ISPRS Journal of Photogrammetry and Remote Sensing 64 (2009) 27–36

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Large area forest inventory using Landsat ETM+: A geostatistical approach

Qingmin Meng [a,*], Chris Cieszewski [a], Marguerite Madden [b]

[a] Warnell School of Forestry and Natural Resources, University of Georgia, Athens, GA 30602, USA

[b] Center for Remote Sensing and Mapping Science, Department of Geography, University of Georgia, Athens, GA 30602, USA

ARTICLE INFO

Article history:
Received 5 February 2007
Received in revised form
7 May 2008
Accepted 2 June 2008
Available online 5 November 2008

Keywords:
Geostatistical approach
Kriging
Regression kriging
Landsat ETM+

ABSTRACT

Large area forest inventory is important for understanding and managing forest resources and ecosystems. Remote sensing, the Global Positioning System (GPS), and geographic information systems (GIS) provide new opportunities for forest inventory. This paper develops a new systematic geostatistical approach for predicting forest parameters, using integrated Landsat 7 Enhanced Thematic Mapper Plus (ETM+) images, GPS, and GIS. Forest parameters, such as basal area, height, health conditions, biomass, or carbon, can be incorporated as a response variable, and the geostatistical approach can be used to predict parameter values for uninventoried points. Using basal area as the response and Landsat ETM+ images of pine stands in Georgia as auxiliary data, this approach includes univariate kriging (ordinary kriging and universal kriging) and multivariable kriging (co-kriging and regression kriging). The combination of bands 4, 3, and 2, as well as the combination of bands 5, 4, and 3, normalized difference vegetation index (NDVI), and principal components (PCs) were used in this study with co-kriging and regression kriging. Validation based on 200 randomly sampling points withheld field inventory was computed to evaluate the kriging performance and demonstrated that band combination 543 performed better than band combination 432, NDVI, and PCs. Regression kriging resulted in the smallest errors and the highest R-squared indicating the best geostatistical method for spatial predictions of pine basal area.

1. Introduction

Large area forest inventories generally are based on field plot sampling, and small area forest inventories usually are processed forest stand units. These two traditional inventories can be integrated by combining ground inventory with Global Positioning System (GPS) and remote sensing data and processing them in geographical information systems (GIS). It is now relatively easy to measure the locations of survey plots, forest stands, and stand boundaries in the field with accuracy of within three meters using differential GPS.

Developments in sensor technology also have allowed the acquisition of remotely sensed data at a range of scales. Remote sensing data are available from satellite sensors providing images with medium spatial resolution of 20–30 m (e.g., Landsat TM, Landsat ETM+, SPOT HRVIR) as well as high spatial resolution of less than 5 m (e.g., Ikonos, QuickBird, LIDAR, and others). Integration of geospatial technologies allows achievements in forest metrics using image data with cell sizes of 30 m, 20 m, 10 m, 5 m, 1 m, or 0.5 m. These forest metrics can be estimated from remote sensing data by modeling the relationships between the image's digital numbers and the forest variables inventoried on the ground with GPS. Geographic information systems and spatial modeling are efficient tools to model, estimate, map, and predict spatial characteristics of stands or trees.

Generally, there are two ways to predict fine scale spatial forest information, nonspatial modeling and spatial modeling. Nonspatial modeling methods widely applied in forest research with linear and nonlinear regressions are the common models applied for estimations of forest variables (Ardö, 1992; Trotter et al., 1997; Dungan, 1998; Cohen et al., 2003; Hudak et al., 2006; Masellj and Chiesi, 2006; Muukkonen and Heiskanen, 2007). K nearest neighbor (KNN) methods for achieving forest metrics using remote sensing data have been applied for forest inventories (Tomppo, 1991; Moeur and Stage, 1995; Franco-Lopez et al., 2001; Holmström and Fransson, 2003; Masellj and Chiesi, 2006; Meng et al., 2007). Artificial neural networks (ANN) also are used for estimating forest variables using remote sensing data (Foody and Boyd, 1999; Foody, 2000; Tatem et al., 2001; Chudamani et al., 2006).

Using the data from Landsat and SPOT as predictors, Tokola et al. (1996) applied both linear regression and the KNN method on forests in the southern boreal vegetation zone in Finland.

* Corresponding address: Center for Applied GI Science, Department of Geography and Earth Sciences, University of North Carolina - Charlotte, 9201 University City Blvd., 28223 Charlotte, NC, USA. Tel.: +1 704 687 5988; fax: +1 704 687 5966.

E-mail address: qmeng1@uncc.edu (Q. Meng).

0924-2716/$ – see front matter © 2008 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.isprsjprs.2008.06.006

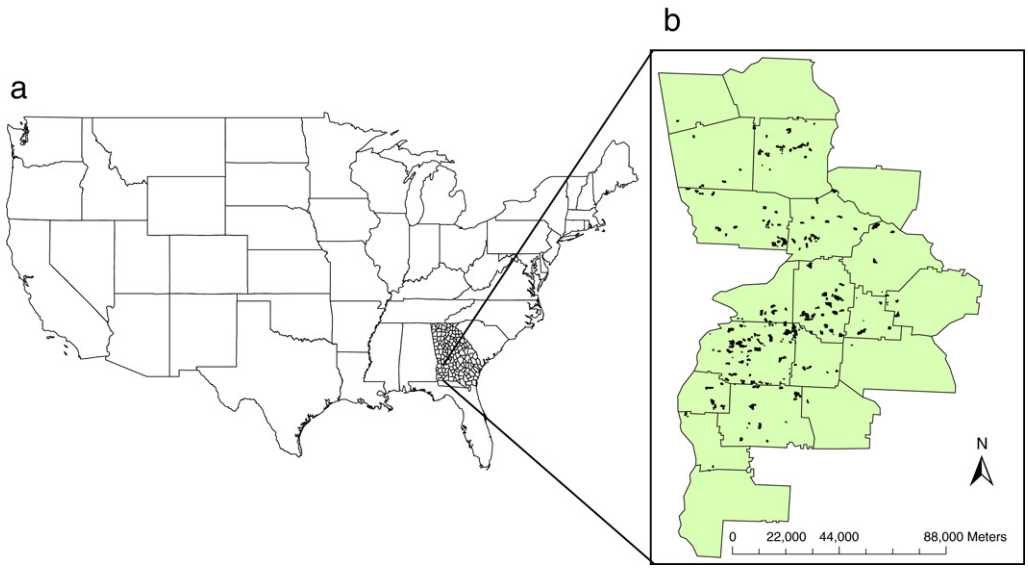

Fig. 1. The study area includes 20 counties (b) in the United States of America (a). The ground inventory locations are indicated as dark points in B.

The authors reported standard errors of stem volume prediction from 70 to 80 m^3/ha (more than 60% of the mean) at the plot level. Trotter et al. (1997) used ordinary least squares to predict stem volume of mature plantations in New Zealand and reported a root mean square error (RMSE) greater than 100 m^3/ha (with a mean stem volume of 413 m^3/ha) for pixel predictions. Using a combination of SPOT 4 and low frequency radar data from the airborne CARABAS system, Holmström and Fransson (2003) applied KNN method to predict forest variables and reported RMSE of 64% (of the mean) of stem volume using optical data and of 53% using the combination of optical and radar data. The stem volume of the sample plots (10 m radius) was in the range of 0–750 m^3/ha with a mean value of 171 m^3/ha. Using Landsat ETM+ data and comparing ANN, multiple linear regression and maximum likelihood classification, Chudamani et al. (2006) concluded that linear regression performed significantly worse than other methods for characterizing forest canopy density.

