Spatial Distribution Characteristics of Species Diversity Using Geographically Weighted Regression Model

Jeongmook Park,[1†] Byoungkoo Choi,[2†] and Jungsoo Lee[1*]

[1]Department of Forest Management, Kangwon National University,
Chuncheon 24341, Republic of Korea
[2]Department of Forest Environment Protection, Kangwon National University,
Chuncheon 24341, Republic of Korea

The objective of this study is to evaluate the spatial distribution patterns of species diversity at different spatial scales, focusing on the Baekdudaegan Protected Area, which is a biodiversity hotspot in the Republic of Korea. The tree species diversity index (Shannon–Weaver index; H') was calculated using tree species data from a 1:5k forest-type map, and the spatial analysis was performed with a 1×1 km^2 grid. Ten factors were selected to estimate the impact of topographic (elevation, slope, northern slope, curvature, wetness, and relief) and geographic (distances from water, road, forest road, and urban area) factors on H' using the ordinary least squares (OLS) and geographically weighted regression (GWR) models. H' increased with the spatial scale. Also, the coefficient of determination (R^2) of the GWR and OLS models increased proportionally and the R^2 of the GWR model was higher than that of the OLS model. Corrected Akaike information criterion (AICc) was lower in the GWR model than in the OLS model, which indicates that the GWR model fits the calculated H' better than the OLS model. Thus, the GWR model is considered to be more practical than the OLS model for understanding the effects of topographic and geographic factors on H' at different scales.

1. Introduction

Rapidly increasing deforestation and the resulting environmental changes have caused a significant reduction in global biodiversity. The Convention on Biological Diversity held in Rio de Janeiro in 1992 led to the establishment of the Nagoya Protocol in 2010, which aims to conserve global biodiversity.[1] To cope with the changing environmental policy in intergovernmental relations on global biodiversity issues, the Republic of Korea established a national plan for forest biodiversity (2013 to 2017) and made related efforts including hosting the 12th Convention on Biological Diversity.[2,3] The Baekdudaegan region is well known for abundant biological and ecological communities and for playing a key role in forest ecosystems in the Republic of Korea. Owing to its unique vegetative characteristics within varied topographic features, the Baekdudaegan region has a diverse range of wildlife and plant species and is a hotspot for biodiversity both ecologically and for academic research.[4,5]

Determining the factors that affect species and biodiversity distributions is critical in conservation perspectives.[6] However, huge gaps in the available information on the spatial distribution of plant species pose a major challenge for regional conservation planning in many parts of the world.[7] Previous studies on domestic forest distribution patterns have examined the relationships of the distribution of *Salicaceae* plant species, topography, and soils with the factors that affect the distribution patterns of plants in the northern Gyeonggi region by principal component analysis (PCA).[8,9] Many studies have investigated the spatial patterns of species diversity and the relationship between species diversity and topographic and geographic factors using various methods such as correlation, regression, and geographic information system (GIS) analyses.[10–12] Various regression, habitat, and fitness models have been developed in an attempt to quantitatively evaluate biodiversity and spatial distribution features.[13] Although the spatial range of biodiversity research studies has been expanded, it is still limited to site-specific studies along with the recognition of environmental issues including climate change.[14,15]

In the Republic of Korea, Han and Lee[16] and Sim *et al.*[17] conducted studies on the dependence of the distribution modeling of species on the spatial scale and distribution patterns of biodiversity. The topographic and geographic factors affecting forest biodiversity were often spatially inconsistent with similar distributions in adjacent areas but different distributions in distant areas.[18,19] In addition, because the vegetation exhibits various communities in a continuous pattern, causing spatial autocorrelation, the spatial location of attribute information should be considered to analyze the quantitative spatial pattern.[20,21]

The geographically weighted regression (GWR) model has been used in forestry sectors, such as for forest resource evaluation (e.g., timber volume estimation), because it allows the estimation of the regression model even in cases where the requirements for spatial correlation, normality, and homoscedasticity are not met.[22,23] Although the distribution of species diversity varies regionally and with the spatial scale within a forest ecosystem, few studies have examined its correlation with the tree species diversity index.[24]

Our study aims to compare the GWR model with the ordinary least squares (OLS) model in terms of the variation with the spatial scale and evaluate major factors affecting tree species diversity by examining the effects of topographic and geographic factors on the tree species diversity index (H') of vegetation communities in the Baekdudaegan region.

We concluded that the GWR model is more practical than the OLS model for understanding the effects of topographic and geographic factors on H' at different spatial scales. It is hoped that this study will be an important reference for future studies on the biodiversity distribution in the region.

2.   Materials and Methods

2.1   Study area

The study site is the Baekdudaegan Protected Area of Gangwon Province, which is well known as a biodiversity hotspot in the Republic of Korea, as shown in Fig. 1. The
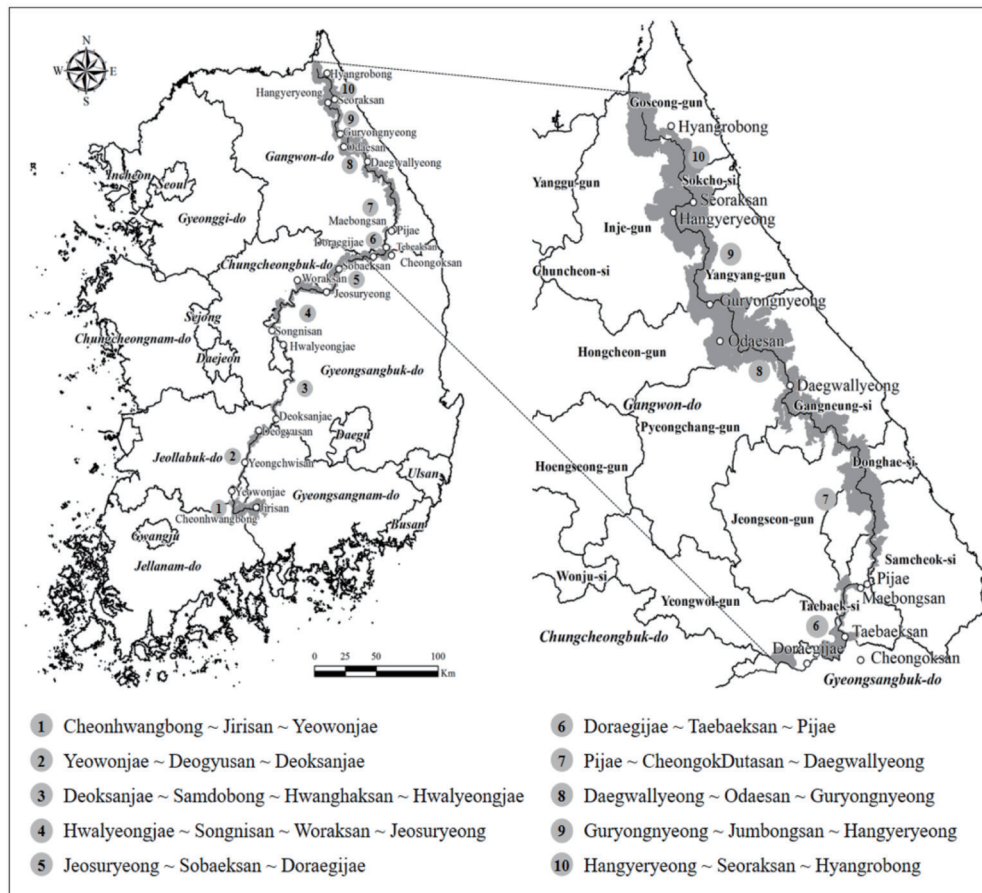
Fig. 1. Map of the Baekdudaegan Protected Area of Gangwon Province, the location of the study. The gray area in the left map is the Baekdudaegan Protected Area and the right map is an enlargement of the Baekdudaegan Protected Area of Gangwon Province.

