values, technically replicating the Landsat data, which integrates reflectance of each of these surfaces within the 30 m pixel. Even studies using lower spatial resolution sensors like AVHRR (*i.e.*, 1 km resolution) or MODIS (*i.e.*, 500 m resolution) obtained correlation coefficients between 0.39 and 0.46 [17].

The most reasonable explanation for the poor performance of optical data in estimating LAI is due to the open canopies and the introduction of alternate understory spectra into the NDVI calculations. One of the failings of the DHP product was that it only provided an estimate of LAI from the focal point of the lens upwards, *i.e.*, 1.3 m. The cover below 1.3 m for the study sites ranged dramatically and included bare ground, thick moss and/or leafy shrubs, thereby impacting the NDVI values derived for these open canopy mixedwoods. As seen in several successful LiDAR studies, variables tend to be used that exclude this understory vegetation by implementing a height threshold of approximately one meter (e.g., [14,15,40]).

Deriving a statistical relationship was also difficult, given the narrow range of the NDVI and LAI values being compared. NDVI values ranged from 0.47 to 0.77 (range of 0.30) and LAI from 0.57 to 4.20 (range of 3.63). This range affected the ability of regression models to accurately depict trends in these data. Previously discussed successful models (e.g., [43]) have ranges of LAI values as large as 0.2 to 8.5 and NDVI ranges from 0.6 to 0.93, and other studies have found that the wider ranges of other, non-normalized SVIs are better suited to LAI analysis [42,44]. As the plot sampling was specifically designed to sample both dense and relatively open areas, as judged by basal area measured the previous summer, this limited range is a product of the forest structure itself. The extensive single species dominance in some areas and physical similarity between several of the dominant tree species (e.g., trembling aspen/balsam poplar, white spruce/black spruce) generate similar canopy conditions in the boreal environment, exhibiting a low and narrow range of LAI.

4.3. Combination Model

It would seem that the addition of other LiDAR variables in conjunction with NDVI does not improve the models sufficiently to warrant further testing. This result was not unexpected considering the extremely poor correlation of NDVI values to LAI. While the model in Equation (3) has the same explanatory power as the LiDAR-only model and a slightly lower RMSE%, the addition of a completely independent optical dataset is not financially feasible for operational implementation. One study that attempted a similar combinatorial approach using LiDAR variables integrated with SPOT-5 SVIs found an improvement from R^2 of 0.75 to 0.79 [15]. This improvement to both explanatory power of the model and residual error (*i.e.*, 0.75 to 0.69) is small considering that the LiDAR-SPOT model used seven parameters, including two SVI (*i.e.*, reduced simple ratio and standard deviation of the red band), and the additional expense of acquiring a second remotely sensed dataset.

The results discussed here present a strong case for LiDAR modeling of LAI as opposed to more traditional optical approaches, particularly for the boreal mixedwood forests of central Ontario. Given the open canopies typical of the Hearst Forest, high resolution optical data tend to integrate surface spectra from all components of the plot (*i.e.*, canopy, understory and ground) and for this region exhibit a narrow range of NDVI values. Conversely, LiDAR allows for a distinction between the forest canopy and the underlying ground cover, allowing it to better estimate LAI for the forest canopy alone.

5. Conclusions

This study was undertaken to examine several potential methods to remotely estimate Leaf Area Index (LAI) in a boreal mixedwood forest of northern Ontario. This research is unique in examining LAI in this particular setting in Ontario, using low density Light Detection and Ranging (LiDAR). Accurate and precise models of LAI allow for the monitoring of a wide array of LAI-dependant variables, e.g., biomass, productivity, general forest health. The ecological and commercial benefits to provincial and federal government agencies, as well as commercial forest managers, are far reaching. These benefits include more accurate predictions of harvest yields, better timing of forest management practices, periodic monitoring of invasive species progression, tracking carbon sequestration by vegetation, and more. The general result from this study shows that LiDAR data provide adequate LAI estimations to predict the variable over large spatial extents at moderate resolution. The final model included DA, COVAR, VCI and crown closure (≥6 m). The variables selected tended to be representative of whole canopy distribution, rather than individual statistical metrics (e.g., percentiles). The overall model produced an adjusted R^2 of 0.53 with a 24.7% RMSE (validation dataset $R^2 = 0.58$, RMSE% = 25.1). The validation dataset produced statistically consistent results.

