Available online at www.sciencedirect.com
SCIENCE DIRECT®
Ecological Modelling 186 (2005) 154–177
ECOLOGICAL MODELLING
www.elsevier.com/locate/ecolmodel

Spatial residual analysis of six modeling techniques

Lianjun Zhang[a,*], Jeffrey H. Gove[b,1], Linda S. Heath[b,2]

[a] *Faculty of Forest and Natural Resources Management, State University of New York, College of Environmental Science and Forestry, One Forestry Drive, Syracuse, NY 13210, USA*
[b] *USDA Forest Service Northeastern Research Station, P.O. Box 640, Durham, NH 03824, USA*

Received 2 August 2004; received in revised form 29 December 2004; accepted 3 January 2005
Available online 6 June 2005

Abstract

In recent years alternative modeling techniques have been used to account for spatial autocorrelations among data observations. They include linear mixed model (LMM), generalized additive model (GAM), multi-layer perceptron (MLP) neural network, radial basis function (RBF) neural network, and geographically weighted regression (GWR). Previous studies show these models are robust to the violation of model assumptions and flexible to nonlinear relationships among variables. However, many of them are non-spatial in nature. In this study, we utilize a local spatial analysis method (i.e., local Moran coefficient) to investigate spatial distribution and heterogeneity in model residuals from those modeling techniques with ordinary least-squares (OLS) as the benchmark. The regression model used in this study has tree crown area as the response variable, and tree diameter and the coordinates of tree locations as the predictor variables. The results indicate that LMM, GAM, MLP and RBF may improve model fitting to the data and provide better predictions for the response variable, but they generate spatial patterns for model residuals similar to OLS. The OLS, LMM, GAM, MLP and RBF models yield more residual clusters of similar values, indicating that trees in some sub-areas are either all underestimated or all overestimated for the response variable. In contrast, GWR estimates model coefficients at each location in the study area, and produces more accurate predictions for the response variable. Furthermore, the residuals of the GWR model have more desirable spatial distributions than the ones derived from the OLS, LMM, GAM, MLP and RBF models.

Keywords: Spatial autocorrelation; Local indicator of spatial autocorrelation (LISA); Ordinary least squares (OLS); Linear mixed model (LMM); Generalized additive model (GAM); Multi-layer perceptron (MLP) neural network; Radial basis function (RBF) neural network; Geographically weighted regression (GWR)

1. Introduction

Ordinary least-squares (OLS) has been a widely applied technique to estimate model parameters in forest and ecological modeling practices. However, one of

* Corresponding author. Tel.: +1 315 470 6558; fax: +1 315 470 6535.
E-mail addresses: lizhang@esf.edu (L. Zhang), jgove@fs.fed.us (J.H. Gove), Lheath@fs.fed.us (L.S. Heath).
[1] Tel.: +1 603 868 7667.
[2] Tel.: +1 603 868 7612.

0304-3800/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2005.01.007

the OLS assumptions, independence of observations, is often violated due to temporal or/and spatial autocorrelations in data, which leads to a biased estimation of the standard errors of model parameters and, consequently, misleading significance tests (Anselin and Griffith, 1988; Fox et al., 2001). Temporal autocorrelation has received early attention since the 1960s. A number of statistical methods have been applied to overcome the problem (e.g., Gregoire et al., 1995). In contrast, the violation of the OLS assumption due to spatial dependency has not drawn much attention in forest and ecological studies until recent years. Modern modeling techniques have increasingly become popular to deal with spatial autocorrelation and heterogeneity for predicting forest composition and attributes, species distributions, biodiversity, forest type and class, insect attack, etc. These techniques include generalized linear model (GLM), linear mixed model (LMM), generalized additive model (GAM), classification and regression tree (CART), multivariate adaptive regression splines (MARS), artificial neural networks (ANN), and geographically weighted regression (GWR) (e.g., Austin and Meyers, 1996; Preisler et al., 1997; Lehmann, 1998; Moisen and Edwards, 1999; Moisen et al., 1999; Guisan and Zimmermann, 2000; Frescino et al., 2001; Gullison and Bourque, 2001; Tappeiner et al., 2001; Austin, 2002; Zaniewski et al., 2002; Guisan et al., 2002; Moisen and Frescino, 2002; Lehmann et al., 2003; Foody, 2003; Zhang et al., 2004; Zhang and Shi, 2004).

These modern modeling techniques have desirable features: they are robust when applied to correlated data, have less restrictions in assumptions, and are flexible in modeling non-linearity and non-constant variance structures (Guisan et al., 2002; Moisen and Frescino, 2002). However, many of them are non-spatial in nature (Laffan, 1999). Although the improvement of model fitting and estimating parameter standard errors is evident in the previous studies, little attention has been paid to the spatial heterogeneity of model performance. Many studies assessed the models in terms of overall model performance, accuracy, and errors (e.g., Moisen and Frescino, 2002; Robertson et al., 2003). Since global assessment summarizes the model errors from many locations, information is not available on where some parts of the study area are better predicted than others. A few studies used maps to locally evaluate the model residuals (Rathert et al., 1999; Anderson et al., 2003). Visualization, however, does not enable us to identify significant clusters of positive or negative model residuals at multiple scales. Others suggested the use of local indicator of spatial association (LISA) to assess the spatial heterogeneity of model residuals (Laffan, 1999; Tiefelsdorf, 2000; Overmars et al., 2003). When data are collected across a large spatial region, it is anticipated that there will be one or more sub-areas where the sizes and shapes of variables or relationships between variables are different from "normal" situations. Unless prior information is available, one may not be able to detect the number, locations, sizes, and shapes of such anomalies by global assessment. LISA is designed to reveal such peculiarities (Boots, 2002; Dale and Fortin, 2002), and has been used to successfully identify clusters in biological datasets, and to identify "hot spots" (positive autocorrelation, or similarity) and "cold spots" (negative autocorrelation, or dissimilarity) of tree growth and competition (Sokal et al., 1998a,b; Shi and Zhang, 2003).

The objectives of this study were (1) to apply six modeling techniques (i.e., OLS, LMM, GAM, MLP, RBF, and GWR) to model the relationships between tree crown area and diameter, (2) to evaluate the performance of above six models using overall model residuals, and residuals across tree size classes, and (3) to spatially assess the performance of the six models in terms of spatial distributions and clustering of positive/negative model errors using the maps of model residuals and local Moran coefficients.

2. Theoretical background

We briefly review the theory for the six modeling techniques used in this study.

2.1. Ordinary least squares (OLS)

Suppose we have a set of n observations ($k = 1, 2, \ldots, n$) on p ($g = 1, 2, \ldots, p$) independent or predictor variables X, and a dependent or response variable Y. The relationship between Y and X can be regressed using OLS as follows:

$$Y = X\beta + \varepsilon \quad (1)$$

where Y is a vector of the observed response variable, X is a known model matrix including a column of 1 (for intercept) and p independent variables, β is a vector of unknown fixed-effects parameters, and ε is a random error term whose distribution is $N(0, \sigma^2 I)$, with I denoting an identity matrix. The OLS estimate of β is obtained by the least-squares method as

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

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

2.2. Linear mixed model (LMM)

A linear mixed model can be expressed as

$$Y = X\beta + Z\gamma + \varepsilon \tag{3}$$

where Y, X, and β are as defined in Eq. (1), Z is a known design matrix, γ is a vector of unknown random-effects parameters, and ε is a vector of unobserved random errors. It is assumed: (1) $E(\gamma) = 0$ and $\mathrm{Var}(\gamma) = G$, (2), $E(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = R$, (3) $\mathrm{Cov}(\gamma, \varepsilon) = 0$, and (4) both γ and ε are normally distributed. The variance of Y is $V = ZGZ^{\mathrm{T}} + R$, and can be estimated by setting up the random-effects design matrix Z and by specifying covariance structures for G and R (Littell et al., 1996). However, OLS is no longer considered as the best approach to estimating LMM. Likelihood-based methods (e.g., maximum likelihood (ML) and restricted/residual maximum likelihood (REML)) are usually used to solve for β and γ, under the assumption that γ and ε are normally distributed. Numerical algorithms such as Newton–Raphson algorithm can be used to obtain

$$\hat{\beta} = (X^{\mathrm{T}}\hat{V}^{-1}X)^{-1}X^{\mathrm{T}}\hat{V}^{-1}Y. \tag{4}$$

LMM can be used to model spatial correlation among observations in data through $R = \mathrm{Var}(\varepsilon)$ such that

$$R = \mathrm{Cov}(\varepsilon_i, \varepsilon_j) = \sigma^2 f(d_{ij}) \tag{5}$$

where d_{ij} is the distance between locations i and j. Different functions $f(d_{ij})$ are available including spherical, exponential, Gaussian, power, etc. The likelihood ratio test can be used to test whether it is necessary to model the covariance structure of the data (Littell et al., 1996). Furthermore, the empirical best linear unbiased predictions (EBLUP) should be used to take spatial autocorrelations into account for predicting the response variable (Schabenberger and Pierce, 2002, p. 683).

2.3. Generalized additive model (GAM)

GAM is a nonparametric extension of GLM (Hastie and Tibshirani, 1990; Guisan et al., 2002). While GLM emphasizes estimation and inference for the parameters of the model, GAM focuses on exploring data nonparametrically. The strength of GAM is its ability to deal with highly non-linear and non-monotonic relationships between the response variable and the set of explanatory variables. Thus, GAM is sometimes referred to as data-driven rather than model-driven (Guisan et al., 2002). In general, GAM can be expressed as:

$$Y = S_0 + \sum_{g=1}^{p} S_g(X_g) + \varepsilon \tag{6}$$

where S_0 is the intercept, and $S_g(X_g)$ is a nonparametric smoothing function for the gth independent variable X. The only underlying assumption is that the smoothing functions in GAM are additive. This additive restriction allows us to interpret a GAM model in a similar way as a traditional linear regression model. GAM can be commonly fitted by numerical algorithms such as so called backfitting, in which the nonparametric smoothing functions $S_g(X_g)$ are sought for minimizing

$$\varepsilon^2 = \left[Y - \left(S_0 + \sum_{g=1}^{p} S_g(X_g) \right) \right]^2. \tag{7}$$

A wide range of nonparametric functions is available for GAM. The combinations of these functions are also possible. Several smoothers are commonly used: (1) cubic spline smoothing, (2) B-spline smoothing, (3) polynomial cubic spline smoothing, (4) locally weighted scatter plot smoothing (LOESS), (5) multivariate LOESS, and (6) bivariate thin-plate spline smoothing (Venables and Ripley, 1997; SAS Institute, Inc. 2002).

2.4. Multi-layer perceptron (MLP) neural network

MLP is the most popular neural network model, which commonly consists of three successive layers: input layer (X), hidden layer (H), and output layer

(Y). The three layers consist of simple computational units called nodes. The nodes from one layer are connected or linked to all nodes in the adjacent layer(s). But there are no lateral connections within any layer, nor are feedback connections possible in a MLP network. Therefore, MLP is also known as a multi-layer feed-forward neural network. Each connection has a set of weights (i.e., coefficients) that represents the strength of the connection. In the input layer, the number of nodes corresponds to the number of input features (i.e., predictor variables, X). The number of nodes in the output layer corresponds to the number of target variables (i.e., response variable, Y). The layer between the input and output is called the hidden layer (H), which has no direct connections outside the neural network. In practice, there is very rarely an advantage in using more than one hidden layer (Lippmann, 1987; Rumelhart et al., 1987). The input features and output variables are known observations in the data. When they are presented to the input nodes and output nodes, respectively, so called activation functions are used to link the net input to the hidden nodes, as well as the hidden nodes to the output nodes. To accomplish the optimization of the weights in the activation functions, an error back-propagation algorithm is commonly used to minimize the objective function defined as

$$E = \frac{1}{2}\sum_{g=1}^{p}(Y - \hat{Y})^2. \tag{8}$$

This is the sum of squared difference between the predicted output ($\hat{Y}$) and the observed output (Y) averaged over all input and output observations (Warner and Misra, 1996). By updating the weights (i.e., coefficients) of the activation functions, the neural network is said to be learning. The connection weights are adjusted using the error back-propagation algorithm based on the generalized delta rule (Rumelhart et al., 1987) such that the weights between two layers are computed iteratively as:

$$\Delta\theta(t+1) = \eta\varepsilon_1 + \alpha\,\Delta\theta(t) \tag{9}$$

$$\Delta\delta(t+1) = \eta\varepsilon_2 + \alpha\,\Delta\delta(t) \tag{10}$$

where $\Delta\theta(t+1)$ is the change of the weights between the input layer and the hidden layer at the $(t+1)$th iteration, $\Delta\theta(t)$ is the change of the weights between the input layer and the hidden layer at the tth iteration, ε_1 is the model error term for the hidden layer, $\Delta\delta(t+1)$ is the change of the weights between the hidden layer and the output layer at the $(t+1)$th iteration, $\Delta\theta(t)$ is the change of the weights between the hidden layer and the output layer at the tth iteration, ε_2 is the model error term for the output layer, η is the learning rate (analogous to the step-size in a gradient-descent-based optimization), and α is the momentum parameter. The above procedure is repeated for all the training samples until the network errors are less than a predefined threshold or stabilized. Training and learning are fundamental to many neural networks (analogous to parameter estimation procedures in statistics) (Rumelhart et al., 1987; Warner and Misra, 1996). The activation functions used in MLP include logistic, Gaussian, linear, hyperbolic tangent, and threshold. A commonly used activation function is the logistic function that introduces nonlinearity into the network (Lek and Guegan, 1999).

