

# Bayesian geographically weighted regression and its application for local modeling of relationships between tree variables

Nirmal Subedi<sup>(1)</sup>,  
Lianjun Zhang<sup>(1)</sup>,  
Zhen Zhen<sup>(2)</sup>

Geographically weighted regression (GWR) has become popular in recent years to deal with spatial autocorrelation and heterogeneity in forestry and ecological data. However, researchers have realized that GWR has some limitations, such as correlated model coefficients across study areas, strong influence of outliers, weak data problem, etc. In this study, we applied Bayesian geographically weighted regression (BGWR) and a robust BGWR (rBGWR) to model the relationship between tree crown and diameter using observed tree data and simulated data to investigate model fitting and performance in order to overcome some limitations of GWR. Our results indicated that, for observed tree data, the rBGWR estimated tree crown more accurate than both BGWR and GWR. For the simulated data, 74.1% of the estimated slope coefficients by rBGWR and 73.4% of the estimated slope coefficients by BGWR were not significantly different ( $\alpha = 0.05$ ) from the corresponding simulated slope coefficients. The estimation of model coefficients by rBGWR was not sensitive to outliers, but the coefficient estimation by GWR was strongly affected by those outliers. The majority of the coefficient estimates by rBGWR and BGWR for weak observations were not significantly ( $\alpha = 0.05$ ) different from the simulated coefficients. Therefore, BGWR (including rBGWR) may be a better alternative to overcome some limitations of GWR. In addition, both BGWR and rBGWR were more powerful than GWR to detect the spatial areas with non-constant variance or spatial outliers.

**Keywords:** Spatial Autocorrelation, Spatial Heterogeneity, Robust Regression, Spatially Varying Coefficients Models

## Introduction

Most tree or plant communities display some degree of spatial structure in the form of geographical patchiness or gradients. Trees or plants at close distances interact to each other (e.g., by competition) more than those farther away (Legendre & Fortin 1989, Legendre 1993, Fortin & Dale 2005). In forestry, young stands commonly exhibit positive spatial autocorrelation due to micro-site patchiness, whereas the competition among trees promotes negative spatial autocorrelation (Reed & Burkhart 1985, Fox et al. 2001). In addition, spatial effects (i.e., spatial autocorrelation and heterogeneity) in tree or plant communities may exist at multiple spatial scales and for

many variables (Wiens 1989, Levin 1992, Ma et al. 2012, Ma & Zhang 2015). Thus, incorporating spatial effects into statistical data analysis and modeling requires creative approaches to describe real-world systems (Green et al. 2005). In the last decade, researchers are increasingly interested in understanding the causes and consequences of spatial autocorrelation and heterogeneity in ecosystem functions (Lovett et al. 2005).

When spatial autocorrelation and heterogeneity exist in forestry and ecological data, the independence and homogeneity assumptions of traditional statistical methods, e.g., ordinary least squares (OLS), may be violated (Green et al. 2005). In recent

years, different spatial models have been developed to deal with spatial effects in the relationships between variables. Based on the spatial scales used in the modeling process, spatial models can be categorized into global and local models (Haining 2004, Schabenberger & Gotway 2005, Lloyd 2006). Global models, such as linear mixed models and spatial autoregressive models (Zhang et al. 2009, Meng et al. 2009), assume that spatial variation is the same everywhere. Thus, a single model is developed using the whole data set and is used for the entire study area. Commonly, a global model requires a device to model spatial autocorrelation among observations in neighboring locations, through either a covariance matrix that can be estimated using a variogram or spatial weight matrix based on spatial proximity of neighbors (Lu & Zhang 2010, 2012, 2013). Obviously, global models do not well represent spatial variations at any individual location and may not be effective to deal with spatial heterogeneity.

In contrast, local models, such as geographically weighted regression (GWR), fit a regression relationship for each spatial location using the neighbors within a given bandwidth (Fotheringham et al. 2002). In recent years, GWR has been applied in various disciplines, including ecology, forestry, real estate, and healthcare (McMillen 2003,

---

□ (1) Department of Forest and Natural Resources Management, State University of New York College of Environmental Science and Forestry, One Forestry Drive, Syracuse, NY 13210 (USA); (2) School of Forestry, Northeast Forestry University, Harbin (China)

@ Zhen Zhen (zhzhen@syr.edu)

Received: Jul 30, 2017 - Accepted: Jun 11, 2018

**Citation:** Subedi N, Zhang L, Zhen Z (2018). Bayesian geographically weighted regression and its application for local modeling of relationships between tree variables. *iForest* 11: 542-552. - doi: 10.3832/ifor2574-011 [online 2018-09-01]

*Communicated by:* Luca Salvati

Foody 2004, Zhang & Shi 2004, Lin & Wen 2011). Local models are more useful to explore locational spatial variation (heterogeneity) in the relationships between variables. In forestry and ecological studies, local models can be a useful tool for evaluating, testing, and modeling the influence of micro-site variation, competition status, growth potential, and management activities on trees or plants. Further, these local model coefficients and model fitting statistics can be readily displayed using GIS or graphic packages to explore local “hot” or “cold” spots across the study area (Zhang & Shi 2004).

However, researchers have realized that GWR has several limitations: (i) lack of a unifying approach to estimate local coefficients; (ii) incorrect estimates of dispersion of model coefficients, as GWR utilizes the same sample data observations (with different spatial weights) to produce a sequence of parameter estimates for all locations in the space. Consequently, the estimated model coefficients are dependent across spatial locations, and the correct expression for the variance of local coefficients could be non-linear; (iii) the influence of outliers (i.e., unusual data observations on Y-axis), because local coefficient estimates can be strongly affected by a single outlier; (iv) weak data (i.e., any observation with insufficient neighbors within its effective bandwidth) problem, especially if the “pre-selected” bandwidth is small; and (v) assumption violation of homogeneous error variance (LeSage 2004, Wheeler & Páez 2010).

To overcome the limitations of GWR, a Bayesian approach was proposed to estimate the local regression coefficients, namely Bayesian geographically weighted regression (BGWR – LeSage 1997, 2004). BGWR introduced the concept of parameter smoothing, which polls the strength among observations. In other words, parameter smoothing defines the relationship between the local regression coefficients of a subject location and the local coefficients of other locations in the study area. Further, a robust BGWR (rBGWR) was also introduced to deal with problems in data such as outliers and weak data (LeSage 1997, 2004).

However, BGWR requires the use of Gibbs sampler, a Monte Carlo simulation, to estimate its parameters. The Gibbs sampler (Gelfand & Smith 1990, Gelfand 2000) is one of the most frequently used methods of Markov Chain Monte Carlo (MCMC). It generates random samples from a (marginal) probability distribution indirectly, without the need to calculate the density itself. By simulating a large enough sample, the sample characteristics such as mean and variance can be computed to the desirable degree of accuracy (Casella & George 1992, Banerjee et al. 2004). Gelfand & Smith (1990) demonstrated how such methods can be useful for a variety of Bayesian inference problems.

To date, limited studies have used the BGWR methods. LeSage (2004) compared GWR and BGWR with econometric data. Furutani (2004) used BGWR to study land prices in Yokohama, Japan. Similarly, Clifton & Romero-Barrutieta (2006) used BGWR to study the role of geography and institutions on growth and development for a region of the United States. However, we are not aware of any applications of BGWR in the fields of ecology or forestry. The objectives of this study were to: (i) fit a relationship between tree crown area and diameter at breast height (dbh) to observed data and simulated data using GWR, BGWR, and rBGWR; (ii) compare model fitting and performance of the three modeling methods; and (iii) investigate the usefulness, advantages, and limitations of BGWR and rBGWR in handling outliers and weak data problems.

## Theoretical background

### Geographically weighted regression

A linear regression relationship between variables can be expressed as (eqn.1):

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

where  $Y$  is an  $n \times 1$  vector of the response variable,  $X$  is an  $n \times p$  matrix with a (first) column of one and  $(p - 1)$  explanatory variables,  $\beta$  is a  $p \times 1$  vector of model coefficients, and  $\varepsilon$  is an  $n \times 1$  vector of random error term with i.i.d.  $N(0, \sigma^2 I)$ , where  $I$  denotes an  $n \times n$  identity matrix. Traditionally, the relationship represented by eqn. 1 is assumed to be universal or constant across the geographical study area.

Now suppose that we have a set of locations  $s_i$  ( $i = 1, 2, \dots, n$ ) with geographic coordinates. The underlying model for geographically weighted regression (GWR) can be expressed as (eqn.2):

$$W(s_i)Y = W(s_i)X\beta(s_i) + W(s_i)\varepsilon(s_i) \quad (2)$$

where  $Y$  and  $X$  are defined as in eqn. 1,  $W(s_i)$  is an  $n \times n$  diagonal matrix of spatial weight,  $\beta(s_i)$  is a  $p \times 1$  vector of model coefficients associated with the subject location  $s_i$ , and  $\varepsilon(s_i)$  is an  $n \times 1$  vector of random error term with i.i.d.  $N(0, \sigma^2 I)$  associated with the subject location  $s_i$ . The aim of GWR is to obtain non-parametric estimates of the regression model for each geographical location  $s_i$ . This can be achieved by using neighboring observations near location  $s_i$  as follows: (i) find a point at one particular location  $s_i$ ; (ii) compute the distance-based spatial weight matrix  $W(s_i)$ ; and (iii) estimate the model coefficients using weighted least-squares regression (Brunsdon et al. 1996, Fotheringham et al. 2002 – eqn. 3):

$$\hat{\beta}(s_i) = (X^T W(s_i) X)^{-1} X^T W(s_i) Y \quad (3)$$

where  $T$  is the transposed matrix and the other terms are as defined above. Further,

using the analogies of local information matrix and hat matrix, Fotheringham et al. (2002) gave the following expression to compute the asymptotic variance of  $\beta(s_i)$  (eqn. 4):

$$\text{Var}(\hat{\beta}(s_i)) = H(s_i) H(s_i)^T \sigma^2 \quad (4)$$

where  $H(s_i) = [X(s_i)^T W(s_i) X(s_i)]^{-1} X(s_i)^T W(s_i)$ . Unfortunately, this expression is incorrect for variance of local coefficients (LeSage 2004, Wheeler & Páez 2010).

