

# Application of Geostatistics to Characterize Leaf Area Index (LAI) from Flux Tower to Landscape Scales Using a Cyclic Sampling Design

S. N. Burrows,<sup>1\*</sup> S. T. Gower,<sup>1</sup> M. K. Clayton,<sup>2</sup> D. S. Mackay,<sup>1</sup> D. E. Ahl,<sup>1</sup> John M. Norman,<sup>3</sup> and George Diak<sup>4</sup>

<sup>1</sup>Department of Forest Ecology and Management, University of Wisconsin–Madison, 120 Russell Laboratories, 1630 Linden Drive, Madison, Wisconsin, 53706, USA; <sup>2</sup>Department of Statistics, University of Wisconsin–Madison, 1210 W. Dayton Street, Madison, Wisconsin, 53706, USA; <sup>3</sup>Department of Soil Science, University of Wisconsin–Madison, 1525 Observatory Drive, Madison, Wisconsin, 53706, USA; and <sup>4</sup>Space Science and Engineering Center, University of Wisconsin–Madison, 1210 W. Dayton Street, Madison, Wisconsin, 53706, USA

## ABSTRACT

Accurate characterization of leaf area index (LAI) is required to quantify the exchange of energy, water, and carbon between terrestrial ecosystems and the atmosphere. The objective of this study was to use a cyclic sampling design to compare the spatial patterns of LAI of the dominant terrestrial ecosystems that comprised the area around the 447-m WLEF television tower, equipped with an eddy flux system, near Park Falls, Wisconsin, USA. A second objective was to compare the efficiency of cyclic, random, and uniform sampling designs in terms of the precision of spatial information derived per unit sampling effort. The vegetation surrounding the tower was comprised (more than 80%) of four major forest cover types: forested wetlands, upland aspen forests, upland northern hardwood forests, and upland pine forests, and a fifth, nonforested cover type, grass (open meadow). LAI differed significantly among the five cover types and averaged 3.45, 3.57, 3.82, 3.99, and 1.14 for northern hardwoods, aspen, forested wetlands, upland conifers, and grass, respectively. The cyclic sampling design maximized information about the variance of veg-

etation characteristics of the heterogeneous landscape and decreased by 60% the number of plots needed to obtain the same confidence interval width using a random sampling design. The range of spatial autocorrelation for LAI was 147 m, but it was decreased to 117 m when vegetation cover was included as a covariate. The cyclic sampling design has several important advantages over other sampling designs. The cyclic sampling design increased the sampling efficiency by optimizing the placement of plots so they were distributed more efficiently for geostatistical analyses such as semi-variograms, correlograms, and spatial regression and can incorporate covariates (for example, vegetation cover, soil properties, and so on) that may explain the sources of spatial patterns. The cyclic sampling design was used to derive a spatial map of LAI and the average LAI for the 3 × 2 km area centered on the flux tower was 3.51 ± 0.89 (with a minimum of 0 and a maximum of 6.35). Airborne and satellite reflectance data have also been used to characterize LAI, but in this region, and many other forests of the world, remotely sensed vegetation indexes saturate in forests with an LAI greater than 3–5. The cyclic sampling design also provides a general ecological sampling approach that can be used at mul-

Received 2 January 2001; Accepted 4 March 2002.

\*Corresponding author; e-mail: burrows@calshp.cals.wisc.edu

tiple scales.

**Key words:** leaf area index (LAI); geostatistics; sampling; variogram; cyclic sampling design; eddy flux; eddy covariance.

## INTRODUCTION

Accurate assessment of regional and global-scale changes in the terrestrial biosphere is essential if human impact on biosphere sustainability is to be understood. There are a myriad of ecosystem attributes to be monitored, but quantifying human habitability will likely include an evaluation of vegetation cover, leaf area index (LAI), and net primary productivity (NPP) (Running and others 1999). These variables are important because vegetation cover is a proxy for land use and biodiversity, LAI is positively correlated to NPP, and NPP is important because it is correlated to fuel, fiber, and food production for human consumption.

Measuring vegetation characteristics such as vegetation cover, LAI, and NPP is not a straightforward task, especially over an area larger than 10 km<sup>2</sup>. How to measure, when to measure, what to measure, and where to measure are all questions that have to be answered before a study can begin. Although all of these questions are essential, the last question—where to measure—is the focus of this paper.

Traditional experimental sampling designs, such as transects (Iachan 1982), random samples (Müller and Zimmerman 1999; Bogaert and Russo 1999), or stratified random sampling schemes (Fassnacht and others 1997), are commonly used to quantify vegetation characteristics and their spatial patterns. These designs, however, may not be the most efficient designs for characterizing spatial patterns of land cover and LAI, especially given the inherent spatial autocorrelation that is often prevalent in ecological systems (Legendre 1993; Aubry and Debouzie 2000; Dutilleul 1993). Many sampling designs are spatially inefficient because they do not account for methods normally used to determine patterns of spatial autocorrelation using geostatistical tools (variograms and correlograms). An efficient sampling design distributes pairs of points at different distances (or lags) to retain the sampling density needed, while decreasing the number of redundant measurements at a constant distance between sample points (Fortin and others 1989).

To estimate the changes and patterns measured across space, geostatistical analysis can be used in a way similar to the use of time-series analysis to study changes and patterns over time. In this paper,

we present a cyclic sampling design that complements the tools used for geostatistical analysis. The cyclic sampling design is illustrated with a case study in northern Wisconsin where LAI was characterized for the footprint of a very tall eddy flux tower as part of NASA's Earth Observing System's Validation Program (Justice and others 1998; Cohen and Justice 1999).

Geostatistical data often exhibit small-scale variations that can be modeled based on spatial correlation. Spatial variability is modeled as a function of distance between sample locations. Locations that are closer to each other are often more similar than locations that are farther apart, and are thus more highly correlated. Spatial variability is often modeled with a semi-variogram instead of a correlation function (Cressie 1993). The semi-variogram represents variance ( $\gamma$ ) as a function of distance between sample locations. Gamma ( $\gamma$ ) is defined as:

$$\gamma(h) = \frac{1}{2|N(h)|} \sum_{N(h)} (z_i - z_j)^2 \quad (1)$$

where  $N(h)$  is the set of all pairs of observations such that the distance between  $i$  and  $j$  is  $h$ .  $|N(h)|$  is the number of distinct pairs in  $N(h)$ , and  $z_i$  and  $z_j$  are data values at locations  $i$  and  $j$ , respectively.

The characteristics of the semi-variogram are particularly important: the nugget, the sill, and the range. The nugget is the estimate of the variance at distance ( $h$ ) equal to 0 ( $\lim_{h \rightarrow 0} \gamma(h)$ ). The sill is the  $\lim_{h \rightarrow \infty} \gamma(h)$  and represents the variance ( $\sigma^2$ ) of the random field (the study area as a whole). Finally, the range is the distance at which data are no longer autocorrelated (Kaluzny and others 1998). Knowing these three characteristics of a semi-variogram (nugget, sill, and range) can provide insight into the spatial patterns on the landscape. A large sill in comparison to the nugget indicates that variance is spatially dependent at scales smaller than the range. Similarly, a longer range indicates that the variance is spatially dependent over longer scales. As the nugget increases, the variation at distance 0 increases, which indicates an increase in the variance not attributable to spatial dependence. This can be caused by measurement errors, or by inherent variations at very small distances. Semi-variograms can be modeled using several different generalized equations that fit the observed patterns, including spherical, exponential, gaussian, rational, and linear equations. Details about fitting semi-variograms can be found in Cressie (1993), Kaluzny and others (1998), and Pinheiro and Bates (2000).