Baekdudaegan Protected Area of Gangwon Province is divided into 10 sections according to the ecological map of Baekdudaegan Mountains; the northern region (Goseong, Inje, Sokcho, and Yangyang) containing section 10; the central region (Hongcheon, Gangneung, and Pyeongchang) contains sections 8 and 9; and the southern region (Jeongseon, Donghae, Samcheok, Taebaek, and Yeongwol) contains sections 6 and 7.[25]

The total study area is 134093 ha, 70% of which is the core area and 30% the buffer zone. The total area of the forest is 126363 ha, 74% of which is deciduous forest, 14% coniferous forest, and 12% mixed coniferous and deciduous forests.[26] The major deciduous species are *Quercus mongolica*, *Fraxinus rhynchophylla*, *Carpinus cordata*, *Betula costata*, *Acer pseudosieboldianum*, and *Lindera obtusiloba*. The major coniferous species are *Pinus densiflora*, *Pinus koraiensis*, *Larix leptolepis*, and *Abies koreana*.[27–31]

The average elevation is higher in the northern (894 m) and southern (863 m) regions than in the central region (758 m). The average slope gradient is also higher in the northern (33°) and southern (29°) regions than in the central region (28°).

2.2   Data and study methods

In this study, the tree species diversity index, which is a dependent variable to be applied to the GWR and OLS models, was calculated for each of the spatial scales 3 by 3 (s3), 5 by 5 (s5), 7 by 7 (s7), and 9 by 9 (s9) using the Baekdudaegan Protected Area map (1:25000 scale) and the forest-type map (1:5000 scale) provided by the Korea Forest Service. Topographic factors were used as independent variables to obtain a digital elevation model [e.g., DEM (30 m) from forest type map (1:5000 scale)] and geographic factors were calculated as independent variable using a river map, a road map, a forest road map, and a building map (1:25000 scale).

The calculated factor values were applied to the OLS and GWR models; then, the corresponding values obtained with both models were compared and the major factors affecting H' were evaluated, as shown in Fig. 2. The ArcGIS 10.1 software was used to calculate H'. The dependent variable H and independent variables, and the topographic and geographic factors were used to construct the GIS database by using the ArcGIS 10.1 program. GWR and OLS were calculated by using "Geographically weighted regression" tool and the "Ordinary least squares" tool of the ArcGIS 10.1 program.

2.3   Tree species diversity index (H')

H' is related to species richness and species evenness. It provides more information than simply the number of species present and makes species richness and evenness easier to see through numerical structures by quantifying the local tree species diversity.[32]
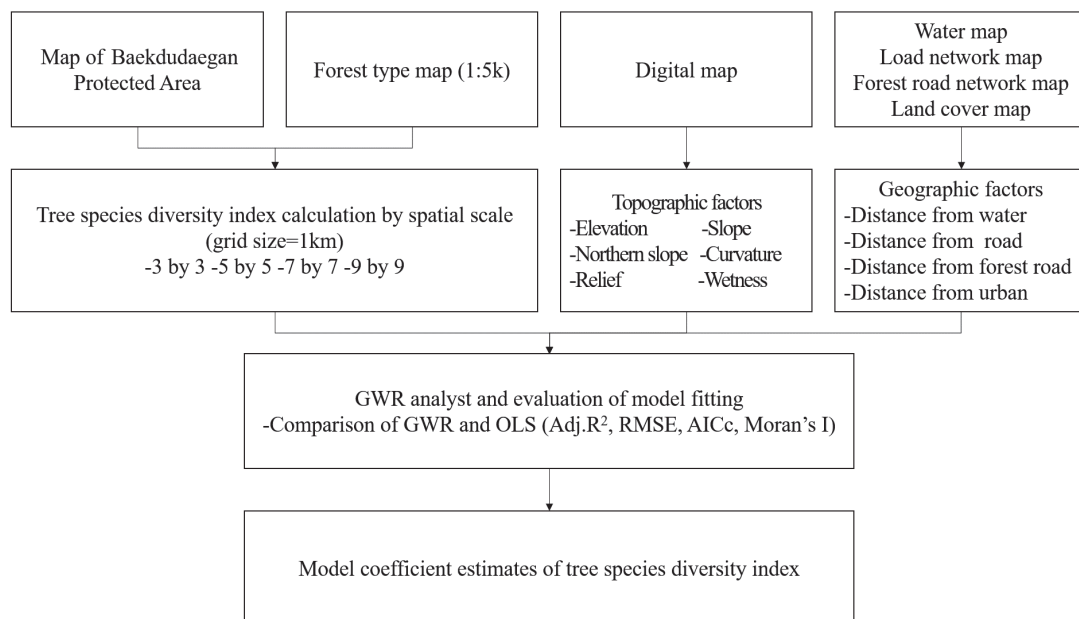

Fig. 2.   Schematic diagram of the research methodology.

H' was estimated using tree species data in the forest-type map. The forest-type map is a spatial digital map that shows the age and diameter classes of trees on a map of 1:5000 scale in the Republic of Korea. The tree species data of the forest-type map is composed of approximately 52 items of forest information obtained from a forest of less than 0.5 ha, which is the minimum area unit of the forest. Tree species data were converted from the polygon form to the grid form except for nonforest and deforested areas. The spatial resolution of the grid was set to 1×1 km^2. Graham and Hijmans analyzed correlation at a resolution using 1×1 km^2, 25×25 km^2, and 50×50 km^2 in the calculation of species richness and found that the calculation of species richness is more correlated at a low resolution than at a high resolution.[15] However, in this study, the forest-type map used in the calculation of the species diversity index had a scale of 1:5000. Therefore, when converting from the polygon form to the grid form, the highest resolution of 1×1 km^2 was used. The spatial scale used to determine the range of population in the species diversity index calculation was set to 3 by 3 (s3), 5 by 5 (s5), 7 by 7 (s7), and 9 by 9 (s9).

As shown in Fig. 3, 3 by 3 means that the value of each cell is calculated by moving the cell from the total of nine populations to determine the value of H' of the central grid at the spatial scale that was calculated by applying the spatial scale technique.[17] This was carried out to determine H' according to the change in spatial scale.

