Evaluation of Parameter Estimation Methods for Fitting Spatial Regression Models

Junfeng Lu and Lianjun Zhang

Abstract: Two types of spatial regression models, a spatial lag model (SLM) and a spatial error model (SEM), were applied to fit the height–diameter relationship of trees. SEM had better model fitting and performance than both SLM and ordinary least squares. Moran’s $I$ coefficients showed that SEM effectively reduced the spatial autocorrelation in the model residuals. Both real data and Monte Carlo simulations were used to compare different parameter estimation methods for the two spatial regression models, including maximum likelihood estimation (MLE), Bayesian methods, two-stage least squares (for SLM) and generalized method of moments (GMM) (for SEM). Our results indicated that GMM was close to MLE in terms of model fitting, much easier in computation, and robust to non-normality and outliers. The Bayesian method with heteroscedasticity did not effectively estimate the spatial autoregressive parameters but produced very small biases for the regression coefficients of the model when few outliers existed. For. Sci. 56(5):505–514.

Keywords: tree height–diameter relationship, spatial autoregressive parameters, maximum likelihood estimation, Bayesian methods, two-stage least squares, generalized method of moments

Because of the rapid development of geographical information techniques, numerous data sets have been collected with references to locations measured as points in space in different study fields. Two problems arise for the geo-referenced sample data: spatial dependence existing among the observations and spatial heterogeneity occurring in the relationships among variables (Anselin and Griffith 1988, LeSage and Pace 2009). It is known that spatial dependence and heterogeneity cause the violations of traditional Gauss-Markov assumptions, which have drawn great attention from forest and ecological modelers (Zhang et al. 2005, 2009). In recent years, spatial regression models have been developed to take the spatial dependence into account and have been widely used in spatial econometrics (LeSage and Pace 2009), whereas local models such as geographically weighted regression and drifted analysis of regression parameters have been used to deal with the spatial heterogeneity (Fotheringham et al. 2002).

The spatial dependence or autocorrelation can be described in different ways in the context of regression modeling, including first-order autoregressive model, mixed regressive-spatial model, and regression model with spatial autocorrelation in the disturbances (LeSage 1998, Anselin 1999). The latter two are commonly called spatial lag model (SLM) and spatial error model (SEM), respectively. Another model is referred to as the general spatial autoregressive model, which includes both a spatial lagged term and a spatially correlated error structure (also called a spatial autoregressive model mixed model).

A rich body of literature on the theoretical discussion of the parameter estimation methods for spatial regression models is available. A maximum likelihood estimator was proposed by Ord (1975) and is still the most widely used method to date. Maximum likelihood estimation (MLE) depends on the assumption that the model errors are normally distributed. Although MLE has the best efficiency among all available estimators when the assumption of error distribution is met, it requires complex computation. Unlike classic regression models, the joint log likelihood for a spatial regression model does not equal the sum of the log likelihoods associated with the individual observations because of the two-directional nature of the spatial dependence, which results in a Jacobian term that is the determinant of a full $n \times n$ matrix, where $n$ is the sample size (Anselin 1999). Spatial two-stage least squares was proposed to make the computation much easier. It was shown that spatial two-stage least squares achieved the consistency and asymptotic normality properties of the standard two-stage least squares (2SLS) (Anselin 1999). Generalized method of moments (GMM) also attracted much attention from researchers. Kelejian and Prucha (1999) showed that GMM was virtually as efficient as MLE while requiring much less computation. Neither 2SLS nor GMM requires the normality of model error distributions. Another approach that is also robust to non-normality is Bayesian methods. It is well known that the Bayesian methods implemented with diffuse prior information can replicate the results of maximum likelihood estimation (LeSage 1998). Furthermore, LeSage (1998) developed a series of Bayesian spatial autoregressive models to deal with spatial heteroscedasticity, another phenomenon violating the assumption of traditional regression models.

Spatial autocorrelation is frequently encountered in ecological data. Many ecological theories and models implicitly assume an underlying spatial pattern in the distributions of organisms and their environment (Legendre and
Spatial autocorrelation is problematic for classic statistical methods such as analysis of variance and ordinary least squares (OLS) regression, which assume independently distributed model errors (Haining 1990, p. 161–166, Legendre 1993). When the response variable is spatially autocorrelated, the assumption of independence is often invalid, and the effects of covariates that are themselves autocorrelated tend to be exaggerated (Gumpertz et al. 1997). Thus, statistical methodologies need to be developed to explicitly incorporate the observed dependence among model errors (Burkhart and Gregoire 1994, Dale et al. 2002).

There have been a number of applications of spatial regression models in ecology in recent years. By using artificial data sets containing virtual organism and environmental variables, Kissling and Carl (2007) showed that SEM was the most reliable and performed well in all cases. In the studies of relationships between environmental variables and species richness, Lichstein et al. (2002) and Tognelli and Kelt (2004) showed that both the conditional autoregressive model and SAR had better model fitting and produced model residuals less spatially autocorrelated than OLS. In forestry, spatial regression models were used as a means of modeling the micro-site-induced spatial dependence prevalent in forest surveys and field experiments (Whittle 1954, 1962, Besag 1974). Fox et al. (2007) showed that the spatial autoregressive moving average model (SARMA) was effective in modeling the spatial dependence observed in the residuals of individual tree growth models. Zhang et al. (2009) compared SLM and SEM against linear mixed models and geographically weighted regression models.

