

Journal of Terrestrial Observation

Volume 2, Issue 1

Winter 2010

Article 5

Assessing Tree Cover in Agricultural Landscapes Using High-Resolution Aerial Imagery

Greg C. Liknes, Charles H. Perry, and Dacia M. Meneguzzo

Assessing Tree Cover in Agricultural Landscapes Using High-Resolution Aerial Imagery

Greg C. Liknes, Charles H. Perry, and Dacia M. Meneguzzo

USDA Forest Service

ABSTRACT

Trees used in agroforestry practices, such as windbreaks, provide a variety of ecosystem benefits and are recognized globally as an important land use. However, efforts to inventory and monitor agroforestry land use have been sporadic, short-lived, or focused on small spatial extents. There are a variety of satellite-derived datasets that provide information about tree cover over broad spatial extents, but most are based on satellite sensors with resolutions too coarse to accurately observe narrow plantings of trees. We derived area estimates of land with tree cover in North Dakota and South Dakota from the National Land Cover Dataset, the Cropland Data Layer, MODIS Vegetative Continuous Fields, and a MODIS land cover product. We compared these image-based estimates to estimates based on *in situ* observations of forest land from the USDA Forest Service's Forest Inventory and Analysis (FIA) program. Satellite-derived estimates of tree cover area differed from FIA forest land estimates by as much as 200,000 ha in both North Dakota and South Dakota. Image data from high resolution satellite sensors can detect small or narrow features, but prohibitively high data costs prevent their use for conducting national inventories. We used freely available, 1-m resolution imagery from the National Agriculture Imagery Program (NAIP) to map tree cover in Pembina County, North Dakota, USA. The approach used image segmentation and Random Forests, an ensemble classification tree algorithm. The Random Forests approach to mapping tree cover resulted in 84.8% agreement between model predictions and the out-of-bag sample. Based on the Gini index, texture attributes were more important predictors of tree cover than spatial or spectral attributes. Variability between flight lines in the NAIP imagery led to over-prediction of tree canopy in particular north/south swaths in the county. While future evaluation is required to develop an optimal training dataset to assess tree cover, the procedure shows promise for application over a broad spatial extent.

ACKNOWLEDGMENTS

We thank Susan Crocker, Dr. Randy Hamilton, and Dr. Kathleen Ward for thoughtful reviews of the early draft of this manuscript. We greatly appreciate the assistance with image interpretation by Cassandra Olson and Paul Sowers.

The Journal of Terrestrial Observation | Volume 2 Number 1 (Winter 2010)

KEYWORDS

image segmentation, Random Forests, NAIP, agroforestry

INTRODUCTION

Trees used in agroforestry practices provide a variety of ecosystem benefits. Windbreaks and shelterbelts reduce wind erosion from crop fields and shield buildings from wind, extend the life of structures, and lower heating costs. Windbreaks can also serve as living snow fences, enhancing wintertime accessibility for landowners and increasing springtime moisture availability for crops. Windbreaks also reduce pesticide drift (Ucar and Hall, 2001). In Florida, tree hedgerows are suggested as a mechanism for enhancing the pollination services of bees (Albrigo and Russ, 2002). Carroll et al. (2004) found that shelterbelts have the potential to lessen flood risk. Under a climate warming scenario, shelterbelts may have an evaporative cooling effect and could help maintain a viable growing season for maize (Easterling et al., 1997). Carucci (2000) determined shelterbelts are an effective way to prevent further desertification in the African Sahel. Windbreaks and shelterbelts provide habitat and travel corridors for a variety of wildlife (Rosenberg et al., 1997).

The use of trees in agroforestry practices also sequesters carbon; linear plantings of some tree species were found to sequester more than 100 metric tons per kilometer in the Canadian Prairie Provinces (Kort and Turnock, 1998). In a valuation of ecosystem services conducted in Canada, carbon sequestration accounted for more than 50% of the value provided by agroforestry (Kulshreshtha and Kort, 2009). With all of these benefits derived from agroforestry practices, there is global recognition of the importance of tracking the extent and condition of trees outside the traditional definition of forests.

There is a modicum of activity focused on the inventory of trees in agroforestry settings. The World Agroforestry Centre (formerly known as the International Centre for Research in Agroforestry) conducted a global inventory of agroforestry practices using a questionnaire in the years from 1982 to 1987 (Oduol et al., 1988). The Global Forest Resource Assessment explicitly defined Trees Outside Forests (TOF) as a category of interest that includes some agroforestry practices (FAO, 2000). An inventory of TOF in Kenya was conducted in the 1990s (Holmgren et al., 1994). India included TOF in its forest monitoring program and used a remote sensing approach to assess their extent (Rawat et al., 2004). In Manitoba, Canada, the Prairie Shelterbelt Program provides millions of tree seedlings to landowners each year and has begun to monitor the extent of agricultural plantings using high-resolution imagery (Wiseman et al., 2008).

In the United States, TOF or trees used in agroforestry are not explicitly inventoried or monitored. Several programs peripherally address the question, but the information they yield is either incomplete or lacks precision and accuracy specifically for TOF. The U.S. Department of Agriculture (USDA) Natural Resource Conservation Service (NRCS) conducts a National Resources Inventory (NRI) on nonfederal lands. Information on windbreaks has been collected in past NRI in-

ventories (Goebel, 1998). Currently, NRI reports areas of Other Rural Land, which includes agroforestry practices, but does not explicitly separate them from farmsteads, farm structures, barren land, and marshland. The USDA National Agriculture Statistics Service (NASS) tracks the area of land used for various agricultural commodities. In support of that effort NASS produces the geospatial Cropland Data Layer¹ (CDL) primarily using Indian Remote Sensing RESOURCESAT-1 (IRS-P6) Advanced Wide Field Sensor (AWiFS) satellite data. The CDL includes forest categories from the National Land Cover Dataset (NLCD) (Homer et al., 2007), a woodland category, Christmas tree plantations, and orchards, but it does not explicitly define any other agroforestry practices.