2.5. Radial basis function (RBF) neural network

RBF is also a general-purpose neural network model. It is a supervised, feed-forward neural network with one hidden layer of artificial neurons (Shin and Goel, 2000). The hidden nodes in the RBF networks contain the "Radial Basis Function", a statistical transformation based on a Gaussian kernel (Fischer, 1997). In a RBF network, the net input to the hidden layer is not a linear combination of the inputs as specified by the weights. Instead the entire input vector is passed directly without changes to the hidden nodes. The hidden nodes have a "basis function" which has the parameters "center" and "width". The center of the basis function is a vector of weights (β) of the same size. There is normally a different center for each hidden node. In the computation of net input, d is measured in terms of the Euclidean distance (radial distance) between the input vector (X) and the weight vectors (β):

$$d = \left[\sum_{i=1}^{n_x}\frac{(\beta - X)^2}{2h}\right]^{1/2} \tag{11}$$

where h is the bandwidth associated with the hidden nodes called the smoothing parameter. The fraction of

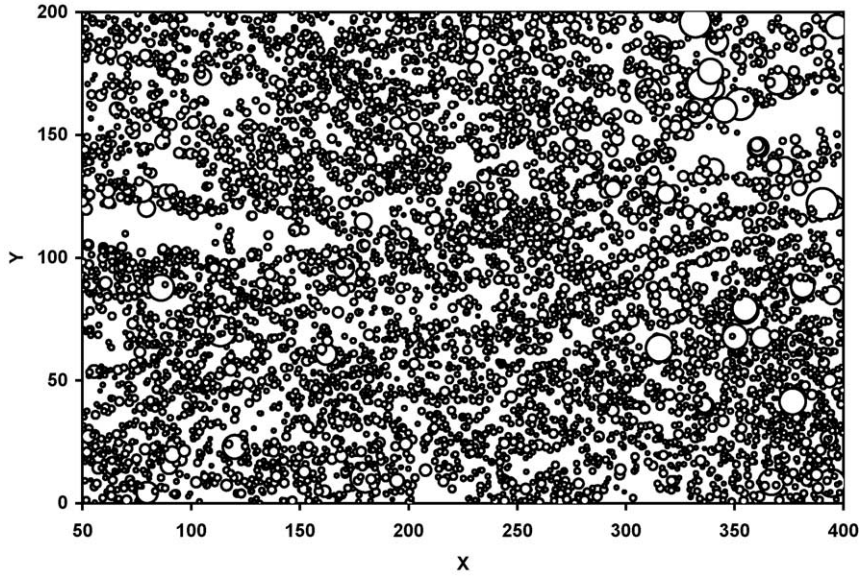


Fig. 1. Location map of tree diameters. The size of the circle is proportional to tree DBH.

overlap between each hidden node and its neighboring nodes is decided by the bandwidth. The output from the hidden nodes is computed by an activation function such as Gaussian, thin-plate spline, linear, multiquadratic or inverse multiquadratic (Shin and Goel, 2000). The output layer computes the predicted output ($\hat{Y}$) as a linear combination of the hidden output values. Since the basis function responds only to a small region of the input space where it is centered, the hidden nodes in RBF are also called localized receptive

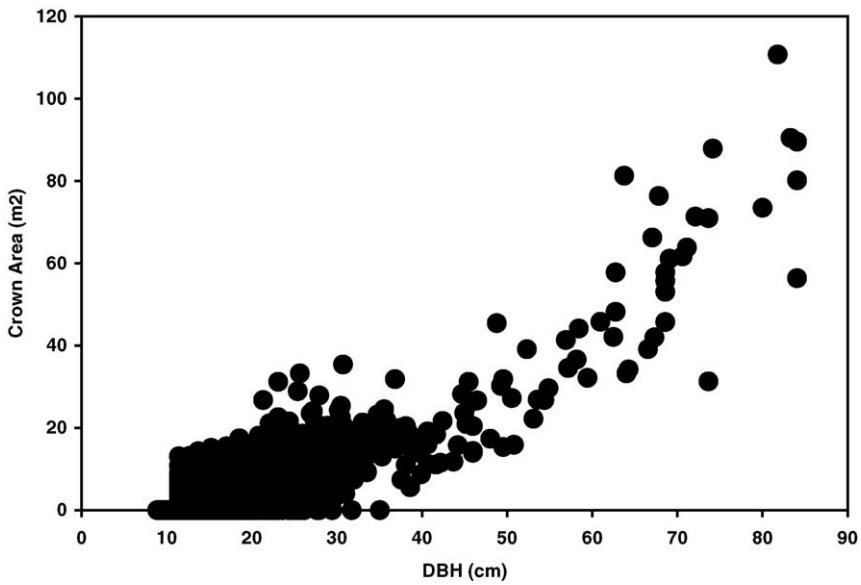


Fig. 2. Scatterplot of tree crown area vs. diameter.

field or locally tuned processing units, which are similar to the "kernel" in kernel regression (Sarle, 1994). Therefore, the RBF networks are localized, while the MLP networks are global (Murnion, 1999). RBF is commonly trained in a two-stage procedure: (1) the input data is used to determine the centers of the basis function by a *k*-means-based clustering algorithm, and (2) the weights for the outputs are found by minimizing the objective function similar to Eq. (8) with an error back-propagation algorithm similar to Eqs. (9) and (10) (Fischer, 1997; Shin and Goel, 2000; Mitra and Basak, 2001).

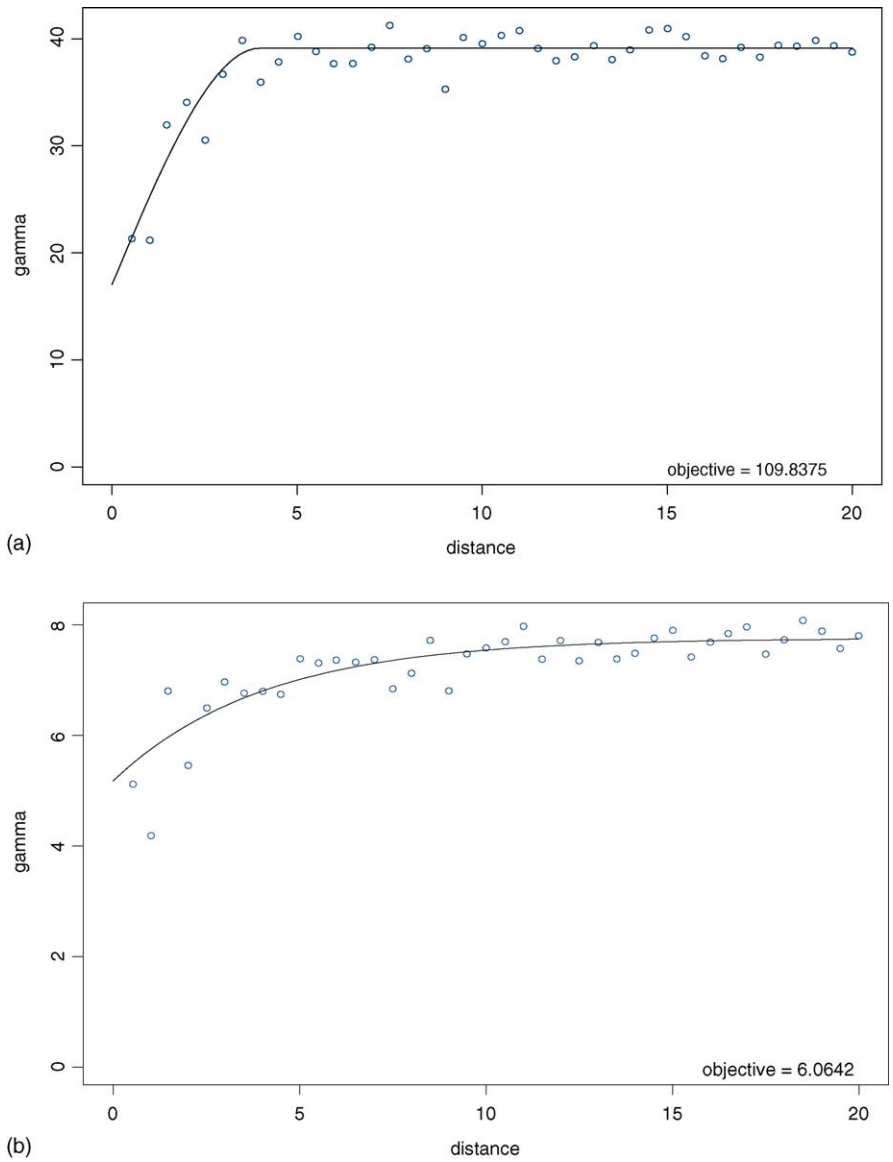


Fig. 3. (a) Variogram of tree DBH (lag = 0.5 m, range = 4.0 m). (b) Variogram of OLS model residual (lag = 0.5 m, range = 4.0 m). (c) Variogram of LMM model residual (lag = 0.5 m, range = 4.4 m). (d) Variogram of GAM model residual (lag = 0.5 m, range = 4.9 m). (e) Variogram of MLP model residual (lag = 0.5 m, range = 4.5 m). (f) Variogram of RBF model residual (lag = 0.5 m, range = 5.4 m). (g) Variogram of GWR model residual (lag = 0.5 m, range = ? m).