Many studies have conducted spatial predictions based on remotely sensed data (Curran, 1988; Atkinson et al., 1994; Dungan et al., 1994; Lark, 1996; Dungan, 1998; Curran and Atkinson, 1998; Addink and Stein, 1999; Atkinson and Lewis, 2000; Chica-Olmo and Abarca-Hernandez, 2000). Few studies have been conducted on estimations of forestry relevant variables using spatial models, although a large number of spatial-statistical and prediction models are available in the literature (e.g. Cressie (1993), Wackernagel (1994), Odeh et al. (1995), Goovaerts (1997) and Odeh and McBratnery (2000). Masellj and Chiesi (2006), Buddenbaum et al. (2005), Berterretche et al. (2005), Tuominen et al. (2003), and Zhang et al. (2004) applied geostatistical models to estimate forest variables, such as leaf area index, and to classify forest lands based on remote sensing data. Gilbert and Lowell (1997) used kriging to predict stem volume in a 1500 ha balsam fir (*Abies balsamea*) dominated forest. Prediction based on 5.6 m and 11.3 m radius plots resulted in a RMSE of 54% (of the mean) and 39%–46%, respectively. Methodologically, the accuracy rate of the predicted variable could be improved by incorporating close field observations as predictors in spatial modeling.

In addition to analyzing spatial characteristics of GIS-integrated ground and remote sensing data, it is also necessary to analyze nonspatial data, for example, the selection of band combinations and data reduction of remotely sensed imagery. What is the association between the response variable and independent variables (i.e., the remotely sensed data)? Distribution tests may be needed, although the derivation of kriging equations does not depend on any distributional assumptions. Correlation diagnostics are important for multivariable geostatistics and variogram models are often fitted to check spatial autocorrelation and dependence. Cross variograms need fitting if multivariable geostatistical approaches are conducted. Additionally, it is important to check whether a spatial trend exists in the data of the response variable. Both universal kriging and regression kriging are efficient to incorporate the trend in geostatistical predictions.

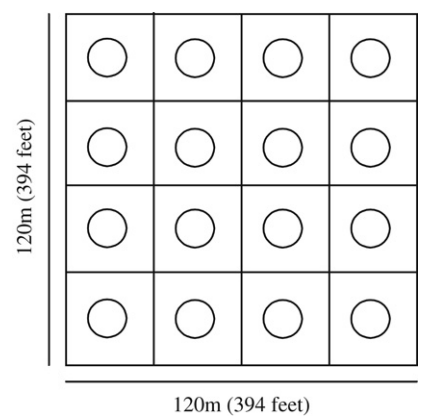

Fig. 2. One example of a field plot.

2. Data sources

2.1. Ground data

Ground data covering 20 counties in west Georgia (USA) were inventoried in 1999 (Fig. 1) by private timber companies. The locations of these ground data were collected using differential Global Positioning System (DGPS) units with accuracy within three meters. One example of a field plot composed of sixteen fixed-radius subplots is indicated in Fig. 2. The radius for the subplots in a given plot was fixed and dependent on the density of a given stand, but the specific distributions of the plots in the study area cannot be given in detail because of the business confidentiality. The coordinates of the ground data were converted to the Universal Transverse Mercator ground coordinate system to match those of

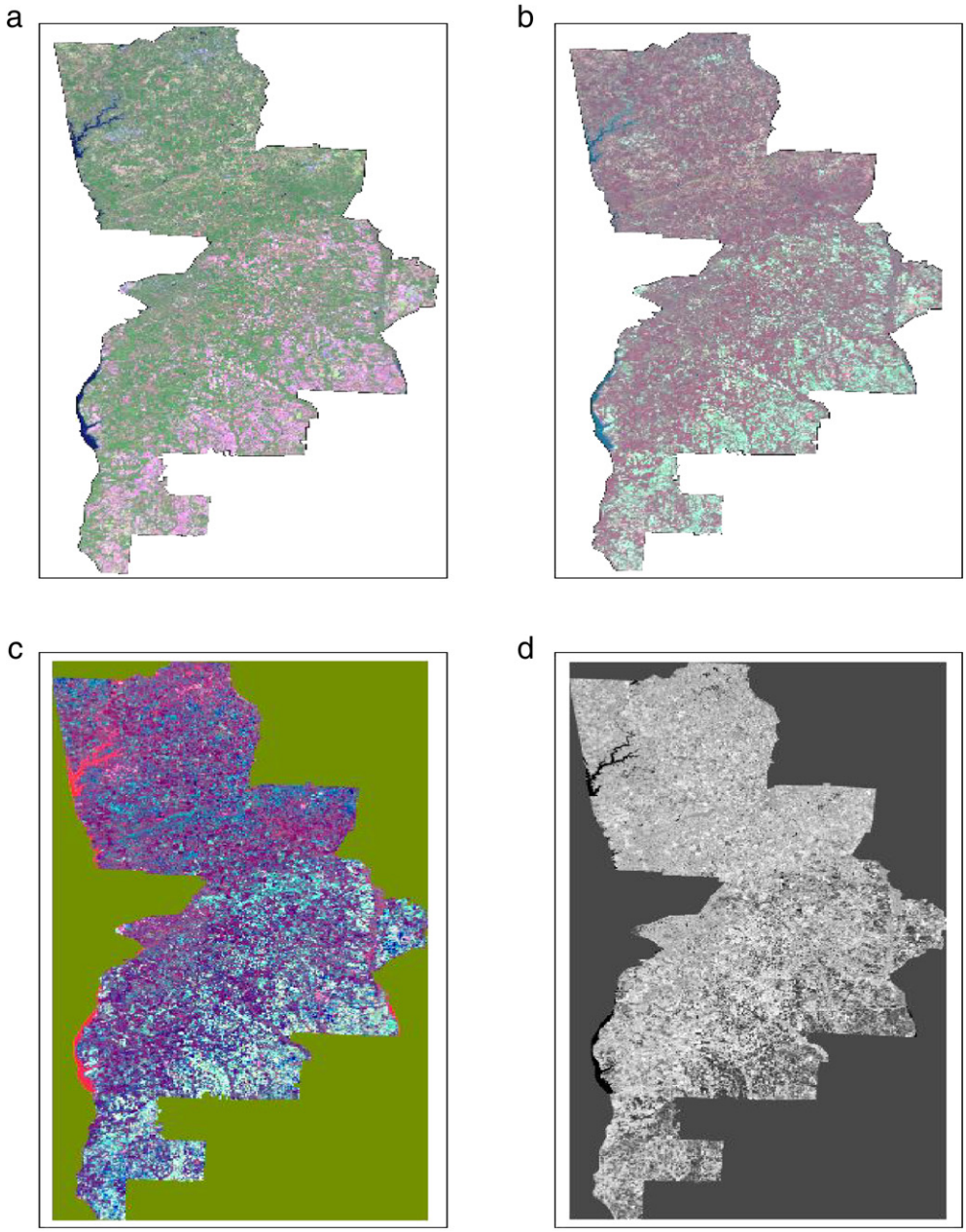

Fig. 3. Landsat ETM+ images used for pine basal area prediction. (a) a 543 band combination; (b) a 432 band combination; (c) the three PCs images; (d) the NDVI images.

the Landsat ETM+ images (Fig. 3). There were 2822 ground records used in this study with a mean of basal area of 13.99 m^2/ha ranging from 0.038 to 29.84 m^2/ha. The basal area and dominant height were measured, and volume of trees was calculated according to tree species. The dominant species are Loblolly pine (*Pinus taeda*) and Slash pine (*Pinus eliotii*), and pine basal area (PBA) was conveniently used as the response variable in this study. A Landsat pixel size (30 m) was maintained in the prediction process of basal area for the uninventoried areas in these 20 counties.

