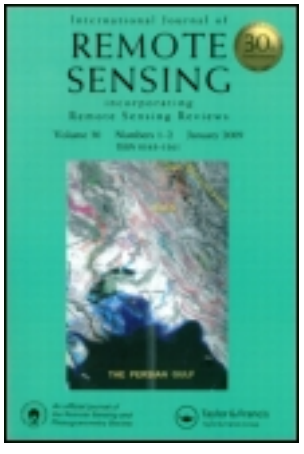

International Journal of Remote Sensing

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tres20

The influence of sampling density on geographically weighted regression: a case study using forest canopy height and optical data

Gang Chen [a], Kaiguang Zhao [b], Gregory J. McDermid [a] & Geoffrey J. Hay [a]

[a] Foothills Facility for Remote Sensing and GIScience, Department of Geography, University of Calgary, Calgary, AB, Canada, T2N 1N4

[b] Department of Biology & Center on Global Change, Duke University, Durham, NC, 27708, USA

Available online: 19 Oct 2011

To cite this article: Gang Chen, Kaiguang Zhao, Gregory J. McDermid & Geoffrey J. Hay (2012): The influence of sampling density on geographically weighted regression: a case study using forest canopy height and optical data, International Journal of Remote Sensing, 33:9, 2909-2924

To link to this article: http://dx.doi.org/10.1080/01431161.2011.624130

International Journal of Remote Sensing
Vol. 33, No. 9, 10 May 2012, 2909–2924

Taylor & Francis
Taylor & Francis Group

The influence of sampling density on geographically weighted regression: a case study using forest canopy height and optical data

GANG CHEN*†, KAIGUANG ZHAO‡, GREGORY J. McDERMID† and GEOFFREY J. HAY†

†Foothills Facility for Remote Sensing and GIScience, Department of Geography, University of Calgary, Calgary, AB, Canada T2N 1N4

‡Department of Biology & Center on Global Change, Duke University, Durham, NC 27708, USA

(*Received 11 October 2010; in final form 1 July 2011*)

Geographically weighted regression (GWR) extends the conventional *ordinary least squares* (OLS) regression technique by considering spatial nonstationarity in variable relationships and allowing the use of spatially varying coefficients in linear models. Previous forest studies have demonstrated the better performance of GWR compared to OLS when calibrated and validated at sampled locations where field measurements are collected. However, the use of GWR for remote-sensing applications requires generating estimates and evaluating the model performance for the large image scene, not just for sampled locations. In this study, we introduce GWR to estimate forest canopy height using high spatial resolution Quickbird (QB) imagery and evaluate the influence of sampling density on GWR. We also examine four commonly used spatial analysis techniques – OLS, *inverse distance weighting* (IDW), *ordinary kriging* (OK) and *cokriging* (COK) – and compare their performance with that using GWR. Results show that (i) GWR outperformed OLS at all sampling densities; however, they produced similar results at low sampling densities, suggesting that GWR may not produce significantly better results than OLS in remote-sensing operational applications where only a small number of field data are collected. (ii) The performance of GWR was better than those of IDW, OK and COK at most sampling densities. Among the spatial interpolation techniques we examined, IDW was the best to estimate the canopy height at most densities, while COK outperformed OK only marginally and produced larger canopy height estimation errors than both IDW and GWR. (iii) GWR had the advantage of generating canopy height estimation maps with more accurate estimates than OLS, and it preserved patterns of geographic features better than IDW, OK or COK.

1. Introduction

Statistical regression models that predict the value of continuously varying biophysical properties represent one of remote sensing's most flexible and widely employed family of analytical techniques (e.g. Franklin 1995, Lefsky *et al*. 1999, Cohen *et al*. 2001). However, the traditional *ordinary least squares* (OLS) regression assumes that the relationship between the dependent and independent variables remains constant across geographical space and thereby exhibits 'stationarity' with respect to model parameters (Fotheringham *et al*. 1998). Unfortunately, this assumption is commonly

*Corresponding author. Email: gangchen@ucalgary.ca

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2012 Taylor & Francis
http://www.tandf.co.uk/journals
http://dx.doi.org/10.1080/01431161.2011.624130

violated in geographic applications, where the processes often display significant patterns of spatial variability. For example, forest growth rate may be strongly linked to elevation on a single mountain slope because high elevations shorten growing seasons, but such a linkage can fail across a more extensive study area when other factors (e.g. the aspect and influence of the surrounding mountain ranges) exert more important controls on microclimatic environments (Kimsey *et al*. 2008). To accommodate such 'nonstationarity' in spatial regression problems, Fotheringham *et al*. (2002) introduced a technique called *geographically weighted regression* (GWR). In geographic applications, GWR has a major advantage over OLS as it incorporates local spatial relationships (Fotheringham *et al*. 2002).

Recognizing the potential value of GWR, the remote-sensing community has successfully experimented with the technique in a variety of forest applications, including species richness and composite mapping (Foody 2004, 2005), net primary production (NPP) estimation (Wang *et al*. 2005), crown closure prediction (Man Shrestha 2006), leaf area index retrieval (Propastin 2009), Douglas-fir site index analysis (Kimsey *et al*. 2008) and tree diameter modelling (Salas *et al*. 2010). However, a few studies have examined the performance of GWR to estimate the canopy height, an important biophysical parameter to measure carbon sequestration in forests.

In this study, our first objective was to introduce GWR as a substitute for OLS for retrieving the canopy height using high spatial resolution optical remote-sensing imagery and small footprint lidar (light detection and ranging) data. The GWR models were compared with another four commonly used spatial analysis techniques – OLS, *inverse distance weighting* (IDW), *ordinary kriging* (OK) and *cokriging* (COK). Although previous works on GWR models have been promising, the technique typically requires large amounts of calibration and validation data that may pose challenges in operational applications (Salas *et al*. 2010). Thus, our second objective was to evaluate the influence of sampling density on GWR.

2. Study area and data

2.1 *Study area*

Our study site (49° 52′ N, 125° 20′ W) was located approximately 10 km southwest of Campbell River on Vancouver Island, British Columbia, Canada (figure 1). The size of the study area is 5.1 km × 5.1 km (2601 ha) and is characterized by conifer and deciduous forests, clearcuts, roads and a river. About 65% of the study area is covered by conifer forests, the bulk of which is dominated by Douglas-fir (*Pseudotsuga menziesii*) trees, along with small proportions of western red cedar (*Thuja plicata*) and western hemlock (*Tsuga heterophylla*) (Morgenstern *et al.* 2004). Another 16% of the study area contains red alder (*Alnus rubra*) while the remainder of the site comprises clearcuts, roads and a river that diagonally bisects the site from southeast to northwest. Topographically, the area is relatively flat, with a gentle slope of 5°–10°. The elevations range from 180 m (southwest) to 440 m (northeast) above sea level.