To date, however, the parameter estimation methods for spatial regression models mentioned above were mostly confined to the theoretical field. In ecology and forestry, research has been focused on the specification of spatial regression models rather than the selection of estimation methods. Although the parameter estimation methods might not make differences as significant as model specification, different estimation methods do require different assumptions, and, thus, are suitable to different data and situations. In this study we attempted to apply spatial regression models (e.g., SLM and SEM) to fit the relationship between heights and diameters of individual trees using both real and simulated data sets for the comparisons of different parameter estimation methods including MLE, GMM, 2SLS, and Bayesian methods.

Theoretical Background

Spatial Regression Models

In this study, spatial regression models include SLM and SEM. An SLM is expressed as follows (Anselin 1999):

\[ y = \rho Wy + X\beta + \epsilon, \]

where \( y \) is a vector of the response variable, \( X \) is a matrix of the explanatory variables, \( \beta \) is a vector of the regression coefficients, \( \rho \) is a spatial autoregressive coefficient, and \( \epsilon \) is a vector of model error terms following \( N(0, \sigma^2 I) \), and \( W \) is a spatial weight matrix. The estimate of \( \rho \) can be considered as an indicator of spatial autocorrelation and is conditional on \( W \) (Anselin 2001).

An SEM is a special case of a regression with a non-spherical error term, in which the off-diagonal elements of the covariance matrix express the structure of spatial dependence. SEM is specified as

\[ y = X\beta + \xi \quad \text{and} \quad \xi = \lambda Wy + \epsilon, \]

where \( \lambda \) is a spatial autoregressive coefficient. However, unlike \( \rho \) in Equation 1, \( \lambda \) is considered as a nuisance parameter, usually of little interest per se, but necessary to correct for the spatial dependence (Anselin and Getis 1993, Anselin 2001).

There are many ways to define the spatial weight matrix \( W \). The elements of \( W \) are nonstochastic and exogenous to the models (Anselin 1999). The most commonly used \( W \) matrix is a binary matrix based on geographic arrangement of the observations, or contiguity. The spatial weights can be also based on distance decay (e.g., inverse distance or inverse distance squared) (Anselin 1980), the structure of a social network (Doreian 1980), an economic distance (Case et al. 1993), k nearest neighbors (Pinkse and Slade 1998), empirical flow matrices (Murdoch et al. 1997), or trade-based interaction measures (Aten 1996, 1997). For the binary matrix the main task is to quantify the spatial continuity. The available options include linear continuity, Rook continuity, Bishop continuity, double linear continuity, double Rook continuity, Queen continuity (LeSage 1998), and Delaunay triangulation (Smirnov and Anselin 2001).

MLE

MLE of SLM and SEM was first outlined by Ord (1975). The underlying assumption is the normal distribution of model errors, i.e., \( \epsilon \sim N(0, \sigma^2 I) \). The log-likelihood function of SLM is (Anselin 1999)

\[ \ln L = \ln |I - \rho W| - \frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{(y - \rho Wy - X\beta)'(y - \rho Wy - X\beta)}{2\sigma^2}. \]

The log-likelihood function of SEM is (Anselin 1999)

\[ \ln L = \ln |I - \lambda W| - \frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{(y - X\beta)'(I - \lambda W)(y - X\beta)}{2\sigma^2}. \]

GMM

A number of approaches were outlined to estimate the coefficients of SEMs under the framework of the method of moments. Kelejian and Prucha (1999) developed a set of moment conditions that yielded the estimation equations for the parameters of SEM. Specifically, if we assume i.i.d.
of the model errors, the following three conditions readily follow:

\[ E \left( \frac{\varepsilon' \varepsilon}{n} \right) = \sigma^2 \]

\[ E \left( \frac{\varepsilon' W W \varepsilon}{n} \right) = \sigma^2 \left( \frac{1}{n} \right) \text{tr}(W W) \]  

(5)

\[ E \left( \frac{\varepsilon' W \varepsilon}{n} \right) = 0. \]

This yields a system of three equations in the parameters \( \lambda, \lambda^2, \) and \( \sigma^2 \) (Anselin 1999). Kelejian and Prucha (1999) suggested the use of nonlinear least squares to obtain the consistent generalized moment estimator from this system.

### 2SLS

The endogeneity of the spatially lagged dependent variable can also be addressed by means of instrumental variables or a 2SLS approach (Anselin 1999). This results in the 2SLS estimator for SLM. Arraiz et al. (2008) showed that 2SLS produced consistent estimates when the disturbances were heteroskedastic, whereas maximum likelihood produced inconsistent estimates. The 2SLS estimator can be produced in three steps:

1. Obtain the consistent estimates of \( \beta \) by the instrumental variables, where \( X, WX, \) and \( W^2X \) are the instruments in SLM.
2. Estimate \( \rho \) and \( \sigma \) by GMM using the samples constructed from the functions of model errors.
3. Use the estimates of \( \rho \) and \( \sigma \) to perform a spatial Cochrane-Orcut transformation of the data and obtain more efficient estimates of \( \beta \).

