An Evaluation of Spatial Autocorrelation and Heterogeneity in the Residuals of Six Regression Models

Lianjun Zhang, Zhihai Ma, and Luo Guo

Abstract: Spatial effects include spatial autocorrelation and heterogeneity. Ignoring spatial effects in a modeling process causes misleading significance tests and suboptimal model prediction. In this study, we used three forest plots with different spatial patterns of tree locations (i.e., clustered, random, and regular patterns) to investigate the spatial distributions and heterogeneity in the model residuals from six regression models with the ordinary least squares (OLS) as the benchmark. Our results revealed that when significant spatial autocorrelations and variations existed in the relationship between tree height and diameter, as in the softwood plot (clustered) and hardwood plot (random), OLS was not appropriate for modeling the relationship between tree variables. Spatial regression models (i.e., spatial lag and spatial error models) were effective for accounting for spatial autocorrelation in the model residuals, but they were insufficient to deal with the problem of spatial heterogeneity. It was evident that the model residuals in both spatial lag and spatial error models had a similar pattern and magnitudes of spatial heterogeneity at spatial scales different from those of the OLS model. In contrast, the linear mixed model and geographically weighted regression incorporated the spatial dependence and variation into modeling processes, and consequently, fitted the data better and predicted the response variable more accurately. The model residuals from both the linear mixed model and geographically weighted regression had desirable spatial distributions, meaning fewer clusters of similar or dissimilar model residuals over space. FOR. SCI. 55(6):533–548.

Keywords: model residuals, ordinary least-squares, spatial lag model, spatial error model, linear mixed model, geographically weighted regression

SPATIAL AUTOCORRELATION (i.e., spatial dependence) and heterogeneity (i.e., spatial nonstationarity) are the two aspects of spatial effects (Anselin and Griffith 1988, Anselin 1990a). Spatial autocorrelation represents the correlations between the values of a random variable at a location and the values of the same variable at "neighboring" locations. The spatial autocorrelation in the error terms of a regression model causes biased estimation of error variance, whereas regression coefficients remain unbiased. Consequently, significance tests and measures of model fit may be misleading (Anselin and Griffith 1988, Dale and Fortin 2002). On the other hand, spatial heterogeneity is defined as structural instability in the form of systematically varying model parameters or different response functions (Anselin and Griffith 1988, Anselin 1990b). Ignoring spatial heterogeneity causes misleading significance tests and suboptimal prediction (Anselin and Griffith 1988). Among different issues with spatial data, spatial heterogeneity has potentially the most damaging effect on the results of data analysis and modeling (Páez and Scott 2004).

Ecologists consider that spatial dependence is caused by underlying spatial processes that result in localized covariation among variables or statistics, and, consequently, clusters of similar or dissimilar values of the variables. In forestry, positive spatial dependence (i.e., trees are surrounded by similar-sized ones) among neighboring trees is mainly due to the microsite effects. It is often observed in young, precanopy closure stands and in older, senescent stands. Intertree competition tends to create a negative autocorrelation (i.e., larger-sized trees are surrounded by smaller trees and vice versa) among neighboring trees (Magnussen 1994, Fox et al. 2001).

A variety of statistical techniques have been developed to quantify spatial dependence such as autocorrelation indices, variogram, correlogram, kriging, Ripley's k-function, and nearest neighbor methods (Dale et al. 2002). These methods have been applied in forestry studies (e.g., Reed and Burkhart 1985, Liu and Burkhart 1994, Kohl and Gertner 1997). Among them, the ones most commonly used are global spatial autocorrelation indices including Moran's I, Geary's C, and Getis' G and G^* (Dale et al. 2002, Páez and Scott 2004). In recent years, local indicators of spatial association have become popular to measure the degree of spatial autocorrelations between locations (Anselin 1995, Boots 2002, Shi and Zhang 2003, Páez and Scott 2004, Wulder et al. 2007). Researchers have also applied a number of statistical regression techniques to incorporate spatial autocorrelation in modeling the relationships between variables, including the spatial lag model (Lichstein et al. 2002, Bullock and Burkhart 2005), spatial error model (Anselin

Lianjun Zhang, Department of Forest and Natural Resources Management, State University of New York, College of Environmental Science and Forestry, One Forestry Drive, Syracuse, NY 13210—Phone: 315-470-6558; Fax: 315-470-6535; lizhang@esf.edu. Zhihai Ma, Department of Forest and Natural Resources Management, State University of New York, College of Environmental Science and Forestry, One Forestry Drive, Syracuse, NY 13210—lizhang@esf.edu. Luo Guo, College of Life and Environmental Sciences, MinZu University of China, 27 Zhong-Guan-Cun South Ave., Beijing 100081, P.R. China—lizhang@esf.edu.

Acknowledgments: We acknowledge the financial support by US Forest Service NE Agenda 2020 Project 05-JV-11242328-009 and by MinZu University of China, 111 Project of MUC (B08044).

Manuscript received November 18, 2008, accepted August 7, 2009

1988), spatial filter model (Borcard and Legendre 2002), moving average autoregression (Fox et al. 2007b), spatial Durbin model (Overmars et al. 2003), linear mixed model and generalized additive model (Zhang and Gove 2005), and others (Dormann et al. 2007). Spatial regression models (i.e., spatial lag and spatial error models) are considered to be effective techniques to directly account for fine-scale spatial autocorrelation in a response variable (Schooley 2006, Dormann et al. 2007).

Li and Reynolds (1995) defined spatial heterogeneity as the spatial complexity and/or variability of the properties of an ecosystem. It may affect many ecological phenomena such as population dynamics, life histories, dispersal patterns, species diversity, predation, and patterns of natural selection (e.g., Legendre 1993, Shiyomi et al. 2000). Wiens (2000) identified four forms of spatial heterogeneity, namely spatial variance, patterned variance, compositional variance, and locational variance, as well as four sources of spatial heterogeneity, namely disturbances, physical processes, geophysical template, and activities of organisms. Therefore, observed spatial heterogeneity may result from more than one biotic (e.g., ecological spatial processes) and abiotic (e.g., environmental factors) process. All sources interact to produce spatial patterns that are characterized as spatial heterogeneity (Wiens 2000, Wagner and Fortin 2005). It is also well-known that spatial heterogeneity is a function of spatial scales, defined as measurement units. Shifting scales may lead to changes in the degree and test for spatial heterogeneity (Dutilleul and Legendre 1993, Li and Reynolds 1995).

Researchers have developed an array of methods and measures for assessing spatial heterogeneity (Dale et al. 2002, Perry et al. 2002). Li and Reynolds (1995) discussed different approaches to quantifying spatial heterogeneity for both categorical and numerical ecological variables. Veldtman (2005) classified various methods of assessing spatial heterogeneity for abundance and occurrence data into three groups: spatially nonexplicit, semiexplicit, and explicit. The commonly used methods and measures include point patterns (randomness indices, aggregation indices, nearest neighbor distance, and quadrat blocking), surface data (trend surface, spectral analysis, variogram, fractal dimension, and autocorrelation indices), and continuous data (variance ratio analysis, correlation analysis, variogram, correlogram, and global and local autocorrelation indices) (e.g., Dutilleul and Legendre 1993, Dale et al. 2002, Wagner and Fortin 2005). However, these methods differ in their information content, biological relevance, and conclusions regarding the forms and degree of spatial heterogeneity (Wiens 2000, Dale et al. 2002, Perry et al. 2002). The problem of incorrectly describing different forms of spatial heterogeneity is partly theoretical and partly methodological, depending on the objectives of study, the methods and measures used, and the dimensionality of spatial heterogeneity (Wiens 2000, Veldtman 2005). In the field of spatial modeling, different regression techniques have been developed to model local variations on complex relationships between random variables over space. These include the spatial expansion method (Anselin 1992), random coefficient model (Fotheringham and Brunsdon 1999), spatial adaptive filtering method (Gorr and Olligschlaeger 1994), and multilevel modeling method (Duncan 1997, Jones 1997). In recent years, geographically weighted regression (GWR) has become popular to explore spatial heterogeneity (Fotheringham et al. 2002, Zhang and Shi 2004, Zhang et al. 2004).

Some consider that spatial heterogeneity may be caused by spatial autocorrelation or spatial dependence (Schooley 2006). Thus, strong spatial autocorrelation (especially significant local indicators of spatial association) implies spatial heterogeneity (Wiens 2000, Perry et al. 2002, Veldtman 2005). Others argue that spatial heterogeneity is not the same as spatial dependence although they may be related (Kupfer and Farris 2007). Spatial autocorrelation and spatial heterogeneity may be observationally equivalent in cross-section data and often occur jointly (Anselin 2001). Therefore, they should be investigated simultaneously for given spatially reference-related data or regression relationships between variables (Graaff et al. 2001).

To date, however, most research assessed model residuals, focusing on either spatial autocorrelation or spatial heterogeneity separately. In this study, we attempted to fit spatial regression models (i.e., spatial lag and spatial error models), the linear mixed-effects model, and GWR to the tree height-diameter relationship, with the ordinary least-squares model as a benchmark. Three forest stands, each representing clustered, random, and regular spatial patterns of tree locations, were used to show the fitting and prediction of these models. Both spatial autocorrelation and heterogeneity in the model residuals were assessed by various measures of spatial statistics. The study provided evaluation and insights on the effectiveness of different regression techniques for dealing with spatial effects (spatial autocorrelation and heterogeneity) in the relationship between tree height and diameter.

Theoretical Background

We describe in brief the regression techniques used in this study.

Ordinary Least Squares

Suppose we have a set of n observations on p independent or predictor variables X and a dependent or response variable y. The relationship between $\mathbf{y}$ and $\mathbf{X}$ can be regressed using ordinary least squares (OLS) as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{y}$ is a vector of the observed response variable, $\mathbf{X}$ is a known model matrix including a column of 1 (for intercept) and p independent variables, $\boldsymbol{\beta}$ is a vector of unknown fixed-effects parameters, and $\boldsymbol{\varepsilon}$ is a vector of random error terms whose distribution is assumed to be $N(0, \sigma^2 I)$, with I denoting an $n \times n$ identity matrix. The OLS estimate of $\boldsymbol{\beta}$ is obtained by (Littell et al. 2006)

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \tag{2}$$

where superscript T denotes the transpose of a matrix. The relationship represented by Equation 1 is assumed to be universal or constant across the geographic area.