2.2. Remote sensing data

Two Landsat 7 Enhanced Thematic Mapper Plus (ETM+) images (Path/Row: 19/37 and 19/38) acquired on 10 September 1999 from the US Geological Survey (USGS) Earth Resource Observation System Data Center were used in this research. Atmospheric conditions were clear at the time of image acquisition, and the data had been corrected for the radiometric and geometric distortions of the images to the standard Level 1G before delivery. Two Landsat images covering this study area were masked after the geometric corrections using USGS digital orthophoto quarterquads (DOQQs) as the source of control (RMSE is less than 10 m). This resulted in 4449 pixel by 9010 row, 6-band (i.e., 1, 2, 3, 4, 5, and 7) images for analysis. The field inventory data were overlaid on the image data, and then the nearest pixel values including Landsat bands and computed NDVI and PCs were attached to the ground records.

2.2.1. Band combinations

Band 1 of Landsat images contributes little for vegetation analysis. Studies indicate that as the leaf coverage changes from 0% to 11.9%, 43.2%, and 87.6%, very little change occurs in the reflectance of band 1 (0.4–0.5 μm) (Short, 1999). The differences of reflectance increase in bands of large wavelength from 0.5 to 0.8 μm as leaves change. The differences of reflectance in the mid-infrared ranges are very close to the differences in the

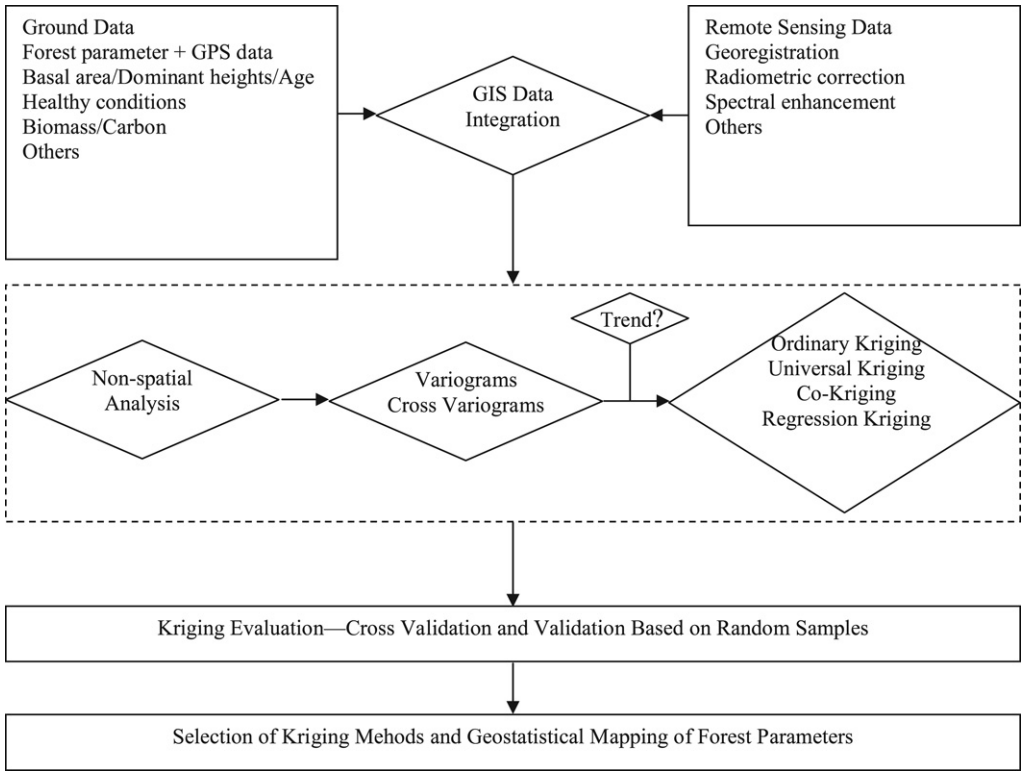

Fig. 4. A systematic geostatistical approach for predicting forest parameters based on remotely sensed data.

near infrared ranges. Band 6 and band 7 with ground resolution of 60 and 120 m are not necessary for spatial prediction of forest parameters with cell size of 30 m. Bands 2, 3, 4, and 5 were applied to estimate pine basal area. The band combination 432 is the standard "false color" composite that is similar to color infrared aerial photography and displays vegetation in shades of red. The band combination 543 provides a display with information on healthy vegetation depicted as bright green. The band combinations 532 and 542 were not applied in this study, since by visual assessment they did not indicate the pine stands as clearly as the combinations of 432 and 543. Landsat band combinations of 432 and 543, therefore, were used for further geosatistical analysis of PBA.

2.2.2. Principal component analysis

Principal component analysis (PCA) is the most frequently used technique for remote sensing data reduction. Generally, remotely sensed data, such as Landsat images, are highly correlated among the adjacent spectral bands (Barnsley, 1999). The Landsat bands are transformed into orthogonal principal components (PCs) with the first PC containing the largest percentage of data variation, and the second PC containing the second largest variance of the data, and so on. The higher the PC is numbered, the less useful information the PC contains. In this research, the six Landsat ETM+ bands used (i.e., bands 1, 2, 3, 4, 5, and 7) were processed using PCA, and the first three PCs were applied for pine basal area analysis because they accounted for more than 95% of the total variance.

2.2.3. Normalized difference vegetation index

NDVI (Eq. (1)) measures both the amount of green vegetation and vegetation health in an

$$NDVI = (NIR - red)/(NIR + red) \quad (1)$$

area, but it also is a basic indicator of changes in vegetation over space and time. It has been extensively applied as a proxy for leaf area index (Tucker, 1979), vegetation biomass (Sellers, 1987), and net primary production (Goward et al., 1985). Therefore, NDVI indicates the spatial characteristics of forest stand development, especially the density and health of trees. It has been proven to be an efficient indicator in detecting and quantifying large-scale changes in plant and ecosystem processes (Braswell et al., 1997; Myneni et al., 1997).

3. Methodology

Rarely has research explored the integration of remote sensing data, GPS, ground data, GIS, and geostatistics to estimate forest parameters at a high spatial resolution for large areas. One systematic geostatistical approach for spatial forest inventory is developed and explored in this research. Compared to the typical ordinary kriging (OK) and universal kriging (UK) using only one variable, this research develops a systematic geostatistical approach – co-kriging (CoK) and regression kriging (RK) using Landsat ETM+ data as predictors – to improve spatial predictions of forest variables by integrating GPS, ground inventory data, Landsat ETM+, and GIS. This systematic geostatistical approach is summarized in a flow chart (Fig. 4), which considers the associations between one forest parameter and DN, and incorporates the spatial dependence of the forest parameter into the process of spatial prediction. In this study, basal area is used as the response variable to conduct this geostatistical approach.