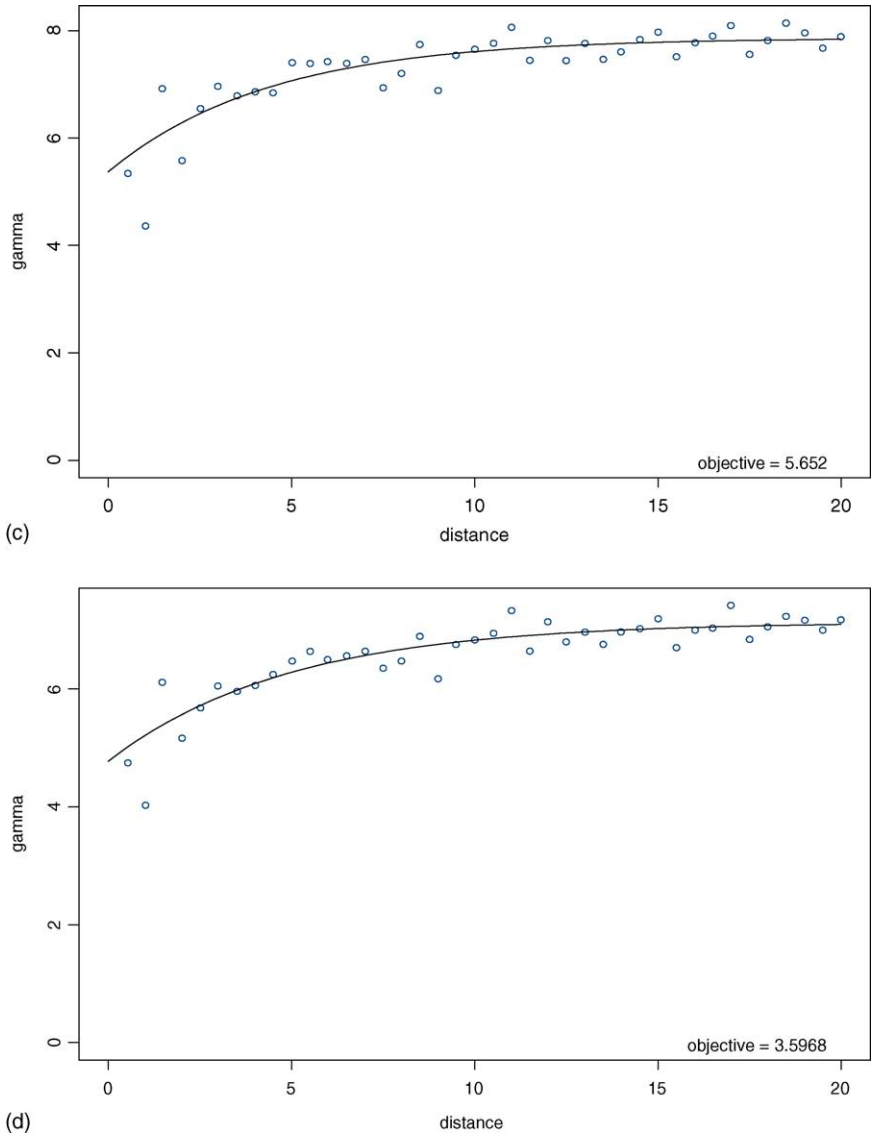


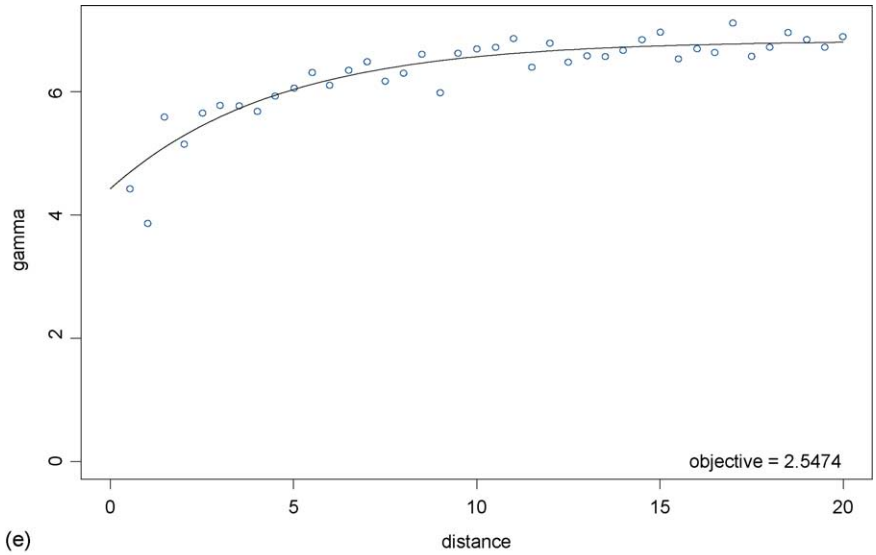
Fig. 3. (*Continued*)

2.6. Geographically weighted regression (GWR)

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

$$Y = \beta_0(u_i, v_i) + \sum_{g=1}^{p} \beta_g(u_i, v_i) X_g + \varepsilon \quad (12)$$

where $\{\beta_0(u_i, v_i), \beta_1(u_i, v_i), \ldots, \beta_p(u_i, v_i)\}$ are $(p+1)$ continuous functions of the location (u_i, v_i) in the study area. Again, ε is the random error term with a distribution $N(0, \sigma^2 I)$. The aim of GWR is to obtain the estimates of these functions for each independent variable X and each geographic location i. This can be achieved by using data near the location i. The estimation procedure of GWR is as follows: (1) draw a circle of a given radius around one particular location i



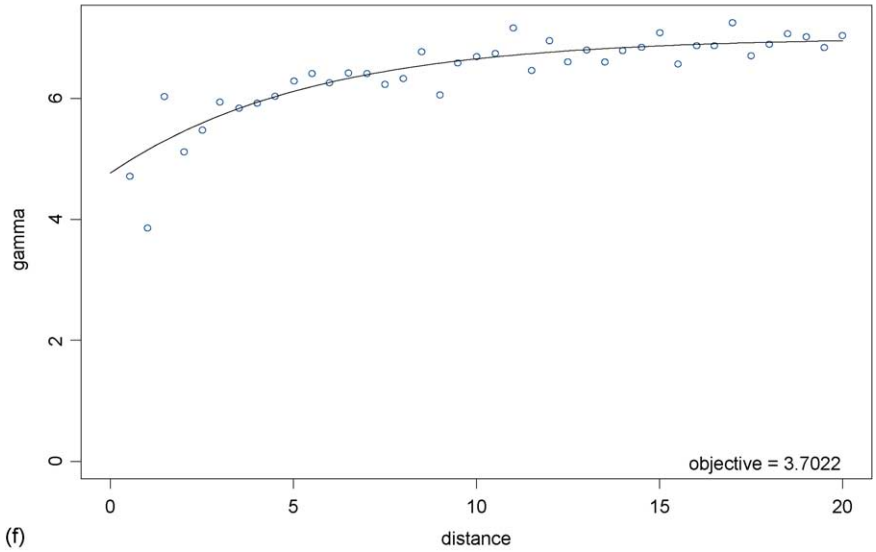


Fig. 3. (*Continued*)

(the center), (2) compute a weight (w_{ij}) for each neighboring observation j according to the distance (d_{ij}) between the location j and the center i, and (3) estimate the model coefficients using weighted least-squares regression such that

$$\hat{\beta}_i = (X^{\mathrm{T}} W_i X)^{-1} X^{\mathrm{T}} W_i y \tag{13}$$

where the weight matrix W_i is

$$W_i = \begin{pmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{\mathrm{in}} \end{pmatrix}. \tag{14}$$

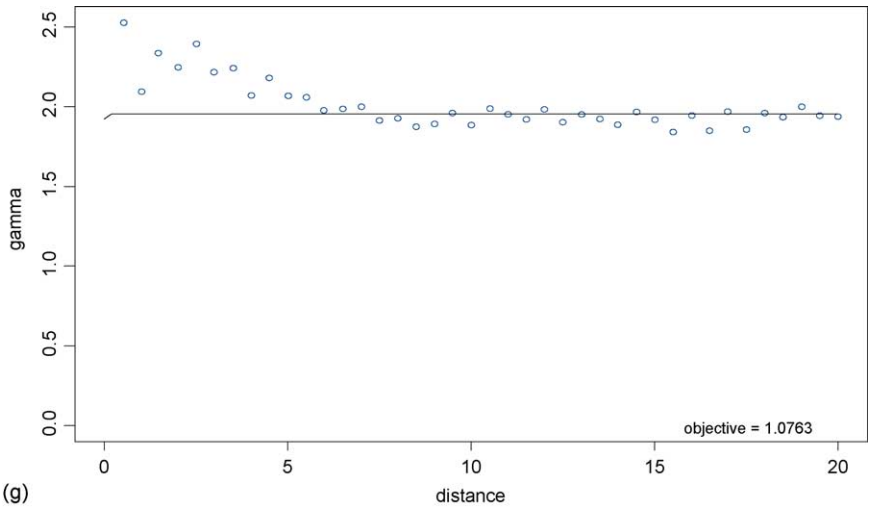


Fig. 3. (*Continued*).

If $W_i = I$ (identity matrix), that is, if each observation in the data has a weight of unity, the GWR model is equivalent to the OLS model. A Gaussian distance-decay-based kernel function is commonly used to compute the weight matrix as follows:

$$w_{ij} = \mathrm{e}^{-(d_{ij}/h)^2} \tag{15}$$

where h is referred to as the bandwidth. This kernel function assumes that the bandwidth at each center i is a constant across the study area (i.e., a fixed kernel). If the locations i and j coincide (i.e., $d_{ij} = 0$), then w_{ij} equals one; otherwise as the distance d_{ij} increases the w_{ij} decreases according to a Gaussian curve. However, the weights are nonzero for all data points, no matter how far they are from the center i (Fotheringham et al., 2002). Note that Eq. (13) is not a single equation but an array of equations, with each β_i corresponding to a row of the matrix whose elements are β_{ij}. It is also possible to compute the standard error for each coefficient estimate. Once each of the w_{ij} has been calculated, the coefficient matrix β can be computed row by row by repeated application of Eq. (13). Therefore, one obtains a set of estimates of spatially varying parameters without specifying a function form for the spatial variation (Brunsdon et al., 1998; Fotheringham et al., 2002).

3. Data and methods

3.1. Data

The data used in this study were the stem map data of a softwood stand located near Sault Ste. Marie, Ontario, Canada (Ek, 1969). It was a mature, second growth, and uneven-aged stand with 6811 trees. Major tree species included balsam fir (*Abice balsamea* (L.) Mill.) (53% in number of trees), black spruce (*Picea mariana* (Mill.) BSP.) (40%), and white spruce (*Picea glauca* (Moench) Voss) (3.7%). Minor species were Tamarack (*Larix laricina* (Du Roi) K. Koch), white pine (*Pinus strobus* L.), balsam poplar (*Populus balsamifera* L.), white birch (*Betula papyrifera* Marsh.), etc. Tree location coordinates, diameters at breast height (DBH), heights, and crown area (CROWN) for trees > 8.9 cm (3.5 in.) were available in the data. Due to insufficient computer memory for handling all trees for estimating the spatial variance matrix in LMM, the whole data set was reduced to 5979 trees by deleting trees between 0 and 5 m on the X-axis. The location map of the tree DBH is shown in Fig. 1, and the descriptive statistics of tree DBH and CROWN are listed in Table 1.

3.2. Regression model

We chose a linear regression model to fit the relationship between tree CROWN and DBH (Fig. 2), in which

Table 1
Descriptive statistics of tree variables for the example plot ($n = 5797$ trees)

Variable	Mean	Std	Minimum	Maximum
Diameter (cm)	15.89	6.77	8.89	84.07
Crown area (m^2)	4.76	5.94	0.00	110.74

CROWN was the response or dependent variable, and the predictor or independent variables included the coordinates (u_i, v_i) of tree locations, tree DBH and DBH^2. We would like to point out that we did not intend to develop a predictive model for tree crown area. Rather, we attempt to investigate the spatial heterogeneity of the model residuals for fitting a simple relationship between tree crown area and DBH by the six modeling techniques. Therefore, no other tree attributes or stand variables were considered as the predictor variables in the model.

3.3. Model fitting and evaluation

In this study, OLS and LMM were fitted using Statistical Analysis System (SAS) 9.0 (SAS Institute, Inc. 2002). GAM was fitted using S-Plus 6.2 (Insightful, Inc. 2003) due to extremely long computing time required by SAS. MLP and RBF were fitted using a public domain package, LNKnet, which is developed at MIT Lincoln Laboratory (Kukoloch and Lippmann, 1999, the software and the manual can be downloaded at http://www.ll.mit.edu/IST/). The GWR model was fitted using a computer software program, GWR 2.0. Detailed information on the GWR software is available at the web site http://www.ncl.ac.uk/geography/GWR (Fotheringham et al., 2002).

For the LMM model, different spatial covariance structures were tried to account for the spatial autocorrelations among trees, including Gaussian, exponential, and spherical functions (Littell et al., 1996). The exponential covariance structure was selected according to the model fitting statistics such as Akaike's Information Criterion (AIC). The EBLUP predictions from the LMM model were used to compute the model residuals (Schabenberger and Pierce, 2002, p. 683). The GAM model was implemented with a multivariate LOESS function for the tree location coordinates (u_i, v_i), and a LOESS function for both DBH and DBH^2, respectively (Insightful, Inc. 2003). The parameters of MLP were set at 24 nodes in one hidden layer, sigmoid activation function, the learning rate $\eta = 0.4$, and a momentum coefficient $\alpha = 0.7$. The settings for RBF were 50 centers with Gaussian activation function. The GWR model was fitted using the GWR 2.0 software with the Gaussian kernel function. The bandwidth in Eq. (15) was determined as $h = 4$ m according to the variograms of tree DBH (Fig. 3(a)) and the OLS model residuals (Fig. 3(b)). This bandwidth was also used for computing local Moran coefficient for the model residuals from the six modeling methods.

