September, 2009

Estimation of forest structural parameters using 5 and 10 meter SPOT-5 satellite data

Peter T. Wolter, *University of Wisconsin-Madison*
Philip A. Townsend, *University of Wisconsin-Madison*
Brian R. Sturtevant, *United States Department of Agriculture Forest Service*

Remote Sensing of Environment 113 (2009) 2019–2036

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Estimation of forest structural parameters using 5 and 10 meter SPOT-5 satellite data

Peter T. Wolter [a,*], Philip A. Townsend [a], Brian R. Sturtevant [b]

[a] Department of Forest and Wildlife Ecology, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA
[b] Forestry Sciences Laboratory, USDA Forest Service, 5985 Highway K, Rhinelander, Wisconsin 54501, USA

ARTICLE INFO

Article history:
Received 26 January 2009
Received in revised form 19 May 2009
Accepted 25 May 2009

Keywords:
Forest structure
Multi-resolution
SPOT-5
Geostatistics
Minnesota
Sub-boreal
Spruce budworm

ABSTRACT

Large areas of forest in the US and Canada are affected by insects and disease each year. Over the past century, outbreaks of the Eastern spruce budworm have become more frequent and severe. The notion of designing a more pest resistant landscape through prescriptive management practices hinges on our ability to effectively model forest–insect dynamics at regional scales. Increasingly, more detailed pixel-wise estimates of forest biophysical parameters are needed for such endeavors. Lidar technology, although promising, is not yet viable for repeated regional accounting, necessitating the development of methods which take advantage of existing spaceborne assets. Our objective is to use one of these assets (SPOT-5) to estimate a large set of forest structural attributes at a finer spatial grain size (5 m and 10 m) over a broader area than is currently available for the purpose of supplying needed input data for disturbance simulation modeling. We employ neighborhood statistics (standard deviation, variance, sill variance, and ratios of these metrics at 5 and 10 m) calculated from SPOT-5 sensor data and derivatives to estimate and map tree canopy diameter (CDIA), bole diameter at breast height (DBH), tree height (HT), crown closure (CC), vertical length of live crown (LC), and basal area (BA). A partial least squares (PLS) regression approach was used with these local statistics and field data to produce models for pixel-wise estimation and mapping of mean values, respectively, for hardwood and coniferous forest CDIA ($R^2 = 0.82$ and 0.93, RMSE 0.62 and 0.47 m), DBH ($R^2 = 0.82$ and 0.90, RMSE 2.92 and 3.75 cm), HT ($R^2 = 0.69$ and 0.92, RMSE 1.27 and 1.59 m), CC ($R^2 = 0.52$ and 0.68, RMSE 5.49 and 6.02%), LC ($R^2 = 0.58$ and 0.81, RMSE 0.96 and 1.25 m), and BA ($R^2 = 0.71$ and 0.74, RMSE 2.47 and 4.58 m^2 ha^{-1}) for a 3600 km^2 area in northeast Minnesota. This approach for quantifying forest structure is robust in the sense that a detailed forest cover type map is not required to stratify analysis at any step in the process. Hence, we show that multi-resolution SPOT-5 data are a practical alternative to lidar for regional characterization of forest biophysical parameters. However, lidar data may potentially be used to calibrate these SPOT-based structure models in the future.

1. Introduction

Large areas of forest in the US and Canada are affected by insects and disease each year. Over the past century, outbreaks of the Eastern spruce budworm (*Choristoneura fumiferana*) have become more frequent and severe as a result of past forest management practices, fire suppression, and pesticide application that favored expansion of host species (Blais, 1983). Because observed changes in insect disturbance history are largely human induced, it may also be possible to undo or at least mitigate these effects through prescriptive forest management (Blais, 1983). Forest ecologists have identified several forest stand characteristics such as tree species composition and basal area (Ghent, 1958; Batzer 1969; Crook et al., 1979; Bergeron et al., 1995; Alfaro et al., 2001; Sturtevant et al., 2004; Hennigar et al., 2008), host needle biomass and terrain position (Magnussen et al., 2004), forest age and crown closure (Alfaro et al., 2001), canopy position (Zhang & Alfaro, 2001), bole diameter (Bergeron et al., 1995), and other structural parameters that are indicative of the relative vulnerability of a stand to a spruce budworm outbreak should one occur. However, the notion of using this information to design a more pest resistant landscape is highly complex and hinges on our ability to effectively model multiple biological disturbance interactions at regional scales (Blais, 1983; Sturtevant et al., 2004). Ideally, spatially explicit landscape succession and disturbance models tailored for these efforts, such as LANDIS and LANDIS II (Mladenoff & He, 1999; Scheller & Mladenoff, 2004; Schumacher et al., 2004; Sturtevant et al., 2004; Scheller et al., 2007), make use of pixel-level information to parameterize the land surface to the extent that these data are available. While rudimentary pixel-level information describing the abundance and distribution of spruce budworm host species on a regional scale is available for some areas (e.g. Wolter et al., 2008), the need for more detailed forest structure information for these purposes and many others is increasingly coveted.

All ecosystem process models require parameterization of the land surface in one form or another. At medium to large spatial scales the

* Corresponding author. Tel.: +1 608 263 9107.
E-mail address: ptwolter@wisc.edu (P.T. Wolter).

0034-4257/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2009.05.009

most realistic possibility for accurate estimation and periodic update of these parameters is satellite remote sensing (Hall et al., 1995; Widlowski et al., 2004). The need to easily extract forest biophysical parameters over large areas at a relatively fine grain size is significant, as it provides a means for the inclusion of previously missing forest parameter data into regional ecosystem models to directly estimate linkages between forest structure and ecosystem functioning (Song, 2007). As such, one of the most persistent objectives of satellite remote sensing has been classification and quantification of forest biophysical properties such as tree species composition (Wolter et al., 1995; Reese et al., 2002), canopy diameter (Li & Strahler, 1985; Woodcock et al., 1997; Cohen & Spies, 1990; Song & Woodcock, 2003; Song, 2007), stem density (Cohen & Spies, 1992; Hudak et al., 2006; McRoberts, 2008), basal area (Franklin, 1986; Franco-Lopez et al., 2001; Hudak et al., 2006; McRoberts et al., 2007; McRoberts, 2008; Wolter et al., 2008), above ground biomass or volume (Franco-Lopez et al., 2001; Santoro et al., 2002; Pulliainen et al., 2003; Zheng et al., 2004; Muukkonen & Heiskanen, 2005; Rauste, 2005; Hall et al., 2006; McRoberts et al., 2007), bole diameter (Greenberg et al., 2005), tree height (Maltamo et al., 2006; Walker et al., 2007), live crown height (Maltamo et al., 2006), crown closure (Li & Strahler, 1985; Cohen et al., 1990; Cohen et al., 1995; Woodcock et al., 1997), stand age (Cohen & Spies, 1992; Cohen et al., 1995; Franklin et al., 2001), disturbance (Vogelmann & Rock, 1989; Healey et al., 2005), health (Vogelmann & Rock, 1988), and other characteristic forest attributes that are commonly sought after to understand forest functional complexity (Mc Elhinny et al., 2005). While many satellite-based efforts have consistently achieved moderate to high levels of success measuring subsets of these parameters, more comprehensive parameter sets describing forest structural complexity beyond small study sites has not yet been achieved (Anderson et al., 2008). Although lidar technology, used by itself or in combination with other sensor data, is considered optimal for estimating many of these forest parameters (Hyyppä & Inkinen, 1999; Anderson et al., 2008; Hudak et al., 2008), automation and extrapolation to larger, regional scales remains a challenge.

Forests of the northern Great Lakes States (Minnesota, Wisconsin, and Michigan) consist largely of second and third growth stands with less than 9% of old growth (>120 years) remaining (Frelich & Reich, 1995). The diversity and smaller stature of these forests effectively precludes application of most Landsat-based techniques for estimating structure that have shown promise for western coniferous forests (Woodcock & Strahler, 1987; Cohen & Spies, 1992; Cohen et al., 1995; Hansen et al., 2001). Alternatively, nearest neighbor techniques, such as the popular *k*-Nearest Neighbor (*k*-NN) method described by McRoberts et al. (2007), have shown promise when used with Landsat data for estimating stand-level forest structure information in the Great Lakes region (McRoberts et al., 2007; McRoberts, 2008, 2009) and northern Europe (Katila & Tomppo, 2001; Tomppo et al., 2009). With *k*-NN, forest parameter predictions, for pixels without ground reference data, are calculated as linear combinations of reference pixel values that are nearest in feature space according to some distance metric (Tomppo et al., 2009). However, arbitrary selection of *k* neighbors, distance metrics, distance cutoff criteria, and neighbor weights are cited as potential limitations of the technique, as well as computation intensity when applied over large areas (Finley et al., 2006; McRoberts et al., 2007; Meng et al., 2009). While data reduction techniques (e.g., principal components analysis) applied to sensor data prior to analysis is a common prescription for increasing the efficiency of the *k*-NN algorithm (Meng et al., 2009; McRoberts et al., 2007), such data reduction may be undesirable if goals include identifying specific spectral regions or indices that best explain variance among dependent forest variables (see Wolter et al., 2008).

Estimates of forest structure made using high spatial resolution (0.6 m–4.0 m) satellite data (Shugart et al., 2000; Song & Woodcock, 2003; Song, 2007; Lamonaca et al., 2008; Wulder et al., 2008), airborne or spaceborne lidar (Lefsky et al., 1999, 2005), or combinations of optical satellite data with airborne lidar (Donoghue & Watt, 2006; Wulder et al., 2007) are increasingly precise, but are limited for regional application due to high cost to coverage area ratios (Zheng et al., 2008) compared to more synoptic satellite sensors such as SPOT (60×60 km), Landsat (185×185 km), or MODIS (2330 km swath). Moreover, airborne lidar coverage represents only a fraction of the regional need for such data, and while it is ideal for measuring tree height, and subsequently, estimating forest biomass, it generally cannot provide direct information on canopy diameter (Song, 2007).

In this study we take advantage of the geospatial relationship between tree canopy size (i.e. diameter) and resulting representations of these canopies when imaged at two different pixel resolutions (Woodcock et al., 1997; Song & Woodcock, 2003; Song, 2007) to estimate mean canopy diameter (CDIA), tree height (HT), bole diameter at breast height (DBH), canopy closure (CC), basal area (BA), and height of live crown (LC) using 5 m and 10 m SPOT-5 (Systeme pour l'Observation de la Terre) satellite sensor data collected over northeast Minnesota. SPOT-5 sensor data is convenient as it represents a reasonable compromise between high and medium spatial resolution, while also having a large coverage area compared to IKONOS or Quickbird satellite data.

1.1. Study objective

The primary goal of this paper is to demonstrate a unique approach for modeling and mapping a set of forest structure parameters (Appendix A) using optical sensor data with a relatively fine spatial resolution (5 m and 10 m), but with large enough coverage area (60 km×60 km) to be practical for repeated, regional studies. We employ a broad suite of predictor variables (Table 1) derived from the SPOT-5 sensor data including panchromatic (PAN, 5 m) and multi-spectral (XS, 10 m) reflectance bands, XS indices, semivariogram sill parameters and sill ratios (Song & Woodcock, 2003; Song, 2007), and

Table 1
Local statistics calculated within Euclidean neighborhoods for SPOT-5 bands and derivatives.

Variables	Descriptions
GRN	Mean of 10 m visible green (July)
RED	Mean of 10 m visible red (July)
NIR	Mean of 10 m near-infrared (July)
SWIR	Mean of 20 m shortwave infrared (July)
P5	Mean of 5 m PANchromatic band (August)
P10P	Mean of simulated 10 m PAN: P5 regularized to10 m
P10X	Mean of simulated 10 m PAN: (GRN + RED)/2
SNIR	Standard deviation of NIR
S5	Standard deviation of P5
S10P	Standard deviation of P10P
S10X	Standard deviation of P10X
VNIR	Variance of NIR
V5	Variance of P5
V10P	Variance of P10P
V10X	Variance of P10X
C5	Semivariogram sill parameter for P5
C10P	Semivariogram sill parameter for P10P
C10X	Semivariogram sill parameter for P10X
NDVI	Normalized difference vegetation index: (NIR-RED) / (NIR+RED)
MSI	Moisture stress index: SWIR/NIR
SVR	Shortwave infrared to visible ratio: SWIR/[(RED+GRN)/2]
S510P	Ratio of standard deviations: S5/S10P
S510X	Ratio of standard deviations: S5/S10X
V510P	Ratio of variances: V5/V10P
V510X	Ratio of variances: V5/V10X
C510P	Ratio of semivariogram sill parameters: C5/C10P
C510X	Ratio of semivariogram sill parameters: C5/C10X

There are four multi-spectral bands (15 July 2006), one 5 m panchromatic (P5) band (31 August 2006), and two simulated 10 m panchromatic bands: one produced from the August panchromatic image (P10P) and the other from the multi-spectral July image (P10X). The remaining image variables were derived using these seven bands that are highlighted in bold. Prefixes *S*, *V*, and *C* (except SVR and SWIR) are used specifically to denote standard deviation, variance, and sill, respectively.

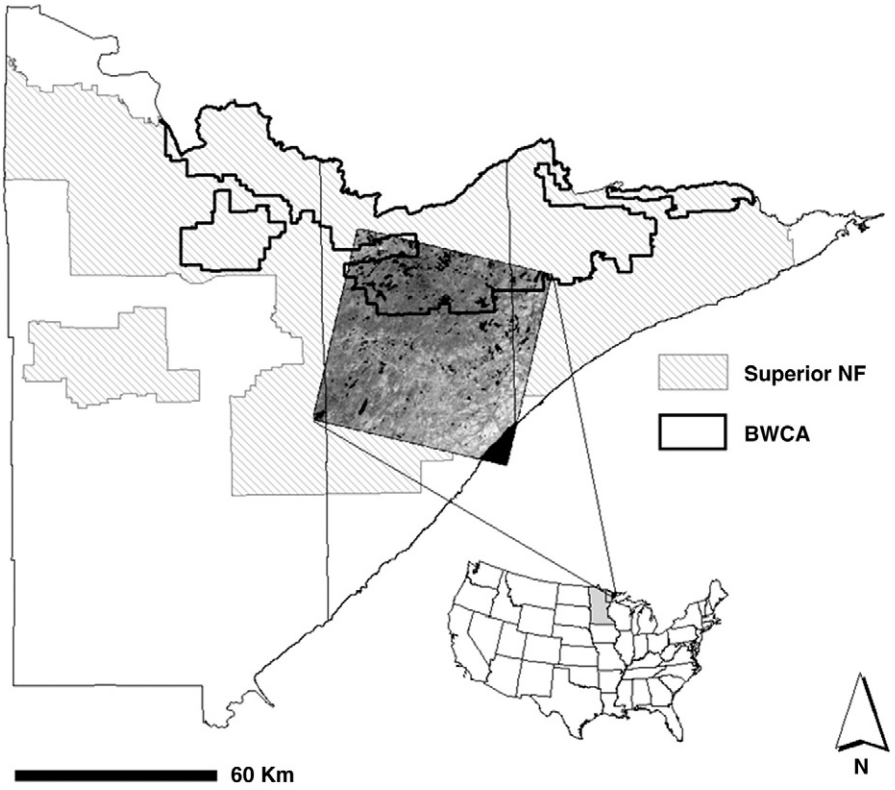

Fig. 1. The study area in northeast Minnesota which consists of a SPOT-5 image footprint (K587, J253) within the Superior National Forest and part of the Boundary Waters Canoe Area (BWCA) wilderness.

first-order texture measures to produce models for pixel-wise estimation and mapping of mean values for tree canopy diameter (CDIA), tree height (HT), vertical length of live crown (LC), canopy closure (CC), bole diameter at breast height (DBH), and bole basal area (BA) for a 3600 km^2 area in northeast Minnesota (Fig. 1).