The Shannon–Weaver index, which is the most widely used index for calculating H' in forest ecosystems, was chosen to calculate H' according to the spatial scale. H' is calculated by applying the tree species diversity index to Shannon's formula. H' is the Shannon–Weaver index, whose value is the number of grids of all species at the spatial scale divided by the total number of grids at the spatial scale.[32] The formula used to calculate H' is as follows.[33]

$$H' = \sum_{i=1}^{k} \left(P_i \times \ln P_i \right)$$

$$\text{where } P_i = \frac{\text{Grid number including species}(i) \text{ in each spatial scale}}{\text{Grid number including all species in each spatial scale}}$$

(1)

| a.3 by 3 | b.5 by 5 | c.7 by 7 | d.9 by 9 |

☐ Tree population

☐ Total populations applied to species diversity index calculation

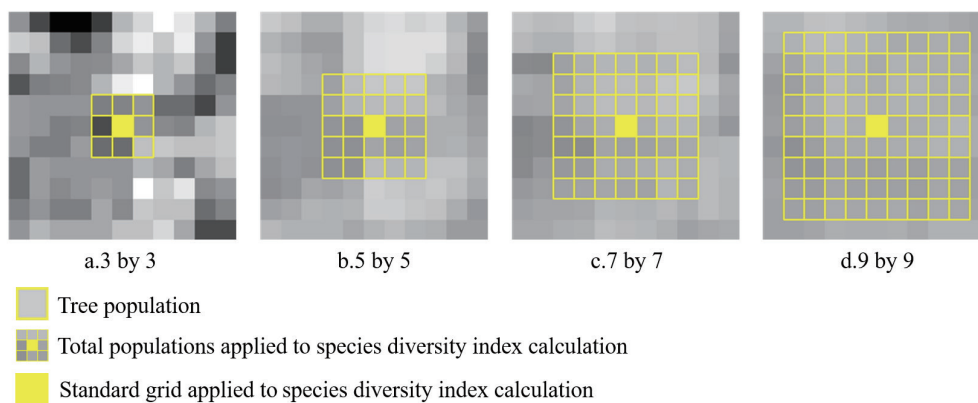☐ Standard grid applied to species diversity index calculation

Fig. 3.   (Color online) Various spatial scale units for calculating tree species diversity index.

2.4　Topographic and geographic factors

As the independent variables to be applied to the GWR and OLS models, six topographic factors and four geographic factors were selected as additional factors that were judged to affect biodiversity by referring to prior studies on the evaluation of species diversity in Baekdudaegan.[8,10,34] Song and Cao reported that elevation and wetness significantly correlate with tree species richness.[35] Gonzalez and Mata reported that tree species richness and diversity correlate with elevation and slope, which are important factors for predicting tree species richness and diversity.[36]

On the other hand, slope is expressed as one of the eight directions, and cannot be digitalized; thus, in this study, calculations were performed using the northern slope, which can be digitalized.

The following factors are considered to affect tree species diversity. The topographic factors included elevation, slope, northern slope, curvature, relief, and wetness data extracted from the DEM, and the geographic factors included distance from water (D_water), distance from a road (D_road), distance from a forest road (D_forest road), and distance from an urban area (D_urban), which were calculated using the Euclidean distance, as shown in Table 1. The spatial analysis unit of the grid for topographic and geographic factors was set to 1×1 km^2 and constructed using the GIS database.

2.5　GWR model

The GWR model is used to determine tree species diversity from topographic and geographic factors.

Table 1
Six topographic and four geographic factors affecting tree species diversity applied to GWR model.

Factor		Note
Topographic factors	Elevation	elevation information using DEM
	Slope	slope extracted from DEM
	Northern slope	$\text{north slope} = \cos\left(\dfrac{aspect \times \pi}{180}\right)$　$360° = 1, 180° = 0$
	Curvature	$\text{curvature} = -2(A + B) \times 100$　$\begin{array}{\|c\|c\|c\|}\hline z1 & z2 & z3 \\\hline z4 & z5 & z6 \\\hline z7 & z8 & z9 \\\hline\end{array}$　$A = \left(\dfrac{z4 + z6}{2} - z5\right) - L^2$　$B = \left(\dfrac{z2 + z8}{2} - z5\right) - L^2$　$L = \text{grid distance (m)}$
	Relief	DEM 3×3 neighborhood standard deviation
	Wetness	$\text{wetness} = \ln\left(\dfrac{flow\,accumulation + 1}{slope + 1}\right)$
Geographic factors	D_water	Euclidean distance
	D_road	
	D_forest road	
	D_urban	

The GWR model provides each spatial position with a regression coefficient, which is unattainable using the global model (OLS). The regression analysis performed with the GWR model selects the center point and the standard distance in each selected area under the assumption that different regression coefficients can be acquired. This means that the regression coefficient is not calculated using a constant value but using the position functions for each geographic location i.[37,38] The formulas used for the OLS and GWR models are given in Eqs. (2) and (3), where Y_i are the response variables. X_k are the set of predictor variables p (k = 1, 2, ..., p); $\beta_0(u_i, v_i)$, $\beta_1(u_i, v_i)$, ..., $\beta_p(u_i, v_i)$ are the regression coefficients for the kth predictor variable and ith location and ε_i are the random error terms. The GWR model calculation was used only when the location coefficient (u_i, v_i) was assigned to the OLS model and the population parameter could be estimated for the position.[39]

$$Y = X\beta + \varepsilon \tag{2}$$

$$Y_i = \beta_0\left(u_i, v_i\right) + \sum_{k=1}^{p} \beta_k\left(u_i, v_i\right) X_{ki} + \varepsilon_i \tag{3}$$

The regression coefficient β_k was assigned a weighted value as a function of the position. A location near geographic location i had a greater effect than a distant area. Although the GWR model followed the weighted least squares (WLS) structure, the regression coefficient was weighted as

$$\hat{\beta}_i = (X'W_iX)^{-1}X'W_iY,$$

$$W_i = \begin{pmatrix} \omega_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \omega_{ik} \end{pmatrix}. \tag{4}$$

The weighted matrix W_i for geographic location i produced different estimated results depending on the spatial composition method. It was generally used as a bandwidth using a different kernel function depending on the density tendency of the data. When the bandwidth that determined the weighted value was fixed, the fixed kernel method was used, whereas when the bandwidth was variable, the adaptive kernel method was used. In the present study, the fixed kernel method was used when the observed sites were distributed regularly throughout the study area, whereas the adaptive kernel method was used when the sites were distributed irregularly or had uncertainty to determine. In addition, the corrected Akaike information criterion (AICc) method [Eq. (5)] was used to determine bandwidth because it considers the difference between observed and estimated values, and the complexity of a model.[23,40] We used the residuals between the regression points to evaluate the models.

2.6   Model evaluation

The coefficient of determination (R^2), the root mean square error (RMSE), the AICc index, and, the global Moran's *I* values were used to assess the OLS and GWR models.[23] The AICc was calculated using the bandwidth between the data point and the regression point, and used to determine the degree of improvement of the GWR model in comparison with the OLS model. If the GWR model had a lower AICc value and the difference was greater than 4, it was considered that the GWR model was better than the OLS model.[40,41]

$$AICc = 2n\log_e(\hat{\sigma}) + n\log_e(2\pi) + n\left(\frac{n + tr(S)}{n - 2 - tr(S)}\right)\left(-S = \frac{\hat{y}}{y}\right) \tag{5}$$

Here, n is the number of features, y is the dependent variable, $\hat{y}$ is the estimation, and $\hat{\sigma}$ is the estimated standard error of residuals.