Linear Mixed Model

The linear mixed model (LMM) is a special case of generalized linear models and can be expressed as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{3}$$

where $\mathbf{y}$, $\mathbf{X}$, and $\boldsymbol{\beta}$ are as defined in Equation 1, $\mathbf{Z}$ is a known design matrix, $\boldsymbol{\gamma}$ is a vector of unknown random-effects parameters, and $\boldsymbol{\varepsilon}$ is a vector of unobserved random errors. It is assumed that (1) $E(\boldsymbol{\gamma}) = 0$ and $Var(\boldsymbol{\gamma}) = \mathbf{G}$, (2) $E(\boldsymbol{\varepsilon}) = 0$ and $Var(\boldsymbol{\varepsilon}) = \mathbf{R}$, (3) $Cov(\boldsymbol{\gamma}, \boldsymbol{\varepsilon}) = 0$, and (4) both $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$ are normally distributed. The variance of $\mathbf{y}$ is $\mathbf{V} = \mathbf{ZGZ}^T + \mathbf{R}$ and can be estimated by setting up the random-effects design matrix $\mathbf{Z}$ and by specifying covariance structures for $\mathbf{G}$ and $\mathbf{R}$ (Littell et al. 2006). In general, OLS is no longer the best approach to estimating an LMM. Likelihood-based methods (e.g., maximum likelihood [ML]) are usually used to solve for $\boldsymbol{\beta}$. The likelihood function for an LMM is

$$\mathbf{L} = -\tfrac{1}{2}\ln|\mathbf{V}| - \tfrac{1}{2}(\mathbf{r}^T\mathbf{V}^{-1}\mathbf{r}) - \frac{n}{2}\ln(2\pi), \tag{4}$$

where ln is natural logarithm, and $\mathbf{r} = \mathbf{y} - \mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$. Thus, the ML estimate of β is obtained by

$$\hat{\beta}_{LMM} = (\mathbf{X}^T\hat{V}^{-1}\mathbf{X})^{-}\mathbf{X}^T\hat{V}^{-1}\mathbf{y}. \tag{5}$$

LMM can be used to characterize or model the spatial covariance structure in the data and remove the effects of spatial autocorrelation to obtain more accurate estimates for the response variable or treatment means (Littell et al. 2006). In principle, spatial autocorrelation can be reflected in either $\mathbf{G}$ or $\mathbf{R}$ or both. For this study, no random effects were considered because we had only one sample plot for each spatial pattern. Thus, $\mathbf{Z} = 0$ and $\mathbf{V} = \mathbf{R}$, which is also referred as a generalized least-squares model (Pinheiro and Bates 2000, p. 204). The spatial autocorrelation among observations is modeled through $\mathbf{R} = Var(\boldsymbol{\varepsilon})$ such that

$$Var(\boldsymbol{\varepsilon}_i) = \sigma^2 + \sigma_1^2 \tag{6}$$

$$Cov(\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j) = \sigma^2 f(d_{ij}, \theta) \tag{7}$$

where d_{ij} is the distance between locations i and j. This is an LMM with a nugget effect in which $\boldsymbol{\sigma}_1^2$, $\boldsymbol{\sigma}^2 + \boldsymbol{\sigma}_1^2$ and $\boldsymbol{\theta}$ can be associated with the geostatistics parameters nugget, sill, and range, respectively. Different covariance models $f(d_{ij}, \boldsymbol{\theta})$ are available, including spherical, exponential, Gaussian, and power. A likelihood ratio test can be used to determine whether it is necessary to model the spatial covariance structure of the data (Littell et al. 2006, SAS Institute, Inc., 2008). Furthermore, the empirical best linear unbiased predictions (EBLUP) should be used to take spatial autocorrelations into account for predicting the response variable (McCulloch and Searle 2001, Schabenberger and Gotway 2005, Zhang and Gove 2005) as follows:

$$\hat{y}_p = X_p\hat{\beta}_{LMM} + \hat{C}_p\hat{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}_{LMM}), \tag{8}$$

where the subscript p stands for prediction, X_p is the data used for prediction, and $\hat{C}_p$ is the estimated covariance matrix between predicted $\mathbf{y}_p$ and observed $\mathbf{y}$ calculated using the distances between the data point and prediction point (Welham et al. 2004, Schabenberger and Gotway 2005, Lark et al. 2006). Both $\hat{C}_p$ and $\hat{V}$ are estimated based on a specific covariance model $f(d_{ij}, \boldsymbol{\theta})$. Conceptually, EBLUP is universal kriging in geostatistics (Goodvaerts 1997, Schabenberger and Gotway 2005).

Spatial Lag Model

Spatial lag model (SLM) is a formal representation of the spatial diffusion process and captures substantive spatial dependence in the data (Anselin 1993, 2001). It is appropriate when the focus of interest is the assessment of the existence and strength of spatial dependence. The SLM is accomplished by including a spatial lag term of the dependent variable $\mathbf{y}$ into the OLS model (Equation 1) such that

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\rho}\mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon} \\ &= (1 - \boldsymbol{\rho}\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (1 - \boldsymbol{\rho}\mathbf{W})^{-1}\boldsymbol{\varepsilon},\end{aligned} \tag{9}$$

where $\mathbf{W}$ is a row-sum standardized weight matrix, $\mathbf{Wy}$ is a spatially lagged response variable, $\boldsymbol{\rho}$ is a spatial autocorrelation parameter, and $\boldsymbol{\varepsilon}$ is a vector of random error terms with distribution $N(0, \boldsymbol{\sigma}^2 I)$. Equation 9 illustrates that the value of $\mathbf{y}$ at each location is not only determined by $\mathbf{X}$ at that location, but also by the $\mathbf{X}$ at neighboring locations through the spatial multiplier $(1 - \boldsymbol{\rho}\mathbf{W})^{-1}$. Note that $\boldsymbol{\varepsilon}$ is no longer uncorrelated with the predictor variable (i.e., spatial lag $\mathbf{Wy}$), as assumed in Equation 1. Therefore, OLS is no longer a proper parameter estimation method because it yields biased and inefficient estimates (Anselin 1988), and the ML method should be used instead. The likelihood function for the SLM is

$$\begin{aligned}\mathbf{L} = \ln|I - \boldsymbol{\rho}\mathbf{W}| - \frac{n}{2}\ln(2\boldsymbol{\pi}) - \frac{n}{2}\ln(\boldsymbol{\sigma}^2) \\ - \frac{(\mathbf{y} - \boldsymbol{\rho}\mathbf{W}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \boldsymbol{\rho}\mathbf{W}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\boldsymbol{\sigma}^2},\end{aligned} \tag{10}$$

where $\boldsymbol{\beta}$, $\boldsymbol{\rho}$, and $\boldsymbol{\sigma}^2$ are parameters to be estimated (Anselin 1988). The ML estimate of $\boldsymbol{\beta}$ is obtained by

$$\hat{\beta}_{SLM} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{I} - \hat{\rho}\mathbf{W})\mathbf{y}. \tag{11}$$

Because $\mathbf{W}$ is a row-sum standardized weight matrix, $\mathbf{Wy}$ represents in fact the weighted averages of neighboring values of $\mathbf{y}$, referred as to a spatial lag. Thus, the inclusion of the $\mathbf{Wy}$ term allows us to assess the degree of spatial autocorrelation. The estimate of $\boldsymbol{\rho}$ can be considered as an indicator of spatial autocorrelation and is conditional on $\mathbf{W}$ (Anselin 1993, 2001). The predicted value of $\mathbf{y}$ is

$$\hat{y} = (I - \hat{\rho}\mathbf{W})^{-1}\mathbf{X}\hat{\beta}_{SLM}. \tag{12}$$

Spatial Error Model

Spatial error model (SEM) assumes that the spatial autoregressive process occurs only in the error term and neither in the response variable nor in the predictor variables

(Anselin 1993, 2001). It is due to omitted spatially correlated variables or the boundaries of spatial regions not coinciding with actual behavior units (Graaff et al. 2001). It is considered as a special case of a regression with a nonspherical error term and is appropriate when the concern is with correcting for the potential influence of spatial autocorrelation due to the use of spatial data, but irrespective of whether the model itself is spatial or not (Anselin 1993, 2001). The SEM is a combination of the OLS regression model and a spatial autoregressive model in the error term $\boldsymbol{\varepsilon}$ such that

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \lambda W\varepsilon + \boldsymbol{\xi} = \mathbf{X}\boldsymbol{\beta} + (1 - \lambda W)^{-1}\xi, \tag{13}$$

where $\mathbf{W}$ is a row-sum standardized weight matrix, $\mathbf{W}\boldsymbol{\varepsilon}$ is a spatially lagged error term, λ is a spatial autocorrelation parameter, and ξ is a well-behaved error term whose distribution is $\mathrm{N}(0, \boldsymbol{\sigma}^2 I)$. Equation 13 indicates that the value of $\mathbf{y}$ for each location is affected by errors of all locations through the spatial multiplier $(1 - \lambda \mathrm{W})^{-1}$. However, unlike ρ in Equation 9, λ is considered as a nuisance parameter, usually of little interest per se but necessary to correct for the spatial dependence. The mean of $\mathbf{y}$ is not affected by the spatial error dependence (Anselin 1993, 2001). The likelihood function for the SEM is

$$\mathbf{L} = \ln|I - \lambda\mathbf{W}| - \frac{n}{2}\ln(2\boldsymbol{\pi}) - \frac{n}{2}\ln(\sigma^2) \tag{14}$$

$$- \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(I - \lambda W)^{\mathrm{T}}(I - \lambda\mathbf{W})(y - \mathbf{X}\boldsymbol{\beta})}{2\boldsymbol{\sigma}^2},$$

where β, λ, and σ^2 are parameters to be estimated (Anselin 1988). The ML estimate of $\boldsymbol{\beta}$ is obtained by

$$\hat{\beta}_{\mathrm{SEM}} =$$

$$((\mathbf{X} - \hat{\lambda}WX)^{\mathrm{T}}(\mathbf{X} - \hat{\lambda}WX))^{-1}(\mathbf{X} - \hat{\lambda}WX)^{\mathrm{T}}(\mathbf{y} - \hat{\lambda}Wy). \tag{15}$$

The predicted values of $\mathbf{y}$ are

$$\hat{y} = \mathbf{X}\hat{\beta}_{\mathrm{SEM}}. \tag{16}$$

Geographically Weighted Regression

Suppose that one has a set of location coordinates (u_i, v_i) for each observation i. The underlying model for geographically weighted regression (GWR) is

$$\mathbf{y} = \boldsymbol{\beta}_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)\mathbf{X}_k + \boldsymbol{\varepsilon}, \tag{17}$$

where $\{\boldsymbol{\beta}_0(u_i, v_i), \boldsymbol{\beta}_1(u_i, v_i), \ldots, \boldsymbol{\beta}_{\mathrm{p}}(u_i, v_i)\}$ are $(p + 1)$ regression coefficients for the location (u_i, v_i) in the study area. Again, $\boldsymbol{\varepsilon}$ is a vector of the random error terms with a distribution $\mathrm{N}(0, \sigma^2 I)$. The aim of GWR is to obtain the estimates of these coefficients for each independent variable X and at each geographic location i using the neighbors within a given bandwidth and weighted least-squares regression such that

$$\hat{\beta}_i = (X^{\mathrm{T}}\mathbf{W}_i\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}_i\mathbf{y}, \tag{18}$$

where $\mathbf{W}_i$ is a geographical weight matrix for the center i such that $\mathbf{W}_i = f(d_i, h)$ in which d_i is the distance vector between the center i and all neighbors and h is the bandwidth or decay parameter. There is a separate geographical weight matrix ($\mathbf{W}_i$) for each observation location in the data. If $\mathbf{W}_i = I$ (identity matrix); that is, if each observation in the data has a weight of unity, the GWR model is equivalent to the OLS model. Note that Equation 18 is not a single equation but an array of equations, with each $\boldsymbol{\beta}_i$ corresponding to a row of the matrix whose elements are β_{ij}. Therefore, one can obtain a set of estimates of spatially varying parameters without specifying a function form for spatial variation (Brunsdon et al. 1998, Fotheringham et al. 2002).

The commonly used weight functions $f(d_i, \theta)$ include a fixed spatial kernel such as a Gaussian distance decay kernel function, in which the distance bandwidth at each center i is a constant across the study area (Foody 2004, Zhang et al. 2004, Bickford and Laffan 2006) and an adaptive spatial kernel such as a bi-square distance decay kernel function, in which the distance bandwidth is selected such that the number of observations with nonzero weights is the same at each center i across the study area. Thus, it has larger bandwidths where the data are sparse and smaller bandwidths where the data are denser (Foody 2003, Farber and Páez 2007, Kupfer and Farris 2007). Changing the spatial kernel function and/or the bandwidth may change the coefficient estimates for the GWR models and result in different model performance and residuals (Guo et al. 2008).