The USDA Forest Service's Forest Inventory and Analysis (FIA) program conducts annual inventories of forest resources on both public and private forest land. The program only records tree measurements on forest land, or land that is occupied by trees and meets minimum area and width requirements (0.4 ha and 36.6 m). Because of this definition of forest, small areas or narrow strips of trees, such as windbreaks, shelterbelts, or riparian corridors, are considered non-forest and therefore are not inventoried by FIA. Using interpretation of aerial imagery, Perry et al. (2009) estimated that treed lands in North Dakota and South Dakota are underestimated by 38% and 30%, respectively, because the FIA definition of forest does not include these agroforestry practices. Other studies conducted in the United States give us an idea of the extent of non-forest trees in states where agriculture is the predominant land use. Hartong and Moessner (1956) estimated reported timberland area in Iowa would be 25% higher if non-forest trees were inventoried. Hansen (1985) used line-intersect sampling and aerial photography to inventory trees in Kansas and found 136,000 ha of wooded strips (compared to approximately 550,000 ha of forest land reported in FIA's 1981 Kansas inventory).

The Great Plains Tree and Forest Invasives Initiative (hereafter referred to as GPI) is a cooperative project of the USDA Forest Service and the state forestry agencies in North Dakota, South Dakota, Nebraska, and Kansas. While this project is intended to prepare Great Plains states for the potential arrival of invasive pests, the combination of photo-interpreted and field data collected on tree resources in the region could provide a useful baseline of information on non-forest trees in agricultural landscapes (Lister et al., 2009). Photo-interpreted points and data from GPI fixed-radius circular plots can be aggregated to produce statistically precise estimates for large areas, such as counties. Data from field plots provide information on tree species, volume, and condition, but do not provide the same spatially explicit information that can be derived from remote-sensing approaches, such as length and width of linear tree plantings.

Satellite-derived data products are attractive for monitoring because of the synoptic view of the landscape they provide. A host of products is available for the conterminous United States that address tree cover, but each has limited utility with regard to narrow linear plantings, or sparse cover, such as pasture or rangeland with trees. NLCD 2001 is based on 30-m resolution Landsat data and does not ex-

PLICITLY map agroforestry land use. The CDL is derived from either Landsat (30-m) or AWiFS (56-m) and, as previously mentioned, does not include windbreaks and shelterbelts as a land use category. The Vegetative Continuous Fields (VCF) data product includes a per-pixel percent tree cover estimate (Hansen et al., 2003). Narrow tree plantings may appear as pixels with very low tree cover because VCF is derived from 500-m Moderate-resolution Imaging Spectroradiometer (MODIS) satellite data. The tradeoff of broad-scale coverage provided by MODIS VCF or NLCD is the pixel size or minimum mapping unit is frequently too large to effectively capture narrow plantings of trees.

The resolution of imagery selected for monitoring should be appropriate to the features to be observed. O'Neill et al. (1996) recommend the grain size for map elements be one-fifth to one-half the size of the features of interest. Woodcock and Strahler (1987) suggest an image resolution of one-half to three-fourths the size of target objects. In order to monitor tree plantings at FIA's forest width requirement (36.6 m), imagery from 7 m to 27 m should be used, with finer resolutions needed for narrower windbreaks. There are many examples of high-resolution imagery used in natural resource monitoring applications of small targets. Laliberte et al. (2004) used QuickBird imagery (61-cm panchromatic and 2.4-m multispectral) to assess shrub encroachment in southern New Mexico. Wiseman et al. (2008) used 62.5 cm resolution imagery to identify shelterbelts and their component tree species in Manitoba, Canada. While these methods were highly effective, they were applied to relatively small areas. However, the approaches show promise for application over broader regions because of the potential to automate parts of the methodology.

Individual landowners may be well aware of the location and condition of agroforestry plantings on their property, yet a coordinated, broad-scale accounting of these trees would be a valuable strategic planning tool for the management of carbon and other ecosystem benefits derived from agroforestry. From a large-scale inventory perspective, assessing TOF or tree cover using high-resolution imagery requires imagery with extensive coverage. The USDA National Agriculture Imagery Program (NAIP) has acquired aerial imagery at 1- and 2-m resolution for much of the United States. The NAIP Program began as a pilot program in 2001 and has been acquiring imagery since 2003 for a multistate area. Imagery is collected during the agricultural growing season in natural color, with an option to add near-infrared information if additional funding is available. For example, color-infrared NAIP imagery was collected for eight states in 2008. NAIP imagery is collected at 1-m resolution, and has been collected in some past years at 2-m resolution. Images can contain up to 10% cloud cover. Data are made available as either compressed county mosaics or uncompressed 3.75 minute by 3.75 minute quarter quadrangles with a 300-m buffer on all sides.

Although NAIP imagery is available for a broad spatial extent, there are some problems to overcome in order to use the data for monitoring trees used in agroforestry. NAIP imagery has a lack of radiometric consistency between flight lines that poses problems for automated processing over large areas. Additionally, availabil-

ity of a near-infrared band (useful for discriminating vegetation from other land cover) is limited to only a few states. However, we hypothesized the texture contrast between trees and their adjacent land uses in the Great Plains would allow us to accurately separate tree cover from the surrounding agricultural landscape.