The spatial weight matrix  $W(s_i)$  is a function of the distance between the subject location ( $s_i$ ) and neighboring observations ( $s_j$ ), and determines the influence of the neighbors on the parameter estimation of local regression coefficients. Three weight functions are commonly used to compute the spatial weight matrix, including Gaussian, exponential, and bi-square weight functions (Fotheringham et al. 2002). The bandwidth of the weight function is either fixed (i.e., Gaussian and exponential functions) or variable (i.e., bi-square function – Guo et al. 2008). The “optimal” bandwidth can be determined using the cross-validation (CV) method, the Akaike’s information criterion (AIC) or pre-defined by the researcher (Farber & Páez 2007).

### Bayesian geographically weighted regression

The underlying model of Bayesian geographically weighted regression (BGWR) is the same as that of GWR in eqn. 2. However, the model error term  $\varepsilon(s_i)$  assumes to follow i.i.d.  $N(0, \sigma^2 V(s_i))$ , where  $V(s_i) = \text{diag}(V_{s_{i1}}, V_{s_{i2}}, \dots, V_{s_{in}})$  is an  $n \times n$  diagonal matrix. Basically,  $V(s_i)$  represents a set of non-constant variances across space and needs to be estimated from the data. BGWR requires the use of a parameter smoothing specification, which is a mathematical function representing the relationship among spatially varying coefficients. For example, a distance smoothing parameter specification is as follows (eqn. 5):

$$\beta(s_i) = (w_{s_{i1}} \otimes I_p, \dots, w_{s_{in}} \otimes I_p) \begin{bmatrix} \beta(s_{i1}) \\ \vdots \\ \beta(s_{in}) \end{bmatrix} + u(s_i)$$

where  $w_{s_{ij}}$  represents normalized distance-based spatial weights, i.e., sum of row vector  $(w_{s_{i1}}, \dots, w_{s_{in}})$  is unity and assumes  $w_{s_{ij}} = 0$ . The stochastic variation term  $u(s_i)$  follows  $N[0, \sigma(s_i)^2 \delta^2 (X(s_i)^T W(s_i) X(s_i))^{-1}]$ . Note the variance of  $u(s_i)$  contains a scale factor  $\delta^2$  which quantifies the amount of variation among the local coefficient vector  $\beta(s_i)$ . The role of this scale factor in estimating local parameters will be discussed later.

The parameter smoothing function and assumption of heterogeneous variance ensure that BGWR captures spatial heterogeneity and autocorrelation. On the other hand, these assumptions add complexity to the modeling process. Thus, a Bayesian approach is required to estimate the model coefficients. The likelihood of BGWR

(LeSage 2004) can be represented as (eqn. 6):

$$\hat{\beta}(s_i) = R(s_i) \cdot \left[ \tilde{X}(s_i)^T V(s_i)^{-1} \tilde{Y}(s_i) + \tilde{X}(s_i)^T \tilde{X}(s_i) J(s_i) \gamma / \delta^2 \right]^{-1} \cdot \left[ \tilde{X}(s_i)^T V(s_i)^{-1} \tilde{X}(s_i) + \tilde{X}(s_i)^T \tilde{X}(s_i) / \delta^2 \right]^{-1} \quad (6)$$

where  $\tilde{Y} = W(s_i) Y(s_i)$ ,  $\tilde{X}(s_i) = W(s_i) X(s_i)$ ,  $J(s_i) = (w_{i1} \otimes I_p, \dots, w_{in} \otimes I_p)$  and  $\gamma^T = (\beta(s_1), \dots, \beta(s_n))$ . Note that when  $V(s_i) = I_n$ , and  $\delta = \infty$ , both  $\tilde{X}(s_i)^T \tilde{X}(s_i) / \delta^2$  and  $\tilde{X}(s_i)^T \tilde{X}(s_i) J(s_i) \gamma / \delta^2$  are zero and the BGWR model is reduced to the GWR model.

Using eqn. 6 and the distribution of  $u(s_i)$ , the conditional distribution of  $\hat{\beta}(s_i)$ , given  $\delta, \sigma(s_i), V(s_i)$ , and  $\gamma$ , is a multivariate normal and can be expressed as (eqn. 7):

$$p(\hat{\beta}(s_i) | \delta, \sigma(s_i), V(s_i), \gamma) \propto N(\hat{\beta}(s_i), \sigma(s_i)^2 R(s_i)) \quad (7)$$

where  $\delta, \sigma(s_i)$  and  $V(s_i)$ , are hyper-parameters (i.e., the parameters of prior distributions), and  $\gamma$  represents the value of other  $\hat{\beta}(s_i)$  for observations  $j \neq i$ . Geweke (1993) proved the existence of the posterior density and moments of this model.

Choosing a prior distribution or prior information for the hyper-parameters is an important issue in Bayesian inference, as it may be sensitive or robust to the choice of the prior distribution (Congdon 2006). In this study, the conditional prior distributions of the BGWR hyper-parameters (i.e.,  $\delta, \sigma(s_i)$  and  $V(s_i)$ ) are as follows:

(1) The conditional prior distribution of scale factor  $\delta$  (eqn. 8):

$$p(\delta | \dots) \propto \text{Gamma}(a, b) \quad (8)$$

where  $a$  and  $b$  are the Gamma distribution parameters. The mean and variance of the Gamma distribution are  $a/b$  and  $a/b^2$ , respectively. Alternatively, the conditional prior distribution of  $\delta$  is  $\chi^2(np)$ , such that (eqn. 9):

$$\omega = \beta(s_i) - J(s_i) \gamma$$

$$p(\delta | \dots) = \delta^{-np} \cdot \exp \left\{ - \sum_{i=1}^n \frac{\omega^T [\tilde{X}(s_i)^T \tilde{X}(s_i)]^{-1} \cdot \omega}{2 \sigma(s_i)^2 \delta^2} \right\} \quad (9)$$

Note that the value of  $\delta$  is very important for defining the relation of  $\hat{\beta}(s_i)$  with other neighboring  $\hat{\beta}(s_j)$ ,  $j \neq i$ . When the inherent dispersion of the GWR parameters is large for a given data set, it will require a larger  $\delta$  value to fit BGWR (LeSage 2004). A very small value of  $\delta$  imposes the constraints on the variation of parameters  $\hat{\beta}(s_i)$ , and it becomes a distance-weighted linear combination of neighboring  $\hat{\beta}(s_j)$ . In contrast, a larger value of  $\delta$  allows the parameter  $\hat{\beta}(s_i)$  to be estimated from the data and less influenced by other neighboring  $\hat{\beta}(s_j)$ . As a result, an informative prior for  $\delta$  is chosen

in this study. It can be estimated using an empirical Bayes method (Carlin & Louis 2000). Once the estimate is obtained, we can choose a set of values of  $\delta$  (e.g., multiply by some constants such as 0.1, 0.25 and 0.5) to explore the interaction of the scale factor with the variation of estimated local coefficients.

(2) The conditional prior distribution of  $\sigma(s_i)$  (eqn. 10):

$$p(\sigma(s_i) | \dots) \propto \sigma(s_i)^{-(m+1)} \cdot \exp \left\{ - \frac{1}{2 \sigma(s_i)^2} [\varepsilon(s_i)^T V(s_i)^{-1} \varepsilon(s_i)] \right\} \quad (10)$$

where  $\varepsilon(s_i) = \tilde{Y}(s_i) - \tilde{X}(s_i) \hat{\beta}(s_i)$ , and  $m$  denotes the number of observations with non-negligible weights. Hence, the posterior distribution of  $\sigma(s_i)$  is  $\chi^2(m)$ . It is a squared sum of  $m$  independent variables with standard normal distributions.

(3) The conditional prior distribution of  $V(s_i)$  (eqn. 11):

$$p \left( \frac{r}{V(s_i)} | \dots \right) \propto \chi^2(r) \quad (11)$$

where  $r$  is a hyper-parameter that controls the amount of dispersion in the  $V(s_i)$  estimates across the observations. This type of prior has been used previously (Lindley 1971, Geweke 1993, LeSage 1997). The motivation for assigning this prior to  $V(s_i)$  is that the mean of prior equals unity and the prior variance is  $2/r$  (LeSage 2004). Thus, introduction of a single hyper-parameter  $r$  can generate  $n^2$  parameter estimates. However, in order to constrain  $V(s_i)$  in relation to  $\varepsilon(s_i)$  and  $\sigma(s_i)$ , the following conditional distribution is used to generate the posterior samples of  $V(s_i)$  (eqn. 12):

$$p \left( \frac{\varepsilon(s_i)^2}{\sigma(s_i)^2 + r} \middle| \dots \right) \propto \chi^2(r+1) \quad (12)$$

which provides greater flexibility to  $v_{ij}$  (the  $j^{\text{th}}$  diagonal of  $V(s_i)$ ), depending on how close the prediction for the  $j^{\text{th}}$  observation was while calibrating the local parameters for the  $i^{\text{th}}$  subject location.

Given the likelihood function and the conditional prior distributions of BGWR hyper-parameters, the posterior distribution of  $\hat{\beta}(s_i)$  is simulated using the Gibbs sampler (Gelfand 2000). As there are a large number of parameters in each draw (iteration), two approaches can be adopted to simulate the model parameters. In the first approach, the parameters are estimated sequentially for each observation in each sim-

ulation draw, and it will be computationally very expensive. In the second approach, the parameters are individually simulated across all draws (iterations) for every observation. One limitation of this approach is that there will be no parameter smoothing relationship. However, a robust parameter estimation (i.e., insensitive to outliers) can be implemented by this approach. The rBGWR uses an independent Student-t linear model, and the parameter estimation is powerful enough to handle outliers and to address robustness concerns in practical settings (Geweke 1993). In addition, this approach is computationally very efficient. Hereafter, this simplified Gibbs sampling approach with robust parameter estimation will be referred to as rBGWR.

