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; Maselli 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; Maselli 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

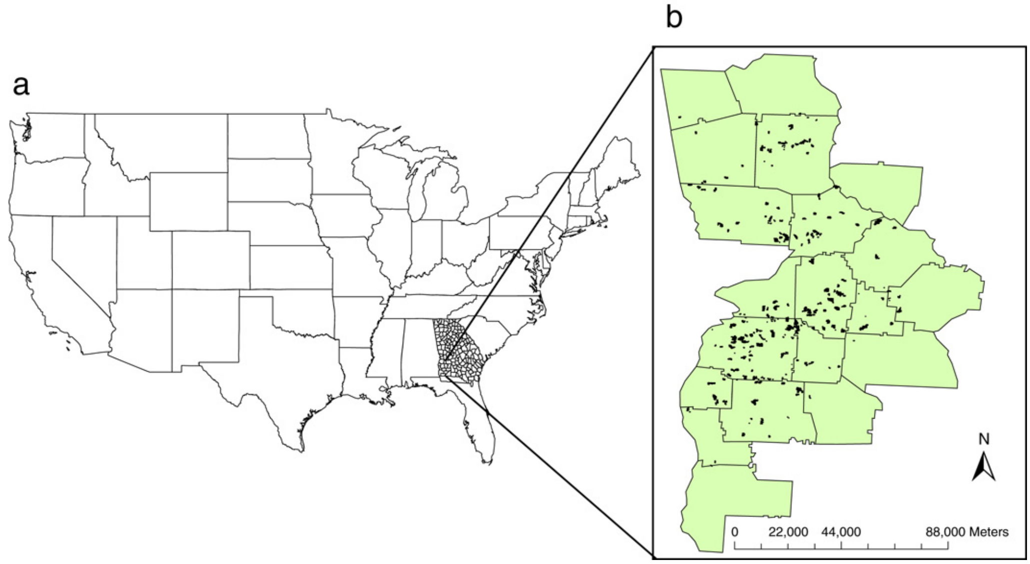

Fig. 1. The study area includes 20 counties (b) in the United States of America (a). The ground inventory locations are indicated as dark points in B.

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.

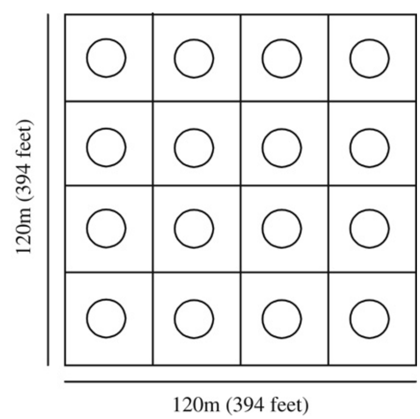

Fig. 2. One example of a field plot.

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