Numerous optimization methods have been proposed to derive efficient sampling designs that help

**Table 1.** Selected Cyclic Sampling Design Patterns

Cycle Definition	Cycle Length (x)	Plots Sampled (n)	Sample Locations (0 to x-1)
1/1	1	1	0
2/3	3	2	0, 1
3/7	7	3	0, 1, 3
4/13	13	4	0, 1, 3, 9
5/21	21	5	0, 1, 4, 14, 16
6/31	31	6	0, 1, 3, 8, 12, 18
7/37	37	7	0, 1, 6, 10, 17, 23, 35

Based on Clinger and Van Ness (1976)

A 3/7 cycle indicates that three plots in every seven are measured. The three plots are spaced such that pairs of plots can be found that are separated by one, two, three, four, five, six, and seven plot widths. The seven-plot cyclic pattern provides additional plot pairs separated by eight, nine, . . . and so on plot widths.

in estimating variograms. Pettitt and McBratney (1993) proposed a transect sampling design developed specifically for the purpose of variogram estimation. Bogaert and Russo (1999) proposed a *D* optimization of a sampling design starting with either a random or systematic base sampling structure. Several other sampling designs have been compared for computational time and optimality in Müller and Zimmerman (1999).

Clinger and Van Ness (1976) proposed a cyclic scheme for sampling over time that uses a pattern of sampled plots to provide information about all lags (time periods between measurements) and repeated the pattern in time. This initial proposal of measuring over time can also be applied to space. A cyclic sampling scheme provides information on the relationship between the characteristics of plot pairs separated by any distance. The number of fundamental unit lengths taken before the sample pattern repeats defines the cycle length. For each cycle length, an optimum sample spacing can be defined that minimizes the number of samples but provides sample pairs separated by any distance (Table 1). Clayton and Hudelson (1995) applied the cyclic sampling design to examine plant disease epidemiology patterns in an agroecosystem and proposed a method for extending the cyclic sampling design to two dimensions. Miller and others (forthcoming) used a two-dimensional, cyclic sampling design to examine spatial patterns of understory vegetation in old-growth northern hardwood forests in northern Wisconsin.

The objective of this study was to use a cyclic sampling design to compare the spatial patterns of LAI of the dominant terrestrial ecosystems that

comprised the area around the 447-m WLEF television tower, equipped with an eddy flux system, near Park Falls, Wisconsin, USA. A second objective was to compare the efficiency of cyclic, random, and uniform sampling designs in terms of the precision of spatial information derived per unit sampling effort.

## METHODS

### Site Description

The study was conducted in northern Wisconsin, near Park Falls (45.9458°N, 90.2723°W). The study area, one of NASA's Earth Observing System Core Validation Sites (<http://modis-land.gsfc.nasa.gov>), was centered on a 447-m-tall communications tower that also supports an eddy flux system to measure energy, water, and carbon exchange between the forest landscape and the atmosphere (Bakwin and others 1998). The tower and surrounding area is located in the Chequamegon-Nicolet National Forest. The study area is in the Northern Highlands physiographic province, a southern extension of the Canadian Shield. The bedrock is comprised of Precambrian metamorphic and igneous rock. It is overlain by 8–90 m of glacial and glaciofluvial material deposited approximately 10,000–12,000 years ago. Topography is slightly rolling, varying by 45 m between highest and lowest elevations within the defined study area. Outwash, pitted outwash, and moraines are the dominant geomorphic landforms. The growing season is short and the winters are long and cold. Mean annual July and January temperatures are 19°C and –12°C, respectively.

The effects of landform on forest ecosystem dynamics and management activities such as thinning and selective and clear-cut harvests have impacted the forest vegetation (Fassnacht and Gower 1997, 1999). Red pine (*Pinus resinosa* Ait) and jack pine (*Pinus banksiana* Lamb) dominate areas of excessively drained, sandy soils derived from glacial outwash. Northern hardwood forests, comprised of sugar maple (*Acer saccharum* Marsh.), red maple (*Acer rubrum* L.), green ash (*Fraxinus Americana* Marsh.), yellow birch (*Betula alleghaniensis* Britton), and basswood (*Tilia americana* L.), occur on the finer-textured soils derived from moraines and drumlins. Soils of intermediate characteristics support a wide variety of broad-leaf deciduous tree species, such as paper birch (*Betula papyrifera* Marsh), quaking aspen (*Populus tremuloides* Michx), bigtooth aspen (*Populus grandidentata* Michx), red maple, and red and white pine (*Pinus strobus* L.).

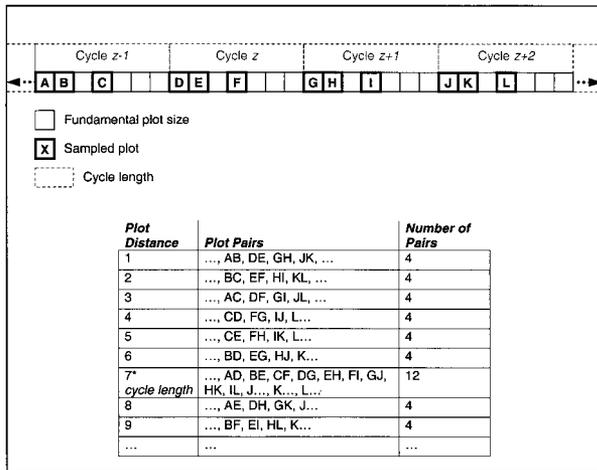


Figure 1. Example of a one-dimensional sampling design based on Clinger and Van Ness (1976) using four 3/7 cycles, illustrated with the dashed gray lines. Thinned lined boxes represent the fundamental plot size. Black boxes represent sampled plots. The table lists the pairs of points found at each plot distance in the four cycles that are used. Equal sampling is achieved at all plot distances except the plot distance that is equal to the cycle length, at which point all pairs of points in one cycle overlap with the pairs of points in another cycle.

The poorly drained organic soils are dominated by white cedar (*Thuja occidentalis* L.), balsam fir (*Abies balsamea* (L.) Mill), white spruce (*Picea glauca* (Moench) Voss), black spruce (*Picea mariana* (Mill) B.S.P.), tamarack (*Larix laricina* (Du Roi) Koch.), and speckled alder (*Alnus regosa* (Du Roi) Spreng.).

### Cyclic Sampling Design

To illustrate the cyclic sampling design, consider a 3/7 design, indicating that three plots in every seven are measured (Table 1). The three plots are spaced within a cycle and the cycle repeats, such that pairs of plots can be found that are separated by one, two, three, four, five, six, and seven plot widths (Figure 1). For example, pair AB is at a plot distance of one, BC is at a distance of two, AC is at a distance of three, and so on. The seven-plot cyclic length provides additional plot pairs separated by eight, nine, and so on plot widths. In the figure, the cycle is repeated four times, but ideally the cycle is repeated numerous times. The only redundancy (or inefficiency) in the design is the frequency at which the cycle is repeated. In the example, it is at a plot distance of seven where there are 12 pairs of plots compared to the four pairs of plots found at all other distances. The cyclic sampling design can also be implemented in multiple dimensions, and the

cycle can differ in the x and y dimension: for example, a 5/21 cycle can be used in the x dimension, and a 4/13 cycle can be used in the y dimension (see Campbell and others 1999).