The model residuals were defined as the difference between the observed and predicted CROWN, and the absolute model residuals were calculated by taking the absolute values of the model residuals. To examine the model residuals across tree sizes, all trees in the stand

Table 2
Model fitting statistics for the six modeling techniques

Model	R^2	SSE	Test[a]	P-value
OLS	0.76	50228		
LMM	0.96	9325	$\chi^2 = 2584$	<0.0001
GAM	0.78	45558	$F = 24.59$	<0.0001
MLP	0.79	44746		
RBF	0.79	44969		
GWR	0.94	11746	$F = 4.69$	<0.0001

[a] *Note*: Hypothesis test for testing the improvement of model fitting over OLS.

Table 3
Characteristics of the model residuals from the six modeling techniques

Model	Mean	Std	Skewness	Kurtosis	Minimum	25% Q	Median	75% Q	Maximum
OLS	0.000	2.90	0.72	16.05	−30.08	−1.49	−0.37	1.48	38.04
LMM	0.000	1.25	0.72	16.05	−12.96	−0.64	−0.16	0.64	16.39
GAM	0.001	2.76	0.43	17.95	−35.67	−1.33	−0.27	1.34	32.08
MLP	0.008	2.74	0.61	13.72	−31.20	−1.26	−0.14	1.26	31.85
RBF	0.025	2.74	0.72	18.57	−34.03	−1.26	−0.13	1.19	35.96
GWR	0.034	1.40	0.58	8.10	−11.32	−0.74	−0.03	0.75	16.44

were grouped into diameter classes. Average model and absolute residuals were calculated for each diameter class.

The spatial distributions of the model residuals from the six modeling techniques were investigated using the global and local Moran coefficients (Anselin, 1995; Tiefelsdorf, 2000; Boots, 2002). The global Moran coefficient (MC) is defined by

$$\mathrm{MC} = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}(h)(e_i-\bar{e})(e_j-\bar{e})}{\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}(h)\sum_{i=1}^{n}(e_i-\bar{e})^2} \quad (16)$$

where e_i and e_j denote the model errors at locations i and j, respectively, $\bar{e}$ is the mean of e_i over the n locations, and $c_{ij}(h)$ is the spatial weight measure within a given distance or bandwidth h (i.e., $h = 4$ m in this study). If location j is a neighbor of the subject location i, then $c_{ij}(h) = 1$; otherwise $c_{ij}(h) = 0$. The expected mean of the MC is $-1/(n-1)$. The Moran coefficient is positive when the observed values of locations within the distance (h) 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. The expected value and variance of the MC for sample size n can be calculated using two sampling assumptions: normality or randomization (Cliff and Ord, 1981; Lee and Wong, 2001). A normal test for the null hypothesis of no spatial autocorrelation between observed values over the n locations can be conducted based on the standardized MC.

Anselin (1995) showed that the global Moran coefficient could be decomposed into local values. The local form of the Moran coefficient is given by:

$$\mathrm{MC}_i = (e_i-\bar{e})\sum_{j=1}^{n}c_{ij}(h)(e_j-\bar{e}) \quad (17)$$

The first component $(e_i - \bar{e})$ is the difference between the model residual e_i at the reference location i and the mean, while the second component $\sum c_{ij}(h)(e_j - \bar{e})$ is the sum of differences between the neighboring model residuals e_j and the mean. A positive local MC_i indicates a cluster of data values around i, similar to that at i, that deviate strongly (either positively or negatively) from $\bar{e}$. A negative local MC_i describes the same situation except that the sign of the model error at i is opposite to that of its neighbors. If either e_i or the values of e_j in the neighborhood of i are close to $\bar{e}$, the local MC_i will indicate no spatial autocorrelation (Boots, 2002). When the local MC_i is standardized by division by the variance $(\sum (e_j - \bar{e})^2/n)$, a pseudo-significant level of MC_i can be obtained by a conditional randomization or permutation approach (Lee and Wong, 2001).

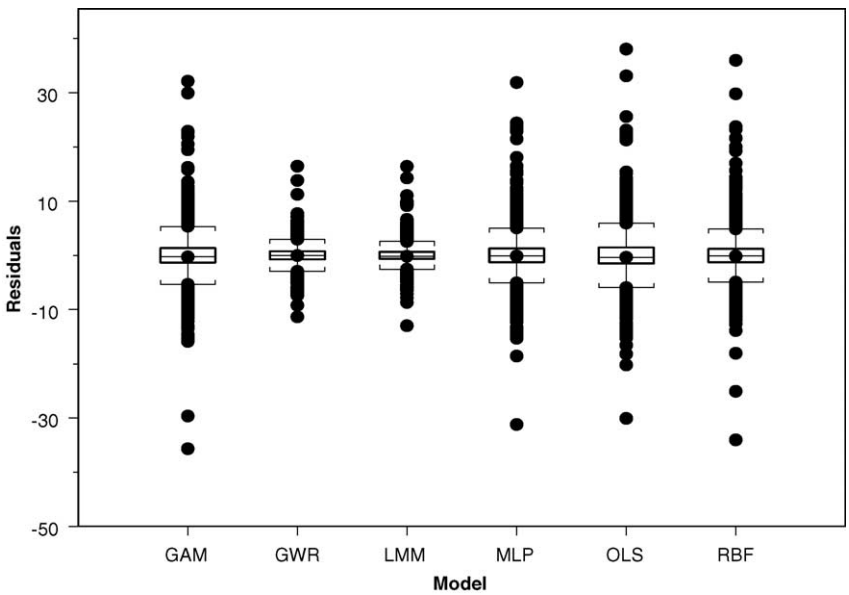


Fig. 4. Box plot of the model residuals from the six modeling techniques.

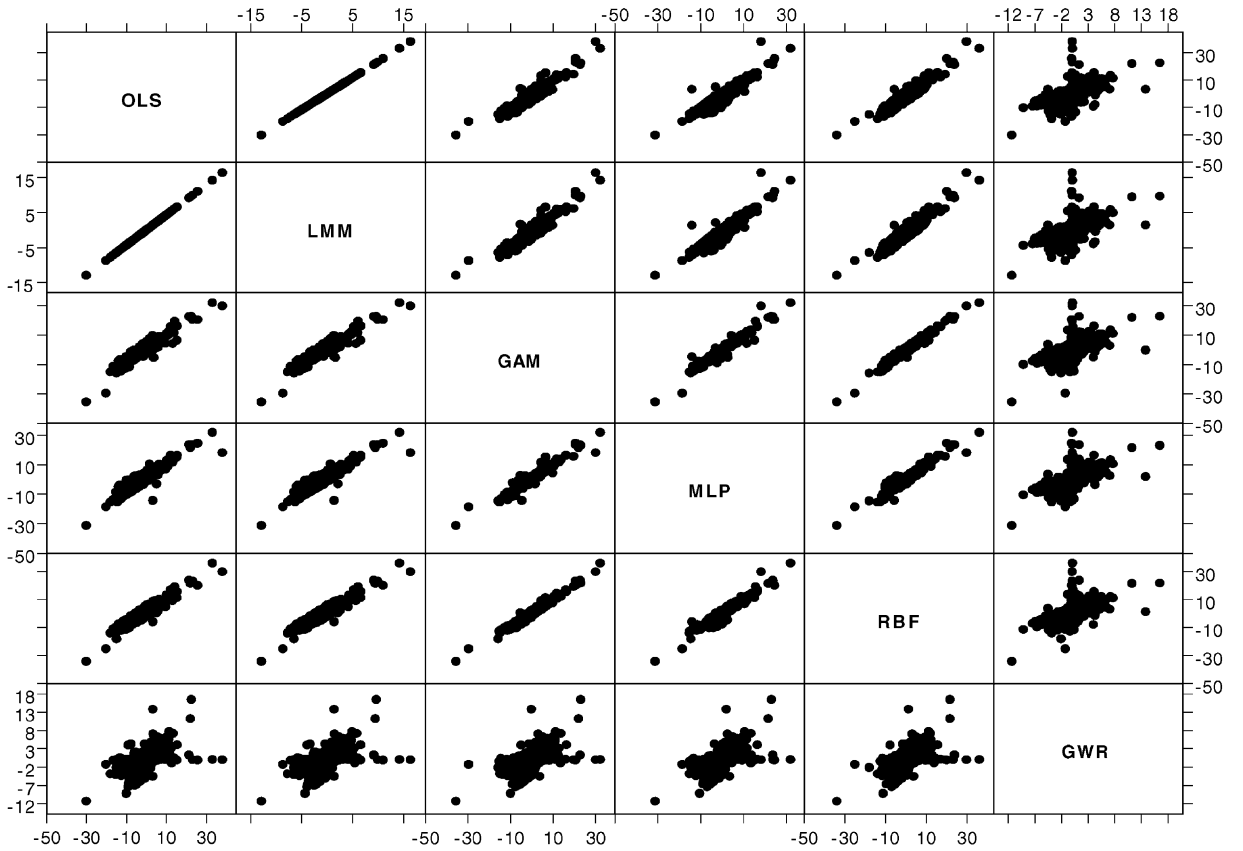


Fig. 5. Matrix plot for the relationships of the model residuals between the six modeling techniques.

4. Results and discussion

4.1. Model fitting

The OLS model fitted the tree data reasonably well, if the violation of the independence assumption was ignored. The OLS model had $R^2 = 0.76$, and the error sum of squares (SSE) = 50,228. All model coefficients were statistically significant ($P < 0.05$), but these tests may be biased due to the spatial autocorrelations between observations. To account for the spatial autocorrelations among the trees, LMM was fitted to the data. The exponential covariance structure was selected since it had the smallest Akaike's AIC compared with other alternative covariance structures. LMM significantly improved model fitting since its SSE (9325) was much smaller than that of OLS. The null model likelihood ratio test was also statistically significant ($P < 0.0001$), indicating that the exponential covariance structure was preferred to the diagonal one of the OLS model (Table 2).

Table 2 also indicated that the GAM model fit the data better than did the OLS model. The GAM model

Table 4
Global Moran coefficient (MC) of the model residuals from the six modeling techniques

Model	Global MC	*Z*-value[a]	*Z*-value[b]
OLS	0.087	10.30	10.31
LMM	0.087	10.30	10.31
GAM	0.087	10.35	10.37
MLP	0.111	13.11	13.13
RBF	0.096	11.40	11.43
GWR	−0.151	−17.82	−17.83

[a] Standard normal test based on the normality assumption.
[b] Standard normal test based on the randomization assumption.

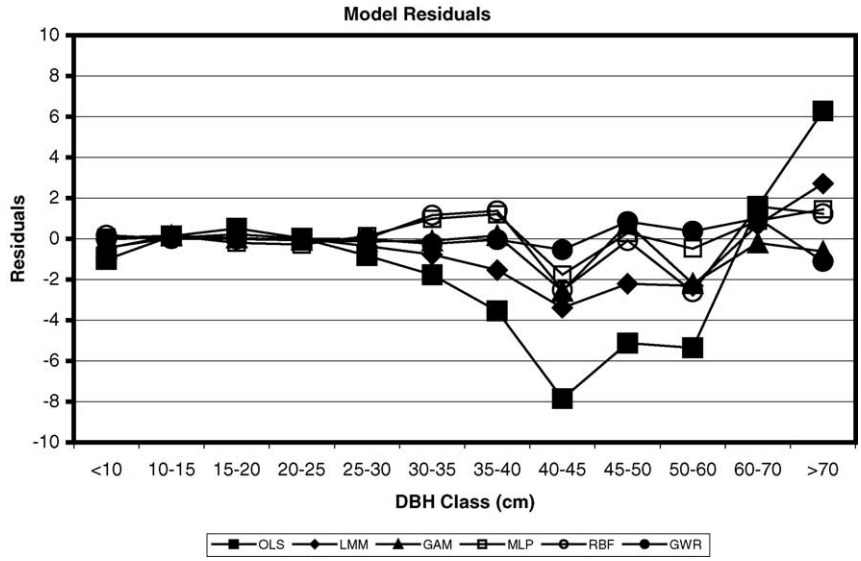


Fig. 6. Model residuals across tree diameter classes.