### Bayesian Methods

It is well-known that the Bayesian methods implemented with diffuse prior information can replicate the results of maximum likelihood estimation. The Bayesian methods have also been used to deal with heteroskedastic and leptokurtic disturbances (LeSage 1998). LeSage (1998) proposed a Bayesian extension of SLM (Equation 1) as

\[ \varepsilon \sim \text{N}(0, \sigma^2 V), \quad \text{where} \quad V = \text{diag}(v_1, v_2, \ldots, v_o) \]  

(6)

and

\[ \rho \sim \text{uniform}(-1, 1), \quad \sigma \sim \Gamma(m, k), \quad \sigma \sim \Gamma(v_0, d_0). \]

To implement a diffuse prior for \( \sigma \) we would set \( v_0 = 0 \) and \( d_0 = 0 \). The prior for \( r \) is based on a \( \Lambda(m, k) \) distribution, which has mean equal to \( mk \) and variance equal to \( mk^2 \).

Small values of \( r \) around 2–7 allow for nonconstant variances and are associated with a prior belief that outliers or nonconstant variances exist. Large values such as \( r = 20 \) or \( 50 \) would produce \( v_i \) estimates close to unity, forcing the model to take on a homoskedastic character and produce parameter estimates equivalent to those by MLE. The Bayesian extension of SEM (Equation 2) has a form similar to that of Equation 6, except that \( \lambda \) is set with the prior distribution uniform\((-1, 1)\) instead of \( \rho \sim \text{uniform}(-1, 1) \) for SLM (LeSage 1998).

### Measures of Model Fitting

The commonly used measure of model fitting, \( R^2 \), can assess the spatial regression models in terms of model prediction. When the models are estimated by MLE, a likelihood ratio test (LRT) can be used to test whether or not SLM or SEM makes a significant improvement in model fitting over OLS, or, in other words, whether or not the improvement warrants the additional computation needed for SLM or SEM (Haining 1990, p. 142–145):

\[ \text{LRT} = 2(\ln L_1 - \ln L_2), \]  

(7)

where \( L_1 \) and \( L_2 \) are the likelihoods of two models. The null hypothesis is \( H_0 : \rho = 0 \) (\( \lambda = 0 \) and the alternative hypothesis is \( H_1 : \rho \neq 0 \) (\( \lambda \neq 0 \)). This LRT statistic approximately follows a \( \chi^2 \) distribution with the degrees of freedom equal to the number of additional parameters in the more complex model (Haining 1990, p. 142–145). For testing either SLM versus OLS (\( H_0 : \rho = 0 \)) or SEM versus OLS (\( H_0 : \lambda = 0 \)), the degrees of freedom for the test is 1. It should be noted that higher likelihood does not mean higher prediction power as reflected by the model \( R^2 \).

### Test for Spatial Effects

Several maximum likelihood–based tests can be used for testing spatial autocorrelation in either model errors or the lagged variables, including the Wald test, likelihood ratio test, and Lagrange Multiplier (LM) test (Buse 1982). The LM test only requires the estimation of model under the null hypothesis. Furthermore, the LM test also allows for the distinction between a spatial error and a spatial lag alternative (Anselin 1999) and, therefore, can be used for the inferences of the spatial autoregressive coefficients. The LM test against a spatial error alternative was originally suggested by Burridge (1980) as

\[ \text{LM}_{\text{err}} = \frac{[\varepsilon' W \varepsilon (\varepsilon' W / n)]^2}{\text{tr}(W^2 + W W)} \]  

(8)

This statistic has an asymptotic \( \chi^2(1) \) distribution. Similarly, the LM test against a spatial lag alternative was outlined in Anselin and Griffith (1988) as

\[ \text{LM}_{\text{lag}} = \frac{[\varepsilon' W \varepsilon (\varepsilon' W / n)]^2}{D} \]  

(9)

where \( D = [WX \beta' (I - X(X'X)^{-1}X') WX \beta / \sigma^2] + \text{tr}(W^2 + W W) \). This statistic also has an asymptotic \( \chi^2(1) \) distribution.