There was no relationship observed when comparing Normalized Difference Vegetation Index (NDVI) derived from WorldView-2 data and LAI derived from digital hemispherical photographs (DHPs). Averaging at the plot level inherently smoothed some of the variability expected from the higher resolution WorldView-2 data. Testing concluded that this lack of relationship was not erroneous, *i.e.*, separate sensors (*i.e.*, Landsat) verified that the NDVI calculations were accurate. The most probable causes of the failure were twofold. First, the low density of the boreal forest (*i.e.*, open canopy) resulted in a significant inclusion of understory spectra contributing to the reflectance, thereby skewing NDVI results. Second, the vegetation of the boreal forest itself has inherently low ranges of LAI and NDVI, making it difficult for regression models to generate strong trends.

The combined model utilizing LiDAR and optical data also proved unsuccessful. Neither the original LiDAR model with NDVI inserted nor a new model built around NDVI improved the explanatory power of the original LiDAR-only model. This lack of success demonstrated that there were no within-data trends that could be exploited by the multiple linear regression (MLR) framework.

It was determined that there were no statistical differences in *in situ* LAI estimates when five or nine DHPs were processed per plot. Future studies could reduce the number of DHPs taken within each plot, dramatically reducing the amount of time and effort spent sampling at each plot. From these photographs, LAI for the final plots had a small dynamic range (0.37–5.01); values typical for the boreal mixedwood forest of northern Ontario.

In order to obtain a better understanding of patterns within the data it may be beneficial to reduce the full dataset down into smaller subsets, *i.e.*, by tree species or even species by age class. In this forest environment, these divisions may consist of black spruce (white spruce and tamarack included), jack pine, and mixedwoods (white birch, trembling aspen and balsam poplar), which, with the current data, would yield sub-groups of 132, 40 and 53 samples, respectively. Separate models may reveal trends not observed in the full dataset.

Since NDVI derived from WorldView-2 data provided little insight into LAI, it may be suitable to investigate spectral vegetation indices (SVIs) that incorporate a soil or understory vegetation baseline

coefficient derived from spectral endmembers. To test other SVIs that incorporate spectral unmixing to distinguish canopy from understory reflectance, would require spectral measurements from beneath the canopy [45,46]. These measurements could be most easily done with a handheld spectroradiometer.

Acknowledgments

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Brandt, J.P. The extent of the North American boreal zone. *Environ. Rev.* **2009**, *17*, 101–161.
2. Forests Branch, Ontario Ministry of Natural Resources (OMNR). *State of the Forest Report 2006*; OMNR: Ottawa, ON, Canada, 2006.
3. Government of Ontario. *Crown Forest Sustainability Act 1994*; Volume Part 1; Law last amended, 1 June 2011.
4. Polis, G.A. Why are parts of the world green? Multiple factors control productivity and the distribution of biomass. *Oikos* **1999**, *86*, 3–15.
5. Deblonde, G.; Penner, M.; Royer, A. Measuring leaf area index with the Li-Cor LAI-2000 in pine stands. *Ecology* **1994**, *75*, 1507–1511.
6. Leblanc, S.G.; Chen, J.M.; Fernandes, R.; Deering, D.W.; Conley, A. Methodology comparison for canopy structure parameters extraction from digital hemispherical photography in boreal forests. *Agric. For. Meteorol.* **2005**, *129*, 187–207.
7. Abuelgasim, A.A.; Fernandes, R.A.; Leblanc, S.G. Evaluation of national and global LAI products derived from optical remote sensing instruments over Canada. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1872–1884.
8. Leverenz, J.W.; Hinckley, T.M. Shoot structure, leaf area index and productivity of evergreen conifer stands. *Tree Physiol.* **1990**, *6*, 135–149.
9. McWilliam, A.L.; Roberts, J.M.; Cabral, O.M.; Leitao, M.V.; Costa, A.C.; Maitelli, G.T.; Zamparoni, C.A. Leaf area index and above-ground biomass of terra firme rain forest and adjacent clearings in Amazonia. *Funct. Ecol.* **1993**, *7*, 310–317.
10. Friedl, M.A.; Schimel, D.S.; Michaelsen, J.; Davis, F.W.; Walker, H. Estimating grassland biomass and leaf area index using ground and satellite data. *Int. J. Remote Sens.* **1994**, *15*, 1401–1420.

11. Hashimoto, H.; Wang, W.; Milesi, C.; White, M.A.; Ganguly, S.; Gamo, M.; Hirata, R.; Myneni, R.B.; Nemani, R.R. Exploring simple algorithms for estimating gross primary production in forested areas from satellite data. *Remote Sens.* **2012**, *4*, 303–326.

12. Verwijst, T.; Telenius, B. Biomass estimation procedures in short rotation forestry. *For. Ecol. Manag.* **1999**, *121*, 137–146.