Because we also wish to gauge the importance of the sill ratio variables when used in combination with other spatial and spectral information for modeling forest structure, we employ partial least squares (PLS) regression. While PLS has been used extensively with hyperspectral data (Ourcival et al., 1999; Smith et al., 2002, 2003; Townsend et al., 2003; Coops et al., 2003; McDonald et al., 2003), Wolter et al. (2008) demonstrated the capability of this approach to estimate forest BA and species composition using broad band, satellite sensor data. PLS regression is convenient in this regard as it allows simultaneous modeling of multiple continuous predictor variables, does not make unrealistic assumptions about spectral or ground measurement error, as in ordinary least-squares regression (Curran & Hay, 1986; Cohen et al., 2003), and addresses the problem of collinearity (dependence) among multiple independent and dependent variables (Helland, 1988).

1.2. Background

1.2.1. Forest structure, sill semivariance, and multi-resolution imagery

Semivariance and other texture measures applied to remote sensing data have been used extensively to identify unique forest structure (Woodcock & Strahler, 1987; Woodcock et al. (1988), Cohen et al., 1990; Franklin et al., 2001; Song & Woodcock, 2002; Coburn & Roberts, 2004). In particular, Woodcock et al. (1988) determined that height of the semivariogram sill parameter (Fig. 2) measured over forests was tightly linked to tree density and percent cover, the range parameter was indicative of crown size, while increased variability in crown sizes produced more rounded semivariograms. The research presented here extends from Song and Woodcock (2003) and Song (2007) in which stand-wise forest canopy diameters have been modeled using a ratio of the semivariogram sill parameters calculated from two different image spatial resolutions.

From a remote sensing perspective, the semivariogram (Fig. 2) is a plot of the expected (E) semivariance (γ) between the electromagnetic radiation (EMR) recorded by the sensor (digital number, DN) at different pixel locations (x_i) separated by the lag (h) distance. In this context, DNs are treated as spatially random, regionalized variables where detected EMR is a function of spatial location, i.e., $DN_i = f(x_i)$ (Jupp et al., 1988; Song, 2007). The general formula for a plot of $\gamma(h)$ against h is

$$\gamma_f(h) = \frac{1}{2}E\left\{\left[f(x_i) - f(x_i + h)\right]^2\right\} = \sigma^2 - \mathrm{Cov}_f(h) \quad (1)$$

where σ^2 and $\mathrm{Cov}_f(h)$ are respective values for variance and covariance between pairs of points separated by distance h in a spatially stationary scene (Jupp et al., 1988, 1989; Song & Woodcock, 2002; Song, 2007). The range parameter of the semivariogram is the lag distance (h) at which the sill is reached, which defines the limit of spatial autocorrelation. The

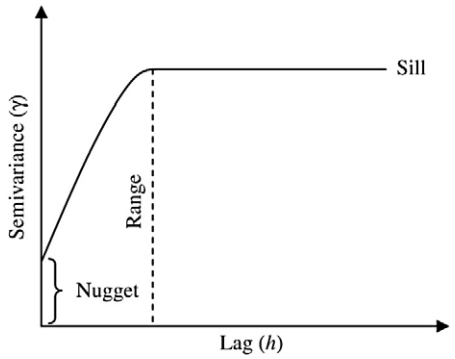

Fig. 2. Classic semivariogram shape for stationary data or imagery. The diagnostic parameters of the semivariogram are the range, sill, and the nugget effect.

sill parameter approximates the overall variance in an image, while the nugget effect represents semivariance at zero lag where variance is spatially independent (Curran, 1988), and is often related to random measurement error or uncorrelated sensor noise.

The Li and Strahler (1985) geometric optical model inversion approach for extracting tree crown shape and density parameters from image data, combined with further theory and model development by Jupp et al. (1988, 1989), provided the framework for landmark studies in satellite-based forest structure estimation using the semivariogram sill parameter (Woodcock et al., 1997; Song & Woodcock, 2002, 2003; Song, 2007). Models in this lineage, used to identify forest parameters that drive spatial and spectral variability in remotely sensed imagery, have presumed that tree crowns are opaque geometric features (disks) shading a contrasting forest floor (background), and that trees are randomly distributed within low or L-resolution (Strahler et al., 1986) pixels. The term "L-resolution" refers to satellite imagery that is composed of pixels that are larger than individual tree crowns, but smaller than a forest stand. The Song and Woodcock (2003) model builds on previous work in which horizontal canopy diameter (D_o) was determined to be the primary factor affecting image variance, assuming tree stem density (λ) was held reasonably constant (Woodcock et al., 1997; Song & Woodcock, 2002). As a result, it was theorized that the mean component spectral signatures (disk and background) remained unchanged when regularized over different pixel sizes (D_P) (Song & Woodcock, 2003; Song, 2007). Thus, division of two formulations for a regularized sills (C) calculated over a stand (Z) at two different spatial resolutions (C_{z1}/C_{z2}) canceled out both the proportion of cover and the mean spectral signature terms, the results of which were found to correlate strongly with disk diameter among two-component images (Song & Woodcock, 2003, Appendix A).

Earlier, Song and Woodcock (2002) determined that as pixel sizes increase, the variogram sill parameter decreases as a function of object size while holding percent cover constant. When cover was allowed to vary between 20 and 80%, the maximum sill variance for any combination of D_p and D_o occurred for stands having 50% cover (Song & Woodcock, 2002). However, they discovered that there was a unique curvilinear relationship between the regularized sill variance and the D_p/D_o ratio, where the rate of decrease in sill variance from the punctual resolution was a fixed function despite the magnitudes of D_p or D_o (Fig. 3). The sill of the variogram in this regard was found to be most sensitive to regularization when ratio $D_p/D_o = 1$, and continued to provide useful information for inferring tree canopy diameter up to $D_p/D_o \leq 3$ (Fig. 3).

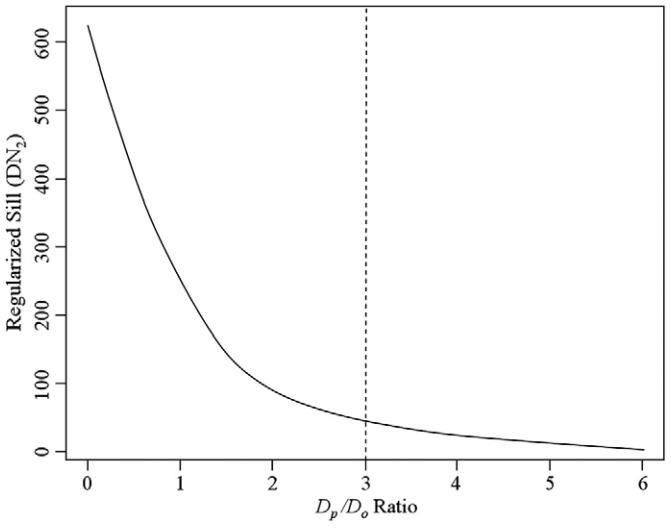

Fig. 3. Negative curvilinear effect of regularization on sill variance as D_p/D_o ratio increases. Sills initially exhibits a steep rate of decrease from punctual variance then become much less sensitive to regularization at $D_p/D_o > 3$. The function remains constant despite either pixel (D_p) or canopy (D_o) size. Adapted from Song and Woodcock (2002).

Song and Woodcock (2003) tested their two-component disk model over conifer canopies (diameter range 1–5 m) using the sill ratio technique. A series of paired resolutions (1–2, 2–4, 3–6, 4–8, 5–10, 10–20, and 15–30 m) derived from high spatial resolution IKONOS imagery were used to calculate sill ratios. When pixel diameters exceeded canopy diameter ($D_p/D_o > 1$) the error terms for crown diameter estimation tended positive; and then negative when $D_p/D_o < 1$. The 3–6 m pixel combination, not the finer resolutions, provided the most accurate estimates of tree crown diameter over the 1–5 m tree size range, with both the lowest error variance and most evenly distributed error bias around zero. Results using the 5–10 m pixel combination had similar error variance, but the bias was slightly more positive.

Song (2007) conducted a similar experiment in North Carolina using IKONOS panchromatic imagery to calculate variogram sills regularized in 1 m increments up to 7 m using a 1-meter separation between the sill ratio numerator and denominator (i.e., C1/C2, C2/C3, …, C6/C7). This time, however, the full complement of 11-bit digital numbers (DNs) was used rather than first converting imagery to a two-component scene as in Song and Woodcock (2003). This strategy worked moderately well for conifers ($R^2 = 0.73$), but poorly for continuous hardwood canopies ($R^2 = 0.43$). Song (2007) concluded that if it is difficult to see individual hardwood crowns from the ground that derivation of a tree crown size parameter from optical satellite data would hardly be possible at any resolution.

Though such conclusions may have stalled interest in this basic technique for use in hardwood forests elsewhere, and along with it the notion of comprehensive structure mapping, it remains virtually untested among forests where hardwood crown closure rarely approaches 100%, such as in northern Minnesota and western Ontario. Thus, the primary objective of this research is to extend the Song (2007) approach from stand-level to pixel-level estimation of mean crown diameter and other forest structural parameters, then operationalize the method for broad application to facilitate regional modeling of insect–host dynamics in northern forests.

1.2.2. Partial least-squares regression

PLS is a predictive, 2-block regression technique that performs simultaneous decomposition of dependent and independent observation matrices (e.g., forest structural parameters and spectra) collected from a single object of interest (Norgaard et al., 2000). The routine strives to balance explanation of variance in both the response and predictor variables (SAS, 2000), while also minimizing potential multicollinearity effects that would otherwise preclude the use traditional multiple linear regression (see Wolter et al., 2008). This is performed by selecting a number of eigenvectors from the independent data matrix that are used to generate score values that best capture predictor variance and correlate strongly with dependent variables (Arenas-Garcia & Camps-Valls, 2007). Since PLS regression does not assume zero error in the predictor data, often mistakenly assumed for remote sensing data (Curran & Hay, 1986), a primary assumption of ordinary least-squares and multiple linear regressions are avoided. What PLS regression does assume is that if well sampled, vectors in the predictor space (e.g., spectral response, regardless of error) should facilitate greater predictive power for additional observations when correlation among predictor variables is high (SAS, 2000).

In theory, it is possible to compute a set of components equal in number to variables in the predictor set; however, the standard procedure is to generate fewer initial components (SAS, 2000). In doing so, the chance of model overfitting is reduced as the lower order components often retain collinearity problems or describe only random measurement error (Geladi & Kowalski, 1986). Cross-validation (leave-one-out) iteratively determines which of the remaining components to use for optimal model complexity (Stone, 1974; Geisser, 1974; Wold et al., 1984). A more detailed discussion of the theory and application of PLS regression may be found in Geladi and Kowalski (1986) and Wolter et al. (2008).

With particular reference to remote sensing data, enhancements to PLS model prediction potential have been achieved by pre-selecting which spectral bands to use for PLS regression analysis (Lindgren et al., 1993; Swierenga et al., 1998; Leardi & Gonzalez, 1998; Forina et al., 1999; Jarvis & Goodacre, 2005; Li et al., 2007; Wolter et al., 2008). Careful design of any pre-selection procedure is essential, as the process aims to exclude spectral bands showing low sensitivity to the response variable. If low or non-responsive bands or variables are retained they have been shown to adversely affect model calibration and result in both large relative bias toward zero and small additive bias away from the origin regardless of signal to noise distribution (Spiegelman et al., 1998). Wolter et al. (2008) successfully used an iterative pre-selection approach where the PLS routine and the associated predicted residual sum of squares (PRESS) statistic were used in a backward stepping fashion to weed out Landsat TM image variables that exhibited low sensitivity to forest basal area (BA). Their use of PLS regression allowed estimation and mapping of *Abies balsamea* and *Picea spp.* BA for a 6.4 million ha area in northern Minnesota and neighboring Ontario. In the research presented here we make use of PLS regression and a band pre-selection approach, following Wolter et al. (2008), using SPOT bands and derivatives to estimate and map forest structure parameters in northeast Minnesota.

2. Methods

2.1. Study region

The 3600 km^2 study region consists of the area of a single 60 × 60 km SPOT-5 footprint (K587 J253, centered at 47.681° latitude and −91.345° longitude) located within the Superior national forest (Fig. 1). Forest cover within the region is diverse (five conifer genera and seven hardwood tree genera) and is considered transitional between the sub-boreal, Great Lakes–St. Lawrence forests and boreal forest (Heinselman, 1973; Baker, 1989). In general, these forests are intensively managed for wood fiber, which has resulted in a dominance of aspen (*Populus tremuloides* and *P. grandidentata*), birch (*Betula papyrifera*), spruce (Picea glauca), and fir (*A. balsamea*) forest associations (Wolter & White, 2002; Pastor et al., 2005; Friedman & Reich, 2005). The northern portion of the region is largely protected (BWCA wilderness), and has an extensive fire history that supports vast stands of pioneer forest dominated by jack pine (*Pinus banksiana*) as well as containing remnants of old-growth white and red pine (*Pinus strobus* and *P. resinosa*) forests (Heinselman, 1973; Frelich & Reich, 1995).

