

An Introduction to Spatial Analysis in Social Science Research

Yanqing Xu ^a, Eugene Kennedy ^{a, b}

^a Department of Geography and Anthropology, Louisiana State University

^b School of Education, Louisiana State University

Abstract ■ For years researchers have recognized the need to consider environmental and contextual variables in the social and behavioral sciences. Multilevel models have grown in popularity in large part because they provide a means to explicitly model the influence of context on many individual level processes. However, in applications of these and other statistical models that incorporate context into the analysis, rarely is physical location or distance between entities considered. In this paper we discuss a variety of spatial analysis techniques and their applications in educational and psychological research. We provide examples with the SAS software package and other more specialized spatial analysis software.

Keywords ■ Spatial Regression; Geographically Weighted Regression; Spatial Lag Model; Spatial Clustering; Global Clustering.

[✉ ekennedy@lsu.edu](mailto:ekennedy@lsu.edu)

Introduction

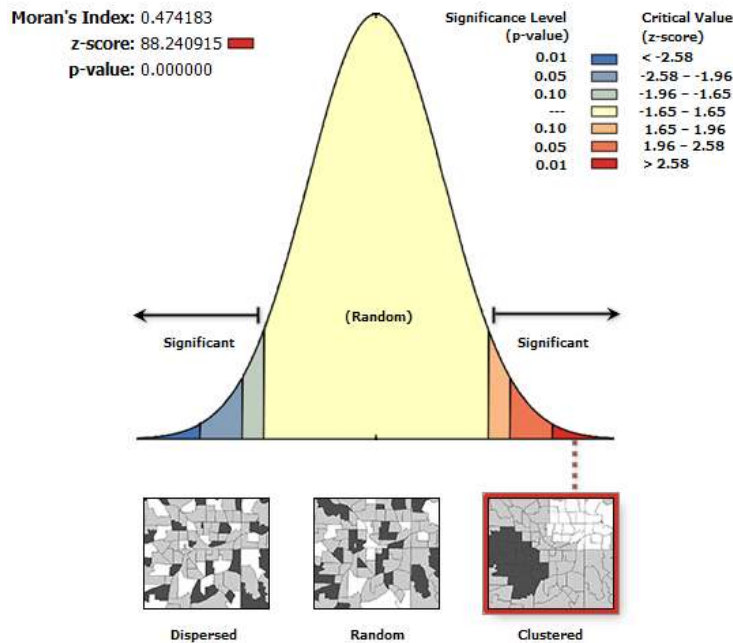
It has long been recognized that to fully understand human behavior researchers must consider the context and environment in which the behavior occurs. Unfortunately, statistical techniques for investigating the role and influence of environment have been limited. For much of the last several decades, researchers faced the equally unattractive choice of either ignoring environment and focusing on individual level variables or ignoring individual variability and conducting analyses at the group or aggregate level. Both approaches have significant limitations (Raudenbush and Bryk, 2002). The former approach may lead to the atomistic fallacy, where inferences about groups are incorrectly drawn from individual-level information (Hox 2002). The latter approach may lead to the ecological fallacy, where relationships observed in groups are assumed to hold for individuals (Freedman 1999). The advent of multilevel modeling techniques in the late 1980s provided tools for incorporating both individual and aggregate data into analyses. These techniques made it possible for researchers to routinely address questions such as “What is the influence of neighborhood characteristics on student achievement”?

Multilevel models, largely because of their ability to model contextual questions, have grown in popularity and have evolved into ever more complicated models (see Raudenbush and Bryk, 2002). However, it has been noted that in many applications researchers make

the untenable assumption that aggregate units close to one another do not affect each other. In other words, in many applications these models do not take proximity into consideration. Municipal boundaries, for example, as defined by law are treated as “real” ecological entities in which social processes in one setting do not “spill over” to another. This assumption is inconsistent with the first law of geography, mainly that objects closer together in space and time are more related than objects further away in space and time. Should spatial relationships exist in a specific dataset and not be reflected in the statistical model, the model is misspecified and the resulting parameter estimates and standard errors are inaccurate. Spatial analysis techniques provide means for testing the tenability of this assumption and of exploring issues of proximity (Haining, 1990). This paper describes a number of techniques of spatial analysis and their potential for applications in psychology and education. Examples using the SAS software are discussed.