2.2 *Lidar data*

We acquired lidar data with an airborne terrain scanning lidar system (Terra Remote Sensing Inc., Sidney, BC, Canada) on 8 June 2004. The system is a discrete return laser scanner (Lightwave Model 110) with a pulse repetition frequency of 10 kHz, a wavelength of 1047 nm, a field of view of 56° and a beam divergence of 3.5 mrad.

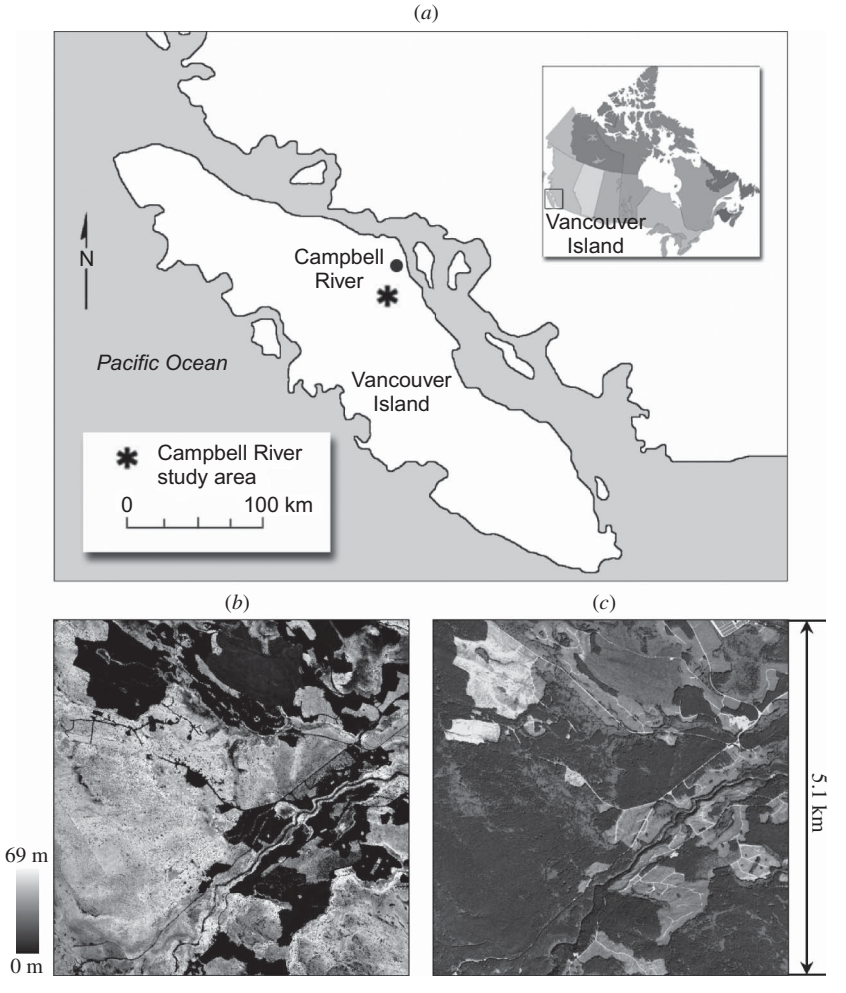

Figure 1. (*a*) Study area located southwest of Campbell River, Vancouver Island, Canada. (*b*) Lidar canopy height model (CHM) of the study area. (*c*) Quickbird greyscale image over the corresponding area converted from a false colour composite using near infrared (NIR), red and green bands.

Continuous-scanning mode in a typical zigzag pattern was used during data acquisition with a footprint size of 0.19 m and an average point density of 0.7 m^{-2} (Hilker *et al*. 2008). The point cloud was transformed by the vendor to a raster-based digital elevation model (DEM) and a digital surface model (DSM) with 1.0 m spatial resolution. We derived a forest canopy height model (CHM) by systematically subtracting the DEM from the DSM.

2.3 *Quickbird data*

A cloud-free Quickbird (QB) image of the study site was acquired on 11 August 2004. Four multispectral bands (i.e. blue, green, red and near infrared (NIR)) and one panchromatic band were used in this study.

We used a principal components spectral sharpening technique (Welch and Ahlers 1987) to combine the spectral information from the QB multispectral bands and the spatial information from the QB panchromatic band. The lidar and pan-sharpened optical data were then geometrically co-registered using the lidar CHM as the base image. We used a second-order polynomial for spatial interpolation with a total of 118 ground control points (root mean square error (RMSE): 0.85 m). The pan-sharpened QB image was resampled to match the spatial resolution of the lidar imagery (1.0 m) using the nearest neighbour digital number (DN) interpolation method.

3. Methods

3.1 *Synthesizing data for model evaluation*

3.1.1 Data sampling. We used a systematic sampling method (McCoy 2005) to assign sampling points to positions at equidistant intervals. There were two main reasons for using systematic sampling: (i) one of our primary emphases was on the influence of the sampling density – not sampling methods – on model performance; and (ii) compared to other sampling methods, systematic sampling was easy to implement and produced consistent samples at different sampling densities. All height measurements were extracted across consistent 20 m × 20 m sample plots.

Figure 2 shows the five different data sampling densities, using plot intervals of 460, 260, 160, 60 and 20 m, representing a density gradient from very low to very high. For illustrative purposes, a histogram-adjusted lidar CHM was used as the base layer, with the overlaid white dots representing sampled plots.

