ELSEVIER

Comparison of regression and geostatistical methods for mapping Leaf Area Index (LAI) with Landsat ETM+ data over a boreal forest

Mercedes Berterretche[a], Andrew T. Hudak[b,*], Warren B. Cohen[c], Thomas K. Maiersperger[a], Stith T. Gower[d], Jennifer Dungan[e]

[a]Department of Forest Science, Oregon State University, 321 Richardson Hall, Corvallis, Oregon 97331
[b]Rocky Mountain Research Station, USDA Forest Service, 1221 South Main Street, Moscow, Idaho 83843
[c]Pacific Northwest Research Station, USDA Forest Service, 3200 SW Jefferson Way, Corvallis, Oregon 97331
[d]Department of Forest Ecology and Management, University of Wisconsin, 1630 Linden Drive, Madison 53706
[e]NASA Ames Research Center, MS 242-4, Moffett Field, CA 94035-1000

Abstract

This study compared aspatial and spatial methods of using remote sensing and field data to predict maximum growing season leaf area index (LAI) maps in a boreal forest in Manitoba, Canada. The methods tested were orthogonal regression analysis (reduced major axis, RMA) and two geostatistical techniques: kriging with an external drift (KED) and sequential Gaussian conditional simulation (SGCS). Deterministic methods such as RMA and KED provide a single predicted map with either aspatial (e.g., standard error, in regression techniques) or limited spatial (e.g., KED variance) assessments of errors, respectively. In contrast, SGCS takes a probabilistic approach, where simulated values are conditional on the sample values and preserve the sample statistics. In this application, canonical indices were used to maximize the ability of Landsat ETM+ spectral data to account for LAI variability measured in the field through a spatially nested sampling design. As expected based on theory, SGCS did the best job preserving the distribution of measured LAI values. In terms of spatial pattern, SGCS preserved the anisotropy observed in semivariograms of measured LAI, while KED reduced anisotropy and lowered global variance (i.e., lower sill), also consistent with theory. The conditional variance of multiple SGCS realizations provided a useful visual and quantitative measure of spatial uncertainty. For applications requiring spatial prediction methods, we concluded KED is more useful if local accuracy is important, but SGCS is better for indicating global pattern. Predicting LAI from satellite data using geostatistical methods requires a distribution and density of primary, reference LAI measurements that are impractical to obtain. For regional NPP modeling with coarse resolution inputs, the aspatial RMA regression method is the most practical option.
Published by Elsevier Inc.

Keywords: Conditional simulation; Kriging; Reduced major axis regression

1. Introduction

1.1. Predicting Leaf Area Index (LAI)

Leaf area index (LAI) is a significant ecological attribute that controls physical and physiological processes in vegetation canopies (Landsberg & Gower, 1997; Waring & Running, 1998) and is widely used as input to biogeochemical process models that predict net primary production (NPP) over extensive terrestrial areas (Running et al., 1999, 1989; Running & Gower, 1991). For such purposes, LAI predictions are often needed as maps, which can be derived from remotely-sensed data using empirically derived regression relationships based on spectral vegetation indices (SVIs).

SVIs are calculated from reflectance data and, through regression, often related to field-based LAI measurements of the dominant canopy (Fassnacht et al., 1997; Peterson et al., 1987; Spanner et al., 1990a,b; Tucker, 1979; Turner et al.,

1999). The most commonly applied SVIs are the normalized difference vegetation index (NDVI) and the simple ratio (SR) (Chen & Cihlar, 1996; White et al., 1997). These and other ratio-based indices, although important, utilize only a fraction of the spectral information available in many image datasets (Cohen et al., 1995) and thus may limit the power of predictive relationships. For situations where understory reflectance and canopy closure are variable (Loechel et al., 1997; Spanner et al., 1990a), and where the understory and background materials contribute substantially to the reflectance signal received by the sensor (Huete et al., 1985; Nemani et al., 1993), the use of shortwave–infrared wavelengths has improved predictive power (Loechel et al., 1997; Nemani et al., 1993). One convenient and powerful means of incorporating multiple spectral bands into a single predictive index is canonical correlation analysis (CCA) (Cohen et al., 2003).

Most commonly, ecologists have estimated variables such as LAI and NPP without much regard for location, and used them in non-spatial compartment ecosystem models (Milne & Cohen, 1999). Among the input variables relevant to canopy processes that can be derived from remote sensing data are land cover type and LAI (Reich et al., 1999). Other spatial surfaces such as temperature and evapotranspiration, or accurate digital elevation models, are also commonly required as model inputs (Running & Nemani, 1987; Wilson et al., 2000). In a few instances, geostatistical methods are used to derive or improve such data layers (Goovaerts, 2000; Kyriakidis et al., 1999). Dungan (1998) compared regression and geostatistical methods for mapping vegetation variables, but using synthetic rather than actual vegetation data.

The goal of this study was to use remote sensing data and field LAI and tree cover data to obtain a maximum growing-season LAI map that could be input into an ecological process model for boreal evergreen needleleaf forest. The work presented is drawn from Berterretche (2002), where a larger set of comparisons was made. These included aspatial methods such as ordinary least squares (OLS) regression, inverse OLS regression, reduced major axis (RMA) regression (Curran & Hay, 1986), and spatial methods such as kriging, cokriging, kriging with an external drift (KED), and sequential Gaussian conditional simulation (SGCS) (Deutsch & Journel, 1998; Goovaerts, 1997). For the sake of economy, we selected from Berterretche (2002) three instructive approaches: RMA, KED, and SGCS to report here.

1.2. Regression methods

Ordinary least square (OLS) regression methods are commonly used to predict LAI from SVIs. OLS regression assumes that the errors are (spatially) independent and that there are no measurement errors in the "independent" variable (X). OLS regression is designed to estimate the "dependent" variable (Y), by minimizing the sum of squares errors in Y with respect to X.

Curran and Hay (1986) described the major measurement errors that should be accounted for in remote sensing variables, and that are generally ignored when using OLS regression. A host of orthogonal regression models exist that take into consideration errors in the measurement of all variables used (Van Huffel, 1997). One in particular, RMA regression, has exhibited promising results in studies by Curran and Hay (1986), Larsson (1993) and Cohen et al. (2003). Cohen et al. (2003) compared traditional OLS (Y on X), inverse OLS (X on Y), and RMA regression models for predicting LAI in an agro-ecosystem and cover percentage at the boreal forest system used in this study; they determined that RMA regression maintained the variance of the observations in the predictions, whereas OLS regression and inverse OLS regression reduced and inflated, respectively, the variance in the predictions. Preserving the variance structure of observed LAI is important if predicted LAI is to be used to drive an ecological process model, as in this study.

1.2.1. Reduced major axis (RMA)

The RMA method minimizes the sum of the cross-products of the differences on both axes, accounting simultaneously for the errors in both dependent and independent variables (Curran & Hay, 1986; Davis, 1986; Miller & Kahn, 1962). The RMA model was developed using the equation $Y = (\alpha + \beta X + \varepsilon)$, where the coefficients α (intercept), β (slope), and γ (error) are estimated. In OLS, α and β are determined by least squares, whereas with RMA, $\alpha = \bar{Y} - (\sigma_Y/\sigma_X)\bar{X}$, and $\beta = \sigma_Y/\sigma_X$. Here, $\bar{Y}$ and $\bar{X}$ are the means of Y and X and σ_Y and σ_X are their standard deviations.

1.3. Geostatistical methods

Geostatistics is concerned with a variety of techniques aimed at understanding and modeling spatial variability through prediction and simulation (Deutsch, 2002; Journel, 1989; Journel & Huijbregts, 1978; Goovaerts, 1997). Geostatistics exploits the presence of spatial autocorrelation and joint dependence in space and time that occur in most natural resource variables (Myers, 1997). In an ecological context, geostatistics have been used to describe the scale and pattern of spatial variability (Burrows et al., 2002; Legendre & Fortin, 1989; Rossi et al., 1992; Woodcock et al., 1988), to characterize canopy structure (Cohen et al., 1990; Hudak & Wessman, 1998; St-Onge & Cavayas, 1997; Wulder et al., 1998), to estimate continuous (Hudak et al., 2002) and categorical variables (Milne & Cohen, 1999; Rossi et al., 1993), and to assess risk (Myers, 1997; Saito & Goovaerts, 2000).