What is Spatial Analysis?

Spatial analysis is analysis of data in which the location or distance between objects is taken into consideration. Spatial analysis includes techniques for visualizing phenomenon, determining if data exhibit spatial autocorrelation, and modeling spatial relationships (Anselin, 1995; Fotheringham, Brunson and Charlton, 2000). These techniques have seen growing use in a



Given the z-score of 88.2409145675, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Figure 1 ■ Spatial autocorrelation pattern

diverse number of disciplines, including demography, epidemiology, political science and sociology (e.g., Kyem, 2000). Researchers have found significant spatial correlations in studies of obesity, crime rates, disease patterns, and the like (e.g., Lin and Wen, 2011). There is also ample reason to believe that these tools are useful in psychology and education. In education, spatial analysis as well as maps of the spatial distribution of phenomena such as school achievement level can be useful to education planners and managers. For example, education planners can use spatial analysis to identify clusters of schools based on characteristics of the student populations served. High concentrations of poverty can be easily identified with visualization techniques. The effects of place are also potentially indirect as, for example, proximity to pollution sources is highly related to student morbidity which leads to increased absenteeism rates (e.g., Engler, 2007). There is growing research along these lines. Falch, Lujala and Storm (2011), for example, found that reduced travel time between students' homes and schools has a positive effect on graduation. Similarly, Bradford (1991) noted that standardized tests results are correlated with neighborhood level socio-economic data. These and other studies suggest that spatial relationships should be considered in

education and psychological research (e.g., Flowerdew and Pearce, 2001). Below, we describe several of the more popular spatial modeling approaches and provide examples from educational research. We begin with the question of whether or not spatial autocorrelation exists and then focus on techniques for modeling spatial relationships. We provide several examples using the SAS software.

Spatial cluster analysis

Spatial cluster analysis detects unusual concentrations or nonrandomness of events in space and time. Since nonrandomness of events indicates the existence of spatial autocorrelation, it is necessary to use techniques such as spatial regression analysis when spatial clustering is found. The methods for spatial cluster analysis can be divided into two groups: global clustering and local clustering.

Global clustering

Moran's I is one of the most common indicators of global clustering. Developed by Patrick Alfred Pierce Moran(1950), it examines whether nearby areas have similar or dissimilar attributes overall. If just one variable or attribute is under consideration, the formula is as follows:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \times \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

where n is the total number of observations (points or polygons), i and j represent different locations, x_i and x_j are values of the variable in the i th and j th locations, and \bar{x} is the mean of the variable. w_{ij} is a measure of spatial proximity for pairs i and j . The values of Moran's I typically fall between -1 and +1. -1 means negative autocorrelation which implies that nearby locations tend to have dissimilar values; +1 means positive autocorrelation which indicated that similar values

Table 1 ■ Global Moran's I summary

Moran's Index	0.474183
Expected Index	-0.000322
Variance	0.000029
z-score	88.240915
p-value	0.000000