It should be noted that the 460 m large interval was chosen to simulate an operational remote-sensing activity of forest research or management, wherein a small

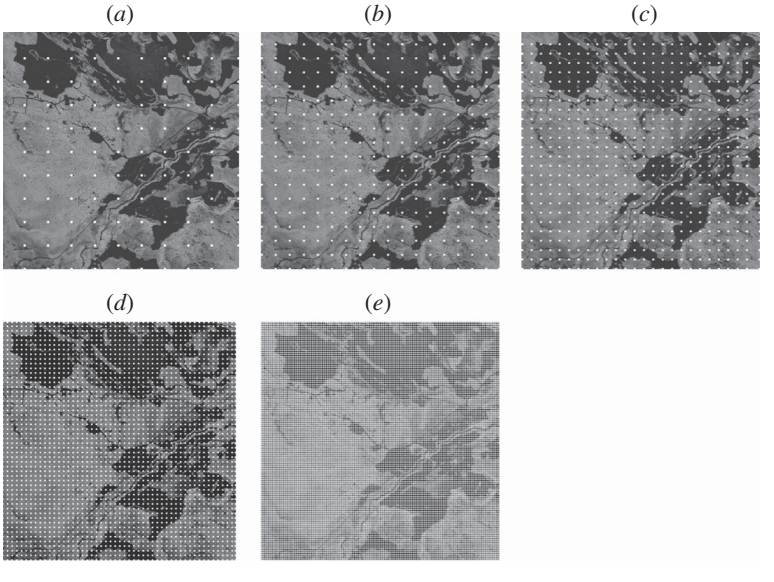

Figure 2. Five data sampling densities using plot intervals of (*a*) 460 m, (*b*) 260 m, (*c*) 160 m, (*d*) 60 m and (*e*) 20 m. For illustrative purposes, histogram-adjusted lidar CHM was used as the backdrop, with the overlaid white dots representing sampled plots.

Table 1. Descriptive statistics of canopy height (from lidar data) for the five sampling densities.

Sampling density (plot interval)	Mean (m)	Standard deviation (m)	Min (m)	Max (m)	Moran's *I*	Spatial heterogeneity, SH (%)
'1' (460 m)	20.9	12.0	0.0	36.6	0.27	46.06
'2' (260 m)	20.7	12.2	0.0	44.7	0.33	53.45
'3' (160 m)	20.8	12.6	0.0	45.1	0.43	51.37
'4' (60 m)	20.8	12.7	0.0	49.7	0.56	54.27
'5' (20 m)	20.5	12.8	0.0	53.0	0.74	57.66

portion of the forest (i.e. 110 plots) of the 2600 ha study area was measured. The 20 m small interval possibly represents the highest sampling density on this site as our plot size is 20 m. The intervals of 260, 160 and 60 m were selected to show the differences in estimating canopy height using various types of models. Based on our initial evaluation, the three densities are sufficient to represent this difference. Descriptive statistics of the canopy height for the five sampling densities are listed in table 1.

It appears that the canopy height mean and standard deviation values remain similar for different plot intervals (table 1). This confirms that the systematic sampling method works in this area. However, plots with smaller intervals have higher chances to cover few tall canopies. To quantify the spatial autocorrelation of the sampled plots, we calculated Moran's *I* by following Goodchild (1986). The positive Moran's *I* means that all samples within the distance tend to be similar. This value is negative when samples are dissimilar. The value close to zero indicates the independence between the samples. The feature of heterogeneity – spatial heterogeneity (SH) – was calculated by following Li and Reynolds (1995), using semivariograms. Large SH values indicate high levels of heterogeneity or low levels of spatial randomness (Zhang *et al*. 2009) and vice versa. Table 1 shows that the samples of canopy height using small plot intervals have stronger spatial autocorrelations and higher SH than those using large ones, although we note that plots with the smallest interval (i.e. 460 m) still exhibit relatively strong spatial autocorrelation (Moran's $I = 0.27$) and high (46.06%).

3.1.2 Extraction of variables. Previous studies have demonstrated the potential of using high spatial resolution (<5.0 m) optical imagery to estimate the forest canopy height (Franklin and McDermid 1993, Donoghue and Watt 2006, Hyde *et al.* 2006, Chen *et al.* 2011). In this research, we extracted two types of QB-derived independent variables: (i) *spectral mean* – the mean of the DNs within the plots for each spectral band (i.e. blue, red, green and NIR); and (ii) *image texture* – the measure of the standard deviation of DNs within a plot for each spectral band. Since small footprint lidar provides highly accurate estimates of a forest's vertical structure (Means *et al.* 1999, Lim *et al.* 2003, Zhao and Popescu 2009), the spatially explicit layer of lidar-derived canopy height was treated as 'pseudo' ground-truth data (Zhao *et al.* 2009).

This also facilitated the field measurement with high sampling densities, for example, 65 536 plots at the 20 m interval. We extracted the canopy height – the dependent variable – for each plot from the lidar CHM by the following two steps. First, all tree tops were extracted through local-maximum (LM) filtering within a fixed 3×3 pixel window. Previous work by Wulder *et al.* (2000) has shown that this approach is

effective when applied to a 1.0 m resolution image over a similar study area. Second, we took a mean of the height values of all tree tops within each plot.

The *minimal-redundancy-maximal-relevance* (mRMR) variable selection approach was used to build models that minimized multicollinearity amongst the independent variables. mRMR is a machine-learning technique that selects variables with the minimal similarity between each other and the maximal relevance with the target class (i.e. canopy height in our case) and has proven to be effective when evaluated with different classifiers (e.g. Bayes and support vector machines) and a variety of data sets (e.g. handwritten digits and cancer cell data) (Peng *et al.* 2005). The mRMR implementation was obtained from open-source code; more technical details can be found in Peng *et al.* (2005). Each of the final models was composed of the top three QB independent variables. We assessed separately for each of the five sampling densities, but they all produced the same variable subset, including the spectral means for the red and NIR bands and the image-texture measure for the NIR band.

3.2 *Models*

3.2.1 Ordinary least squares. In remote-sensing studies, the OLS has been widely used as a standard linear regression procedure. Consider the training samples as $(\boldsymbol{x}_i, y_i)$, $i = 1, \ldots, n$, where $\boldsymbol{x}_i$ is an independent p-dimensional vector $[x_{i1}, \ldots, x_{ip}]$, y_i is a scalar dependent variable and n is the number of training samples. The relationship between $\boldsymbol{x}_i$ and y_i is formulated as

$$y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + \varepsilon_i, \tag{1}$$

where β_0 is the intercept, β_j is the slope for x_{ij} and ε_i is an error term. The estimated model coefficients are calculated using the least-squares method:

$$\hat{\boldsymbol{\beta}} = (\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathrm{T}}\boldsymbol{Y}, \tag{2}$$

where $\hat{\boldsymbol{\beta}}$ denotes the estimate of a $(p + 1) \times 1$ vector of model coefficients (β_0, $\beta_1, \ldots, \beta_p$), $\boldsymbol{Y}$ denotes an $n \times 1$ vector of dependent variables, $\mathbf{x}$ is an $n \times (p + 1)$ matrix of independent variables and $\mathbf{x}^{\mathrm{T}}$ denotes the transpose of matrix $\mathbf{x}$. Apparently, such OLS models assume that the relationships between variables are spatially stationary. The model parameters $\boldsymbol{\beta}$ estimated using the training samples are considered globally equivalent and suitable to other parts of the study area.