Using satellite-derived data products and aerial imagery, we set out to determine area estimates of tree cover in the northern Great Plains in the United States. First, we compared estimates of tree cover for North and South Dakota derived from NLCD, CDL, MODIS VCF, MODIS MOD12Q1 land cover product, and FIA. Second, we developed a method for mapping tree cover from widely available, high-resolution NAIP imagery using image segmentation in conjunction with a data mining approach. Because our objective was to develop a procedure that is feasible over a broad spatial scale, image segmentation and model development took speed and simplicity into account, as well as accuracy.

METHODS

Study Area

The focus of our first objective was the Great Plains states of North Dakota and South Dakota. These states lie primarily within the Level II west-central semi-arid prairies and temperate prairies ecoregions (Figure 1a). To address our second objective, we developed a map of tree cover for Pembina County, North Dakota, which shares a border with Canada (Figure 1a). The region experiences a wide annual variation in temperature, with average monthly temperatures ranging from approximately -13°C in January to 22°C in July. The regional climate is also characterized by moderate precipitation, periodic drought, and high winds. Naturally occurring tree cover is sparse, and land use is dominated by row-crop agriculture and rangeland grazing. Pembina County has approximately 24,600 ha of forestland meeting the FIA definition of forest land (Miles, 2009). Most of this forest occurs in riparian corridors along the Red River of the North and along the Pembina and Tongue rivers. Of the 289,755 ha in Pembina County, 90% are used for farming (USDA National Agricultural Statistics Service, 2007).

Statewide estimates of tree cover

We derived estimates of tree cover from four satellite-derived datasets for the states of North Dakota and South Dakota. Specifically, the area of forest- or tree-related land cover was calculated from each dataset by summing the area of pixels in representative categories. We used the NLCD 2001, NASS CDL, MODIS VCF, and the MODIS MOD12Q1 land cover product² at 30-m, 56-m, 500-m, and 1-km resolutions, respectively. A threshold of 25% tree cover for the MODIS VCF was selected to separate forested pixels from non-forest pixels. This threshold was chosen because it creates a match between nationwide VCF forest area estimates and those of the

Resources Planning Act assessments (Nelson, 2005). For context, we also report FIA estimates of forest land derived from *in situ* measurements from the years 2003 to 2007. Details of FIA sampling scheme and estimation procedures can be found in Bechtold and Patterson (2005).

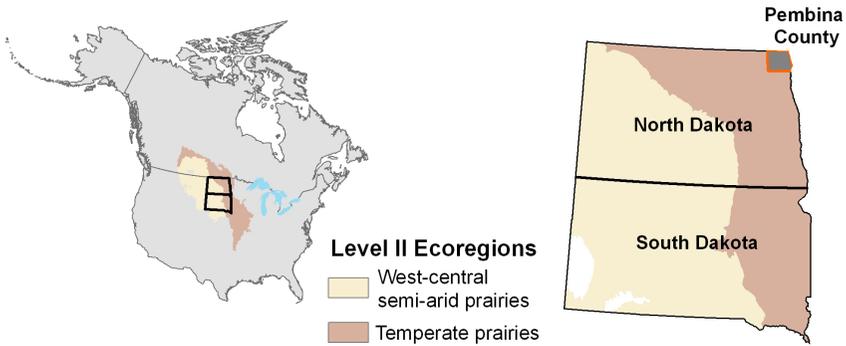


Figure 1a. Location of North Dakota and South Dakota relative to the west-central semi-arid prairies and temperate prairies (left). Pembina County is located in the northeast corner of North Dakota (right).



Figure 1b. 2003 NAIP imagery for Pembina County, North Dakota (left). Arrangement of NAIP image quarter quadrangles (QQs) for Pembina County, North Dakota (right). The QQ with diagonal hatching was used to develop the predictive model for the rest of the QQs in the flight path.

ESTIMATING TREE COVER FROM IMAGE SEGMENTATION OF HIGH-RESOLUTION IMAGERY

Image data

Because our objective was to develop a feasible method for estimating tree cover over a broad area, we elected to work with NAIP imagery (Figure 1b). For Pembina County, current 1-m NAIP imagery was collected in 2003 and 2005. After examination of both datasets, the 2003 imagery was selected because of a more natural appearance and better color contrast between tree cover and agricultural land use

(Figure 2). We elected to work with uncompressed quarter quadrangles (QQs) in GeoTIFF format because of the compatibility with the image segmentation software. The imagery was collected along north/south flight paths, and there are visible radiometric differences across the county. The imagery has three visible bands (red, green, and blue).

Image segmentation

Image segmentation was implemented using Definiens Developer Professional v. 7 (Definiens AG, 2008). Fine-scale (small area) segments were created during a first pass to allow the local variability in areas with tree cover to be captured in separate segments (segmentation parameters: scale = 30; shape = 0.1; compactness = 0.9). Adjacent segments were iteratively merged based on their standard deviation and brightness using a range of segmentation parameters. For example, the Definiens scale parameter was varied between 30 and 300. This process resulted in larger segments in agricultural fields that are brighter and visually homogeneous and smaller segments in areas with tree cover where the visual appearance is darker and more heterogeneous. In order to adhere to our objective of developing a process that is feasible for application over a large spatial extent, a balance was struck between creating segments that perfectly captured the variation in the image and minimizing the time required to create the segments. Once the segmentation process was developed using a small test area, it was applied to all quarter quadrangle images that intersected the county (Figure 1b). The resultant segmentation dataset contained 384,520 segments with a range of 1,416 to 8,587 segments per quarter quadrangle. The first two panels in Figure 3 show a representative image in Pembina County and the resulting image segments.

Predicting presence/absence of tree cover

All image segments from a single QQ (Figure 1b) were labeled via image interpretation as one of three categories: tree cover, no tree cover, and mixed. A reference dataset with 3,554 labeled image segments was created, representing less than 1% of the Pembina County area. In Table 1, information about the reference dataset is presented, showing the breakdown of the three categories with respect to the number of image segments and the percentage of area in the sample.