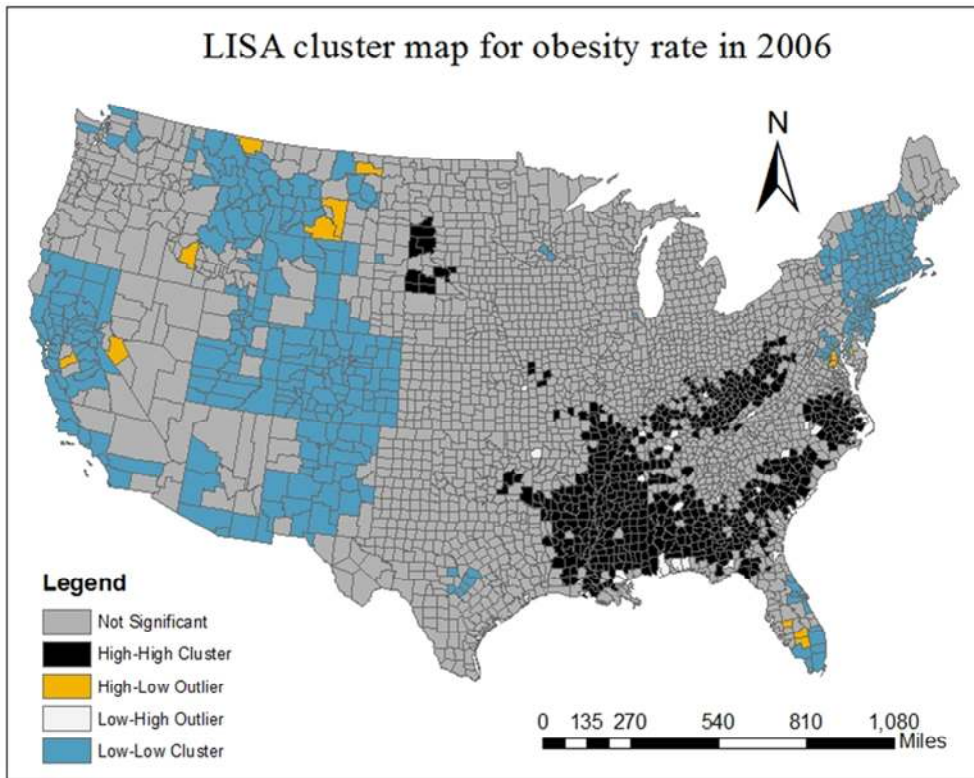


Figure 2 ■ An example of LISA cluster map

tended to occur in adjacent areas. If no spatial autocorrelation is found, then the spatial arrangement would be completely random. The transformation between Moran's I to Z score is:

$$Z = \frac{I - E(I)}{SE_I}$$

where $E(I) = -1/(n-1)$ is the expected index, I is the calculated value for Moran's I and SE_I is the standard error (ArcGIS Resources 2014). Figure 1 and Table 1 present graphic and numeric results for Moran's I by using the obesity rate in 2006 in the contiguous United States, including 49 states and Washington D.C.

In Figure 1, three patterns are depicted: dispersed, random and clustered. These patterns represent obesity rate for counties in the counterminous United States. The clustered pattern suggests a strong spatial autocorrelation. In Table 1 Moran's I is presented, along with the associated Z score. This value can be compared to any standard normal Z table and the value here of 88.24 is strong evidence that in this sample dataset there is substantial spatial autocorrelation.

Similar to Moran's I , Geary's C (Geary 1954) is another form of global clustering which is based on the

diversity in responses among observations:

$$C = \frac{n-1}{2 \sum_{i=1}^n \sum_{j=1}^n w_{ij}} \times \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - x_j)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

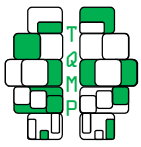
The value of Geary's C ranges from 0 to 2. If the value is equal or close to 1, it means that there is no autocorrelation. If the value of Geary's C is larger than 1, it indicates negative spatial autocorrelation; values less than one indicate a positive spatial autocorrelation.

Local clustering

Local Indicators of Spatial Association (LISA) are used to detect local clusters (Anselin, 1995). It can be defined as

$$I_i = z_i \sum_j w_{ij} z_j$$

where z_i, z_j are in deviations from the mean, w_{ij} is the spatial weight. A positive value indicates either a high value surrounded by high values (high-high) or a low value surrounded by low values (low-low), while a negative value indicates either a high value surrounded by low values (high-low) or low value surrounded by



high values (low-high). Figure 2 shows an example of a LISA cluster map for the obesity rate in the conterminous United States at the county level in the year of 2006. We can see that hot spots (high-high regions) are mostly found in the southern areas and most of the cold spots (low-low regions) are situated in the west and northeast areas.

For both examples presented there is strong evidence of spatial autocorrelation. This is the first step in spatial analysis—in the examples below, Moran's I will be computed with the SAS software. The next step is to attempt to incorporate the spatial autocorrelation structure in the actual analysis.

Spatial regression

Multiple regression is a widely known technique among psychologist, educators and other social science researchers. In this section we discuss three expansions of the regression model for purposes of accommodating spatial relationships among data elements.