### Data and methods

We briefly discuss the observed data and simulated data, crown area regression model, model fitting process for each method (GWR, BGWR, and rBGWR), and the evaluation of the three methods.

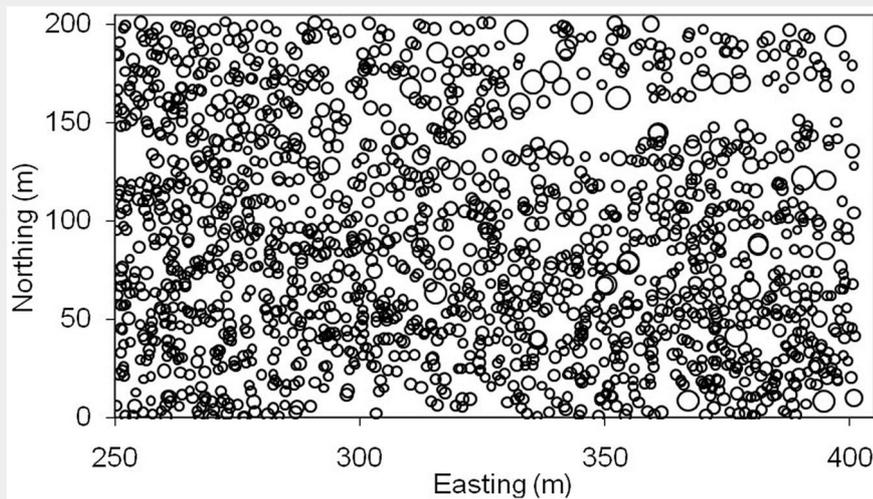
#### Ontario softwood data

The observed data used in this study were a subset of the stem map data of a mature, second growth, uneven-aged softwood stand located near Sault Ste. Marie, Ontario, Canada (Ek 1969). It was 150 × 200 m in size and had a total of 1698 trees. The dominant tree species included balsam fir (*Abies balsamea* [L.] Mill. – 57.6% in number of trees), black spruce (*Picea mariana* [Mill.] BSP. – 33.5%), and white spruce (*Picea glauca* [Moench.] Voss. – 4.6%). Other tree species were in smaller percentages and included white pine (*Pinus strobus* L.), balsam poplar (*Populus balsamifera* L.), Tamarack (*Larix laricina* [Du Roi] K. Koch), and white birch (*Betula papyrifera* Marsh.). The information on measurements included tree location (coordinate), diameter at breast height (dbh), total height, and crown area (crown) for trees dbh > 8.9 cm (Zhang et al. 2008, 2009).

Tab. 1 shows the descriptive statistics of tree dbh and crown, as well as their spatial autocorrelation (Moran's I), relative spatial heterogeneity (SH%), and nearest-neighbor distance (Zhang et al. 2008, 2009). Fig. 1 illustrates the stem map of trees with circle size proportional to the tree dbh. Note that the Ontario softwood data had some special features: (i) clustered spatial pattern of trees; (ii) large variation in tree dbh (e.g., potential outliers); and (iii) large gaps between trees in the upper portion of the

**Tab. 1** - Descriptive statistics of observed tree data (Ontario softwood stand). Moran's I and SH% were calculated at lag distance of 4.48 m and 2.24 m, respectively.

Variable	n	Mean	SD	Min	Max	Moran's I	SH%	Distance (m)
dbh (cm)	1698	18.68	8.47	11.43	84.07	0.087	91.64	2.2 (0.3-11.0)
Crown area (m <sup>2</sup> )	1698	7.79	7.80	0.93	90.48	0.115	80.62	-



**Fig. 1** - Map of tree locations for Ontario softwood data. Circles are proportional to the diameter at breast height of trees.

plot (e.g., potential weak data observations).

#### Regression model

Based on the relationship between tree dbh and crown in the data, we chose the following linear regression model (eqn. 13):

$$\ln(\text{crown}) = \beta_0 + \beta_1 \ln(\text{dbh}) + \varepsilon \quad (13)$$

where crown is the tree crown area (m<sup>2</sup>), dbh is tree diameter at breast height (cm),

$\beta_0$  and  $\beta_1$  are regression coefficients to be estimated,  $\ln$  is the natural logarithm, and  $\varepsilon$  is the model random error. The crown-dbh relationship was used in this study because crown is considered the “engine” of a tree, but it is difficult and time consuming to be measured in field. However, we emphasize that we did not intend to develop a predictive crown-dbh model. Rather, we attempted to demonstrate that BGWR is a flexible alternative modeling approach which can be used to investigate spatial

heterogeneity in the relationships between response variable (tree crown) and covariate (dbh).

#### Simulated data

To objectively verify the performance of BGWR, a simulated tree data set with geographic coordinates, tree dbh and crown area, and known local regression coefficients was generated for each tree. The procedure of generating the simulated data is as follows:

(1) The AMORPHYS software (Valentine et al. 2000) was used to generate tree locations with a clustered spatial pattern. The clustered spatial pattern was chosen because local models such as GWR would have more advantages over global models than random or regular spatial patterns (Zhang et al. 2009). Then, the values of tree dbh were randomly generated from a two-parameter Weibull distribution, and other tree attributes such as tree crown area were calculated using a process-based stand model (Valentine et al. 2000).

(2) The generated tree data were used to fit eqn. 13 by GWR 3.0 (Fotheringham et al. 2002) to obtain the two local regression coefficients ( $\beta_0$  and  $\beta_1$ ) for each tree / location. These local regression coefficients were transformed into a unit scale [0, 1], and used to calculate the cumulative distribution function. The purpose of this step was to estimate the scale-invariant measure of association between  $\beta_0$  and  $\beta_1$  using a bivariate Gaussian copula (Nelsen 1999, Wang et al. 2010).

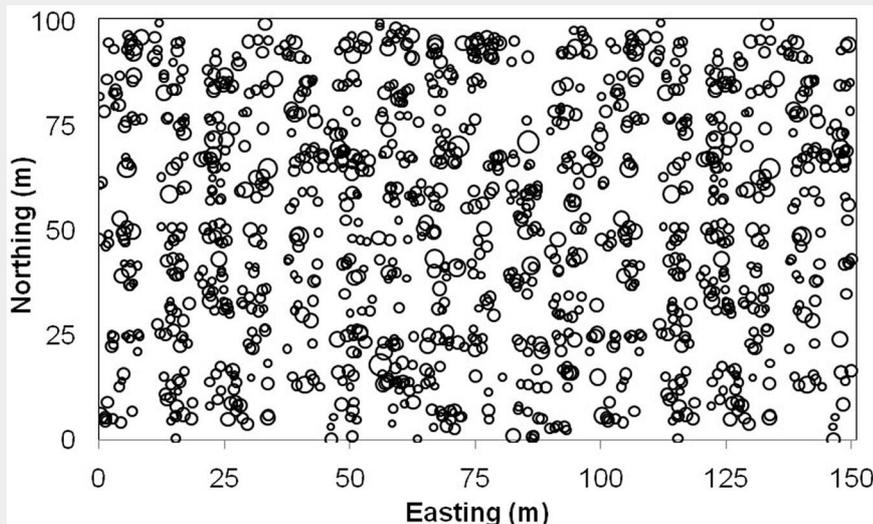
(3) Given the estimated association between  $\beta_0$  and  $\beta_1$ , the bivariate Gaussian copula was used to randomly simulate the two coefficients ( $\beta_0$  and  $\beta_1$ ) in a unit scale [0, 1], and then transform these coefficients back to the original scale and assign the coefficients to the spatial locations generated in step (1).

(4) Create 5 outliers by artificially making the  $\beta_1$  coefficient very large (e.g., two times larger). Create scenarios of weak data (any observation with insufficient neighbors within its effective bandwidth) by deleting a few trees at the locations of low trees density.

(5) Given the randomly generated regression coefficients ( $\beta_0$  and  $\beta_1$ ) and tree dbh generated in step (1), tree crown area was computed using eqn. 13 for each tree / location. Finally, a total of 1230 trees were simulated with geographic coordinates, tree dbh and crown, and known local regression coefficients for each location. Tab. 2 provides a summary of the simulated data. Fig. 2 illustrates the spatial pattern of these simulated trees with circle size proportional to tree dbh.

**Tab. 2** - Descriptive statistics of simulated tree data (n=1230).

Variable	Mean	SD	Min	Max	Moran's I	SH%
DBH (cm)	18.23	9.44	7.43	80.45	0.03	86.40
Crown area (m <sup>2</sup> )	12.94	7.00	4.61	74.91	0.05	86.90
Intercept $\beta_0$	0.201	0.143	-0.260	0.580	0.01	87.80
Slope $\beta_1$	0.807	0.130	0.600	1.233	0.01	92.60
Nearest neighbor distance	1.4	1.26	0.15	6.47	-	-



**Fig. 2** - Map of tree locations for simulated data. Circles are proportional to the diameter at breast height of trees.

#### Model fitting

Three modeling methods, GWR, BGWR, and rBGWR, were used to fit the linear relationship between  $\ln(\text{crown})$  and  $\ln(\text{dbh})$  (eqn. 13) for both Ontario softwood data and simulated data. The GWR model was

fitted using GWR 3.0 software (Fotheringham et al. 2002). The cross validation (CV) method was used to compare three spatial weigh functions, i.e., Gaussian (fixed kernel), exponential (fixed kernel), and tri-cube functions (Farber & Páez 2007). In this study, the same Gaussian spatial weight function (fixed kernel) and bandwidth (10 m) were used for all three methods in order to make compatible model comparison (Fotheringham et al. 2002, Guo et al. 2008).