2.2. Field data

Field plot data used to model forest structural parameters were collected during the summers of 2006 and 2007 and distributed evenly within the SPOT-5 footprint to facilitate comprehensive mapping (see Wolter et al., 2008). Each of the plots ($n = 120$) consisted of a total of five subplots: one located at the intersection and one at each of the four end points of two crossing 60 m transect lines placed near center of large (≥4.4 ha), homogenous stands (Fig. 4). Sufficient stand size and homogeneity assured that stand edge effects were minimized during analysis, and that image misregistration errors were inconsequential. Estimates of basal area (BA) by species were collected at each subplot using angle count sampling with a metric basal area factor two prism (Grosenbaugh, 1952). Estimates of percent canopy cover (hence forth refereed to as canopy closure) were calculated from densitometer measurements made every meter along each of the two 60 m axes within a plot (Fig. 4). Densitometer measurements made in 2006 were recorded as either open sky (O) or tree (T), where the number of tree canopy hits (T) divided by the total number of densitometer observations from a plot ($n = 121$) provided plot-level estimates for canopy closure. In 2007 densitometer measurements were recorded as open sky (O), conifer (C), or hardwood (H), which allowed the proportion of canopy closure associated with conifer and hardwood tree components to be calculated ($n = 79$ plots). In addition, photos of each subplot (one vertical and four horizontal) were taken and general site information (e.g., slope, aspect, etc.) was recorded. It should be noted that use of the term hardwood refers to broad-leaf angiosperms.

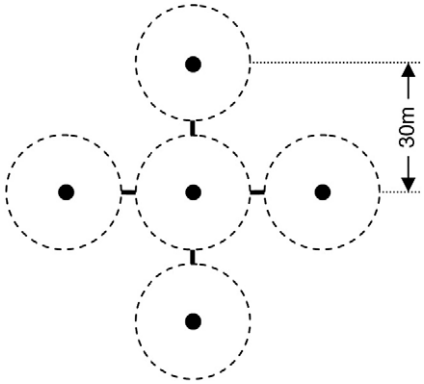

Fig. 4. A single ground plot composed of five variable radius subplots along two perpendicular axes, where subplots are separated by 30 m. The area surrounding each subplot center, within which trees are tallied, varies in radius in accordance with the bole diameter of the trees involved (Grosenbaugh, 1952).

For 106 of the 120 field plots, trees that were identified in basal area prism sweeps for the center subplot were split into two groups based on visual assessment of canopy diameter: those having smaller and larger canopies. In each group, a median tree representing canopy diameter was measured: canopy diameter (CDIA), bole diameter at breast height (DBH), tree height (HT), and vertical length of the live crown (LC), where LC is tree height minus the height to the lowest live branch (Appendix A). Nine of the 106 plots were placed in early regeneration stands (<5 years old) and were not used for modeling purposes. Both HT and LC measurements were made using an Impulse 200 laser range finder (Laser Technology Inc., Edgewood, CO) mounted on a monopod. Measurement of CDIA was based on the average of two canopy diameter measurements per tree (maximum and minimum canopy width) by using a densitometer to accurately estimate the vertical projection of the canopy edge on ground. Measurements were averaged across the two trees to produce single metrics for HT, LC, DBH, and CDIA per plot. Ground-based estimates of CDIA were cross-checked against estimates derived from low altitude (300 m above ground level) aerial photographs collected on 5 October 2006.

2.3. SPOT image data

Two SPOT-5 images were used in this study: 1) a 15 July 2006 multi-spectral (XS) image with four bands including visible green (500–590 nm), visible red (610–680 nm), near-infrared (780–890 nm), and shortwave infrared (1580–1750 nm), and 2) a 31 August 2006 5-meter panchromatic (PAN) image (480–710 nm). The 15 July 2006 XS image was acquired at 17:05:14 Zulu with a westward looking (284.36°) earth incidence angle of 8.17° at scene center, while sun elevation and azimuth angles at time of overpass of were 60.97° and 147.32°, respectively. The 31 August 2006 PAN image was acquired at 17:01:24 Zulu with sun elevation and azimuth angles of 49.04° and 155.23°, respectively. The PAN image was also westward looking (283.43°), but with an earth incidence angle 13.55° off nadir. It should be noted that SPOT-5's XS image bands are recorded at a resolution of 10 m with the exception of the SWIR band, which is actually recorded at 20 m then resampled to 10 m.

Both the XS and PAN images were orthorectified and geocorrected using a sensor-specific model (Erdas Imagine, Leica GeoSystems, 2006), 1-meter National Agricultural Imagery Program (NAIP) data from 2003–04 (source: www.lmic.state.mn.us/chouse/), and a 10 m digital elevation model (DEM) (source: www.usgsquads.com/

elevationdata.htm), after which each 10 m XS pixel was converted to four 5-meter pixels (XS_5) using a direct linear transform. This was done so that the two images could later be stacked into one file for easier processing and data extraction. The XS_5 and PAN images were then coregistered using the Erdas Imagine AUTOSYNC routine (Leica GeoSystems, 2006). The routine automatically identified 500 evenly distributed image-to-image tie points using sub-pixel correlation, after which a 1-meter RMSE threshold was set prior to AUTOSYNC's calculation of a second-order block triangulation to co-register the two images.

2.4. Local statistics and image predictor variables

Differentially corrected GPS locations for each field plot were used to identify SPOT image pixels closest to plot centers. For model development, field plots were categorized as conifer or hardwood based on composition of species in the forest overstory. A region-growing function based on spectral Euclidean distance (DN = 10) was then used to build local neighborhoods from which image predictor variables were calculated. Spectral Euclidean distances were calculated using XS bands 1–4 and the PAN band regularized to 10 m. The original 5 m PAN band was intentionally left out of this procedure to avoid any fine-scale bias. The region-growing algorithm ceased when a maximum radial distance of 97.5 m from the center pixel was reached. This strategy for calculating local image statistics was used for three reasons: 1) to circumvent problems associated with boundary pixels or edge effects; 2) to represent an area large enough so variance measures could be used as *de facto* proxies for the semivariogram sill parameter (see Coburn & Roberts, 2004); and 3) to eliminate dependency on detailed forest maps to guide these analyses.

The variables and local image statistics used for PLS model development are listed in Table 1. Image derivative bands such as NDVI (Rouse et al., 1974; Tucker, 1979) and MSI (Rock et al., 1986) were included as they are commonly used to study forest structural parameters (see Asner et al., 2003). Also, indices that make use of SWIR (e.g., MSI and SVR, Table 1) are important as this region of the electromagnetic spectrum has been shown to be sensitive to forest density, tree size, and BA (Cohen & Spies, 1992; Cohen et al., 1995; Hansen et al., 2001; Lu et al., 2004; Wolter et al., 2008). The SVR index (Wolter et al., 2008) was originally developed for use with Landsat TM and ETM+ data to provide a SWIR-based index that specifically excluded NIR. Subsequently, the SVR was found to be more sensitive to the relative BA of spruces and fir in northern Minnesota than other routinely used SWIR-based indices (Wolter et al., 2008). Lastly, local semivariogram sill parameters for the 5 m PAN band (P5) and two simulated 10 m PAN bands (P10P and P10X, Table 1) were calculated using a spherical model as it has been found to produce a better fit than other standard models when applied to remotely sensed images of forest cover (Coburn & Roberts, 2004; Lévesque & King, 1999; St-Onge & Cavayas, 1995).

2.5. PLS regression model development and mapping

PLS regression models were developed for six dependent forest structure variables (CDIA, DBH, HT, LC, BA, and CC) for both hardwood and conifer dominated field plots. A recursive backward elimination band selection procedure, described by Wolter et al. (2008), was used to pre-select the most relevant image predictor variables from the full set available (Table 1). At each step in this band selection process (including final models) a leave-one-out cross-validation was performed by recursively removing individual ground plot data points until each point was withheld once. Retained and withheld ground plot data points were then used to compute respective residuals and prediction error for each level of model complexity. The cumulative sum of these individual point prediction errors, known as the predicted residual sum of squares (PRESS), provided the measure of model predictive capacity (Wold et al., 1984, 2008). However, an additional assessment of the 12 final models was performed via jackknife cross-validation (see Coops et al., 2003) where 50 randomizations withholding 20% of ground plot data was performed at each step.

After PLS regression model development, the best models were used with SPOT-5 image data to estimate and map the six forest structure parameters across the study area on a per pixel basis. To do this, an algorithm was developed to automatically extract necessary image predictor variables for all forested areas within the study region. First, all non-forested areas were masked out of the SPOT-5 imagery using an unsupervised classification consisting of 100 classes, where 90 randomly distributed points across 57 aerial photographs served as validation data for accuracy assessment. Then, for each forest pixel, a local neighborhood of pixels was grown outward (radius ≤ 97.5 m) using a spectral Euclidean distance of 10 DN. The number of neighbors identified at each pixel location was written to a separate band to be used later to identify any potentially problematic cells, e.g. edge or isolated forest patches that do not contain a sufficient number of pixels to calculate a semivariogram. Once the image predictor variables were generated (Table 1), the SPOT-5 near-infrared (NIR) band was used as a discriminant at each pixel location to determine whether conifer (DNs < 105) or hardwood (DNs ≥ 105) structure models would be applied. The NIR threshold value (105) is the average DN between pure hardwood (mean 121.98) and pure conifer (mean 88.79) pixels within local neighborhoods (~20,000 pixels total) associated with hardwood and conifer field plot locations (44 and 53 plots, respectively). The NIR band was used as it has been shown to be a robust discriminant for these major forest types in this region (Shen et al., 1985). The accuracy of discriminating conifer and hardwood dominated stands in this fashion was assessed using densitometer-based canopy closure estimates from the 2007 field season, where the proportion of conifer and hardwood canopy elements were recorded ($n = 79$ plots). Cases where conifer elements constituted >50% of the plot's canopy closure estimate were considered conifer dominated, otherwise the stand was considered hardwood dominated. Using this rule, 39 of the plots were classified as conifer and 40 as hardwood.

3. Results

3.1. PLS regression models

3.1.1. Combined structure models

The development of a single, viable PLS regression model for all hardwood structure variables was unsuccessful—even when models were highly simplified. However, separate PLS regression models for each hardwood dependent variable were successful. On the other hand, development of PLS regression models for all conifer structure parameters was feasible; explaining ~72% of variation among all parameters (Table 2A). Coefficient of determination (R^2) values for

Table 2
Two sets of PLS regression statistics for combined conifer structure models.

			A			B
Image variables used			13			12
Factors used			4			4
PRESS			0.59			0.48
Model σ^2 explained (%)			84.93			84.04
Ground σ^2 explained (%)			71.75			81.81
Model Pr > F			0.0001			0.0001
	R^2	Adj. R^2	RMSE	R^2	Adj. R^2	RMSE
CDIA (m)	0.87	0.87	0.64	0.88	0.88	0.63
DBH (cm)	0.86	0.86	4.40	0.86	0.86	4.41
HT (m)	0.86	0.86	2.05	0.86	0.86	2.03
CC (%)	0.69	0.68	6.36	0.70	0.69	6.32
LC (m)	0.80	0.79	1.30	0.79	0.79	1.31
BA (m^2/ha)	0.22	0.21	6.20	–	–	–

Model A includes predictions for all forest structure parameters, while model B excludes basal area.

the six conifer structure parameters ranged from 0.22 for BA (RMSE 6.2 m^2 ha^{-1}) to 0.87 for CDIA (RMSE 0.64 m). Due to poor model performance for the conifer BA parameter (Table 2A, Fig. 5), a second PLS model was generated that excluded BA (Table 2B), which accounted for 82% of the variation in the remaining five parameters. However, while excluding BA had little effect on R^2 and RMSE values among retained parameters, overall explanation of variance in the field data improved by 10% and PRESS by 11%. For each of these combined PLS models, CDIA, DBH, and HT had similar R^2 values between 0.86–0.88, with RMSE values of ~0.64 m (8.1% of range), ~4.41 cm (8.8% of range), and ~2.04 m (9.4% of range), respectively, while CC (R^2 0.70, RMSE 6.36% or ~11.4% of range) and LC (R^2 0.80, RMSE 1.30 m or ~8.7% of range) results were moderate by comparison. The final combined PLS regression model for conifer structure used 13 image predictor variables that were represented by four latent factors (Table 2A).

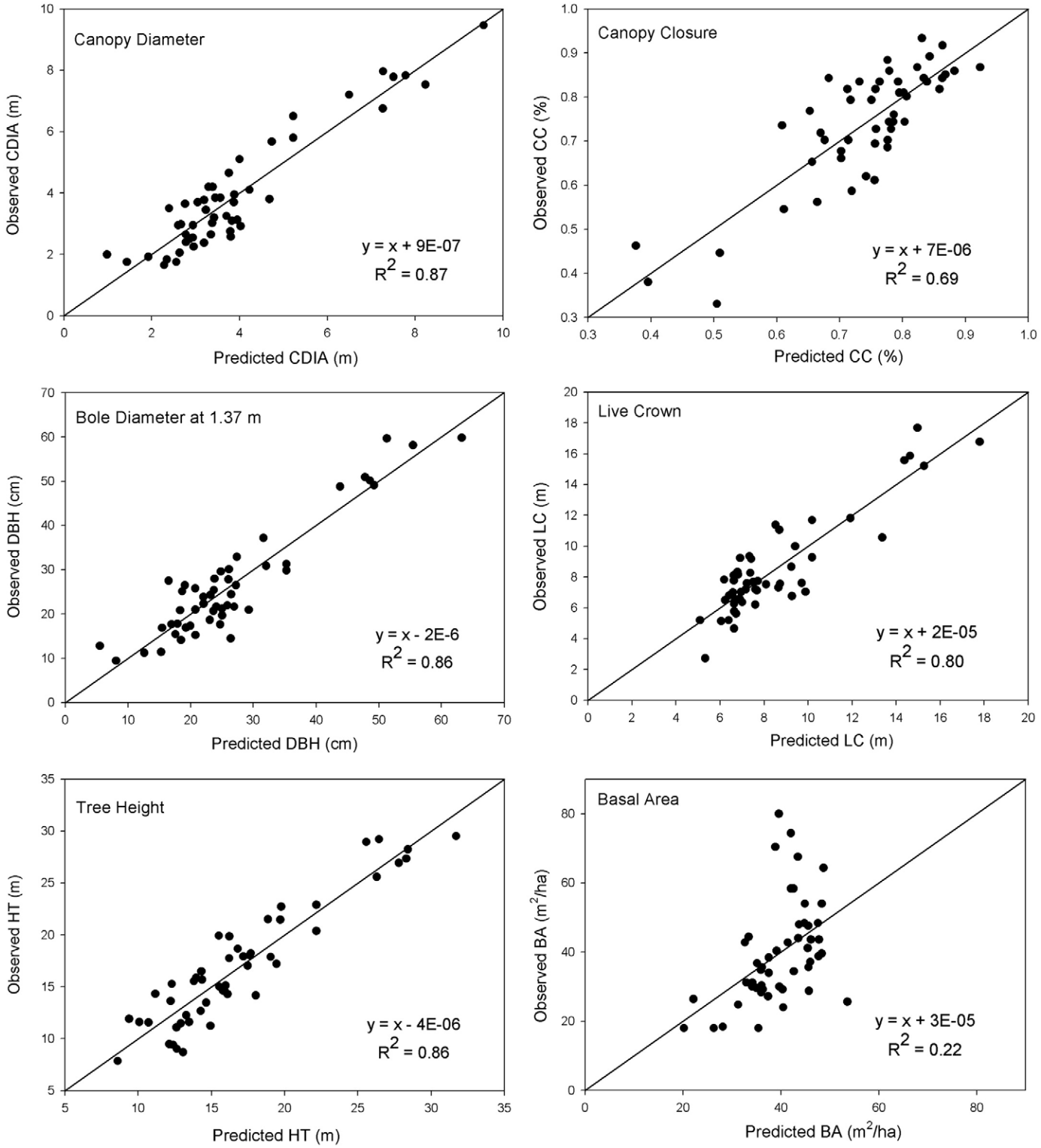

Fig. 5. PLS regression results for the combined conifer structure model. Note that the basal area (BA) results were poor when modeled in combination with the other parameters.