Spatial lag model

This model is also called the spatially autoregressive model which includes means of the dependent variable in neighboring areas as an extra explanatory variable. For the case where there is one explanatory or predictor variable, the model is expressed as

$$y_i = \beta_0 + \beta_1 x_i + \rho w_i y + \varepsilon_i, \quad i = 1, \dots, n$$

where y_i is the value of the dependent variable for case i , β_0 is the intercept, x_i is the predictor variable, w_i is a $1 \times n$ row vector representing the proximity of other cases to case i (coefficients sum to 1), y is a $n \times 1$ column vector of values of the dependent variable for all cases, and ρ is the lag coefficient. The model shows that the value of y_i at each location i is determined not only by x_i at that location, but also by the value of y for neighboring units.

Spatial error model

This model is different from the spatial lag model in that it doesn't treat the dependent variable as autoregressive, it considers the error term as autoregressive (Ward and Gleditsch, 2007). The model is expressed as

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

where ε_i is the usual residual in regression. In this case, however, the residual is divided into a random part ε_i , and a portion which reflects the spatial covariance between units

$$e_i = \lambda w_i \xi_i + \varepsilon_i$$

In this model, λ is a coefficient which reflects the degree of spatial covariance between units, w_i as above is a proximity measure and ξ_i is the portion of the residual/error that is spatially correlated between units. The full model can be represented by

$$y_i = \beta_0 + x_i \beta + \lambda w_i \xi_i + \varepsilon_i$$

Both the spatial lag model and spatial error model are estimated by the maximum likelihood method.

Geographically Weighted Regression model

A Geographically Weighted Regression (GWR) model is a localized multivariate regression that allows the parameters of a regression model to change locally (Lin and Wen 2011). The model can be expressed as:

$$y_i = \beta_{0i} + \beta_{1i} x_{1i} + \beta_{2i} x_{2i} + \dots + \beta_{ni} x_{ni} + \varepsilon$$

where β_{ni} refers to the estimated regression coefficients which are location variant and $x_{1i}, x_{2i} \dots$ refers to the explanatory variables at location i . The spatial variability of an estimated local regression coefficient is examined to determine whether the underlying process exhibits spatial autocorrelation (Fotheringham, Brunsdon and Charlton 2000).

GWR is calibrated by multiplying the geographically weighted matrix W_i consisting of geo-referenced data (Fotheringham, Brunsdon and Charlton 2002). The estimated local coefficient β_i is given by

$$\beta_i = (X^T W_i X)^{-1} X W_i Y$$

The W_i s are defined by the spatial neighboring relations between points, which is defined as:

Louisiana County Level Poverty Rate & School Achievement Level

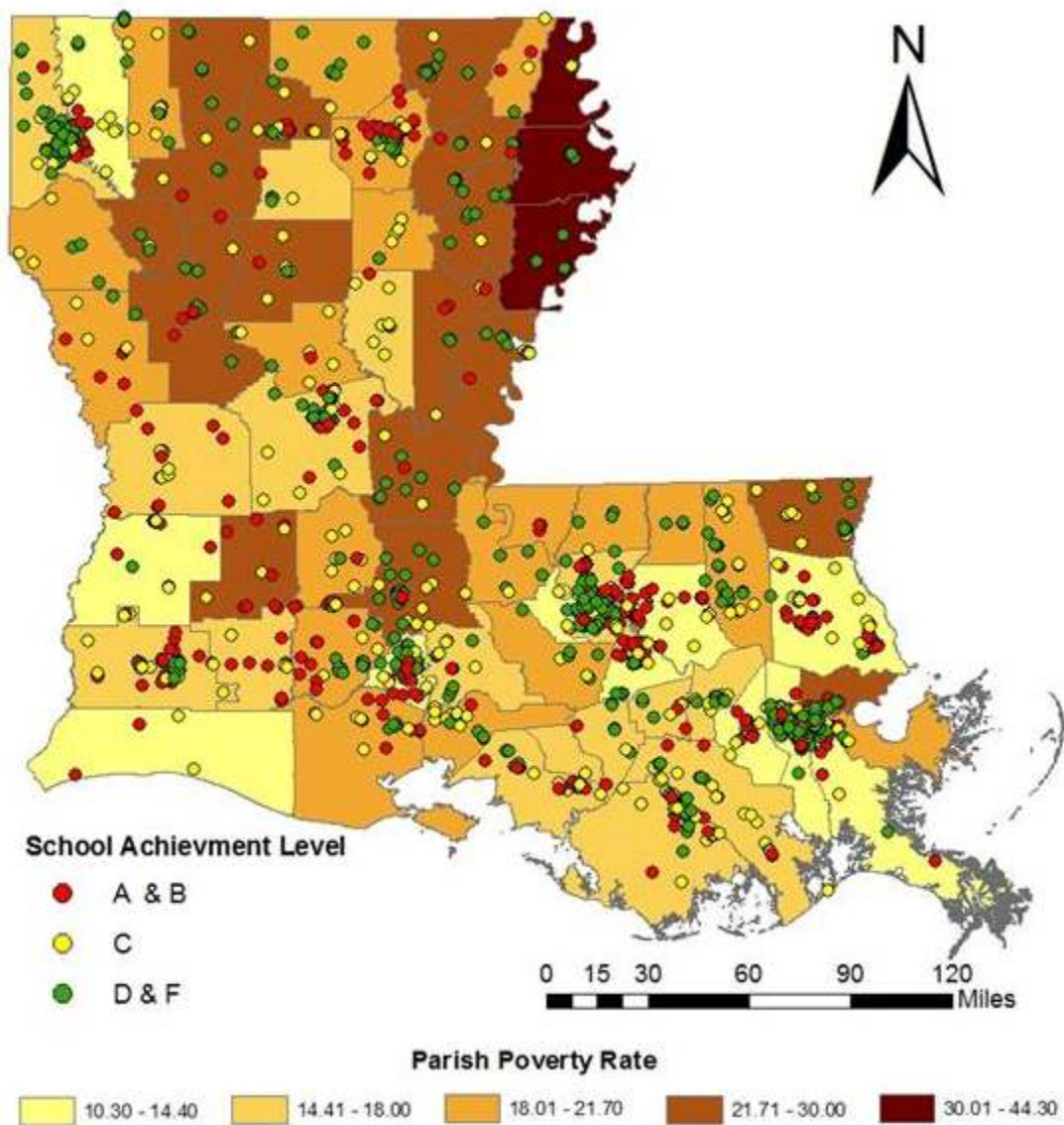


Figure 3 ■ Poverty rates at the county level and school achievement level in Louisiana (2009)

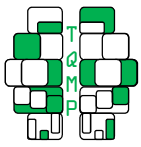
$$W_i = \begin{pmatrix} w_{i1} & 0 & 0 & \dots & 0 \\ 0 & w_{i2} & 0 & \dots & 0 \\ 0 & 0 & w_{i3} & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & 0 & 0 & \dots & w_{in} \end{pmatrix},$$

where w_{in} is the impact between position i and position n and the values are between 0 and 1. The larger the w_{in} , the closer geographically data points are, and the stronger impact they have on each other. The spatial variability of an estimated local regression coefficient is examined to determine whether the underlying

process exhibits spatial autocorrelation. Unlike the traditional ordinary least squares model, the GWR model is a localized multivariate regression that allows the parameters of a regression model to change locally which means it could access the spatial influences among neighboring areas.

Spatial Analysis with Multilevel Models

The multilevel model provides a natural mechanism for studying contextual and organizational effects on human behavior. The basic model as described below provides for fixed and random effects. In other words,



the basic multilevel model accounts for within-cluster correlations, but assumes no meaningful correlations between clusters. This is handled, in some respect, by having cluster level covariates in the model in which between cluster variation is explained. However, the spatial perspective allows for the possibility that the relationships between clusters change, the greater the spatial distance between units.

A criticism of the multilevel model for studying contextual effects is that the traditional models make the untenable assumption that context effects are not related across contexts. Individuals, for example, in one neighborhood are not affected by characteristics of neighboring communities. There is substantial evidence that this is not the case and that ignoring correlations among contextual units can lead to misleading results. One strategy for addressing this is to relax the assumption that individuals are limited to just one community, as is the case with the traditional model. The cross-classified model assumes that individuals move across communities or organizations and attempts to model this additional complexity. In spatial analysis, this is modeled in terms of a decreasing correlation between communities as they grow in distance from one another — a fact consistent with the first law of geography which demonstrates that “everything is related to everything else, but near things are more related than distant thing” (Tobler 1970). Below we discuss a simple model for incorporating spatial autocorrelation into multilevel models.