Before installing plots in the field, a preliminary study was conducted to determine the approximate scale at which spatial patterns of vegetation characteristics occurred. Normalized difference vegetation index (NDVI) has been used to characterize LAI for this region (Fassnacht and others 1997). We developed NDVI maps for the study area from several Landsat midsummer images and quantified the spatial patterns (S-PLUS Spatial Statistics Module 1.5 for Windows; Mathsoft Inc., Cambridge, MA, USA). The range of spatial autocorrelation was approximately 500–600 m depending on the Landsat image used. We tested other vegetation indexes (Tasselled caps, Simple Ratio, individual bands) because they have been found to be better predictors of LAI for northern hardwoods (Fassnacht and others 1997; Gower and others forthcoming), but we found that all the vegetation indexes produced similar estimates for the range of variation. Therefore, we used NDVI to determine which cycle length was most likely to yield spatial patterns and minimize the number of field measurements. An iterative process was used to identify an optimal design. We sampled the image at the locations prescribed by each design and estimated the spatial variations (using variograms) to determine if it matched our population estimates derived from using all pixels within the NDVI scene. Confidence intervals, based on Cressie (1993) and Kabrick and others (1997), were calculated for gamma to ensure that the width of a 95% confidence interval was within 10% of the gamma estimate from the variogram calculation. We ensured that the confidence interval width stayed below 10% for distances up to two times the range estimates generated from remote sensing. This conservative assessment was done because our initial range estimates from remote sensing were only approximations of LAI ranges.

We selected the 6/31 cyclic sampling design pattern (Table 1) and implemented it in a 3 cycle by 3 cycle pattern across the landscape, except that one cycle had to be moved and 12 plots excluded to avoid private land, leaving a total of 312 plots (Figure 2). Plot size was defined at 30 × 30 m to be consistent with Landsat TM pixel size. Comparisons of NDVI-based variograms and the changes in confidence interval widths showed that the modified design still complied with the predefined selection criteria. We assessed our final sampling design against the vegetation cover of the surrounding area, using the WISCLAND statewide vegetation

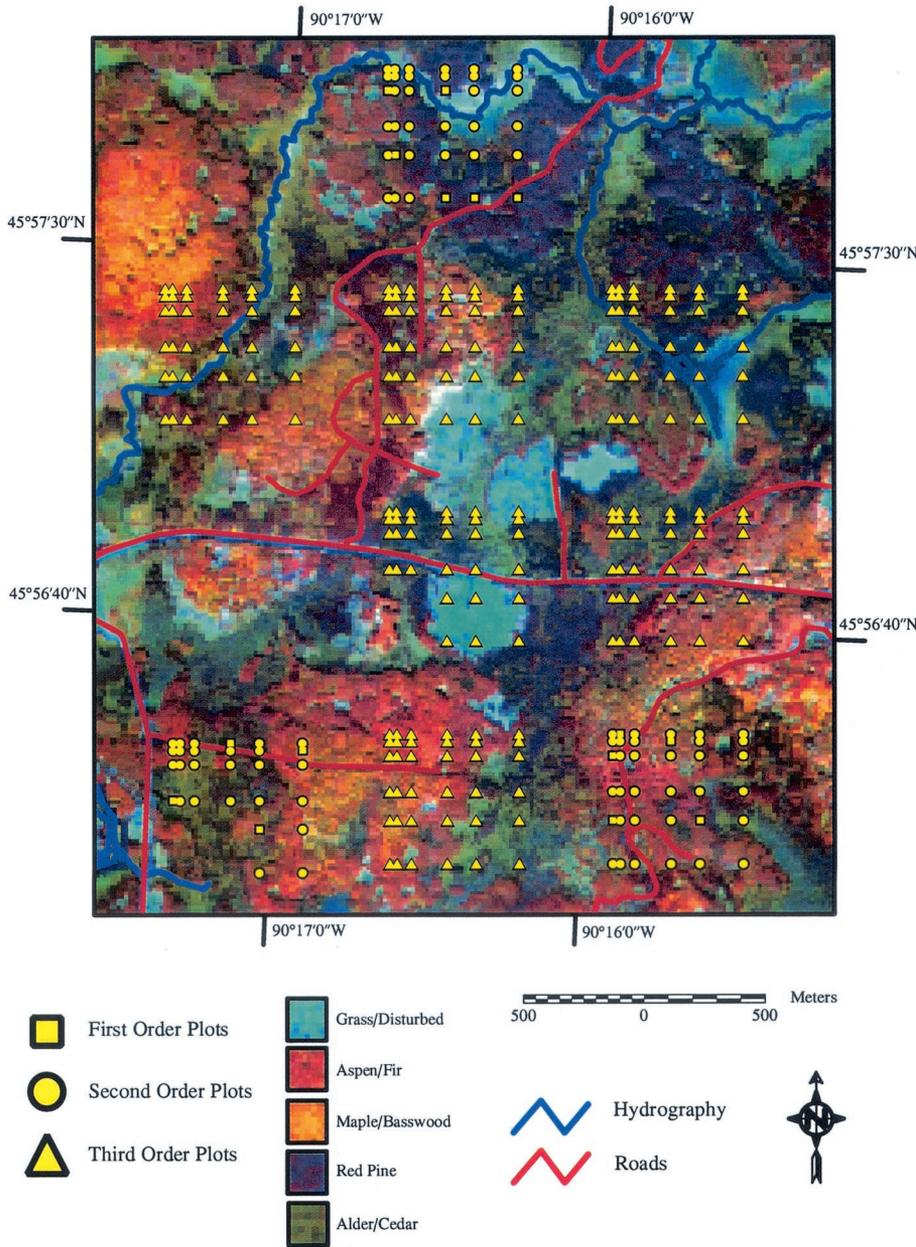


Figure 2. Image illustrating the cyclic sampling design used to characterize the vegetation characteristics for the WLEF tall eddy flux tower in the Chequamegon-Nicolet National Forest, Price County, Wisconsin, USA. First, second, and third order plots are overlaid on top of an airborne multispectral image from September 1998 that shows the common spectral characteristics of the major cover types found in the site. First and second order plots had optical LAI and direct LAI measurements taken; third order plots only had optical LAI measurements taken.

cover map (WiDNR 1998), to ensure that we were sampling vegetation cover types in percent coverage that were representative of the percent coverage found in the study area (see Table 2).

**Field Measurements**

Plot centers were located using an Ashtech GG-24 Surveyor (Magellan Inc., Sunnyvale, CA, USA), a single-frequency real-time kinematic (RTK) global positioning system (GPS) that used both the GPS and GLONASS satellites (Van Diggelen 1997). A base station and a rover unit were used. The base station was equipped with a 35-W radio modem

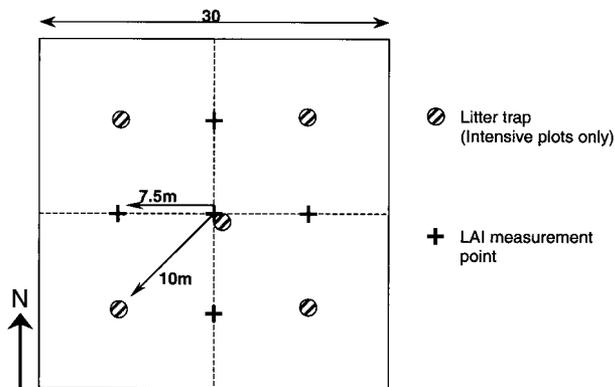
(Pacific Crest, San Jose, CA, USA) to broadcast the real-time differential correction information to the rover unit. All 312 plots were surveyed to within 50-cm horizontal root mean square error, and the center of each plot was marked with a wooden stake.