3.1. Correlation analysis

Correlation analysis was applied to measure the strength of association between the response variable and the independent variables. Pearson's product-moment correlation coefficient

(r_{xy}, Eq. (2)) and the Pearson partial correlation coefficient ($r_{xy\bullet zq}$, Eq. (3)) were used

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{2}$$

$$r_{xy\bullet zq} = \frac{r_{xy\bullet z} - r_{xq\bullet z} r_{yq\bullet z}}{\sqrt{(1 - r^2_{xq\bullet z})(1 - r^2_{yq\bullet z})}} \tag{3}$$

to measure the association between the response variable (e.g., pine basal area) and the independent variables (e.g., Landsat bands). Pearson's product-moment correlation measures the association without considering the correlation contributions from other associated variables. The Pearson partial correlation in this study measures the strength of a relationship between pine basal area and one Landsat band, while controlling the effects of two additional Landsat bands. Therefore, Eq. (3) is called the second-order partial correlation indicating the partial correlation between pine basal area and a given Landsat band, because this partial correlation is conditioned on two additional Landsat bands of z and q. For example, we calculate the second-partial correlation between pine basal area (e.g., y) and Landsat band 2 (e.g., x) by controlling band 3 (e.g., q) and band 4 (e.g., z). Likewise, a first order partial correlation is the partial correlation that is conditioned on only one variable (e.g. $r_{xy\bullet z}$ has a single control variable z), and we can call a typical Pearson correlation coefficient without controls "a zero order correlation".

3.2. Geostatistical approach

Geostatistical methods are based on the theory of regionalized variables (Matheron, 1965), which assume that observations are stochastic variables. A spatial property Z at location x is assumed to be a realization of a random function $Z(x)$. The stationarity of the first- and second-order moments of $Z(x) - Z(x+h)$ is the intrinsic hypothesis that allows us to define the semivariogram.

Semivariogram is one of the key steps in geostatistical modeling. Semivariogram describes the spatial dependence of spatial variables. Semivariogram has been used widely in remote sensing to determine spatial structures (Curran, 1988; Warren et al., 1990; Atkinson and Lewis, 2000). Based on the semivariogram, the geostatistical process derives optimal linear unbiased spatial prediction methods (i.e., kriging) by minimizing mean-squared prediction error. However, the assumptions of stationarity, which often are not met by the field-sampled data sets, and the requirement of a large dataset to define the spatial autocorrelation result in the limitations of univariate kriging. Fortunately, geostatistical methods also provide optimal prediction methods using auxiliary data. Large volumes of auxiliary data like remote sensing data for forest research are available now. Incorporating the auxiliary data, co-kriging and regression kriging, as described below, can increase prediction accuracy. The gstat package (Pebesma, 2005) is mainly referenced for variogram and kriging methods as follows.

3.2.1. Variograms

The direct variogram generally is computed from Eq. (4),

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_{i+h})]^2 \tag{4}$$

where x_i is a data location, h is a vector of distance, $Z(x_i)$ is the data value of one kind of attribute at location x_i, $N(h)$ is the number of data pairs for a certain distance and direction of h units. This equation is used for determining the spatial autocorrelation of the univariate variable.

A typical cross variogram is calculated using Eq. (5). It is applied for the joint spatial

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{[z(x_i) - z(x_{i+h})] \cdot [y(x_i) - y(x_{i+h})]\} \tag{5}$$

variability between two types of spatial variables. It is defined as half of the average product of the lag distance relative to the two variables Z and Y with the same notations as Eq. (4).

When direct and cross variogram models are fitted, they also can guarantee that the fitted models follow the linear model of coregionalisation (Goovaerts, 1997). This ensures the cross covariance matrices are always positive.

3.2.2. Kriging

3.2.2.1. Ordinary kriging and universal kriging. Ordinary kriging (OK) is identical to multiple linear regression, with a couple of important differences. The ordinary kriging model is as in Eq. (6).

$$\hat{z}(s_0) = \sum_{i=1}^{n} \lambda_i z(s_i). \tag{6}$$

$\hat{Z}(s_0)$ is the value to be interpolated at location s_0, $z(s_i)$ are the sampled values at their locations, and λ_i are the weights to be assigned to each sampled value. Universal kriging is applied when a trend exists. Universal kriging is often fitted using a polynomial equation, which can be represented in the similar way as the Eq. (6) to analyze the trend across the study area.

3.2.2.2. Cokriging. For forest applications, a few studies using remote sensing data have been conducted using the geostatistical approach. Dungan et al. (1994) and Dungan (1998) applied co-kriging and a stochastic simulation method for forest management using synthetic remote sensing datasets.

Co-kriging (CoK) is an extension of kriging, and is a method for estimating one or more variables of interest using data from several variables by incorporating not only spatial correlation but also inter-variable correlation. Co-kriging is a very versatile and rigorous statistical technique for spatial point estimation when both primary and auxiliary attributes are available. It is defined as in Eq. (7).

$$\hat{z}(s_0) = \sum_{j=1}^{n} z(s_j) \Lambda_{j\bullet}. \tag{7}$$

If each component of $z(s_j)$ satisfies the intrinsic hypothesis (Journel and Huijbregts, 1978), then Eq. (7) is unbiased if

$$\sum_{j=1}^{n} \Lambda_{j\bullet} = I_l \tag{8}$$

where I_l is an identity matrix with the size of l that is the number of variables, and $\Lambda_{j\bullet}$ are the weights associated with prediction. Eq. (7) is

$$\sum_{\phi=1}^{v} \Gamma(s_i, s_j) + \Psi = \Gamma(s_i, s_0) \tag{9}$$

where $z(s_j)$ is the vector $z_1(s_j) \ldots z_m(s_j)$. $\Gamma(s_i, s_j)$ and $\Gamma(s_i, s_0)$ are the cross variograms, and Ψ is the Lagrange Multiplier for i from 1 to n.

According to the sample relations between the primary variable and the auxiliary variables, CoK could be described in several ways as follows. The efficient way to predict the primary variable may be to use the auxiliary variables to cokrige it into dense grid locations. This is named heterotopic cokriging (Wackernagel, 1994). Isotopic

Table 1
Pearson correlation matrix for the variables analyzed for pine basal area estimation.

	PINEBA	Band 2	Band 3	Band 4	Band 5	NDVI	PC1	PC2	PC3
PINEBA	1								
Band 2	−0.3917	1							
Band 3	−0.5417	0.8364	1						
Band 4	0.3456	0.1221	−0.0724	1					
Band 5	−0.5964	0.8067	0.9312	−0.0488	1				
NDVI	0.6365	−0.6517	−0.8794	0.5202	−0.8187	1			
PC1	−0.5195	0.8623	0.9384	0.1197	0.9766	−0.7417	1		
PC2	−0.6520	0.7129	0.9022	−0.3508	0.9448	−0.9269	0.8784	1	
PC3	−0.0315	−0.1852	−0.1287	−0.7872	−0.3163	−0.2450	−0.3835	−0.0441	1

Table 2
Partial correlation between pine basal area and Landsat bands.

	432 band combination			543 band combination		
	Band 2	Band 3	Band 4	Band 3	Band 4	Band 5
r_{xy}	0.0129	−0.3350	0.3436	0.0828	0.3997	−0.3429
P value	0.4976	<.0001	<.0001	<.0001	<.0001	<.0001

cokriging requires that data on both the target variable and covariables be measured at all sample locations. A variant of both is generalized cokriging (Myers, 1982) that involves simultaneous prediction of all the correlated variables into more dense locations. The complete case is the case where the covariates and the primary variable do not share any common locations. A more general type applied to remote sensing data is collocated cokriging, where covariates are available at all interpolation locations, although the primary variable is available at only a few locations. When CoK is compared to univariate kriging, no new concept is added, but there is heavier notation associated with having several variables (Goovaerts, 1997).