The BGWR and rBGWR models were fitted using `bgwr.m` and `bgwr.m` Matlab functions available in the Econometric Toolbox (LeSage 1997). These functions were modified to accommodate the options of thinning for longer MCMC chains and calculation of deviance. Both BGWR methods require the prior distributions for the hyper-parameter  $r$  (related to  $V(s_i)$ ) and the scale factor  $\delta^2$ . In this study, three priors (4, 5 and 6) were chosen for the hyper-parameter  $r$ . For each  $r$ , three priors on the scale factor  $\delta^2$  were selected for the BGWR method, which were 10%, 25% and 50% of estimated using the empirical Bayes method with the sample data. The priors on these scale factors larger than 50% of the diffuse prior were not included, because larger priors reduce the efficacy of parameter smoothing and result in BGWR coefficients similar to GWR coefficients (LeSage 2004). Hence, there were nine BGWR models with three priors of  $r$  combined by three priors of  $\delta^2$ . However, the rBGWR models used only three priors of  $r$ .

Posterior distributions of model parameters for each model were sampled using the Gibbs sampler. A single long chain of MCMC was used to generate samples of each coefficient. The convergence of sampled coefficients was diagnosed using the method of Raftery & Lewis (1992) and other exploratory tools. The convergence diagnostic suggested that most of the coefficients estimated by BGWR converged within 4000 samples after a sufficient burn in. However, those estimated coefficients with specification of smaller priors on the scale factor required almost 15,000 samples to converge at the precision of 0.005 (quartile = 0.025 and probability = 0.95).

Using the post convergence values of model coefficients, the Deviance Information Criterion (DIC – Spiegelhalter et al. 2002) was calculated for all the models as an assessment measure of model fitting, such that  $DIC = D(\beta) + pD$ , where  $D(\beta)$  is the average deviance of the mean of posterior model coefficients, and  $pD$  is the effective number of model coefficients, which describes the complexity of the model and serves as a penalization term that corrects deviance's propensity toward models with more parameters. Therefore, DIC is a Bayesian alternative to Akaike's information criterion (AIC). A model with a smaller DIC value is desirable suggesting improved model fitting without excessive parameterization. In order to compare against GWR,

one BGWR model and one rBGWR model with the smallest DIC value were chosen among all fitted models.

**Model evaluation**

The local coefficients estimated by the three methods (GWR, BGWR, and rBGWR) were summarized using descriptive statistics. For the Ontario softwood data, the two regression coefficients estimated by each method were interpolated using inverse distance weighting method and contour maps of these coefficients were generated for each method. Further, the model residuals from each method were assessed across a range of the explanatory variable (dbh).

The local regression coefficients were also evaluated for spatial heterogeneity using intra-block variance ( $V_{intra}$ ) and inter-block variance ( $V_{inter}$  – Li & Reynolds 1995, Garrigues et al. 2006, Zhang et al. 2009) as defined below (eqn. 14, eqn. 15):

$$V_{intra} = \frac{1}{B} \sum_{i=1}^B \frac{1}{n_i} \sum_{j=1}^{n_i} (\beta_{ij} - \bar{\beta}_i)^2 \tag{14}$$

$$V_{inter} = \frac{1}{B} \sum_{i=1}^B (\bar{\beta}_i - \bar{\beta})^2 \tag{15}$$

where  $B$  is the number of blocks,  $n_i$  is the number of observations in the  $i^{th}$  block, and  $\beta_{ij}$ ,  $\bar{\beta}_i$ , and  $\bar{\beta}$  are the estimated local coefficients for  $j^{th}$  observation in the  $i^{th}$  block, the mean value of local coefficients in the  $i^{th}$  block, and the overall mean of coefficients of the whole plot, respectively. In general,  $V_{inter}$  indicates the regional spatial variability, while  $V_{intra}$  quantifies the local spatial variability (Zhang et al. 2009). These measures of spatial heterogeneity were compared among the three modeling methods for the Ontario softwood data.

For the simulated data, the “true” local coefficients of each observation were known. Hence, the 95% of credible limits for each coefficient estimated by BGWR and rBGWR were compared with the corresponding known coefficients for each location. If the 95% credible limits of the posterior samples of a coefficient estimated for

a particular location included the known coefficient, the coefficient estimated by the method for the location was declared to be unbiased. The percentages of unbiased intercept, slope, and both (intercept and slope) coefficients for all locations by BGWR and rBGWR were calculated. However, this assessment was not used for GWR because the estimation of variance of local coefficients was incorrect (LeSage 2004, Wheeler & Páez 2010). Similarly, the estimation bias (defined as simulated–estimated) was computed for the two local coefficients as well as the response variable  $[\ln(\text{crown})]$  for each location and each modeling method.

Model fitting and performance were evaluated for special cases such as outliers and weak data (any observation with insufficient neighbors within its effective bandwidth) by comparing the estimated coefficients and predicted crown area by each modeling method. For the Ontario softwood data, the comparison was only for weak observations (no obvious outliers in the data), while for simulated data, the modeling fitting and performance were evaluated for both outliers and weak observations in terms of the 95% credible limits interval of the corresponding BGWR and rBGWR coefficients, and the estimation bias of model coefficients and tree crown area.

**Results**

*Ontario softwood data*

**Model fitting**

The identification of the “best” BGWR and rBGWR models for different priors of the hyper-parameters  $r$  and  $\delta^2$  represented the first step of this study. Tab. 3 shows the model fitting statistics of the nine BGWR models (with the combination of three priors of  $r$  (4, 5, and 6) and three priors of  $\delta^2$  (10%, 25% and 50% of the estimated  $\delta^2$  by the empirical Bayes method from the data), as well as the three rBGWR models (for three priors of  $r$  4, 5, and 6). The model deviances  $D(\beta)$ , which measured model fit-

**Tab. 3** - Model fitting statistics of nine BGWR models and three rBGWR models for the Ontario softwood data.

Method	Model	Prior Mean of $r$	Scale Factor $\delta^2$ (%)	DIC	$D(\beta)$	$pD$
BGWR	1	4	10	4183.1	893.9	3289.2
	2	4	25	4149.6	899.7	3249.9
	3	4	50	4115.8	904.0	3211.8
	4	5	10	3891.4	1039.6	2851.8
	5	5	25	3852.6	1042.2	2810.3
	6	5	50	3826.7	1044.2	2782.5
	7	6	10	4516.4	1160.0	3356.4
	8	6	25	4500.0	1160.3	3339.7
	9	6	50	3751.1	1161.1	2590.1
rBGWR	1	4	-	4620.2	1279.2	3341.1
	2	5	-	4193.0	1307.0	2886.0
	3	6	-	5273.4	1323.7	3949.7

**Tab. 4** - Summary of local regression coefficients by the three modeling methods (Ontario softwood data).

Coefficient	Method	n	Mean	Median	SD	Skewness	Kurtosis	Min	Max
Intercept $\beta_0$	GWR	1698	-2.258	-2.293	1.245	0.541	4.845	-6.087	3.672
	BGWR	1698	-2.250	-2.225	1.407	0.231	4.877	-8.395	5.098
	rBGWR	1698	-2.242	-2.212	1.480	0.227	4.942	-8.665	5.384
Slope $\beta_1$	GWR	1698	1.426	1.440	0.424	-0.596	4.845	-0.736	2.774
	BGWR	1698	1.428	1.425	0.482	-0.252	4.877	-1.113	3.485
	rBGWR	1698	1.426	1.421	0.507	-0.255	4.942	-1.216	3.576

ting, were very similar across the three priors of scale factors  $\delta^2$  within a particular prior of  $r$ . However, the model complexity (pD) decreased as the prior of scale factor  $\delta^2$  increased. Thus, the smallest DIC values were observed for the BGWR models at the scale factor  $\delta^2 = 50\%$  across the three priors of  $r$ . This finding implies that the observed tree data favored a larger dispersion of local regression coefficients, and consequently it indicates the evidence of spatial heterogeneity in the sample data.

Interestingly, the BGWR and rBGWR models had consistently smaller DIC values for  $r = 5$ , which was consistent with the finding previously reported by LeSage (2004). Therefore, the BGWR models with  $r = 5$  and  $\delta^2 = 50\%$  and the rBGWR models with  $r = 5$  had the smallest DIC value, which were chosen to compare with the GWR model. While the model fitting statistics similar to  $D(\beta)$  and pD were also available for GWR, these values were not directly comparable (Waller et al. 2007). Thus, they were not reported for GWR. The coefficient of determination for the GWR model was 0.7824.

Tab. 4 compares the local regression coefficients estimated by GWR with the posterior means of local coefficients estimated by BGWR and rBGWR. It appears that the three modeling methods produced very similar mean and median values for the two coefficients ( $\beta_0$  and  $\beta_1$ ). But the rBGWR model yielded the largest variation for the two coefficients (the largest standard deviations and ranges), followed by the BGWR

model, while the GWR had much smaller variation for the two coefficients (Tab. 4). The distributions of the intercept coefficient ( $\beta_0$ ) were positively skewed and the distributions of the slope coefficient ( $\beta_1$ ) were negatively skewed. Both were leptokurtic (kurtosis  $> 4.8$ ) for the three modeling methods.

#### Model prediction

Fig. 3 illustrates the average prediction errors [= observed  $\ln(\text{crown})$  - predicted  $\ln(\text{crown})$ ] from the three modeling methods across a range of tree size (dbh) classes. It is clear that rBGWR produced the most accurate prediction for the response variable, followed by BGWR and GWR. On the other hand, the trend in prediction errors was similar for the three modeling methods across tree sizes (dbh). All yielded larger prediction errors for the dbh classes from 18 to 27 cm.

#### Assessment of coefficient heterogeneity

The two local regression coefficients were used to generate contour maps across the study area for the three modeling methods (Fig. 4). In general, the contour maps show that GWR produced more “hot” or “cold” spots of the two coefficients estimated than those obtained by BGWR and rBGWR. The spatial pattern and trend of the two coefficients were similar between BGWR and rBGWR.