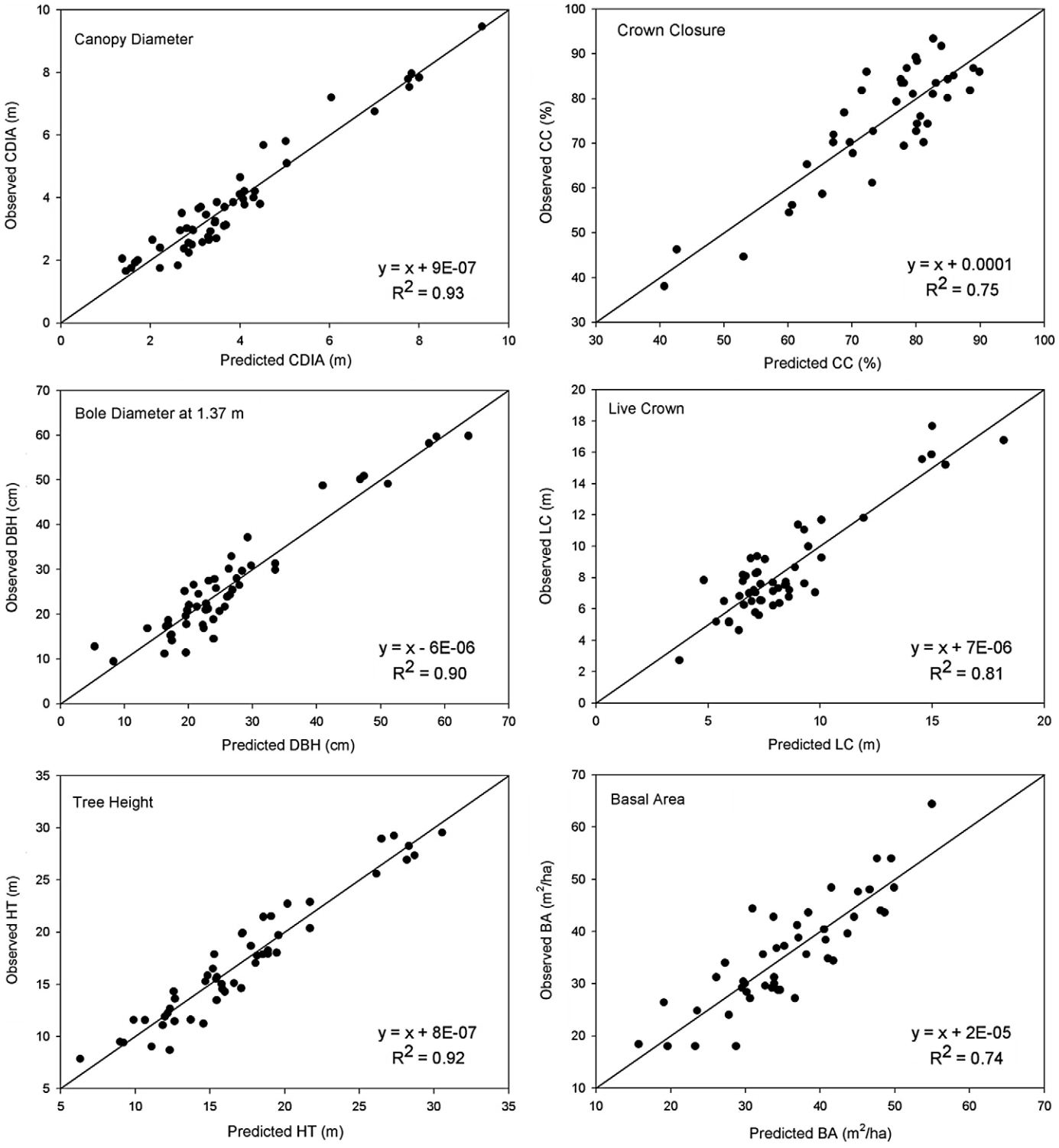

Fig. 6. PLS regression results for the six individual conifer structure models. In general, individual models out perform the combined models; especially for basal area (see Fig. 5).

3.1.2. Parameter-specific PLS structure models

Each of the six parameter-specific PLS regression models for estimating conifer stand structure (Fig. 6) resulted in lower RMSE and higher R^2 values compared to the combined-model estimates (Tables 2A and 3A). The greatest improvements in model performance came from CDIA, BA, HT, and DBH, where RMSE values were reduced by 26.6%, 26.1%, 22.4%, and 14.8%, respectively, over the combined PLS regression model results (Tables 2A and 3A). CDIA and HT models had the lowest PRESS values (lower = better) at 0.30 and 0.32, respectively, followed closely by DBH (0.37). Conversely, CC and BA had the highest PRESS values at 0.65 and 0.55, respectively, while LC was intermediate among the extremes (Table 3A). In general, conifer model performance exceeded that of hardwood models in terms of R^2, but not RMSE (Table 3).

Component loading results for individual conifer structure models (Fig. 7) show that of the 27 image predictor variables tested (Table 1) the CDIA model used the fewest at eight, the CC model used the most at 22, and the remaining conifer structure models used between 10 and 12 (Table 3A). Component loadings for conifer CDIA were

Table 3
Parameter-specific PLS regression model statistics for conifers (A) and hardwoods (B).

	CDIA (m)	DBH (cm)	HT (m)	BA (m2/ha)	CC (%)	LC (m)
A						
Image variables used	8	11	11	10	22	12
Factors used	8	10	7	4	4	2
PRESS	0.30	0.37	0.32	0.55	0.65	0.48
Model variance explained (%)	100	100	90.51	87.92	70.60	54.17
Ground variance explained (%)	93.47	90.48	92.06	74.05	67.98	81.48
Model Pr>*F*	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
LCV R^2	0.93	0.90	0.92	0.74	0.68	0.81
LCV RMSE	0.47	3.75	1.59	4.58	6.02	1.25
JCV R^2	0.91	0.83	0.88	0.58	0.52	0.77
JCV RMSE	0.59	5.51	2.04	6.66	7.66	1.57
B						
Image variables used	17	18	7	10	11	7
Factors used	8	8	4	8	1	4
PRESS	0.55	0.56	0.66	0.65	0.74	0.74
Model variance explained (%)	97.61	96.66	90.41	99.94	30.91	66.27
Ground variance explained (%)	82.26	82.25	68.51	71.18	51.99	57.84
Model Pr>*F*	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
LCV R^2	0.82	0.82	0.69	0.71	0.52	0.58
LCV RMSE	0.62	2.92	1.27	2.47	5.49	0.96
JCV R^2	0.74	0.79	0.52	0.63	0.52	0.53
JCV RMSE	0.65	5.06	1.83	4.28	7.66	1.42

Coefficient of determination (R^2) and root mean squared error (RMSE) are reported for both leave-one-out cross-validation (LCV) and jackknife cross-validation (JCV) with 50 randomizations withholding 20% of the plot data at each step.

strongest, in decreasing order, from SNIR, SVR, and RED with the S510X ratio fifth in importance. The top three loading weights for conifer DBH came from MSI, NDVI, and SVR, while ratios S510X and S510P were ninth and 10th in importance, respectively. Among the height metrics, S10X, V510X, and SNIR represented the top three component loadings for conifer HT, while V5, C5, and V510X were strongest for conifer LC. Conifer BA and CC models each were loaded most strongly on the MSI predictor, while the ratio C510X was third and 12th in importance for BA and CC, respectively (Fig. 7).

Individual PLS regression models of hardwood forest structure explained between 52% (CC) and 82% (both CDIA and DBH) of the variation in field plot data for the six dependent variables studied (Table 3B, Fig. 8). BA, HT, and LC were intermediate in this regard accounting for 71%, 69%, and 58% of the variation in field plot data, respectively. As mentioned earlier, RMSE of prediction values for all the hardwood structure variables, except CDIA, were lower than the respective individual conifer structure model values (Table 3).

Component loadings for hardwood CDIA were strongest, in descending order, from C10P, S10P, and SVR, while the ratios S510P and C510X were eighth and 15th in importance, respectively, among the 17 predictor variables used (Fig. 9). For the hardwood DBH model, SVR, C10P, and S5 had the strongest component loadings, while the ratio S510P was 13th among 18 predictor variables used. Hardwood HT and LC models each used seven image predictor variables which was the least among the six models. Component loadings for the hardwood HT model were strongest, in decreasing order, for C10P, SWIR, and S5 with the C510P ratio fifth in importance, while strongest loadings for the LC model came from SVR, S5, and the V510P ratio. Hardwood BA and CC models used 10 and 11 image predictor variables, respectively, with BA component loadings weighted heaviest on SVR, NIR, and MSI, while CC component loadings were strongest from NIR, SWIR, and NDVI. All individual hardwood and conifer PLS structure models, except hardwood CC, made use of the 5–10 m ratio data in some form (i.e., standard deviation (*S*), variance (*V*), or sill (*C*), Table 1).

3.1.3. Pixel-wise structure mapping

Pixel-wise semivariogram sill parameter data, calculated from the PAN image bands (P5, P10P, and P10X, Table 1), were successfully generated for over 90% of the forest pixels in the SPOT scene. The remaining forest pixels either produced unreasonable sill values that grossly exceeded the neighborhood local variance (e.g. when local neighborhoods consisted of less than 50 pixels) or were returned as zero values when a sill parameter was not attainable. Pixels in the former case were often associated with road–forest edges or other forest–non-forest edges, while pixel in the latter case, that failed to produce any sill values at all, appeared as single pixel locations distributed evenly over the study area, and were not associated with edge features. Pixels of the later nature were filled in the final structure maps (Fig. 10) using a 3×3 median filter, whereas pixels having erroneous sill values due to proximity with forest–non-forest edges were recoded to zero.

The accuracy of the initial classification used to mask out non-forest areas was 94.4% ($n=90$, khat $=0.89$), while the accuracy of using the SPOT-5 NIR band to discriminate between conifer and hardwood dominated pixels (DN$<$105 and DN$\geq$105, respectively) was 92.4% ($n=79$, khat $=0.85$).

4. Discussion

4.1. Sill ratios, proxies, and SWIR-based indices

For northern Minnesota forests, contrast among sunlit and shaded canopy components was sufficient to use ratios of semivariogram sill parameters calculated at 2 pixel resolutions to characterize CDIA. Hardwood forest types in this study (primarily aspen, birch, ash, and maple) are composed of canopies that are clearly discernable from the ground and from aerial photographs (Fig. 11A). Conifer canopies, while also fulfilling this requirement, appear to exhibit greater contrast, likely due to what we suspect is a more complete shading of the forest floor (Fig. 11B). In addition, the mean tree crown diameter within the study area is approximately 4 m ($\sigma=1.7$ m, range 1.6–9.6 m), translating to D_p/D_o ratio values of 1.3 and 2.5 for 5 m and 10 m data, respectively, which are well within the effective limit of the 3.0 threshold defined by Song and Woodcock (2002).

Initial predictor-wise ordinary least-squares (OLS) regressions against each of the dependent structure variables provided an indication of the importance of the 31 August 2006 PAN-based local variability measures: S5, S10P, V5, V10P, C5, and C10P. All of these variables were strongly related (R^2 range of 0.60–0.77) to the conifer structure variables CDIA, HT, DBH, and LC, which themselves are highly correlated to one another (Table 4). We had expected that a subset of these spatial variability measures would be strongly indicative of either conifer CC or hardwood CC in accordance with Woodcock et al. (1988), but these relationships did not emerge. In any event, while some of these OLS relationships between image predictor variables and dependent structure variables were quite good, simultaneous use of all the image predictor variables yielded far superior results by comparison (Table 3). PLS regression provided a convenient way to both circumvent collinearity issues among multiple image-based predictors and assess the importance of these variables for estimating specific forest parameters.