3.2.2.3. Regression kriging. Regession kriging (RK) is a hybrid method that combines either a simple or multiple-linear regression model (or a variant of the generalized linear model (GLM) and regression trees) with kriging (Odeh et al., 1995; Goovaerts, 1997). In the process of RK, kriging with uncertainty introduces the regression residuals (i.e., the model uncertainty) into the kriging system, which is then applied directly to predict the primary variable. The predictions are combined from two parts; one is the estimate $\hat{m}(S_0)$ obtained by regressing the primary variable on the k auxiliary variables $q_k(s_0)$ and $q_0(s_0) = 1$; the second part is the residual estimated from the ordinary kriging. Regression kriging is estimated as follows:

$$\hat{z}_{rk}(s_0) = \hat{m}(s_0) + \hat{\ell}(s_0) \tag{10}$$

$$\hat{z}_{rk}(s_0) = \sum_{k=0}^{v} \hat{\beta}_k q_k(s_0) + \sum_{i=1}^{n} \omega_i(s_0)\ell(s_i) \tag{11}$$

where $\hat{\beta}_k$ are trend model coefficients, optimally estimated using generalized least squares; ω_i are weights determined by the semivariance function, and ℓ are the regression residuals. The gstat package is used to conduct the regression kriging (Pebesma, 2004, 2005).

3.2.3. Model evaluation

In this study, different geostatistical models are developed and applied for pine basal area prediction. There are always discrepancies between true and predicted values. It is necessary to validate the models and to check which is more efficient. In addition to cross validation that is used to validate whether the model fits the training data, validation based on random samples outside of the training data set is applied to assess these kriging approaches for spatial estimation. We developed 200 random points to check which model is more efficient in spatial predictions of pine basal area.

There are many different measures for checking discrepancies, and each has its advantages and weaknesses. Details about forecast evaluation were discussed by Murphy and Katz (1985). Typically, root mean square error (RMSE) is often used, other indices including standard deviation (SD), bias error (BE), and mean-absolute error (MAE) are used to make a relatively complete comparisons. The measures are listed as Eqs. (12)–(15).

$$\mathrm{SD} = \left[\frac{1}{N-1}\sum_{n=1}^{N}(X_n - \bar{X})^2\right]^{1/2} \tag{12}$$

$$\mathrm{BE}(X) = \frac{1}{N}\sum_{n=1}^{N}(X_f - X_o) \tag{13}$$

$$\mathrm{RMSE}(X) = \left[\frac{1}{N}\sum_{n=1}^{N}(X_f - X_o)^2\right]^{1/2} \tag{14}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{n=1}^{N}\left|X_f - X_o\right| \tag{15}$$

where N is the size of the sample, X_n is the sample values and $\bar{X}$ is the mean of the sample, X_f is the forecast value, and X_o is the observed value. A positive BE indicates a tendency to overpredict, while a negative BE implies underprediction.

4. Results

Based on the 2822 field measurements, various kriging methods are applied to estimate basal area across the 20 counties, in which remotely sensed data are used as predictor variables when spatial estimation is predicted through cokriging and regression kriging. Pearson correlation is used to understand the contributions from predictors and semivariogram models are fitted to explore spatial dependence and variability.

4.1. Correlations between pine basal area and predictors

A relatively high correlation between auxiliary bands and PBA indicates that the auxiliary variables can play important roles in the prediction process of PBA (Tables 1 and 2). A partial correlation shows the contribution of an auxiliary variable to the prediction of PBA when several auxiliary variables are used in

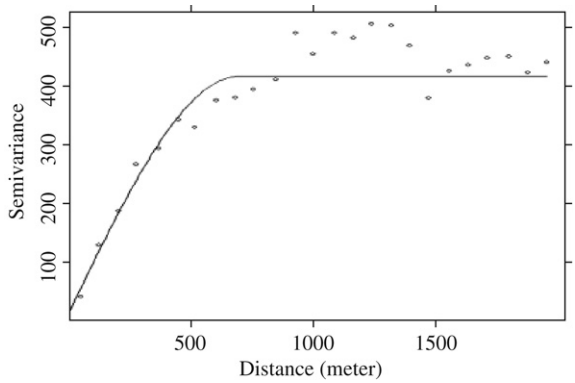

Fig. 5. Semivariogram analysis for pine basal area.

model fitting. Therefore, kriging with contributions from highly correlated auxiliary variables can perform better than kriging only based on the dependent variable of PBA.

Predictors are grouped into four groups: a 432 band combination; a 543 band combination; a three-PCs combination; and an NDVI image. The general Pearson correlation coefficients were calculated and summarized in Table 1. Considering the absolute values of these coefficients, for the correlations between pine basal area and different independent variables, PC2 has the highest correlation, the second one is NDVI, the third one is band5, and then, band3, PC1, band2, band4, and PC3. The spectral characteristics of the bands of Landsat ETM+ typically determine the correlation between PBA and the auxiliary variables. PBA in a given area describes the degree to which pine trees occupy in this area. Band 2 of Landsat ETM+ is usually applied to assess the vegetation vigor, while band 3, 4, and 5 are designed respectively for chlorophyll absorption, biomass surveys, and vegetation measurements. The high correlation between PC2 and band 3 and 5 and the relatively high correlation between PC2 and band 4 determine that PC2 is highly correlated with PBA. We can expect if a PC or a band index is highly correlated with band 3, 4, and 5 then basal area would be highly correlated with this PC or band index.

Since different combinations of predictors were used, the Pearson partial correlation coefficients were calculated and tested in the combinations of bands and PCs in order to better understand the associations between pine basal area and the predictors (Table 2). In the 432 band combination, band 3 and band 4 have similar degree correlations but in different directions; one is positive, and the other is negative; band 2 is little correlated with the pine basal area, and the coefficient is not significantly different from 0. In the 543 band combination, band 4 and band 5 have similar correlations with pine basal area. However, band 4 is positively correlated, and band 5 is negatively correlated. Band 3 is little correlated with pine basal area. PC2 is highly correlated with pine basal area. The coefficient of PC1 is much smaller. The correlation between PC3 and pine basal area is not statistically significant, since its P value is around the boundary of 0.05.

4.2. Variograms and spatial dependence

Variograms were used to spatially analyze the surface properties of pine basal area. Based on the variogram cloud, the empirical semivariogram model was created. The different types of semivariogram models used to fit the points include exponential, Gaussian, circular, spherical, tetraspherical, pentaspherical, Hole effect, K-Bessel, and J-Bessel models. The spherical model had the best fits and was selected as the theoretical model applied for spatial predictions (Fig. 5). The fit of the spherical model has a nugget of 5, a partial sill of 450, and a range of 750 m. Also, there was no obvious trend existing among the pine basal area across the study area.

The characteristics of the semivariogram also may be affected by the directions, which result from a special geographic phenomenon. For example, a certain kind of species exists and crosses the area in a certain direction. Anisotropy therefore is often checked before further analysis. Semivariogram analyses at directions 0, 45, 90, 135, 180, 225, 270, and 315 were conducted, and the results indicated similar spatial dependence at a scale about 750 m in eight directions (Fig. 6). It is not necessary to consider anisotropic effects in spatial estimation using different kriging models.

4.3. Assessment of pine basal area estimation

We first applied Univariate kriging (i.e., OK and UK) to estimate the pine basal area based on the 2822 ground inventory points. The UK was used to check whether it is effective compared to the OK, though there was no obvious trend of pine basal area existing across the study area. Four types of co-kriging were applied using the 432 band combination, the 543 band combination, NDVI, and PCs as the auxiliary data. At last, four groups of regression kriging were conducted using the 432 band combination, the 543 band combination, NDVI, and PCs as predictors.