Fig. 5 shows the intra-block  $V_{\text{intra}}$  (eqn. 14) and inter-block  $V_{\text{inter}}$  (eqn. 15) variances of

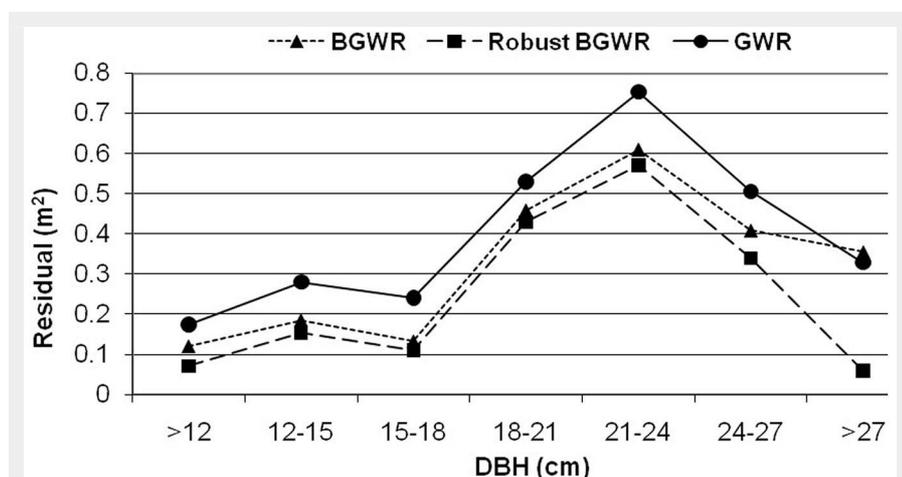
the two coefficients ( $\beta_0$  and  $\beta_1$ ) from the three modeling methods. In general,  $V_{\text{inter}}$  indicates the regional spatial variability, while  $V_{\text{intra}}$  quantifies the local spatial variability. When the block size increases,  $V_{\text{intra}}$  increases and  $V_{\text{inter}}$  decreases. Fig. 5 indicates the three modeling methods yielded similar inter-block variances (regional spatial variability) for the two coefficients across different block sizes. In contrast, rBGWR produces the largest localized spatial variability for both coefficients (i.e., the largest  $V_{\text{intra}}$ ), followed by BGWR, while GWR had the smallest  $V_{\text{intra}}$  for both coefficients, implying that the two regression coefficients were more similar within blocks (probably appeared as “hot” or “cold” spots in Fig. 4).

#### Simulated data

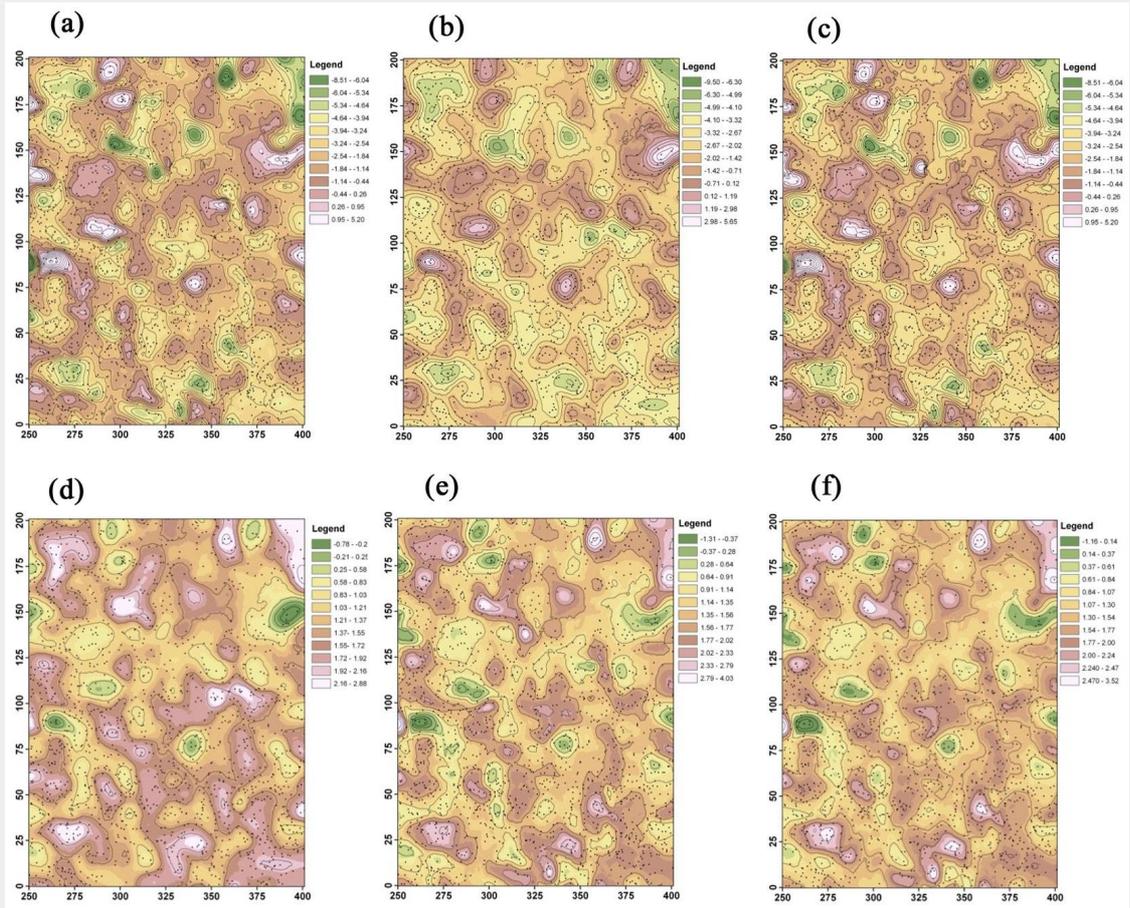
##### Model fitting

Tab. 5 shows the model fitting statistics and the range of the heteroscedastic variance  $V(s_i)$  by BGWR (with  $r = 4, 5$  and 6, and  $\delta^2 = 50\%$  only) and rBGWR (with  $r = 4, 5$  and 6). It seems that for BGWR increasing the prior of  $r$  increased the model deviance  $D(\beta)$ , but the model complexity (pD) decreased. In the case of rBGWR, as the prior mean for  $r$  increased, the model deviance increased, but the model complexity remained flat. However, the DIC indicated both BGWR and rBGWR fitted the data better with smaller prior of  $r$ . Because the simulated data had outliers, the choice of smaller mean prior on  $r$  allowed for larger variation in  $V(s_i)$  and consequently improved the model fit (Geweke 1993, LeSage 1997).

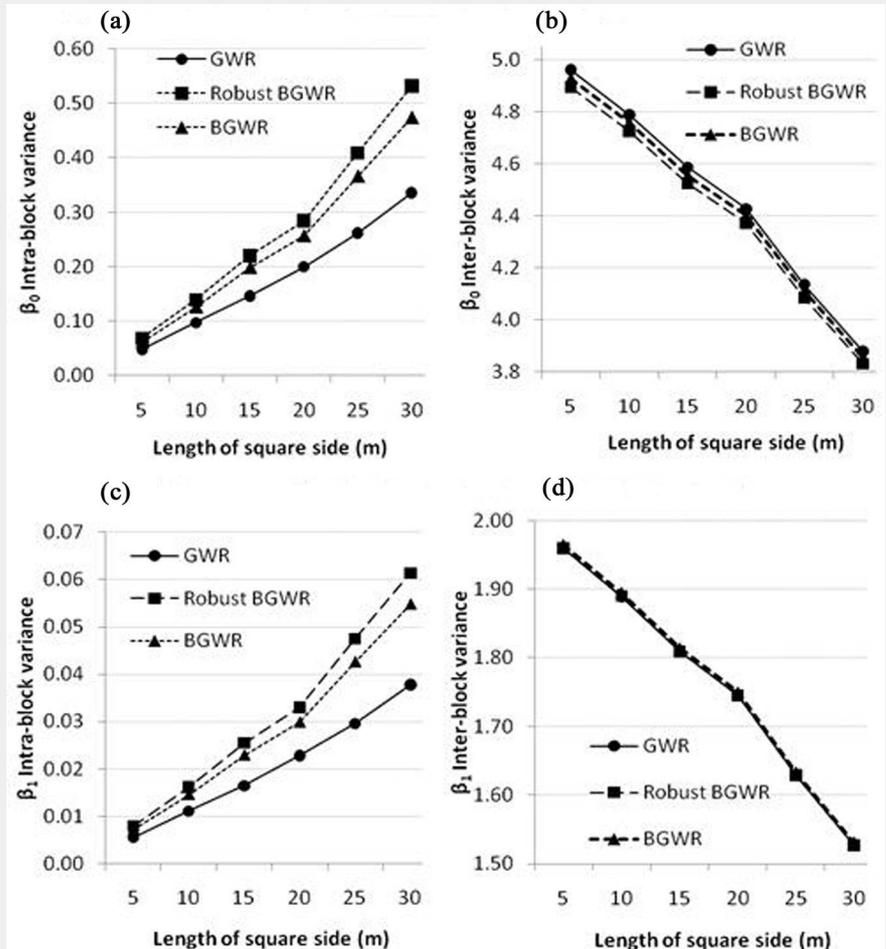
Similar to the results obtained for the observed data, the posterior mean of  $V(s_i)$  was sensitive to the priors of  $r$  for rBGWR, but not so sensitive for BGWR (Tab. 5). Based on the model fitting statistics, the BGWR and rBGWR models with  $r = 4$  were chosen to compare with GWR. As mentioned earlier, the DIC value was not calculated for GWR. However, the coefficient of determination of the GWR model was 0.75 for fitting the simulated data. It was not surprising that the coefficient of determinations of GWR were very similar for the observed and simulated data since the two datasets had similar descriptive statistics (see Tab. 1 and Tab. 2) and features (e.g., cluster pattern, large variation in dbh). The coefficient of determination of GWR for

**Fig. 3** - Average model prediction errors by GWR, BGWR, and rBGWR (Ontario softwood data).

**Fig. 4** - Contour maps of local regression coefficients. (a) GWR -  $\beta_0$ ; (b) BGWR -  $\beta_0$ ; (c) rBGWR -  $\beta_0$ ; (d) GWR -  $\beta_1$ ; (e) BGWR -  $\beta_1$ ; and (f) rBGWR -  $\beta_1$ .