### Moran’s I Coefficient

Spatial autocorrelation measures the similarity between sampling units for a given variable as a function of spatial distance (Sokal and Oden 1978a, 1978b, Griffith 1987, Legendre 1993, Rossi and Quénéhervé 1998). For quantitative or continuous variables, Moran’s \( I \) coefficient is the
most commonly used measure in univariate autocorrelation analyses (Diniz and Bini 2003):

\[
I = \left(\frac{n}{s}\right) \left[ \frac{\sum_{i} \sum_{j} (y_i - \bar{y})(y_j - \bar{y})w_{ij}}{\sum_{i} (y_i - \bar{y})^2} \right].
\]

(10)

where \(n\) is the number of sampling units, \(y_i\) and \(y_j\) are the observed values of variable, \(\bar{y}\) is the average of \(y\) across the region, and \(w_{ij}\) is the element of matrix \(W\), in which \(w_{ij} = 1\) if the observations \(i\) and \(j\) are within a given distance class interval, \(w_{ij} = 0\) otherwise, and \(s\) denotes the total number of entries (connections) in \(W\). Moran’s \(I\) ranges between \(-1\) and \(+1\) for maximum negative and positive autocorrelations, respectively. The geographical distances can be partitioned into discrete classes, creating successive \(W\) matrices and allowing computation of different Moran’s \(I\) values for the same variable. This allows one to evaluate the behavior of autocorrelation as a function of spatial distance in a form of graph called a spatial correlogram (Isaaks and Srivastava 1989).

Data and Methods

Data

The data used in this study were a part of the stem map data of a softwood stand located near Sault Ste. Marie, Ontario, Canada (Ek 1969). The stand was mature, second-growth, and uneven-aged. A plot of \(100 \times 100\) m in size with 659 trees was used. The average tree dbh was 17.9 cm (ranging from 10.2 to 74.2 cm), and the average total height was 13.1 m (ranging from 6.4 to 32.9 m). The position of every tree is recorded in terms of spatial coordinates. The species in the plot include white birch (Betula papyrifera Marsh.), white spruce (Picea glauca [Moench] Voss), black spruce (Picea mariana [Mill.] BSP), white pine (Pinus strobus L.), and balsam fir (Abies balsamea [L.] Mill.).

Implementation of Models

Both SLM and SEM were fitted to the plot data above. Because the scatterplots of the tree heights (HT) against diameters (dbh) were quadratic in shape, we chose the following model as the OLS model to fit the HT-dbh relationship:

\[
\ln(\text{HT}) = \beta_0 + \beta_1 \ln(\text{dbh}) + \epsilon,
\]

(11)

where \(\ln\) is natural logarithm, \(\beta_0\) and \(\beta_1\) are regression coefficients to be estimated, and \(\epsilon\) is the model error term. Model residuals were defined as the difference between the observed and predicted \(\ln(\text{HT})\). When the lagged \(\ln(\text{HT})\) term or autocorrelated error term was added to Equation 11, SLM or SEM was generated for the tree height–diameter model, respectively.

Parameter Estimation

It is usually desirable to know the properties of an estimator. Specifically, it is of interest to analyze the bias and variance of MLE, GMM, 2SLS, and Bayesian methods for the spatial regression models using Monte Carlo simulations. Kelejian and Prucha (1999) used an “idealized” spatial weight matrix as well as a real-world spatial weight matrix in their simulations for analyzing the properties of 2SLS and MLE. In this study, we used the tree height–diameter plot data and Delaunay triangulation to define the spatial continuity weight matrices. A Delaunay triangulation for a set of \(P\) points in the plane is a triangulation \(DT(P)\) such that no point in \(P\) is inside the circumcircle of any triangle in \(DT(P)\). The Delaunay triangulations maximize the minimum angle of all angles of the triangles in \(DT(P)\) (Smirnov and Anselin 2001). Actually, the Delaunay triangulation is also an option in ArcGIS for creating the spatial weights matrices.

Each parameter estimator has its unique properties. Theoretically, MLE is most efficient (producing the lowest variance) if the normality assumption is met. Both 2SLS and GMM are robust to non-normality by default and have been proved to be robust to heteroscedasticity. The Bayesian methods are designed to tackle outliers and heteroscedasticity. For the Bayesian methods, the value of \(r\) determines the prior distribution: a small \(r\) allows outliers and nonconstant variance, whereas a large \(r\) produces homoscedastic prior distribution. To compare the model fitting and performance under different circumstances, we set \(r = 2\) to take the possible heteroscedasticity of model errors into account, and the model errors \(\epsilon\) following \(N(0, \sigma^2 I)\), no parameter \(r\), and no prior distribution of \(\epsilon\); thus, the model errors were assumed to be homoscedastic.

Monte Carlo Simulations

In this study, we simulated three scenarios for SLM and SEM, respectively, with the sample size = 100 for each scenario: the data with the normal distribution of model errors, with the parameters set as \(\beta_0 = -1, \beta_1 = 1, \rho = 0.7\) (for SLM), or \(\lambda = 0.7, \alpha^2 = 0.5\); the above normal data with three outliers; and the data with the \(t\) distribution of model errors with \(df = 10\).

Results

Model Fitting

First, the OLS model was fitted to the tree height–diameter data. The \(R^2\) of OLS was 0.6336. The SLM and SEM models were then applied using different parameter estimation methods. Recall that 2SLS was used for SLM and GMM was used for SEM. The \(R^2\) of each model with different parameter estimation methods is shown in Table 1. The results indicated that SLM was worse than OLS in terms of model \(R^2\), whereas SEM fitted the data better (higher model \(R^2\)) than both OLS and SLM. However, for the same model (i.e., SLM or SEM), there was little difference among different parameter estimation methods. The \(R^2\) of MLE was very close to that of the Bayesian method with homoscedasticity. 2SLS and GMM, which are free of distributional assumption for the model errors, had model \(R^2\) comparable to that of MLE. The Bayesian method with heteroscedasticity had slightly smaller model \(R^2\) than other parameter estimation methods (Table 1).