The results were evaluated using cross validation (Table 3). Bias errors using the kriging methods indicated the values of BE were close to 0, and almost unbiased estimations of pine basal area were obtained. For RMSE, there was not much difference between OK, UK, and the four kinds of co-kriging. However, the RMSEs of the estimations using regression kriging were much smaller than those from OK, UK, and co-kriging.

In order to further assess these geostatistical approaches, validation based on 200 random sample points outside of the training dataset was used to compare these kriging methods (Table 4). The regression kriging methods had the smallest BE, MAE, RMSE, and SDe, which indicated that regression kriging was more efficient than other kriging methods. Pine basal area predictions based on RK resulted in the prediction BE of 27.9%–31.5% of the mean (13.99 m^2/ha), the prediction MAE of 39.3%–42.1% of the mean, the prediction RMSE of 63.5%–68.6% of the mean, and the prediction SDe of 59.3%–62.1% of the mean using the 200 random points outside the training datasets. Additionally, using the 200 random sampled points, scatter plots of observations versus predictions were listed in Fig. 7 and the corresponding R-squared was attached. R-squared is calculated using, $R^2 = 1 - \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / \sum_{i=1}^{n}(y_i - \bar{y})^2$ where y_i is the field values of PBA, $\bar{y}$ is the mean of field values of PBA, and $\hat{y}$ is the estimated PBA value. The scatter plots and the R-squared also show that regression kriging is the most powerful approach to spatially predicting pine basal area.

5. Discussion

Challenges still exist in the field of large area forest inventory using remotely sensed data (Tokola et al., 1996; Trotter et al., 1997; Holmström and Fransson, 2003). Spatial diversity of forest stands and landscape makes the spatial prediction of forest parameters a major challenge, although the remote sensing data are highly associated with forest features. For example, forest stands may have very similar values of biomass/carbon but have different spectral characteristics, because of differences in species. The differences of spectral characteristics between plantations and natural stands might exist although the stands have many of the same characteristics, such as same species, same age, and same density. These differences will add noise when the prediction

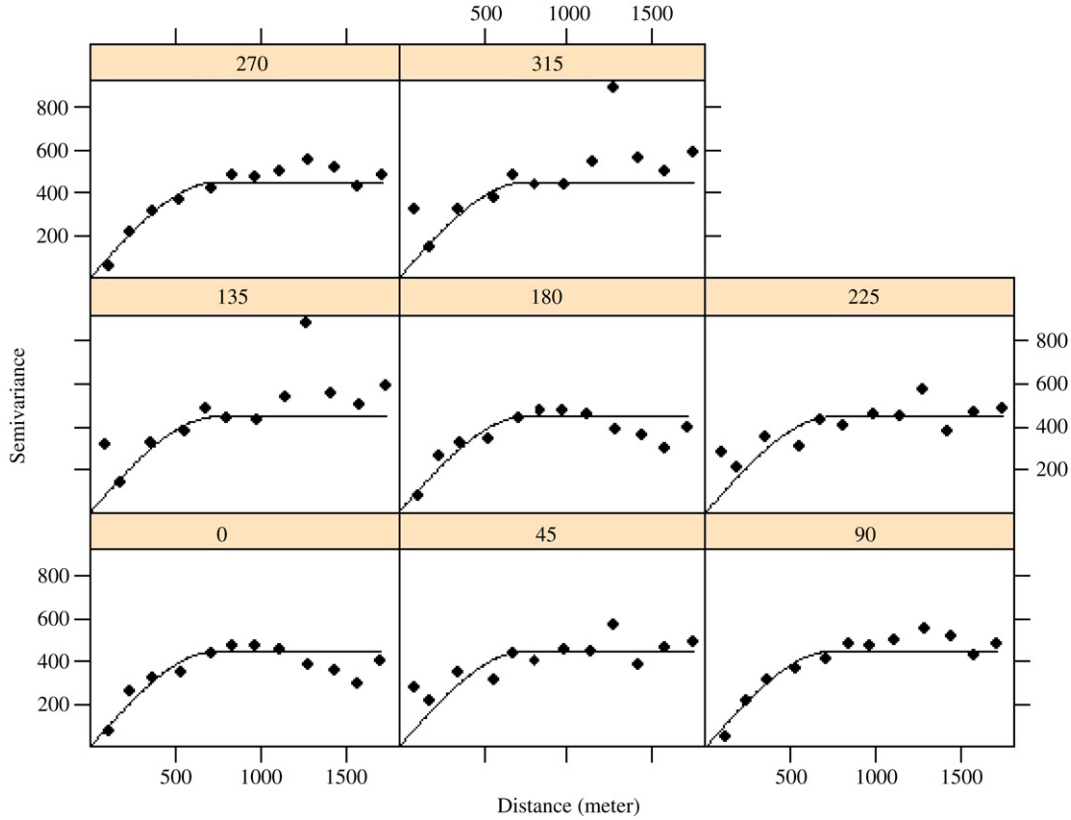

Fig. 6. Semivariogram modeling of directional effects.

Table 3
Model evaluation using cross validation.

	OK	UK	CoK432	CoK543	CoKndvi	CoKPCs	RK432	RK543	RKndvi	RKPCs
BE	−0.076	−0.078	−0.099	−0.100	−0.095	−0.095	−0.078	−0.067	−0.066	−0.070
RMSE	11.310	11.290	10.970	11.000	11.010	11.020	7.020	7.000	7.220	6.890

Ordinary kriging (OK), universal kriging (UK), Co-kriging (Cok), and regression kriging (RK) are used to predict basal area. CoK432 means using the 432 band combination as predictors to krige the basal area, likewise CoK543, CoKndvi, CoKPCs, RK432, RK543, RKndvi, and RKPCs; bias error (BE) and root mean square error (RMSE) are used to measure the discrepancy between observations and predictions.

Table 4
Model and forecast evaluation based on random samples.

	OK	UK	CoK432	CoK543	CoKndvi	CoKPCs	RK432	RK543	RKndvi	RKPCs
BE	10.120	10.130	4.990	4.980	4.760	4.660	4.460	4.010	4.432	3.964
RMSE	13.320	13.390	10.550	10.320	10.560	10.010	9.655	8.980	9.601	9.161
SDe	8.660	8.770	9.300	9.260	9.310	9.210	8.583	8.601	8.700	8.280
MAE	10.330	10.470	6.310	6.290	6.310	6.280	5.929	5.502	5.900	5.727

Stand deviation of errors (SDe), mean-absolute errors (MAE); other notations as Table 3.

models are fitted based on the associations between remotely sensed data and ground-inventoried data.

Mutivariable kriging is more robust compared with univariate kriging as indicated in this research. Regression kriging is powerful compared with other kriging models. As we know the performance of typical kriging models are best for spatial interpolation, while these models can be significantly poor for spatial extrapolation because the kriging coefficients depend on the spatial variation. Regression kriging can be applied for spatial extrapolation because its main coefficients except the coefficient for the residual part just depend on the linear correlation between dependent variable and independent variables. Multivariable kriging can be applied for almost all kinds of forest parameters. Little research has applied multivariable kriging to estimate forest variables for forest resource management, although either numerical or categorical data can be used in the process of kriging, i.e., any kind of variable can be used as auxiliary data or predictors.