Variable radius subplots were established at plot center and at 7.5 m from plot center in each of the four cardinal directions (Figure 3). A detailed description of the plot design is summarized in Campbell and others (1999). All trees greater than 2.5 cm diameter at breast height (DBH, 1.37 m) were identified to species and the diameter was measured.

**Table 2.** Summary of the Percent Vegetation Cover Types Sampled for the Entire Study Area Population, Uniform Sampling Scheme (with 315 Points), Random Sampling Scheme (312 Points), and Implemented Cyclic Sampling Scheme (312 points)

WISCLAND Cover Type	WISCLAND Cover (%)	Uniform Sampling Design (%)	Random Sampling Design (%)	Cyclic Sampling Design (%)
Agric. other	0.3	0.3	1.3	0.3
Grass	1.8	1.9	1.0	1.6
Jack pine <sup>a</sup>	6.7	7.0	3.5	7.7
Red pine	6.5	7.3	9.0	3.5
Mixed/other conifer	2.4	1.9	1.3	1.3
Aspen	23.8	22.2	25.3	22.1
Sugar maple	7.8	9.8	8.7	12.5
Mixed/other deciduous	5.9	7.0	7.7	6.1
Mixed/deciduous/conifer	5.0	7.0	5.4	7.4
Open water	0.2	0.0	0.0	0.0
Wetland: emergent/meadow	0.3	0.3	0.0	1.6
Wetland: lowland shrub	3.7	2.5	4.5	2.6
Wetland: shrub deciduous	10.6	8.3	10.3	7.7
Wetland: broad-leaved evergreen	2.8	2.2	1.9	1.3
Forested wetland: conifer	6.2	8.3	5.1	8.7
Forested wetland: mixed	9.3	9.5	9.9	12.5
Barren	3.5	1.3	1.9	0.6
Shrub land	3.2	3.2	3.2	2.6
<b>Total</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>

Based on the WISCLAND land cover classification (WIDNR 1998)  
<sup>a</sup>All jack pine in study area clear-cut in 1996–97



**Figure 3.** Plot-level sampling scheme for leaf area measurements. Crosses represent optical LAI measurements; hashed circles represent litter basket leaf area estimates.

Litterfall traps were installed to measure leaf litterfall and LAI. Five 60 × 40 cm litterfall traps were used at 100 of the 312 plots. The 100 plots were located in three distinct corners of the study area, each with different vegetation types present.

Plots were defined as first, second, or third order

depending on the intensity of measurements at the plot. LAI was measured optically at all plots (first, second, and third order plots) using a Li-Cor LAI-200 Plant Canopy Analyzer (Li-Cor Inc., Lincoln, NE, USA) and measured directly from litterfall traps at the first and second order plots. Standard field measurement methods were used (Gower and Norman 1991; Fassnacht and others 1994; Chen and others 1997).

LAI is normally calculated as  $LAI = (1-\alpha)Le_{\gamma_E} / \Omega_E$ , where  $\alpha$  is the ratio of wood area to total (wood + foliage area) plant area,  $Le$  is the effective LAI,  $\gamma_E$  is the needle-to-shoot area ratio that quantifies clumping at the shoot level, and  $\Omega_E$  is the site-specific clumping correction factor for clumping at the branch to tree level (Chen and others 1997; Gower and others 1999).  $\gamma_E$  is calculated as  $A_n/A_s$ , where  $A_n$  is the ratio of one-half the total area (all sides) of needles in a shoot and  $A_s$  is one-half the total shoot area (Gower and others 1999). Because of the mixed nature of many of the plots, the  $\gamma_E$  parameter for each species was weighted by the basal area that each species contributes to total plot basal area. Because of the high

**Table 3.** Summary of July 1999 Leaf Area Index Estimates by Dominant Cover Type, with a Spherical Spatial Covariance Structure

Vegetation Cover Type	Mean	SE	5th %	95th %	Range (m)	Sill	Nugget
Aspen	3.57	0.19	0.51	5.72	80.20	1.90	0.2585
Forested wetlands	3.82	0.12	1.32	5.66	158.99	1.29	0.5144
Grass	1.14	0.35	0.10	3.80	22.49	1.23	0.3431
Northern hardwoods	3.45	0.16	1.31	5.44	90.63	1.61	0.2268
Upland conifers	3.99	0.23	0.98	6.57	53.19	2.37	0.3240

SE, standard error of the estimated mean

The sill is the estimated variance ( $\hat{\sigma}^2$ ) as  $\lim_{h \rightarrow \infty} \gamma(h)$ , the nugget is the estimated variance at distance zero, and the range is the estimated distance at which the sill begins.

levels of heterogeneity at the site, it was difficult to ascertain the correct value for  $\Omega_E$ . We therefore assumed a value of one, so that LAI was only corrected for shoot clumping. Excluding the  $\Omega_E$  correction factor should not affect the results because it is typically near one for northern hardwoods forests (Gower and others 1999) and will not affect general spatial relationships, which is the focus of this paper.

### Design Comparison Methodology

We compared the cyclic sampling design to uniform and random sampling schemes for the 312 plots used. To do this, we sampled from a Landsat TM image that served as our "population." The uniform sampling design had plots located 180 m apart in the x and y dimensions, and the random sampling design used plots that were randomly located with a minimum of 30 m between plot centers. The test consisted of a series of simulations to quantify the number of points needed for the random sampling design to achieve a similar estimate of variance and confidence interval of gamma as the cyclic sampling design. Using ERDAS Imagine 8.4 (ERDAS, Inc., Atlanta, GA, USA), cells were randomly sampled from that population. Estimates of spatial autocorrelation parameters and confidence intervals were calculated from those samples.

### Statistical Analyses

All of the analyses presented are based on the subplot data. Statistical analyses were performed using S-Plus 2000 for Windows (Mathsoft, Inc., Cambridge, MA, USA.), the S-Plus Spatial Statistics Module version 1.5 for Windows (Mathsoft, Inc., Cambridge, MA, USA), and the NLME Library v3.3 (Pinheiro and Bates 2000). Spatial regression was accomplished using SAS's PROC MIXED procedure (SAS release 8.1 for Windows; SAS Institute, Inc.,

Cary, NC, USA). We used variograms because their results can easily be integrated into the mixed-effects model's covariance structure. The LME procedure of Pinheiro and Bates (2000) has spatial covariance structure built into the model-fitting routines. Neither SAS nor S-Plus alone had the capacity to accomplish all that we needed, so a combination of the tools available in each package was used.

A spatially explicit prediction (that is, a map) was created using the built-in Kriging function in SAS's PROC MIXED. Kriging is a linear interpolation that allows predictions of unknown values in the study area based on information from measurements made at sample locations. The process works by taking the local information measured at each plot and applying it to each cover type based on the variogram models (Table 3). It incorporates the model of spatial variance derived from those observations, as well as the individual measurements, to produce a map that predicts both the values across the landscape and the standard errors of those predictions. For additional information on Kriging, see Cressie (1993), Kaluzny and others (1998), and Pinheiro and Bates (2000). We used the range, sill, and nugget values of the different cover types and applied those to a cover type map of the site (Mackay and others forthcoming).