Figure 2. Samples of NAIP imagery collected in North Dakota in 2005 (left) and 2003 (right).

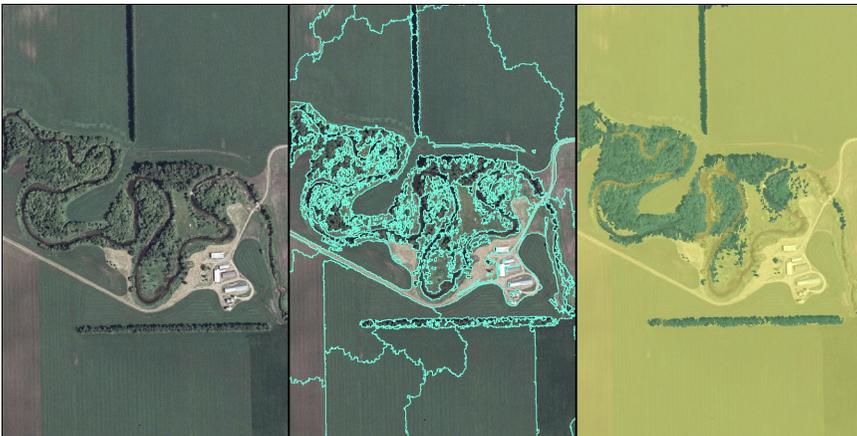


Figure 3. Illustration of the image segmentation/classification workflow. The image on the left is a 2003 NAIP sample image from Pembina County, North Dakota. The middle image shows the image segmentation borders in light blue for the same image. In the right image, segments with predicted tree cover are fully transparent, while areas without tree cover are covered by an opaque, yellow mask.

Table 1: Image segment properties by class label for a reference dataset in Pembina County, North Dakota, derived from 2003 1-m NAIP imagery and visual image interpretation.

Segment property	Tree cover	No tree cover	Mixed
Number of segments	2,010	1,305	239
Area (%)	7	92	1
Mean area (ha)	1.5	29.4	1.9
Median area (ha)	0.09	4.2	0.07

In addition to creating image segments, Definiens Developer software calculated a series of spectral, spatial, and textural attributes for every image segment. Table 2 provides a complete listing of the attributes used in this study. While vegetation indices or band ratios are frequently used in vegetation classification, many effective indices require a near-infrared band (Bannari et al., 1995). Because only visible bands were available, we elected not to use band ratios.

The gray-level co-occurrence matrix (GLCM) and gray-level difference vector (GLDV) attributes are texture measures based on the work of Haralick et al. (1973). Whereas standard deviation gives an idea of the overall variability across an image segment, GLCM and GLDV texture measures provide more information about the spectral differences between neighboring pixels. Attribute definitions and details of how the texture measures are calculated in the software can be found in the Reference Book (Definiens AG, 2007). Because each band of the NAIP imagery contains 8-bit data, the “quick 8/11” version of GLCM/GLDV texture calculations was used. Several of the Haralick texture measures are strongly correlated; therefore, a representative subset of available GLCM/GLDV measures was selected for inclusion in the predictor dataset.

In order to predict the presence of tree cover based on the attributes associated with each image segment, Random Forests³ was used. Random Forests (Breiman, 2001) is an extension of earlier work on Classification and Regression Trees® (CART®) (Breiman et al., 1984). Random Forests builds a series of classification trees, withholding predictor variables and observations for each tree. We used the Random Forests algorithm (RF) in the freely available R statistical computing platform⁴ (specifically, the randomForest package⁵). RF provides advantages that were desirable for our objectives. Specifically, the algorithm can handle many predictor variables simultaneously and can provide relative measures of importance for predictor variables. Additionally, RF provides several useful diagnostics to assess its performance. Because observations are withheld during the tree building process, a sample exists from which classification accuracy can be assessed. This is referred to as the out-of-bag (OOB) sample.

Using the reference dataset of 3,554 image segments, a model was developed using RF to classify segments into three classes: tree cover, no tree cover, and mixed. Class predictions using RF were based on an ensemble of classification trees with final class assignments determined by a plurality of predictions across trees. Therefore, RF can also assign a probability to each prediction. This was particularly use-

ful in the overlap region between quarter quadrangles where mean probability was used to determine class assignments.

During this model development phase, an attempt was made to arrive at a set of predictor variables that was a good compromise between classification accuracy and conceptual simplicity. This was necessary because each additional variable adds to the time required to calculate spatial, spectral, and textural attributes for the image segments and our objective is to develop a procedure that can be implemented quickly and simply over a broad area.

Table 2: Image segment attributes used to develop a predictive model of tree cover in Pembina County, North Dakota, using NAIP imagery.

Spectral attributes	Short name(s)
Brightness ^{i,a}	Bright, Mn_I1, Mn_I2, Mn_I3
Contrast ⁱ	Cn_nl1, Cn_nl2, Cn_nl3
mean difference to neighboring segment	Md_nl1, Md_nl2, Md_I3
mean difference to the scene ⁱ	Md_snl1, Md_snl2, Md_snl3
minimum pixel brightness ⁱ	Min_I1, Min_I2, Min_I3
maximum pixel brightness ⁱ	Max_I1, Max_I2, Max_I3
maximum difference ^a	Max_diff
standard deviation ⁱ	Sd_I1, Sd_I2, Sd_I3
Spatial attributes	
area	Area
asymmetry	Asymm
border index	Bord_ind
border length	Bord_len
compactness	Compact
density	Density
elliptical fit	Elipfit
length	Length
length/width ratio	Lwratio
main direction	Main_dir
radius of smallest enclosing ellipse	Radelips
radius of largest enclosed ellipse	Radelipl
rectangular fit	Rectfit
roundness	Round
shape index	Shpind
width	Width
Haralick texture attributes^{i,a}	
GLCM Angular 2 nd Moment	Glc_a2, _a2l1, _a2l2, _a2l3
GLCM Entropy	Glc_e, _el1, _el2, _el3
GLCM Homogeneity	Glc_h, _h1, _h2, _h3
GLCM Mean	Glc_m, _m1, _m2, _m3
GLCM Standard Deviation	Glc_s, _s1, _s2, _s3
GLDV Angular 2nd Moment	Gld_a2, _a2l1, _a2l2, _a2l3
GLDV Entropy	Gld_e, _el1, _el2, _el3
GLDV Mean	Gld_m, _ml1, _ml2, _ml3
GLDV Contrast	Gld_c, _cl1, _cl2, _cl3

