USING FOREST INVENTORY DATA ALONG WITH SPATIAL LAG AND SPATIAL ERROR REGRESSION TO DETERMINE THE IMPACT OF SOUTHERN PINE PLANTATIONS ON SPECIES DIVERSITY AND RICHNESS IN THE CENTRAL GULF COASTAL PLAIN

Andrew J. Hartsell[1]

Abstract.—This study investigates the impacts of southern yellow pine plantations on species evenness and richness in the gulf coastal plain. This process involves using spatial lag and spatial error regression techniques using GeoDa software and U.S. Forest Service's Forest Inventory and Analysis data. The results indicate that increasing plantation area is negatively correlated to species evenness and richness. Preliminary results indicate that for every 10 percent increment increase in southern yellow pine plantation area, Shannon's E decreases by 0.02 and species richness declines by 1.6 species. However, these models account for less than 50 percent of the data's variance, an indication that the models are incomplete and more research is needed.

INTRODUCTION

Biodiversity, synonymous with biological diversity, can be defined as "the variety and variability among living organisms and the ecological complexes in which they occur" (OTA 1987). Humans perceive regions with a multitude of diverse species to have more value than those that don't (Ehrlich 1991, Wilson 1993). Possible reasons that species diversity is valued by humans are: larger number of plant species means a greater variety of crops and life; greater species diversity helps assure natural sustainability for all life forms; diverse ecosystems can better withstand and recover from a variety of disasters; and finally the planets complex systems, ecological networks, and energy flows are dependent upon numerous organisms and interactions (Gaston 1996, SCBD 2006, Wilson 1993).

However, global biodiversity may be threatened by anthropogenic sources. The main factors responsible for potential biodiversity loss include: land use change; habitat change such as forest fragmentation and conversion; invasive alien species; overexploitation; and pollution. Plantations, which are artificially regenerated forests that are often composed of genetically modified or alien species, satisfy two of these factors. It is important that science ascertains the positive and negative impacts of this management regime to facilitate public discourse and planning.

STUDY AREA

The initial study area was limited to the states of Texas, Louisiana, Mississippi, and Alabama. Only counties in those states having the majority of their area in the gulf coastal plain, as defined by Bailey (1998), were considered. This population was thinned further by two more factors. First, the Mississippi River and its associated alluvial basin bisect the study area. Counties in this region were removed. Second, any county with less than 200,000 acres of forest land was removed from the dataset due to FIA's sampling intensity. This assures that at least 30 forested plots are in each county, providing a reasonable estimate of species diversity and richness at the county level. Additionally, any "island" counties that were isolated

[1] Research Forester, U.S. Forest Service, Southern Research Station, 4700 Old Kingston Pike, Knoxville TN 37919. To contact, call 865-862-2032 or email at ahartsell@fs.fed.us.

and not attached to the larger study area were also removed. The final dataset was composed of data from 158 counties (Fig. 1).

DEFINITIONS AND CONCEPTS

Measuring Biodiversity

Shannon-Wiener (Shannon's) evenness index (E) and diversity index (H) come from information theory and measure the order and disorder within a population (Shannon and Weaver 1971). Shannon's diversity index is derived by calculating the proportion of species *i* relative to the total number of species (p_i), and then multiplying by the natural logarithm of this proportion ($\ln p_i$). The result is summed across species, and multiplied by −1:

$$H = -\sum_{i=1}^{R} p_i \ln p_i$$

Species richness (R) is the number of different species found in a region or study area. For this study, species richness is a count of all tree species found in each county. Species richness does not take into account the relative abundance distributions of species.

Spatial Statistics

Detecting Spatial Autocorrelation

One of the most common ways of detecting spatial autocorrelation in group-level data is the Moran's I statistic. Moran's I is a weighted correlation coefficient used to detect departures from randomness such as clusters. The formula for Moran's I is:

$$I = \sum_i \sum_j w_{ij} (x_i - \mu)(x_j - \mu) / \sum_i (x_i - \mu)^2$$

where: μ is the mean of the *x* variable
w_{ij} are the elements of the spatial weights matrix.