**Fig. 5** - Intra-block variances ( $V_{intra}$ ) and inter-block variances ( $V_{inter}$ ) of local regression coefficients from the three modeling methods. (a) Intra-block -  $\beta_0$ ; (b) Inter-block -  $\beta_0$ ; (c) Intra-block -  $\beta_1$ ; (d) Inter-block -  $\beta_1$ .



**Tab. 5** - Model fitting statistics for BGWR and rBGWR to the simulated data. BGWR models were fitted with scale factor = 50%.

Method	Model	Prior Mean of $r$	DIC	D( $\beta$ )	pD	V( $s_i$ )		
						Min	Max	STD
BGWR	1	4	1567.24	632.04	935.21	0.9949	1.0212	0.0022
	2	5	1587.61	699.24	888.36	0.9957	1.0204	0.0019
	3	6	1651.40	750.07	875.07	0.9964	1.0196	0.0017
rBGWR	4	4	1550.99	746.42	804.57	1.0400	30.8000	0.8891
	5	5	1568.63	764.23	804.40	1.0300	14.7800	0.4192
	6	6	1583.89	776.21	807.68	1.0000	8.3100	0.2276

**Tab. 6** - Summary of local regression coefficients from GWR, BGWR and rBGWR methods for the simulated data. Q1, Q3, P2.5 and P97.5 are lower quartile, upper quartile, 2.5<sup>th</sup> percentile, and 97.5<sup>th</sup> percentile, respectively. (SD): standard deviation.

Coefficient	Method	n	Mean	Median	Q1	Q3	P2.5	P97.5	Min	Max	SD
Intercept $\beta_0$	Simulated	1230	0.201	0.198	0.096	0.312	-0.08	0.478	-0.260	0.58	0.143
	GWR	1230	0.231	0.265	-0.124	0.685	-1.526	1.683	-7.656	4.801	0.917
	BGWR	1230	0.300	0.302	-0.131	0.788	-1.436	2.002	-7.702	3.303	0.847
	rBGWR	1230	0.298	0.302	-0.16	0.812	-1.456	2.035	-7.805	3.462	0.87
Slope $\beta_1$	Simulated	1230	0.811	0.800	0.696	0.903	0.610	1.079	0.600	2.248	0.154
	GWR	1230	0.797	0.781	0.622	0.939	0.278	1.418	-0.448	3.613	0.334
	BGWR	1230	0.766	0.765	0.583	0.941	0.158	1.401	-0.349	3.272	0.313
	rBGWR	1230	0.766	0.764	0.576	0.948	0.143	1.423	-0.430	3.304	0.322

simulated data (0.75) was slightly lower than that for observed data (0.78) because simulated data have more outliers than observed data.

A summary of descriptive statistics on the local regression coefficients estimated by the three modeling methods are given in Tab. 6, along with the summary for the simulated “true or known” coefficients for each location. It appears that the posterior mean and median of the slope coefficient ( $\beta_1$ ) estimated by the three modeling methods are very similar from each other, as

well as similar to the “known  $\beta_1$ ”. On the other hand, the posterior mean and median of the intercept coefficient ( $\beta_0$ ) were larger than the “known  $\beta_0$ ”. In addition, the estimated two local coefficients by the three modeling methods had much larger standard deviations and ranges than the “known” simulated coefficients (Tab. 6). The reasons were that the simulated “true or known” coefficients were individually generated using the bivariate Gaussian copulas with a given measure of association between the two coefficients  $\beta_0$  and  $\beta_1$ ,

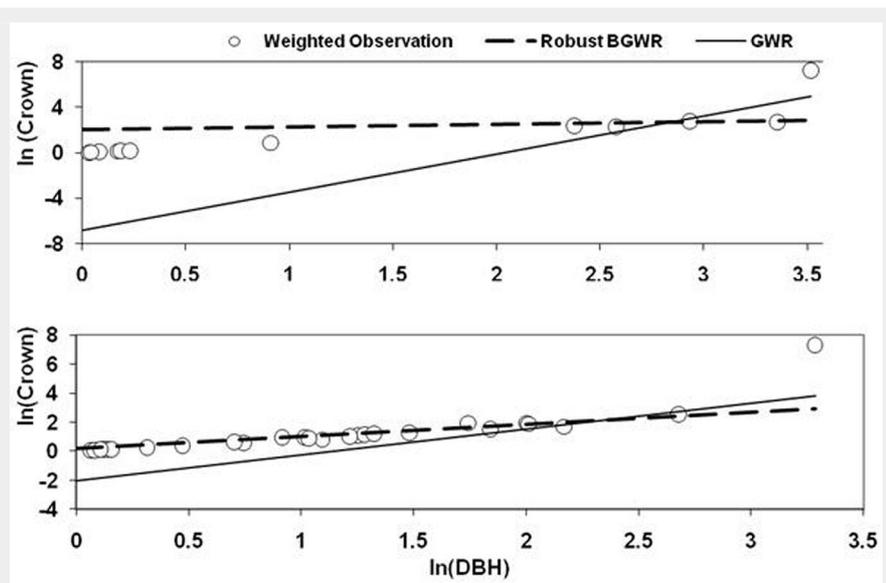
while the local coefficients estimated by the three modeling methods were based on the response variable  $\ln(\text{crown})$  and explanatory variable  $\ln(\text{dbh})$  of the neighboring trees for each subject location. Hence, larger variances in the estimated local coefficients by the three modeling methods were not surprising.

**Accuracy and bias in estimates of local coefficients**

The 95% credible limits of each model coefficient for both BGWR and rBGWR were used to assess the accuracy of estimates of the local model coefficients. If the 95% credible limit of a coefficient for a particular location included the “known” simulated coefficient, the coefficient estimated by that method for that location was declared to be unbiased. The results indicated that, out of 1230 comparisons, 853 (69.3%) of  $\beta_0$  and 903 (73.4%) of  $\beta_1$  by BGWR, and 864 (70.2%) of  $\beta_0$  and 912 (74.1%) of  $\beta_1$  by rBGWR, respectively, were unbiased. When the two coefficients were compared simultaneously, the rBGWR method (839 or 68.2%) was slightly better than the BGWR method (826 or 67.2%).

**Sensitivity to outliers**

To assess the sensitivity of GWR and rBGWR to outliers in the data, two outliers (Tree ID 141 and 541) were selected to compare the local coefficients by the two methods. Fig. 6 shows the scatterplot of  $\ln(\text{crown})$  and  $\ln(\text{dbh})$ , including all observations within the given bandwidth. The regression lines of GWR and rBGWR were generated using coefficients of these methods. It was clear that the rBGWR coefficients were indeed “robust” to the outlier, while the GWR coefficients were



**Fig. 6** - The outlier and neighboring observations within the bandwidth. The two regression lines of GWR and rBGWR were generated using the estimated slope and intercept coefficients. The response variable (crown area) and explanatory variable (dbh) are weighted by proximity to outlier observations and expressed in logarithmic scale.

**Tab. 7** - Simulated and geographically weighted regression (GWR) estimated coefficients, and Bayesian geographically weighted regression (BGWR) and robust BGWR (rBGWR) estimated 95% credible limits of coefficients for weak observations in simulated data. (‡): indicates the 95% of credible limits of BGWR and rBGWR coefficients that do not contain simulated coefficients.

Coefficients	Tree ID	Location		Simulated	GWR	BGWR		rBGWR	
		Easting	Northing			95% Credible limits		95% Credible limits	
Intercept	133	75.1	15.5	-0.023	-0.392	-0.661 <sup>‡</sup>	-0.094 <sup>‡</sup>	-0.553 <sup>‡</sup>	-0.179 <sup>‡</sup>
	174	37.5	21.9	0.014	1.556	-0.544	2.607	-0.438	2.583
	265	59.9	31.8	0.080	-0.708	-6.598	5.564	-4.230	3.138
	620	32.7	74.7	0.020	-0.880	-2.913	1.048	-2.623	0.619
	649	85.2	76.6	0.340	-0.040	-0.189 <sup>‡</sup>	0.132 <sup>‡</sup>	-0.190 <sup>‡</sup>	0.130 <sup>‡</sup>
	924	142.6	23.1	0.135	-0.120	-1.103	0.871	-0.918	0.757
	925	137.5	21.9	0.399	-1.681	-4.007 <sup>‡</sup>	-0.286 <sup>‡</sup>	-4.034 <sup>‡</sup>	-0.484 <sup>‡</sup>
	1132	132.7	74.7	0.119	0.130	-0.802	0.791	-0.738	0.679
Slope	133	75.1	15.5	0.925	1.045	0.945 <sup>‡</sup>	1.136 <sup>‡</sup>	0.974 <sup>‡</sup>	1.096 <sup>‡</sup>
	174	37.5	21.9	0.971	0.341	-0.082	1.191	-0.070	1.153
	265	59.9	31.8	0.951	1.235	-1.083	3.409	-0.174	2.543
	620	32.7	74.7	1.034	1.301	0.653	1.995	0.827	1.903
	649	85.2	76.6	0.700	0.864	0.793 <sup>‡</sup>	0.929 <sup>‡</sup>	0.792 <sup>‡</sup>	0.930 <sup>‡</sup>
	924	142.6	23.1	0.853	0.941	0.584	1.313	0.620	1.248
	925	137.5	21.9	0.649	1.496	0.931 <sup>‡</sup>	2.433 <sup>‡</sup>	1.008 <sup>‡</sup>	2.444 <sup>‡</sup>
	1132	132.7	74.7	0.837	0.824	0.601	1.140	0.650	1.119

strongly affected by the outlier that pulled the regression line of GWR up and resulted in smaller slope coefficient and larger and negative intercept coefficient. This result was expected as GWR uses weighted least square method with constant variance to estimate the local coefficients. In contrast, the rBGWR method assumes heterogeneous variances and gave larger variance to the outlier. In addition, rBGWR used  $W(s_i)/V(s_i)$  as the spatial weight for estimating the coefficients. For example, rBGWR estimates  $V(s_i)$  were 6.87 for Tree ID 141 and 7.68 for Tree ID 541, respectively. Due to the larger variances, the spatial weights of the outliers were diminished heavily to 0.145 for Tree ID 141 and 0.13 for Tree ID 541, respectively.