Data and Methods

Data

The data used in this study were the stem map data collected from three forest stands located near Sault Ste. Marie, Ontario, Canada (Ek 1969). The three stands represented distinctly different forest types with different spatial patterns of tree locations. The first stand was a mature, second growth, and uneven-aged softwood stand (approximately 400 × 200 m in size with 6,881 trees). Tree species included balsam fir (*Abice balsamea* [L.] Mill.), black spruce (*Picea mariana* [Mill.] BSP.), a few scattered large white pine (*Pinus strobus* L.), and other minor species. The second stand was a mature and uneven-aged hardwood stand (approximately 800 × 200 m in size with 5,503 trees), which was predominantly sugar maple (*Acer saccharum* Marsh.) with some yellow birch (*Betula alleghaniensis* Britt.). The third stand was a red pine (*Pinus resinosa* Ait.) plantation (approximately 470 × 150 m in size with 9,066 trees), in which the red pine trees were originally planted in 1929 with a spacing of approximately 1.8 × 1.8 m and were thinned in 1956 by removing alternate rows (Ek 1969).

Ek (1969) computed the Pielou (1959) index of nonrandomness for the three stands: the softwood stand had a Pielou index of 1.35, representing a clustered spatial pattern of tree locations, the hardwood stand had a Pielou index of 0.98, indicating a random spatial pattern; and the pine plantation had a Pielou index of 0.77, showing a regular/uniform spatial pattern. Measurements on trees included tree location coordinates, dbh, total height (HT), and crown area (not available for the pine plantation). The

Table 1. Descriptive statistics of tree variables in the three plots

Plot	Variable	Mean	SD	Min	Max	Moran's *I*	SH%	Distance (m)*
Softwood	HT (m)	13.27	2.98	3.96	32.92	0.21	75.2	2.1 (0.3–11.0)
	dbh (cm)	17.76	6.82	8.90	84.07	0.11	62.2	
Hardwood	HT (m)	19.39	3.95	6.10	39.01	0.15	52.5	4.1 (0.3–13.0)
	dbh (cm)	32.76	12.49	8.89	100.6	0.06	50.3	
Plantation	HT (m)	15.92	1.29	6.65	19.20	−0.01	—	1.9 (0.3–4.1)
	dbh (cm)	7.60	1.67	2.00	12.80	0.03	—	

* Data are mean (range).

records were taken for trees ≥8.9 cm (3.5 in.) in dbh for both softwood and hardwood stands and for trees ≥2.54 cm (1.0 in.) in dbh for the pine plantation. Tree height-diameter equations were also developed based on sample trees and were used to calculate the tree total heights if the height measurements were not available (Ek 1969).

However, we encountered difficulties of insufficient computer memory and extremely long computing hours when we used all tree records in the pine plantation to fit regression models. Thus, we decided to select a subplot of 150 × 150 m in size with 2,350 trees, which was a close representation of the stand. Descriptive statistics of the tree HT (m) and dbh (cm) are listed in Table 1, and the location maps of the tree dbh are shown in Figure 1 for the three plots.

Regression Model

Because the scatterplots of the tree HT against dbh were quadratic in shape for the three plots (not shown), we chose the following model to fit the HT-dbh relationship:

$$\ln(\mathrm{HT}) = \beta_0 + \beta_1 \cdot \ln(\mathrm{dbh}) + \varepsilon, \qquad (19)$$

where ln is natural logarithm, β_0 and β_1 are regression coefficients to be estimated from data, and ε is the model error term. Model residuals were defined as the difference between the observed and predicted ln(HT). However, we emphasize that we did not intend to develop a predictive HT-dbh model. Rather, we attempted to investigate the spatial autocorrelation and heterogeneity of the model errors from the six regression methods for fitting the HT-dbh relationship, given the three forest types with different spatial patterns.

Model Fitting

In this study, OLS and LMM were fitted using SAS (SAS Institute, Inc., 2008). For LMM, the exponential covariance structure was selected from available structures (e.g., spherical, exponential, Gaussian, and power) according to Akaike's information criterion (AIC) (Littell et al. 2006). The predictions for ln(HT) were obtained based on the EBLUP to account for the spatial autocorrelations among trees (Zhang and Gove 2005). Both SLM and SEM models were fitted by GeoDa software (Anselin et al. 2006, GeoDa Center for Geospacial Analysis and Computing 2009). Three types of spatial weights were tried for each model including first-order rook contiguity, distance bandwidth, and *k*-nearest neighbors. For the distance bandwidth $h = 10$ m was used for the softwood plot, $h = 18$ m was used for the hardwood plot, and $h = 7$ m was used for the plantation plot (which were the same as $\mathrm{GWR_h}$ defined below). For the k-nearest neighbors $n = 20$ trees was used (which was the same as $\mathrm{GWR_n}$ defined below).

The GWR model was fitted by GWR 3.0 (Charlton et al. 2003, Fotheringham et al. 2009) using two distance-decay functions: a fixed Gaussian kernel function with a fixed distance bandwidth (namely $\mathrm{GWR_h}$) and an adaptive bi-square kernel function with a fixed number of observations (namely $\mathrm{GWR_n}$) (Fotheringham et al. 2002, Guo et al. 2008). To make $\mathrm{GWR_h}$ and $\mathrm{GWR_n}$ compatible in terms of the neighboring trees included for fitting the models, we selected the following bandwidth for $\mathrm{GWR_h}$: $h = 10$ m for the softwood plot, $h = 18$ m for the hardwood plot, and $h = 7$ m for the plantation plot, which roughly included 20 trees within the range of the given bandwidth. Thus, the bandwidth for $\mathrm{GWR_n}$ was $n = 20$ trees for all three plots.

The overall model fitting was evaluated by three statistics including the coefficient of determination (R^2), mean sum of squares of error (MSE), and AIC. Although both model R^2 and MSE may be biased because they are based on the assumption of independent observations. We used them as the approximation measures of the model fitting. In this case, the AIC criterion is more appropriate because the likelihood function does not rely on the assumption of independent error terms (Littell et al. 2006).

Model Assessment

The spatial autocorrelations among the model residuals were investigated using both global and local Moran coefficients (Anselin 1995, Boots 2002). Moran's *I* is positive when the observed values of locations within the distance tend to be similar, negative when they tend to be dissimilar, and approximately zero when the observed values are arranged randomly and independently over space (Lee and Wong 2001, Zhang and Gove 2005). A Moran scatterplot was used to investigate the spatial autocorrelations among model residuals. Then, spatial correlograms with a 5-m lag increment were obtained to explore the changes in Moran's *I* of model residuals across different lag distances (Isaaks and Srivastava 1989). The spatial distribution of the local Moran's *I* was used to show the "hot spots" of residual clusters (i.e., the same sign of model residuals) and "cold spots" of residual clusters (i.e., the opposite sign of model residuals) (Zhang and Gove 2005, Zhang et al. 2005). For

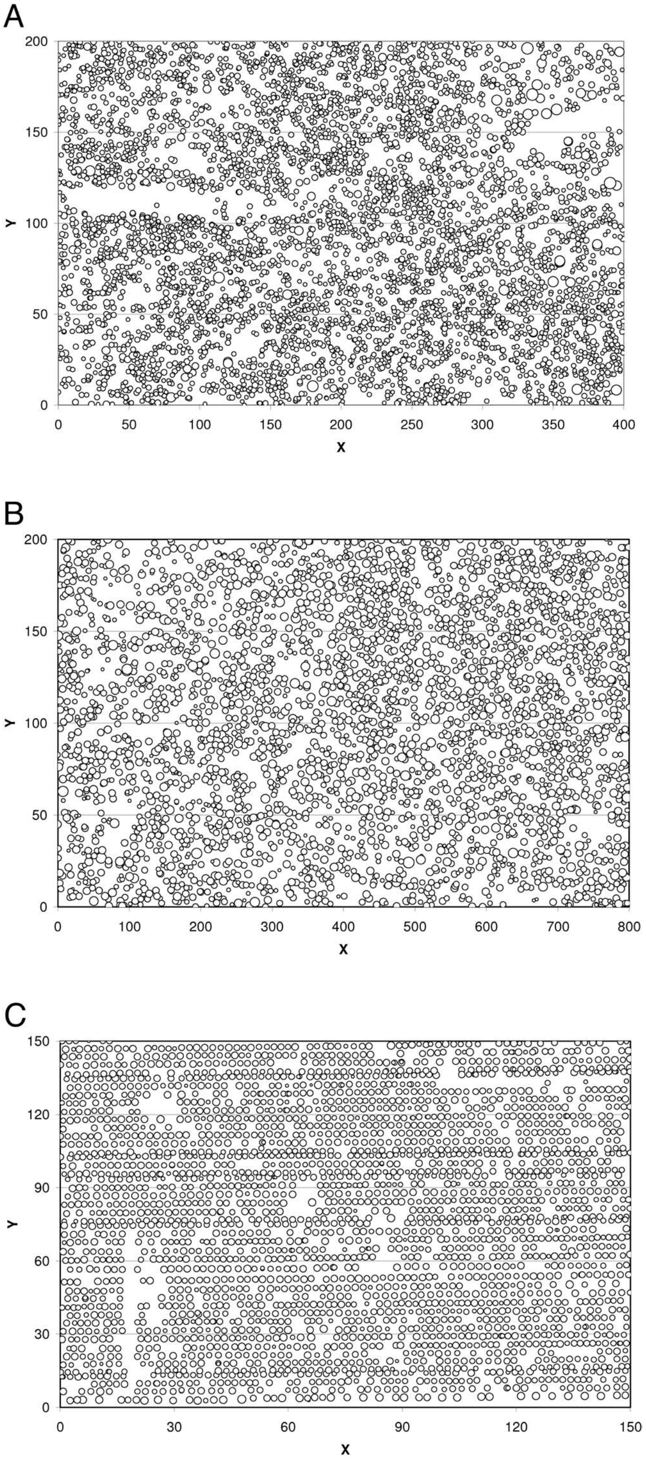

Figure 1. Map of tree locations for (A) softwood (clustered), (B) hardwood (random), and (C) pine plantation (regular). The circles are proportional to the sizes of tree dbh.

the spatial regression models, several regression diagnostics in the GeoDa software were used for testing non-normality, heteroscedasticity, and spatial autocorrelation in model residuals (Anselin 2005).

To quantify the spatial heterogeneity of model residuals, a semivariogram was fit to the residuals from each model. Two of the parameters of each semivariogram, sill variance and nugget variance, were obtained. The difference between the sill and nugget is called partial sill (Schabenberger and Gotway 2005). The relative spatial heterogeneity (SH%) was calculated as (Biondi et al. 1994, Li and Reynolds 1995)

$$\text{SH\%} = \left(\frac{\text{sill} - \text{nugget}}{\text{sill}}\right) \cdot 100. \tag{20}$$

It is also called relative structured variability (Schabenberger and Gotway 2005) or structured spatial variation (Fox et al. 2007a). Because sill represents the maximum or total spatial variation and nugget represents the random spatial variation (e.g., measurement error), SH% represents the proportion of the autocorrelated spatial heterogeneity (Li and Reynolds 1995) or the degree of spatially structured variability (Schabenberger and Gotway 2005, Fox et al. 2007a). A large value of SH% indicates a high level of spatial heterogeneity or a low level of spatial randomness.