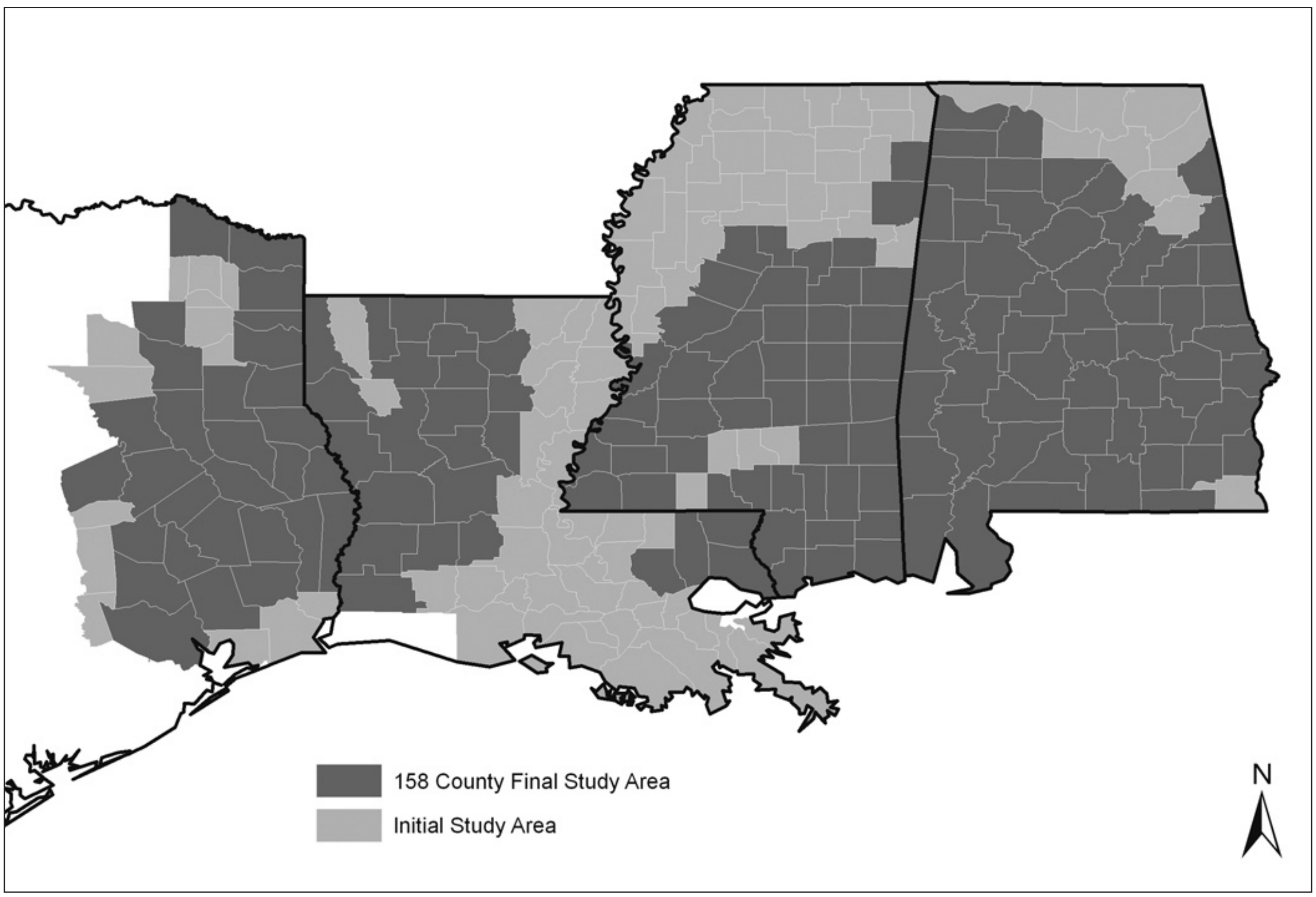

Figure 1.—Study area.

Geographically Weighted Regression: Spatial Lag and Spatial Error Models

Geographically weighted regression (GWR) can be performed in the presence of spatial autocorrelation. GWR accounts for distinctions between spatial similarity between the dependent and independent variables. Ordinary least squares (OLS) and other simple statistics do not do this. The basic formula for GWR is:

$$y = \chi\beta + e_i$$

where: χ is an n×p matrix of regressors
β is a p×1 vector of unknown parameters
e is unobserved scalar random variables (errors).