#### Weak data scenario

Tab. 7 showed the simulated coefficients and the GWR local coefficients, and the 95% credible limits of the two coefficients by BGWR and rBGWR for eight selected weak data observations. BGWR and rBGWR produced unbiased coefficients (i.e., the simulated coefficients were within the 95% credible limits of estimated coefficients) for five (5) observations among the eight (8) weak observations compared. The rBGWR method produced the smallest prediction error for these weak observations, followed by BGWR, and GWR (results not shown), implying that rBGWR and BGWR are more accurate in estimating the local coefficients and predicting the response variable than GWR when the weak data problems exists.

#### Discussion

In this study, we established a relationship between tree crown and dbh for observed and simulated data using GWR, BGWR and rBGWR, and demonstrated the

advantages of Bayesian approach to improve the model fitting and performance of GWR. Firstly, BGWR and rBGWR not only provided better estimation of local regression coefficients, but also correct estimation of the variances of the coefficients because Bayesian procedures (e.g., Gibbs sampler) are not affected by the lack of independence in the sample data. Thus, correct statistical inference and credible limits (or confidence intervals) can be computed and used for statistical testing and model interpretation. The average prediction errors yielded by BGWR and rBGWR were only about 50% of that produced by GWR across all DBH classes. Secondly, the estimates of local coefficients by rBGWR were not affected by potential outliers, while these estimates by GWR were “contaminated” by the outliers (e.g., producing smaller slope and larger and negative intercept). It is not uncommon that forestry and/or ecological data have potential outliers or unusual observations. If the coefficient estimates of a model are sensitive to the outliers, it could have serious implication for interpreting the meaning of the coefficients as well as the regression relationships between variables. BGWR and rBGWR can automatically detect and down-weight the outliers' influence in the parameter estimation processes. Furthermore, when the spatial pattern of trees or plants is clustered, there are good chances that the number of trees could be sparse in some locations, resulting in very small number of effective observations that can be used to estimate the model parameters (i.e., the weak data problem). BGWR and rBGWR are able to produce unbiased estimation on model parameters and prediction by incorporating prior information, such as expert opinion or existing knowledge, to impose restrictions on the spatial

nature of parameter variation (LeSage 2004). According to our results, rBGWR produced the smallest prediction error in weak data scenario, followed by BGWR and GWR. When large spatial data sets are involved in data analysis and modeling, heterogeneous variance may be a serious issue for model parameter estimation. BGWR combines the distance-based spatial weight matrix, non-constant variance in model error term and any prior information into a parameter smoothing specification as a unified parameter estimation approach to overcome the limitations of GWR (LeSage 2004, Wheeler & Páez 2010).

However, the BGWR methods need to be used with caution. Given their nature, the local coefficients estimated by the BGWR methods may not be used to make prediction outside the study area. Attention should be given to assess potential model misspecification, because the spatial heterogeneities of the relationships between variables might have been produced as a result of model misspecification (Cressie 1993, Fotheringham et al. 2002). The assessment of residuals may be useful for identifying model misspecification (Zhang & Gove 2005).

One of the challenges on modeling spatial effects (i.e., spatial autocorrelation and heterogeneity) is to identify an appropriate spatial scale (Fortin & Dale 2005). Changes in the bandwidth or number of neighbors will result in change of support (Gotway & Young 2002), and consequently will produce different model fitting and parameter estimates (Guo et al. 2008). Larger bandwidth will delineate a broader region as neighborhood and will smooth the spatial heterogeneity rather too much, while smaller bandwidth will define a limited area as neighborhood and will produce too “spiky” spatial heterogeneity (Wiens 1989,

Bailey & Gatrell 1995). For the choice of scale (in this study bandwidth and number of neighbors), Wagner & Fortin (2005) suggested that the scale of analysis needs to be determined in an ecologically meaningful and methodologically sound way.

Another challenge to the BGWR methods is related to the convergence of parameters. In this study, longer MCMC chains were used to sample the posterior distributions of local coefficients. While sampling posterior of coefficients for nine BGWR models (with a combination of the priors on variance and scale factor) and for three rBGWR models (with three priors on variance), the slope and intercept coefficients estimated by these models converged at the desired level of precision (quartile = 0.025, precision = 0.005 and probability = 0.95) for most of the observations within the 4,000 runs after the few thousands initial burn-in and storing one sample out of 10 samples. However, for a few observations, those that fall at the edge of the study area required almost 15,000 iterations (at the maximum for the BGWR method with prior on scale factor at the smallest level) to converge. The assessment of histogram of coefficients and its movement across iterations also suggested the desirability of longer runs for the latter observations. One of the reasons for the convergence problem may be the correlation between the parameters within the parameter set (Congdon 2006). However, the autocorrelation among the posterior sample of coefficients can be reduced by thinning, e.g., keeping one coefficient out of 10 sampled coefficients. In contrast, the BGWR method posterior sample of coefficients with large (the largest level) prior on scale factor converged more quickly than those with small (the smallest level) prior on scale factor. In the case of the rBGWR method, which samples all the posterior samples of coefficients for an observation before moving to the next observation, this method took less time to run. Besides this, the computation time was directly related to the precision level required for parameter convergence; e.g., for smaller precision (quartile = 0.025, precision = 0.01 and probability = 0.95), the estimates of coefficients by rBGWR and BGWR converged within a few thousands of iterations after sufficient burn-in.

Despite these limitations, BGWR methods have been recommended as an effective tool for investigating spatial heterogeneity of covariate effects across a study area. Future studies on developing an optimization method as a function of both bandwidth and covariates are recommended. The current approach of bandwidth optimization (e.g., the CV method) optimizes only bandwidth for a given set of covariates. A better method that optimizes a combination of covariates (similar to forward or backward selection in regression) and bandwidth simultaneously is more desirable. In addition, a window-driven software for param-

eter estimation by the BGWR methods would certainly be an asset for wider adaptations of these Bayesian methods.

## Conclusion

The relationship between tree crown area and diameter at breast height was modeled using GWR, BGWR, and rBGWR models. Both observed and simulated tree data were used to compare model fitting and performance. For the observed Ontario softwood stand, BGWR and rBGWR produced more accurate parameter estimates and better model prediction than GWR. Further, rBGWR produced the largest localized spatial variability for both coefficients, followed by BGWR, while GWR had the smallest localized spatial variability for both coefficients. Thus, the contour maps of local regression coefficients show more “hot” or “cold” spots of the two coefficients estimated by GWR than those obtained by BGWR and rBGWR.

For the simulated data, the percentage of the unbiased parameter estimation for the two local coefficients by rBGWR was slightly higher than those by BGWR. However, the credible limits cannot be computed for GWR due to the lack of an accurate expression of the variance of the model coefficients in GWR. Further, the coefficients estimates using the rBGWR method were not affected by outliers, while these estimates by GWR were “contaminated” by the outliers, resulting in erroneous parameter estimates. Similarly, for the weak data, the majority of the coefficients estimated by the rBGWR and BGWR methods were not different from the simulated “known” coefficients. Therefore, BGWR and rBGWR are better alternatives to overcome the limitations of GWR and are more powerful than GWR to detect the spatial areas with non-constant variance or spatial outliers.

## Acknowledgements

This work was supported by the National Natural Science Foundation of China, proj. no. 31400491: “Three dimensional individual tree crown delineation based on multiple remotely sensed data sources and tree competition mechanism”, and by the Fundamental Research Funds for the Central Universities, proj. no. 2572017CA02.

## References

- Bailey TC, Gatrell AC (1995). Interactive spatial data analysis. Longman, Essex, UK, pp. 413. [online] URL: <http://www.personal.psu.edu/faculty/f/k/fkw/rsoc597/introduction.pdf>
- Banerjee S, Carlin BP, Gelfand AE (2004). Hierarchical modeling and analysis for spatial data. Chapman and Hall/CRC, Boca Raton, FL, USA, pp. 584.
- Brunsdon C, Fotheringham AS, Charlton ME (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis* 28: 281-298. - doi: [10.1111/j.1538-4632.1996.tb00936.x](https://doi.org/10.1111/j.1538-4632.1996.tb00936.x)
- Carlin BP, Louis TA (2000). Bayes and empirical