The spatial variability can also be decomposed at three spatial scales (Myers 1997, Chiles and Delfiner 1999, Garrigues et al. 2006) as follows:

(1) Intrablock spatial variance:

$$S_{\text{intra}} = \frac{1}{B}\sum_{g=1}^{B}\frac{1}{n_g}\sum_{h=1}^{n_g}(e_{gh} - \bar{e}_g)^2 \tag{21}$$

(2) Interblock spatial variance:

$$S_{\text{inter}} = \frac{1}{B}\sum_{g=1}^{B}(\bar{e}_g - \bar{e})^2 \tag{22}$$

(3) Total spatial variance:

$$S_{\text{total}} = S_{\text{inter}} + S_{\text{intra}} \tag{23}$$

where e_{gh} is the hth model residual in the gth block, $\bar{e}_g$ is the mean of model residuals in the gth block, and $\bar{e}$ is the overall mean of model residuals of the whole plot. Thus, S_{inter} indicates the regional spatial variability, whereas S_{intra} quantifies the local spatial variability. Both dispersion variances are a function of block size; i.e., S_{inter} decreases and S_{intra} increases as the size of the blocks increases (Wackernagel 2003, Garrigues et al. 2006). In this study we evaluated both S_{intra} and S_{inter} of model residuals from SLM, SEM, LMM, and GWR against those from OLS at a range of block sizes (spatial scales), i.e., 5, 10, 15, 20, and 30 m.

Results and Discussion

Spatial Characteristics of Three Plots and Tree Variables

The softwood stand had a clustered spatial pattern of tree locations (Figure 1A) with a few large-sized trees and two big gaps on the northeast corner and west side of the plot. The average distance between trees was relatively small (mean = 2.1 m and range = 0.3–11.0 m). The trees in the hardwood plot had a random spatial pattern (Figure 1B), as well as larger tree sizes and longer distance from each other (mean = 4.1 m, and range = 0.3–13.0 m). It appeared that HT in the softwood plot had larger spatial autocorrelations (Moran's I = 0.21) and higher spatial heterogeneity (SH% = 75.2%) than those in the hardwood plot (Moran's I =

Table 2. Model fitting statistics for the six regression models and measures of spatial autocorrelation and heterogeneity for the model errors

Plot	Model	R^2	MSE	AIC	Moran's I	SH%	OLS%*
Softwood (clustered)	OLS	0.61	0.01781	−5,819.9	0.30	74.9	100
	SLM	0.67	0.01526	−6,384.7	0.09	73.5	98.1
	SEM	0.70	0.01393	−6,650.5	−0.02	69.7	93.1
	LMM	0.91	0.00423	−6,929.0	−0.12	29.7	39.7
	GWR_h	0.72	0.01363	−6,847.4	0.12	60.5	80.8
	GWR_n	0.84	0.01017	−6,562.5	−0.06	30.3	40.5
Hardwood (random)	OLS	0.64	0.01948	−3,576.3	0.28	55.0	100
	SLM	0.67	0.01758	−3,852.7	0.11	43.3	78.7
	SEM	0.71	0.01592	−4,069.7	−0.03	42.1	76.5
	LMM	0.81	0.01031	−4,218.6	−0.06	25.6	46.5
	GWR_h	0.75	0.01343	−4,423.6	0.07	35.4	47.3
	GWR_n	0.81	0.01020	−4,343.7	−0.10	17.2	31.3
Plantation (regular)	OLS	0.91	0.00079	−12,319.3	0.05	—	—
	SLM	0.90	0.00079	−12,317.3	0.05	—	—
	SEM	0.90	0.00078	−12,335.2	0.01	—	—
	LMM	0.93	0.00065	−12,298.0	−0.04	—	—
	GWR_h	0.93	0.00058	−12,891.6	−0.01	—	—
	GWR_n	0.95	0.00040	−13,137.0	−0.11	—	—

* OLS% is the percentage of model SH% against the SH% of the OLS model.

0.15 and SH% = 52.5%) (Table 1). In contrast, the trees in the pine plantation with a regular spatial pattern were relatively smaller in diameter than trees in other two plots and were not only regularly distributed in space (mean = 1.9 m, and range = 0.3–4.1 m) but also more uniform in size (Table 1; Figure 1). There was almost no spatial autocorrelation for HT (Moran's $I < -0.01$), and SH% was not available because no semivariogram could be computed. The correlograms (not shown) of the two tree variables indicated that HT had relatively a higher Moran's I across a range of lag distances than did dbh for both softwood and hardwood plots.

Model Fitting

The OLS model fitted the data of the three plots reasonably well, if the violation of the independence assumption was ignored. The R^2 of the OLS models was 0.61 for the softwood plot, 0.64 for the hardwood plot and 0.91 for the plantation plot (Table 2).

Three types of spatial weights (i.e., first-order rook contiguity, distance bandwidth, and k-nearest neighbors) in the GeoDa software (Anselin et al. 2006) were used to fit the SLM and SEM. However, there was almost no difference among the three types for either the SLM or SEM in terms of spatial autocorrelation and heterogeneity in the model residuals. Thus, we chose to show only the results from the SLM and SEM with the rook spatial weight. Table 3 provides several regression diagnostics on spatial dependence for the OLS model residuals and the results were as follows: the OLS model residuals had significant spatial autocorrelation for the three plots, but the spatial autocorrelation (global Moran's I) was 0.30 for the softwood plot, 0.28 for the hardwood plot, and only 0.05 for the plantation plot and the values of the testing statistics for the SEMs [i.e., Lagrange multiplier (LM) (error) and robust LM (error)] were higher than those for the SLMs [i.e., LM (lag) and robust LM (lag)], indicating that the SEM was a more appropriate alternative for the HT-dbh relationship in the three plots than was the SLM (Anselin 2005, p. 199). Nevertheless, we fitted both SLMs and SEMs for the purpose of comparison because they deal with spatial autocorrelation in different ways. To account for the spatial autocorrelations, the SLM uses the spatial lag of the dependent variable as one of the predictor variables, whereas the SEM deals with spatial autocorrelation in the model error term. Both SLMs and SEMs fitted the data of the three plots better (i.e., larger R^2, smaller MSE, and smaller AIC) than the OLS models (Table 2), suggesting improvement in model fitting for both the SLM and SEM. In general, the SEM fitted the data slightly better than did the SLM (i.e., smaller MSE, AIC, global Moran's I, and SH%) for the three plots (Table 2).

The LMM is capable of incorporating the spatial dependence among data observations into parameter estimation and model prediction (Littell et al. 2006). Table 2 reveals that the LMM fitted the data much better (i.e., larger R^2, smaller MSE, and smaller AIC) than did the OLS model, SLM, and SEM, especially for the softwood and hardwood

Table 3. Spatial dependence diagnostics for the OLS model residuals

Plot	Test statistics	Value	P
Softwood (clustered)	Moran's I	0.30	<0.0001
	LM (lag)	805.2	<0.0001
	Robust LM (lag)	29.1	<0.0001
	LM (error)	1268.8	<0.0001
	Robust LM (error)	492.7	<0.0001
Hardwood (random)	Moran's I	0.28	<0.0001
	LM (lag)	328.3	<0.0001
	Robust LM (lag)	0.3124	0.5762
	LM (error)	708.3	<0.0001
	Robust LM (error)	380.3	<0.0001
Plantation (regular)	Moran's I	0.05	<0.0001
	LM (lag)	0.00003	0.9960
	Robust LM (lag)	1.14	0.2850
	LM (error)	16.6	<0.0001
	Robust LM (error)	17.8	<0.0001

plots. Table 4 shows that the standard errors of the LMM parameters were smaller than those of the OLS model for both softwood and hardwood plots owing to the positive spatial autocorrelations in HT. In contrast, the standard errors of the LMM parameters were the same as those of the OLS model for the plantation plot because of the trivial spatial autocorrelations in HT (Table 1).

Both the GWR_h and GWR_n models fitted the data of the three plots much better than the OLS model with higher R^2, lower MSE, and smaller AIC for each of the three plots (Table 2). The Monte Carlo test for spatial stationarity (Fotheringham et al. 2002) indicated that the two coefficients of the GWR_h and GWR_n models were significantly nonstationary over space for both softwood and hardwood plots, whereas they displayed nonsignificance for the plantation plot. The averages of the two regression coefficients derived from either the GWR_h or GWR_n models were similar to those of the OLS models for the three plots (Table 4). On the other hand, the ranges of the two regression coefficients derived from the GWR_n models were much wider than those from the GWR_h models, indicating that the GWR_h model produced smoother spatial distributions for the regression coefficients (Guo et al. 2008). In addition, the GWR_n model fitted the data of the three plots better than did the GWR_h model in terms of model R^2 and MSE, whereas its AIC values were relatively larger because it had a larger effective number of model parameters to fit (de Smith et al. 2007, p. 247).

Figure 2 shows the absolute residuals from the six models across a range of tree diameter classes. For the softwood plot, the LMM model produced the smallest model absolute residuals for different sizes of trees, followed by the GWR_n model (Figure 2A). For the hardwood plot and plantation plot, the GWR_n model yielded the smallest model absolute residuals, especially for small-sized and large-sized trees (Figure 2B and C).

Spatial Autocorrelation of Model Residuals

The SLM, SEM, LMM, and two GWR models produced significantly smaller global Moran's I values than did the OLS model for all three plots, except the SLM and GWR_n model for the plantation. Among them, the SEM, LMM, and GWR_n model generally produced negative global Moran's I values, whereas other three models produced positive ones (Table 2). Figure 3 illustrates the spatial correlograms of the model residuals from the six models for the three plots. For the softwood plot, the SEM produced relatively consistent Moran's I (<0.05) across different lag distances. At the 5-m lag distance, the LMM and GWR_n model have negative Moran's I and the GWR_h model has positive Moran's I, which quickly decreases to zero at the 10-m lag distance. The SLM has larger Moran's I and the OLS model has the largest Moran's I across the lag distances (Figure 3A). For the hardwood plot, the model residuals of the six regression models behave similarly as for the softwood plot, except that the Moran's I of the SEM, GWR_n model, and GWR_h model approaches zero at approximately the 20-m lag distance (Figure 3B). In contrast, for the plantation plot, the Moran's I of the mode residuals of the six regression models is very small across different lag distances and quickly approaches zero after the 5-m lag distance (Figure 3C). In fact, the lag distance at which the Moran's I approaches zero for the three plots is compatible with the bandwidth we chose for the GWR_h model (i.e., $h = 10$ m for the softwood plot, $h = 18$ m for the hardwood plot, and $h = 7$ m for the plantation plot).

The local Moran's coefficient (I_i) was computed for each model residual from each of the six regression models. The spatial distributions of the local Moran's I_i are illustrated in Figures 4, 5, and 6 for the three plots. Recall that a positive local Moran's I_i indicates hot spots of the clusters with the same sign of the model residuals (black dots in the figures), whereas a negative local Moran's I_i shows cold spots of the

Table 4. Regression coefficient estimates and standard errors of the coefficients of the six models

Plot	Model	Regression coefficient $\hat{\beta}_0$ ($S_{\hat{\beta}_0}$)	Regression coefficient $\hat{\beta}_1$ ($S_{\hat{\beta}_1}$)	Spatial parameter
Softwood (clustered)	OLS	0.9217 (0.0187)	0.5803 (0.0066)	—
	SLM	0.1038 (0.0342)	0.5525 (0.0063)	$\hat{\rho} = 0.3498$
	SEM	0.9677 (0.0059)	0.5650 (0.0169)	$\hat{\lambda} = 0.5451$
	LMM	0.9633 (0.0181)	0.5363 (0.0060)	—
	GWR_h	0.9716 (−0.2228∼1.9271)	0.5636 (0.2361∼0.9850)	—
	GWR_n	0.9750 (−1.5533∼4.5182)	0.5637 (−0.8291∼1.4095)	—
Hardwood (random)	OLS	1.4139 (0.0202)	0.4477 (0.0059)	—
	SLM	0.5572 (0.0511)	0.4388 (0.0057)	$\hat{\rho} = 0.3016$
	SEM	1.4122 (0.0054)	0.4960 (0.0217)	$\hat{\lambda} = 0.4960$
	LMM	1.4023 (0.0215)	0.4503 (0.0053)	—
	GWR_h	1.4048 (0.35167∼2.3369)	0.4514 (0.2065∼0.7264)	—
	GWR_n	1.4840 (−1.7916∼3.2273)	0.4289 (−0.0337∼1.2820)	—
Plantation (regular)	OLS	2.0762 (0.0043)	0.3441 (0.0021)	—
	SLM	2.0761 (0.0224)	0.3441 (0.0021)	$\hat{\rho} = 4.06E{-5}$
	SEM	2.0749 (0.0042)	0.3447 (0.0021)	$\hat{\lambda} = 0.1201$
	LMM	2.0745 (0.0043)	0.3446 (0.0021)	—
	GWR_h	2.0923 (1.7581∼2.3134)	0.3368 (0.2251∼0.5031)	—
	GWR_n	2.1318 (1.1446∼2.5817)	0.3175 (0.1084∼0.7655)	—

Numbers in parentheses are the standard errors of the regression coefficients for OLS, SLM, SEM, and LMM and the ranges of the coefficients for the GWR_h and GWR_n models.