Unlike aspatial regression techniques mentioned above, geostatistics predicts values of a primary variable of interest, in this case LAI, using both measured values of that primary variable and a model of its spatial structure. Several geostatistical techniques exist to bring in information about

a related secondary variable or variables, in this case information from remote sensing, to help predict the primary. Along with this different formulation of the prediction approach come expanded models uncertainty that depend on the data values in addition to data configuration (Deutsch & Journel, 1998). Stochastic simulation is an example of a probabilistic geostatistical approach that provides a distribution of possible values for each cell of the surface, characterizing uncertainty. These uncertainty measurements can improve ecological interpretation, help assess error in a spatial context, and decrease losses and risks in policy and management decision-making (Rossi et al., 1993).

1.3.1. Kriging with an external drift (KED)

KED is a variant of kriging that allows for the use of secondary information known at every location (exhaustive), which is assumed to reflect the local spatial trend of the primary variable (Deutsch & Journel, 1998; Goovaerts, 1997). In a landscape, spatial variation can be decomposed into two components: large-scale variation and small-scale variation. The KED trend represents the large-scale variability of the primary variable. The residuals from the trend represent the small-scale variability, and the final KED result combines both. KED models the trend under the assumptions of a linear relationship between primary and secondary variables and smooth variation in the secondary variable. The distinctive feature of KED is that the algorithm employs a non-stationary random function model, where stationarity is limited within each search neighborhood, yielding more local detail than with ordinary kriging (Deutsch & Journel, 1998). The KED estimator is

$$Z^*_{\text{KED}}(u) = \sum_{\alpha=1}^{n(u)} \lambda^{\text{KED}}_\alpha(u) Z(u_\alpha)$$

where $Z^*_{\text{KED}}(u)$ is the KED estimator at location u, $\lambda^{\text{KED}}_\alpha(u)$ are the KED weights corresponding to the n samples at location u, and $Z(u_\alpha)$ are the sample values within the search neighborhood.

1.3.2. Sequential Gaussian conditional simulation (SGCS)

Stochastic simulation is a probabilistic approach that provides a distribution of multiple, equally probable realizations of the joint distribution of one or more variables in space, generating a model of spatial uncertainty (Goovaerts, 1997; Rossi et al., 1993). The collocated simple cokriging estimate of the primary variable (in this case, LAI) is

$$Z^*_{\text{SCK}}(u) = \sum_{\alpha_1=1}^{n_1(u)} \lambda^{\text{SCK}}_{\alpha_1}(u)[Z_1(u_{\alpha_1}) - m_1]$$
$$+ \lambda^{\text{SCK}}_{\alpha_2}(u)[Z_2(u) - m_2] + m_1,$$

where $Z^*_{\text{SCK}}(u)$ is the collocated simple cokriging estimator at location u, $\lambda^{\text{SCK}}_{\alpha i}$ are the collocated simple kriging weights, and m_1 and m_2 are global means of the primary and secondary variables, respectively.

Collocated cokriging uses the lag correlation function between the primary and secondary variables, following the Markov-type approximation, $\rho_{12}(h) \approx \rho_{12}(0) * \rho_{11}(h)$ where $\rho_{12}(h)$ is the lag cross-correlation function of the primary and secondary variables, $\rho_{12}(0)$ is the correlation coefficient between the primary and secondary variables, and $\rho_{11}(h)$ is the lag correlation function of the primary variable. In collocated cokriging, the dependence of the secondary variable on the primary one is limited to the collocated data, to avoid matrix instability problems caused by highly redundant secondary information and to speed up the process (Goovaerts, 1997).

2. Methods

2.1. Study site

The site is an approximately 1 km^2 area centered around the eddy covariance flux tower at the northern old black spruce (NOBS) site within the Boreal Ecosystem-Atmosphere Study (BOREAS) (Sellers et al., 1997), near Thompson, Manitoba, Canada. It is typical of the extreme northern boreal forest, having gentle terrain, containing a few lakes, and with abundant permafrost. The soils are derived from parent material deposited by Glacial Lake Agassiz. Vegetation consists primarily of black spruce (*Picea mariana*), with scattered tamarack (*Larix laricina*), balsam poplar (*Populus balsamifera*), and jack pine (*Pinus banksiana*). Canopy closure is highly variable, but generally less than 60% (Cohen et al., 2003). Stand ages are variable, up to 80 years old. A distinctive feature of the site is the abundance and diversity of bryophytes covering the soil (e.g., feathermoss, *Pleurozium* spp.; reindeer lichen, *Cladina mitis*), and a variety of grasses and shrubs (e.g., willow, *Salix* spp.; Labrador tea, *Ledum groenlandicum*). A more complete description of this site is given by Gower et al. (1997).

2.2. Sampling design

This study used a systematic spatial cluster sampling design based on Clinger and Van Ness (1976). The design was a spatial application of unequal but periodic intervals from a discrete time series that distribute pairs of plots at all separation distances (lags) (Burrows et al., 2002). The design maximizes information about the spatial variability of vegetation in heterogeneous landscapes by decreasing redundant measurements at constant lags (Burrows et al., 2002), and is more efficient than random samples (Fortin et al., 1989). The lags sampled in this nested design ranged from 25 to 500 m.
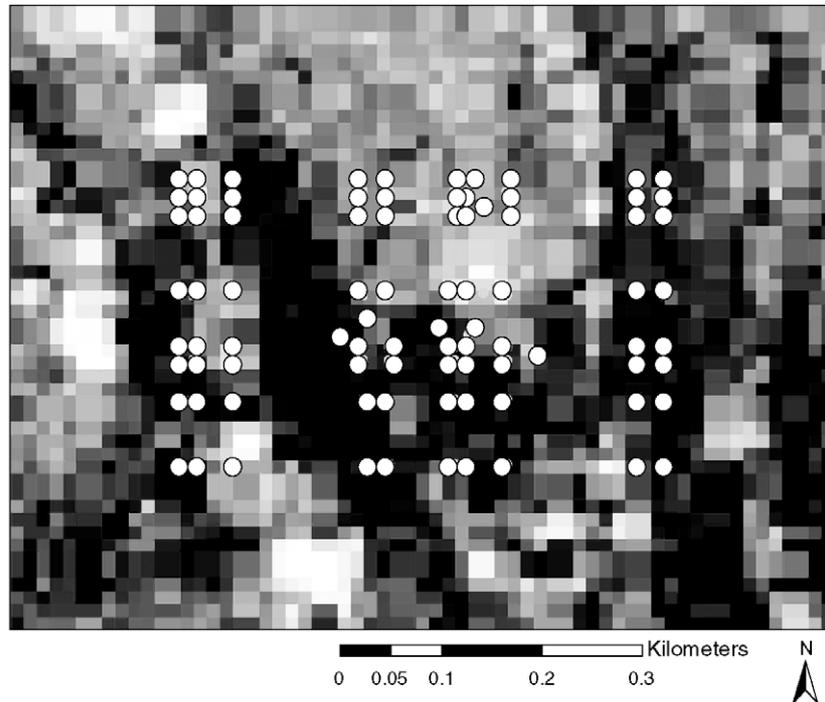
Fig. 1. Sample plot locations (dots) surrounding the eddy flux tower (not shown) at NOBS. Background image is a black and white transformation from a color composite (5,4,3) of the Landsat ETM+ image.

LAI and cover were measured at each of 86 plots (Fig. 1). Each plot was 25 m × 25 m with multiple sub-plots, over which measurements were averaged to produce a unique value per plot. All plots were geolocated using a real-time differential GPS with an accuracy of <0.5 m in both the *x* and *y* directions.