- Bayes methods for data analysis. Chapman and Hall/CRC, Boca Raton, FL, USA, pp. 440.
- Casella G, George EI (1992). Explaining the Gibbs sampler. *The American Statistician* 46: 167-174. - doi: [10.1080/00031305.1992.10475878](https://doi.org/10.1080/00031305.1992.10475878)
- Clifton EV, Romero-Barrutieta A (2006). Institutions versus geography: subnational evidence from the United States. IMF Working Papers WP/06/169, pp. 1-26. [online] URL: <http://ssrn.com/abstract=926224>
- Congdon P (2006). Bayesian statistical modeling. John Wiley and Sons, New York, USA, pp. 581.
- Cressie N (1993). Statistics for spatial data. John Wiley and Sons, Toronto, Canada, pp. 461.
- Ek AR (1969). Stem map data for three forest stands in northern Ontario. Information Report no. OX-113, Canadian Forest Service Publications, Marie, Canada, pp. 23.
- Farber S, Páez A (2007). A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations. *Journal of Geographical Systems* 9: 371-396. - doi: [10.1007/s10109-007-0051-3](https://doi.org/10.1007/s10109-007-0051-3)
- Foody GM (2004). Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. *Global Ecology and Biogeography* 13: 315-320. - doi: [10.1111/j.1466-822X.2004.00097.x](https://doi.org/10.1111/j.1466-822X.2004.00097.x)
- Fortin MJ, Dale MRT (2005). Spatial analysis: a guide to ecologists. Cambridge University Press, Cambridge, UK, pp. 365.
- Fotheringham AS, Brunsdon C, Charlton M (2002). Geographically weighted regression: the analysis of spatially varying relationships. John Wiley and Sons, West Sussex, UK, pp. 284.
- Fox JC, Ades PK, Bi H (2001). Stochastic structure and individual-tree growth models. *Forest Ecology and Management* 154: 261-276. - doi: [10.1016/S0378-1127\(00\)00632-0](https://doi.org/10.1016/S0378-1127(00)00632-0)
- Furutani T (2004). Bayesian geographically weighted regression model and its application for land price model estimation. *Journal of City Planning Institute of Japan* 39: 787-792.
- Garrigues S, Allard D, Baret F, Weiss M (2006). Quantifying spatial heterogeneity at the landscape scale using variogram models. *Remote Sensing of Environment* 103 (1): 81-96. - doi: [10.1016/j.rse.2006.03.013](https://doi.org/10.1016/j.rse.2006.03.013)
- Gelfand AE (2000). Gibbs sampling. *Journal of the American Statistical Association* 95: 1300-1304. - doi: [10.1080/01621459.2000.10474335](https://doi.org/10.1080/01621459.2000.10474335)
- Gelfand AE, Smith AFM (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85: 398-409. - doi: [10.1080/01621459.1990.10476213](https://doi.org/10.1080/01621459.1990.10476213)
- Geweke J (1993). Bayesian treatment of the independent student-t linear model. *Journal of Applied Econometrics* 8 (S1): S19-S40. - doi: [10.1002/\(ISSN\)1099-1255](https://doi.org/10.1002/(ISSN)1099-1255)
- Gotway CA, Young J (2002). Combining incompatible spatial data. *Journal of the American Statistical Association* 97: 632-648. - doi: [10.1198/016214502760047140](https://doi.org/10.1198/016214502760047140)
- Green JL, Hastings A, Arzberger P, Ayala FJ, Cottingham KL, Cuddington K, Davis F, Dunne JA, Fortin M-J, Gerber L, Neubert M (2005). Complexity in ecology and conservation: mathematical, statistical, and computational challenges.

- Bioscience 55: 501-510. - doi: [10.1641/0006-3568\(2005\)055\[0501:CLEACM\]2.o.CO;2](https://doi.org/10.1641/0006-3568(2005)055[0501:CLEACM]2.o.CO;2)
- Guo L, Ma Z, Zhang L (2008). Comparison of bandwidth selection in application in geographically weighted regression: a case study. *Canadian Journal of Forest Research* 38: 2526-2534. - doi: [10.1139/X08-091](https://doi.org/10.1139/X08-091)
- Haining R (2004). *Spatial data analysis: theory and practice*. Cambridge University Press, Cambridge, UK, pp. 432.
- Legendre P, Fortin MJ (1989). Spatial pattern and ecological analysis. *Plant Ecology* 80: 107-138. - doi: [10.1007/BF00048036](https://doi.org/10.1007/BF00048036)
- Legendre P (1993). Spatial autocorrelation: trouble or new paradigm? *Ecology* 76: 1659-1673. - doi: [10.2307/1939924](https://doi.org/10.2307/1939924)
- LeSage JP (1997). Bayesian estimation of spatial autoregressive models. *International Regional Science Review* 20: 113-129. - doi: [10.1177/016001769702000107](https://doi.org/10.1177/016001769702000107)
- LeSage JP (2004). A family of geographically weighted regression models. In: "Advances in Spatial Econometrics" (Anselin LR, Florax JGM, Rey SJ eds). Springer-Verlag, Berlin, Germany, pp. 240-264. - doi: [10.1007/978-3-662-05617-2\\_11](https://doi.org/10.1007/978-3-662-05617-2_11)
- Levin SA (1992). The problem of pattern and scale in ecology. *Ecology* 73: 1943-1967. - doi: [10.2307/1941447](https://doi.org/10.2307/1941447)
- Li H, Reynolds JF (1995). On definition and quantification of heterogeneity. *Oikos* 73: 280-284. - doi: [10.2307/3545921](https://doi.org/10.2307/3545921)
- Lin CH, Wen TH (2011). Using geographically weighted regression (GWR) to explore spatial varying relationships of immature mosquitoes and human densities with the incidence of dengue. *International Journal of Environmental Research and Public Health* 8: 2798-2815. - doi: [10.3390/ijerph8072798](https://doi.org/10.3390/ijerph8072798)
- Lindley DV (1971). The estimation of many parameters. In: "Foundation of Statistical Inference" (Godambe VP, Sprott DA eds). Holt, Rinehart and Winston, New York, USA, pp. 435-455. - doi: [10.1002/j.2333-8504.1970.tb00411.x](https://doi.org/10.1002/j.2333-8504.1970.tb00411.x)
- Lloyd CD (2006). *Local models for spatial analysis*. CRC Press, Boca Raton, FL, USA, pp. 244.
- Lovett GM, Jones CG, Turner MG, Weathers KC (2005). *Ecosystem function in heterogeneous landscapes*. Springer, New York, USA, pp. 489.
- [online] URL: [http://www.caryinstitute.org/sites/default/files/public/reprints/Lovett\\_et\\_al\\_Ecosystem\\_Function\\_2005.pdf](http://www.caryinstitute.org/sites/default/files/public/reprints/Lovett_et_al_Ecosystem_Function_2005.pdf)
- Lu J, Zhang L (2010). Evaluation of parameter estimation methods for fitting spatial regression models. *Forest Science* 56: 505-514. - doi: [10.17221/127/2010-JFS](https://doi.org/10.17221/127/2010-JFS)
- Lu J, Zhang L (2012). Geographically local linear mixed models for tree height-diameter relationship. *Forest Science* 58: 75-84. - doi: [10.5849/forsci.09-123](https://doi.org/10.5849/forsci.09-123)
- Lu J, Zhang L (2013). Evaluation of structure specification in linear mixed models for modeling the spatial effects in tree height-diameter relationships. *Annals of Forest Research* 56: 137-148.
- Ma Z, Zuckerberg B, Porter WF, Zhang L (2012). Use of localized descriptive statistics for exploring the spatial pattern changes of bird species richness at multiple scales. *Applied Geography* 32 (2): 185-194. - doi: [10.1016/j.apgeog.2011.05.005](https://doi.org/10.1016/j.apgeog.2011.05.005)
- Ma Z, Zhang L (2015). Modeling bird species richness at multiple spatial scales using 2D wavelets analysis. *Forest Science* 61: 1-16. - doi: [10.5849/forsci.11-041](https://doi.org/10.5849/forsci.11-041)
- McMillen DP (2003). Neighbourhood house price indexes in Chicago: a Fourier repeat sales approach. *Journal of Economic Geography* 3: 57-73. - doi: [10.1093/jeg/3.1.57](https://doi.org/10.1093/jeg/3.1.57)
- Meng Q, Cieszewski CJ, Strub MR, Borders BE (2009). Spatial regression modeling of tree height-diameter relationships. *Canadian Journal of Forest Research* 39 (12): 2283-2293. - doi: [10.1139/X09-136](https://doi.org/10.1139/X09-136)
- Nelsen RB (1999). *An introduction to copulas*. Springer, New York, USA, pp. 315.
- Raftery A, Lewis S (1992). How many sampler in the Gibbs sampler? *Bayesian Statistics* 4: 763-774.
- Reed DD, Burkhart HE (1985). Spatial autocorrelation of individual tree characteristics in loblolly pine stands. *Forest Science* 31: 575-587. [online] URL: <http://academic.oup.com/forestscience/article-abstract/31/3/575/4656995>
- Schabenberger O, Gotway CA (2005). *Statistical methods for spatial data analysis*. Chapman and Hall/CRC Press, Boca Raton, FL, USA, pp. 488.
- Spiegelhalter DJ, Best NG, Carlin Pet B (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society* 64: 583-639. - doi: [10.1111/1467-9868.00353](https://doi.org/10.1111/1467-9868.00353)
- Valentine HT, Herman DA, Gove JH, Hollinger DY, Solomon DS (2000). Initializing a model stand for process-based projection. *Tree Physiology* 20 (5-6): 393-398. - doi: [10.1093/treephys/20.5-6.393](https://doi.org/10.1093/treephys/20.5-6.393)
- Wagner HH, Fortin MJ (2005). Spatial analysis of landscapes: concepts and statistics. *Ecology* 86: 1975-1987. - doi: [10.1890/04-0914](https://doi.org/10.1890/04-0914)
- Waller LA, Zhu L, Gotway CA, Gorman DM, Grunewald PJ (2007). Quantifying geographic variations in associations between alcohol distribution and violence: a comparison of geographically weighted regression and spatially varying coefficient models. *Stochastic Environmental Research and Risk Assessment* 21 (5): 573-588. - doi: [10.1007/s00477-007-0139-9](https://doi.org/10.1007/s00477-007-0139-9)
- Wang M, Upadhyay A, Zhang L (2010). Trivariate distribution modeling of tree diameter, height, and volume. *Forest Science* 56: 290-300. [online] URL: <http://academic.oup.com/forestscience/article-abstract/56/3/290/4604468>
- Wheeler DC, Páez A (2010). Geographically weighted regression. In: "Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications" (Fisher MM, Getis A eds). Springer, Heidelberg, Germany, pp. 461-486. - doi: [10.1007/978-3-642-03647-7\\_22](https://doi.org/10.1007/978-3-642-03647-7_22)
- Wiens JA (1989). Spatial scaling in ecology. *Functional Ecology* 3: 385-397. - doi: [10.2307/2389612](https://doi.org/10.2307/2389612)
- Zhang L, Gove JH (2005). Spatial assessment of model errors from four regression techniques. *Forest Science* 51: 334-346.
- Zhang L, Shi H (2004). Local modeling of tree growth by geographically weighted regression. *Forest Science* 50: 225-244.
- Zhang L, Ma Z, Guo L (2008). Spatially assessing model errors of four regression techniques for three types of forest stands. *Forestry* 81: 209-225. - doi: [10.1093/forestry/cpn014](https://doi.org/10.1093/forestry/cpn014)
- Zhang L, Ma Z, Guo L (2009). Spatial autocorrelation and heterogeneity in the relationships between tree variables. *Forest Science* 55: 533-548.