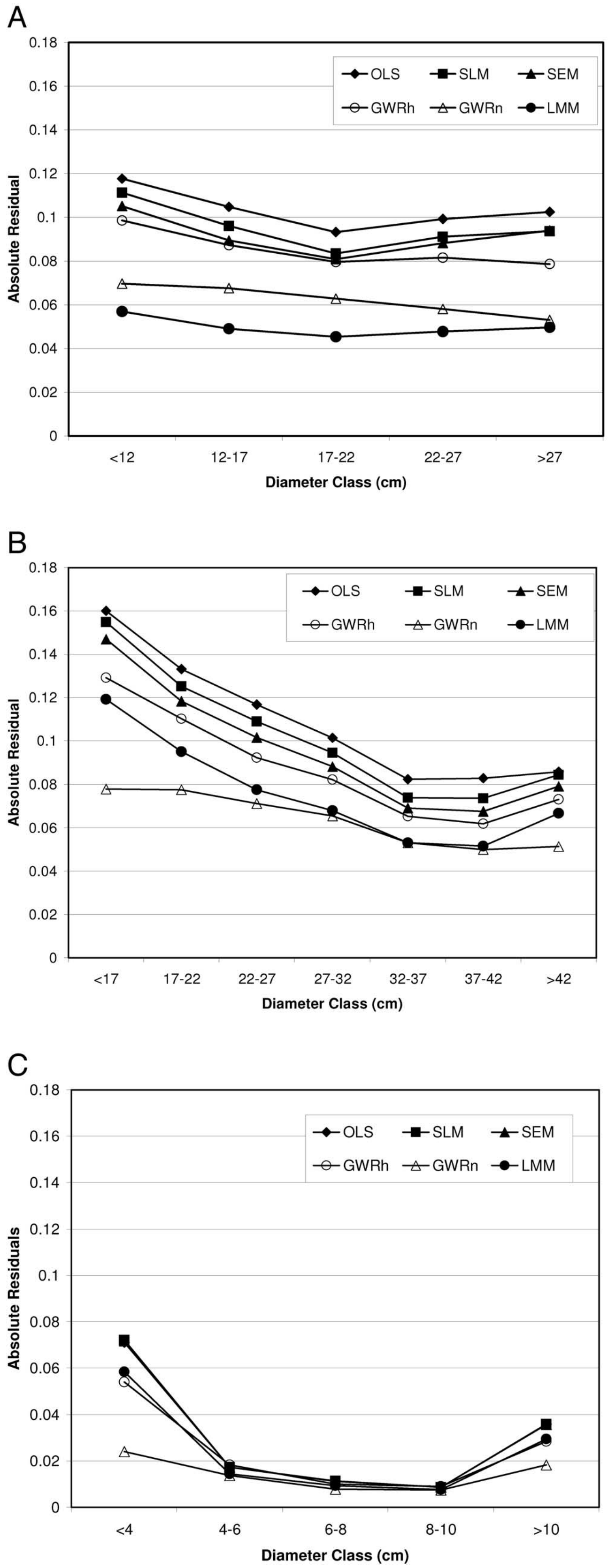

Figure 2. Absolute model residuals across dbh classes for (A) softwood (clustered), (B) hardwood (random), and (C) pine plantation (regular).

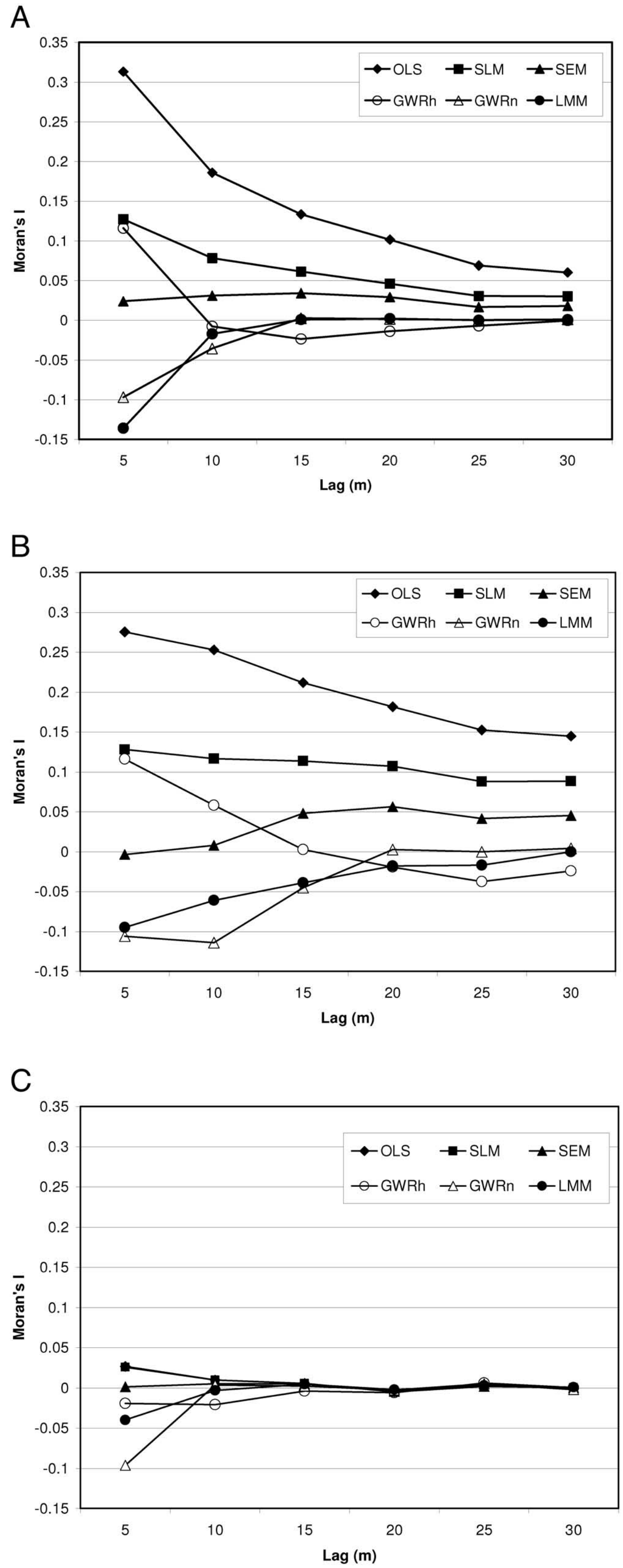

Figure 3 Correlogram of model residuals for (A) softwood (clustered), (B) hardwood (random), and (C) pine plantation (regular).

clusters with the opposite sign of the model residuals (circles in the figures). In general, the local Moran's I_i for the model residuals from the OLS model and SLM has similar spatial patterns with more and larger positive local Moran's I_i values (black dots) for the three plots, indicating the clusters of either positive or negative model residuals (Figures 4A and B, 5A and B, and 6A and B). The SEM and

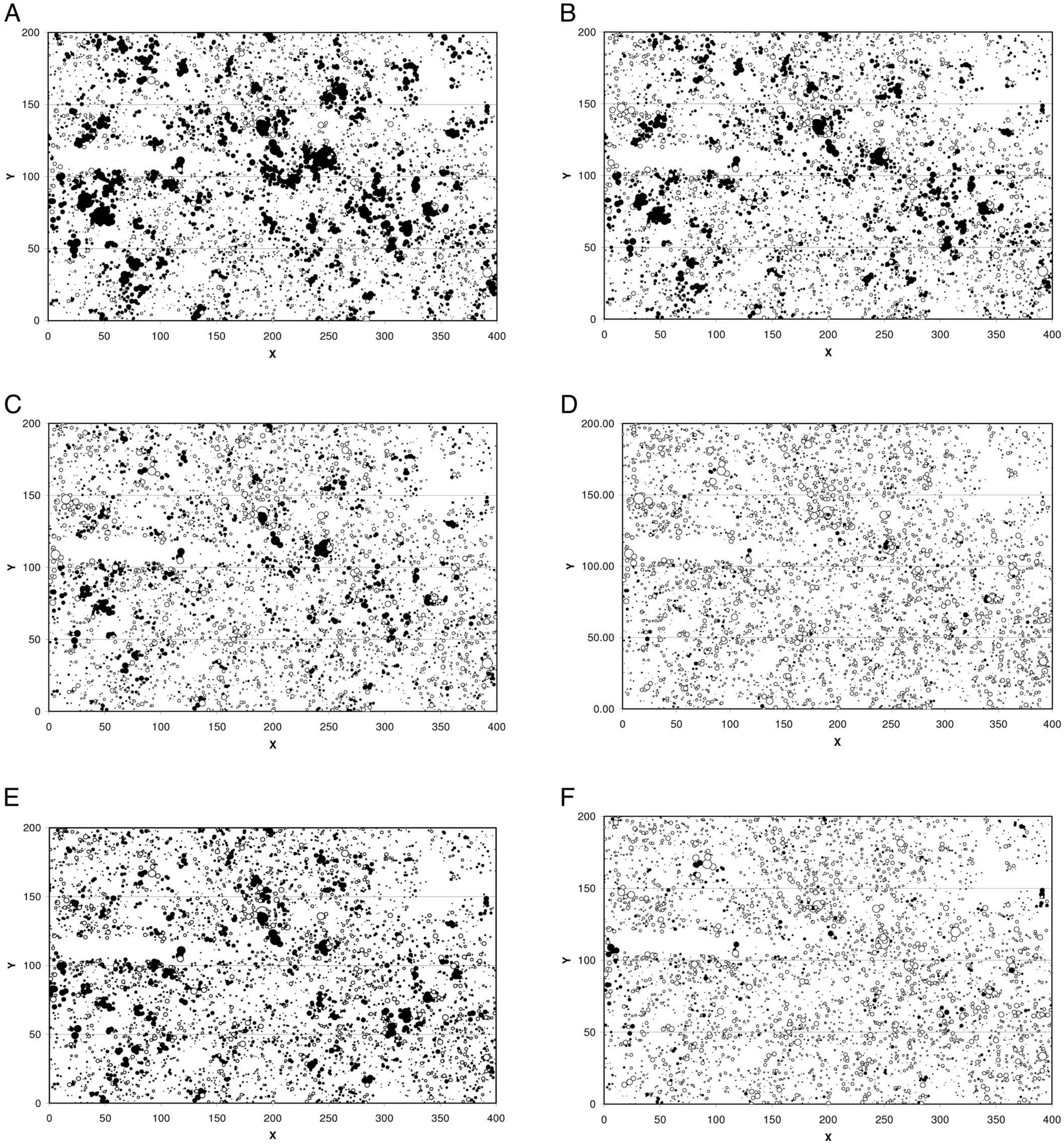

Figure 4 Distribution of local Moran's *I* of model residuals for the softwood plot (A) OLS, (B) SLM, (C) SEM, (D) LMM, (E) GWR_h, and (F) GWR_n.