To determine the spatial autocorrelation of residuals in spatial data, we used global Moran's *I* values. Also, Moran's *I* value was used to determine the similarities between the attribute data of the surrounding feature and the average attribute data of the total study area from the target feature. As Moran's *I* value approaches 0, spatial autocorrelation rarely appears and the observed value becomes more random; as it approaches +1, a positive spatial autocorrelation exists and the observed value becomes more clustered; and as it approaches −1, a negative spatial autocorrelation exists and the observed value becomes more dispersed.[42,43] For the spatial weight matrix W_{ij}, the inverse distance method was used, in which the spatial heterogeneity increases with the distance between spatial unit features i and j. The Euclidean distance was measured as the distance between features.[21] The weighted distance was set as 1 km so that the grids facing the surrounding area could be included.

$$Moran's\ I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} -W_{ij}(d)(x_i - \overline{x})(x_j - \overline{x})}{\sum_{i=1}^{n}\sum_{j=1}^{n} -W_{ij}(d)\sum_{i}^{n}(x_i - \overline{x})^2} \tag{6}$$

Here, n is the number of observations (grids), $\overline{x}$ is the mean of the variable, x_i is the value of the variable at a particular location, x_j is the value of the variable at another location, and W_{ij} is the weight index for the location of i relative to j.

3.   Results and Discussion

3.1   Distribution of *H′* at different scales

There are 12 tree species within the current study area, *Pinus densiflora*, *Pinus koraiensis*, *Larix leptolepis*, *Abies holophylla*, *Quercus mongolica*, *Quercus variabilis*, other *Quercus spp.* (*Quercus aliena*, *Quercus dentata*, and *Quercus serrata*), *Betula platyphylla var. japonica*, *Fraxinus rhynchophylla*, and *Juglans sinensis*, as well as other deciduous forests (deciduous broad-leaved trees), and mixed coniferous and deciduous forest.

The average H' value at the different spatial scales ranged from 0.991 (s3) to 1.371 (s9). These values were similar to the average H' value calculated for areas from Hyangnobong to Gitdaebaggybong in the Baekdudaegan region of Gangwon Province by Jeong and Oh.[44] Gabriel *et al.*[45] reported that there were significant differences in the tree species diversity distribution according to the spatial scale. In this study, the average H' increased with the spatial scale and was 1.37 times higher at s9 than at s3. Moreover, the minimum H' increased from s7, but the maximum H' hardly changed with the spatial scale.

The average H' was higher in the northern and southern regions, whose elevation and slope were higher than those of the central region regardless of the spatial scale. In particular, in the southern region, H' had the highest value of 1.879 at s9, similar to the value of H' calculated by Hwang *et al.*[28] as shown in Table 2 and Fig. 4.

3.2   Selection of major factors affecting *H'*

The correlation coefficients between the major factors, except for that between the slope and the relief, ranged from 0.005 to −0.511. The correlation coefficient between the slope and the relief was 0.811 but the variance inflation factors (VIFs) for the slope and relief factors were lower than 10 and multicollinearity did not occur, as shown in Table 3. Therefore, we included all 10 factors when constructing the regression model.

3.3   Evaluating model fitness

In this study, the fitness of the OLS and GWR models was evaluated to analyze the spatial relationship between the dependent and independent variables. OLS is a global model that estimates regression coefficients at the overall level, and if there is geographic irregularity, it cannot distinguish spatial changes. GWR is a regional model that is used to analyze the difference according to the spatial location.

By comparing the fitness between OLS and GWR, it was determined which model was more suitable for analyzing the spatial relationship between the dependent and independent variables. The coefficient of determination of the OLS and GWR models increased with the spatial scale. In the case of s9, the coefficient of determination was 0.41 for the GWR model, which was 3.4 times higher than that in the OLS model. Also, the coefficient of determination was higher for GWR than for OLS regardless of the spatial scale. This was consistent with the results of other studies. Liu *et al.* used GWR to determine the changes in microregions.[23]

The RMSE tended to decrease for both models as the spatial scale increased. Also, the RMSE was lower for the GWR model than for the OLS model regardless of the spatial scale,

Table 2
Summary of statistics of H′ at different spatial scales.

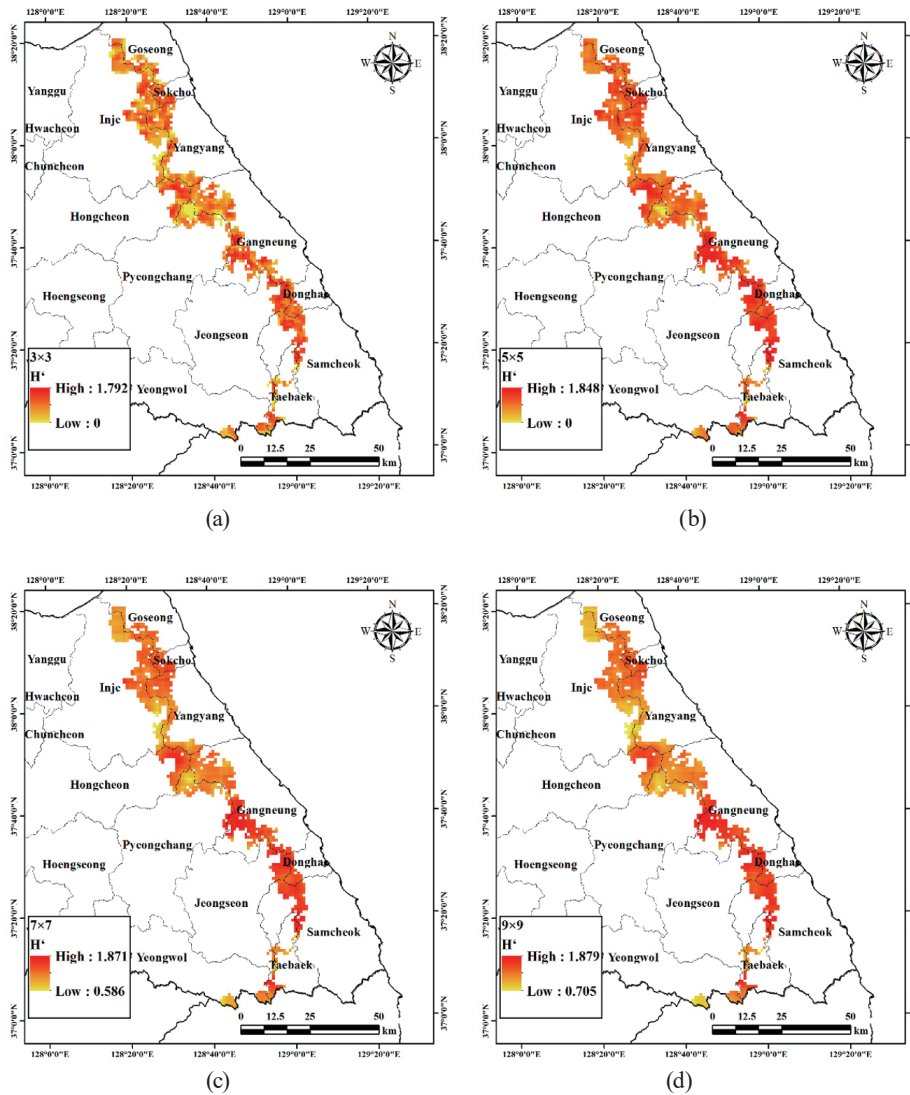| Spatial scale	Min.	Max.	Avg.	St. Dev.
s3	0.000	1.792	0.991	0.333
s5	0.000	1.848	1.214	0.285
s7	0.586	1.871	1.309	0.248
s9	0.705	1.879	1.361	0.230

Fig. 4.   (Color online) Distribution maps of *H′* for four spatial scales in the Baekdudaegan Protected Area of Gangwon Province. (a) s3, (b) s5, (c) s7, and (d) s9.