2.3. Field measurements

LAI of a vegetation canopy is the one-sided green leaf area per unit ground area; LAI is a critical parameter for modeling canopy gas and energy exchange with the atmosphere (Monteith & Unsworth, 1990). In this study, LAI was based on tree diameter-at-breast-height by density measurements and allometric equations developed by Gower et al. (1999). LAI was measured in a 5 point pattern like that on a die and cover in a 9 point (3 × 3) pattern. In both cases, spacing among plots was maximized and edges avoided by one-half the distance between sub-plots. Tree cover was quantified with an upward-looking, digital true-color camera fixed to a 1.70 m high monopod. Nine photographs were taken at each plot at a 30° view angle. The digital photos were enhanced with adjustments for color balance, contrast and saturation using standard, commercially available software. On each photo a reticular grid with 96 intersection points was used for cover determination. At each intersection point presence or absence of live tree cover was noted, the results transformed into percent live tree cover, then averaged over the nine non-overlapping photos per plot (Cohen et al., 2003).

2.4. Satellite image processing

Image processing was performed using ERDAS Imagine software (Atlanta, GA). This study used a Landsat ETM+ image (path 33/row 21) acquired on July 10, 1999; it had level 1 G processing, a 30 m cell size, and was projected in UTM coordinates (WGS84 datum). A 1 m panchromatic IKONOS image acquired on May 20, 2000 was georectified with the same projection parameters, using GPS points collected in the field with real-time differential correction and accurate to <0.5 m. This IKONOS image served as the base image for the ETM+ image, coregistered with a root mean square error <7 m using an automated routine (Kennedy & Cohen, 2003), then radiometrically resampled to 25 m using cubic convolution. The ETM+ raw digital numbers were transformed to percent reflectance using the COST model developed by Chavez (1996).

2.5. Canonical index calculation

Single pixel values were extracted from the six ETM+ reflectance bands at the 86 plots, and CCA was performed to develop spectral-based canonical indices (Cohen et al., 2003). Two canonical indices were developed: one relating LAI to the ETM+ bands and the other relating cover to the ETM+ bands. Each canonical index represented the linear combination of ETM+ bands having the greatest correlation with either LAI or cover (Johnson, 1998; Ramsey & Shafer, 1997), which facilitated subsequent RMA regression modeling (Cohen et al., 2003).
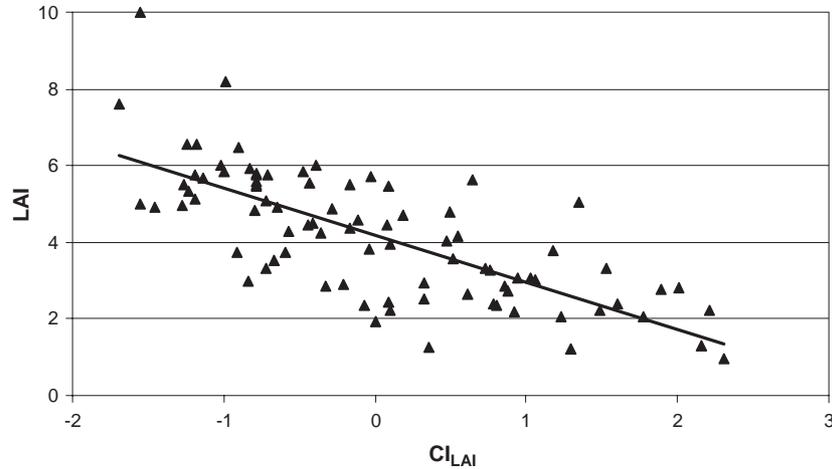
Fig. 2. Reduced Major Axis (RMA) regression model (trend line) for LAI and its Canonical Index (CI_{LAI}). See text for model coefficients.

Maps of the two canonical indices (CI_{LAI} and CI_{cover}) were created based on the equation

$$CI_v = \sum SCC_{b_x} \frac{b_x - \bar{b}_x}{s_{b_x}}$$

where CI_v is the canonical index for variable v (LAI or cover), s_{b_x} is the standardized canonical coefficient for band x, $\bar{b}_x$ is the mean of the band x values at 86 plots, and s_{b_x} is the standard deviation about this mean. The linear regression model used to predict LAI from CI_{LAI} by the RMA method (Fig. 2) was LAI = (4.19 − 1.68 * CI_{LAI}).

2.6. Geostatistical modeling

Geostatistical procedures were performed with GSLIB software (Deutsch & Journel, 1998). Data were transformed to normal scores prior to geostatistical analyses. The normal score transformation is non-linear and rank-preserving, matching the original data to a standard normal distribution (zero mean, unit variance). The purpose of assigning normal scores is to transform potentially skewed data distributions into Gaussian distributions to improve the applicability of a multi-Gaussian assumption in geostatistical modeling (Deutsch & Journel, 1998; Goovaerts, 1997). Prior to mapping results, predictions were back-transformed to approximate the original data distribution; that the predictions only approximate the original data is a caveat of the back-transformation, which unfortunately introduces bias (Saito & Goovaerts, 2000). Given that the original data in this study were not strongly skewed, the bias is not considered serious, but predictions must be interpreted with this caveat.

Location maps for LAI and cover measurements, histograms, scatterplots, semivariograms, cross-correlograms and Pearson's correlation coefficients for all variable pairs were explored to reveal important spatial and aspatial properties of the data. Directional semivariograms of the LAI normal scores were computed at 0, 10, 30, 50, 70, 90, 100, 120, 140, and 160°. The rose diagram (Fig. 3), a plot that indicates the distances at which directional semivariograms

reach a predetermined semivariance in selected directions, shows directions of maximum and minimum continuity and was used to construct anisotropic models.

Experimental semivariograms were modeled with positive linear combinations of three effects: nugget, spherical, and hole (Deutsch & Journel, 1998; Goovaerts, 1997). Geostatistical algorithms preferentially weight data closer to the location being predicted, within a specified search neighborhood. These local search neighborhoods limit the stationarity assumption to small areas, allow the calculation
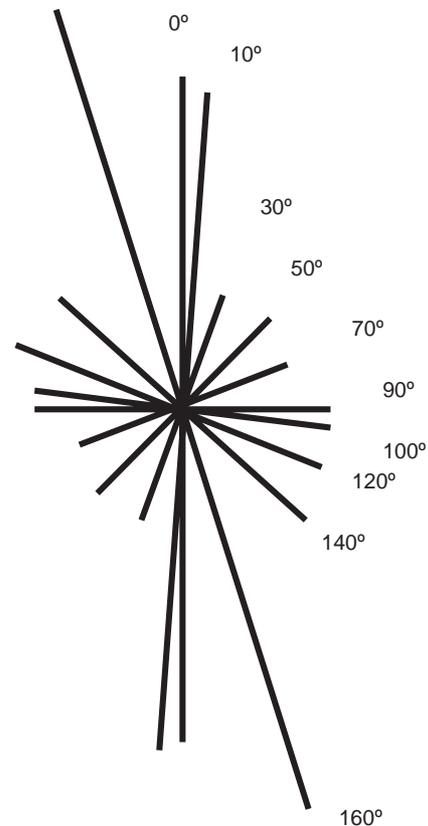

Fig. 3. Rose diagram, where each segment represents the same semivariance magnitude, allowing for visualization of the principal anisotropy axes.

of local trends and means required for some of the methods, and decrease computational time. GSLIB allows the user to manage search neighborhoods by changing their size, shape, and the amount of data to be used in the calculation.