GWR_h model produce fewer hot spots than the OLS model and SLM and have similar spatial patterns for the local Moran's I_i (Figures 4C and E, 5C and E, and 6C and E). In contrast, the LMM and GWR_n model produce more negative local Moran's I_i values (circles) or more cold spots of dissimilar model residuals (Figures 4D and F, 5D and F, and 6D and F), which is more desirable because neighboring residuals have opposite signs and may result in a zero average for any local area.

Spatial Heterogeneity of Model Residuals

Because SH% is calculated from the semivariograms of model residuals, it represents the overall spatial heterogeneity or the degree of spatially structured variability in the model residuals. We computed the percentage of the SH% of other five models against the SH% of the OLS model (i.e., the OLS% column in Table 2). For the softwood plot, it was evident that the SH% values of both the SLM and

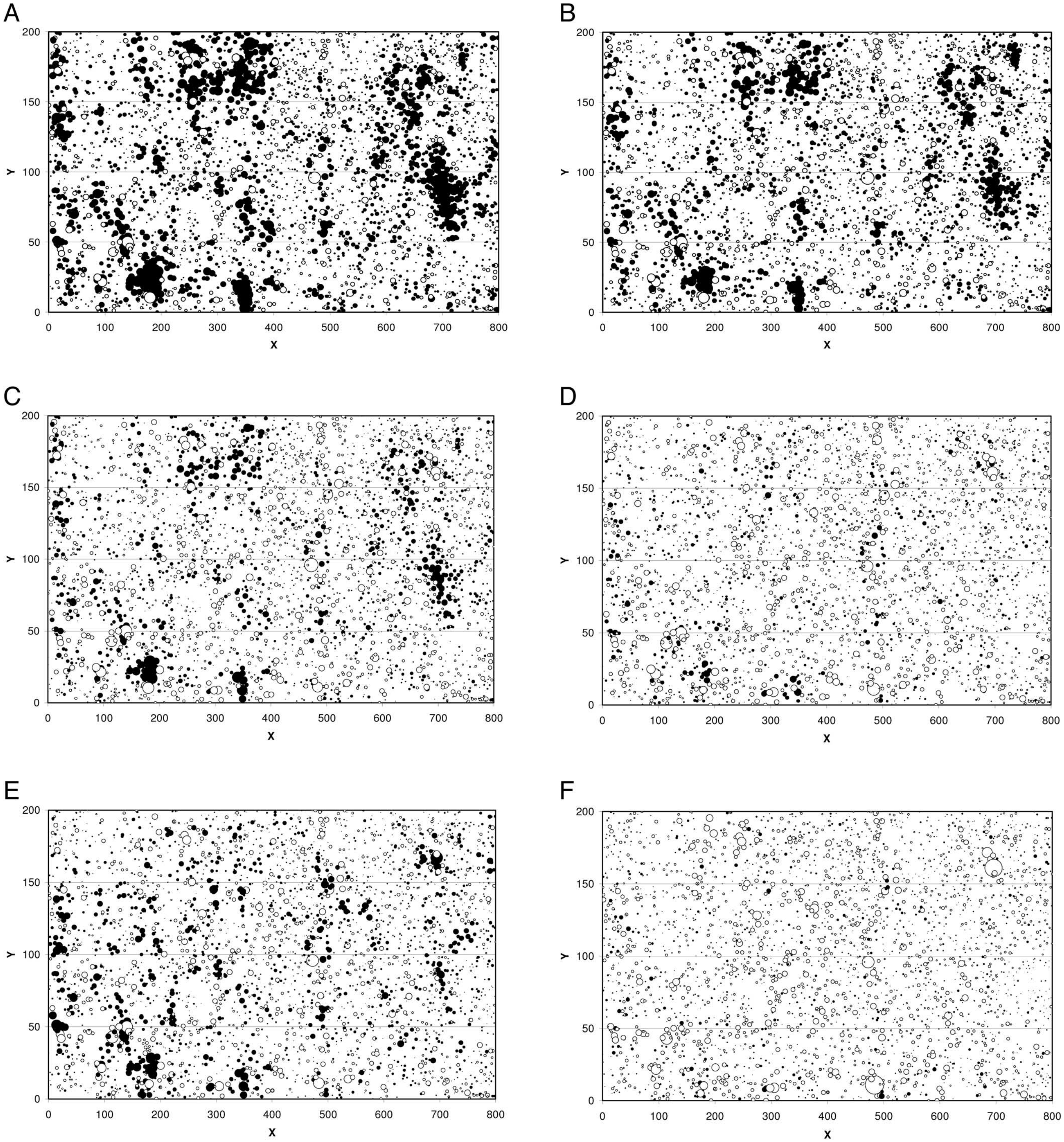

Figure 5. Distribution of local Moran's *I* of model residuals for the hardwood plot (A) OLS, (B) SLM, (C) SEM, (D) LMM, (E) GWR_h, and (F) GWR_n.

SEM were similar to the SH% of the OLS model, the SH% of the GWR_h model was reduced to approximately 81% of the SH% of the OLS model, and the SH% of the LMM and GWR_n model was decreased to approximately 40% of the SH% of the OLS model (Table 2). For the hardwood plot, the SLM and SEM reduced their SH% to approximately 78% of the SH% of the OLS model, the LMM and GWR_h model decreased their SH% to approximately 47% of the SH% of the OLS model, and the GWR_n model significantly reduced SH% to approximately 31% of the SH% of the OLS model (Table 2).

Figure 7 illustrates the intrablock variance in model residuals, which represents the average local spatial variability for a given block size. For the softwood plot, the LMM has the smallest intrablock variance (local spatial variability) across different block sizes, followed by the GWR_n model. The OLS model, SLM, and SEM have much larger and similar intrablock variances across different block sizes. The intrablock variance of the GWR_h model is smaller than that of the latter three models (Figure 7A). However, the differences of the intrablock variance among the six models are relatively small for the 5-m block size

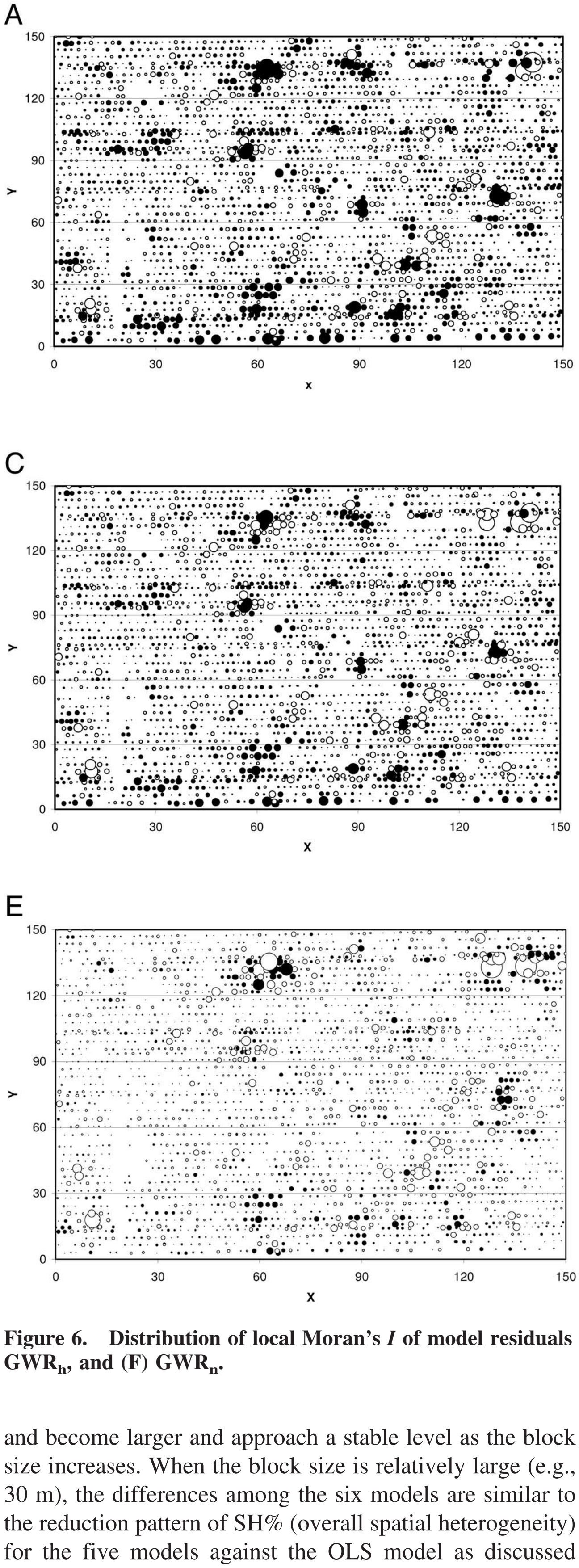

Figure 6. Distribution of local Moran's *I* of model residuals for the plantation plot (A) OLS, (B) SLM, (C) SEM, (D) LMM, (E) GWR_h, and (F) GWR_n.

and become larger and approach a stable level as the block size increases. When the block size is relatively large (e.g., 30 m), the differences among the six models are similar to the reduction pattern of SH% (overall spatial heterogeneity) for the five models against the OLS model as discussed above (Table 2).

For the hardwood plot, the GWR_n model has the smallest local spatial variability across different block sizes, followed by the LMM and GWR_h model. Again, the OLS model, SLM, and SEM have much larger and similar intrablock variances across different block sizes (Figure 7B). However, the intrablock variances of the six models are still increasing as the block size increases to 30 m. In contrast, for the plantation plot, the intrablock variances in model residuals of the six models behave similarly to those of the hardwood plot, except these intrablock variances approach a stable level at the block size of 10 m (Figure 7C). Note that the *y*-axis range of the intrablock variance for the plantation plot is much smaller than those of the softwood and hardwood plots. The pattern and behavior of interblock variance

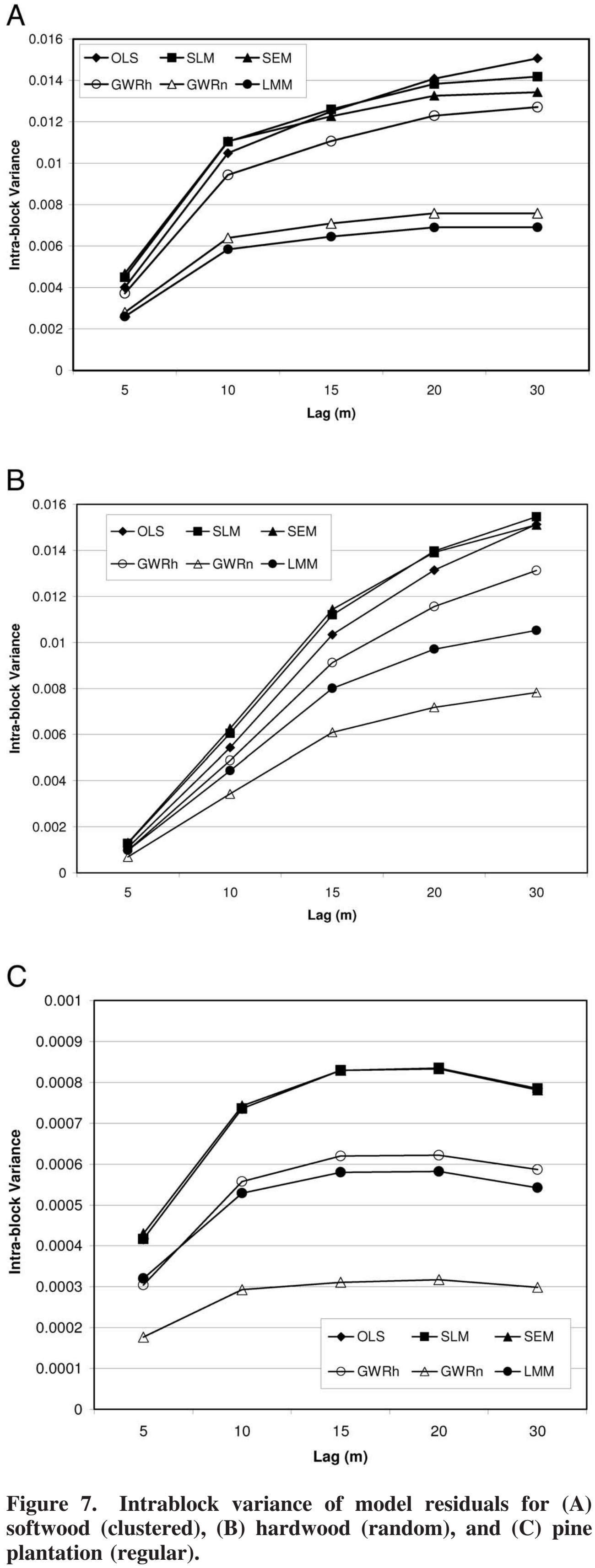

Figure 7. Intrablock variance of model residuals for (A) softwood (clustered), (B) hardwood (random), and (C) pine plantation (regular).