Remote sensing data and ground inventory data are collected and stored in different data structures. The discrepancy between remotely sensed data and ground sampling data might be the source of errors in forest predictions (Tokola et al., 1996; Gilbert and Lowell, 1997). The ground inventory data are usually collected at the forest plot level or forest stand level. The plot size may be from several meters to 10 or 20 m. The stand size may be from 10 m to dozens of meters, and the stands are assumed to be homogenous. Some ground data may be finer than remote sensing data in spatial resolution, but generally, remote sensing data has a finer spatial resolution than ground inventory data. This may result in some noise added to the geostatistical modeling and cause bias errors. On the other hand, remote sensing data processing

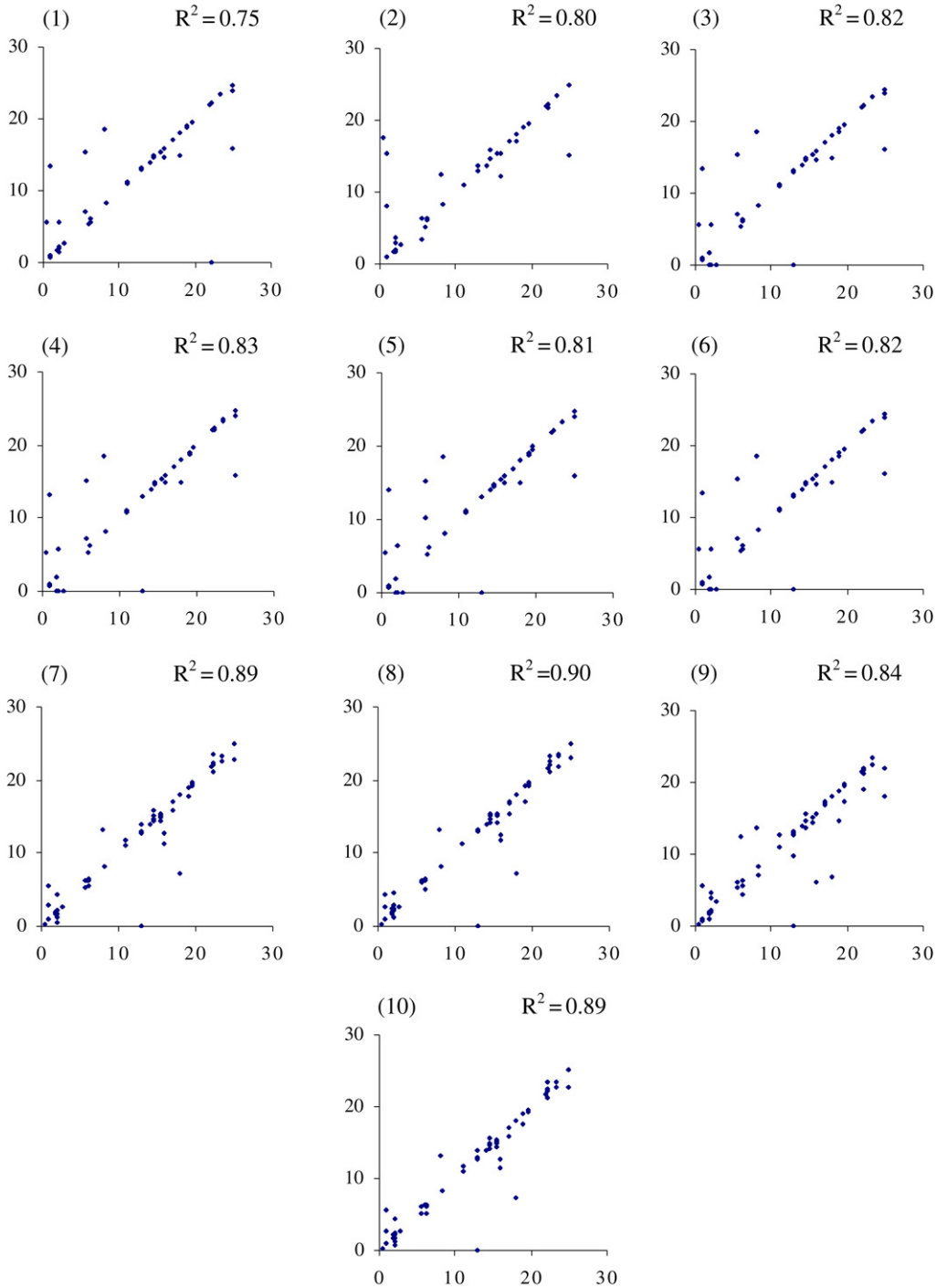

Fig. 7. Scatter plots of observations (*y* axis) versus predictions (*x* axis). Kriging: (1) ordinary kriging, (2) universal kriging, (3) cokriging with band 234, (4) cokriging with band 345, (5) cokriging with NDVI, (6) cokriging with PCs, (7) regressiogn kriging with band 234, (8) regression kriging with band 345, (9) regression kriging with NDVI, (10) regression kriging with PCs.

especially high accurate classification of the interesting variable can help reduce the estimation errors.

6. Conclusions

The systematic approach of geostatistical prediction developed by integrating Landsat ETM+ data, ground inventory, and GPS data provides a new way to spatially estimate forest parameters based on remotely sensed data. It has many applications in forest or natural resource management. Forest metrics, such as stand density, dominant height, species, stand age, forest health conditions, the probability of forest fire, biomass, carbon, and so on, can be incorporated in these models for forest inventory in this study area. This geostatistical approach can be applied in other research regions, while correlation analysis, semivariogram modeling, and kriging models need to be examined for spatial estimation.

Providing high spatial information is essential for large area timber, biomass, and carbon budget management and planning. Kriging is an optimum method for spatial interpolation. Regression

kriging is the powerful one among the different kriging methods in this research. It was used to predict the pine basal area at 30 m for these 20 counties (about 35 000 km^2) using 2822 ground inventory data points. Four groups of independent variables are used in RK. The 543 band combination resulted in the smallest BE, RMSE, MAE, and had a relatively smaller SDe. Therefore, Compared with OK, UK and CoK using different auxiliary data, RK resulted in the smallest BE, RMSE, SDe, and MAE; RK using the 543 and combination is the best method for pine basal area predictions. For other forest parameters, such as dominant height, timber volume, or biomass/carbon, other band combinations, such as PCs or NDVI need to be applied again to check which will result in best estimations.

Acknowledgements

The authors thank the anonymous reviewers' helpful comments on an earlier version of this paper. The authors thank Drs. E. Lynn Usery and Bruce E. Borders for reviewing this paper. Mr. Roger Lowe's help in supplying part of the data in this research also is appreciated.