Although we effectively modeled six forest parameters using these SPOT-5 data, PLS regression results show only a few cases (conifer HT, LC, BA, and hardwood LC) where ratio data (C510X, V510X, and V510P, see Table 1) could be considered strong image predictor variables among the 27 variables tested (Figs. 7, 9). Unexpectedly, the sill ratio variables (C510X and C510P) were not dominant predictor variables in either the hardwood CDIA or conifer CDIA models, as we had presumed based on previous work (Song & Woodcock, 2003; Song, 2007). Rather, SVR and MSI often had greater factor loadings than the ratio variables. However, of the 12 reduced PLS models generated, 10 models included at least one ratio variable in the final model, with hardwood CC and BA

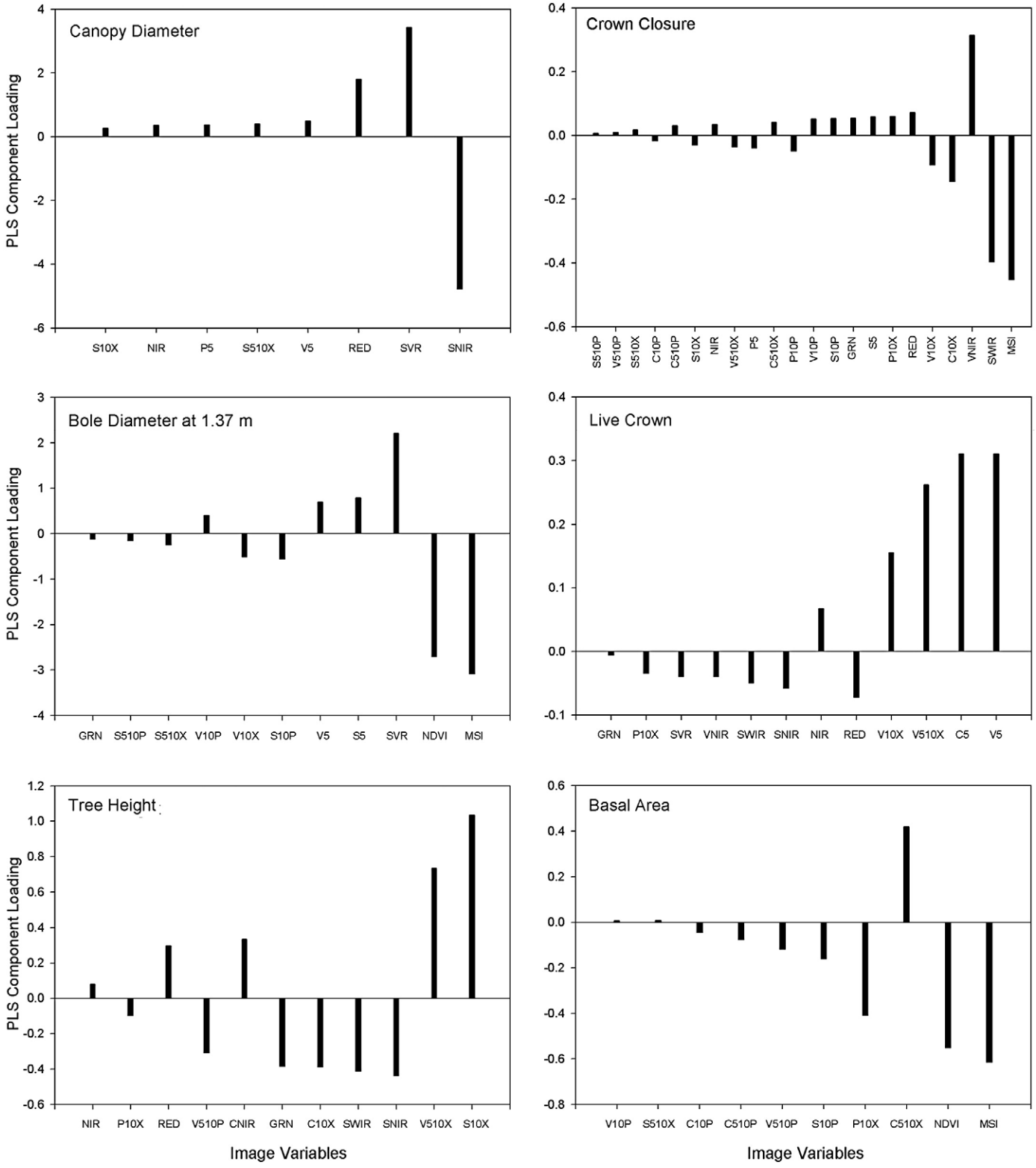

Fig. 7. Component loadings for image variables automatically selected for each of the six conifer structure models using iterative PLS regression according to Wolter et al. (2008). Loadings are arranged from least weight on the left to greatest weight on the right. Image variables are listed in Table 1.

models being the exceptions. Notably, the final hardwood CDIA model, which represents canopies with lower overall spectral contrast than conifer canopies, retained C510X, S510X, and S510P ratio variables, while the final conifer CDIA model retained only one (S510X) of the six ratios variables, indicating that S510X (Table 1) may be a fair proxy for the sill ratio in this case. Otherwise, a clear overall proxy for the sill ratio data did not emerge. In fact, exclusion of any of these ratio variables could compromise the results of this study given potentially suboptimal sensor and illumination geometry (discussed below).

This study extends Wolter et al. (2008) by showing that the SPOT-5 formulation of SVR (Table 1), while not equivalent to Landsat-based SVR (SVR_L), is still a strong predictor of BA, as well as DBH and LC for

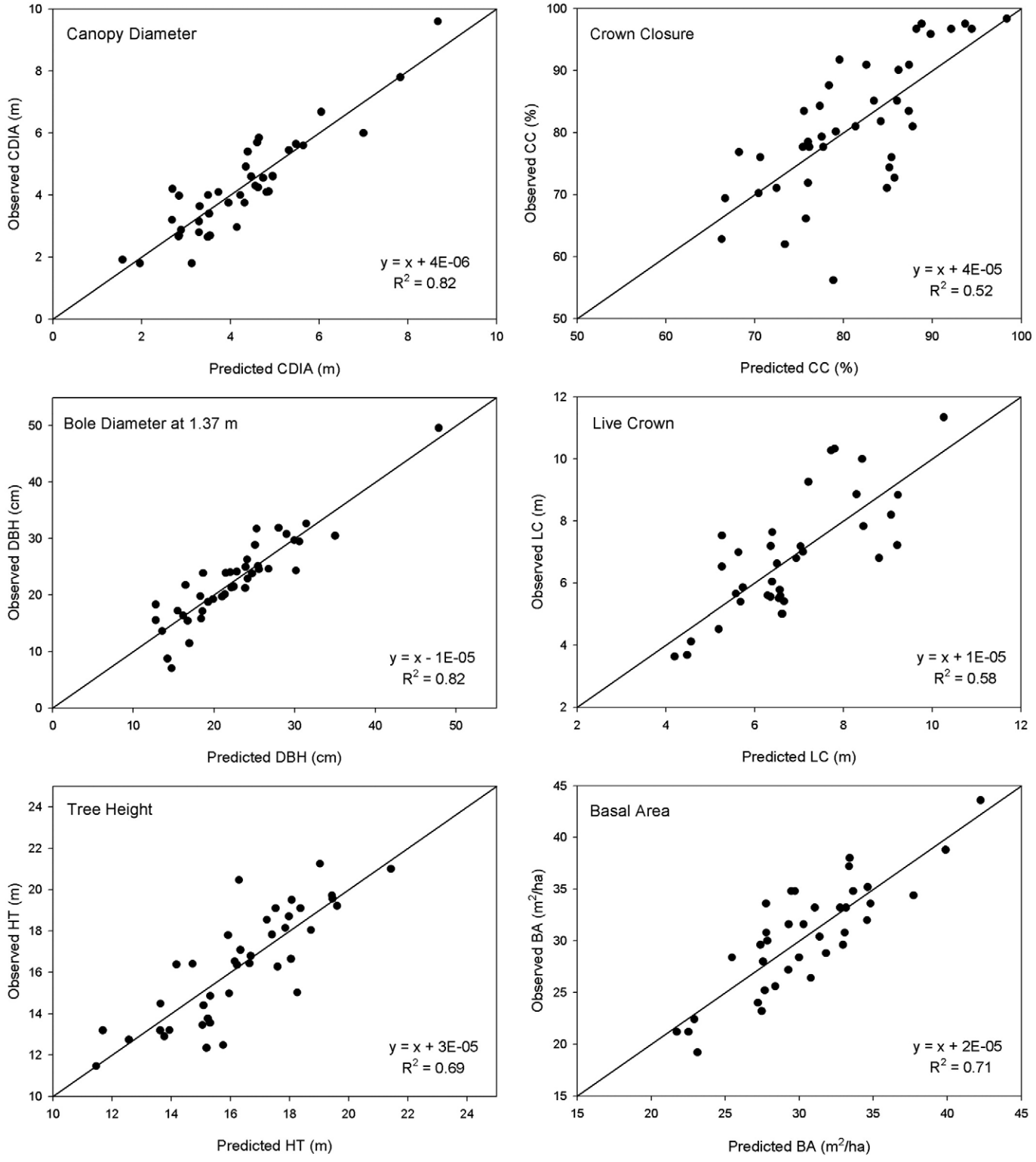

Fig. 8. PLS regression results for the six individual models of hardwood forest structure.

hardwood forests. However, MSI was dominant for predicting BA, DBH, and CC within coniferous forests (Figs. 7 and 9). Among the three strongest component loadings reported for each individual conifer and hardwood structure model sets, the SVR predictor appeared most frequently (six), followed by MSI (four), and all 5 m to 10 m ratio data combined (four) (Figs. 7 and 9). The fact that SVR_L and the SPOT-5 SVR (SVR_S) are not equivalent formulations may explain the apparent difference in sensitivities of this ratio to conifer BA among the two studies. SVR_L is the average of Landsat's two SWIR bands (SWIR5 1.55–1.75 μm, SWIR7 2.09–2.35 μm) divided by the average of Landsat's three visible bands (blue 0.45–0.515 μm, green 0.525–0.605 μm, red 0.63–0.69 μm), whereas SVR_S consists of SPOT-5's single SWIR band (1.58–1.75 μm) divided the average of its two visible bands (green 0.50–0.59 μm, red 0.61–0.68 μm). Since Landsat band 7 is known to

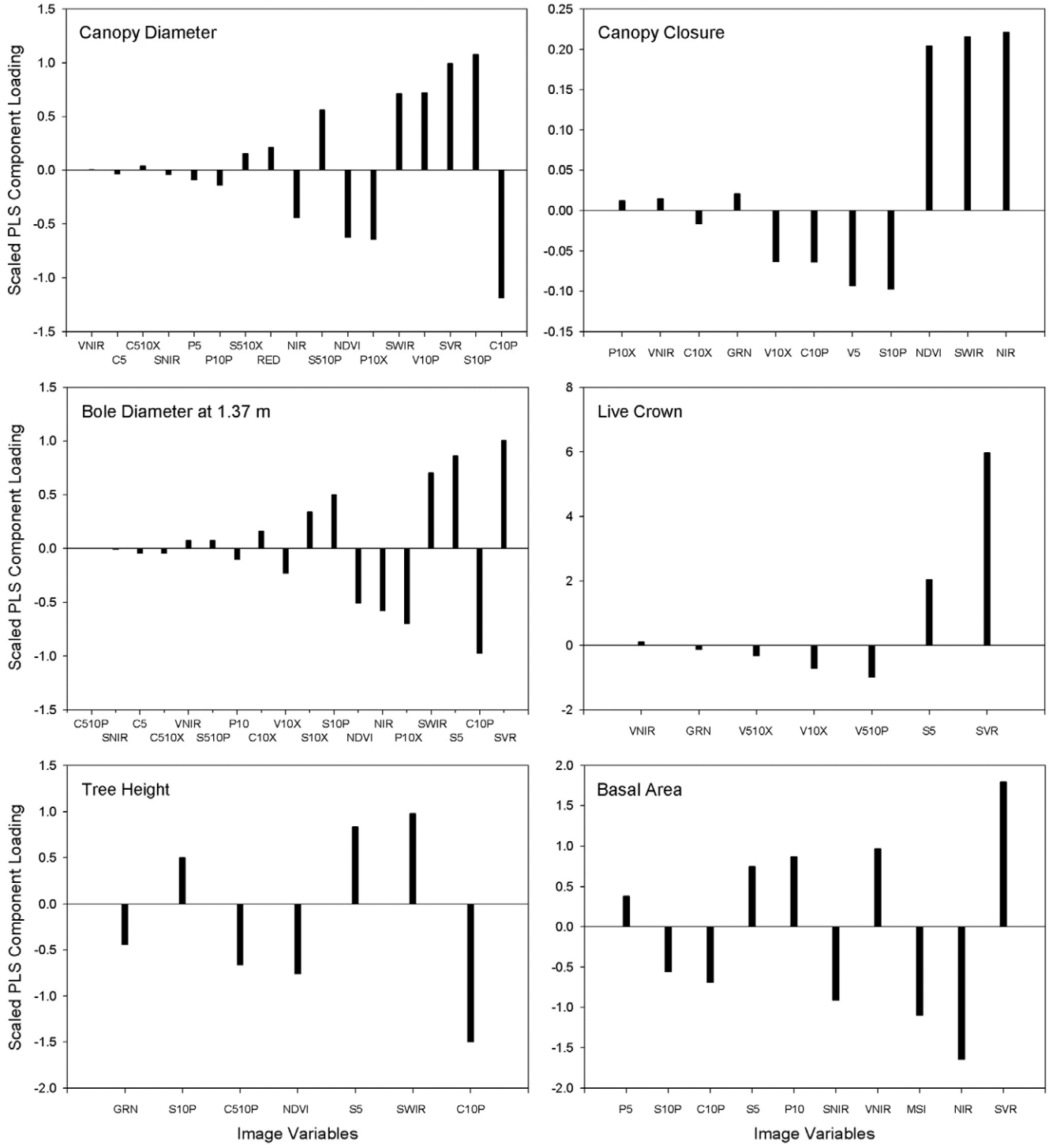

Fig. 9. Component loadings for image variables automatically selected for each of the six hardwood structure models using iterative PLS regression according to Wolter et al. (2008). Loadings are arranged from least weight on the left to greatest weight on the right.

correlate inversely with conifer BA ($r = -0.48$, Brockhaus & Khorram, 1992), we suspect the absence of this information in SVR_S plays a role in the sensitivity of this ratio to conifer BA, as well as providing additional evidence of the importance of this wavelength interval for modeling conifer structure.