(average regional variability) were similar among the six models for the three plots, but in an opposite direction; i.e., the interblock variance decreases as the block size increases (not shown).

Discussion

We investigated the spatial autocorrelation and heterogeneity of the model residuals from six regression techniques. The model performance depended on their characteristics and nature. Spatial regression models (e.g., SLM and SEM) are designed to obtain a significant improvement over a traditional OLS model by incorporating spatial autocorrelation into the modeling process. The SLM is applicable if the dependent variable is affected by the values of the dependent variable in nearby locations. In contrast, the SEM is applicable when spatial autocorrelation occurs as a nuisance resulting from misspecification of model. The SEM is considered a linear regression model with a spatial autoregressive error term. Our results indicated that, for the three forest plots with different spatial patterns of tree locations, the SEM was more suitable for dealing with the spatial autocorrelation in the tree height-diameter relationship according to the LM tests (Table 3) and Moran's I in model residuals (Figure 3). The dependence in tree height between neighboring trees was due to the attributional similarity among them rather than a direct effect from neighbors' size because tree height is less influenced by competition and the similarity in tree height is mainly due to microsite environment (Liu and Burkhart 1994, Bullock and Burkhart 2005).

In spatial regression models the spatial autocorrelation is captured by the weight matrix $\mathbf{W}$ in Equations 9 and 13. It is a row-sum standardized matrix and represents, in fact, the weighted averages of neighboring values of $\mathbf{y}$ in the SLM or the weighted averages of neighboring model errors in the SEM. Changing the definition of neighbors affects the number of neighboring observations used for computing the weighted averages. Even though the number of neighbors involved in computation varies for different types of spatial weight matrices, the differences between the weighted averages may be relatively small. The influence of the weighted averages on parameter estimation and model prediction is further reduced by the spatial autocorrelation parameter ρ in the SLM or λ in the SEM. In this study we found that there was almost no difference between three types of spatial weight (e.g., rook, distance, and k-nearest neighbors) for either the SLM or SEM in terms of spatial autocorrelation and heterogeneity in the model residuals.

Although the SEM and SLM are able to directly and effectively correct for the spatial autocorrelation in the spatially correlated data, they are insufficient to deal with the problem of spatial heterogeneity (Table 2; Figure 7) because spatial regression models generally assume that relationships between variables are spatially stationary. Their parameter estimates are global and are applied equally to all data points regardless of locations (Jetz et al. 2005, Bickford and Laffan 2006). Our results showed that the model residuals in both the SLM and SEM had degrees of spatially structured variability (SH%) and patterns of intrablock variance (local spatial variability) across different block sizes similar to those of the OLS model for the three plots (Table 2; Figure 7).

The LMM is also a global model, but it deals with spatial correlation in two ways: characterization and adjustment.

The characterization involves estimating spatial covariance parameters, e.g., nugget, partial sill, and range of the semivariogram. The adjustment involves obtaining more accurate predictions of the response variable through the EBLUP. Conceptually, EBLUP is universal kriging in geostatistics (Goodvaerts 1997, Schabenberger and Gotway 2005). The kriging process finds a set of weights for estimating the value of a response variable at a focal point from the values at a set of neighboring data points. The weight on each data point generally decreases with increasing distance to that focal point, in accordance with the decreasing data-to-estimation covariance $\hat{C}_p$ of Equation 8. However, the set of weights is also designed to account for the redundancy among the neighboring data points, represented in the data point-to-point covariance $\hat{V}$ of Equation 8. Note that the covariances ($\hat{C}_p$ and $\hat{V}$) and the kriging weight are determined entirely by the data spatial correlation and the covariance model [i.e., $f(d_{ij}, \boldsymbol{\theta})$ of Equation 7] for the semivariogram, not the actual data values (Welham et al. 2004, Schabenberger and Gotway 2005, Lark et al. 2006). In this study the exponential spatial covariance structure was used for the LMM. It was similar to the geographical weight function of the GWR_h model for computing the spatial weights of neighboring trees. Therefore, the LMM actually emphasizes "local" information that is determined by the semivariogram parameters (i.e., nugget, sill, and range); even the LMM model per se was fitted globally. We found that the LMM fitted data much better than the SLM and SEM and provided more desirable model residuals in terms of spatial autocorrelation and heterogeneity (Table 2; Figures 3 and 7).

GWR is a spatially local model, which uses a moving window over a spatially distributed set of observations and produces a set of model coefficients from subsamples of data around specific points in space (Páez and Scott 2004). Although GWR does not directly incorporate spatial autocorrelation into the modeling process because it assumes $N(0, \sigma^2 I)$ for the model error term and uses the weighted least-squares method for parameter estimation, it takes spatial locations explicitly into account and emphasizes local variation in the relationships between tree variables. Therefore, GWR not only is able to accommodate spatial heterogeneity but also significantly reduces spatial autocorrelation in model residuals as found in other studies (e.g., Zhang et al. 2005, Kupfer and Farris 2007). However, the model fitting and performance depend on the kernel function of spatial weight and size of bandwidth. In this study, two GWR models were fitted using either the fixed kernel function (i.e., GWR_h) or the adaptive kernel function (i.e., GWR_n). However, the two types of the kernel function selected different number of neighbors in the modeling process. The GWR_n model selected the neighbors close to the subject location and may cover a relatively small geographical area, especially at the locations where the density of trees is high. This nature of the adaptive spatial kernel function may provide the GWR_n model a higher ability to account for the local spatial heterogeneity and autocorrelation among the neighbors (Fotheringham et al. 2002, Bickford and Laffan 2006, Guo et al. 2008). Thus, the GWR_n model fitted the data better and produced more desirable model residuals in terms of spatial autocorrelation and heterogeneity than did the GWR_h model for the three plots (Table 2; Figures 3 and 7).

Conclusion

In this study, we used three forest plots to investigate the spatial distributions and heterogeneity in the model residuals from five regression models with OLS as the benchmark. The three plots represented different spatial patterns of tree locations, i.e., clustered, random, and regular patterns. These spatial patterns may be found in different forest types such as uneven-aged, irregular natural stands, even-age stands, or plantations.

When significant spatial autocorrelations and variations exist in the relationships between tree variables, as in the softwood plot (clustered) and hardwood plot (random), OLS is not appropriate for modeling the relationships between tree variables. Spatial regression models, especially the SEM, are very effective for directly accounting for spatial autocorrelation, but they are insufficient to deal with the problem of spatial heterogeneity because they are global in nature. This may also be true for other global models such as generalized additive models and neural network models (Zhang and Gove 2005, Zhang et al. 2005). In contrast, LMM and GWR incorporate the spatial dependence and variation into modeling processes, and consequently, fit the data better and predict the response variable more accurately. The model residuals from both the LMM and GWR have desirable spatial distributions, meaning fewer clusters of similar model errors or clusters of dissimilar model errors over space. One should apply both global and local models to spatially correlated data because they provide different perspectives and insights for data analysis and model prediction. GWR is a useful tool for exploring spatial nonstationarity at different spatial scales in the relationships between variables and assisting a modeler to identify correct specifications for a global model (Jetz et al. 2005, Foody 2005).

Literature Cited

Anselin, L. 1988. *Spatial econometrics: Methods and models.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 284 p.

Anselin, L. 1990a. What is special about spatial data? Alternative perspectives on spatial data analysis. P. 63–77 in *Spatial statistics: Past, present, and future,* Griffith, D. (ed.). Institute of Mathematical Geography, Ann Arbor, MI.

Anselin, L. 1990b. Spatial dependence and spatial structure instability in applied regression analysis. *J. Reg. Sci.* 30:185–207.

Anselin, L. 1992. Spatial dependence and spatial heterogeneity: Model specification issues in the spatial expansion paradigm. P. 334–354 in *Applications of the expansion method.* Jounes, J.P., and E. Casetti (eds.). Routledge, London, UK.

Anselin, L. 1993. Discrete space autoregressive models. P 454–469 in *Environmental modeling with GIS,* Goodchild, M. F., B.O. Parks, and L.T. Steyaert (eds.). Oxford University Press, New York, NY.

Anselin, L. 1995. Local indicator of spatial association—LISA. *Geog. Anal.* 27:93–115.