3.2.2 Geographically weighted regression. Built upon the OLS regression technique, the GWR model is formulated as

$$y_i = \beta_0(\boldsymbol{u}_i) + \sum_{j=1}^{p} x_{ij}\beta_j(\boldsymbol{u}_i) + \varepsilon_i, \tag{3}$$

where $\boldsymbol{u}_i$ represents the location of the ith sample; the other variables and parameters are the same as in equation (1). Apparently, the major difference between GWR and

OLS lies in the calculation of the model coefficients β, which are spatially varying in GWR. The estimation of β at each location $\boldsymbol{u}_i$ is derived from

$$\hat{\boldsymbol{\beta}}(\boldsymbol{u}_i) = \left[\mathbf{x}^{\mathrm{T}}\mathbf{w}(\boldsymbol{u}_i)\mathbf{x}\right]^{-1}\mathbf{x}^{\mathrm{T}}\mathbf{w}(\boldsymbol{u}_i)\boldsymbol{Y}, \tag{4}$$

where $\mathbf{w}(\mathbf{u}_i)$ is an $n \times n$ matrix that represents the geographical weighting of the ith sample with its diagonal elements. The calculation of $\mathbf{w}(\boldsymbol{u}_i)$ is dependent on the kernel chosen and the bandwidth for that kernel. There are typically two types of kernels: fixed and adaptive. A fixed kernel, centred at each regression point, applies a constant bandwidth across the entire study area. Due to its simplicity, the fixed kernel was used in this study. Readers may refer to Fotheringham *et al.* (2002) for details of GWR principles.

3.2.3 Inverse distance weighting, ordinary kriging and cokriging. Spatial interpolation is a procedure that typically applies mathematical functions to estimate the values at unsampled/unmeasured locations by using the existing observations from sampled areas. In this study, three commonly used spatial interpolation techniques were evaluated, that is, IDW, OK and COK.

IDW determines each unknown value by using a linearly weighted combination of values at sampled locations. The equation is typically formulated as

$$\hat{z}(\boldsymbol{u}^*) = \frac{\sum_{i=1}^{n} z(\boldsymbol{u}_i) w_i^p}{\sum_{i=1}^{n} w_i^p}, \tag{5}$$

where $\hat{z}(\boldsymbol{u}^*)$ is the value to be estimated at the location $\boldsymbol{u}^*$, $z(\boldsymbol{u}_i)$ is the measured value for the ith sample point at an observed location $\boldsymbol{u}_i$, n is the total number of samples, the ith weight w_i is the inverse of distance from $\boldsymbol{u}_i$ to $\boldsymbol{u}^*$ and p is the weighting power. A commonly used value of 2 was used for p in this study. For details of IDW, refer to Watson and Philip (1985).

Kriging is a statistically strict spatial regression approach that explicitly considers spatial autocorrelation within data sets by constructing semivariograms from sample points. The estimation of an unknown value is typically formulated as

$$\hat{z}(\boldsymbol{u}^*) = \sum_{i=1}^{n} z(\boldsymbol{u}_i) w_i. \tag{6}$$

Calculation of the weight parameter w_i is based on a semivariogram, which represents the semivariance between samples as a function of distance:

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [z(\boldsymbol{u}_i + h) - z(\boldsymbol{u}_i)]^2, \tag{7}$$

where $\gamma(h)$ is the semivariance and $n(h)$ is the number of paired samples at a separation distance of h. Because the jagged semivariogram derived from samples is not suitable for the real application, a mathematical model is often used instead to parameterize

the semivariogram. In this study, the exponential model was found to perform better than others. As a widely used kriging model, OK assumes the data set has a constant but unknown mean. For details, refer to Oliver and Webster (1990).

COK is often used to improve the prediction for the entire surface by taking into account not only the primary variable but also one or more secondary variables. The estimation of an unknown value is formulated as

$$\hat{z}(\boldsymbol{u}^*) = \sum_{i=1}^{n} z(\boldsymbol{u}_i) w_i + \sum_{k=1}^{p} \sum_{j=1}^{m} v_{jk}(\boldsymbol{u}_{jk}) w_{jk}, \tag{8}$$

where $v_{jk}(\boldsymbol{u}_{jk})$ is the kth secondary variable for the jth sample point at the location $\boldsymbol{u}_{jk}$, w_{jk} is the weight assigned to the secondary variable $v_{jk}(\boldsymbol{u}_{jk})$, p is the number of secondary variables at each location and m is the total number of locations for the secondary variables. To determine all the weight parameters, a total number of $(p + 1)(p + 2)/2$ semivariograms and cross-semivariograms have to be calculated. A cross-semivariogram is defined as

$$\gamma_{ef}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [z_e(\boldsymbol{u}_i + h) - z_e(\boldsymbol{u}_i)][z_f(\boldsymbol{u}_i + h) - z_f(\boldsymbol{u}_i)], \tag{9}$$

where $\gamma_{ef}(h)$ is the cross-semivariance between the variables e and f and z_e and z_f represent the measured values for the variables e and f. Please refer to McBratney and Webster (1983) for details. To limit the calculation intensity in practice, the number of secondary variables is often less than four. In this study, the three most significant QB-derived variables, which have been selected using the mRMR feature selection algorithm (see §3.1.2), were used as the secondary variables in COK. They are the spectral means for the red and NIR bands and the image-texture measure for the NIR band.

3.3 *Model evaluation*

The aforementioned models were evaluated in three ways. (i) OLS versus GWR: the two types of models were assessed and compared at the five sampling densities in terms of adjusted R^2 (coefficient of determination), Akaike's information criterion (AIC) (Akaike 1974), RMSE and spatial autocorrelation (Moran's I) and SH in model residuals. (ii) GWR versus the other three spatial interpolation techniques (i.e. IDW, OK and COK): comparisons were made at the five sampling densities in terms of mean error, mean (absolute) error (i.e. average of errors' absolute values), RMSE and standard deviation of errors. (iii) Comparison of the canopy height estimation maps for all the five types of models at the five sampling densities. All plot data over the entire study area were used for our model evaluation.