Table 3
Correlation between six topographic and four geographic factors in the Baekdudaegan Protected Area of Gangwon Province.

	Elevation	Slope	North slope	Curvature	Wetness	Relief	D_water	D_road	D_forest_road	D_urban	VIF
Elevation	1										1.698
Slope	−.158***	1									6.449
North slope	−.168***	.099***	1								1.055
Curvature	.475***	−.085***	−.052*	1							1.873
Wetness	−.338***	−.478***	.026	−.511***	1						2.463
Relief	−.062**	.881***	.092***	−.026	−.388***	1					4.880
D_water	.205***	−.061**	−.007	.012	−.026	−.045	1				1.102
D_road	.229***	−.011	.030	−.003	.048*	.021	.046	1			1.082
D_forest road	.050*	.018	−.006	−.029	.005	.034	.185***	.046	1		1.047
D_urban	.097***	−.068**	.074***	−.025	.063**	−.073***	.117***	.022	.054*	1	1.050

*$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$

and the difference in RMSE between the models increased with the spatial scale. In addition, the AICc of the models tended to decrease as the spatial scale range increased; it was lower for GWR than for OLS. Jo[46] and Kim and Jun[38] reported the improvement factors evaluated on the basis of the difference in AICc, and in the present study, as the spatial scale increased, the difference in AICc between models occurred, indicating that GWR performed better than OLS.

Moreover, the Moran's *I* of the two models increased with the spatial scale. In the case of s3, Moran's *I*, values of the OLS and GWR models were close to 0, and thus, there was no spatial autocorrelation. Similarly to that reported by Kim *et al.*,[47] Moran's *I* was lower in GWR than in OLS and the vegetation distribution characteristics were differently composed with clusters. Therefore, the GWR model is more suitable when considering the spatial effect in the current study area, as shown in Table 4.

3.4   Dependence of distribution of local R^2 and residual values on the GWR results

3.4.1   Distribution of local R^2 of tree diversity index calculated by incorporating six topographic factors and four geographic factors

The main advantage of the GWR model is that the local R^2 and residual values can be determined for each feature,[48] and the coefficient of determination was used to measure the fitness of the regression equations at each spatial scale. When the coefficient of determination is near 1, ten factors have a strong effect on H'.[49] Figure 5 shows the distribution of the local R^2 at each spatial scale.

The distributions of local R^2 values were calculated using each grid, which is the spatial unit in the GWR model, as shown in Fig. 5. The difference between the maximum and minimum local R^2 values was the highest (0.46) for the s9 spatial scale. The differences in the R^2 values for the s3, s5, and s7 spatial scales were 0.11, 0.24, and 0.40, respectively, indicating that as the spatial scale increased, the difference in the coefficient of determination also increased.

In terms of spatial distribution characteristics, the coefficient of determination of the southern region (Samcheok, Taebaek, and Yeongwol) was the highest in s3, but that of the central region (Gangneung and Pyeongchang) was the highest in s5 and s7. Moreover the local R^2 of s9 was high in the southern region (Taebaek and Yeongwol) as well as in the central region (Gangneung and Pyeongchang).

Table 4
Comparison of model fitting between the OLS model and the GWR model using R^2, RMSE, AICc, and Moran's *I*.

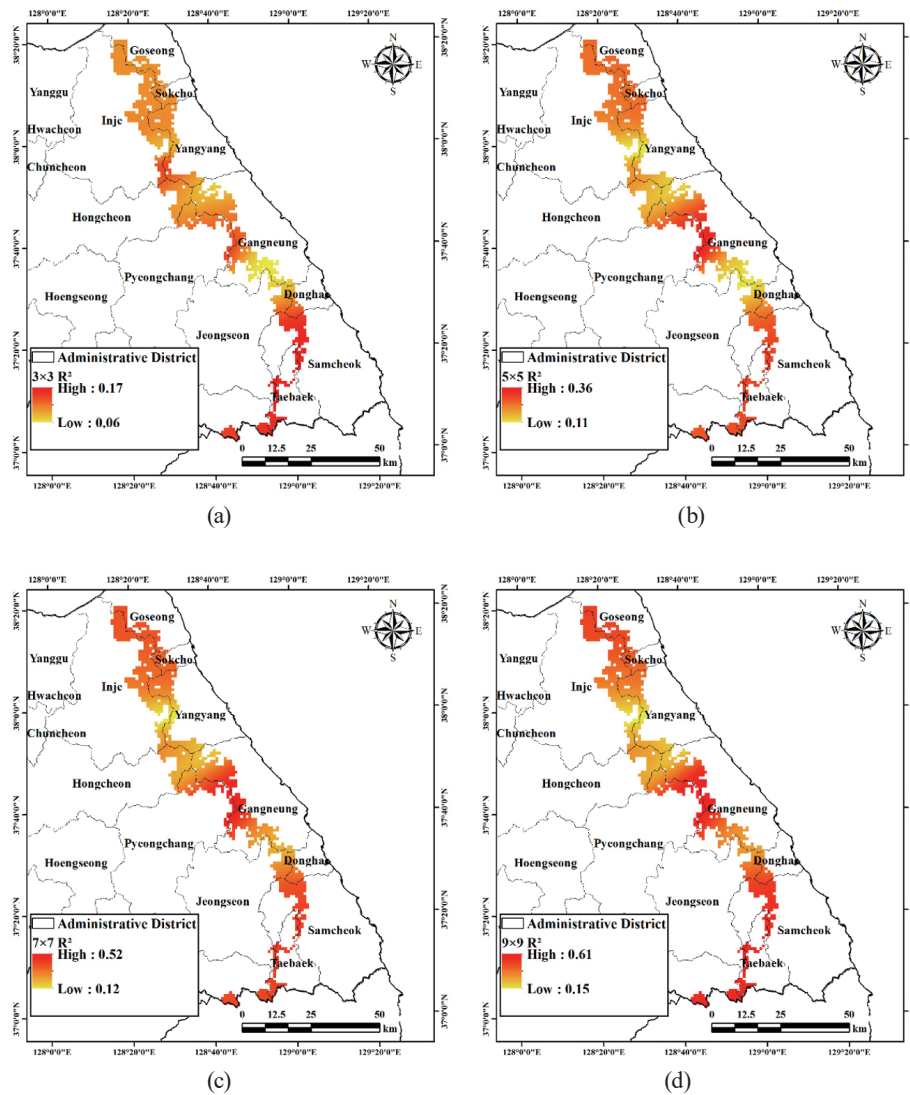| Spatial scale	OLS				GWR			
	R^2	RMSE	AICc	Moran's *I* (Z-Score)	R^2	RMSE	AICc	Moran's *I* (Z-Score)
s3	0.02	0.3281	797	0.1931 (47.19)	0.11	0.2974	620	0.0916 (22.49)
s5	0.06	0.2754	353	0.3741 (93.69)	0.22	0.2207	−137	0.1609 (39.38)
s7	0.09	0.2364	−35	0.5298 (129.12)	0.33	0.1634	−899	0.2102 (51.38)
s9	0.12	0.2146	−281	0.6345 (154.57)	0.41	0.1290	−1500	0.2773 (67.71)