2.6.1. Kriging with an external drift (KED)

The procedure for KED consisted of three steps (Goovaerts, 1997, page 197). First, the trend coefficients a_0^* and a_1^* of the trend model $m_{KED}^*(u)$ were evaluated within the search neighborhood from the $n(u)$ data pairs $(z_1(u), z_2(u))$, where $z_1(u)$ were the primary (LAI) sample data and $z_2(u)$ were the secondary (CI$_{cover}$) data. These

coefficients were estimated through the kriging system. Then, the trend components $m(u)$ were estimated at all primary sampled locations and at all other locations. Finally, simple kriging was performed on the residuals of the trend:

$$Z_{KED}^*(u) - m_{KED}^*(u) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha^{SK}(u)[Z(u_\alpha) - m_{KED}^*(u_\alpha)]$$

where $m_{KED}^*(u)$ was the trend component, estimated as

$$m_{KED}^*(u) = a_0^*(u) + a_1^*(u)z_2(u) \quad \vee \alpha = 1, \ldots \ldots, n(u)$$

where $Z(u_\alpha)$ are the sample data and λ_α^{SK} are the simple kriging weights. The anisotropic model used for the KED
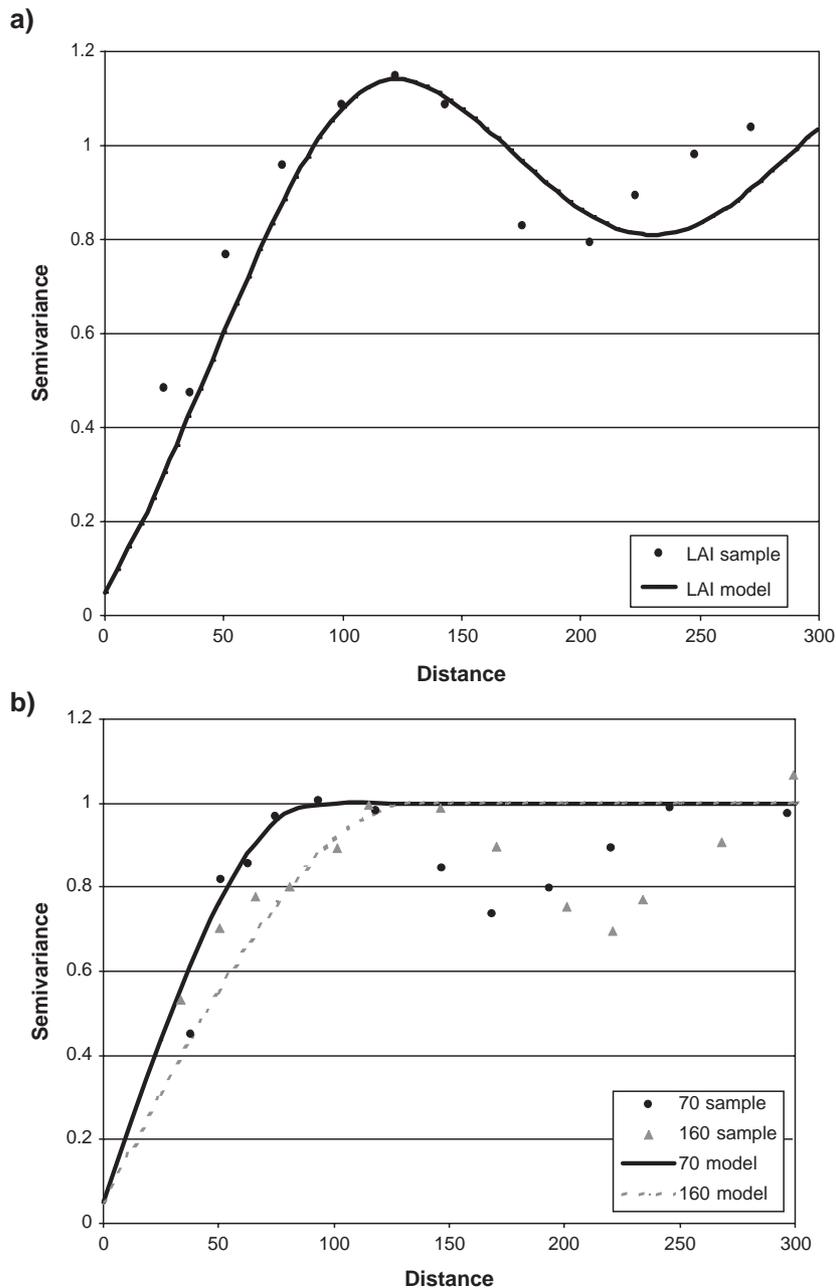

Fig. 4. (a) Experimental omnidirectional semivariogram used for isotropic KED and SGCS models; (b) experimental directional LAI semivariograms in 70° and 160° used for anisotropic Kriging with an External Drift (KED) and Sequential Gaussian Conditional Simulation (SGCS) models.

Table 1
Summary statistics for field LAI and cover data ($N=86$ plots)

Variable	Min	Max	Mean	Median	SD
LAI ($m^2\ m^{-2}$)	0.98	9.98	4.19	4.26	1.68
Cover (%)	0.00	63.89	38.55	38.77	16.10

was given by the combination of one model at azimuth 70° (minimum direction of continuity) and the other at azimuth 160° (maximum direction of continuity) (Fig. 4):

$$\gamma_{LAI_{KED}}(h) = 0.05 + 0.95*sph_{(h\ min=90,\ h\ max=135)}$$

2.6.2. Sequential Gaussian conditional simulation (SGCS)

SGCS with simple collocated cokriging of one secondary variable was performed using CI_{cover} as the secondary variable. The same anisotropic model used in KED (see above) was used in the SGCS algorithm to model LAI and 51 realizations were generated. A variance reduction factor of 0.85 was applied to ensure that the LAI normal score variance was close to one, to match the normal score variance of the KED model and thus facilitate their comparison.

The SGCS model produced a set of joint realizations of the spatial distribution of LAI, conditional to both the primary and collocated secondary variables. A Markov-type approximation for the linear model of coregionalization was used, so that only the semivariogram of the primary variable needed to be modeled. The cross-semivariogram model was derived as a linear rescaling of the primary variable semivariogram model, using the correlation coefficient between LAI and cover ($r=0.83$).

2.7. Model evaluation

Sample data values are preserved at their measured locations in geostatistical predictions, but are not preserved in the RMA regression. Therefore, summary statistics of the predictions cannot be used to directly compare the different methods. Also, partly because of the efficient field sampling design, there were an insufficient number of sample values to withhold for testing purposes. To check the consistency of the methods, we used cross-validation, a procedure in which each sample value is removed one at a time from the data set, and that location is predicted from the remaining data (Deutsch & Journel, 1998; Isaaks & Goovaerts, 1997; Srivastava, 1989; Wackernagel, 1998). For RMA and KED, cross-validation results were assessed with scatterplots of the observed versus predicted values, and of residuals versus predicted values to check for unbiasedness and homoscedasticity.

Cross-validation of the SGCS method was too cumbersome due to the many realizations (51), so the evaluation of the SGCS predictions was handled differently. Conditional variance was calculated to assess uncertainty. The conditional variance $\sigma^2(u_1)$ measures the spread of the conditional probability distribution around its mean $z_E^*(u_i)$, where

$$\sigma^2(u_i) = \sum_{l=1}^{N} \frac{\left(z^l(u_i)z^l(u_i)\right)}{N} - \left(\sum_{l=1}^{N} \frac{z^l(u_i)}{N}\right)^2$$

and $z^l(u_i)$ refers to a single realization of N realizations. Uncertainty increases as the spread of the probability distribution increases. The uncertainty of a probabilistic model was defined by Deutsch (2002) as the average conditional variance of all n locations in the area of interest

$$\varepsilon = \frac{1}{n} \sum_{i=1}^{n} \sigma^2(u_i); \varepsilon$$

is between 0 and 1 in normal space.

In summary, no single metric sufficed for comparing the RMA, KED and SGCS methods, yet a desirable overall evaluation was achieved by combining three approaches. First, the basic statistics of the predicted LAI distributions were compared to those from the measured LAI distribution. Second, omnidirectional and directional (70° and 160°) semivariograms were used to assess anisotropy and global variance. Third, the two spatial methods (KED and SGCS) were compared for how well they preserved the degree and pattern of spatial uncertainty.

3. Results

Measured LAI values ranged from 1 to 10 (Table 1). Mean LAI was 4.2, relative to a median of 4.3, indicating only a slight positive skew in the data. Tree cover ranged from 0% to 64% with a mean and median of 39%. The spatial distribution of LAI and cover values revealed a slight north–south trend in these attributes (Figs. 3 and 5).