4. Results and discussion

4.1 *Comparison of GWR and OLS*

Separate OLS models were developed for the five sampling densities (table 2). The three independent variables used were those selected from the mRMR feature selection algorithm, including the mean spectral values for the red and NIR bands and

Table 2. OLS models for the five sampling densities.

Sampling density (plot interval)	Model
'1' (460 m)	CH = 47.345 – 0.066(DN3) – 0.049(DN4) + 0.077(TXIT4)
'2' (260 m)	CH = 46.935 – 0.103(DN3) – 0.035(DN4) + 0.102(TXIT4)
'3' (160 m)	CH = 53.796 – 0.137(DN3) – 0.028(DN4) + 0.066(TXIT4)
'4' (60 m)	CH = 50.588 – 0.128(DN3) – 0.029(DN4) + 0.088(TXIT4)
'5' (20 m)	CH = 53.636 – 0.135(DN3) – 0.030(DN4) + 0.076(TXIT4)

Note: CH is the estimated canopy height; DN3 is the mean spectral value for the red band; DN4 is the mean spectral value for the NIR band; and TXIT4 is the image-texture value for the NIR band.

the image-texture value extracted from the NIR band. They proved more significant than the other QB-derived variables estimating the forest canopy height in this study. Specifically, the mean spectral values for the red and NIR bands have a negative correlation with the canopy height, whereas a positive correlation can be found using the image-texture value extracted from the NIR band. Table 3 shows five GWR models developed for the five sampling densities. Different from OLS models, the model coefficients derived from GWR are spatially nonstationary, that is, varying with locations. The minimum, median and maximum values for each parameter, as well as the selected optimal bandwidths, are presented in table 3. Apparently, the variability of each coefficient varies with the increase in the sampling density. When the sampled plots are far from each other, the optimal bandwidth for GWR is large. With the increase in the sampling density, the parameters tend to reveal more spatial details, that is, higher variability (table 3).

Table 4 is the comparison of the canopy height estimation performance using GWR and OLS. When a very low sampling density is used (i.e. plot interval of 460 m), GWR

Table 3. GWR models for the five sampling densities.

Sampling density (plot interval)	Intercept (min/median/ max)	DN3 (min/median/ max)	DN4 (min/median/ max)	TXIT4 (min/median/ max)	Selected bandwidth (m)
'1' (460 m)	39.543	−0.188	−0.064	−0.015	3420
	57.508	−0.087	−0.050	0.075	
	71.168	−0.007	−0.028	0.118	
'2' (260 m)	21.147	−0.328	−0.071	−0.093	2180
	50.645	−0.122	−0.019	0.071	
	80.160	0.045	0.026	0.206	
'3' (160 m)	−0.901	−0.751	−0.099	−0.236	770
	54.928	−0.169	−0.012	0.014	
	143.301	0.097	0.152	0.272	
'4' (60 m)	7.800	−0.491	−0.091	−0.144	640
	50.193	−0.154	−0.011	0.044	
	106.604	0.102	0.121	0.210	
'5' (20 m)	11.794	−0.533	−0.079	−0.068	200
	51.862	−0.156	−0.009	0.042	
	103.128	0.107	0.130	0.158	

Note: Canopy height is the model estimate; DN3 is the mean spectral value for the red band; DN4 is the mean spectral value for the NIR band; and TXIT4 is the image-texture value for the NIR band.

Table 4. Comparison of the canopy height estimation performance using GWR and OLS for the five sampling densities.

Sampling density (plot interval)	Model	Adjusted R^2	RMSE (m)	Akaike's information criterion	Moran's I	Spatial heterogeneity, SH (%)
'1' (460 m)	OLS	0.42	9.31	882.77	0.64	40.69
	GWR	0.50	8.89	876.06	0.61	34.56
'2' (260 m)	OLS	0.51	8.96	2318.60	0.60	38.44
	GWR	0.65	8.16	2248.59	0.53	28.35
'3' (160 m)	OLS	0.51	8.85	4867.46	0.60	39.74
	GWR	0.75	8.13	4661.02	0.47	22.58
'4' (60 m)	OLS	0.53	8.83	19388.45	0.61	38.38
	GWR	0.74	6.93	18077.57	0.50	29.17
'5' (20 m)	OLS	0.53	8.83	117706.86	0.61	38.50
	GWR	0.72	5.81	109796.69	0.29	*

Note: *SH value was not available, as no semivariogram could be calculated.

and OLS models have a similar performance in terms of adjusted R^2, RMSE, AIC, Moran's I and SH. As for AIC, a decrease of 3.0 is typically considered a significant improvement of model performance (Fotheringham *et al*. 2002); however, the value of AIC gets higher with the increase in the sample size. With the increase in the sampling density, GWR models performed consistently better than the corresponding OLS models, that is, with lower RMSE, higher adjusted R^2, smaller AIC, smaller Moran's I and smaller SH. Our result that GWR outperformed OLS is consistent with the findings of many previous studies (Foody 2003, 2004, Wang *et al*. 2005, Kimsey *et al*. 2008, Propastin 2009). Furthermore, as the sampling density gets higher (i.e. smaller plot interval), GWR has a higher ability to improve the canopy height estimation accuracy than OLS (table 4). For example, there are relatively small differences between GWR and OLS at the sampling density '1', with Δ(adjusted R^2) = 0.08, Δ(RMSE) = 0.42 m, Δ(AIC) = 6.71, Δ(Moran's I) = 0.13 and Δ(SH) = 6.13%. Here, 'Δ' represents the difference between two values. However, large differences are found at a higher sampling density '5', with Δ(adjusted R^2) = 0.19, Δ(RMSE) = 3.02 m, Δ(AIC) = 7910.17 and Δ(Moran's I) = 0.32. This phenomenon can be explained by the fact that GWR considers the local spatial relationships. When samples get denser in a specific area, the neighbouring plots can describe better the spatial variability and define more reliable weights for each centre plot. Unlike GWR, OLS is a global regression technique and, thus, the increase in the sampling density only helps to increase the OLS model performance to some extent. For example, table 4 shows a relatively large change in adjusted R^2, RMSE, Moran's I and SH for OLS models when the plot interval decreases from 460 to 260 m, but the model performance is barely influenced by the increase in the sampling density if the plot interval is smaller than 260 m.