SSE (45,558) was smaller than the OLS model's SSE (50,228). The F-test ($F = 24.59$) for testing the improvement of GAM over OLS was highly significant ($P < 0.0001$) (Hastie and Tibshirani, 1990; Venables and Ripley, 1997). In addition, the three LOESS functions were statistically significant ($P < 0.0001$). The MLP and RBF models produced similar model R^2 and SSE as the GAM model that were better than those of the OLS model (Table 2). Since there were no degrees of freedom available in the output of LNKnet software for the SSEs of MLP and RBF, no statistical test was performed for the improvement of the two neural network models over OLS. However, the SSEs of MLP and RBF were even smaller than that of GAM (Table 2).

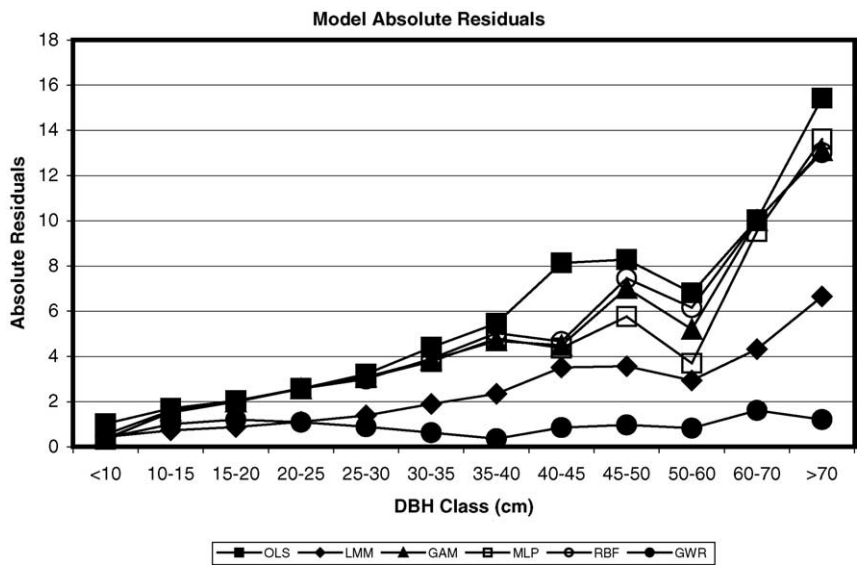


Fig. 7. Absolute model residuals across tree diameter classes.

Table 5
Local Moran coefficient (MC_i) of the model residuals from the six modeling techniques

Model	Mean	Std	Skewness	Kurtosis	Minimum	25% Q	Median	75% Q	Maximum
OLS	0.406	4.22	1.36	157.00	−94.72	−0.38	0.04	0.67	80.42
LMM	0.406	4.22	1.36	157.00	−94.72	−0.38	0.04	0.67	80.42
GAM	0.409	4.13	1.45	159.08	−92.88	−0.32	0.03	0.64	83.11
MLP	0.518	4.39	−0.36	177.37	−118.27	−0.25	0.05	0.69	64.02
RBF	0.450	4.23	0.61	184.66	−110.94	−0.26	0.05	0.65	84.09
GWR	−0.706	3.97	−26.99	1043.86	−167.49	−0.79	−0.09	0.12	12.30

Table 6
Z-value of the local Moran coefficient of the model residuals from the six modeling techniques

Model	Mean	Std	Skewness	Kurtosis	Minimum	25% Q	Median	75% Q	Maximum
OLS	0.188	2.00	2.42	149.41	−40.31	−0.18	0.02	0.32	40.28
LMM	0.188	2.00	2.42	149.41	−40.31	−0.18	0.02	0.32	40.28
GAM	0.189	2.02	3.27	170.00	−43.74	−0.16	0.02	0.32	39.95
MLP	0.235	2.12	1.61	163.55	−48.37	−0.13	0.02	0.33	42.23
RBF	0.206	2.06	3.12	174.83	−45.39	−0.13	0.02	0.32	43.30
GWR	−0.340	1.77	−23.09	811.12	−74.99	−0.38	−0.04	0.06	6.15

We suspected that the MLP and RBF also improved the data fitting significantly over the OLS model.

The GWR model fit the data much better than the OLS model. The average model R^2 was 0.94, and the error sum of squares was 11,746 (Table 2). A *F*-test test was used for testing the improvement of the GWR model over the OLS model (Fotheringham et al., 2002). The results indicated that the GWR model improved model fitting significantly ($F = 4.69$ and $P < 0.0001$) over the OLS model (Table 2). This implies the rela-

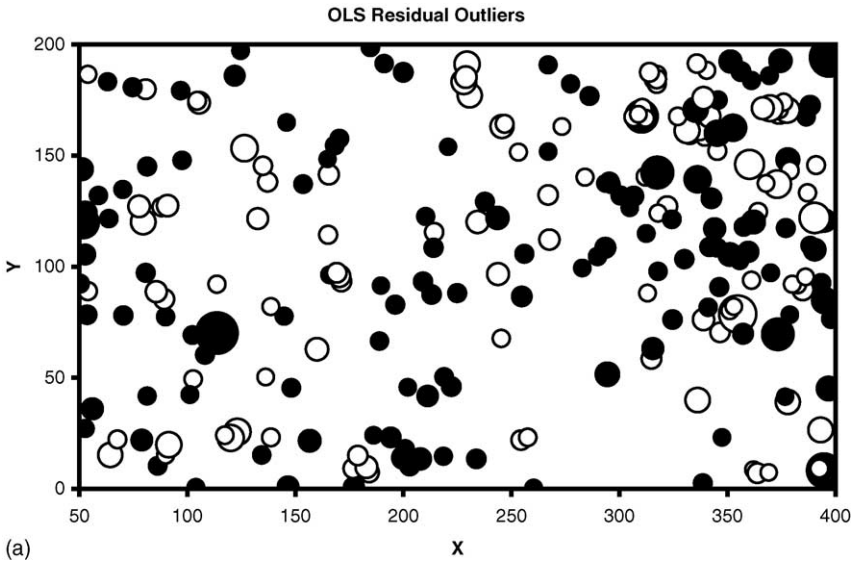


Fig. 8. (a) Plot of OLS model residual outliers. The size of the symbols (black dot and circle) is proportional to the model residuals. The black dots represent positive residuals, and the circles represent negative residuals. Plot of (b) LMM, (c) GAM, (d) MLP, (e) RBF and (f) GWR model residual outliers.

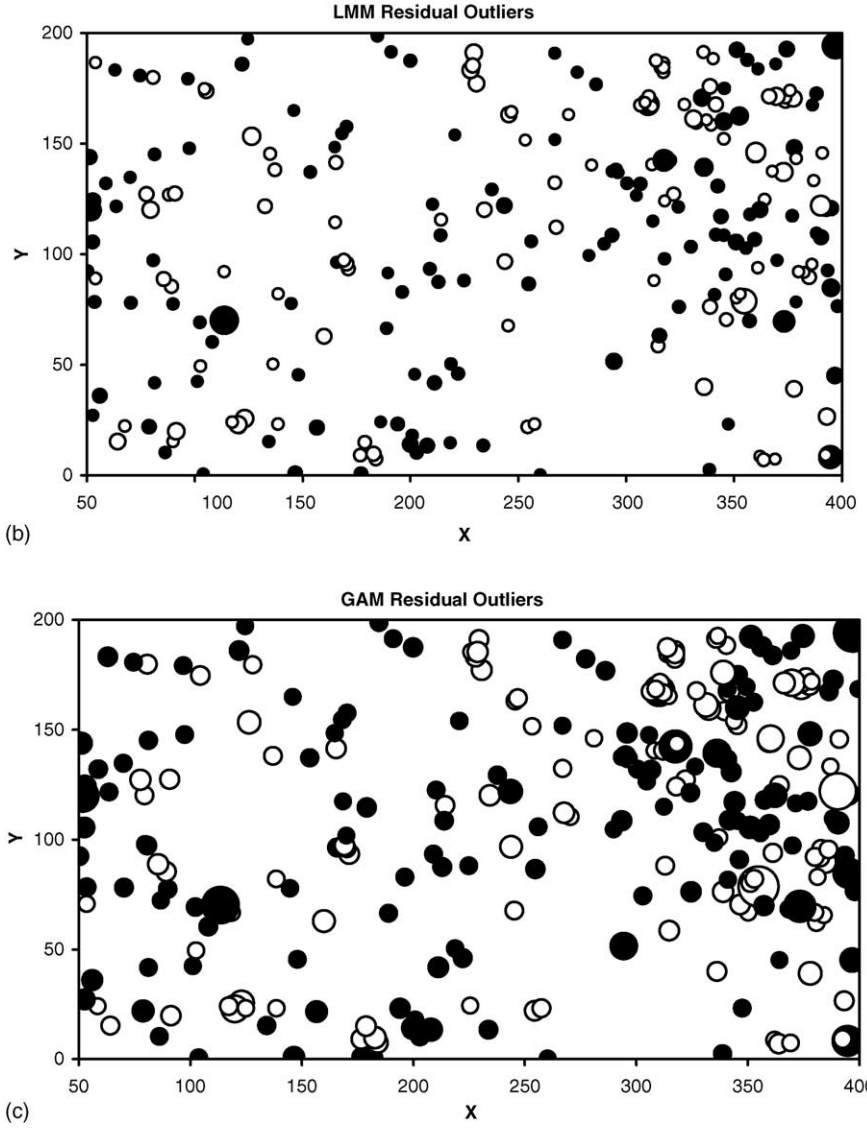


Fig. 8. (*Continued*)

tionship between CROWN and DBH was not constant across the stand under study and the model parameters should vary from sub-area to sub-area within the stand.

4.2. Conventional analysis of model residuals

First we assessed the six models by examining overall model residuals, residual distributions, and residuals across tree size classes. Table 3 and Fig. 4 show that the model residuals from OLS, GAM, MLP, and RBF had similar average, standard deviation, skewness, kurtosis, and quartiles. The above four models produced positive skewness and large positive kurtosis for the residual frequency distributions. On the other hand, the LMM model yielded a smaller range for the model residuals, although its residual frequency distribution had similar skewness and kurtosis to those of the last four models. In contrast, the GWR model produced positive skewness, much smaller kurtosis (about two times smaller), and smaller range for the model residuals than the OLS

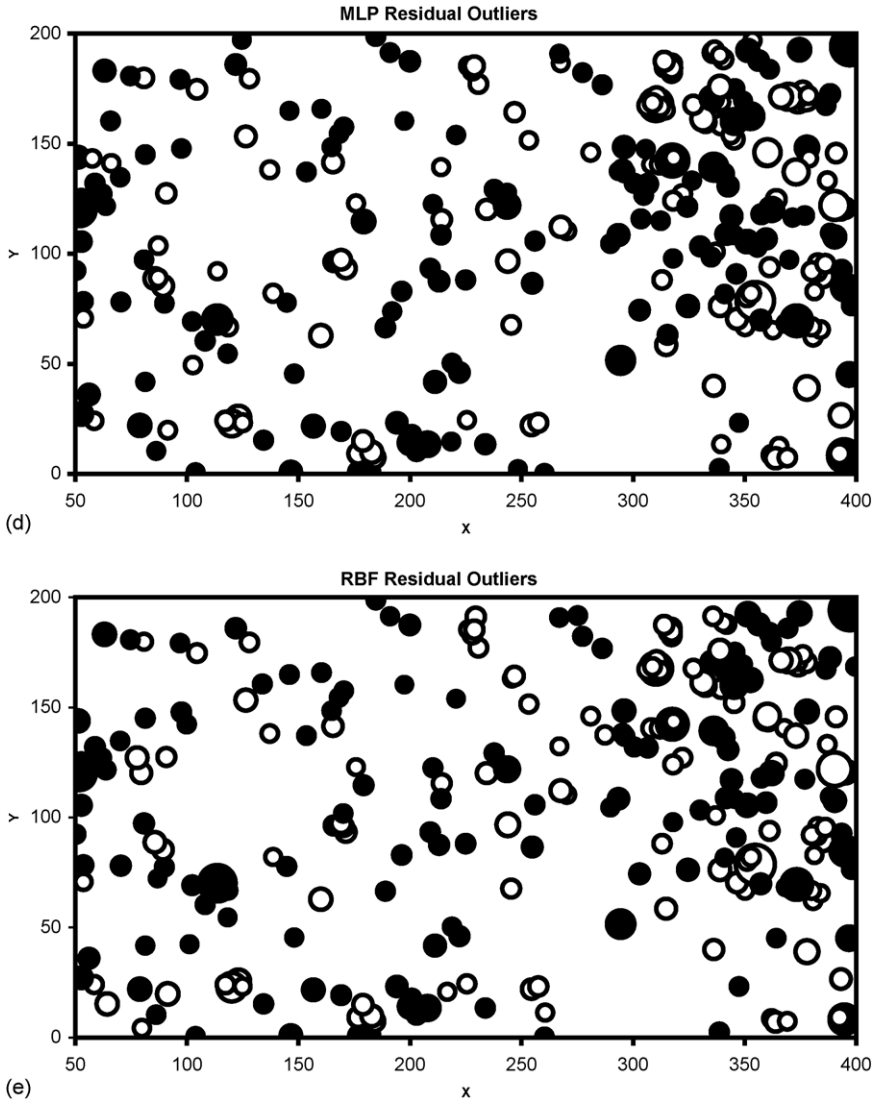


Fig. 8. (*Continued*)

model (Table 3). Fig. 5 illustrates that the residuals from OLS, LMM, GAM, MLP, and RBF have strong linear relationships with each other, while the model residuals from GWR are different from the other five models.