Incomplete canopy closure permitted a high influence of non-tree cover components in the Landsat ETM+ reflectance signal. This, combined with the reflectance properties of black spruce crowns and shadows, yielded negative linear relationships for observed LAI and cover with all ETM+ reflective bands and canonical indices, CI_{LAI} and CI_{cover}

Table 2
Pearson correlation coefficients between Landsat ETM+ reflectance data and LAI and cover at 86 plots, and canonical weights of Landsat ETM+ bands on CI_{LAI} and CI_{cover}

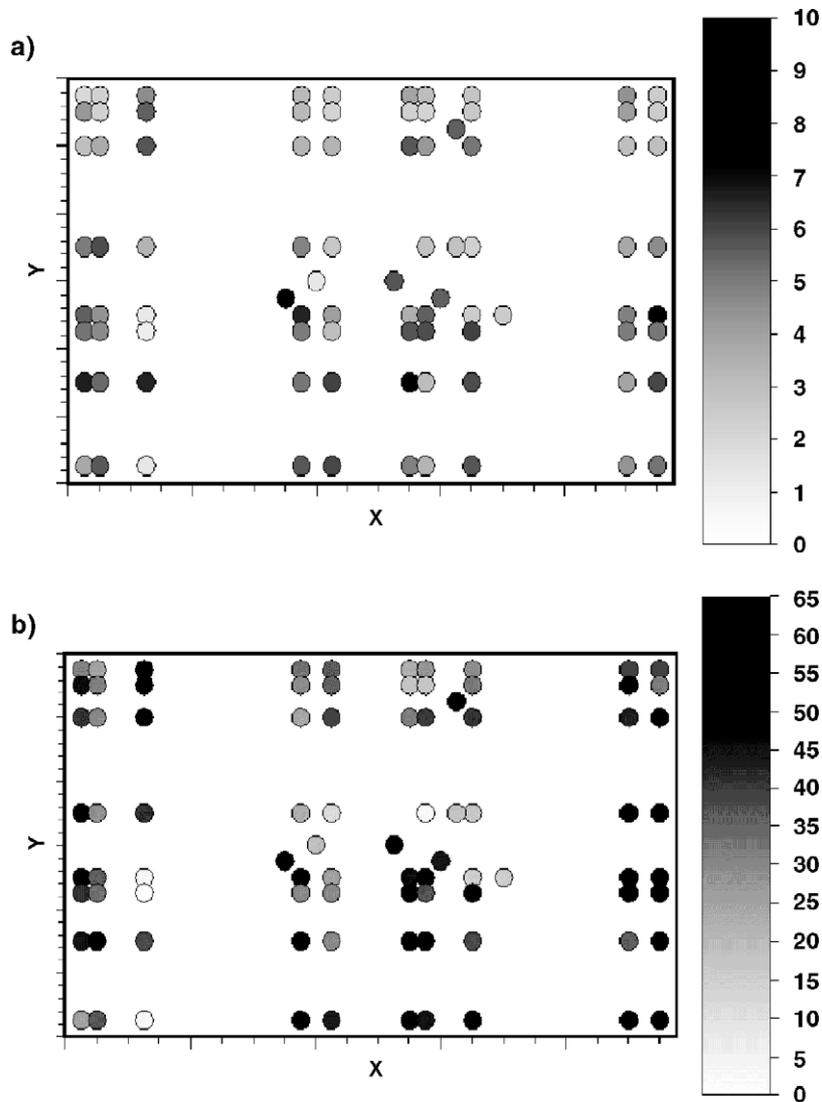| Variable	LAI	Cover	CI_{LAI}	CI_{cover}
Band 1	−0.29	−0.33	−0.05	−0.07
Band 2	−0.60	−0.73	−0.08	0.17
Band 3	−0.58	−0.70	−0.12	0.06
Band 4	−0.63	−0.74	−0.05	0.25
Band 5	−0.73	−0.78	1.15	0.20
Band 7	−0.69	−0.78	0.09	0.41
CI_{LAI}	−0.73	−0.77	–	–
CI_{cover}	−0.70	−0.80	–	–

Fig. 5. Location and magnitude of (a) LAI, and (b) cover, for each plot.

(Table 2). ETM+ bands 5 and 7 had the strongest correlations with LAI and cover and consequently the heaviest influence on CI_{LAI} and CI_{cover} (Table 2).

Cross-validation of RMA and KED model predictions (Fig. 6) revealed that RMA was a better overall model for LAI values up to 7. For values higher than 7, the RMA
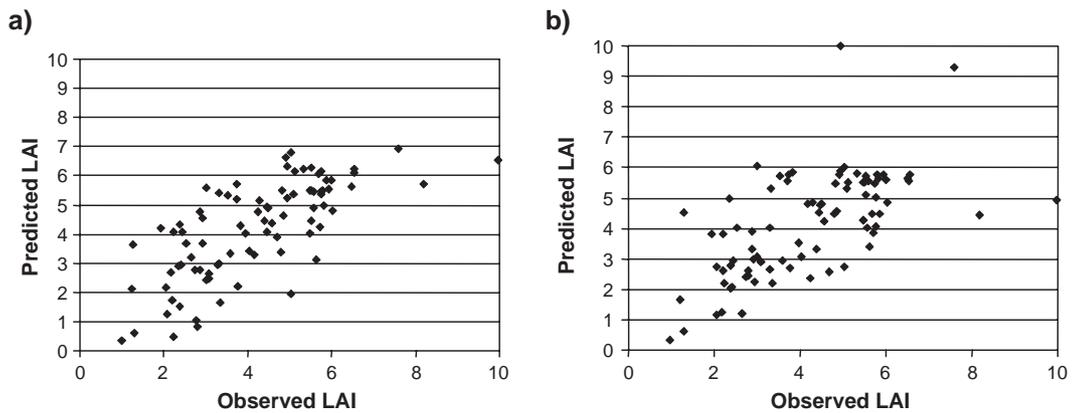

Fig. 6. Cross-validation results for (a) Reduced Major Axis (RMA) regression, and (b) Kriging with an External Drift (KED).

Table 3
Summary statistics of LAI predictions from the RMA and KED models (Fig. 7), a single SGCS realization (Fig. 7), and across all 51 SGCS realizations

Method	Min	Max	Mean	Median	SD	RMSE	r^2
RMA	0.37	6.94	4.19	4.42	1.65	1.23	0.72
Anisotropic KED	0.33	9.98	4.15	4.36	1.69	1.44	0.63
SGCS (1 realization)	0.07	10.00	4.19	4.25	1.68	–	–
SGCS (51 realizations)	0.74	9.98	4.16	4.09	1.27	–	–

model was asymptotic, whereas the KED model was extremely variable. Summary statistics for these predictions (Table 3) indicate that both models retained the mean LAI (4.2) and standard deviation (1.7) of the field measurements (Table 1). The median values were within 0.2 of the field-measured median. Minimum values were somewhat lower than the field-measured minimum value, but only RMA had a reduced maximum. Simulation data at plot locations from a randomly selected SGCS realization (and for the mean of all 51 realizations) were used for comparison, yielding quite similar statistics to those of the other two methods (Table 3).

LAI surfaces were developed for each model (Fig. 7) by applying the equations described in Section 2.5. For comparison, the CI$_{LAI}$ surface is shown. In terms of overall patterns observed, KED best exhibited the observed global, north–south trend in the LAI data (Fig. 5). All three methods revealed anisotropy, as observed in field-measured LAI (Fig. 3) and the CI$_{LAI}$ semivariograms (Fig. 8). However, KED showed a reduced anisotropy and lower global variance (i.e., lower sill), relative to RMA and SGCS. The SGCS and RMA surfaces more closely reflected the actual anisotropic pattern evident in the CI$_{LAI}$ surface (Fig. 8) and the original image (Figs. 1 and 7).