## RESULTS AND DISCUSSION

### Comparison of Different Sampling Designs

We focus on gamma and the confidence interval calculated for the gamma estimates at different distances to evaluate the sampling plans. Specifically, if the confidence interval width was less than 10% of gamma, the sampling design was considered effective for that distance. The uniform sampling design provided spatial information only at intervals

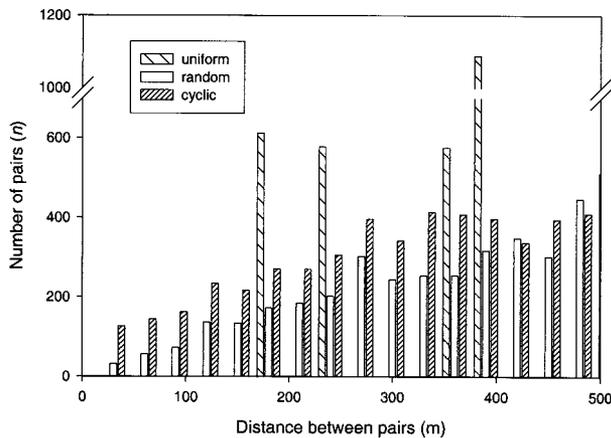


Figure 4. Comparison of the effectiveness of three different sampling schemes at sampling distances between 30 and 500 m given a plot size of 30 m.

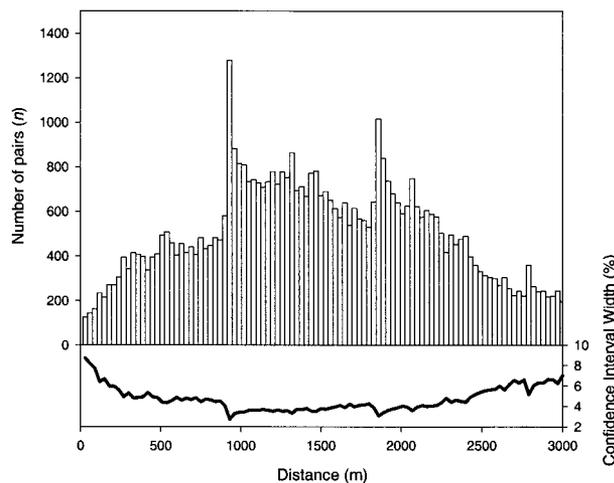


Figure 5. Histogram of the number of pairs of points sampled using the cyclic sampling design from 30 to 3000 m. Below are the confidence interval widths based on the number of pairs of points sampled at each lag distance shown.

of the sample spacing (that is, 180, 360, 540, and so on). The uniform sampling scheme provided no spatial information for some distances at less than 500 m (Figure 4). The random sampling scheme sampled all lag distances better than the uniform design, but the number of plot pairs separated by less than 100 m was inadequate and therefore not effective (based on the confidence interval width). The cyclic sampling design sampled an adequate confidence interval width at all lag distances to quantify spatial variation with confidence interval widths less than 10% of gamma at scales ranging from 30 to 3000 m (Figure 5).

The cyclic sampling design included pairs of re-

dundant plots at specific lags because the diagonal pairs provided more information at certain distances that were not in the original one dimension and also some redundancy where the lag distance equals the cycle length (or multiples thereof). In contrast, all plots in the uniform sampling design occurred at a constant distance from each other. The random sampling design produced a spatial coverage comparable to the cyclic sampling design, but the random plot locations would be very difficult to relocate in a study site that is 900 ha in extent, even using GPS navigation. Also, the random sampling design required 1000 random plots to achieve confidence interval widths comparable to the 312 plots used in the cyclic sampling design. Because all of the sample plots are equally spaced, the uniform sampling design can only achieve the same spatial resolution by two means—either the spatial domain of the study has to be reduced so that all of the plots are only 30 m apart from each other, making the study area size 1/36th the original size, or the number of sampled plots must be increased to  $36 \times 312$  (11,232 plots).

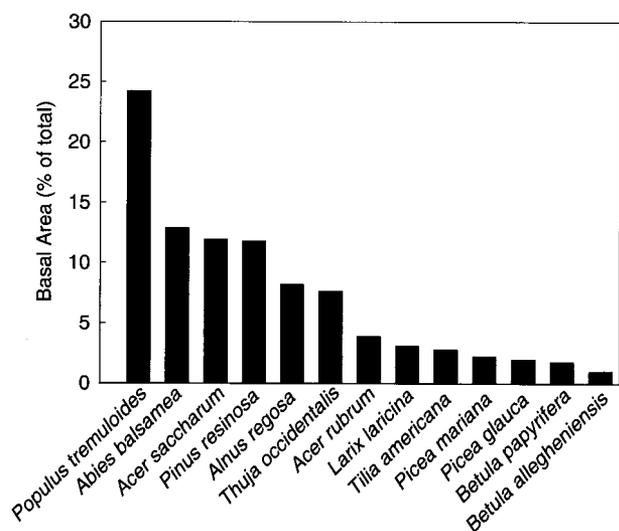
All three of the sampling designs adequately sampled the different vegetation cover types, as classified by WISCLAND (WIDNR 1998). Cover types that made up less than 5% of the landscape were often undersampled in all three sampling designs. If project goals were to sample all vegetation cover types, this underrepresentation may have been a concern. Given the errors in the WISCLAND classification (WIDNR 1998) and that management activities have changed the vegetation since the classification was performed (all jack pine stands in the study area were clear-cut in 1996 and 1997), we thought that we sampled the vegetation adequately with the cyclic sampling design. For other ecosystems, where small areas may be of larger concern, such as quantifying boreal wetlands and their trace gas emissions (Roulet and others 1997), a stratified cyclic sampling design that captures each cover type may be more desirable.

The benefits of the cyclic sampling design must be weighed against potential drawbacks (Table 4). There is sampling redundancy when the cycle repeats itself, but a uniform sampling design is far more redundant and the random design less so. A more important concern is that if the cycle length is the same as the spatial patterning on the ground, that pattern may not be detected. This problem can be avoided both with adequate sampling and by ensuring unbiased placement of sample locations. The cyclic sampling design is also more complicated to implement because designing the optimal scheme requires a priori information, thereby requiring users to understand the intricacies of the

**Table 4.** Qualitative Benefits and Drawbacks of the Three Sampling Schemes: Uniform, Random, and Cyclic

Issue	Uniform	Random	Cyclic
Ease of design and implementation	-	-	--
Problem if sampling pattern exactly matches pattern on landscape	--	0	--
Plots near each other produce hotspots of information on a Kriged prediction surface	0	++	-
No redundant pairs of points	--	0	+
Equal number of pairs of points across all lag distances	--	0	+
Maximizes variance of samples	--	0	++
Distance traveled to reach all points	--	--	++
Ease of navigation to each point	++	--	++

--, large negative attribute; -, small negative attribute; 0, neutral impact; +, small positive attribute; ++, large positive attribute



**Figure 6.** Percent basal area for the overstory tree species in 1248 variable radius subplots located in the tower footprint.

calculations and design. The additional costs associated with developing a strong cyclic sampling design are justified. Because some points are located near each other, hotspots of increased knowledge are created on a prediction surface where multiple points improved the local predictions.

### Definitions of Plots and Subplots

Subplot measurements, made at distances from 7.5 to 15 m apart from each other, were expected to influence lag distances only up to the 30-m plot size, but there were unexpected consequences. The

estimated range of spatial autocorrelation of LAI decreased from 205 to 147 m when we switched from plot-level average to subplot measurements, suggesting the individual subplot measurements provided useful information for describing the variogram shape. We concluded that it is important to maintain as much spatial information as possible and not to aggregate or average data either in the field or during analysis until after the geostatistical analyses have been completed.

### Site Vegetation Characteristics

Four forest cover types comprised 80% of the total area of the 312 plots used to characterize the footprint of the WLEF tower (Figure 6). The four dominant forest types were (a) lowland forested wetlands consisting of alder and northern white cedar; (b) trembling aspen, often mixed with a balsam fir understory; (c) northern hardwoods dominated by sugar maple, red maple, and basswood; and (d) upland red and jack pine plantations. Grass (meadows, clear-cuts, and so on) was the fifth major cover type. Roads, houses, and a mixture of different lesser cover types comprised the remaining land area.