4.2. Comparison with lidar forest structure studies

Maximum tree height and mean canopy height are among the most accurate and frequently estimated forest parameters using airborne lidar systems (Falkowski et al., 2006; Hyyppä & Inkinen, 1999; Anderson et al., 2008), as they are closely related to the quantities directly measured by these sensors (Lim et al., 2003b). Other lidar-based structure estimates such as canopy openness (Lim et al., 2003a), canopy diameter (Falkowski et al., 2006; Popescu et al., 2003), basal area (Lefsky et al., 1999; Næsset, 2002; Lim et al., 2003a; Holmgren, 2004; Jensen et al., 2006), bole diameter at breast height (Lefsky et al., 1999; Næsset 2002; Lim et al., 2003a; Jensen et al., 2006; Anderson et al., 2008), and wood volume or biomass (Lim et al., 2003a; Lim & Treitz, 2004; Anderson

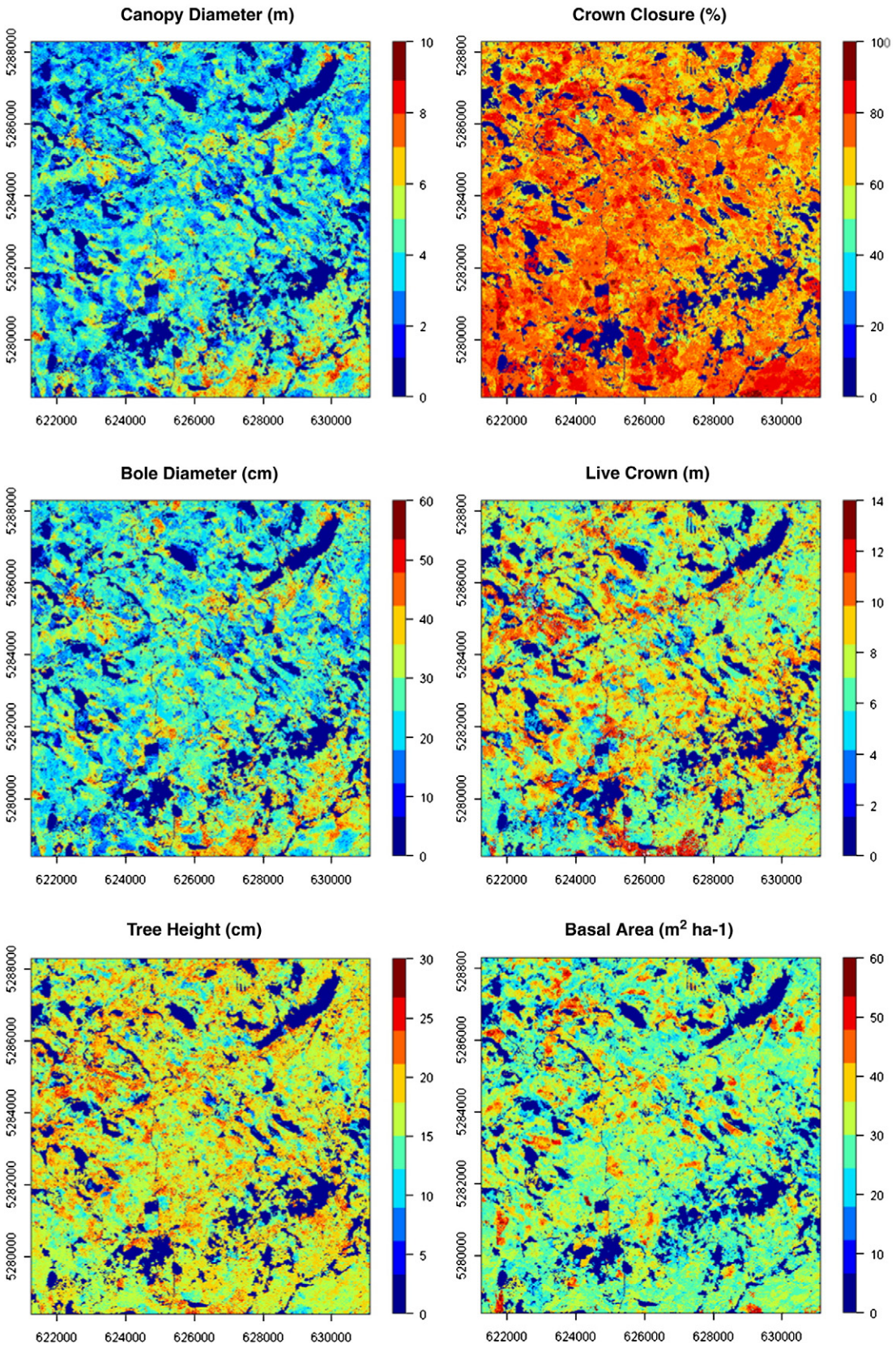

Fig. 10. Pixel-wise forest structure mapping results from a 10 × 10 km subset of the study area for canopy diameter (CDIA), bole diameter at breast height (DBH), tree height (HT), crown closure (CC), height of live crown (LC), and basal area (BA). Coordinates in both the *X* and *Y* directions are given in meters (UTM zone 15).

et al., 2008) are either modeled using combinations of lidar-derived variables or via fusion of these lidar variables with optical sensor data.

The accuracy of the forest structure parameters modeled here using SPOT-5 data compare favorably with similar parameter estimates made using lidar data (Table 5). In some instances, the SPOT-5 results were

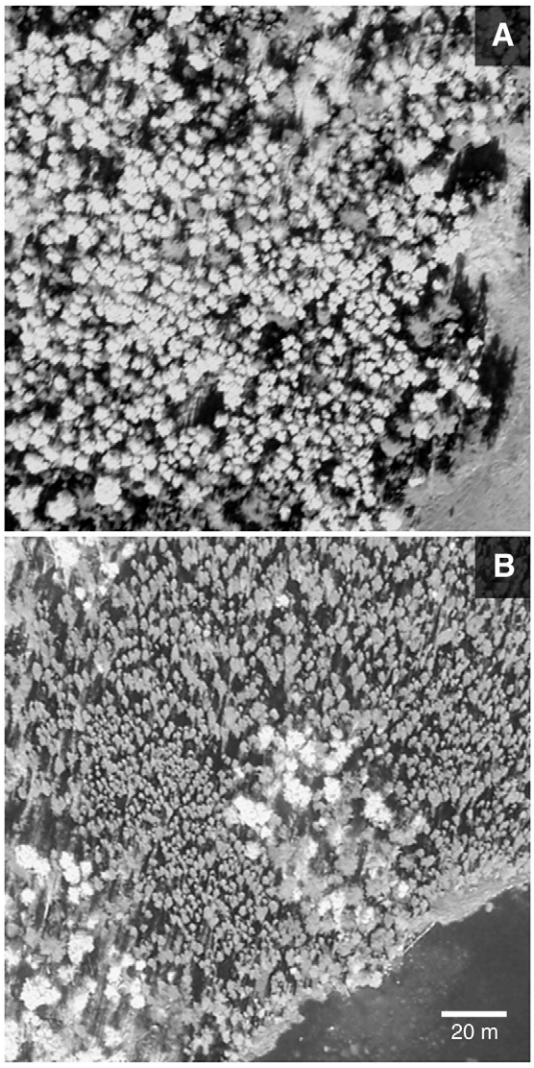

Fig. 11. Subsets of aerial photograph taken on 5 October 2006 showing typical hardwood (A) and conifer (B) forest cover within the study area.

Table 5
Comparisons between Minnesota forest structure estimates, modeled using SPOT-5 sensor data, and estimates modeled in other regions and forest types using lidar data.

Parameter	Forest type	R^2	RMSE	Units	n	Reference
CDIA	Conifer	0.93	0.47	m	53	The authors
	Conifer	0.86	1.35	m	29	Falkowski et al. (2006)
	Conifer	0.63	1.36	m	31	Popescu et al. (2003)
	Hardwood	0.82	0.62	m	44	The authors
	Hardwood	0.62	1.41	m	33	Popescu et al. (2003)
DBH	Conifer	0.90	3.75	cm	53	The authors
	Conifer	0.46	1.93	cm	38	Anderson et al., 2008)
	Conifer	0.61	6.31	cm	64	Jensen et al. (2006)
	Conifer	0.89	1.57	cm	61	Næsset (2002)
	Conifer	0.61	9.77	cm	22	Lefsky et al. (1999)
	Hardwood	0.82	2.92	cm	44	The authors
	Hardwood	0.43	2.22	cm	55	Anderson et al. (2008)
	Hardwood	0.63*	3.17	cm	48	Lim et al. (2003b)
HT	Conifer	0.92	1.58	m	53	The authors
	Conifer	0.94	2.64	m	29	Falkowski et al. (2006)
	Conifer	0.91	3.03	m	64	Jensen et al. (2006)
	Conifer	0.99	0.59	m	144	Holmgren (2004)
	Conifer	0.92	0.92	m	61	Næsset (2002)
	Hardwood	0.69	1.27	m	44	The authors
	Hardwood	0.94	2.29	m	39	Anderson et al. (2008)
	Hardwood	0.87*	1.65	m	49	Lim et al. (2003b)
CC	Conifer	0.75	6.02	%	53	The authors
	Conifer	0.79_a	8.00	%	64	Jensen et al. (2006)
	Hardwood	0.52	5.49	%	44	The authors
	Hardwood	0.76*	7.45	%	36	Lim et al. (2003b)
LC	Conifer	0.81	1.25	m	53	The authors
	Conifer	0.87_b	0.17	m	31	Maltamo et al. (2006)
	Hardwood	0.58	0.96	m	44	The authors
BA	Conifer	0.74	4.58	m^2	53	The authors
	Conifer	0.91	2.99	m^2	64	Jensen et al. (2006)
	Conifer	0.88	2.70	m^2	144	Holmgren (2004)
	Conifer	0.86	2.23	m^2	61	Næsset (2002)
	Conifer	0.87	11.81	m^2	21	Lefsky et al. (1999)
	Hardwood	0.71	2.47	m^2	44	The authors
	Hardwood	0.25	8.05	m^2	158	Anderson et al. (2008)
	Hardwood	0.85*	6.42	m^2	44	Lim et al. (2003b)

Structure parameters are canopy diameter (CDIA), bole diameter at breast height (DBH), tree height (HT), canopy closure (CC), vertical length of live crown (LC), and basal area (BA). Of the lidar studies listed, only results derived through use of solely lidar sensor data are presented. Studies that developed relationships from log transformed independent and dependent variables are indicated with asterisks (*). The subscript, a, under the R^2 column indicates that canopy openness was estimated, which is the inverse of CC, while subscript, b, indicates estimates of height to the bottom of the live crown and not LC per se.

notably better than their lidar-based counterparts in terms of R^2 and/or RMSE (e.g., conifer and hardwood CDIA and DBH), although lidar-based estimates out performed SPOT for several hardwood structure variables (e.g., HT and BA). The ratio-based image variables that were hypothesized to be most sensitive to CDIA did not turn out to be the most significant predictors. However, the superior performance of the structure models for conifers (especially HT, LC, and CC), compared to hardwoods, is linked principally to greater overall canopy contrast (e.g., Fig. 11) and to stronger correlations between CDIA and all but one (BA) of the remaining conifer structural parameters (Table 4).

Table 4
Correlation matrices for forest structure parameters among hardwood and conifer field plot data.

	CDIA	DBH	HT	LC	BA	CC
Hardwood						
CDIA	1.00					
DBH	0.89	1.00				
HT	0.46	0.53	1.00			
LC	0.37	0.35	0.55	1.00		
BA	0.27	0.30	0.32	0.12	1.00	
CC	0.27	0.20	0.31	0.21	0.31	1.00
Conifer						
CDIA	1.00					
DBH	0.93	1.00				
HT	0.85	0.89	1.00			
LC	0.80	0.80	0.80	1.00		
BA	0.10	0.25	0.08	0.05	1.00	
CC	0.35	0.45	0.43	0.27	0.69	1.00

4.3. Viewing geometry, solar illumination, and repeatability

The proportion of shaded versus sunlit crown seen by an optical sensor is a function of sensor look angle and solar illumination angle (Li & Strahler, 1985). As such, extrapolation of the results from this study to neighboring areas without recalibration depends primarily on holding illumination factors more or less constant in successive image acquisitions, assuming any atmospheric differences are correctable. In an ideal case, all imagery (XS and PAN) would be nadir looking and collected on the same date, with successive image acquisitions scheduled for anniversary dates to hold illumination angle constant. In this study, however, both the XS and PAN image acquisitions were westward looking with earth incidence angles that differed by 5.4°, while sun elevation angles differed by 11.9°. This sunny-side viewing geometry means a lower proportion of shaded canopy components are seen by the sensor, which may partially

explain why the ratio data were not stronger components in our CDIA models. As such, two scenarios would likely strengthen these relationships: 1) collection of imagery having eastward or shadow-side viewing geometry or 2) nadir imagery collected late in the growing season. Either case would enhance contrast between sunlit and shaded canopy components by increasing the proportion of shaded canopy visible to the sensor.

4.4. Ramifications and suggestions for future research

Song and Woodcock (2003) state that the sill ratio technique is only good if stand maps are available for calculating local statistics, but stand maps may often be outdated or unreliable. Thus, pixel-wise definition of Euclidean neighborhoods circumvents the stand map requirement by producing a dense network of pixel-wise estimates of forest structure which may be used as is or, if one wishes, in combination with an existing stand map to predict average structural dimensions for a set of stands or a region.

However, extraction of necessary image predictor information was not always possible due to cases where Euclidean neighborhoods surrounding each pixel were either too small to produce a semivariogram or did not yield a reasonable sill parameter from the input data. While this is somewhat inconvenient, it is not prohibitive to the process. In such cases, we simply chose to fill the holes in the final maps by using a 3×3 median filter applied only at these locations. One might also consider modeling the sill parameters (C5, C10P, and C10X) for dropout pixel locations by using the relationship of these sill metrics to V5, V10P, and V10X, respectively, over areas that did produce good sill parameter values.

Extension of forest structure models to similarly forested regions depends on whether SPOT-5 image data (XS & PAN) with the same approximate sun elevation and look angle geometry can be acquired. For this study, the chances of finding suitable archived image data in other areas within northern Minnesota that fit our specific geometry constraints are remote. However, recalibration of our models using nadir or close to nadir imagery (concurrent XS and PAN) from late in the growing season would be far easier to plan future image acquisitions around. Data from the Panchromatic Remote-sensing Instrument for Stereo Mapping (PRISM) instrument aboard the Advanced Land Observing Satellite (ALOS) could also be used to calibrate such structure models. PRISM's 2.5 m nadir-looking sensor provides a 70 km swath width, while the forward (+24°) and backward (−24°) sensors provide 35 km swaths. Adding 2.5 m sill (C2.5), variance (V2.5), and standard deviation (S2.5) variables, as well as ratios incorporating five as well as 10 m data, may improve modeling accuracies.

Evaluation of the effects of observation geometry on PLS factor loadings offers the potential to semi-automate mapping of forest structure on a regional scale. If successful, we stand to improve biome-specific parameterization of biophysical processes as they relate to insect disturbance or carbon flux estimates, while also providing improved assessment of the efficacy of management strategies aimed at reducing the periodicity and severity of insect outbreaks, suggested by Blais (1983), or carbon sequestration conservation (Widlowski et al., 2004). This also offers the opportunity to extrapolate limited lidar-based estimates to larger regions covered by multi-scale imagery.