ANSELIN, L. 2001. Spatial econometrics. P. 310–330 in *A companion to theoretical econometrics*, Baltagi, B. (ed.). Blackwell, Oxford, UK.
ANSELIN, L. 2005. *Exploring spatial data with GeoDa^{TM}: A workbook*. Available online at www.csiss.org; last accessed Mar. 21, 2008.
ANSELIN, L., AND D.A. GRIFFITH. 1988. Do spatial effects really matter in regression analysis. *Papers Reg. Sci. Assoc.* 65:11–34.
ANSELIN, L., I. SYABRI, AND Y. KHO. 2006. GeoDa: An introduction to spatial data analysis. *Geogr. Anal.* 38:5–22.
BICKFORD, S.A., AND S.W. LAFFAN. 2006. Multi-extent analysis of the relationship between pteridophyte species richness and climate. *Global Ecol. Biogeogr.* 15:588–601.
BIONDI, F., D.E. MYERS, AND C.C. AVERY. 1994. Geostatistically modeling stem size and increment in an old-growth forest. *Can. J. For. Res.* 24:1354–1368.
BOOTS, B. 2002. Local measures of spatial association. *EcoScience* 9:168–176.
BORCARD, D., AND P. LEGENDRE. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol. Model.* 153:51–68.
BRUNSDON, C.A., A.S. FOTHERINGHAM, AND M.E. CHARLTON. 1998. Geographically weighted regression—Modeling spatial non-stationary. *Statistician* 47:431–443.
BULLOCK, B.P., AND H.E. BURKHART. 2005. An evaluation of spatial dependency in juvenile loblolly pine stands using stem diameter. *For. Sci.* 51:102–108.
CHARLTON, M., S. FOTHERINGHAM, AND C. BRUNSDON. 2003. *GWR 3. Software for geographically weighted regression.* Spatial Analysis Research Group, Department of Geography, University of Newcastle upon Tyne, UK.
CHILES, J., AND P. DELFINER. 1999. *Geostatistics: Modeling spatial uncertainty.* John Wiley & Sons, New York, NY. 695 p.
DALE, M.R.T., AND M. FORTIN. 2002. Spatial autocorrelation and statistical tests in ecology. *EcoScience* 9:162–167.
DALE, M.R.T., P. DIXON, M. FORTIN, P. LEGENDRE, D.E. MYERS, AND M.S. ROSENBERG. 2002. Conceptual and mathematical relationships among methods for spatial analysis. *Ecography* 25:558–577.
DE SMITH, M.J., M.F. GOODCHILD, AND P.A. LONGLEY. 2007. *Geospatial analysis: A comprehensive guide to principles, techniques and software tools.* 2nd ed. Matador, Leicester, UK.
DORMANN, C.F., J.M. MCPHERSON, M.B. ARAÚJO, R. BIVAND, J. BOLLIGER, G. CARL, R.G. DAVIES, A. HIRZEL, W. JETZ, W.D. KISSLING, I. KÜHN, R. OHLEMÜLLER, P.R. PERES-NETO, B. REINEKING, B. SCHRÖDER, F.M. SCHURR, AND R. WILSON. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography* 30:609–628.
DUNCAN, C. 1997. Applying mixed multivariate multilevel models in geographical research. P. 100–117 in *Place and people: Multilevel modeling in geographical research*, Wstert, P., and R.N. Verhoeff (eds.). Nederlandse Geografische Studies 227, Univ. of Utrecht, Utrecht, The Netherlands.
DUTILLEUL, P., AND P. LEGENDRE. 1993. Spatial heterogeneity against heteroscedasticity: An ecological paradigm versus a statistical concept. *OIKOS* 66:152–171.
EK, A.R. 1969. *Stem map data for three forest stands in northern Ontario.* Information Report O-X-113. For. Res. Lab., Sault Ste. Marie, ON, Canada. 23 p.
FARBER, S., AND A. PÁEZ. 2007. A systematic investigation of cross-validation in GWR model estimation: Empirical analysis and Monte Carlo simulations. *J. Geogr. Sys.* 9:371–396.
FOODY, G.M. 2003. Geographical weighting as a further refinement to regression modeling: An example focused on the NDVI-rainfall relationship. *Remote Sens. Environ.* 88:283–293.
FOODY, G.M. 2004. Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. *Global Ecol. Biogeogr.* 13:315–320.
FOODY, G.M. 2005. Clarification on local and global analysis. *Global Ecol. Biogeogr.* 14:99–100.
FOTHERINGHAM, A.S., AND C.A. BRUNSDON. 1999. Local forms of spatial analysis. *Geog. Anal.* 31:340–358.
FOTHERINGHAM, A.S., C.A. BRUNSDON, AND M.E. CHARLTON. 2002. *Geographically weighted regression: The analysis of spatially varying relationships.* John Wiley & Sons, New York, NY.
FOTHERINGHAM, A.S., M.E. CHARLTON, AND C.A. BRUNSDON. 2009. *GWR 3.0.* Available online at ncg.nuim.ie/ncg/GWR/; last accessed Mar. 21, 2008.
FOX, J.C., P.K. ADES, AND H. BI. 2001. Stochastic structure and individual-tree growth models. *For. Ecol. Manag.* 154:261–276.
FOX, J.C., H. BI, AND P.K. ADES. 2007a. Spatial dependence and individual-tree growth models: I. Characterizing spatial dependence. *For. Ecol. Manag.* 245:10–19.
FOX, J.C., H. BI, AND P.K. ADES. 2007b. Spatial dependence and individual-tree growth models: II. Modeling spatial dependence. *For. Ecol. Manag.* 245:20–30.
GARRIGUES, S., D. ALLARD, F. BARET, AND M. WEISS. 2006. Quantifying spatial heterogeneity at the landscape scale using variogram models. *Remote Sens. Environ.* 103:81–96.
GEODA CENTER FOR GEOSPACIAL ANALYSIS AND COMPUTING. 2009. *GeoDa*. Available online at www.geoda.uiuc.edu/; last accessed Mar. 21, 2008.
GOODVAERTS, P. 1997. *Geostatistics for natural resources evaluation.* Oxford University Press, New York, NY. 483 p.
GORR, W.L., AND A.M. OLLIGSCHLAEGER. 1994. Weighted spatial adaptive filtering: Monte Carlo studies and application to Illicit Drub Market Modeling. *Geog. Anal.* 26(1):67–87.
GRAAFF, T. DE, R.J.G.M. FLORAX, P. NIJKAMP, AND A. REGGIANI. 2001. A general misspecification test for spatial regression models: Dependence, heterogeneity, and nonlinearity. *J. Region. Sci.* 41:255–276.
GUO, L., Z. MA, AND L. ZHANG. 2008. Comparison of bandwidth selection in application of geographically weighted regression: A case study. *Can. J. For. Res.* 38:2526–2534.
ISAAKS, E.H., AND R.M. SRIVASTAVA. 1989. *An introduction to applied geostatistics.* Oxford University Press, New York, NY.
JETZ, W., C. RAHBEK, AND J.W. LICHSTEIN. 2005. Local and global approaches to spatial data analysis in ecology. *Global Ecol. Biogeogr.* 14:97–98.
JONES, J. 1997. Multilevel approaches to modeling contextuality: From nuisance to substance in the analysis of voting behavior. P. 19–40 in *Place and people: Multilevel modeling in geographical research,* Wstert, P., and R.N. Verhoeff (eds.). Nederlandse Geografische Studies 227, Univ. of Utrecht, Utrecht, The Netherlands.
KOHL, M., AND G. GERTNER. 1997. Geostatistics in evaluating forest damage surveys: Considerations on methods for describing spatial distributions. *For. Ecol. Manag.* 95:131–140.
KUPFER, J.A., AND C.A. FARRIS. 2007. Incorporating spatial nonstationarity of regression coefficients into predictive vegetation models. *Landscape Ecol.* 22:837–852.
LARK, R.M., B.R. CULLIS, AND S.J. WELHAM. 2006. On spatial prediction of soil properties in the presence of a spatial trend: The empirical best linear unbiased predictor (EBLUP) with REML. *Eur. J. Soil Sci.* 57:787–799.
LEE, J., AND D.W.S. WONG. 2001. *Statistical analysis with ArcView GIS.* John Wiley & Sons, New York, NY, 192 p.

LEGENDRE, P. 1993. Spatial autocorrelation: Trouble or new paradigm? *Ecology* 74:1659–1673.

LI, H., AND J.F. REYNOLDS. 1995. On definition and quantification of heterogeneity. *OIKOS* 73:280–284.

LICHSTEIN, J.W., T.R. SIMONS, S.A. SHRINER, AND K.E. FRANZREB. 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecol. Monogr.* 72:445–463.

LITTELL, R.C., G.A. MILLIKEN, W.W. STROUP, R.D. WOLFINGER, AND O. SCHABENBERGER. 2006. *SAS for mixed models.* 2nd ed. SAS Institute, Inc., Cary, NC. 814 p.

LIU, J., AND H.E. BURKHART. 1994. Spatial characteristics of diameter and total height in juvenile loblolly pine (*Pinus taeda* L.) plantations. *For. Sci.* 40:774–786.

MAGNUSSEN, S. 1994. A method to adjust simultaneously for spatial microsite and competition effects. *Can. J. For. Res.* 24: 985–995.

MCCULLOCH, C.E., AND S.R. SEARLE. 2001. *Generalized, linear, and mixed models.* John Wiley & Sons, New York, NY.

MYERS, D.E. 1997. Statistical models for multiple-scaled analysis. P. 273–293 in *Scale in remote sensing and GIS,* Quattrochi, D.A., and M.F. Goodchild (eds.). Lewis Publishers, Boca Raton, FL. 406 p.

OVERMARS, K.P., G.H.J. DE KONING, AND A. VELDKAMP. 2003. Spatial autocorrelation in multi-scale land use models. *Ecol. Model.* 164:257–270.

PÁEZ, A., AND D.M. SCOTT. 2004. Spatial statistics for urban analysis: A review of techniques with examples. *GeoJournal* 61:53–67.

PERRY, J.N., A. LIEBHOLD, M.S. ROSENBERG, J. DUNGAN, M. MIRITI, A. JAKOMULSKA, AND S. CITRON-POUSTY. 2002. Illustration and guidelines for selecting statistical methods for quantifying spatial patterns in ecological data. *Ecography* 25: 578–600.

PIELOU, E.C. 1959. The use of point-to-plant distance in the study of the pattern of plant population. *J. Ecol.* 47:607–613.

PINHEIRO, J.C. AND D.M. BATES. 2000. *Mixed-effects models in S and S-PLUS.* Springer, New York, NY.

REED, D.D., AND H.E. BURKHART. 1985. Spatial autocorrelation of individual tree characteristics in loblolly pine stands. *For. Sci.* 31:575–587.

SAS INSTITUTE, INC. 2008. *SAS/STAT users' guide.* Version 9.1.3. SAS Institute, Inc. Cary, NC.

SCHABENBERGER, O., AND C.A. GOTWAY. 2005. *Statistical methods for spatial data analysis.* Chapman and Hall/CRC, Boca Raton, FL. 488 p.

SCHOOLEY, R.L. 2006. Spatial heterogeneity and characteristic scales of species-habitat relationships. *BioScience* 56:533–537.

SHI, H., AND L. ZHANG. 2003. Local analysis of tree competition and growth. *For. Sci.* 49(6):938–955.

SHIYOMI, M., S. TAKAHASHI, AND J. YOSHIMURA. 2000. A measure for spatial heterogeneity of a grassland vegetation based on the beta-binomial distribution. *J. Veg. Sci.* 11:627–632.

VELDTMAN, R. 2005. *Dimensions of spatial heterogeneity: A classification of non-, semi- and explicit spatial heterogeneity.* Available at upetd.up.ac.za/thesis/available/etd-02092006-083709/unrestricted/03chapter3.pdf; last accessed Mar. 21, 2008.

WACKERNAGEL, H. 2003. *Multivariate geostatistics: An introduction with applications.* 3rd ed. Springer, New York, NY. 387 p.

WAGNER, H.H., AND M. FORTIN. 2005. Spatial analysis of landscapes: Concepts and statistics. *Ecology* 86:1975–1987.

WELHAM, S.J., B.R. CULLIS, B. GOGEL, A. GILMOUR, AND R. THOMPSON. 2004. Prediction in linear mixed models. *Aust. NZ J. Statist.* 46:325–347.

WIENS, J.A. 2000. Ecological heterogeneity: Ontogeny of concepts and approaches. P. 9–31 in *Ecological consequences of environmental heterogeneity,* Hutchings, M.J., E.A. John, and A.J.A. Stewart (eds.). Blackwell Science, Oxford, UK.

WULDER, M.A., J.C. WHITE, N.C. COOPS, T. NELSON, AND B. BOOTS. 2007. Using local spatial autocorrelation to compare outputs from a forest growth model. *Ecol. Model.* 209:264–276.

ZHANG, L., H. BI, P. CHENG, AND C.J. DAVIS. 2004. Modeling spatial variations in tree diameter-height relationships. *For. Ecol. Manag.* 189:317–329.

ZHANG, L., AND J.H. GOVE. 2005. Spatial assessment of model errors from four regression techniques. *For. Sci.* 51:334–346.

ZHANG, L., J.H. GOVE, AND L. HEATH. 2005. Spatial residual analysis of six modeling techniques. *Ecol. Model.* 186:154–177.

ZHANG, L., AND H. SHI. 2004. Local modeling of tree growth by geographically weighted regression. *For. Sci.* 50:225–244.