Assuming there is no correlation between the model errors or no correlation between the lagged response
parameter estimates (i.e., with homoscedasticity produced similar results in the parameter estimation methods (Table 2). For the Bayesian method with homoscedasticity had similar estimates from those with other methods, whereas MLE and the Bayesian method with homoscedasticity had slightly different parameter estimates (Table 2). For example, the posterior distributions of \( \hat{\rho} \) and \( \hat{\gamma} \) from the two Bayesian methods are approximately normal despite the fact that the prior distributions were set as the uniform distribution. For example, the posterior distributions of \( \hat{\rho} \) and \( \hat{\gamma} \) from the Bayesian method with heteroscedasticity are illustrated in Figure 1a and b, respectively. The posterior distributions of \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) remained normal as defined by the prior distributions. For example, the posterior distributions of \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) for the Bayesian method with heteroscedasticity for the SEM model are illustrated in Figure 1c and d, respectively.

Figure 2 shows the mean of the 1,000 draws for the parameters \( v_i \) as defined in Equation 6. The large values of \( v_i \) can be used to detect the heteroscedasticity in some observations. It seems that the heteroscedasticity in SLM (Figure 2a) is relatively higher than that in SEM (Figure 2b) because there are several observations with the \( v_i \) values much larger than 50.

**LM Test**

The LM test was used to test OLS against the SLM and/or SEM alternatives. The testing statistic was 44.18 for the SLM alternative, which was much higher than the critical value \( \chi^2(1) = 17.61 \). Thus, the null hypothesis of no spatial lag dependence was rejected. Similarly, the testing

### Table 1. Model \( R^2 \) of spatial lag models (SLM) and spatial error models (SEM) by different parameter estimation methods

<table>
<thead>
<tr>
<th>Models</th>
<th>MLE</th>
<th>GMM (SLS)</th>
<th>Bayesian homoscedasticity</th>
<th>Bayesian heteroscedasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLM</td>
<td>0.613</td>
<td>0.617</td>
<td>0.613</td>
<td>0.607</td>
</tr>
<tr>
<td>SEM</td>
<td>0.692</td>
<td>0.692</td>
<td>0.693</td>
<td>0.686</td>
</tr>
</tbody>
</table>

MLE, maximum likelihood estimation; GMM (SLS), generalized method of moments (2-stage least squares).

### Table 2. Model parameter estimates of spatial lag models by different parameter estimation methods

<table>
<thead>
<tr>
<th>Model parameters</th>
<th>MLE</th>
<th>SLS</th>
<th>Bayesian homoscedasticity</th>
<th>Bayesian heteroscedasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho )</td>
<td>0.2710</td>
<td>-0.0949</td>
<td>0.2683</td>
<td>0.2158</td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>0.2831</td>
<td>1.2137</td>
<td>0.2909</td>
<td>0.4045</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0.5580</td>
<td>0.5591</td>
<td>0.5577</td>
<td>0.5711</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>0.0172</td>
<td>0.0193</td>
<td>0.0173</td>
<td>0.0069</td>
</tr>
</tbody>
</table>

MLE, maximum likelihood estimation; SLS, 2-stage least squares.

variables, MLE certainly produces the same results as OLS. The log-likelihood of OLS was 380.2, and the log-likelihoods of SLM and SEM were 627.9 and 650.8, respectively. Using LRT with the significance level \( \alpha = 0.05 \), we concluded that both SLM and SEM fitted the data significantly better than OLS (\( P < 0.0001 \)).

**Model Coefficients**

The estimates of two model coefficients of OLS were \( \hat{\beta}_0 = 0.9723 \) and \( \hat{\beta}_1 = 0.5588 \). The estimates of model coefficients are listed in Table 2 for SLM and Table 3 for SEM, respectively. The two-sample \( z \) test was used to test the differences of model coefficient estimates between OLS and SLM or SEM. It was evident that SLM had \( \hat{\beta}_1 \) significantly different from both OLS and SEM across different parameter estimation methods (Table 2), whereas the \( \hat{\beta}_0 \) of SEM was close to that of OLS for each parameter estimation method (Table 3). In contrast, the \( \hat{\beta}_1 \) of both SLM and SEM was not significantly different from that of OLS. In general, SEM had smaller \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) than OLS (Table 3). For the SLM model, 2SLS produced estimates for some parameters (i.e., \( \hat{\rho} \) and \( \hat{\beta}_0 \)) that were significantly different from those with other methods, whereas MLE and the Bayesian method with homoscedasticity had similar estimates. The Bayesian method with heteroscedasticity yielded slightly different parameter estimates (Table 2). For the SEM model, MLE, GMM, and the Bayesian method with homoscedasticity produced similar results in the parameter estimates (i.e., \( \hat{\gamma}, \hat{\beta}_0, \) and \( \hat{\beta}_1 \)). Again, the Bayesian method with heteroscedasticity yielded slightly different parameter estimates (Table 3).