13. Riaño, D.; Valladares, F.; Condés, S.; Chuvieco, E. Estimation of leaf area index and covered ground from airborne laser scanner (Lidar) in two contrasting forests. *Agric. For. Meteorol.* **2004**, *124*, 269–275.

14. Morsdorf, F.; Kötz, B.; Meier, E.; Itten, K.I.; Allgöwer, B. Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. *Remote Sens. Environ.* **2006**, *104*, 50–61.

15. Jensen, J.L.R.; Humes, K.S.; Vierling, L.A.; Hudak, A.T. Discrete return lidar-based prediction of leaf area index in two conifer forests. *Remote Sens. Environ.* **2008**, *112*, 3947–3957.

16. Chen, J.M.; Cihlar, J. Retrieving leaf area index of boreal conifer forests using Landsat TM images. *Remote Sens. Environ.* **1996**, *55*, 153–162.

17. Wang, Q.; Adiku, S.; Tenhunen, J.; Granier, A. On the relationship of NDVI with leaf area index in a deciduous forest site. *Remote Sens. Environ.* **2005**, *94*, 244–255.

18. Yang, W.; Dong, H.; Bin, T.; Stroeve, J.C.; Shabanov, N.V.; Knyazikhin, Y.; Nemani, R.R.; Myneni, R.B. Analysis of leaf area index and fraction of PAR absorbed by vegetation products from the terra MODIS sensor: 2000–2005. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1829–1842.

19. Propastin, P.; Kappas, M. Retrieval of coarse-resolution leaf area index over the republic of kazakhstan using NOAA AVHRR satellite data and ground measurements. *Remote Sens.* **2012**, *4*, 220–246.

20. Asner, G.P.; Powell, G.V.N.; Mascaro, J.; Knapp, D.E.; Clark, J.K.; Jacobson, J.; Kennedy-Bowdoin, T.; Balaji, A.; Paez-Acosta, G.; Victoria, E.; *et al*. High-resolution forest carbon stocks and emissions in the Amazon. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16738–16742.

21. Hearst Forest Management Inc. Current Forest Condition. Available online: http://www.hearstforest.com/english/current.html (accessed on 16 June 2011).

22. Hearst Forest Management Inc. *Forest Management Plan for the Hearst Forest*; Hearst Forest Management Inc.: Hearst, ON, Canada, 2007.

23. Woods, M.; Pitt, D.; Penner, M.; Lim, K.; Nesbitt, D.; Etheridge, D.; Treitz, P. Operational implementation of a LiDAR inventory in Boreal Ontario. *For. Chron.* **2011**, *87*, 512–528.

24. North West Geomatics. *The Hearst Forest-North Western Ontario*; North West Geomatics: Calgary, AB, Canada, 2008.

25. Environment Canada. Canadian Climate Normals 1971–2000: Geraldton. Available online: http://www.climate.weatheroffice.gc.ca/climate_normals/index_e.html (accessed on 16 June 2011).

26. Leblanc, S. *DHP-TRACWin Manual*; Natural Resources Canada-Canada Centre for Remote Sensing: Québec, QC, Canada, 2008.

27. Chen, J.M.; Rich, P.M.; Gower, S.T.; Norman, J.M.; Plummer, S. Leaf area index of boreal forests: Theory, techniques, and measurements. *J. Geophys. Res.* **1997**, *102*, 29429–29443.

28. Thomas, V.; Noland, T.; Treitz, P.; McCaughey, J.H. Leaf area and clumping indices for a boreal mixed-wood forest: Lidar, hyperspectral, and Landsat models. *Int. J. Remote Sens.* **2011**, *32*, 8271–8297.

29. Gower, S.T.; Kucharik, C.J.; Norman, J.M. Direct and indirect estimation of leaf area index, fAPAR, and net primary production of terrestrial ecosystems. *Remote Sens. Environ.* **1999**, *70*, 29–51.

30. Chen, J.; Leblanc, S.; Thomas, V. Personal Commumication. January 2012.

31. Van Ewijk, K. Personal Communication. April 2011.

32. Barr, A.G.; Black, T.A.; Hogg, E.H.; Kljun, N.; Morgenstern, K.; Nesic, Z. Inter-annual variability in the leaf area index of a boreal aspen-hazelnut forest in relation to net ecosystem production. *Agric. For. Meteorol.* **2004**, *126*, 237–255.