Spatial lag models (SLM) and spatial error models (SEM) are two types of GWR. Spatial lag models produce a spatially lagged variable on the right hand side of a regression equation. A spatial error model (SEM) considers the estimation of maximum likelihood of a spatial regression model that includes a spatial autoregressive error term on the right hand side of the regression equation.

METHODS

Species richness, Shannon's E, total forest area, and percent of forest area in southern yellow pine plantations (SYP) were computed for each county in the study area. Ordinary least squares (OLS) analysis was performed on the both Shannon's E and species richness using GeoDa version 0.95 software (OpenGeoDa version 1.2 is now available from the GeoDa Center at Arizona State University). Moran's I was calculated to determine if spatial dependence was an issue. If the data was determined to be spatially autocorrelated, then a series of LaGrange multiplier (LM) test statistics were computed. The results of the LM would then indicate which GWR model, spatial lag model or spatial error model, would be used in the final analysis.

RESULTS

Shannon's evenness index (E) was the first dependent variable investigated. The OLS regression of Shannon's E was performed using percent of forest land per county in southern yellow pine plantations (PCT_SYP_PL) as one independent variable, and a dummy variable that indicated if a county was on the east side of the Mississippi River. The average Shannon's evenness index was 0.695 (Table 1). The R^2 and adjusted R^2 were 0.368 and 0.360, respectively. The F-statistic and associated p-value indicated that the model was statistically significant. All three predictor variables, the intercept and two independent variables, were significant as well.

Table 1.—Results of ordinary least squares analysis on species evenness index using percent southern yellow pine plantations per county and location flag

Dependent Variable:	**SHANNONS_E**	Number of Observations:	158
Mean dependent var:	0.694942	Number of Variables:	3
S.D. dependent var:	0.0565841	Degrees of Freedom:	155
R-squared:	0.368228	F-statistic:	45.1709
Adjusted R-squared:	0.360076	Prob(F-statistic):	3.49482e-016
Sum squared residual:	0.319599	Log likelihood:	265.867
Sigma-square:	0.00206193	Akaike info criterion:	-525.734
S.E. of regression:	0.0454085	Schwarz criterion:	-516.546
Sigma-square ML:	0.00202278		
S.E. of regression ML:	0.0449753		

Variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	0.7422254	0.01048335	70.80043	0.0000000
PCT_SYP_PL	-0.002851605	0.0003366784	-8.469818	0.0000000
EAST_FLAG	0.03198344	0.007769342	4.116622	0.0000623

Tests for multicollinearity, normality, and heteroskedasticity proved to be insignificant (Table 2). However, Moran's I proved to be highly significant (p value =0.000000) indicating that spatial autocorrelation was an issue with the data. The first two tests (LM-error and LM-lag) were both significant, indicating that the robust models are more appropriate. The robust versions were to be considered only if the standard versions were significant. In this instance, both LM-lag and LM-error were significant, so the robust versions were then used. The Robust LM-error statistic was not significant (p value=0.8675), but the Robust LM-lag statistic was (p value=0.0087). Therefore, a spatial lag model is needed to remove any spatial autocorrelation. Table 3 shows the results of the spatial lag regression model on Shannon's evenness.

Table 2.—Regression diagnostics on ordinary least squares analysis of Shannon's species evenness index

REGRESSION DIAGNOSTICS			
MULTICOLLINEARITY CONDITION NUMBER		5.998419	
TEST ON NORMALITY OF ERRORS			
TEST	DF	VALUE	PROB
Jarque-Bera	2	0.2463868	0.8840927
DIAGNOSTICS FOR HETEROSKEDASTICITY			
RANDOM COEFFICIENTS			
TEST	DF	VALUE	PROB
Breusch-Pagan test	2	1.014434	0.6021692
Koenker-Bassett test	2	1.02169	0.5999884
SPECIFICATION ROBUST TEST			
TEST	DF	VALUE	PROB
White	5	N/A	N/A
DIAGNOSTICS FOR SPATIAL DEPENDENCE			
FOR WEIGHT MATRIX: **Queen** (row-standardized weights)			
TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.277114	5.5791937	0.0000000
Lagrange Multiplier (lag)	1	33.2970224	0.0000000
Robust LM (lag)	1	6.8814084	0.0087097
Lagrange Multiplier (error)	1	26.4434443	0.0000003
Robust LM (error)	1	0.0278303	0.8675084
Lagrange Multiplier (SARMA)	2	33.3248527	0.0000001