Table 4 shows the standard errors of the parameter estimates \( \hat{\beta}_0, \hat{\beta}_1, \) and \( \hat{\rho} \) (for SLM) or \( \hat{\gamma} \) (for SEM) from different parameter estimation methods. For the SLM model, 2SLS produced much larger standard errors for the parameter estimates than other methods, whereas both Bayesian methods had smaller standard errors than MLE. For the SEM model, MLE yielded larger standard errors for the parameter estimates than other methods, except GMM had much larger standard error for \( \hat{\gamma} \) (Table 4).

The posterior distributions of \( \hat{\rho} \) and \( \hat{\gamma} \) from the two Bayesian methods are approximately normal despite the fact that the prior distributions were set as the uniform distribution. For example, the posterior distributions of \( \hat{\rho} \) and \( \hat{\gamma} \) from the Bayesian method with heteroscedasticity are illustrated in Figure 1a and b, respectively. The posterior distributions of \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) remained normal as defined by the prior distributions. For example, the posterior distributions of \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) for the Bayesian method with heteroscedasticity for the SEM model are illustrated in Figure 1c and d, respectively.

**LM Test**

The LM test was used to test OLS against the SLM and/or SEM alternatives. The testing statistic was 44.18 for the SLM alternative, which was much higher than the critical value \( \chi^2(1) = 17.61 \). Thus, the null hypothesis of no spatial lag dependence was rejected. Similarly, the testing
statistic was 116.97 for the SEM alternative and the null hypothesis of no dependence among model residuals can be rejected. Further, the LM test was performed to compare the SLM and SEM alternatives. The testing statistic was 402.42, implying that the SLM alternative did not account for the spatial dependence as effectively as the SEM alternative.

Residual Analysis

The spatial correlograms for the model residuals from OLS, SLM, and SEM are shown in Figure 3. When the lag distance is less than 25 m, the residuals from all models show positive spatial autocorrelation (i.e., positive Moran’s I coefficients). For large lag distance (>30 m), the residuals have negative Moran’s I coefficients. The Moran’s I coefficients are much larger at small spatial lag distance, say 5 m, and then approach zero after spatial lag distance 20 m. In general, SEM has much smaller spatial autocorrelation in the model residuals than either OLS or SLM for different parameter estimation methods. SLM has even larger Moran’s I coefficients than OLS except 2SLS. Among the parameter estimation methods for SEM, the Bayesian method with homoscedasticity produced the smallest Moran’s I coefficients, whereas the Bayesian method with heteroscedasticity yielded much larger Moran’s I coefficients than other three methods (Figure 3).

### Table 4. Estimates of standard errors of the parameters by different parameter estimation methods

<table>
<thead>
<tr>
<th>Model parameters</th>
<th>SLM MLE</th>
<th>SLM 2SLS</th>
<th>SLM Bayesian homoscedasticity</th>
<th>SLM Bayesian heteroscedasticity</th>
<th>SEM MLE</th>
<th>SEM GMM</th>
<th>SEM Bayesian homoscedasticity</th>
<th>SEM Bayesian heteroscedasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho )</td>
<td>0.0455</td>
<td>0.0742</td>
<td>0.0409</td>
<td>0.0371</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.0659</td>
<td>0.1197</td>
<td>0.0449</td>
<td>0.0489</td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>0.1478</td>
<td>0.1947</td>
<td>0.1140</td>
<td>0.1014</td>
<td>0.0577</td>
<td>0.0436</td>
<td>0.0430</td>
<td>0.0532</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0.0217</td>
<td>0.1697</td>
<td>0.0163</td>
<td>0.0140</td>
<td>0.0210</td>
<td>0.0151</td>
<td>0.0148</td>
<td>0.0186</td>
</tr>
</tbody>
</table>

Figure 1. The posterior distributions of model parameters from the Bayesian method with heteroscedasticity for (a) spatial autoregressive parameter \( \rho \) (SLM), (b) spatial autoregressive parameter \( \lambda \) (SEM), (c) regression coefficient \( \beta_0 \), and (d) regression coefficient \( \beta_1 \).
Monte Carlo Simulations

The bias and variance of each parameter estimator for the parameters $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\rho}$, or $\hat{\gamma}$ were computed for SLM (Table 5) and SEM (Table 6) using Monte Carlo simulation data sets. Three different distributions were used for the model errors, i.e., normal, normal with outliers, and $t$ distribution. Under the normality, MLE produced larger bias for $\hat{\rho}$ than 2SLS for the SLM model (Table 5), whereas MLE yielded much smaller bias for $\hat{\gamma}$ than GMM for the SEM model (Table 6). The Bayesian method with heteroscedasticity generally had larger biases for the spatial autoregressive parameters $\hat{\rho}$ or $\hat{\gamma}$. It should be noted that GMM produced positive bias for $\hat{\gamma}$, whereas the other methods had negative biases. The three estimators produced similar variances of the parameter estimates.