The SGCS model run using an anisotropic semivariogram model and CI$_{cover}$ as a secondary variable resulted in a mean LAI uncertainty of 0.39. To illustrate the different variance behaviors between predicted and simulated surfaces, an east–west transect that goes through 10 field plots was sliced from both the KED and SGCS surfaces (Fig. 9). Consistent with theory, the KED estimated error variance was dependent on the data configuration only, where the variance decreased when approaching the sample locations, and increased away from them. In contrast, SGCS condi-
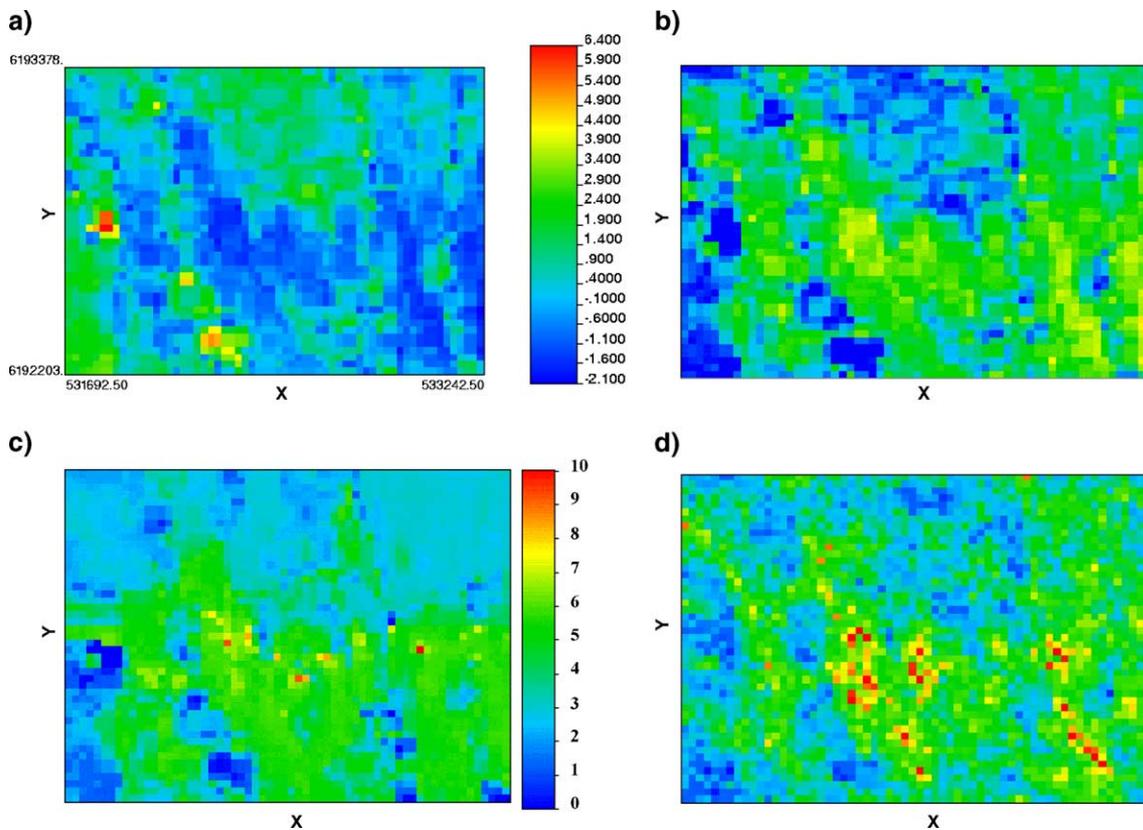

Fig. 7. Maps of the (a) Canonical Index for LAI (CI$_{LAI}$) image and LAI surfaces predicted using: (b) Reduced Major Axis (RMA) regression, (c) Kriging with an External Drift (KED), and (d) Sequential Gaussian Conditional Simulation (SGCS). The CI$_{LAI}$ image indicates the UTM X and Y extent of the study area (meters) and helps for visually assessing the patterns in the predicted LAI surfaces.
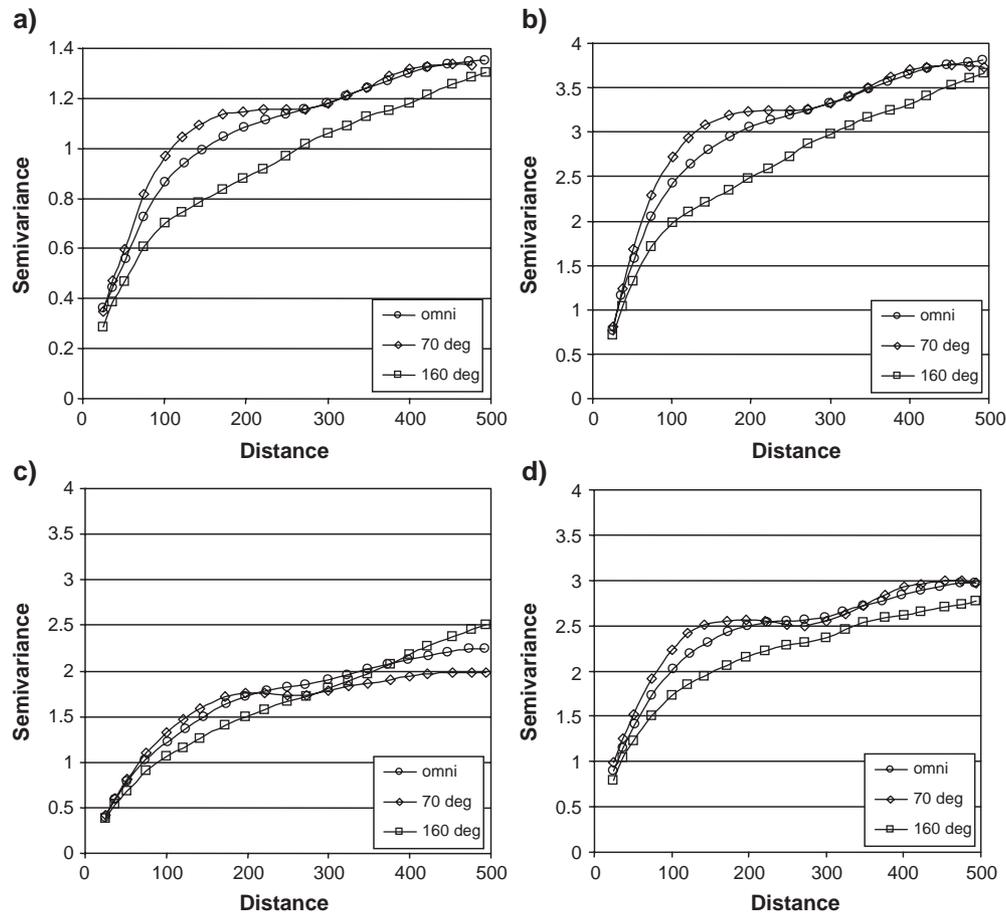
Fig. 8. (a) Omnidirectional and (b) directional (70° and 160°) semivariograms derived from the Canonical Index for LAI (CI_{LAI}) image and predicted LAI surfaces in Fig. 7.

tional variance depended on the data configuration and the data values, as is also predicted by theory. Conditional variance was greater when two adjacent samples had dissimilar LAI values, and smaller when adjacent values were more alike.

4. Discussion and conclusions

4.1. Relationship of field and remote sensing data

Chen and Cihlar (1996), Loechel et al. (1997) and Turner et al. (1999) stated that when canopy closure is low, LAI and near-infrared reflectance have virtually no relationship. In this study, bands 3 and 4 both showed a negative relationship with LAI, thus limiting the utility of common SVIs like NDVI and SR. The canonical indices improved markedly the ability of the spectral data to account for the variability observed in LAI over SR and NDVI; e.g., the R^2 improved from <0.01 for LAI and NDVI to 0.54 for LAI and CI_{LAI}. Both geostatistical models used CI_{cover} as a secondary variable and reproduced an acceptable pattern, suggesting that the information provided by CI_{cover} was useful in accounting for LAI

spatial variability. We conclude that canonical indices are much more useful than SVIs for empirical modeling strategies as were tested in this study.

4.2. Model comparison

Comparing the basic statistics of the predicted LAI distributions was not a very useful evaluation technique in this study. All three methods preserved the basic statistics of the field measurements (i.e., mean, median, and standard deviation), but this contradicted theory in the case of KED, which would normally lower the standard deviation. The minimum value was under-predicted by all three methods and RMA under-predicted the upper range of LAI. Both the RMA and KED models provided acceptable RMSE values and correlations, but SGCS could not be evaluated in the same way. From these statistics alone, it is not clear that there are meaningful advantages and disadvantages among the methods for deriving LAI surfaces to be used as input to NPP models. For that, we must examine the spatial properties of the surfaces.