References

Addink, E., Stein, A., 1999. A comparison of conventional and geostatistical methods to replace clouded pixels in NOAA-AVHRR images. International Journal of Remote Sensing 20 (5), 961–977.
Ardö, J., 1992. Volume quantification of coniferous forest compartments using spectral radiance recorded by Landsat Thematic Mapper. International Journal of Remote Sensing 13 (9), 1779–1786.
Atkinson, P.M., Lewis, P., 2000. Geostatistical classification for remote sensing: An introduction. Computers & Geosciences 26 (4), 361–371.
Atkinson, P.M., Webster, R., Curran, P.J., 1994. Cokriging with airborne MSS imagery. Remote Sensing of Environment 50 (3), 335–345.
Barnsley, M.J., 1999. Digital remote sensing data and their characteristics. In: Longley, P., Rhind, D.W., Goodchild, M. (Eds.), Principles of Geographical Information Systems, second edition. Longman, pp. 451–466.
Berterretche, M., Hudak, A.T., Cohen, W.B., Maiersperger, T.K., Gower, S.T., Dungan, J., 2005. Comparison of regression and geostatistical methods for mapping Leaf Area Index (LAI) with Landsat ETM+ data over a boreal forest. Remote Sensing of Environment 96 (1), 49–61.
Braswell, B.H., Schimel, D.S., Linder, E., Moore III., B., 1997. The response of global terrestrial ecosystems to interannual temperature variability. Science 278 (5339), 870–872.
Buddenbaum, H., Schlerf, M., Hill, J., 2005. Classification of coniferous tree species and age classes using hyperspectral data and geostatistical methods. International Journal of Remote Sensing 26 (24), 5453–5456.
Chica-Olmo, M., Abarca-Hernandez, F., 2000. Computing geostatistical image texture for remotely sensed data classification. Computer & Geosciences 26 (4), 373–383.
Chudamani, J., Leeuw, J.D., Skidmore, A.K., Duren, I.C., Oosten, H., 2006. Remotely sensed estimation of forest canopy density: A comparison of the performance of four methods. International Journal of Applied Earth Observation and Geoinformation 8 (1), 84–95.
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 (4), 561–571.
Cressie, N.A.C., 1993. Statistics for Spatial Data. John Wiley & Sons, New York.
Curran, P.J., Atkinson, P.M., 1998. Geostatistics and remote sensing. Progress in Physical Geography 22 (1), 61–78.
Curran, P.J., 1988. The semivariogram in remote sensing: An introduction. Remote Sensing of Environment 24 (3), 493–507.
Dungan, J.L., Peterson, D.L., Curran, P.J., 1994. Alternative approaches for mapping vegetation quantities using ground and image data. In: Michener, W., Stafford, S., Brunt, J. (Eds.), Environmental Information Management and Analysis: Ecosystem to Global Scales. Taylor and Francis, London, UK, pp. 237–261.
Dungan, J.L., 1998. Spatial prediction of vegetation quantities using ground and image data. International Journal of Remote Sensing 19 (2), 267–285.
Foody, G.M, Boyd, D.S., 1999. Fuzzy mapping of tropical land cover along an environmental gradient from remotely sensed data with an artificial neural network. Journal of Geographical Systems 1 (1), 23–35.
Foody, G.M., 2000. Mapping land cover from remotely sensed data with a softened feed forward neural network classification. Journal of Intelligent and Robotic Systems 29, 433–449.
Franco-Lopez, H., Ek, A.R., Bauer, M.E., 2001. Estimation and mapping of forest stand density, volume and cover type using k-nearest neighbors method. Remote Sensing of Environment 77 (3), 251–274.
Gilbert, B., Lowell, K., 1997. Forest attributes and spatial autocorrelation and interpolation: Effects of alternative sampling schemata in the boreal forest. Landscape and Urban Planning 37, 235–244.
Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.
Goward, S.N., Tucker, C.J., Dye, D.G., 1985. North American vegetation patterns observed with the NOAA-7 advanced very high resolution radiometer. Vegetatio 64 (1), 3–14.
Holmström, H., Fransson, J.E.S., 2003. Combining remotely sensed optical and radar data in *k*NN estimation of forest variables. Forest Science 49 (3), 409–418.
Hudak, A.T., Crookston, N.L., Evans, J.Z.S., Falkowski, M.J., Smith, A.M.S., Gessler, P.E., Morgan, P., 2006. Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral satellite data. Canadian Journal of Remote Sensing 32 (2), 126–138.
Lark, R.M., 1996. Geostatistical description of texture on an aerial photograph for discriminating classes of land cover. International Journal of Remote Sensing 17 (11), 2115–2133.
Masellj, F., Chiesi, M., 2006. Evaluation of statistical methods to estimate forest volume in a Mediterranean region. IEEE Transactions On Geoscience And Remote Sensing 44 (8), 2239–2250.
Matheron, G., 1965. Les Variable Regionalisees et leur Estimation. Masson, Paris.
Meng, Q., Cieszewski, C.J., Madden, M., Borders, B.E., 2007. K nearest Neighbor method for forest inventory using remote sensing data. GIScience & Remote Sening 44 (2), 149–165.
Moeur, M., Stage, A.R., 1995. Most similar neighbor: An improved sampling inference procedure for natural resource planning. Forest Science 41 (2), 337–359.
Murphy, A.H., Katz, R.W., 1985. Probability, Statistics, and Decision Making in the Atmospheric Sciences. Westview Press, Boulder, Colo.
Myers, D.E., 1982. Matrix formulation of cokriging. Mathematic Geology 14 (3), 249–257.
Myneni, R.B., Keeling, C.D., Tucker, C.J., Asrar, G., Nemani, R.R., 1997. Increased plant growth in the northern high latitudes from 1981 to 1991. Nature 386 (6626), 698–702.
Muukkonen, P., Heiskanen, J., 2007. Biomass estimation over a large area based on standwise forest inventory data and ASTER and MODIS satellite data: A possibility to verify carbon inventories. Remote Sensing of Environment 107 (4), 617–624.
Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1995. Further results on prediction of soil properties from terrain attributes: Heterotopic cokriging and regression-kring. Geoderma 67 (3-4), 215–226.
Odeh, I.O.A., McBratnery, A.B., 2000. Using AVHRR images for spatial prediction of clay content in the lower Namoi Valley of eastern Australia. Geoderma 97 (3-4), 237–245.
Pebesma, E.J., 2004. Multivariable Geostatistics in S: The Gstat Package. Computer & Geosciences 30 (7), 683–691.
Pebesma, E.J., (2005). The Gstat Package. http://cran.r-project.org/doc/packages/gstat.pdf. (accessed 25. 11. 07).
Short, N.M., (1999). The remote sensing tutorial. http://www.fas.org/irp/imint/docs/rst/Sect3/Sect3_1.html. (accessed on 01. 01. 08).
Sellers, P.J., 1987. Canopy reflectance, photosynthesis and transpiration. II. The role of biophysics in the linearity of their interdependence. Remote Sensing of Environment 21 (2), 143–183.
Tatem, A.J., Lewis, H.G., Atkinson, P.M., Nixon, M.S., 2001. Land cover mapping from remotely sensed images at the sub-pixel scale using a Hopfield neural network. IEEE Transactions on Geoscience and Remote Sensing 39 (4), 781–796.
Tokola, T., Pitkänen, S., Partinen, S., Muinonen, E., 1996. Point accuracy of a non-parametric method in estimation of forest characteristics with different satellite materials. International Journal of Remote Sensing 17 (12), 2333–2351.
Tomppo, E., 1991. Satellite imagery-based national inventory of Finland. International Archives of Photogrammetry and Remote Sensing 28 (7-1), 419–424.
Trotter, C.M., Dymond, J.R., Goulding, C.J., 1997. Estimation of timber volume in a coniferous plantation forest using Landsat TM. International Journal of Remote Sensing 18 (10), 2209–2223.
Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment 8 (2), 127–150.
Tuominen, S., Fish, S., Poso, S., 2003. Combining remote sensing, data from earlier inventories, and geostatistical interpolation in multisource forest inventories. Canadian Journal of Forest Research 33 (4), 623–634.
Warren, B.C., Spies, T.A., Bradshaw, G.A., 1990. Semivariograms of digital imagery for analysis of conifer canopy structure. Remote Sensing of Environment 34 (3), 167–178.
Wackernagel, H., 1994. Multivariate spatial statistics. Geoderma 62 (1), 83–92.
Zhang, C., Franklin, S.E., Wulder, M.A., 2004. Geostatistical and texture analysis of aribore-acquired images used in forest classification. International Journal of Remote Sensing 25 (4), 859–865.