Fig. 6 shows the average model residuals across the diameter classes for the six models. It appears that the six models produce similar model residuals for trees up to 30 cm in diameter. However, the OLS model produces larger negative residuals for large-sized trees (40–60 cm in diameter) and larger positive residuals for trees > 70 cm in diameter. The LMM, GAM, MLP, and RBF models generate similar residual patterns across tree diameter classes. Fig. 7 illustrates the average model absolute residuals across the diameter classes for the six models. It is clear that the GWR model consistently yields much smaller absolute residuals across the diameter classes, especially for large-sized trees. The LMM model is the second best model in terms of the average model absolute residuals across the diameter classes.

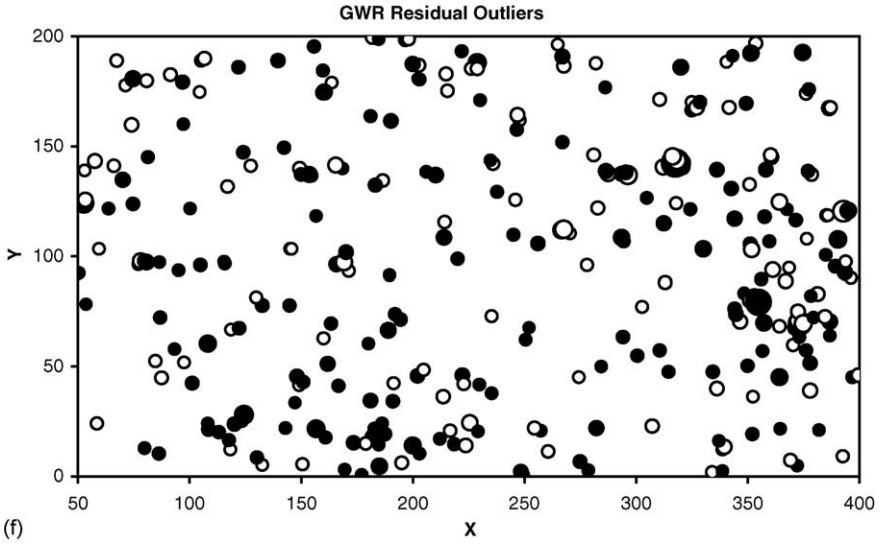


Fig. 8. (*Continued*).

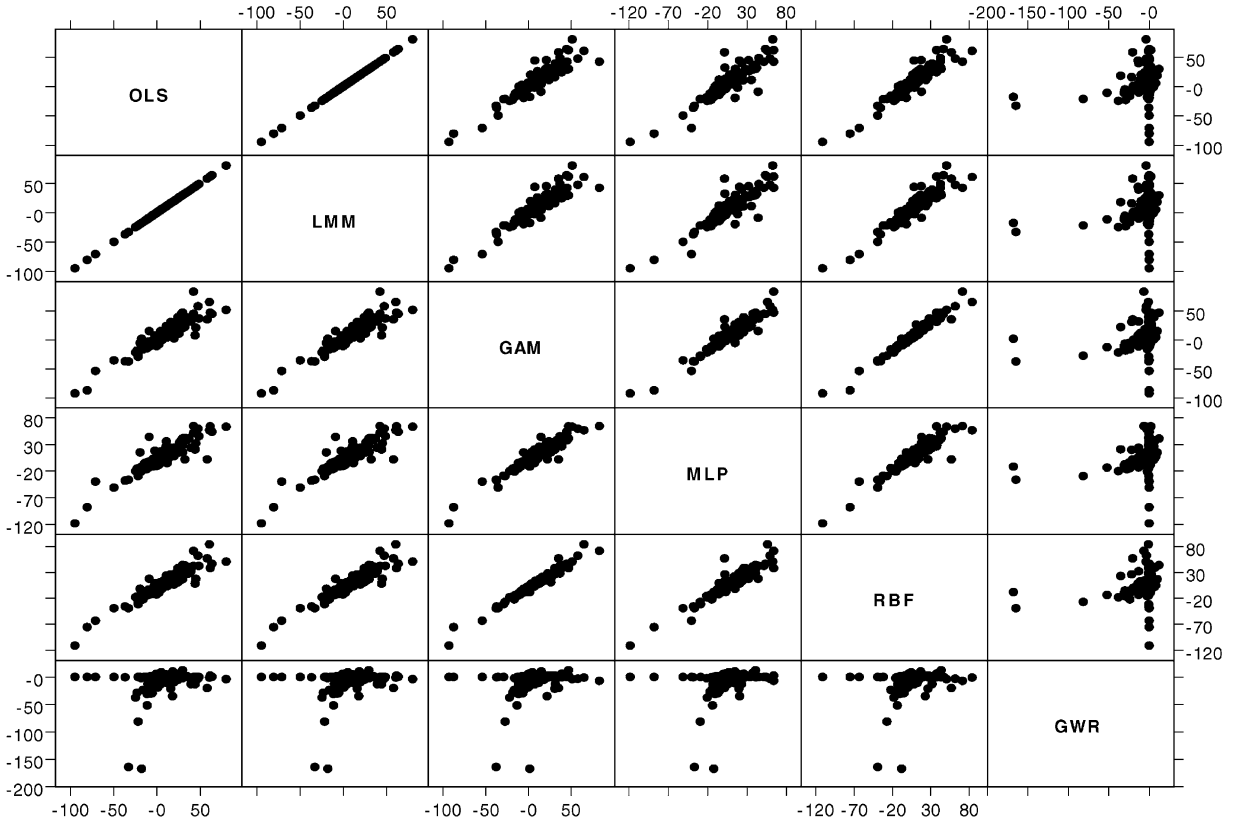


Fig. 9. Matrix plot for the relationships of the local Moran coefficients between the six modeling techniques.

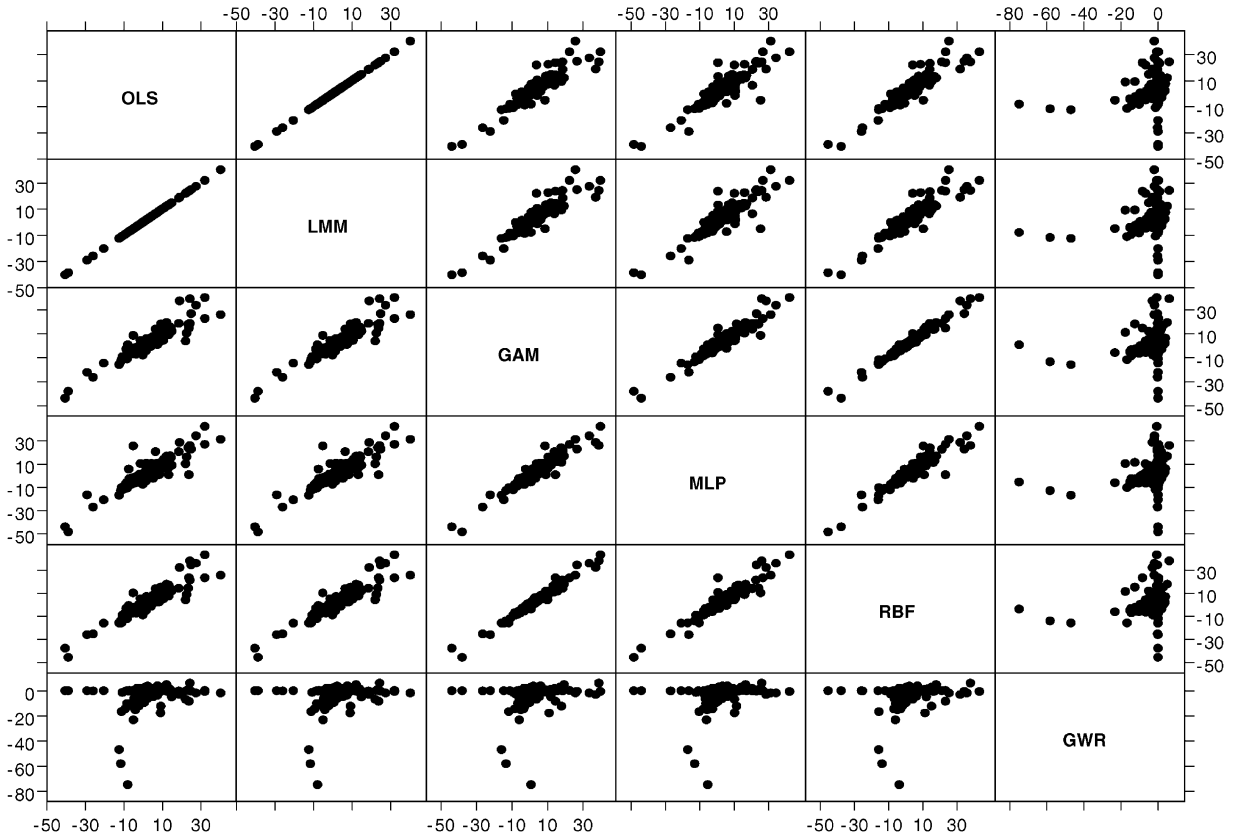


Fig. 10. Matrix plot for the relationships of the *Z*-values of the local Moran coefficients between the six modeling techniques.

Table 7
Comparison of the significant *Z*-values for the local Moran coefficient

Model	# of significant $\|Z\| = 1.96$, $n = 5979$ (%)	Among the significant *Z*-values	
		$Z \leq -1.96$ (%)	$Z \geq 1.96$ (%)
OLS	335 (5.6)	91 (27.2)	244 (72.8)
LMM	335 (5.6)	91 (27.2)	244 (72.8)
GAM	341 (5.7)	105 (30.8)	236 (69.2)
MLP	385 (6.4)	112 (29.1)	273 (70.9)
RBF	361 (6.0)	107 (29.6)	254 (70.4)
GWR	295 (4.9)	277 (93.9)	18 (6.1)
	# of significant $\|Z\| = 3.30$, $n = 5979$ (%)	$Z \leq -3.30$ (%)	$Z \geq 3.30$ (%)
OLS	153 (2.6)	42 (27.5)	111 (72.5)
LMM	153 (2.6)	42 (27.5)	111 (72.5)
GAM	159 (2.7)	45 (28.3)	114 (71.7)
MLP	177 (3.0)	49 (27.7)	128 (72.3)
RBF	168 (2.8)	50 (29.8)	118 (70.2)
GWR	121 (2.0)	117 (96.7)	4 (3.3)

Number in parenthesis is the percentage.