We further evaluated the significance of the spatial variability in GWR parameter estimates using a Monte Carlo test (table 5). At low sampling densities (e.g. 460 m plot interval), the nonstationary relationship between the QB variables/intercept and the canopy height tends to be insignificant. With the increase in the sampling densities, the relationship moves to statistically significant. This is consistent with the results in table 4 that GWR only slightly outperforms OLS using the large plot interval of 460 m. However, GWR can fit the data better than OLS when there exists a high degree of nonstationarity in parameter estimates.

Table 5. Comparison of the significance of the spatial variability in GWR parameter estimates using Monte Carlo test for the five sampling densities.

Sampling density (plot interval)	Parameter	*p*-Value
'1' (460 m)	Intercept	n/s
	DN3	n/s
	DN4	n/s
	TXIT4	n/s
'2' (260 m)	Intercept	n/s
	DN3	<0.001
	DN4	n/s
	TXIT4	<0.01
'3' (160 m)	Intercept	<0.05
	DN3	<0.001
	DN4	<0.001
	TXIT4	<0.01
'4' (60 m)	Intercept	<0.001
	DN3	<0.001
	DN4	<0.001
	TXIT4	<0.001
'5' (20 m)	Intercept	<0.001
	DN3	<0.001
	DN4	<0.001
	TXIT4	<0.001

Note: n/s, not significant.

4.2 *Comparison of GWR, IDW, OK and COK*

Table 6 is the comparison of canopy height estimation performance over the entire study area using the five types of models – GWR, IDW, OK, COK and OLS – for the five sampling densities. It appears that (i) all models' mean errors are close to zero; (ii) the mean (absolute) errors that represent the error magnitude have a similar trend as RMSE and error standard deviation; and (iii) RMSE and error standard deviation representing the error variance are very similar (table 6). Clearly, the performance of the three spatial interpolation techniques depends critically on the sampling density. Higher densities result in lower errors. Although they have a similar performance using relatively low density samples (i.e. plot intervals of 460 and 260 m), the errors of IDW decrease dramatically at higher densities (i.e. plot intervals of 160, 60 and 20 m). It is beyond our expectation that the COK, which used both primary height variable and three secondary QB variables, only slightly outperformed OK and produced larger canopy height estimation errors than both IDW and GWR at high densities. A possible reason for this is the difficulties associated with the reliable estimation of a relatively large number of semivariograms.

Compared to the spatial interpolation methods, GWR outperforms IDW, OK and COK at most sampling densities (table 6). An exception occurs at the smallest plot interval of 20 m, where IDW seems more suitable to estimate the canopy height than GWR. In contrast to those models considering spatial dependence, aspatial OLS produced relatively consistent results. Apparently, OLS fits better in the condition when the sample spatial autocorrelation is very low. Overall, table 6 suggests that the regression models (including GWR and OLS) are much better than the spatial interpolation techniques at relatively low sampling densities, while the performance of GWR and spatial interpolation tends to be similar with the increase in the sample numbers.

Table 6. Comparison of the model performance using OLS, GWR, IDW, OK and COK for the five sampling densities.

Sampling density (plot interval)	Model	Mean (m)	Mean (absolute) (m)*	RMSE (m)	Standard deviation (m)
'1' (460 m)	OLS	−0.17	7.36	9.31	9.15
	GWR	0.37	6.85	8.89	8.89
	IDW	0.27	7.89	10.69	10.73
	OK	0.35	8.32	10.53	10.52
	COK	0.34	8.09	10.30	10.30
'2' (260 m)	OLS	0.18	6.74	8.96	8.80
	GWR	0.30	5.92	8.16	8.15
	IDW	0.01	7.33	10.07	10.13
	OK	0.04	7.84	10.11	10.11
	COK	0.05	7.72	9.94	9.94
'3' (160 m)	OLS	−0.09	6.67	8.85	8.77
	GWR	−0.31	5.45	8.13	8.13
	IDW	0.11	6.19	8.81	8.84
	OK	0.13	7.14	9.31	9.31
	COK	0.13	6.84	9.00	9.00
'4' (60 m)	OLS	0.08	6.62	8.83	8.76
	GWR	0.05	4.75	6.93	6.93
	IDW	0.20	4.79	7.24	7.22
	OK	0.19	6.11	8.23	8.22
	COK	0.19	5.88	7.95	7.95
'5' (20 m)	OLS	0.21	6.59	8.83	8.76
	GWR	−0.05	3.83	5.81	5.89
	IDW	−0.08	3.15	5.15	5.15
	OK	−0.08	4.50	6.52	6.52
	COK	−0.08	4.37	6.36	6.36

Note: *This mean error was calculated using model errors' absolute values.

4.3 *Mapping of model estimates*

Figure 3 is the comparison of the estimated canopy height maps of the entire study area using OLS, IDW, OK, COK and GWR for the five sampling densities. On visual examination, the maps derived from OLS and GWR provide sharp boundaries among non-vegetation features (e.g. cutblocks, river and roads in black colour). This is because QB-derived variables were used in the regression models and the original shapes of those features are, therefore, well maintained as shown in the QB image (figure 1). However, the maps derived from OLS and GWR have different abilities to describe the canopy height details. As GWR also considered the spatial dependence between the neighbouring samples and the centre plots, the canopy height estimate at each location is more accurate than that using OLS. As to the spatial interpolation techniques, the maps derived from IDW, OK and COK reveal distinct interpolation characteristics: IDW maps show a bull's eye pattern, while OK and COK maps have a smoothing effect. Compared to GWR and OLS maps, all the maps derived from IDW, OK and COK appear blurry at the boundaries of non-vegetation features, and this is due to their limitation in estimating abrupt changes of geographic features. At high sampling densities (e.g. plot intervals of 160, 60 and 20 m), however, large forest stands were better delineated using spatial interpolation than OLS.