While regional lidar data sets are routinely assembled for terrain mapping purposes, use airborne lidar technology as a sole data source for repeated, regional forest structure estimation may be cost prohibitive. However, lidar data, especially full waveform, could be used to supplement or supplant field measurement data (e.g., CDIA, HT, LC, etc.) for the purpose of calibrating SPOT-based structure models. Lidar returns could be used directly for estimating some variables (e.g., HT and LC), and modeled for others (e.g., CDIA, CC, BA, etc.) where direct estimation is not possible. Many such algorithms exist for estimating tree height and various crown dimensions using lidar data (e.g., Falkowski et al., 2006; Holmgren 2004; Lim et al., 2003a; Maltamo et al., 2006), which can be as accurate as plot-level ground data (Maltamo et al., 2006). PLS regression modeling could also be investigated for this purpose.

In any event, for tree height, lidar sampling density would have to be significantly greater than what is typically collected to increase the probability of capturing points of maximum tree height. Many studies tend to underestimate this parameter due to the high probability that samples miss tree tops (Popescu et al., 2002; Yu et al., 2004; Chen et al., 2006). Naturally, concerns of this nature are greatest for strongly excurrent coniferous canopies where the area of maximum tree height is often quite small (e.g. *Abies balsamifera*).

5. Conclusions

This study examined the potential for modeling and mapping forest biophysical parameters on a regional scale using high-resolution optical satellite imagery, as increasingly detailed forest structure data are needed to support ongoing modeling efforts which strive to understand insect–host dynamics of the spruce budworm in Minnesota and neighboring Ontario. Pixel-wise extraction of local neighborhood statistics from multi-resolution (5 and 10 m) SPOT-5 sensor data, including semivariogram sill ratio information following Song and Woodcock (2002, 2003), provided an efficient and accurate basis for mapping structural properties using PLS regression. While 5 m and 10 m SPOT-5 sill ratio variables (C510X, C510P), and candidate proxies (S510X, S510P, V510X, V510P), were not among the strongest independent predictors of forest structural parameters, they were necessary components in 10 of the 12 final structure models. In addition, we expect the relationship between crown diameter (CDIA) and the sill ratio variables and proxies would improve if solar illumination and sensor look angles favored greater visibility of shaded forest canopy components. Consequently, SVR (Wolter et al., 2008) and MSI (Rock et al., 1986) were frequently of greater importance for predicting forest structure than any of the local variability measures or associated ratios.

In general, models of conifer structure outperformed hardwood models, but both sets of results compare favorably with forest structure studies in which lidar data were used to estimate similar forest parameters. We attribute the greater accuracy of conifer over hardwood models to both greater overall contrast between the sunlit and shaded components of conifer canopies and to stronger correlations among conifer dependent structural variables — especially between CDIA, DBH, HT, and LC. Ultimately, extrapolation of these results outside their calibration area will depend on the availability of imagery which captures similar illumination and sensor geometry. If successful, the implications of this study to forest ecology and management are substantial, as it provides a means to supply sorely needed information into regional ecosystem models, and thus advancing of our knowledge of the linkages between forest structure and ecosystem functioning.

Acknowledgments

This research was supported by funding from the U.S. Forest Service via the National Fire Plan (Project 01.NCS.C.2) and a grant from the USDA Cooperative State Research, Education and Extension Service (CSREES) Managed Ecosystems program (2005-35101-16342), as well as fellowship support from the Department of Forest and Wildlife Ecology at the University of Wisconsin. Thanks to David Helmers and Peter Crump for programming support, to the Natural Resources Research Institute of the University of Minnesota–Duluth for transportation assistance, and to the four anonymous reviewers who's comments greatly improved this paper.

Appendix A

List of abbreviations and definitions.

PLS	Partial least-squares	CDIA	Canopy diameter (m)
PRESS	Predicted residual sum of squares	DBH	Bole diameter at breast height (cm)
D_p	Diameter of a pixel	HT	Tree height (m)
D_o	Diameter of an object	CC	Canopy and/or crown closure (%)
C	Semivariogram sill parameter	LC	Vertical live crown length (m)
Z	Forest stand	BA	Bole basal area (m^2 ha^{-1})
C_z	Sill parameter from forest stand Z	DN	Digital number
XS, X	Multi-spectral	SVR_L	Landsat-based SWIR/visible ratio
PAN, P	Panchromatic	SVR_S	SPOT-based SWIR/visible ratio

References

Alfaro, R. I., Taylor, S., Brown, R. G., & Clowater, J. S. (2001). Susceptibility of northern British Colombia forests to spruce budworm defoliation. *Forest Ecology and Management*, *145*(3), 181–190.

Anderson, J. E., Plourde, L. C., Martin, M. E., Braswell, B. H., Smith, M-L., Dubayah, R. O., Hofton, M. A., & Blair, J. B. (2008). Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest. *Remote Sensing of Environment*, *112*, 1856–1870.

Arenas-Garcia, J., & Camps-Valls, G. (2007). Feature extraction from remote sensing data using Kernel Orthonormalized PLS. *Proceedings of IEEE Geoscience and Remote Sensing Symposium (IGARSS), 23–28 July, Barcelona, Spain* (pp. 258–261).

Asner, G. P., Hicke, J. A., & Lobell, D. B. (2003). Per-pixel analysis of forest structure: Vegetation indices, spectral mixture analysis and canopy reflectance modeling. In M. Wulder & S. E. Franklin (Eds.), *Methods and applications for remote sensing of forests: Concepts and case studies* (pp. 209–254). New York: Kluwer Academic Publishers.

Baker, W. L. (1989). Landscape ecology and nature reserve design in the Boundary Waters Canoe Area, Minnesota. *Ecology*, *70*(1), 23–35.

Batzer, H. O. (1969). Forest character and vulnerability of balsam fir to spruce budworm in Minnesota. *Forest Science*, *15*(1), 17–25.

Bergeron, Y., Leduc, A., Morin, H., & Joyal, C. (1995). Balsam fir mortality following the last spruce budworm outbreak in Northwestern Quebec. *Canadian Journal of Forest Research*, *25*, 1375–1384.

Blais, J. R. (1983). Trends in the frequency, extent, and severity of spruce budworm outbreaks in eastern Canada. *Canadian Journal of Forest Research*, *13*(4), 539–547.

Brockhaus, J. A., & Khorram, S. (1992). A comparison of SPOT and Landsat-TM data for use in conducting inventories of forest resources. *International Journal of Remote Sensing*, *13*(6), 3035–3043.

Chen, Q., Baldocchi, D., Gong, P., & Kelly, M. (2006). Isolating individual trees in a Savanna woodland using small footprint lidar data. *Photogrammetric Engineering and Remote Sensing*, *72*(8), 923–932.

Coburn, C. A., & Roberts, A. C. B. (2004). A multiscale texture analysis procedure for improved forest stand classification. *International Journal of Remote Sensing*, *25*(20), 4287–4308.

Cohen, W. B., Maiersperger, T. K., Gower, S. T., & Turner, D. P. (2003). An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sensing of Environment*, *84*, 561–571.

Cohen, W. B., & Spies, T. A. (1992). Estimating structural attributes of Douglas-fir/western hemlock forest stands from Landsat and SPOT imagery. *Remote Sensing of Environment*, *41*, 1–17.

Cohen, W. B., Spies, T. A., & Bradshaw, G. A. (1990). Semivariograms of digital imagery for analysis of conifer canopy structure. *Remote Sensing of Environment*, *34*, 167–178.

Cohen, W. B., Spies, T. A., & Fiorella, M. (1995). Estimating the age and structure of forests in a multi-ownership landscape of western Oregon, USA. *International Journal of Remote Sensing*, *16*, 721–746.

Coops, N. C., Smith, M. L., Martin, M. E., & Ollinger, S. V. (2003). Prediction of eucalypt foliage nitrogen content from satellite-derived hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(6), 1338–1346.

Crook, G. W., Vezina, P. E., & Hardy, Y. (1979). Susceptibility of balsam fir to spruce budworm defoliation as affected by thinning. *Canadian Journal of Forest Research*, *9*(3), 428–435.

Curran, P. (1988). The semivariogram in remote sensing: An introduction. *Remote Sensing of Environment*, *24*, 493–507.

Curran, P., & Hay, A. (1986). The importance of measurement error for certain procedures in remote sensing at optical wavelengths. *Photogrammetric Engineering and Remote Sensing*, *52*(2), 229–241.

Donoghue, D. N. M., & Watt, P. J. (2006). Using LiDAR to compare forest height estimates from IKONOS and Landsat ETM+ data in Sitka spruce plantation forests. *International Journal of Remote Sensing*, *27*(11), 2161–2175.

Falkowski, M. J., Smith, A. M. S., Hudak, A. T., Gessler, P. E., Vierling, L. A., & Crookston, N. L. (2006). Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of lidar data. *Canadian Journal of Remote Sensing*, *32*(2), 153–161.

Finley, A. O., McRoberts, R. E., & Ek, A. (2006). Applying an efficient k-Nearest Neighbor search to forest attribute imputation. *Forest Science*, *52*(2), 130–135.

Franklin, J. (1986). Thematic mapper analysis of coniferous forest structure and composition. *International Journal of Remote Sensing*, *7*(10), 1287–1301.

Franklin, S. E., Wulder, M. A., & Gerylo, G. R. (2001). Texture analysis of IKONOS panchromatic data for Douglas-fir age separability in British Colombia. *International Journal of Remote Sensing*, *22*(13), 2627–2632.

Forina, M., Casolino, C., & Pizarromillan, C. (1999). Iterative predictor weighting (IPW) PLS: A technique for the elimination of useless predictors in regression problems. *Journal of Chemometrics*, *13*(2), 165–184.

Franco-Lopez, H., Ek, A. R., & Bauer, M. E. (2001). Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote Sensing of Environment*, *77*, 251–274.

Frelich, L. E., & Reich, P. B. (1995). Spatial patterns and succession in a Minnesota southern-boreal forest. *Ecological Monographs*, *65*, 325–346.

Friedman, S. K., & Reich, P. B. (2005). Regional legacies of logging: Departure from presettlement forest conditions in northern Minnesota. *Ecological Applications*, *15*(2), 726–744.

Geisser, S. (1974). A predictive approach to the random effect model. *Biometrica*, *61*, 101–107.

Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, *185*, 1–17.

Ghent, A. W. (1958). Mortality of overstory trembling aspen in relation to outbreaks of the forest tent caterpillar and the spruce budworm. *Ecology*, *39*(2), 222–232.

Greenberg, J. A., Dobrowski, S. Z., & Ustin, S. L. (2005). Shadow allometry: Estimating tree structural parameters using hyperspatial image analysis. *Remote Sensing of Environment*, *97*(1), 15–25.

Grosenbaugh, L. R. (1952). Plotless timber estimates—New, fast, easy. *Journal of Forestry*, *50*, 322–337.

Hall, R. J., Skakun, R. S., Arsenault, E. J., & Case, B. S. (2006). Modeling forest stand structure attributes using Landsat ETM+ data: Application to mapping of aboveground biomass and stand volume. *Forest Ecology and Management*, *225*, 378–390.

Hall, F. G., Townshend, J. R., & Engman, E. T. (1995). Status of remote sensing algorithms for estimation of land surface state parameters. *Remote Sensing of Environment*, *51*(1), 138–156.

Hansen, M. J., Franklin, S. E., Woudsma, C., & Peterson, M. (2001). Forest structure classification in the north Columbia Mountains using the Landsat TM Tasseled Cap Wetness component. *Canadian Journal of Remote Sensing*, *27*(1), 20–32.

Healey, S. P., Cohen, W. B., Yang, Z., & Krankina, O. N. (2005). Comparison of Tassled Cap-based Landsat data structures for use in forest disturbance. *Remote Sensing of Environment*, *97*, 301–310.

Heinselman, M. L. (1973). Fire in the virgin forests of the Boundary Waters Canoe Area, Minnesota. *Quaternary Research*, *3*, 329–382.

Helland, I. S. (1988). On the structure of partial least squares regression. *Communications in Statistics — Simulation and Computation*, *17*(2), 581–607.

Hennigar, C. R., MacLean, D. A., Quiring, D. T., & Kershaw, J. A., Jr. (2008). Differences in spruce budworm defoliation among balsam fir and white, red, and black spruce. *Forest Science*, *54*(2), 158–166.

Holmgren, J. (2004). Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning. *Scandinavian Journal of Forest Research*, *19*, 543–553.

Hudak, A. T., Crookston, N. L., Evans, J. S., Falkowski, M. J., Smith, A. M. S., Gessler, P. E., & Morgan, P. (2006). Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral satellite data. *Canadian Journal of Remote Sensing*, *32*(2), 126–138.

Hudak, A. T., Crookston, N. L., Evans, J. S., Hall, D. E., & Falkowski, M. J. (2008). Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sensing of Environment*, *112*, 2232–2245.

Hyyppä, J., & Inkinen, M. (1999). Detecting and estimating attributes for single trees using laser scanner. *The Photogrammetric Journal of Finland*, *16*, 27–42.

Jarvis, R. M., & Goodacre, R. (2005). Genetic algorithm optimization for pre-processing and variable selection of spectroscopic data. *Bioinformatics*, *21*(7), 860–868.

Jensen, J. L. R., Humes, K. S., Conner, T., Williams, C. J., & DeGroot, J. (2006). Estimation of biophysical characteristics for highly variable mixed-conifer stands using small-footprint lidar. *Canadian Journal of Forest Research*, *36*, 1129–1138.

Jupp, D. L. B., Strahler, A. H., & Woodcock, C. E. (1988). Autocorrelation and regularization in digital images. I. Basic theory. *IEEE Transactions on Geoscience and Remote Sensing*, *26*, 463–473.

Jupp, D. L. B., Strahler, A. H., & Woodcock, C. E. (1989). Autocorrelation and regularization in digital images. II. Simple image models. *IEEE Transactions on Geoscience and Remote Sensing*, *27*, 247–258.

Katila, M., & Tomppo, E. (2001). Selecting estimation parameters for the Finnish multisource National Forest Inventory. *Remote Sensing of Environment*, *76*, 16–32.

Lamonaca, A., Corona, P., & Barbati, A. (2008). Exploring forest structural complexity by multi-scale segmentation of VHR imagery. *Remote Sensing of Environment*, *112*, 2839–2849.

Leardi, R., & Gonzalez, A. L. (1998). Genetic algorithms applied to feature selection in PLS regression: how and when to use them. *Chemometrics and Intelligent Laboratory Systems*, *41*(2), 195–207.

Lefsky, M. A., Cohen, W. B., Acker, S. A., Parker, G. G., Spies, T. A., & Harding, D. (1999). Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sensing of Environment*, *70*, 339–361.