The estimated autocorrelation range for effective LAI (as measured with the Li-Cor LAI-2000 and corrected for shoot-level clumping) was 147 m with an estimated sill and nugget of 2.16 and 0.53, respectively (Figure 7a). Including vegetation cover as a covariate decreased the overall estimated range to 117 m, with an estimated sill and nugget of 1.78 and 0.55, respectively (Figure 7b), implying a relationship between vegetation cover and LAI, as indicated in Figure 7b. Most of the oscillations in

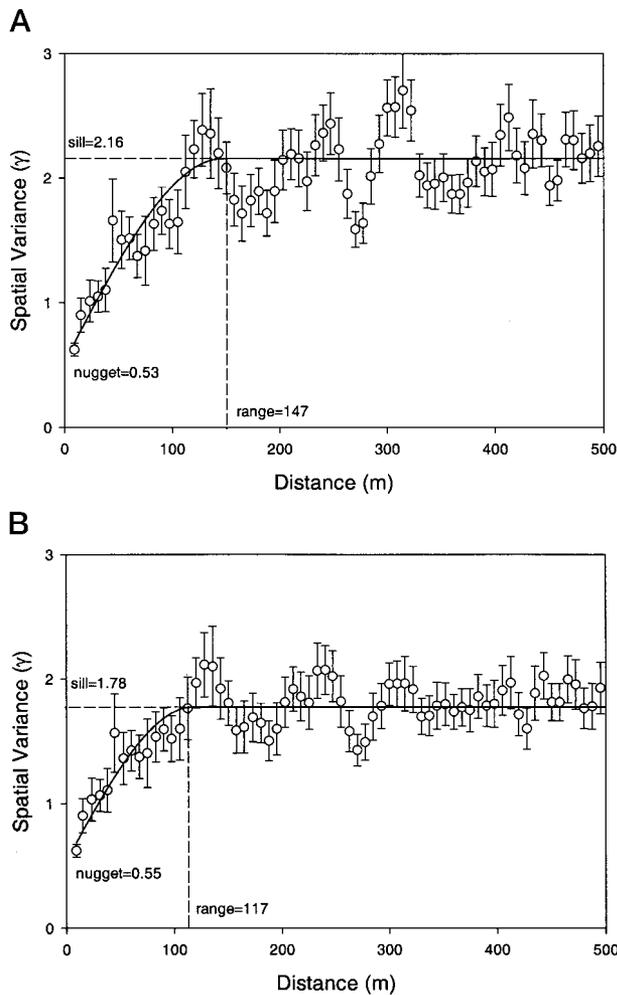


Figure 7. (A) Semi-variogram for the model, independent of vegetation, of effective leaf area index measured in July 1999. Open circles represent gamma estimates at different lag distances, with 99% confidence intervals; solid line represents fitted spherical model. (B) Semi-variogram for the vegetation dependant model of effective leaf area index measured in July 1999. Open circles represent gamma estimates at different lag distances, with 99% confidence intervals; solid line represents fitted spherical model.

gamma along the sill (Figure 7a and b) were attributed to random variation; note the 99% confidence intervals where very few points lie completely off the estimated spherical model. The oscillations in the variogram could be caused by spatial patterns on the landscape caused by glacial deposition, soil types, management activities, management unit size, and so on; however, the cause(s) cannot be determined directly (Radeloff and others 2000). Future research will use covariates such as management activities and biophysical variables to examine the potential causes of the spatial patterns of LAI.

The spatial autocorrelation patterns differed among the five main vegetation cover types: aspen had a mean LAI of 3.57 and a range of 80 m, forested wetlands had a mean LAI of 3.82 and a range of 159 m, grass had a mean LAI of 1.14 and a range of 22.5 m, northern hardwood forests had a mean LAI of 3.45 and a range of 91 m and, upland conifer forests had a mean LAI of 3.99 and a range of 53 m (Table 2). Recall that the range represented the distance at which LAI was correlated. Grasses had a very short range, suggesting that there are mostly fine-scale differences in LAI values, probably due to the short stature of the vegetation and the location where it occurred. The three upland forest cover types (aspen, hardwoods, and conifers) had moderate range estimates (53–91 m), probably caused by frequent management (thinning, timber stand improvement), which fragments a continuous canopy into patches. Forested wetlands had the largest range estimate (159 m) because human disturbances do not fragment the canopy as much.

Fassnacht and Gower (1997) measured the LAI of the six major forest types in north central Wisconsin and reported that the LAIs for jack pine (AQV and QAE habitat types), mixed conifer–hardwood forests (PMV and AVVib habitat types), and northern hardwood forests (ATD and AViO habitat types) averaged 2.16, 4.73, and 6.48, respectively. Gower and others (Forthcoming), using the 24 plots from Fassnacht and Gower (1997) and an additional 24 plots that included a greater combination of forest leaf habit (evergreen versus deciduous) and soil drainage classes, reported that LAI ranged from 2.1 to 8.4. The lower LAI values reported in this study are likely due to the lower water-holding capacity of the soils in the area as well as recent harvesting and thinning management practices.

### Spatial Estimates of LAI

An important aspect of the sampling design and spatial regression analysis is to develop predictions and to derive error estimates for the predictions. Based on our cyclic sampling design, we developed a map of the predicted LAI for the study site (Figure 8). The average LAI for the  $3 \times 2$  km area was  $3.51 \pm 0.89$  (with a minimum of 0 and a maximum of 6.35). Error estimates are an important tool for evaluating both the predictions and the sampling design itself. Therefore, we derived a map of standard errors using the same spatial regression as used to generate the map of predicted LAI (Figure 9). The standard errors decrease near sampled locations, especially when there are several sampled plots close together. Partly for this reason, the design is best used for interpolation within the study area, not extrapolation outside this region. If