ⁱ attribute was calculated for each band individually

^a attribute was calculated using all bands

I1, I2, I3 represent the attributes for layers 1, 2, and 3 (red, green, and blue), respectively

Note: While standard deviation can be considered a measure of texture, it is grouped here with spectral attributes, as it is an indication of the spectral variability in an image segment. Texture, in this instance, is reserved for GLCM and GLDV measures.

RESULTS

Statewide estimates

Satellite-derived areal estimates of tree cover varied widely within both North Dakota and South Dakota (Table 3). In North Dakota, the estimates of forest cover from the satellite-derived products were substantially higher than the FIA estimate (as much as 73%), with the exception of the MODIS-derived land cover product, MOD12Q1, which was 67% lower. In South Dakota, satellite-derived estimates matched the FIA estimate much more closely, ranging from the MODIS VCF's 29% underestimate to the Cropland Data Layer's 12% overestimate. However, the largest absolute difference was similar in both states, approximately 200,000 ha.

Table 3: Statewide estimates of land with tree cover. Estimates are from satellite-derived land cover products and Forest Inventory and Analysis data.

	North Dakota area (ha)	South Dakota area (ha)
NLCD 2001 ^a	451,285	807,048
CDL ^b	490,318	814,405
MODIS VCF ^c	389,975	517,723
MODIS MOD12Q1 ^d	94,400	725,897
FIA ^e	283,375	724,478

The following land use/land cover categories were used to derive the area estimates:

NLCD 2001	Deciduous, coniferous, and mixed forests; woody wetlands
Cropland Data Layer	NLCD forest (as above) and woodland
MODIS VCF	Percent tree cover ≥ 25
MODIS MOD12Q1	Evergreen, deciduous, and mixed forests; closed shrublands woody savannas
FIA	Areas that are at least 10% stocked with trees, 0.4 ha in area, and 36.6 m wide

^a National Land Cover Dataset 2001

^b Cropland Data Layer

^c MODIS Vegetative Continuous Fields

^d MODIS land cover product

^e Forest Inventory and Analysis

Pembina County estimate

We begin with an exploration of the performance of the RF model. According to the OOB sample, 3,105 image segments out of 3,554 were correctly assigned to either tree cover, no tree cover, or mixed—an overall agreement of 84.8% (Table 4). Agreement for the tree cover class was 95.9%, whereas the highest error rate occurred in the mixed category, for which segments were frequently mislabeled as tree cover. These commission errors would lead to an overestimate of tree cover, with an entire image segment assigned to the tree cover class while only a very small portion of the segment may be tree-covered.

Table 4: Confusion matrix for a Random Forests model used to predict tree cover classes in Pembina County, North Dakota, based on the out-of-bag sample from one quarter quadrangle.

Actual	Predicted			Agreement (%)
	Tree cover	No tree cover	Mixed	
Tree cover	1,928	75	7	95.9
No tree cover	232	1,067	6	81.8
Mixed	184	35	20	8.4
Overall				84.8

It should be noted that confusion matrices are somewhat more intuitive when provided for pixel-based classifications for which equal-sized units are compared. In this case, it may be more appropriate to weight the confusion matrix for each class by the mean area of the classes involved in the training dataset. Simple multiplication by the mean segment area from Table 1 reveals areal class agreement of 56.6%, 98.8%, and 2.8% for the tree cover, no tree cover, and mixed classes, respectively. Overall, the RF model correctly classified 34,300 of 41,836 ha, or 82% agreement with respect to area.

As RF builds each classification tree, the algorithm attempts to minimize the Gini index, $G(t) = \sum p(i)p(j)p(k)$, where $p(i,j,k)$ are the probabilities of the three classes at node t in the classification tree. The mean amount each predictor variable reduces the Gini index across all trees in the RF model is a measure of variable importance. The Haralick texture measures provided large decreases in the Gini index relative to the spectral and spatial attributes (Figure 4). In particular, gray-level co-occurrence matrix homogeneity calculated for all 3 bands (Glc_h) had a mean Gini index decrease of 181 while the largest decrease for a spectral attribute was 28 (Max_12—maximum pixel value in the red band) and the largest decrease for a spatial attribute was 24 (rectfit—rectangular fit). Twelve Haralick texture measures were among the top 20 most important predictor variables (Figure 4).

The RF model was applied to nearly 385,000 image segments across all of Pembina County. There is significant radiometric variability across the NAIP imagery within the county and the model appears not to have performed well in north/south flight paths for which the imagery radiometry differed substantially from that in the QQ used for the training dataset. In areas where this image difference is greatest, agricultural fields are predicted to be tree cover with high probability (Figure 4). The resulting tree cover estimate for the county is 37,963 ha, substantially higher than the FIA forest land estimate of 24,600 ha.