4.3. Spatial assessment of model residuals

The global MC was computed for the model residuals from the six models. Table 4 showed that the global MC for the residuals of the OLS, LMM, GAM, MLP and RBF models were all significantly positive (Z-values > 1.96), which indicated that, across the stand, the above five models produced model residuals in clusters of similar values (i.e., either positive or negative values). On the other hand, the global MC for the GWR model residual was a significant negative value (Z-value < −1.96). It means that the model

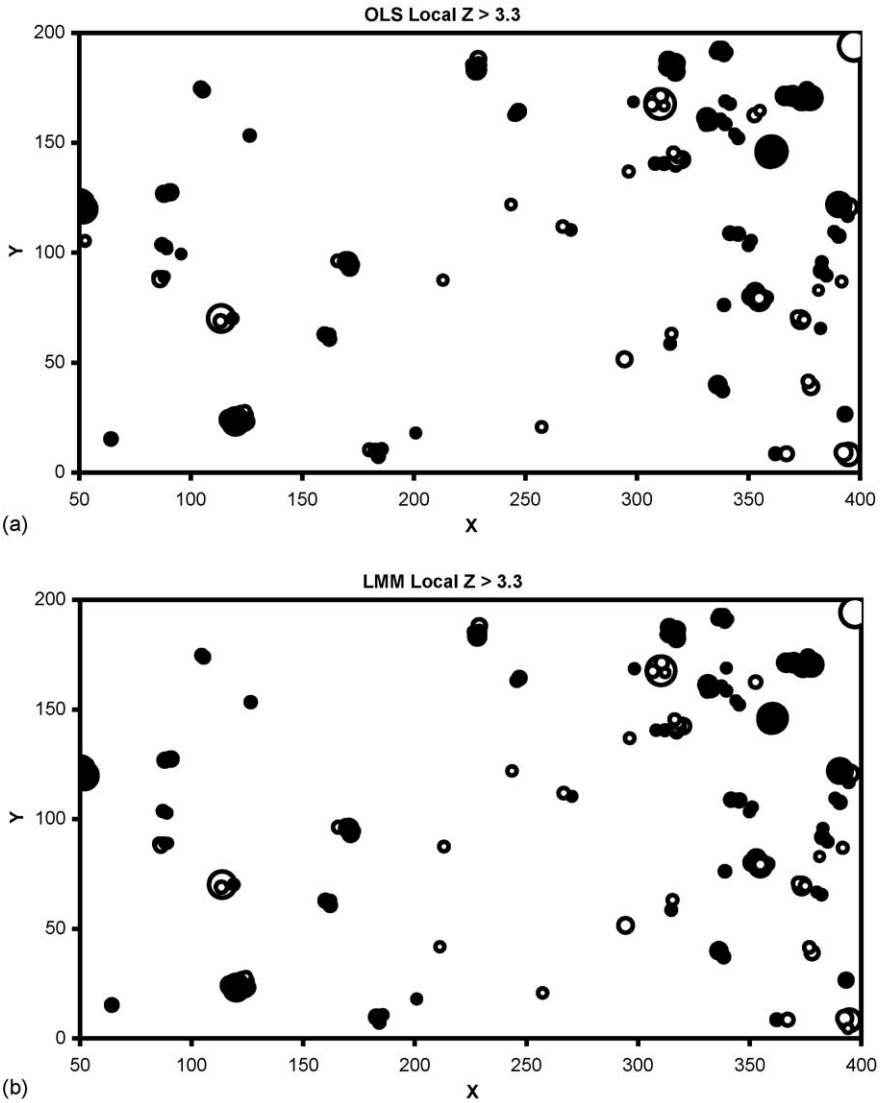


Fig. 11. (a) Plot of Z-value outliers of the local Moran coefficient for the OLS model. The size of the symbols (black dot and circle) is proportional to the Z-value. The black dots represent positive local Z-value, and the circles represent negative local Z-value. (b) Plot of Z-value outliers of the local Moran coefficient for the LMM model. (c) Plot of Z-value outliers of the local Moran coefficient for the GAM model. (d) Plot of Z-value outliers of the local Moran coefficient for the MLP model. (e) Plot of Z-value outliers of the local Moran coefficient for the RBF model. (f) Plot of Z-value outliers of the local Moran coefficient for the GWR model.

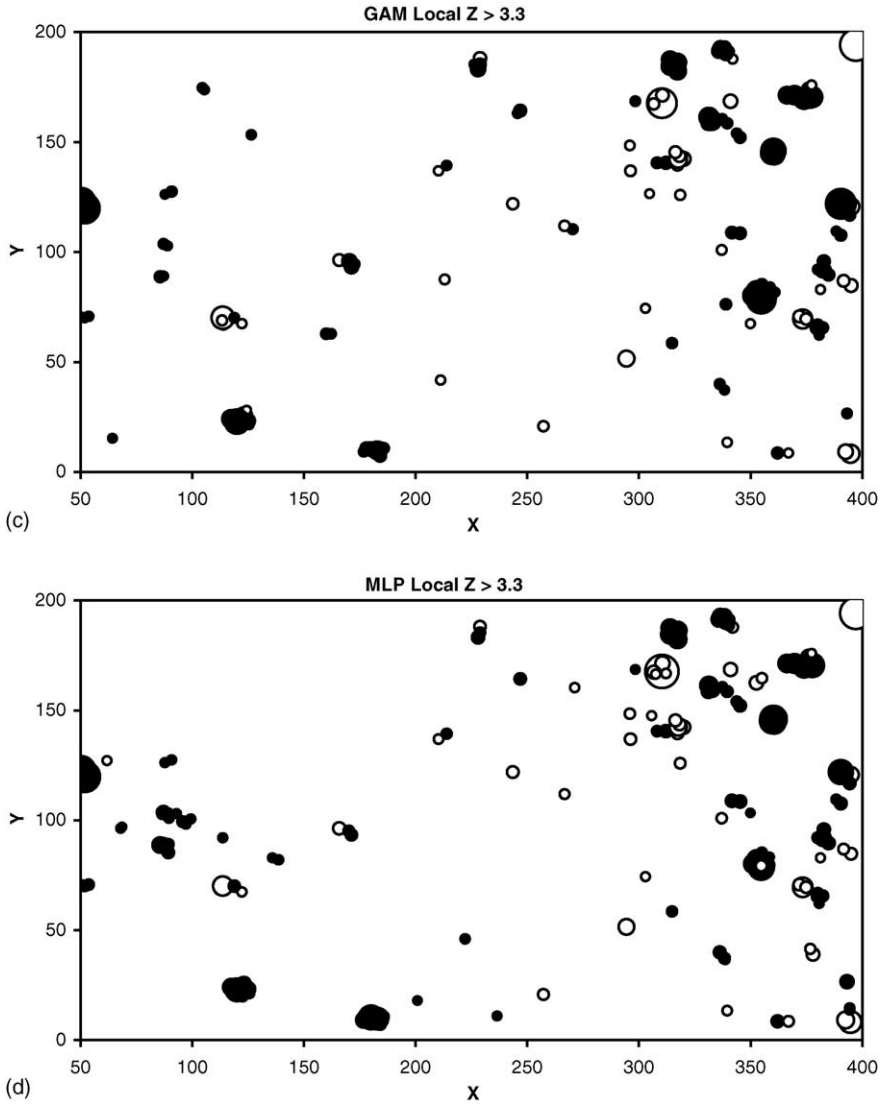


Fig. 11. (*Continued*)

residuals from the GWR model were, on average, in clusters of dissimilar values (i.e., a positive residual was surrounding by negative residuals, and vice versa).

Variograms were also used to investigate the spatial variability of model residuals among distance classes. The range of the variogram indicates that there is no spatial autocorrelation between model residuals beyond this distance (Isaaks and Srivastava, 1989; Kohl and Gertner, 1997). Fig. 3(b) shows that the OLS model residuals have a range of 4 m, the same as that of tree DBH (Fig. 3(a)). The ranges of the variograms are 4.4 m for LMM (Fig. 3(c)), 4.9 m for GAM (Fig. 3(d)), 4.5 m for MLP (Fig. 3(e)), and 5.4 m for RBF (Fig. 3(f)), respectively.

It is difficult to show every model residual in a map due to a large number of trees involved in the study. We chose to show the outliers of model residuals, i.e. the residuals larger than two times standard deviation of the residuals in magnitude. It is clear that the

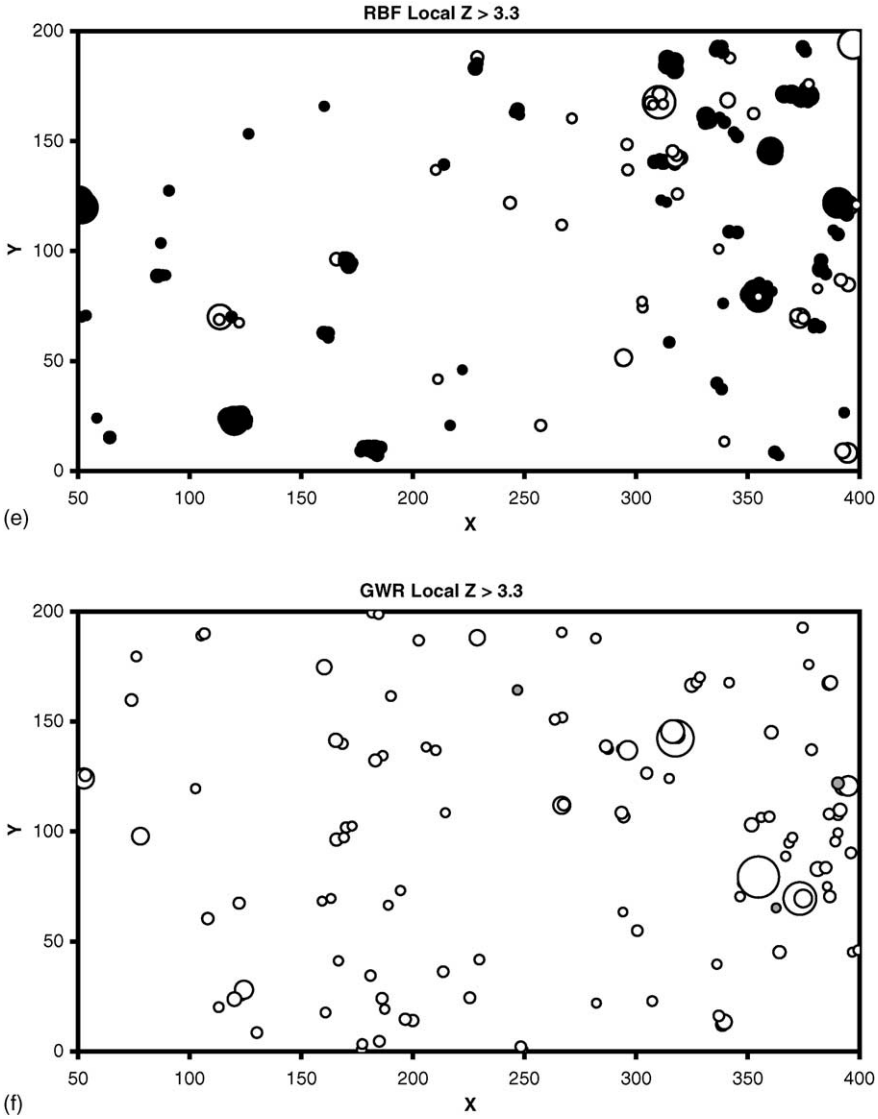


Fig. 11. (*Continued*).

spatial distributions of the residual outliers from GAM (Fig. 8(c)), MLP (Fig. 8(d)), and RBF (Fig. 8(e)) are almost identical to those of OLS (Fig. 8(a)). Although the magnitude of the residual outliers from LMM is much smaller than those from OLS, they have a similar spatial pattern to the above four models in terms of size, sign and clustering (Fig. 8(b)). On the other hand, Fig. 8(f) illustrates that the residual outliers from the GWR model are much smaller in magnitude and have, in general, a different spatial pattern across the stand from the other five models.

4.4. Comparison of six models

Local indicator of spatial association has been proved to be a useful tool to identify "hot spots" (positive autocorrelation, or similarity) and "cold spots" (negative autocorrelation, or dissimilarity) of values

(Anselin, 1995; Boots, 2002; Shi and Zhang, 2003). Therefore, the local Moran coefficient (MC_i) was computed for each model residual from each of the six models with the bandwidth of 4.0 m, and *Z*-value was also computed for each corresponding local MC_i. Table 5 indicated that the local MC_i for the model residuals from the OLS, LMM, GAM, MLP and RBF models had similar averages, standard deviations, ranges and percentiles. Evidently, the above five models produced larger and more frequent positive local MC_i values with strong linear relationships among the models (Fig. 9). In contrast, the GWR model produced more negative local MC_i, or more "cold spots" of dissimilar model residuals (Table 5 and Fig. 9). The *Z*-values of the local MC_i had similar patterns as the local MC_i values for the six modeling techniques (Table 6 and Fig. 10).