Lefsky, M. A., Harding, D. J., Keller, M., Cohen, W. B., Carabajal, C. C., Espirito-Santo, F. D. B., Hunter, M. O., & de Oliveira, R. (2005). Estimates of forest canopy height and aboveground biomass using ICESat.*Geophysical Research Letters*, *32* art.no.L22 S02.

Leica Geosystems (2006). *Imagine AutoSync*. White paper available at: http://gi.leica-geosystems.com/documents/pdf/IMAGINEAutoSyncWhitePapeFeb06.pdf, 1–20.

Lévesque, J., & King, D. J. (1999). Airborne digital camera image semivariance for evaluation of forest structural damage at an acid mine site. *Remote Sensing of Environment*, *68*, 112–124.

Li, X., & Strahler, A. H. (1985). Geometric-optical modeling of a conifer forest canopy. *IEEE Transactions on Geoscience and Remote Sensing, GE-23*, 705—721.
Li, L., Ustin, S. L., & Riaño, D. (2007). Retrieval of fresh leaf fuel moisture content using Genetic Algorithm Partial Least Squares (GA–PLS) modeling. *IEEE Transactions on Geoscience and Remote Sensing Letters*, *4*(2), 216—220.
Lim, K. S., & Treitz, P. M. (2004). Estimation of above ground forest biomass from airborne discrete return laser scanner date using canopy-based quantile estimators. *Scandinavian Journal of Forest Research*, *19*, 558—570.
Lim, K., Treitz, P., Baldwin, K., Morrison, I., & Green, J. (2003). Lidar remote sensing of biophysical properties of tolerant hardwood forests. *Canadian Journal of Remote Sensing*, *29*(5), 658—678.
Lim, K., Treitz, P., Wulder, M., St-Onge, B., & Flood, M. (2003). Lidar remote sensing of forest structure. *Progress in Physical Geography*, *27*(1), 88—106.
Lindgren, F., Geladi, P., & Wold, S. (1993). The kernel algorithm for PLS. *Journal of Chemometrics*, *7*(1), 45—59.
Lu, D., Mausel, P., Brondízioc, E., & Moran, E. (2004). Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. *Forest Ecology and Management*, *198*(1–3), 149—167.
Magnussen, S., Boudewyn, P., & Alfaro, R. (2004). Spatial prediction of the onset of spruce budworm defoliation. *Forestry Chronicle*, *80*(4), 485—494.
Maltamo, M., Hyyppa, J., & Malinen, J. (2006). A comparative study of the use of laser scanner data and field measurements in prediction of crown height in boreal forests. *Scandinavian Journal of Forest Research*, *21*, 231—238.
McDonald, S., Niemann, K.O., Goodenough, D.G., Dyk, A., West, C., Han, T., & Murdoch, M. (2003). Hyperspectral Remote Sensing of Conifer Chemistry and Moisture. Proceedings of IEEE Geoscience and Remote Sensing Symposium (pp. 1, 552–554), 21–25 July, Victoria University, British Columbia, Canada.
Mc Elhinny, C., Gibbons, P., Brack, C., & Bauhus, J. (2005). Forest and woodland stand structural complexity: Its definition and measurement. *Forest Ecology and Management*, *218*, 1—24.
McRoberts, R. E. (2008). Using satellite imagery and the k-nearest neighbors technique as a bridge between strategic and management forest inventories. *Remote Sensing of Environment*, *112*, 2212—2221.
McRoberts, R. E. (2009). A two-step nearest neighbors algorithm using satellite imagery for predicting forest structure within species composition classes. *Remote Sensing of Environment*, *113*, 532—545.
McRoberts, R. E., Tomppo, E. O., Finley, A. O., & Heikkinen, J. (2007). Estimating areal means and variances of forest attributes using the k-Nearest Neighbors technique and satellite imagery. *Remote Sensing of Environment*, *111*, 466—480.
Meng, Q., Cieszewski, C. J., Madden, M., & Borders, B. E. (2009). K Nearest Neighbor method for forest inventory using remote sensing data. *GIScience & Remote Sensing*, *44*(2), 149—165.
Mladenoff, D. J., & He, H. S. (1999). Design, behavior and application of LANDIS, an object oriented model of forest landscape disturbance and succession. In D. J. Mladenoff & W. L. Baker (Eds.), *Spatial modeling of forest landscape change: Approaches and applications* (pp. 125—162). Cambridge, UK: Cambridge University Press.
Muukkonen, P., & Heiskanen, J. (2005). Estimating biomass for boreal forests using ASTER satellite data combined with standwise forest inventory data. *Remote Sensing of Environment*, *99*, 434—447.
Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, *80*, 88—99.
Norgaard, L., Saudland, A., Wagner, J., Nielsen, J. P., Munck, L., & Engelsen, S. B. (2000). Interval (iPLS): A comparative chemometric study with an example from near-IR spectroscopy. *Applied Spectroscopy*, *54*(3), 413—419.
Ourcival, J. M., Joffre, R., & Rambal, S. (1999). Exploring the relationships between reflectance and anatomical and biochemical properties in *Quercus ilex* leaves. *New Phytologist*, *143*(2), 351—364.
Pastor, J., Sharp, A., & Wolter, P. (2005). An application of Markov models to the dynamics of Minnesota's forests. *Canadian Journal of Forest Research*, *35*, 3011—3019.
Popescu, S. C., Wynne, R. H., & Nelsen, R. F. (2002). Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. *Computers and Electronics in Agriculture*, *37*, 71—95.
Popescu, S. C., Wynne, R. H., & Nelson, R. F. (2003). Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing*, *29*(5), 564—577.
Pulliainen, J., Engdahl, M., & Hallikainen, M. (2003). Feasibility of multi-temporal interferometric SAR data for stand-level estimation of boreal forest stem volume. *Remote Sensing of Environment*, *85*, 397—409.
Rauste, Y. (2005). Multi-temporal JERS SAR data in boreal forest biomass mapping. *Remote Sensing of Environment*, *97*, 263—275.
Reese, H. M., Lillesand, T., Nagel, D. E., Stewart, J. S., Goldmann, R. A., Simmons, T. E., Chipman, J. W., & Tessar, P. A. (2002). Statewide land cover derived from multiseasonal Landstat TM data—A retrospective of the WISCLAND project. *Remote Sensing of Environment*, *82*, 224—237.
Rock, B. N., Vogelmann, J. E., Williams, D. L., Vogelmann, A. F., & Hoshizaki, T. (1986). Remote detection of forest damage. *BioScience*, *36*(7), 439—445.
Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *Proceedings of the 3rd ERTS Symposium* (pp. 1, 309—317). NASA SP-351, Washington D.C.: U.S. Government Printing Office.
Santoro, M., Askne, J., Smith, G., & Fransson, J. E. S. (2002). Stem volume retrieval in boreal forests from ERS-1/2 interferometry. *Remote Sensing of Environment*, *81*, 19—35.
SAS Institute Inc. (2000). *SAS Users Guide, Version 8* (pp. 2693—2734). Cary, North Carolina: SAS Institute Incorporated.
Scheller, R. M., Domingo, J. B., Sturtevant, B. R., Williams, A. R., Gustafson, E. J., & Mladenoff, D. J. (2007). Design, development, and application of LANDIS-II, a spatial landscape simulation model with flexible temporal and spatial resolution. *Ecological Modelling*, *201*, 409—419.
Scheller, R. M., & Mladenoff, D. J. (2004). A forest biomass module for a landscape simulation model, LANDIS: design, validation, and application. *Ecological Modelling*, *180*, 211—229.
Schumacher, S., Bugmann, H., & Mladenoff, D. J. (2004). Improving the formulation of tree growth and succession in a spatially explicit landscape model. *Ecological Modelling*, *180*, 175—194.
Shen, S. S., Badhwar, G. D., & Carnes, J. G. (1985). Separability of boreal forest species in the Lake Jennette area, Minnesota. *Photogrammetric Engineering and Remote Sensing*, *51*(11), 1775—1783.
Shugart, H. H., Bourgeau-Chavez, L., & Kasischke, E. S. (2000). Determination of stand properties in boreal and temperate forests using high-resolution imagery. *Forest Science*, *46*(4), 478—486.
Smith, M. L., Martin, M. E., Plourde, L., & Ollinger, S. V. (2003). Analysis of hyperspectral data for estimation of temperate forest canopy nitrogen concentration: Comparison between an airborne (AVIRIS) and a spaceborne (Hyperion) sensor. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(6), 1332—1337.
Smith, M. L., Ollinger, S., Martin, M., Aber, J., Hallett, R., & Goodale, C. (2002). Direct estimation of aboveground forest productivity through hyperspectral remote sensing of canopy nitrogen. *Ecological Applications*, *12*, 1286—1302.
Song, C. (2007). Estimating tree crown size with spatial information of high resolution optical remotely sensed imagery. *International Journal of Remote Sensing*, *28*(15), 3305—3322.
Song, C., & Woodcock, C. E. (2002). The spatial manifestation of forest succession in optical imagery: The potential of multi-resolution imagery. *Remote Sensing of Environment*, *82*, 271—284.
Song, C., & Woodcock, C. E. (2003). Estimating tree crown size from multiresolution remotely sensed imagery. *Photogrammetric Engineering and Remote Sensing*, *69*(11), 1263—1270.
Spiegelman, C. H., McShane, M. J., Goetz, M. J., Motamedi, M., Yue, Q. L., & Coté, G. L. (1998). Theoretical justification of wavelength selection in PLS calibration: Development of a new algorithm. *Analytical Chemistry*, *70*(1), 35—44.
Stone, M. (1974). Cross-validitory choice and assessment of statistical predictions. *Journal of Royal Statistical Society*, *36*, 111—133.
St-Onge, B. A., & Cavayas, F. (1995). Estimating forest stand structure from high resolution imagery using the directional variogram. *International Journal of Remote Sensing*, *16*(11), 1999—2021.
Strahler, A. H., Woodcock, C. E., & Smith, J. A. (1986). On the nature of models in remote sensing. *Remote Sensing of Environment*, *20*, 121—139.
Sturtevant, B. R., Gustafson, E. J., Li, W., & He, H. S. (2004). Modeling biological disturbances in LANDIS: a module description and demonstration using spruce budworm. *Ecological Modelling*, *180*(1), 153—174.
Swierenga, H., de Groot, P. J., de Weijer, A. P., Derksen, M. W. J., & Buydens, L. M. C. (1998). Improvement of PLS model transferability by robust wavelength selection. *Chemometrics and Intelligent Laboratory Systems*, *41*(2), 237—248.
Tomppo, E. O., Gagliano, C., De Natale, F., Katila, M., & McRoberts, R. E. (2009). Predicting categorical variables using an improved k-Nearest Neighbors estimator. *Remote Sensing of Environment*, *113*, 500—517.
Townsend, P. A., Foster, J. R., Chastain, R. A., & Currie, W. S. (2003). Application of imaging spectroscopy to mapping canopy nitrogen in the forests of the central Appalachian Mountains using Hyperion and AVIRIS. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(6), 1347—1354.
Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*, 127—150.
Vogelmann, J. E., & Rock, B. N. (1988). Assessing forest damage in high-elevation coniferous forests in Vermont and New Hampshire using Thematic Mapper data. *Remote Sensing of Environment*, *24*, 227—246.
Vogelmann, J. E., & Rock, B. N. (1989). Use of Thematic Mapper data for the detection of damage caused by the pear thrips. *Remote Sensing of Environment*, *30*(3), 217—225.
Walker, W. S., Kellndorfer, J. M., LaPoint, E., Hoppus, M., & Westfal, J. (2007). An empirical InSAR-optical fusion approach to mapping vegetation canopy height. *Remote Sensing of Environment*, *109*, 482—499.
Widlowski, J. L., Pinty, B., Gobron, N., Verstraete, M. M., Diner, D. J., & Davis, A. B. (2004). Canopy structure parameters derived from multi-angular remote sensing data for terrestrial carbon studies. *Climatic Change*, *67*, 403—415.
Wold, S., Ruhe, A., Wold, H., & Dunn, W. J. (1984). The collinearity problem in linear regression. The Partial Least Squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing*, *5*(3), 735—743.
Wolter, P. T., Mladenoff, D. J., Host, G. E., & Crow, T. R. (1995). Improved forest classification in the northern Lake State using multi-temporal Landsat imagery. *Photogrammetric Engineering and Remote Sensing*, *61*(9), 1129—1143.
Wolter, P. T., Townsend, P. A., Kingdon, C. C., & Sturtevant, B. R. (2008). Remote sensing of the distribution and abundance of host species for spruce budworm in northern Minnesota and Ontario. *Remote Sensing of Environment*, *112*, 3971—3982.
Wolter, P. T., & White, M. A. (2002). Recent forest cover type transitions and landscape structural changes in northeast Minnesota. *Landscape Ecology*, *17*, 133—155.
Woodcock, C. E., Collins, J. B., Jakabhazy, V. D., Li, X., Macomber, S. A., & Wu, Y. (1997). Inversion of the Li–Strahler canopy reflectance model for mapping forest structure. *IEEE Transactions on Geoscience and Remote Sensing*, *35*, 405—414.
Woodcock, C. E., & Strahler, A. H. (1987). Factors of scale in remote sensing. *Remote Sensing of Environment*, *21*, 311—332.
Woodcock, C. E., Strahler, A. H., & Jupp, D. L. B. (1988). The use of variograms in remote sensing: II. real digital images. *Remote Sensing of Environment*, *25*, 349—379.

Wulder, M. A., Han, T., White, J. C., Sweda, T., & Tsuzuki, H. (2007). Integrating profiling LIDAR with Landsat data for regional boreal forest canopy attribute estimation and change characterization. *Remote Sensing of Environment*, *110*, 123−137.

Wulder, M. A., White, J. C., Hay, G. J., & Castilla, G. (2008). Towards automated segmentation of forest inventory polygons on high spatial resolution satellite imagery. *Forestry Chronicle*, *84*(2), 221−230.

Yu, X., Hyyppa, J., Kaartinen, H., & Maltamo, M. (2004). Automatic detection of harvested trees and determination of forest growth using airborne laser scanning. *Remote Sensing of Environment*, *90*, 451−462.

Zhang, Q., & Alfaro, R. I. (2001). Periodicity of two-year cycle spruce budworm outbreaks in central British Columbia: A dendro-ecological analysis. *Forest Science*, *48*(4), 722−731.

Zheng, D., Heath, L. S., & Ducey, M. J. (2008). Satellite detection of land-use change and effects on regional forest aboveground biomass estimates. *Environmental Monitoring and Assessment*, *144*, 67−70.

Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M., Le Moine, J., & Ryu, S. (2004). Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. *Remote Sensing Environment*, *93*, 402−411.