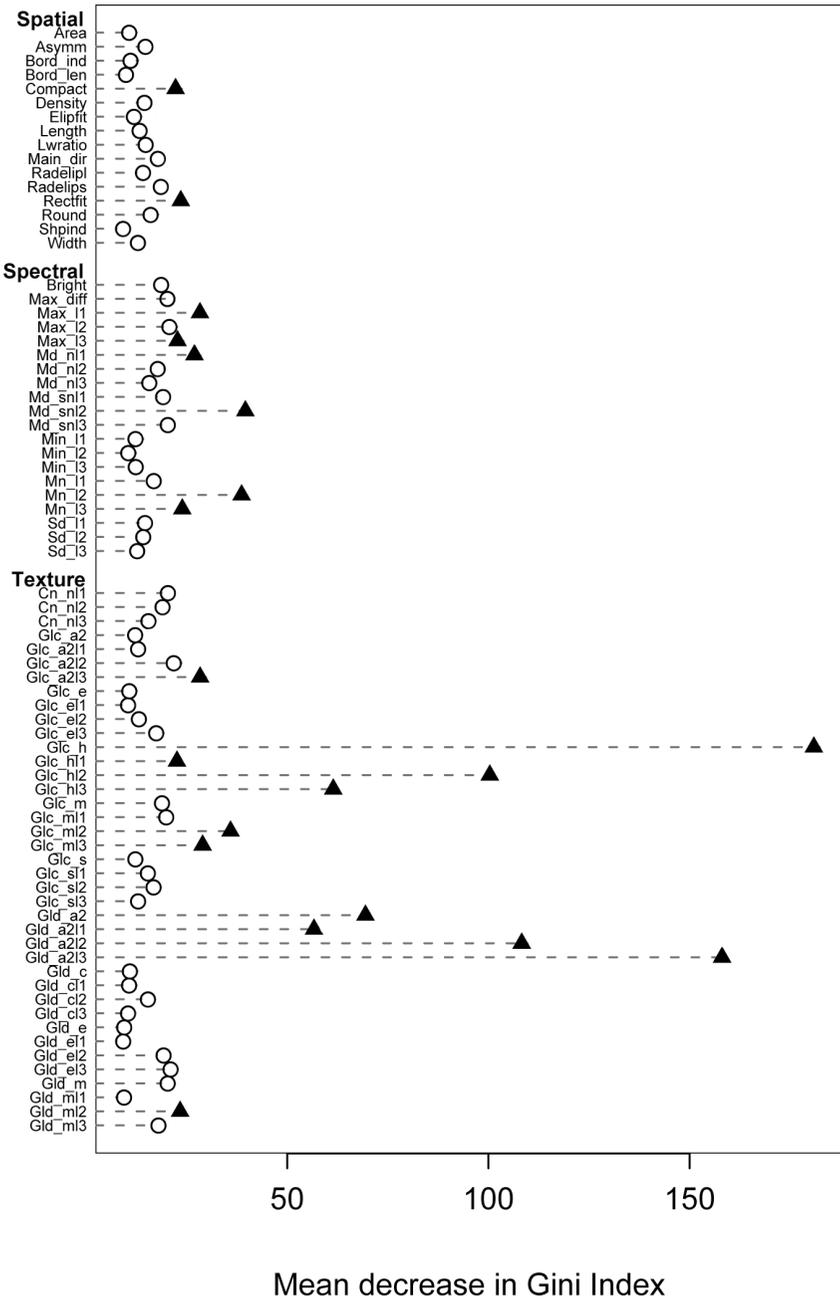


Figure 4. Mean decrease in Gini index using a Random Forests model of tree cover with a variety of spatial, spectral, and textural predictor attributes from image segments derived from one quarter quadrangle of 2003 NAIP imagery from Pembina County, North Dakota. Solid triangles indicate the 20 most significant predictors.

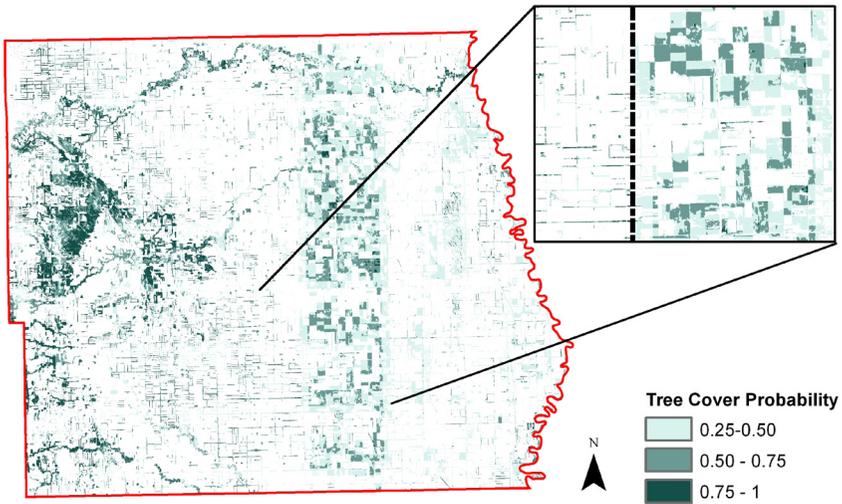


Figure 5. A 1-m resolution tree cover probability map for Pembina County, North Dakota, based on 2003 NAIP imagery. A north/south stripe just to the east of center is noticeable, probably due to radiometric differences in the imagery. In the enlarged area, agricultural fields are erroneously assigned a high tree cover probability to the east of the vertical, dashed black line.

DISCUSSION

We examined statewide areal estimates of land with trees from a variety of satellite-derived datasets. Direct comparison between the various sources is difficult due to differences in definition of land cover categories, yet the wide range of areas illuminates the challenge of monitoring tree cover in sparsely forested regions. Differences across data sources in North Dakota were greater than those in South Dakota, which has a slightly higher proportion of forest. Other patterns of difference could be attributed to a variety of sources, including differences in resolution, sensor characteristics, imagery vintage, and application of land use versus land cover definitions. Additionally, area estimation of tree cover in North Dakota and South Dakota is complicated by the abundance of tree plantings that are much narrower than the resolution of the satellite sensor, or in the case of FIA, do not meet the minimum width requirements for forest land.