Table 3.—Spatial lag regression model on Shannon's species evenness index

Spatial Weight:	**Queen**		
Dependent Variable:	**SHANNONS_E**	Number of Observations:	158
Mean dependent var:	0.694942	Number of Variables:	4
S.D. dependent var:	0.0565841	Degrees of Freedom:	154
Lag coeff. (Rho):	0.510154		
R-squared:	0.518880	Log likelihood:	282.131
Sq. Correlation:	-	Akaike info criterion:	-556.262
Sigma-square:	0.00154043	Schwarz criterion:	-544.012
S.E. of regression:	0.0392483		

Variable	Coefficient	Std. Error	z-value	Probability
W_SHANNONS_E	0.5101542	0.07463184	6.83561	0.0000000
CONSTANT	0.3820561	0.05341811	7.152183	0.0000000
PCT_SYP_PL	-0.002135705	0.0003113954	-6.8585	0.0000000
EAST_FLAG	0.01523912	0.007230402	2.107645	0.0350616

The same process was repeated for species richness (R). The OLS regression on species richness included another independent variable, the amount of forest land in a county. This variable was labeled Forest_K, as each whole unit represents 1,000 acres of forest land. The average species richness for the study area was 50.5 (Table 4), indicating that each county in the study area has an average of 50 distinct tree species greater than 1.0 inch diameter at breast height (d.b.h.). The R^2 was 0.298 and the adjusted R^2 was 0.284, indicating that less than 30 percent of the dataset's variation was captured in the model. However, the model and all variables were statistically significant.

Tests for multicollinearity and normality indicated that neither was a problem. However, both tests for heteroskedasticity revealed that variances may not be equal. Furthermore, Moran's I shows that the data are spatially dependent (Table 5). The LM statistics indicated that the Robust LM-lag was insignificant.

Table 4.—Results of ordinary least squares analysis on Shannon's species richness using percent southern yellow pine plantations per county, amount of forested acres per county, and location flag

Dependent Variable:	**RICHNESS**	Number of Observations:	158
Mean dependent var:	50.5506	Number of Variables:	4
S.D. dependent var:	8.09715	Degrees of Freedom:	154
R-squared:	0.297692	F-statistic:	21.759
Adjusted R-squared:	0.284011	Prob(F-statistic):	8.38957e-012
Sum squared residual:	7275.27	Log likelihood:	-526.734
Sigma-square:	47.242	Akaike info criterion:	1061.47
S.E. of regression:	6.87328	Schwarz criterion:	1073.72
Sigma-square ML:	46.046		
S.E. of regression ML:	6.78572		

Variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	36.267	2.222544	16.31779	0.0000000
PCT_SYP_PL	-0.1072605	0.05378244	-1.994341	0.0478798
EAST_FLAG	3.60517	1.195587	3.015397	0.0030027
FOREST_K	0.04105437	0.005197799	7.898414	0.0000000

Table 5.—Regression diagnostics on ordinary least squares analysis of Shannon's species richness index

REGRESSION DIAGNOSTICS			
MULTICOLLINEARITY CONDITION NUMBER	9.475663		
TEST ON NORMALITY OF ERRORS			
TEST	DF	VALUE	PROB
Jarque-Bera	2	3.64542	0.1615873
DIAGNOSTICS FOR HETEROSKEDASTICITY			
RANDOM COEFFICIENTS			
TEST	DF	VALUE	PROB
Breusch-Pagan test	3	11.23643	0.0105138
Koenker-Bassett test	3	13.29254	0.0040448
SPECIFICATION ROBUST TEST			
TEST	DF	VALUE	PROB
White	9	N/A	N/A
DIAGNOSTICS FOR SPATIAL DEPENDENCE			
FOR WEIGHT MATRIX: **Queen** (row-standardized weights)			
TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.432723	8.6153092	0.0000000
Lagrange Multiplier (lag)	1	51.3552467	0.0000000
Robust LM (lag)	1	0.0432789	0.8352011
Lagrange Multiplier (error)	1	64.4791145	0.0000000
Robust LM (error)	1	13.1671467	0.0002849
Lagrange Multiplier (SARMA)	2	64.5223934	0.0000000

Therefore, a spatial error model must be created to counter these issues. Anselin notes that the spatial error model is also useful for reducing heteroskedasticity as well (Anselin 1992, 2005). A spatial error regression was performed to correct for these issues (Table 6).