Simulations were developed initially to provide measures of spatial uncertainty, but these have increasingly been used as maps of the variable of interest in cases
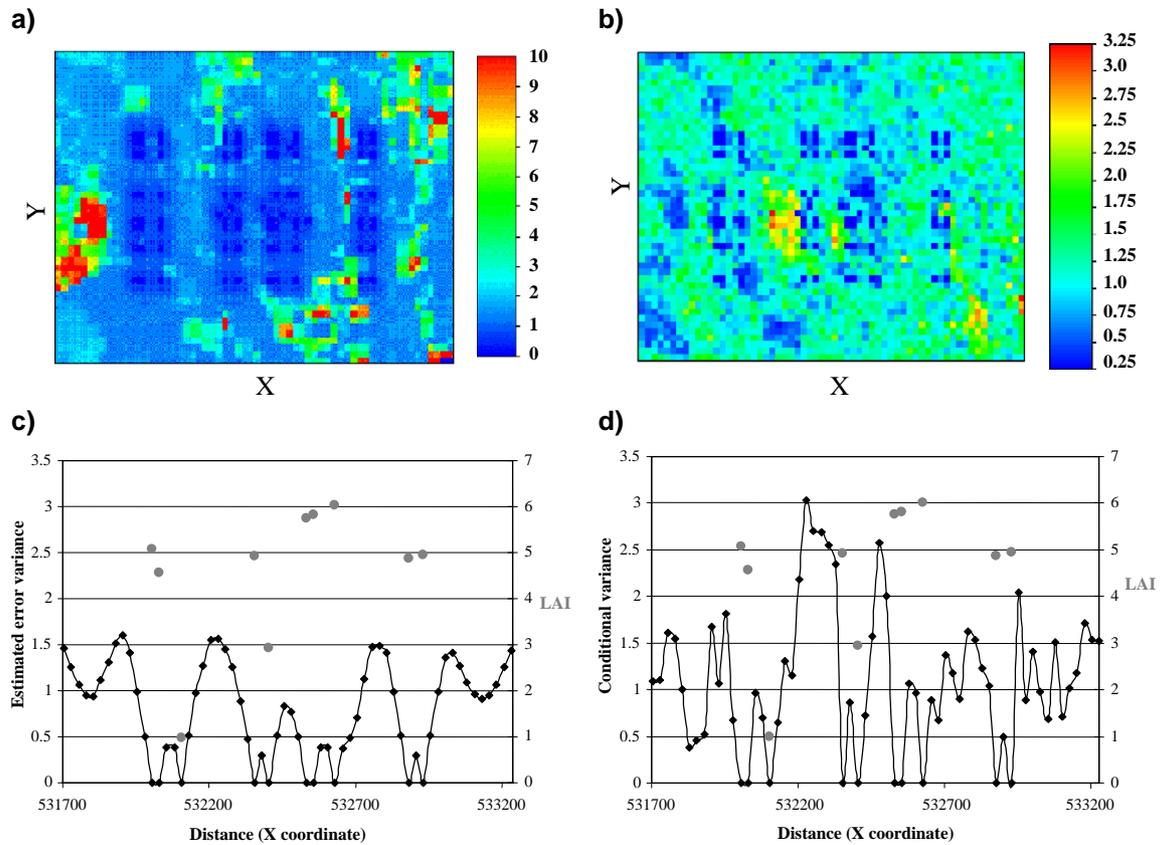
Fig. 9. (a) Kriging with an External Drift (KED) variance, and (b) Sequential Gaussian Conditional Simulation (SGCS) conditional variance; values along the transect (shown as a black line) are plotted below. (c) LAI sample values (dots) and the corresponding KED estimated error variance, and (d) LAI sample values (dots) and the corresponding SGCS conditional variance.

where the reproduction of the spatial variability is more important than local accuracy and where sample and exhaustive data are available (Deutsch, 2002). A good spatial uncertainty map should depend on the values of the samples, the distances among them and their geometry. The conditional variance of the whole set of realizations (Fig. 9, right) provided a visual and quantitative spatial uncertainty measure. While kriging may be useful in cases where local accuracy is important, simulation was better for this application where global continuity was more important.

4.3. LAI maps for NPP models

The production of LAI surfaces as inputs for NPP modeling requires a reliable mean and variance reproduction for each biome. The sampling design plays a relevant role in this task and we strongly recommend a basic a priori anisotropy analysis of the data, such as a rose diagram, and an unsupervised classification and/or a feature space analysis to evaluate the representativeness of the sampling design as first steps in similar studies.

This study tried to match the ground-based measurement support to Landsat ETM+ pixel size. Averages over the support surface were used, which "regularizes" the semi-

variograms. The effects of regularization are the same as those for aggregation: the variance is reduced, the range increases, and the mean remains unchanged (Isaaks & Srivastava, 1989; Woodcock et al., 1988). The reduction of the variance may be the most important aspect among these three, and the choice to use prediction or simulation to produce the LAI surfaces would consequently play an important role in the final quality of the NPP surfaces.

The semivariogram sills are associated with variability in the spatial patterns of the LAI surfaces derived from the three models (Fig. 8). RMA and SGCS did a better job of preserving the anisotropy and variance structure of the data, because of the smoothing effect KED has on the variance (Figs. 8 and 9). As a result, RMA and SGCS appear more promising for scaling up from Landsat ETM+ (30 m) to Moderate Resolution Imaging Spectrometer (MODIS, 250–1000 m) spatial resolutions. The set of SGCS realizations could also serve as input for sensitivity analysis to assess how LAI variability affects NPP process models.

Looking at the semivariogram ranges (Fig. 4), processes regulating LAI at this particular forest are relevant at distances up to 150 m. Aggregating Landsat-based LAI maps to MODIS resolutions (250 to 1000 m) would imply a certain amount of unknown error that would need to be quantified to assess the relevance of the final NPP maps.

This would depend on adequate support in the observations (Heuvelink, 1998), which is problematic for two reasons. First, it would be impractical to replicate beyond the NOBS tower footprint the density of adequately distributed LAI observations required to use geostatistical methods. Second, the controls on LAI variability are heavily localized to sub-pixel (MODIS) areas, so attempting to predict LAI directly from MODIS imagery using SGCS or any other geostatistical approach would result in a pure nugget effect, leaving RMA regression as the best alternative. For characterizing spatial pattern in LAI it is necessary to model at the intermediate resolution of Landsat before aggregating LAI predictions to predict NPP using the coarse resolution of MODIS.

Acknowledgements

References

Berterretche, M. (2002). *Comparison of regression and geostatistical methods to develop LAI surfaces for NPP modeling.* Master of Science thesis in Forest Science (219 p.). Corvallis, Oregon State University.

Burrows, S., Gower, S., Clayton, M., Mackay, D., Ahl, D., Norman, J., et al. (2002). Application of geostatistics to characterize leaf area index (LAI) from flux tower to landscape scales using a cyclic sampling design. *Ecosystems, 5,* 667–679.

Chavez, P. (1996). Image-based atmospheric corrections—revisited and improved. *Photogrammetric Engineering & Remote Sensing, 62,* 1025–1036.

Chen, J., & Cihlar, J. (1996). Retrieving LAI of boreal conifer forests using Landsat TM images. *Remote Sensing of Environment, 55,* 153–162.

Clinger, W., & Van Ness, J. (1976). On equally spaced time points in time series. *The Annals of Statistics, 4,* 736–745.

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

Cohen, W. B., Spies, T. A., & Bradshaw, G. (1990). Semivariograms of digital imagery for analysis of conifer canopy structure. *Remote Sensing of Environment, 35,* 167–178.

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

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

Davis, J. (1986). *Statistics and data analysis in Geology* (2nd ed.). New York: John Wiley & Sons.

Deutsch, C. (2002). *Geostatistical reservoir modeling.* New York: Oxford University Press.

Deutsch, C., & Journel, A. (1998). *GSLIB: Geostatistical Software Library and user's guide* (2nd ed.). New York: Oxford University Press.

Dungan, J. (1998). Spatial prediction of vegetation quantities using ground and image data. *International Journal of Remote Sensing, 19,* 267–285.

Fassnacht, K., Gower, S., MacKenzie, M., Nordheim, E., & Lillesand, T. (1997). Estimating the LAI of North Central Wisconsin forests using the Landsat thematic mapper. *Remote Sensing of Environment, 61,* 229–245.