Fig. 5.   (Color online) Distribution map of local R^2 values calculated from the GWR model using four spatial scales in the Baekdudaegan Protected Area of Gangwon province. (a) s3, (b) s5, (c) s7, and (d) s9.

3.4.2   Distribution of local R^2 of H' calculated by separating each factor

In this study, the tree species diversity distribution obtained using the GWR model depended on the region owing to the effects of topographic and geographic factors.

The correlation was determined by regression analysis to examine the effects of topographic and geographic factors on the tree species diversity. Table 5 gives the regression coefficient distribution of the six topographic factors and four geographic factors affecting H', which includes the coefficients, average, minimum, and maximum values of each factor in the study area.

Table 5
Relationships of species diversity index with six topographic factors and four geographic factors according to spatial scales using GWR model.

Statistics	s3			s5			s7			s9		
	Min.	Max.	Avg.	Min.	Max.	Avg.	Min.	Max.	Avg.	Min.	Max.	Avg.
Intercept	0.70516	1.49188	1.10347	0.85179	1.98074	1.38954	0.87986	2.21285	1.49941	0.89178	2.37775	1.56644
Elevation	−0.00029	0.00023	−0.00009	−0.00033	0.00023	−0.00010	−0.00039	0.00035	−0.00009	−0.00042	0.00042	−0.00007
Slope	−0.01016	0.03539	0.00763	−0.01241	0.02691	0.00187	−0.01740	0.02497	−0.00189	−0.01441	0.02065	−0.00207
North slope	−0.09988	0.10645	−0.01051	−0.07687	0.06050	−0.00447	−0.05099	0.02830	0.00332	−0.03872	0.02602	0.01030
Curvature	−0.49519	1.09952	0.51848	−0.95276	0.55400	0.12887	−0.35150	0.60976	0.15992	−0.46532	0.60145	0.10466
Wetness	−0.05008	0.06741	0.01135	−0.03436	0.06742	0.00465	−0.03699	0.05851	−0.00275	−0.02667	0.04744	−0.00190
Relief	−0.00310	0.00283	0.00006	−0.00304	0.00313	0.00058	−0.00309	0.00306	0.00076	−0.00284	0.00259	0.00057
D_water	−0.00004	0.00004	0.00000	−0.00005	0.00004	−0.00001	−0.00005	0.00004	−0.00001	−0.00005	0.00004	−0.00001
D_road	−0.00004	0.00001	−0.00001	−0.00006	0.00001	−0.00001	−0.00005	0.00001	−0.00001	−0.00004	0.00001	−0.00001
D_forest road	−0.00006	0.00002	0.00000	−0.00003	0.00003	0.00001	−0.00004	0.00004	0.00000	−0.00004	0.00005	0.00000
D_urban	−0.00003	0.00001	−0.00001	−0.00004	0.00001	−0.00001	−0.00005	0.00001	−0.00002	−0.00005	0.00001	−0.00002

The GWR model was used to determine H', which is a variable dependent on the topographic and geographic factors. H' was calculated to determine which topographic and geographic factors affected the tree species diversity.

The average coefficient of regression for elevation was low for all four spatial scales with values of less than –0.0007. The coefficients of regression for the average slope were 0.00763 and 0.00187 for s3 and s5, respectively, indicating that H' increased with the slope. For s7 and s9, the coefficients of regression for the average slope were −0.00189 and −0.00207, respectively, demonstrating that H' decreased with decreasing slope. The coefficients of regression for the average northern slope were −0.01051 and −0.00447 for s3 and s5, respectively, indicating that H' increased as far as the northern slope of high value. For s7 and s9, the coefficients of regression were 0.00332 and 0.01030, respectively, indicating that H' increased as close to the northern slope of high value. For the curvature, the average coefficient of regression was positive for all four spatial scales (s3, 0.51848; s5, 0.12887; s7, 0.15992; and s9, 0.10466), showing that H' increased with the curvature. By comparing the curvature with the other factors, the average coefficient of regression for the curvature was high and had the greatest impact on H'. The coefficients of regression for average wetness were 0.01135 and 0.00465 for s3 and s5, respectively. H' increased with the wetness, whereas for s7 and s9, H' increased with decreasing wetness. The coefficient of regression for the average relief had positive values for all four spatial scales (s3, 0.00006; s5, 0.00058; s7, 0.00076; and s9, 0.00057), indicating that H' increased with the difference in the elevation of the surrounding terrain. However, the effect of elevation on H' was minimal. The coefficient of regression for the four geographic factors of D_water, D_road, D_forest road, and D_urban ranged from −0.00002 to 0.00000, showing that their effects on H' were negligible.

The major topographic factors determining H' for the GWR model were curvature, northern slope, slope, and wetness. In particular, the curvature and northern slope were positive, and when the curvature and northern slope increased or when the slope and wetness decreased, H' increased. The relief was also positive, and when the relief increased, H' also increased. However, the regression coefficient of the relief is low, meaning that it has a negligible effect on H'. The effect of geographic factors was lower in the H' prediction model.

Our study found that topographic factors significantly correlate with tree species diversity, proving the practicality of topographic factors for studying the tree species diversity distribution for each grid (region).

On the basis of the above discussion, we can predict how tree species diversity will change with topographic factors. In Table 5, each number gives the impact of a factor on the dependent variable H' for the corresponding grid unit (region).

Figure 6 shows the regional coefficient distribution of the explanatory variables for tree species diversity calculated by the regression formula at spatial scale s9. The effect of each coefficient on tree species diversity varies from region to region. The coefficient for each factor has positive and negative values, and the magnitude of the coefficient is different for each grid (region).