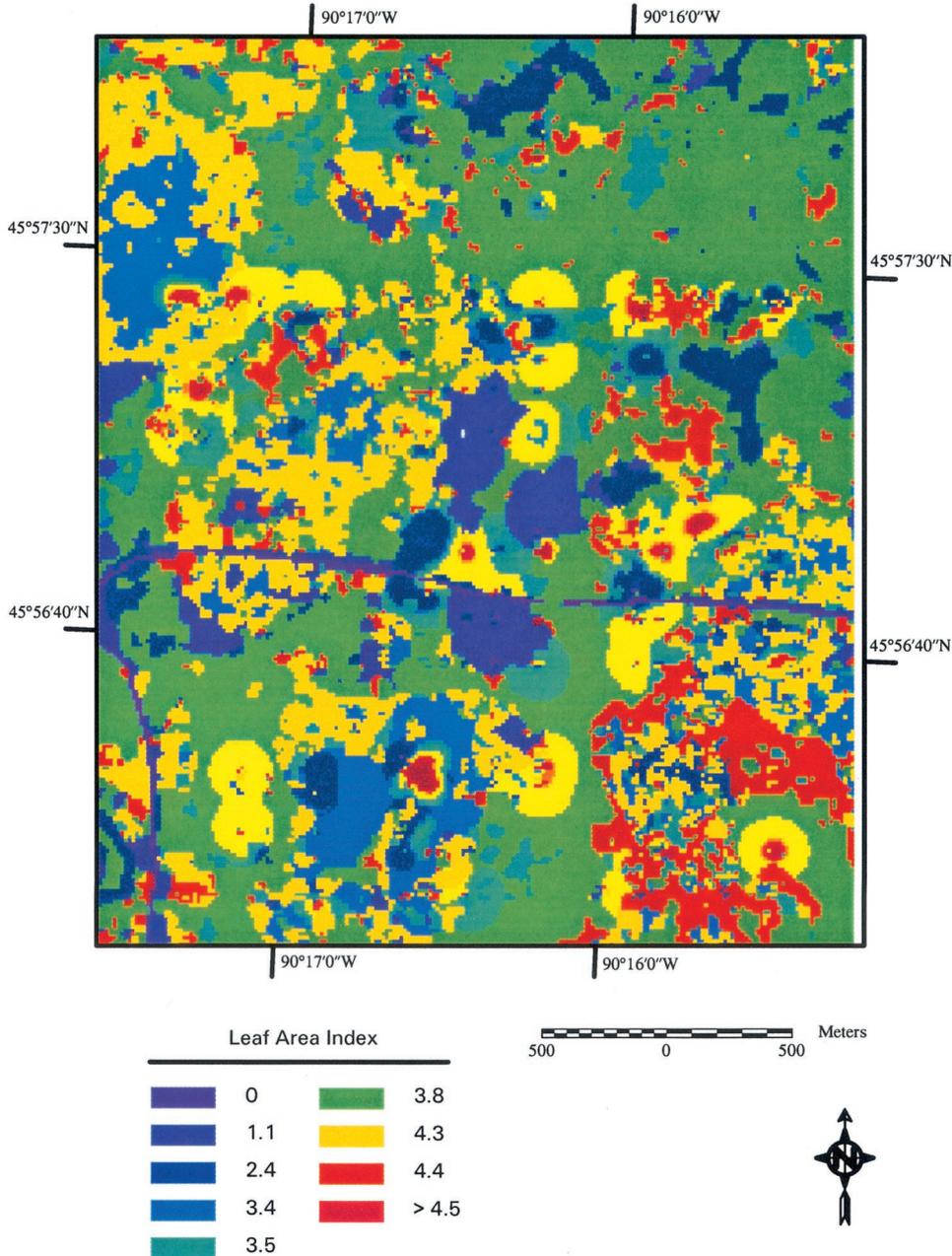


Figure 8. Kriged LAI surface map based on estimates of spatial patterns of effective LAI in four major forest cover types and the minor nonforest cover type.

the goal was only prediction, and the variogram was known (which was not the case in this study), then a different design may be more efficient.

### Remote Sensing and Geostatistics

Quantifying spatial heterogeneity is an essential component of the validation of MODIS products. Cover types that comprise a small fraction of the total landscape but are of importance are omitted as the scale of observation increases in a heterogenous landscape (Benson and MacKenzie 1995; Roulet and others 1997). The application of geostatistical analysis of data

will play an important role in increasing the understanding of how spatial heterogeneity differs among biomes, in accurately compositing pixels of the same size from a single sensor, or using different-sized pixels from different sensors. The approach described in this paper provides an alternative method to remote sensing for estimating LAI and vegetation cover for the footprint of eddy flux towers. Our approach may be a good alternative for terrestrial ecosystems that have LAIs that exceed the threshold at which many remotely sensed vegetation indexes saturate. Fassnacht and others (1997) showed that in the same

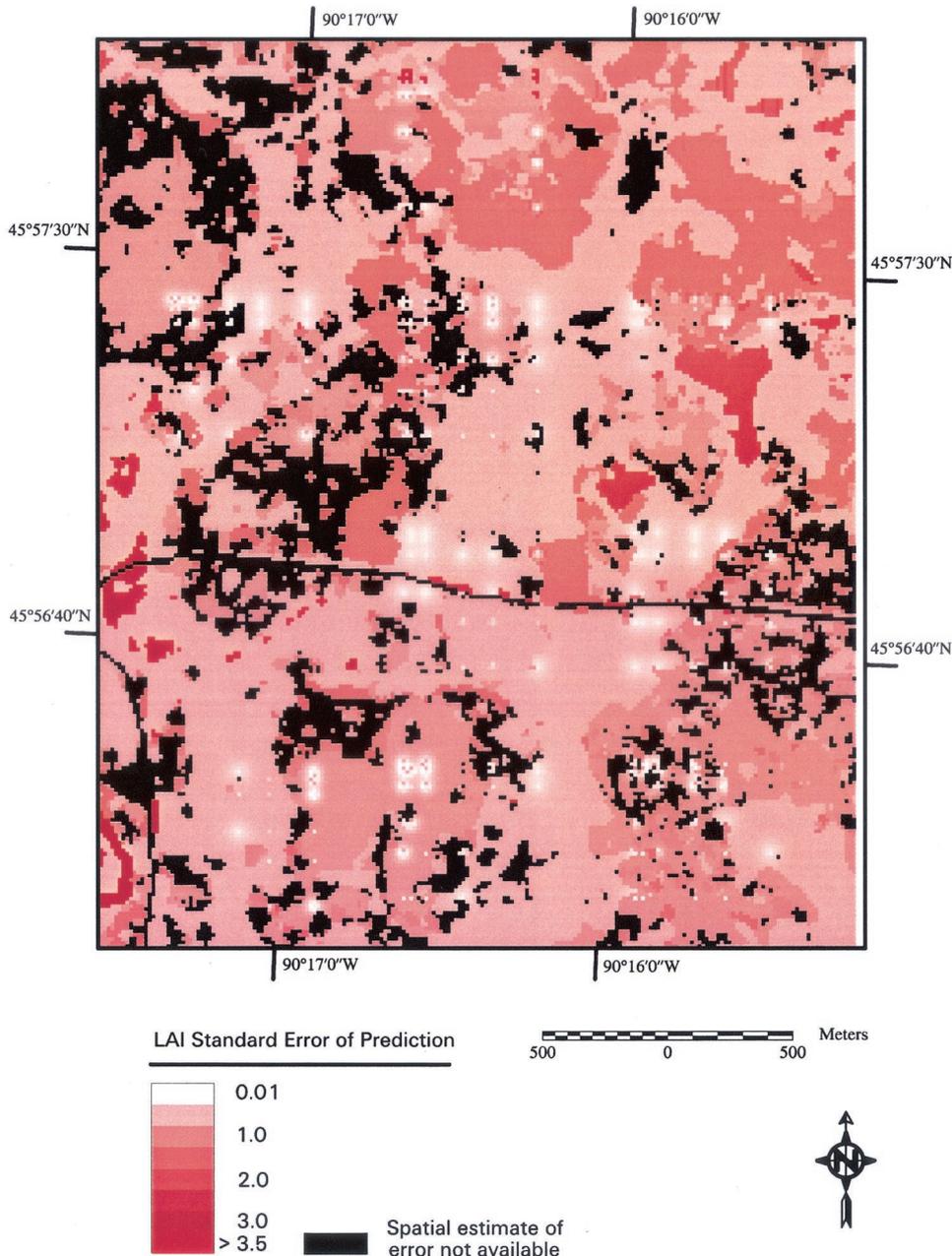


Figure 9. Kriged LAI standard error estimates based on estimates of spatial patterns of effective LAI in four major forest cover types and the minor nonforest cover type. Black regions contain cover types that were not spatially Kriged.

region as this study, high LAI values (more than 5) are difficult to predict using NDVI or SR (simple ratio) indexes derived from Landsat TM 5 images because of sensor saturation. Alternatively, geostatistics can be used in combination with remote sensing to adjust for locally high or low areas of LAI. The cyclic sampling design could provide a new approach for deriving other important ecosystem characteristics that cannot be estimated accurately via remote sensing. For example, soil carbon content is an important parameter that is needed to parameterize ecosystem models used to simulate carbon and water exchange for areas surrounding eddy flux towers.