Fortin, M., Drapeau, P., & Legendre, P. (1989). Spatial autocorrelation and sampling design in plant ecology. *Vegetatio, 83,* 209–222.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation.* New York: Oxford University Press.

Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology, 228,* 113–129.

Gower, S. T., Kucharik, C., & Norman, J. (1999). Direct and indirect estimation of leaf area index, fAPAR, and net primary production of terrestrial ecosystems. *Remote Sensing of Environment, 70,* 29–51.

Gower, S. T., Vogel, J. G., Norman, J. M., Kucharik, C. J., Steele, S. J., & Stow, T. K. (1997). Carbon distribution and aboveground net primary production in aspen, jack pine, and black spruce stands in Saskatchewan and Manitoba, Canada. *Journal of Geophysical Research, 102,* 29029–29041.

Heuvelink, G. (1998). *Error propagation in environmental modeling with GIS.* London: Taylor & Francis.

Hudak, A., Lefsky, M., Cohen, W., & Berterretche, M. (2002). Integration of lidar and landsat ETM data for estimating and mapping forest canopy height. *Remote Sensing of Environment, 82,* 397–416.

Hudak, A., & Wessman, C. (1998). Textural analysis of historical aerial photography to characterize woody plant encroachment in South African savanna. *Remote Sensing of Environment, 66,* 317–330.

Huete, A., Jackson, R., & Post, D. (1985). Spectral response of a plant canopy with different soil backgrounds. *Remote Sensing of Environment, 17,* 37–53.

Isaaks, E., & Srivastava, R. (1989). *An introduction to applied geostatistics.* New York: Oxford University Press.

Johnson, D. (1998). *Applied multivariate methods for data analysis* (1st ed.). Pacific Grove, CA: Duxbury Press.

Journel, A. (1989). *Fundamentals of geostatistics in five lessons.* Washington, DC: American Geophysical Union.

Journel, A., & Huijbregts, C. (1978). *Mining geostatistics.* New York: Academic Press.

Kennedy, R. E., & Cohen, W. B. (2003). Automated designation of tie-points for image-to-image registration. *International Journal of Remote Sensing, 24,* 3467–3490.

Kyriakidis, P., Shortridge, A., & Goodchild, M. (1999). Geostatistics for conflation and accuracy assessment of digital elevation models. *International Journal of Geographical Information Science, 13,* 677–707.

Landsberg, J., & Gower, S. (1997). *Applications of physiological ecology to forest management.* San Diego, CA: Academic Press.

Larsson, H. (1993). Linear regressions for canopy cover estimation in Acacia woodlands using Landsat-TM, -MSS, and SPOT HRV XS data. *International Journal of Remote Sensing, 14,* 2129–2136.

Legendre, P., & Fortin, M. (1989). Spatial pattern and ecological analysis. *Vegetatio, 80,* 107–138.

Loechel, S., Walthall, C., Brown de Colstoun, E., Chen, J., Markham, B., & Miller, J. (1997). Variability of boreal forest reflectances as measured from a helicopter platform. *Journal of Geophysical Research, 102,* 29495–29503.

Miller, R., & Kahn, J. (1962). Statistical Analysis in the Geological Sciences. New York: John Wiley & Sons.

Milne, B., & Cohen, W. B. (1999). Multiscale assessment of binary and continuous landcover variables for MODIS validation, mapping, and modeling applications. *Remote Sensing of Environment, 70,* 82–98.

Monteith, J., & Unsworth, M. H. (1990). *Principles of environmental physics.* London: Edward Arnold, 291 pp.

Myers, J. (1997). *Geostatistical error management: Quantifying uncertainty for environmental sampling and mapping*. New York: Van Nostrand Reinhold.

Nemani, R., Pierce, L., Running, S., & Band, L. (1993). Forest ecosystem processes at the watershed scale: Sensitivity to remotely sensed leaf area index estimates. *International Journal of Remote Sensing, 14*, 2519–2534.

Peterson, D., Spanner, M., Running, S., & Teuber, K. (1987). Relationship of thematic mapper simulator data to leaf area index of temperate coniferous forests. *Remote Sensing of Environment, 22*, 323–341.

Ramsey, F., & Shafer, D. (1997). *The statistical sleuth: A course in methods of data analysis* (1st ed.). Pacific Grove, CA: Duxbury Press.

Reich, P., Turner, D., & Bolstad, P. (1999). An approach to spatially distributed modeling of net primary production (NPP) at the landscape scale and its application in validation of EOS NPP products. *Remote Sensing of Environment, 70*, 69–81.

Rossi, R., Borth, P., & Tollefson, J. (1993). Stochastic simulation for characterizing ecological spatial patterns and appraising risk. *Ecological Applications, 3*, 719–735.

Rossi, R., Mulla, D., Journel, A., & Franz, E. (1992). Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs, 62*, 277–314.

Running, S., Baldocchi, D., Gower, S., Bakwin, P., & Hibbard, K. (1999). A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS data. *Remote Sensing of Environment, 70*, 108–127.

Running, S., & Gower, S. (1991). FOREST-BGC, a general model of forest ecosystem processes for regional applications: II. Dynamic carbon allocation and nitrogen budgets. *Tree Physiology, 9*, 147–160.

Running, S., & Nemani, R. (1987). Extrapolation of synoptic meteorological data in mountainous terrain and its use in simulating forest evapotranspiration rate and photosynthesis. *Canadian Journal of Forest Research, 17*, 472–483.

Running, S., Nemani, R., & Peterson, D. (1989). Mapping regional forest evapotranspiration and photosynthesis by coupling satellite data with ecosystem simulation. *Ecology, 70*, 1090–1101.

Saito, H., & Goovaerts, P. (2000). Geostatistical interpolation of positively skewed and censored data in a dioxin-contaminated site. *Environmental Science & Technology, 34*, 4224–4235.

Sellers, P., Hall, F., Kelly, R., Black, A., Baldocchi, D., Berry, J., et al. (1997). BOREAS in 1997: Experiment overview, scientific results, and future directions. *Journal of Geophysical Research, 102*, 28731–28769.

Spanner, M., Pierce, L., Peterson, D., & Running, S. (1990a). Remote sensing of temperate coniferous forest leaf area index: The influence of canopy closure, understory vegetation and background reflectance. *International Journal of Remote Sensing, 11*, 95–111.

Spanner, M., Pierce, L., Running, S., & Peterson, D. (1990b). The seasonality of AVHRR data of temperate coniferous forests: Relationships with LAI. *Remote Sensing of Environment, 33*, 97–112.

St-Onge, B., & Cavayas, F. (1997). Automated forest structure mapping from high resolution imagery based on directional semivariogram estimates. *Remote Sensing of Environment, 61*, 82–95.

Tucker, C. (1979). Red and infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment, 8*, 127–150.

Turner, D., Cohen, W., Kennedy, R., Fassnacht, K., & Briggs, J. (1999). Relationships between LAI and Landsat TM spectral vegetation indices across three temperate zone sites. *Remote Sensing of Environment, 70*, 52–68.

Van Huffel, S. (1997). *Recent advances in total least squares techniques and errors-in-variables modeling*. Philadelphia: Society for Industrial and Applied Mathematics.

Wackernagel, H. (1998). *Multivariate geostatistics: An introduction with applications* (2nd ed.). New York: Springer-Verlag.

Waring, R., & Running, S. (1998). *Forest ecosystems: Analysis at multiple scales* (2nd ed.). San Diego: Academic Press.

White, J., Running, S., Nemani, R., Keane, R., & Ryan, K. (1997). Measurement and remote sensing of LAI in rocky mountain montane ecosystems. *Canadian Journal of Forest Research, 27*, 1714–1727.

Wilson, J., Mitasova, H., & Wright, D. (2000). Water resource applications of geographic information systems. *URISA Journal, 12*, 61–79.

Woodcock, C., Strahler, A., & Jupp, D. (1988). The use of variograms in remote sensing: I. Scene models and simulated images. *Remote Sensing of Environment, 25*, 323–343.

Wulder, M., LeDrew, E., & Lavigne, M. (1998). Aerial image texture information in the estimation of northern deciduous and mixed wood forest LAI. *Remote Sensing of Environment, 64*, 64–76.