The northern region of the study area has high elevation and slope. As the elevation and slope increased, H' also increased. In contrast, the central and southern regions have low values, which led to low H'. In addition, the northern slope and curvature are high in most areas, and as the northern slope and curvature increased in most areas of the study area, H' also increased.
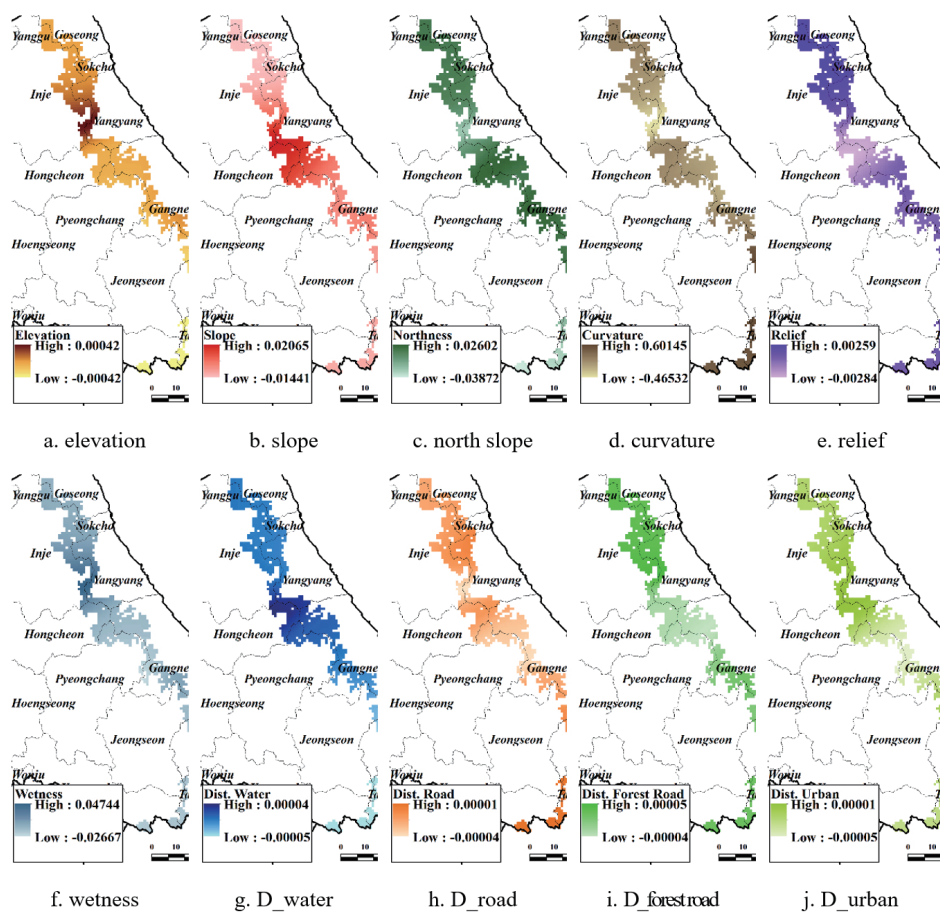

Fig. 6. (Color online) Distribution maps of local R^2 values calculated from the GWR model at spatial scale s9 in the Baekdudaegan Protected Area of Gangwon province.

From the effects of each factor on R^2 for the s9 spatial scale, which had the highest R^2 among the four spatial scales, the elevation, wetness, and D_water had the greatest effects in the northern areas; the slope and northern slope had the greatest effects in the central area; the curvature and D_road had the greatest effects in the southern area; and the relief, D_forest road, and D_urban had the greatest effects in the northern and southern areas. These results indicate that H' was affected differently according to differences in the regional distribution of topographic and geographic factors, as shown in Fig. 6.

4.   Conclusions

This study was carried out to determine the spatial distribution characteristics and effects of topographic and geographic factors on H' at four spatial scales (s3, s5, s7, and s9) in the Baekdudaegan Protected Area using the GWR model.

The method of calculating H' at each spatial scale using a forest-type map was effective for understanding and displaying the spatial distribution characteristics of species diversity in the Baekdudaegan Protected Area. In addition, OLS and GWR were compared in terms of their effectiveness in determining the relationship between topographic and geographic factors, and species diversity. Two models were compared using R^2, RMSE, AICc, and Moran's I.

The GWR model was more suitable for examing the relationship of H' with topographic and geographic factors because R^2 was higher in the GWR model than in the OLS model, and AICc and RMSE were lower in the GWR model than in the OLS model. Because the GWR model shows the variations of properties caused by spatial specificity, it is possible to determine the variations of species diversity depending on the relationship of variables. By calculating H' at each spatial scale, the coefficient of determination between the 10 factors and H' was obtained, and the effect of each factor on tree species diversity and the effects of the different regions were determined.

In a future study, if the variables affecting tree species diversity are predicted for each grid (region) through the GWR model, or if other factors are added to the factors proposed in this study, it will be important to increase the efficiency of the investigation of tree species diversity by identifying the spatial changes in the region. Localized and detailed information of factors affecting tree species diversity can be useful for the management and planning of the Baekdudaegan Protected Area by estimating the distribution of tree species diversity in each region, which is expected to be important for improving the forest protection system.