We developed a method for estimating tree cover using image segmentation of widely available, high-resolution (1-m) NAIP imagery and an ensemble classification tree approach. A model was created that performed with an overall 84.8% classification agreement on the out-of-bag sample of image segments. Image segments for the reference dataset were labeled using visual image interpretation in about 8 hours. Most other procedures in the workflow required relatively little manual work. Image segmentation and the calculation of attributes required 30 to 40 minutes for each of Pembina County's 110 quarter quadrangles on a desk-

top PC. Applying the model to the image segments across the entire county took approximately 30 minutes of processing time. The methodology is relatively easy to implement. However, further investigation is required to determine how to efficiently work with the image variability across flight paths.

Early reviews of this manuscript raised concerns over the decision to develop a training dataset from a single QQ. As a result, we created a second training dataset with a minimum of 5 image segments in each of the 110 QQ's, thus distributing samples throughout the study area. A new RF model was created based on this training dataset, and class agreement based on the OOB sample fell to 71.3%. By grouping the Pembina County QQ's into 8 north/south strips that parallel flight paths, a new predictive factor was created with each flight path assigned a value ranging from 1 to 8. When this factor was added to the RF model, OOB class agreement improved to 74.2%. The reduced class agreement (relative to the RF model created using the single QQ training dataset) is very likely due to differences in the number of image segments in the training dataset (714 versus 3,554). However, it is illustrative to note when the flight path predictor variable was included with the spatially distributed training dataset, the Gini index indicates it is the most important predictor. Further evaluation is required to determine how much OOB class agreement can be improved by adding additional image segments to the training dataset and whether or not adding flight path or perhaps longitude of the image segment centroid would allow RF to accurately predict tree cover across images with radiometric variability.

Our approach builds on the work of Laliberte et al. (2007) and Wiseman et al. (2008). We extend their image segmentation approaches to NAIP imagery available over a wide area with a few differences. Our focus was on developing an operational methodology for assessing trees in sparsely forested regions, and we were able to map tree cover for nearly 290,000 ha in North Dakota (although with limited accuracy). Wiseman et al. (2008) focused on shelterbelts with distinctive shape and area characteristics, and used this knowledge to isolate shelterbelts from a database of spectral and spatial attributes using Boolean logic in a manual, iterative process. They found attributes such as area, asymmetry, and shape index to be important predictors of shelterbelts. In contrast, we were more generically focused on tree cover, and found Haralick texture attributes to be the most important predictors. Additionally, after the development of a training dataset, using RF was a relatively automated approach to classification.

Laliberte et al. (2007) used CART® in conjunction with image segmentation of QuickBird imagery. They found the mean near-infrared values of image segments to be one of the most important predictors of vegetative cover. In contrast, we used RF, which is related to CART®, but produces a series of classification trees and has been found to produce better classification accuracy (Gislason et al., 2006). The 2003 NAIP imagery for North Dakota did not include a near-infrared band, so our model was more dependent on texture measures rather than spectral information.

In summary, satellite-derived data products provide a means for monitoring trees over large areas. However, estimates of forest land are inconsistent across various datasets in the Great Plains. Trees in agricultural settings are undoubtedly under-represented in satellite-derived datasets because sensor resolutions are too coarse to consistently capture narrow linear plantings of trees. If we are to account for very narrow tree plantings, such as single- or double-row windbreaks, remote sensing approaches that utilize existing imagery sources with resolutions finer than 5 m are needed. We presented an incremental step toward monitoring trees in agricultural landscapes. Although NAIP is a nationwide source of imagery, the lack of consistency between flight paths is an operational challenge to mapping tree cover at a high-resolution on a broad scale. With additional development, the approach presented using Random Forests could be a viable operational tool for mapping tree cover from existing imagery sources.

NOTES

- 1 <http://www.nass.usda.gov/research/Cropland/SARS1a.htm>.
- 2 <http://www-modis.bu.edu/landcover/userguide/lc.html>.
- 3 Random Forests is a trademark of Leo Breiman and Adele Cutler and is licensed to Salford Systems.
- 4 The R Project for Statistical Computing--<http://www.r-project.org/>.
- 5 <http://cran.r-project.org/web/packages/randomForest/index.html>.

REFERENCES

- Albrigo, L.G., and Russ, R.V. 2002. Considerations for improving honeybee pollination of citrus hybrids in Florida. *Proceedings of the Florida State Horticultural Society* 115:27-31.
- Bannari, D. Morin, Bonn, F., and Huete, A.R. 1995. A review of vegetation indices. *Remote Sensing Reviews* 13:95-120.
- Bechtold, W.A., and Patterson, P.L. (eds.). 2005. *The enhanced forest inventory and analysis program--national sampling design and estimation procedures*. General Technical Report SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 pp.
- Breiman, L. 2001. Random forests. *Machine Learning* 45:5-32.
- Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth International Group. 358 pp.
- Carroll, Z.L., Bird, S.B., Emmett, B.A., Reynolds, B., and Sinclair, F.L. 2004. Can tree shelterbelts on agricultural land reduce flood risk? *Soil Use & Management* 20(3):357-359.
- Carucci, R. 2000. Trees outside forests: an essential tool for desertification control in the Sahel. *Unasylva* 200. <http://www.fao.org/docrep/x3989e/x3989e05.htm>.
- Definiens AG. 2007. Definiens Developer 7 Reference Book, URL: <http://www.definiens.com>.
- Definiens AG. 2008. Definiens Developer 7 User Guide, URL: <http://www.definiens.com>.