33. Rencher, A.C.; Pun, F.C. Inflation of R^2 in best subset regression. *Technometrics* **1980**, *22*, 49–53.

34. STATISTICA. StatSoft, Inc.: Tulsa, OK, USA, 2012.

35. De Veaux, R.; Velleman, P.; Bock, D. *Stats-Data and Models*; Pearson Addison Wesley: Toronto, ON, Canada, 2005.

36. Moorthy, I.; Miller, J.; Hu, B.; Chen, J.; Li, Q. Retrieving crown leaf area index from an individual tree using ground-based lidar data. *Can. J. Remote Sens.* **2008**, *34*, 320–332.

37. Van Ewijk, K.Y.; Treitz, P.M.; Scott, N.A. Characterizing forest succession in central Ontario using lidar-derived indices. *Photogramm. Eng. Remote Sens.* **2011**, *77*, 261–269.

38. Woods, M.; Lim, K.; Treitz, P. Predicting forest stand variables from LiDAR data in the Great Lakes-St. Lawrence forest of Ontario. *For. Chron.* **2008**, *84*, 827–839.

39. Treitz, P.; Lim, K.; Woods, M.; Pitt, D.; Nesbitt, D.; Etheridge, D. LiDAR sampling density for forest resource inventories in Ontario, Canada. *Remote Sens.* **2012**, *4*, 830–848.

40. Peduzzi, A.; Wynne, R.H.; Fox, T.R.; Nelson, R.F.; Thomas, V.A. Estimating leaf area index in intensively managed pine plantations using airborne laser scanner data. *For. Ecol. Manag.* **2012**, *270*, 54–65.

41. Chasmer, L.; Hopkinson, C.; Treitz, P.; McCaughey, H.; Barr, A.; Black, A. A lidar-based hierarchical approach for assessing MODIS fPAR. *Remote Sens. Environ.* **2008**, *112*, 4344–4357.

42. Stenberg, P.; Rautiainen, M.; Manninen, T. Reduced simple ratio better than NDVI for estimating LAI in Finnish pine and spruce stands. *Silva Fenn.* **2004**, *38*, 3–14.

43. Davi, H.; Soudani, K.; Deckx, T.; Dufrene, E.; le Dantec, V.; Francois, C. Estimation of forest leaf area index from SPOT imagery using NDVI distribution over forest stands. *Int. J. Remote Sens.* **2006**, *27*, 885–902.

44. Eklundh, L.; Hall, K.; Eriksson, H.; Ardö, J.; Pilesjö, P. Investigating the use of Landsat thematic mapper data for estimation of forest leaf area index in southern Sweden. *Can. J. Remote Sens.* **2003**, *29*, 349–362.

45. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309.

46. Keshava, N.; Mustard, J.F. Spectral unmixing. *IEEE Signal Process. Mag.* **2002**, *19*, 44–57.

Appendix

Table A1. LiDAR predictor variable descriptions.

LiDAR Predictor	Description
LG%Hwd	Percent Hardwood by Basal Area for trees ≥ 10 cm
LG%Con	Percent Conifer by Basal Area for trees ≥ 10 cm
Sm%Hwd	Percent Hardwood by Basal Area for trees < 10 cm
Sm%Con	Percent Conifer by Basal Area for trees < 10 cm
MEAN	Mean height (m)
STD_DEV	Standard Deviation
ABS_DEV	Absolute Standard Deviation
SKEW	Skewness
KURTOSIS	Kurtosis
MIN	Minimum height (m)
P10	First Decile (10th Percentile) LiDAR Height (m)
P20	Second Decile (20th Percentile) LiDAR Height (m)
...	...
P90	Ninth Decile (90th Percentile) LiDAR Height (m)
MAX	Maximum height (m)
D1	Cumulative percentage of the number of returns found in Bin 1 of 10
D2	Cumulative percentage of the number of returns found in Bin 2 of 10
...	...
D9	Cumulative percentage of the number of returns found in Bin 9 of 10
DA	First returns/All Returns
DB	First and only return/All Returns
DV	First Vegetation Returns/All Returns
MEDIAN	Median Height (m)
VDR	Vertical Distribution Ratio = VDR=[Max − Median]/Max
COVAR	Coefficient of variation (STD/Mean)
CanCOVAR	Coefficient of variation (STD/Mean) of first returns only
H	Shannon-Weaver Index
VCI	Vertical Complexity Index
FIRST	Number of First Returns
ALLRETURNS	Number of all Returns
FIRSTVEG	Number of First Vegetation Returns only
ALLGROUND	Number of Ground Returns
cc0	Crown closure > 0 m
cc2	Crown closure ≥ 2 m
cc4	Crown closure ≥ 4 m
…	…
cc28	Crown closure ≥ 28 m