Acknowledgments

References

1   S. K. Kim: Public Land Law Rev. **60** (2013) 449. http://uci.or.kr/G704-000326.2013.60..015
2   Korea Forest Service: Basic Biodiversity Plan 2013–2017 (Korea Forest Service, Daejeon, 2011) p. 7.
3   Ministry of Environment: 12th Meeting of the Conference of the Parties to the Convention on Biological Diversity (COP 12) Pyeongchang, Republic of Korea: Guidebook (Ministry of Environment, Sejong, 2014) p. 1.
4   H. S. Cho, S. D. Lee, and M. J. Kim: Korean. J. Environ. Ecol. **26** (2012) 770. http://uci.or.kr/G704-000479.2012.26.5.017
5   Ministry of Environment: The Biodiversity of Korea (Ministry of Environment, Sejong, 2012) p. 24.
6   J. H. Lawton, S. Nee, A. J. Letcher, and P. H. Harvey: Large Scale Ecology and Conservation Biology (Blackwell Scientific Publications, Boston, 1994) p. 41.
7   S. Ferrier: Syst. Biol. **51** (2002) 331. https://doi.org/10.1080/10635150252899806
8   Y. H. Ahn, Y. C. Yang, and S. H. Chun: Korean J. Environ. Ecol. **15** (2001) 213. http://db.koreascholar.com/article.aspx?code=262204
9   D. S. Kang and W. K. Paik: Korean J. Plant Tax. **37** (2007) 41. http://uci.or.kr/G704-000487.2007.37.1.005
10 K. S. Lee and D. S. Cho: Korean J. Environ. Ecol **23** (2000) 255. https://kiss.kstudy.com/thesis/thesis-view.asp?key=2353376
11 I. H. Park and Y. K. Seo: Korean J. Environ. Ecol **14** (2003) 296. http://kiss.kstudy.com/thesis/thesis-view.asp?key=2686028
12 C. B. Lee, J. H. Chun, and H. H. Kim: J. Agric. Life Sci. **48** (2014) 93. http://doi.org/10.14397/jals.2014.48.3.93
13 H. S. Kwon, C. W. Seo, and C. H. Park: Korea Soc. Geospatial Inf. Syst. **20** (2012) 11. http://doi.org/10.7319/kogsis.2012.20.3.011
14 F. Corsi, E. Duprè, and L. Boitani: Conserv. Biol. **13** (1999) 150. https://doi.org/10.1046/j.1523-1739.1999.97269.x
15 C. H. Graham and R. J. Hijmans: Global Ecol. Biol. **15** (2006) 578. https://doi.org/10.1111/j.1365-2486.2006.01256.x
16 M. D. Han and G. B. Lee: Korean J. Environ. Ecol. **22** (2008) 347. http://db.koreascholar.com/article.aspx?code=262521
17 W. D. Sim, J. W. Park, and J. S. Lee: J. Forest Environ. Sci. **30** (2014) 91. http://doi.org/10.7747/JFS.2014.30.1.91
18 D. G. Jo: Survey Res. **10** (2009) 1. http://uci.or.kr/G704-001342.2009.10.3.006
19 D. J. Choi and Y. C. Suh: Korea Soc. Geospatial Inf. Syst. **20** (2012) 57. http://doi.org/ 10.7319/kogsis.2012.20.3.057
20 D. A. Griffith and L. J. Layne: A Casebook for Spatial Statistical Data Analysis: a Compilation of Analyses of Different Thematic Data Set (Oxford University Press, New York, 1999) p. 524.
21 H. W. Choi, K. H. Kim, and C. Y. Lee: Korean Assoc. Geogr. Inf. Stud. **10** (2007) 60. http://uci.or.kr/G704-000890.2007.10.4.002
22 W. Quan, N. Jian, and J. Tenhunen: Global Ecol. Biogeogr. **14** (2005) 379. https://doi.org/10.1111/j.1466-822X.2005.00153.x
23 C. Liu, L. Zhang, F. Li, and X. Jin: J. Forest. Res. **25** (2014) 269. https://doi.org/10.1007/s11676-014-0458-x
24 W. Xugao, H. Zhanqing, Y. Ji, Z. Jian, L. Buhang, and Y. Xiaolin: Ecol. Res. **23** (2008) 709. https://doi.org/10.1007/s11284-007-0430-8
25 Korea Forest Service: The Ecological Map of the Baekdudaegan Mountains (Korea Forest Service, Daejeon, 2013) p. 6.
26 Korea Forest Service: The Baekdudaegan Mountains Forest Biodiversity R&D Center (Korea Forest Service, Daejeon, 2014) p. 884.
27 H. S. Cho, G. T. Kim, and G. C. Choo: Korean J. Environ. Ecol. **19** (2005) 55. http://uci.or.kr/G704-000479.2005.19.1.007
28 K. M. Hwang, J. M. Lee, and J. H. Kim: J. Agric. Life Sci. **46** (2012) 41. http://uci.or.kr/G704-001926.2012.46.4.012
29 Y. S. Kim, N. Y. Kim, Y. S. Kim, H. B. Lee, S. C. Kim, J. H. Kim, and W. G. Park: J. Korean For. Soc. **101** (2012) 226. http://uci.or.kr/G704-000375.2012.101.2.017
30 S. D. Lee, S. H. Hong, and J. S. Kim: Korean J. Environ. Ecol. **26** (2012) 719. http://db.koreascholar.com/article.aspx?code=262821
31 G. G. Oh and S. G. Park: Korean J. Environ. Ecol. **15** (2002) 330. http://db.koreascholar.com/article.aspx?code=262212
32 M. Beals, L. Gross, and S. Harrell: The Institute for Environmental Modelling (TIEM) (2000) http://www.tiem.utk.edu/~gross/bioed/bealsmodules/shannonDI.html

33 R. Sagar, A. S. Raghubanshi, and J. S. Singh: For. Ecol. Manage. **186** (2003) 61. https://doi.org/10.1016/S0378-1127(03)00235-4
34 H. K. Song, S. K. So, M. Y. Kim, J. M. Park, S. H. Lee, and G. S. Park: Korean J. Environ. Ecol. **21** (2007) 82. http://uci.or.kr/G704-000479.2007.21.1.005
35 C. Song and M. Cao: Forests **8** (2017) 344. https://doi.org/10.3390/f8090344
36 A. S. Gonzalez and L. L. Mata: Divers. Distrib. **11** (2005) 567. https://doi.org/10.1111/j.1366-9516.2005.00186.x
37 S. Fotheringham, C. Brunsdon, and M. Charlton: Geographically Weighted Regression: the Analysis of Spatially Varying Relationships (Wiley, Hoboken, 2002) p. 282.
38 H. Y. Kim and C. M. Jun: Korean Assoc. Geogr. Inf. Stud. **15** (2012) 35. http://uci.or.kr/G704-000890.2012.15.2.004
39 B. Huang, B. Wu, and M. Barry: Int. J. Geogr. Inf. Sci. **24** (2010) 383. https://doi.org/10.1080/13658810802672469
40 Q. Zhou, C. Wang, and S. Fang: Atmos. Pollut. Res. **10** (2019) 835. https://doi.org/10.1016/j.apr.2018.12.012
41 C. M. Hurvich, J. S. Simonoff, and C. L. Tsai: J. R. Stat. Soc. **60** (1998) 271. https://doi.org/10.1111/1467-9868.00125
42 P. A. P. Moran: J. R. Stat. Soc. **10** (1948) 243. https://www.jstor.org/stable/2983777
43 L. Anselin: Geog. Anal. **27** (1995) 93. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x
44 B. K. Jeong and C. H. Oh: Korean J. Environ. Ecol. **27** (2013) 449. http://uci.or.kr/G704-000479.2013.27.4.009
45 D. Gabriel, I. Roschewitz, T. Tscharntke, and C. Thies: Ecol. Appl. **16** (2006) 2011. https://doi.org/10.1890/1051-0761(2006)016[2011:BDADSS]2.0.CO;2
46 D. G. Jo: Korea J. Popul. Stud. **32** (2009) 1. http://uci.or.kr/G704-000152.2009.32.3.008
47 K. M. Kim, Y. H. Roh, and E. S. Kim: Korean Assoc. Geogr. Inf. Stud. **17** (2014) 69. http://doi.org/10.11108/kagis.2014.17.4.069
48 C. D. Kang: Korea Real Estate Rev. **20** (2010) 107. http://uci.or.kr/G704-001886.2010.20.2.010
49 J. Gao and S. Li: Appl. Geogr. **31** (2011) 292. https://www.sciencedirect.com/science/article/pii/S0143622810000676?via%3Dihub

About the Authors

Jeongmook Park received his B.S, M.S, and Ph.D. degrees from Kangwon National University, Korea, in 2011, 2013, and 2019, respectively. Since 2019, he has been a postdoctoral researcher at Kangwon National University. His research interests include forest spatial analysis using GIS and RS. (jm_park@kangwon.ac.kr)

Byoungkoo Choi received his B.S. and M.S. degrees from Kangwon National University, Republic of Korea, in 2002 and 2004, respectively, and his Ph.D. degree from Mississippi State University, USA, in 2011. From 2014 to 2015, he was a research scientist at the National Institute of Ecology, Republic of Korea. Since 2015, he has been an assistant professor at Kangwon National University. His research interests include ecohydrology, watershed management, and forestry BMPs. (bkchoi@kangwon.ac.kr)

Jungsoo Lee received his B.S. degree from Kangwon National University, Republic of Korea, in 1998 and his M.S. and Ph.D. degrees from Tokyo University, Japan, in 2000 and 2003, respectively. Since 2006, he has been a professor at Kangwon National University. His research interests include GIS and remote sensing related to forest management. (jslee72@kangwon.ac.kr)