The R^2 improved to 0.56, but as with the SLM model, this is a pseudo statistic and probably not directly comparable to OLS R^2. The best way to determine an improvement of goodness of fit over the OLS model is to compare LL, AIC, and SC. For the SLE model on species richness, all three improved.

DISCUSSION

The results of this study indicate that the area of southern yellow pine plantations in a county has a negative impact on species evenness and richness. Based on the spatially lagged regressions, Shannon's evenness (E) will decrease by 0.02 for every 10 percent increment increase in SYP plantation area. Likewise, species richness will drop by 1.6 species for the same change in plantation area.

However, while both models are statistically significant, they fail to account for over half of the variation in the dataset. This indicates that there are explanatory variables not accounted for. Further research needs to be performed to determine what these variables may be. Possible sources are: population estimates, road densities, land fragmentation patterns, or other socioeconomic factors.

LITERATURE CITED

Anselin, L. 1992. **Spatial data analysis with GIS: an introduction to application in the social sciences.** Tech. Rep. 92-10. Santa Barbara, CA: University of California-Santa Barbara, National Center for Geographic Information and Analysis. 53 p. Available at http://www.ncgia.ucsb.edu/Publications/Tech_Reports/92/92-10.PDF.

Anselin, L. 2005. **Exploring spatial data with GeoDatm: a workbook.** Urbana-Champaign, IL: University of Illinois, Urbana-Champaign, Center for Spatially Integrated Social Science, Spatial Analysis Laboratory. 244 p. Available at http://geodacenter.asu.edu/system/files/geodaworkbook.pdf.

Bailey, R.G. 1998. **Ecoregions: The ecosystem geography of the oceans and continents.** New York, NY: Springer-Verlag. 176 p.

Ehrlich, P.R.; Wilson, E.O. 1991. **Biodiversity studies: science and policy.** Science. 253(5021): 758-762.

Table 6.—Spatial error regression model on Shannon's species evenness index

Spatial Weight:	**Queen**		
Dependent Variable:	**RICHNESS**	Number of Observations:	158
Mean dependent var:	50.550633	Number of Variables:	3
S.D. dependent var:	8.097153	Degree of Freedom:	155
Lag coeff. (Lambda):	0.675756		
R-squared:	0.559861	R-squared (BUSE):	-
Sq. Correlation:	-	Log likelihood:	-500.089426
Sigma-square:	28.857203	Akaike info criterion:	1006.18
S.E. of regression:	5.37189	Schwarz criterion:	1015.366637

Variable	Coefficient	Std. Error	z-value	Probability
CONSTANT	39.73594	2.177353	18.24966	0.0000000
PCT_SYP_PL	-0.1636192	0.04842778	-3.378622	0.0007286
FOREST_K	0.04129515	0.004798334	8.606143	0.0000000
LAMBDA	0.6757558	0.06527036	10.35318	0.0000000

Gaston, K.J. 1996. **Biodiversity: a biology of numbers and difference.** Oxford, UK: Blackwell. 396 p.

Secretariat of the Convention on Biological Diversity (SCBD). 2006. **Global biodiversity outlook 2.** [Summary of the second Global Biodiversity conference]. Montreal, Canada: Secretariat of the Convention on Biological Diversity. 81 p. Available at www.biodiv.org/GBO2.

Shannon, C.E.; Weaver, W. 1971. **A mathematical theory of communication.** Champaign, IL: University of Illinois Press. 144 p.

U.S. Congress Office of Technology Assessment (OTA). 1987. **Technologies to maintain biological diversity.** Washington, DC: U.S. Congress, Office of Technology Assessment.

Wilson, E.O. 1993. **The diversity of life.** New York, NY: W.W. Norton & Co. 448 p.