Due to the problems of multiple comparisons for the local MC_i for the entire data set, it is necessary to adjust the significance levels for testing the significance of the local MC_i for each location. One possibility is to apply the Bonferroni adjustment in which the significance level for each individual location is α/n where n is the sample size. However, this adjusted local significance level is too conservative for a large sample size (i.e., $n = 5979$ in this study), and may not be appropriate for testing local LISA (Anselin, 1995; Boots, 2002). Therefore, the local *Z*-values for the local MC_i were evaluated for the significance levels of 0.05 ($Z_{\alpha/2} = 1.96$), and 0.001 ($Z_{\alpha/2} = 3.30$). Table 7 indicated that the six models produced similar percentage of significant local *Z*-values out of the 5979 locations for the two significance levels. Among the significant *Z*-values, there were about 70% positive *Z*-values and 30% negative *Z*-values for the OLS, LMM, GAM, MLP and RBF models, indicating these five models tended to generate more clusters of either positive or negative model residuals in some sub-areas of the stand. Trees in those sub-areas were either all underestimated (positive residuals) or all overestimated (negative residuals) for the response variables. On the other hand, the majority (about 95%) of the local *Z*-values there were negative *Z*-values among the significant *Z*-values for the GWR model (Table 7). If there are clusters of the model residuals existing, a large error tends to be surrounded by smaller neighboring errors and a small error tends to be surrounded by larger neighboring errors.

Fig. 11 illustrates the spatial distributions of the local *Z*-values larger than $|Z_{\alpha/2}| = 3.30$ for the six models. It is obvious that the OLS, LMM, GAM, MLP and RBF produced similar spatial patterns of *Z*-values in terms of size, sign and clustering (Fig. 11(a)–(e)). On the other hand, Fig. 11(f) indicates that the local *Z*-values from the GWR model are much smaller in magnitude (except a few spots) and have, in general, a different spatial pattern across the stand from the other five models.

5. Conclusion

In recent years there has been an increasing interest in applying modern modeling techniques to account for spatial autocorrelations among data observations. Although these techniques do have desirable features such as less restrictive model assumptions, many of them are non-spatial in nature. For example, both RBF and GWR use a Gaussian "kernel" function to process the data. The Gaussian functions in GWR are located in two-dimensional geographical space, whereas the RBF Gaussian functions are located in multi-dimensional space of predictor variables (Murnion, 1999). The LOESS functions used in GAM or other kernel regression methods process the data in a manner similar to RBF. Therefore, many modern modeling techniques do not provide truly spatial error measures (Laffan, 1999). However, current practice in the assessment of model performance focuses on model fitting and overall model accuracy and errors (e.g., Moisen and Frescino, 2002). Little attention has been paid to the spatial heterogeneity of model errors.

In this study, we utilize local Moran coefficients to investigate spatial distribution and heterogeneity in model residuals from six modeling techniques, with ordinary least-squares (OLS) as the benchmark. The results indicate (1) modern modeling techniques such as LMM, GAM, MLP and RBF may improve model fitting to the data and provide better prediction for the response variable than the OLS model, but they produce similar spatial patterns for the model residuals as the OLS model does, (2) OLS, LMM, GAM, MLP and RBF models yield more residual clusters of similar values, indicating that trees in some sub-areas were either all underestimated or all overestimated for the response variable, and (3) GWR, a local modeling

method, produces more accurate predictions for the response variable, as well as more desirable spatial distribution for the model residuals than the ones derived from other five modeling techniques.

Acknowledgment

This research was supported by funds provided by the U.S. Department of Agriculture, Forest Service, Northeastern Research Station, RWU 4104. The authors would like to thank the Editor and two anonymous reviewers for their valuable suggestions and helpful comments on the manuscript.

References

Anderson, R.P., Lew, D., Peterson, A.T., 2003. Evaluating predictive models of species distributions: criteria for selecting optimal models. Ecol. Model. 162, 211–232.

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

Anselin, L., Griffith, D.A., 1988. Do spatial effects really matter in regression analysis. Pap. Reg. Sci. Assoc. 65, 11–34.

Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. Ecol. Model. 157, 101–118.

Austin, M.P., Meyers, J.A., 1996. Current approaches to modeling the environmental niche of eucalypts: implication for management of forest biodiversity. For. Ecol. Manage. 85, 95–106.

Boots, B., 2002. Local measures of spatial association. EcoScience 9, 168–176.

Brunsdon, C.A., Fotheringham, A.S., Charlton, M.E., 1998. Geographically weighted regression — modeling spatial non-stationary. Statistician 47, 431–443.

Cliff, A.D., Ord, J.K., 1981. Spatial Processes: Models and Applications. Pion, London, 266 pp.

Dale, M.R., Fortin, M.-J., 2002. Spatial autocorrelation and statistical tests in ecology. EcoScience 9, 162–167.

Ek, A.R., 1969. Stem map data for three forest stands in northern Ontario. For. Res. Lab., Sault Ste. Marie, Ontario. Information Report O-X-113, 23 pp.

Fischer, M.M., 1997. Computational neural networks: a new paradigm for spatial analysis. Environ. Plan. A 29, 1873–1891.

Foody, G.M., 2003. Geographical weighting as a further refinement to regression modeling: an example focused on the NDVI-rainfall relationship. Remo. Sens. Environ. 88, 283–293.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. John Wiley & Sons, New York, 269 pp.

Fox, J.C., Ades, P.K., Bi, H., 2001. Stochastic structure and individual-tree growth models. For. Ecol. Manage. 154, 261–276.

Frescino, T.S., Edwards Jr., T.C., Moisen, G.G., 2001. Modeling spatially explicit forest structure attributes using generalized additive models. J. Veg. Sci. 12, 15–26.

Gregoire, T.G., Schabenberger, O., Barrett, J.P., 1995. Linear modeling of irregular spaced, unbalanced, longitudinal data from permanent-plot measurements. Can. J. For. Res. 25, 137–156.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135, 147–186.

Guisan, A., Edwards Jr., T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Model. 157, 89–100.

Gullison, J.J., Bourque, C.P.A., 2001. Spatial prediction of tree and shrub succession in a small watershed in Northern Cape Breton Island, Nova Scotia. Ecol. Model. 137, 181–199.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman & Hall, New York, 335 pp.

Insightful, Inc., 2003. S-Plus 6.2 Guide to Statistics, vols. I and II. Insightful Corporation, Seattle, WA.

Isaaks, E.H., Srivastava, R.M., 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York, 561 pp.

Kohl, M., Gertner, G., 1997. Geostatistics in evaluating forest damage surveys: considerations on methods for describing spatial distributions. For. Ecol. Manage. 95, 131–140.

Kukoloch, L., Lippmann, R., 1999. LNKnet User's Guide, MIT Lincoln Laboratory. (The software and the manual can be downloaded at http://www.ll.mit.edu/IST.).

Laffan, S.W., 1999. Spatially assessing model error using geographically weighted regression. GeoComputation 99. Available at http://www.geocomputation.org/1999/086/gc_086.htm.

Lee, J., Wong, D.W.S., 2001. Statistical Analysis with ArcView GIS. John Wiley and Sons Inc., New York, 192.

Lehmann, A., 1998. GIS modeling of submerged macrophyte distribution using generalized additive models. Plant Ecol. 139, 113–124.

Lehmann, A., Overton, J.McC., Leathwick, J.R., 2003. GRASP: generalized regression analysis and spatial prediction. Ecol. Model. 160, 165–183.

Lek, S., Guegan, J.F., 1999. Artificial neural networks as a tool in ecological modeling. Ecol. Model. 120, 65–73.

Lippmann, R.P., 1987. An introduction to computing with neural nets. IEEE ASSP Mag, April 4–22.

Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., 1996. SAS system for mixed models, SAS Institute Inc., Cary, NC, 1996, 633 pp.

Mitra, S., Basak, J., 2001. FRBF: a fuzzy radial basis function network. Neural Comp. Appl. 10, 244–252.

Moisen, G.G., Edwards Jr., T.C., 1999. Use of generalized linear models and digital data in a forest inventory of northern Utah. J. Agric. Biol. Environ. Stat. 4, 372–390.

Moisen, G.G., Frescino, T.S., 2002. Comparing five modeling techniques for predicting forest characteristics. Ecol. Model. 157, 209–225.

Moisen, G.G., Culter, D.R., Edwards Jr., T.C., 1999. Generalized linear mixed models for analyzing error in a satellite-based vegetation map of Utah. In: Mowrer, H.T., Congalton, R.G. (Eds.), Quantifying spatial uncertainty in natural resources. Theory

and application for GIS and remote sensing. Ann Arbor Press, Chelsea, Michigan, USA, pp. 37–44.

Murnion, S., 1999. Exploring spatial non-stationarity with radial basis function neural networks. Geo. Environ. Model. 3, 35–45.

Overmars, K.P., de Koning, G.H.J., Veldkamp, A., 2003. Spatial autocorrelation in multi-scale land use models. Ecol. Model. 164, 257–270.

Preisler, H.K., Rappaport, N.G., Wood, D.L., 1997. Regression methods for spatially correlated data: an example using beetle attacks in a seed orchard. For. Sci. 43, 71–77.

Rathert, D., White, D., Sifneos, J.C., Hughes, R.M., 1999. Environmental correlates of species richness for native freshwater fish in Oregon, USA. J. Biogeog. 26, 1–17.

Robertson, M.P., Peter, C.I., Villet, M.H., Ripley, B.S., 2003. Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modeling techniques. Ecol. Model. 164, 153–167.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1987. Learning internal representations by error propagation. In: Rumelhart, J.L., McClelland, PDP Research Group (Eds.), Parallel Distributed Processing—Explorations in the Microstructure of Cognition, vol. 1. The MIT Press, MA, pp. 318–362.

Sarle, W.S., 1994. Neural networks and statistical methods. In: Proceedings of the Nineteenth Annual SAS Users Group International Conference, Cary, NC. SAS Institute.

SAS Institute, Inc., 2002. SAS/STAT Users' guide. Version 9.0. SAS Institute, Inc., Cary, NC.

Schabenberger, O., Pierce, F.J., 2002. Contemporary Statistical Models for the Plant and Soil Sciences. CRC Press, Boca Raton, FL, 738 pp.

Shi, H., Zhang, L., 2003. Local analysis of tree competition and growth. For. Sci. 49, 938–955.

Shin, M., Goel, A.L., 2000. Empirical data modeling in software engineering using radial basis functions. IEEE Trans. Soft. Eng. 26, 1–10.

Sokal, R.R., Oden, N.L., Thomson, B.A., 1998a. Local spatial autocorrelation in a biological model. Geog. Anal. 30, 331–354.

Sokal, R.R., Oden, N.L., Thomson, B.A., 1998b. Local spatial autocorrelation in a biological variables. Biol. J. Linnean Soc. 65, 41–62.

Tappeiner, U., Tappeiner, G., Aschenwald, J., Tasser, E., Ostendorf, B., 2001. GIS-based modeling of spatial pattern of snow cover distribution in an alpine area. Ecol. Model. 138, 265–275.

Tiefelsdorf, M., 2000. Modeling spatial processes: the identification and analysis of spatial relationships in regression residuals by means of Moran's I. Lecture Notes Earth Sci. 87, Springer, 167 pp.

Venables, W.N., Ripley, B.D., 1997. Modern applied statistics with S-Plus, second ed. Springer, New York, 548 pp.

Warner, B., Misra, M., 1996. Understanding neural networks as statistical tools. Am. Stat. 50 (4), 284–293.

Zaniewski, A.E., Lehmann, A., Overton, J.M., 2002. Predicting species spatial distribution using presence-only data: a case study of native New Zealand ferns. Ecol. Model. 157, 261–280.

Zhang, L., Shi, H., 2004. Local modeling of tree growth by geographically weighted regression. For. Sci. 50, 225–244.

Zhang, L., Bi, H., Cheng, P., Davis, C.J., 2004. Modeling spatial variations in tree diameter-height relationships. For. Ecol. Manage. 189, 317–329.