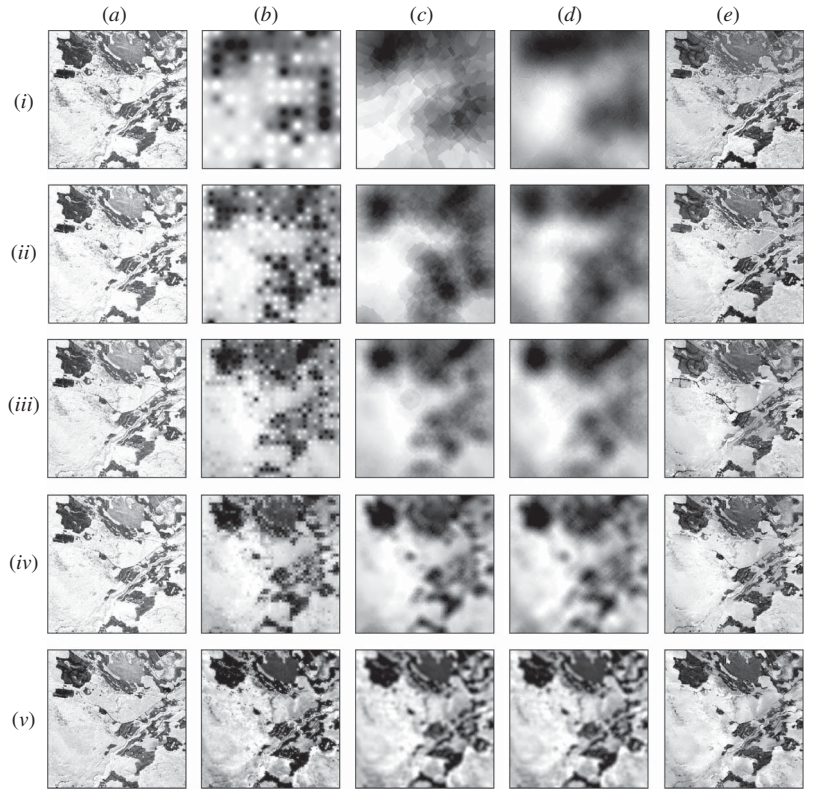

Figure 3. Comparison of the estimated canopy height maps of the entire study area using the five types of models, that is, OLS (*a*), IDW (*b*), OK (*c*), COK (*d*) and GWR (*e*), for the five sampling densities, that is, 460 m (*i*), 260 m (*ii*), 160 m (*iii*), 60 m (*iv*) and 20 m (*v*).

5. Conclusion

The major advantage of GWR over OLS is the explicit consideration of spatial non-stationarity in modelling. By using neighbouring samples to calibrate the model coefficients associated with a focal point, GWR models can yield a better fit during model calibration than OLS models. However, it should be noted that, although remotely sensed data have the capacity to monitor geographic features in large extents, samples from field measurements can only cover a small portion. The operational use of the GWR technique (especially in forest management) requires generating all estimates for the entire study site, including the large areas where field data are absent. In this study, GWR was introduced to estimate the forest canopy height, with the influence of using five sampling densities. These densities represent plot intervals from very large to very small: 460, 260, 160, 60 and 20 m. GWR and another four commonly used analytical techniques – OLS, IDW, OK and COK – were also compared. As a high sampling density requires a large number of ground-truth data (e.g. 65 536 plots at the 20 m interval), the lidar-derived CHM was used to simulate those data, as it has a full cover of the study area with a relatively high accuracy in capturing tree vertical structures. The following conclusions can be drawn:

- For model calibration that assesses model goodness of fit (i.e. R^2 and AIC) based on training samples, and validation that evaluates model performance (i.e. RMSE, Moran's I and SH) based on the data over the entire study area, GWR outperformed OLS, especially at high sampling densities. Typically, GWR benefits more from the increase of measurements than OLS.
- Compared to the three spatial interpolation techniques – IDW, OK and COK – GWR has a better performance, especially at low sampling densities. However, their canopy height estimation results tend to become more similar as more samples are collected. Using a very small plot interval of 20 m, IDW even outperformed GWR. Although COK is a favoured geostatistical technique with strong generalization ability by considering both primary and secondary variables, it only outperformed OK marginally and produced larger errors than GWR at most sampling densities.
- Due to the limitation of estimating abrupt changes of geographic features, the three spatial interpolation techniques tend to predict more blurry surfaces than those using regression techniques. At high sampling densities (i.e. plot intervals of 160, 60 and 20 m), however, large forest stands were more accurately delineated using IDW, OK or COK, than OLS. Compared to the other techniques tested in this study, GWR has advantages in producing maps with more accurate estimates of the canopy height and it also maintained better the shapes of geographic features. However, it should be noted that the above conclusions were made based on the reality that our study area is relatively flat with a gentle slope of 5°–10°.
- Many remote-sensing studies require field measurements to accurately calibrate models. However, compared to the large image scene, field data typically cover just a small portion (i.e. a low sampling density). In our study, GWR and OLS produced similar results at the large plot interval of 460 m. Therefore, we suggest that the use of GWR in remote-sensing operational applications should be carefully considered, as GWR may not necessarily produce significantly better results than the simple OLS when a small portion of field data is used.
- Similar to GWR, another regression approach based on spatially varying model coefficients is the expansion method (Cassetti 1972). However, the model coefficients in the expansion method are redefined by other equations of variables (e.g. latitude) (Cassetti 1972, Fitch *et al.* 2010). For GWR, the coefficients are calculated based on the spatial correlations between the independent and dependent variables within certain spatial extents (i.e. bandwidths). Compared to the expansion method, GWR lacks the flexibility to incorporate extra variables in modelling; however, it has the advantage of improving the model performance by directly considering the local spatial relationship – the nonstationarity – between the variables.
- As pointed out by Jetz *et al.* (2005), we should also be aware that the spatial nonstationarity relationships between variables discovered in GWR models may be 'in fact global', but 'appear to vary locally due to missing variables or interactions terms'. However, it is difficult to thoroughly test all potential variables. Jetz *et al.* (2005) also argued that an arbitrary variable used in GWR may still increase the model prediction performance more than using OLS does. To verify that the variables used in our study are more effective to predict forest height than an arbitrarily selected one, we performed a simple test by using the QB panchromatic band as a 'randomly selected' variable in GWR modelling. The model's RMSE increased by 19% compared to the results using the red and NIR

bands, which supports our decisions for variable selection made in this research. Although there remain limitations, GWR has been widely acknowledged as a very useful tool to identify the relationships between variables (Jetz *et al.* 2005).

References

Akaike, H., 1974, A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, pp. 716–723.

Cassetti, E., 1972, Generating models by the expansion method: applications to geographical research. *Geographical Analysis*, **4**, pp. 81–91.

Chen, G., Hay, G.J., Castilla, G., St-Onge, B. and Powers, R., 2011, A multiscale geographic object-based image analysis (GEOBIA) to estimate lidar-measured forest canopy height using Quickbird imagery. *International Journal of Geographical Information Science*, **25**, pp. 877–893.