- Easterling, W.E., Hays, C.J., McKenney Easterling, M., and Brandle, J.R. 1997. Modelling the effect of shelterbelts on maize productivity under climate change: an application of the EPIC model. *Agriculture, Ecosystems & Environment* 61(2-3):163-176.
- FAO. 2000. *FRA 2000 on definitions of forest and forest change*. Forest Resources Assessment Working Paper No. 33. Rome: FAO Forestry Department. 15 pp.
- Gislason, P.O., Benediktsson, J.A., and Sveinsson, J.R. 2006. Random Forests for landcover classification. *Pattern Recognition Letters* 27:294-300.
- Goebel, J.J. 1998. The National Resources Inventory and its role in U.S. agriculture. In Holland, T.E.; Van den Broecke, M.P.R. (eds.). *Agricultural Statistics 2000: Proceedings of the international conference on agricultural statistics*. Voorburg: International Statistical Institute. 181-192.
- Hansen, M. 1985. Line intersect sampling of wooded strips. *Forest Science* 31(2):282-288.
- Hansen, M., DeFries, R., Townshend, J.R., Carroll, M., Dimiceli, C., and Sohlberg, R. 2003. *Vegetation Continuous Fields MOD44B, 2001 Percent Tree Cover, Collection 3*. College Park, MD: University of Maryland.
- Haralick, R., Shanmugan, K., and Dinstein, I. 1973. Textural features for image classification. *IEEE Transactions on Systems, Man & Cybernetics* 3(1):610-621.
- Hartong, A.L., and Moessner, K.E. 1956. *Wooded strips in Iowa*. Forest Survey Release 21. Columbus, OH: U.S. Department of Agriculture, Forest Service, Central States Forest Experiment Station.
- Holmgren, P., Masakha, E.J., and Sjöholm, H. 1994. Not all African land is being degraded: a recent survey of trees on farms in Kenya reveals rapidly increasing forest resources. *Ambio* 23(7):390-395.
- Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J.N., and Wickham, J. 2007. Completion of the 2001 National Land Cover Database for the conterminous United States. *Photogrammetric Engineering & Remote Sensing* 73:337-341.
- Kort, J., and Turnock, R. 1998. Carbon reservoir and biomass in Canadian prairie shelterbelts. *Agroforestry Systems* 44(2-3):175-186.
- Kulshreshtha, S., and Kort, J. 2009. External economic benefits and social goods from prairie shelterbelts. *Agroforestry Systems* 75(1):39-47.
- Laliberte, A.S., Rango, A., Havstad, K.M., Paris, J.F., Beck, R.F., McNeely, R., and Gonzalez, A.L. 2004. Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico. *Remote Sensing of Environment* 93(1-2):198-210.
- Labiberte, A.S., Fredrickson, E.L., and Rango, A. 2007. Combining decision trees with hierarchical object-oriented image analysis for mapping arid rangelands. *Photogrammetric Engineering and Remote Sensing* 73(2):197-207.

- Lister, A., Scott, C.T., and Rasmussen, S. 2009. Inventory of trees in nonforest areas in the Great Plains States. In McWilliams, W., Moisen, G., and Czaplewski, R. (comp.). *Forest Inventory and Analysis (FIA) Symposium 2008*; 21-23 Oct 2008; Park City, UT. Proceedings RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.
- Miles, P.D. 2009. Forest inventory mapmaker web-application version 3.0. St. Paul, MN: U.S. Department of Agriculture, Forest Service. [online]. fia.fs.fed.us/tools-data/other/default.asp (accessed 30 Mar 2009).
- Nelson, M.D. 2005. Satellite remote sensing for enhancing national forest inventory. [Ph.D. dissertation]. St. Paul: University of Minnesota. 167 pp.
- Oduol, P.A., Muraya, P., Fernandes, E.C.M., Nair, P.K.R., Oktingati, A., and Maghembe, J. 1988. The agroforestry systems database at ICRAF. *Agroforestry Systems* 6(1-3):253-270.
- O'Neill, R.V., Hunsaker, C.T., Timmins, S.P., Jackson, B.L., Jones, K.B., Riitters, K.H., and Wickham, J.D. 1996. Scale problems in reporting landscape pattern at the regional scale. *Landscape Ecology* 11:169-180.
- Perry, C.H., Woodall, C.W., Liknes, G.C., and Schoeneberger, M.M. 2009. Filling the gap: improving estimates of working tree resources in agricultural landscapes. *Agroforestry Systems* 75(1):91-101.
- Rawat, J.K., Dasgupta, S., and Kumar, R. 2004. Assessment of tree resources outside forest based on remote sensing satellite data. Proceedings of Map India 2004. [online]. www.gisdevelopment.net/application/environment/conservation/mi04179.htm (accessed 30 Mar 2009).
- Rosenberg, D.K., Noon, B.R., and Meslow, E.C. 1997. Biological corridors: form, function, and efficacy. *Bioscience* 47:677-687.
- Ucar, T., and Hall, F.R. 2001. Windbreaks as a pesticide drift mitigation strategy: a review. *Pest Management Science* 57(8):663-675.
- U.S. Department of Agriculture, National Agriculture Statistics Service. 2007. Census of Agriculture. [online]. Washington, DC. www.agcensus.usda.gov (accessed 30 Mar 2009).
- Wiseman, G, Kort, J., and Walker, D. 2008. Quantification of shelterbelt characteristics using high-resolution imagery. *Agriculture, Ecosystems & Environment* 131(1-2):111-117.
- Woodcock, C.E., and Strahler, A.H. 1987. The factor of scale in remote sensing. *Remote Sensing of Environment* 21:311-332.