When few outliers existed, GMM or 2SLS and the Bayesian method with heteroscedasticity yielded smaller biases than MLE. GMM or 2SLS was good at estimating the spatial autoregressive parameters with smaller biases, whereas the Bayesian method with heteroscedasticity had smaller biases on both $\hat{\beta}_0$ and $\hat{\beta}_1$. In addition, the Bayesian method with heteroscedasticity yielded smaller variances of the parameter estimates of $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\rho}$, or $\hat{\gamma}$ than the other two methods (Tables 5 and 6).

Under non-normality ($t$ distribution in this study), GMM or 2SLS was better in terms of the biases than MLE for the SEM model. GMM produced the positive bias for $\hat{\gamma}$, whereas the other two methods yielded negative biases. The Bayesian method with heteroscedasticity had the highest biases but similar variances of the parameter estimators (Tables 5 and 6).

Discussion

Fox et al. (2001) showed that among different spatial autoregressive models SEM was most effective for modeling individual tree growth. Our study further confirmed...
their conclusion in terms of model fitting and the spatial autocorrelation in model residuals. It should be noted that Fox et al. (2001) used a moving average error process for the SEM model, whereas we used an autoregressive error process. Our results indicated that SEM was superior to SLM not only in the likelihood of model fitting but also in model predictions. On the other hand, SLM fitted the data significantly better than OLS, but produced poorer predictions than OLS. The LM test and Moran’s I coefficients of model residuals showed that SEM reduced the spatial autocorrelation of the OLS model errors much better than SLM (Zhang et al. 2009). Nevertheless, a general version of the spatial regression model may include both a lagged response variable and autocorrelated error structure (LeSage 1998), which was originally proposed for the case in which the LM test detected the spatial dependence among the model residuals of the SLM model. However, this general spatial regression model required more complicated computation and did not necessarily lead to a better model fitting (Fox et al. 2007).

Spatial dependence among the individual trees of a forest stand is typically positive and operates at the scale of micro-site variation. However, it may be confounded by a negative spatial dependence over small intertree distances caused by the competition among immediate neighbors (Fox et al. 2007). Unlike the finding in Fox et al. (2007), we did not find the apparent evidence of a competition effect in the correlogram of the OLS residuals. Although the Moran’s I coefficients of the model residuals became negative at large distances (e.g., >30 m), the absolute values of the negative Moran’s I coefficients were close to zero. Therefore, we can conclude that the micro-site variation was the dominant effect in the data used in our study.

In this study, we used one spatial weight matrix for both SLM and SEM. Fox et al. (2007) suggested that a spatial dependence matrix should be capable of simultaneously modeling the consequences of confounded spatial mechanisms. Thus, it might be a good idea to define separate spatial weight matrices based on the influence of competition and micro-site variation, which can be used to further explore the performance of SLM and SEM as well as different parameter estimation methods.

Our study indicated that different parameter estimation methods led to different model fitting and performance for spatial regression models based on the tree height–diameter data and Monte Carlo simulations. MLE is currently the most commonly used parameter estimation method. However, it often involves intensive computation and may have the problem of sparseness of spatial weighting matrices. More importantly, it needs the normality assumption. GMM and 2SLS, as the alternatives to MLE, can produce model fitting and parameter estimation similar to MLE under the normality. For the situations for which the data do not follow normal distributions or have outliers, GMM or 2SLS

### Table 5. Bias and variance of estimators in SLM from Monte Carlo simulations

<table>
<thead>
<tr>
<th>Simulation scenario</th>
<th>Model parameter</th>
<th>Bias</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLE</td>
<td>SLS</td>
<td>Bayesian heteroscedasticity</td>
</tr>
<tr>
<td>Normality</td>
<td>ρ</td>
<td>−0.0373</td>
<td>−0.0294</td>
</tr>
<tr>
<td></td>
<td>β₀</td>
<td>−0.0126</td>
<td>−0.0125</td>
</tr>
<tr>
<td></td>
<td>β₁</td>
<td>0.0050</td>
<td>0.0050</td>
</tr>
<tr>
<td>Outliers</td>
<td>ρ</td>
<td>−0.5515</td>
<td>−0.4821</td>
</tr>
<tr>
<td></td>
<td>β₀</td>
<td>−0.3691</td>
<td>−0.3712</td>
</tr>
<tr>
<td></td>
<td>β₁</td>
<td>0.3884</td>
<td>0.3860</td>
</tr>
<tr>
<td>t distribution</td>
<td>ρ</td>
<td>−0.0605</td>
<td>−0.0454</td>
</tr>
<tr>
<td></td>
<td>β₀</td>
<td>0.0096</td>
<td>0.0094</td>
</tr>
<tr>
<td></td>
<td>β₁</td>
<td>−0.0030</td>
<td>−0.0030</td>
</tr>
</tbody>
</table>

MLE, maximum likelihood estimation; SLS, 2-stage least squares.