### CONCLUSION

The sampling design first proposed by Clinger and Van Ness (1976) is applicable for ecological sampling schemes to quantify spatial patterns of vegetation characteristics. Accurate characterization of LAI is essential for parameterizing surface vegetation atmosphere transfer models that are routinely calibrated at eddy flux towers and used to simulate water, energy, and gas exchange between terrestrial ecosystems and the atmosphere. This study demonstrated the strength of the cyclic sampling design at the landscape level to quantify ecosystem charac-

teristics and adds additional knowledge and value to ground-based measurements.

## ACKNOWLEDGMENTS

The research was supported by a NASA EOS Validation grant (NAG5-6457) to S. T. G., J. M. N., and G. D. and a NASA BigFoot grant (NAG5-8069) to S. T. G. A McIntire-Stennis Grant provided additional support to S. T. G. and D. S. M. Linda Parker, Jud Isebrands, and other US Forestry Service scientist and managers at the Chequamegon-Nicolet National Forest and the North Central Research Station facilitated our research on national forests. Numerous undergraduate students assisted with the installation of the plots, periodic LAI measurements, litterfall collections, and processing. Dr. Edward Rastetter provided a wealth of comments that greatly improved the manuscript.

## REFERENCES

- Aubry P, Debouzie D. 2000. Geostatistical estimation variance for the spatial mean in two-dimensional systematic sampling. *Ecology* 81:543–53.
- Bakwin PS, Tans PP, Hurst DF, Zhao CL. 1998. Measurements of carbon dioxide on very tall towers: results of the NOAA/CMDL program. *Tellus Ser B-Chem Phys Meteorol* 50:401–15.
- Benson BJ, MacKenzie MD. 1995. Effects of sensor spatial-resolution on landscape structure parameters. *Landscape Ecol* 10:113–20.
- Bogaert P, Russo D. 1999. Optimal spatial sampling design for the estimation of the variogram based on a least squares approach. *Water Resources Res* 35:1275–89.
- Campbell JL, Burrows S, Gower ST, Cohen WB. 1999. Bigfoot: characterizing land cover, LAI, and NPP at the landscape scale for EOS/MODIS validation. Field manual version 2.1. Oak Ridge (TN): Environmental Sciences Division, Oak Ridge National Laboratory. 104 p.
- Chen J, Rich PW, Gower ST, Norman JM, Plummer S. 1997. Leaf area index of boreal forests: theory, techniques and measurements. *J. Geophys Res* 102(D24):29429–29443.
- Clayton MK, Hudelson BD. 1995. Confidence-intervals for autocorrelations based on cyclic samples. *J Am Stat Assoc* 90:753–7.
- Clinger W, Van Ness JW. 1976. On unequally spaced time points in time series. *Ann Stat* 4:736–45.
- Cohen WB, Justice CO. 1999. Validating MODIS terrestrial ecology products: linking in situ and satellite measurements. *Remote Sens Environ* 70:1–3.
- Cressie NA. 1993. *Statistics for spatial data*. New York: Wiley.
- Dutilleul P. 1993. Spatial heterogeneity and the design of ecological field experiments. *Ecology* 74:1646–58.
- Fassnacht KS, Gower ST. 1999. Comparison of the litterfall and forest floor organic matter and nitrogen dynamics of upland forest ecosystems in north central Wisconsin. *Biogeochemistry* 45:265–84.
- Fassnacht KS, Gower ST. 1997. Interrelationships among edaphic and stand characteristics, leaf area index, and aboveground net primary productivity for upland forest ecosystems in north central Wisconsin. *Can J For Res* 27:1058–67.
- Fassnacht KS, Gower ST, MacKenzie MD, Nordheim EV, Lillesand TM. 1997. Estimating the leaf area of north central Wisconsin forests using the Landsat Thematic Mapper. *Remote Sens Environ* 61:229–45.
- Fassnacht KS, Gower ST, Norman JM, McMurtrie RE. 1994. A comparison of optical and direct-methods for estimating foliage surface-area index in forests. *Agric For Meteorol* 71:183–207.
- Fortin M-J, Drapeau P, Legendre P. 1989. Spatial autocorrelation and sampling design in plant ecology. *Vegetatio* 83:209–22.
- Gower ST, Fassnacht KS, Mladenoff DJ, Burrows SN. Forthcoming. Landscape patterns of leaf area index, aboveground net primary production, and light use efficiency of forests in northern Wisconsin. *Landscape Ecol*.
- Gower ST, Kucharik CJ, Norman JM. 1999. Direct and indirect estimation of leaf area index,  $f_{APAR}$ , and net primary production of terrestrial ecosystems. *Remote Sens Environ* 70:29–51.
- Gower ST, Norman JM. 1991. Rapid estimation of leaf area index in conifer and broad-leaf plantations. *Ecology* 72:1896–900.
- Iachan R. 1982. Systematic-sampling — a critical-review. *Intl Sta Rev* 50:293–303.
- Justice CO, Starr D, Wickland D, Privette J, Suttles T. 1998. EOS validation coordination: an update. *Earth Observer* 10:55–60.
- Kabrick JM, Clayton MK, McBratney AB, McSweeney K. 1997. Cradle-knoll patterns and characteristics on drumlins in northeastern Wisconsin. *Soil Sci Soc Am J* 61:595–603.
- Kaluzny SP, Vega SC, Cardoso TP, Shelly A, editors. 1998. *S+Spatialstats: user's manual for Windows and Unix*. New York: Springer-Verlag.
- Legendre P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74:1659–73.
- Mackay DS, Ahl DE, Ewers BE, Gower ST, Barrows SN, Samanta J, Davis, KJ. Forthcoming. Effects of aggregated classifications of forest composition on estimates of evapotranspiration in a Northern Wisconsin forest. *Global Change Biology* (In press).
- Miller TF, Mladenoff DJ, Clayton MK. Forthcoming. Spatial autocorrelation and patterns of understory vegetation and environment in old-growth northern hardwood forests. *Ecol Monogr*
- Müller WG, Zimmerman DL. 1999. Optimal designs for variogram estimation. *Environmetrics* 10:23–37.
- Pettitt AN, McBratney AB. 1993. Sampling designs for estimating spatial variance-components. *Appl Stati* 42:185–209.
- Pinheiro JC, Bates DM. 2000. *Mixed-effects models in S and S-PLUS*. New York: Springer-Verlag.
- Radeloff VC, Miller TF, He HS, Mladenoff DJ. 2000. Periodicity in spatial data and geostatistical models: autocorrelation between patches. *Ecography* 23:81–91.
- Roulet NT, Crill PM, Comer NT, Dove A, Boubonniere RA. 1997. CO<sub>2</sub> CH<sub>4</sub> flux between a boreal beaver pond and the atmosphere. *J Geophys Res* 102:29313–9.
- Running SW, Baldochii DD, Cohen W, Gower ST, Turner D, Bakwin P, Hibbard K. 1999. A global terrestrial monitoring network, scaling tower fluxes with ecosystem modeling and EOS satellite data. *Remote Sens Environ* 70:108–27.
- Van Diggelen F. 1997. GPS and GPS+GLONASS RTK. Available online at: <http://www.ashtec.com>.
- [WIDNR] Wisconsin Department of Natural Resources. 1998. WISCLAND land cover (WLCGW930). Madison (WI): Wisconsin Department of Natural Resources.