Cohen, W.B., Maiersperger, T.K., Spies, T.A. and Otter, D.R., 2001, Modelling forest cover attributes as continuous variables in a regional context with Thematic Mapper data. *International Journal of Remote Sensing*, **22**, pp. 2279–2310.

Donoghue, D.N.M. and Watt, P.J., 2006, Using LiDAR to compare forest height estimates from IKONOS and Landsat ETM+ data in Sitka spruce plantation forests. *International Journal of Remote Sensing*, **27**, pp. 2161–2175.

Fitch, D., Stow, D., Hope, A. and Rey, S., 2010, MODIS vegetation metrics as indicators of hydrological response in watersheds of California Mediterranean-type climate zones. *Remote Sensing of Environment*, **114**, pp. 2513–2523.

Foody, G.M., 2003, Geographical weighting as a further refinement to regression modelling: an example focused on the NDVI–rainfall relationship. *Remote Sensing of Environment*, **88**, pp. 283–293.

Foody, G.M., 2004, Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. *Global Ecology and Biogeography*, **13**, pp. 315–320.

Foody, G.M., 2005, Mapping the richness and composition of British breeding birds from coarse spatial resolution satellite sensor imagery. *International Journal of Remote Sensing*, **26**, pp. 3943–3956.

Fotheringham, A.S., Brunsdon, C. and Charlton, M., 2002, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, 269 pp. (New York: Wiley).

Fotheringham, A.S., Charlton, M.E. and Brunsdon, C., 1998, Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, **30**, pp. 1905–1927.

Franklin, J., 1995, Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, **19**, pp. 474–499.

Franklin, S.E. and McDermid, G.J., 1993, Empirical relations between digital SPOT HRV and CASI spectral response and lodgepole pine (*Pinus contorta*) forest stand parameters. *International Journal of Remote Sensing*, **14**, pp. 2331–2348.

Goodchild, M.F., 1986, *Spatial Autocorrelation (Catmog 47)*, 56 pp. (Norwich: Geo Books).

Hilker, T., Wulder, M.A. and Coops, N.C., 2008, Update of forest inventory data with lidar and high spatial resolution satellite imagery. *Canadian Journal of Remote Sensing*, **34**, pp. 5–12.

Hyde, P., Dubayah, P., Walker, W., Blair, J.B., Hofton, M. and Hunsaker, C., 2006, Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. *Remote Sensing of Environment*, **102**, pp. 63–73.

Jetz, W., Rahbek, C. and Lichstein, J.W., 2005, Local and global approaches to spatial data analysis in ecology. *Global Ecology and Biogeography*, **14**, pp. 97–98.

Kimsey, M.J., Moore, J. and McDaniel, P., 2008, A geographically weighted regression analysis of Douglas-fir site index in north central Idaho. *Forest Science*, **54**, pp. 356–366.

Lefsky, M.A., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A. and Harding, D., 1999, Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sensing of Environment*, **70**, pp. 339–361.

Li, H. and Reynolds, J.F., 1995, On definition and quantification of heterogeneity. *Oikos*, **73**, pp. 280–284.

Lim, K., Treitz, P., Baldwin, K., Morrison, I. and Green, J., 2003, Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. *Canadian Journal of Remote Sensing*, **29**, pp. 658–678.

Man Shrestha, P., 2006, Comparison of ordinary least square regression, spatial autoregression, and geographically weighted regression for modeling forest structural attributes using a geographical information system (GIS)/remote sensing (RS) approach. Master thesis, University of Calgary, Canada, 204 pp.

McBratney, A.B. and Webster, R., 1983, Optimal interpolation and isarithmic mapping of soil properties: V. Coregionalization and multiple sampling strategy. *Journal of Soil Science*, **34**, pp. 137–162.

McCoy, R.M., 2005, *Field Methods in Remote Sensing*, 159 pp. (New York: The Guilford Press).

Means, J.E., Acker, S.A., Harding, D.J., Blair, J.B., Lefsky, M.A., Cohen, W.B., Harmon, M.E. and McKee, W.A., 1999, Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the Western Cascades of Oregon. *Remote Sensing of Environment*, **67**, pp. 298–308.

Morgenstern, K., Black, T.A., Humphreys, E.R., Griffis, T.J., Drewitt, G.B., Cai, T.B., Nesic, Z., Spittlehouse, D.L. and Livingstone, N.J., 2004, Sensitivity and uncertainty of the carbon balance of a Pacific Northwest Douglas-fir forest during an El Nino La Nina cycle. *Agricultural and Forest Meteorology*, **123**, pp. 201–219.

Oliver, M.A. and Webster, R., 1990, Kriging: a method of interpolation for geographical information systems. *International Journal of Geographic Information Systems*, **4**, pp. 313–332.

Peng, H., Long, F. and Ding, C., 2005, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**, pp. 1226–1238.

Propastin, P.A., 2009, Spatial non-stationarity and scale-dependency of prediction accuracy in the remote estimation of LAI over a tropical rainforest in Sulawesi, Indonesia. *Remote Sensing of Environment*, **113**, pp. 2234–2242.

Salas, C., Ene, L., Gregoire, T.G., Næsset, E. and Gobakken, T., 2010, Modelling tree diameter from airborne laser scanning derived variables: a comparison of spatial statistical models. *Remote Sensing of Environment*, **114**, pp. 1277–1285.

Wang, Q., Ni, J. and Tenhunen, J., 2005, Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems. *Global Ecology and Biogeography*, **14**, pp. 379–393.

Watson, D.F. and Philip, G.M., 1985, A refinement of inverse distance weighted interpolation. *Geoprocessing*, **2**, pp. 315–327.

Welch, R. and Ahlers, W., 1987, Merging multiresolution SPOT HRV and Landsat TM data. *Photogrammetric Engineering and Remote Sensing*, **53**, pp. 301–303.

Wulder, M., Niemann, K.O. and Goodenough, D.G., 2000, Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. *Remote Sensing of Environment*, **73**, pp. 103–114.

Zhang, L., Ma, Z. and Guo, L., 2009, An evaluation of spatial autocorrelation and heterogeneity in the residuals of six regression models. *Forest Science*, **55**, pp. 533–548.

Zhao, K. and Popescu, S., 2009, Lidar-based mapping of leaf area index and its use for validating GLOBCARBON satellite LAI product in a temperate forest of the southern USA. *Remote Sensing of Environment*, **113**, pp. 1628–1645.

Zhao, K., Popescu, S. and Nelson, R., 2009, Lidar remote sensing of forest biomass: a scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment*, **113**, pp. 182–196.