### Table 6. Bias and variance of estimators in spatial error model from Monte Carlo simulations

<table>
<thead>
<tr>
<th>Simulation scenario</th>
<th>Model parameter</th>
<th>Bias</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLE</td>
<td>GMM</td>
<td>Bayesian heteroscedasticity</td>
</tr>
<tr>
<td>Normality</td>
<td>λ</td>
<td>−0.0013</td>
<td>0.0105</td>
</tr>
<tr>
<td></td>
<td>β₀</td>
<td>0.0073</td>
<td>0.0108</td>
</tr>
<tr>
<td></td>
<td>β₁</td>
<td>−0.0061</td>
<td>−0.0104</td>
</tr>
<tr>
<td>Outliers</td>
<td>λ</td>
<td>−0.4380</td>
<td>−0.0259</td>
</tr>
<tr>
<td></td>
<td>β₀</td>
<td>−0.4570</td>
<td>−0.3220</td>
</tr>
<tr>
<td></td>
<td>β₁</td>
<td>0.4192</td>
<td>0.2822</td>
</tr>
<tr>
<td>t distribution</td>
<td>λ</td>
<td>−0.0088</td>
<td>0.0029</td>
</tr>
<tr>
<td></td>
<td>β₀</td>
<td>−0.0083</td>
<td>−0.0060</td>
</tr>
<tr>
<td></td>
<td>β₁</td>
<td>−0.0031</td>
<td>−0.0044</td>
</tr>
</tbody>
</table>

MLE, maximum likelihood estimation; GMM, generalized method of moments.
proved to be more effective, i.e., have smaller biases than MLE according to the Monte Carlo simulations. Therefore, GMM or 2SLS is relatively robust to the violation of normality as shown by Arraiz et al. (2008).

The Bayesian methods provide a different methodology for estimating the parameters of spatial regression models. If the prior information on the parameters is available, the Bayesian method can be more effective. The Bayesian spatial autoregressive models developed by LeSage (1998) can deal with spatial heteroscedasticity and outliers. The application on the height–diameter data in this study showed that the Bayesian method with homoscedasticity produced results of model fitting and performance similar to those of MLE, although there were some small differences in the estimates of the model coefficients. Further, the Bayesian methods can detect the possible existence of heteroscedasticity using the values of $t$ defined in Equation 6. It appeared that SLM showed more heteroscedasticity than the SEM model. For dealing with the outliers in the Monte Carlo simulations, the Bayesian methods had much smaller biases for the model coefficients $\beta_0$ and $\beta_1$ than MLE and GMM or 2SLS. However, the Bayesian methods were not good at estimating the spatial autoregressive parameters $\hat{\phi}$ and $\gamma$. For non-normality such as the $t$ distribution in the simulation study, the Bayesian methods underperformed MLE in terms of bias. In all of the circumstances, the Bayesian methods generally produced the smallest variances of the parameter estimates. The posterior distributions of spatial autoregressive parameters were close to normal even if their prior distributions were set as the uniform distribution. In practice, we commonly have no idea of the distributions of model parameters. However, if previous studies can provide some information on the distribution of parameters, it will make the Bayesian methods more effective.

In this study, we focused on the small-sample properties of different parameter estimators using empirical and simulation data. The large-sample properties, especially the consistency and asymptotic distribution, need to be analyzed by theoretical studies. More theoretical works are needed to compare MLE, GMM, 2SLS, and Bayesian methods for spatial regression models. In fact, an efficient GMM estimation proposed by Lee (2007) can also be used to estimate SLM and has a better efficiency than 2SLS. Lin and Lee (2005) developed a robust GMM to tackle the unknown heteroscedasticity of spatial autoregressive models. These new theoretical achievements can be applied in different study fields including ecological and forest modeling.

Conclusions

We fitted spatial regression models to a plot of tree height–diameter data with OLS as a benchmark. Our results indicated that SEM was significantly better than either OLS or SLM, in terms of model fitting and spatial autocorrelation in model residuals. The positive spatial dependence among the individual trees was obvious as shown by the spatial correlogram and spatial autoregressive parameters. It indicated that the micro-site variation was dominant over the competition among trees in the plot.

For modeling the tree height–diameter relationship, MLE was superior to other parameter estimation methods because the assumptions of normality and homoscedasticity were basically held for the sample data at hand. For the Monte Carlo simulation data, however, the GMM method was robust to non-normality and produced much smaller bias for the spatial autoregressive parameter when few outliers existed. Furthermore, GMM required much less intensive computation than MLE and the Bayesian methods. The Bayesian method with heteroscedasticity did not effectively estimate the spatial autoregressive parameters but yielded very small biases for the intercept and slope coefficients of the regression model. Therefore, the selection of the parameter estimation methods may depend on the distribution of data and variables, as well as the purpose of the specific research.

To our knowledge, one of the most promising topics in spatial modeling is the definition of spatial weight matrix, which guarantees its exogeneity to the model as well as consistent and asymptotically normal estimators (Anselin 1999). In practice, different schemes can be used to derive the spatial weight matrix such as spatial covariance and Kriging for the spatial regression models.

Literature Cited