In the following section we present examples of fitting several spatial models using the statistics described above. The examples use student achievement and poverty data in Louisiana in 2009. The distribution of schools in Louisiana, by parish (county) poverty level and school performance label (A, B, C, D, F) is presented in Figure 3. As is obvious, there is great variation in the state with regard to the distribution of poverty, with the greatest occurring in the northeast corridor. High performing schools are spread across the state, but in many areas there are concentrations of poor performing schools. The map also shows that many schools are isolated and remotely located. Spatial techniques have the potential to explore the potential of location to impact school processes and outcomes. For example, do schools located in the middle of high poverty parishes experience greater difficulty attracting qualified teachers than those located on the borders? In the following sections we

address these and similar issues.

Examples with Cross Sectional Data

We present two examples.

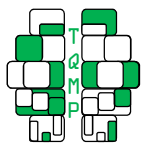
Spatial Regression at the School level

We begin with a regression model which includes spatial autocorrelations at the school level. The outcome is **avgmathss** (average math score) and the predictor is **pctlunch** (percentage of students who receive free lunch). The first step is to produce residuals and then use **ProcVariogram** to produce Moran's *I*. The syntax is presented below. Note that both the *latitude* and *longitude* of each school is used in the analysis.

For these analyses we use two SAS system files **All.sas**, which has all the school level data and **Coordinate.SAS** which has the latitude and longitude data for each school. The contents of these files are listed in the Appendix. We first run the regression procedure (see Line 1 below) and create an output dataset of predicted values (*yhat*) and residuals (*yresid*). We will examine these residuals for the possibility of spatial correlation. To accomplish this, we merge the geographic coordinates of each school with the residuals from the regression run (Lines 2 through 5), and then execute the ProcVariogram routine, which will produce Moran's *I*. (Lines 6).

```
Procreg;  
Model avgmathss=pctlunch; output out=residuals  
  p=yhat r=yresid;  
Data gene.spresid;set residuals;  
Data one;set gene.coordinate;  
Data two;set gene.spresid;  
Data three;merge one(in=ii) two(in=jj);by  
  sitecode; if ii and jj;  
procvariogram data=three;  
Compute novarautoc (weights=distance);  
Coordinates xc=longitude yc=latitude;  
Var yresid;
```

The results from the ProcVariogram procedure are presented in Figure 4. Both Moran's *I* and Geary's *C* are significant at the .05 probability. These results support the conclusion that there are spatial correlations among the residuals when school level math achievement is predicted from the percent of students receiving free school lunch.



Assumption	Coefficient	Autocorrelation Statistics				
		Observed	Expected	StdDev	Z	Pr> Z
Normality	Moran's I	0.00179	-0.00138	0.000813	3.9	<.0001
Normality	Geary's c	1.01305	1.00000	0.005941	2.2	0.0281

Figure 4 ■ Results for Moran’s I and Geary’s c

Given that there is spatial autocorrelation present in these data, a model which ignores this would be misspecified. In the next stage of our study, we use the Proc Mixed procedure to compare various spatial autocorrelation structures. The syntax for the baseline model is as follows:

Baseline Model

```
proc mixed;
model avgmathss = pctlnch/solution;
```

The output below presents estimates of the fixed effects. These show that percent free lunch is a significant predictor of math performance. Our next step in this process is to compare various spatial autocorrelation structures. To facilitate this, we also include the model fit statistics: AIC, etc. Decisions regarding which autocorrelation structure best fits these data are based on these fit statistics— smaller values being better.

Figure 5 presents output from the initial regression model. We established earlier that the residuals from this model exhibit spatial correlations and we now began in investigation of various possible structures. SAS offers several possible spatial covariance

Covariance Parameter Estimates					
CovParm	Estimate				
Residual	421.38				

Fit Statistics	
-2 Res Log Likelihood	6422.6
AIC (smaller is better)	6424.6
AICC (smaller is better)	6424.6
BIC (smaller is better)	6429.2

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr> t
Intercept	369.20	2.5328	722	145.77	<.0001
pctlnc	-94.222	3.3780	722	-27.89	<.0001

Figure 5. Output from Initial Regression of Math Achievement on Percent Free Lunch

structures: Exponential, Gaussian, Linear, Spherical, etc. For this model, each represents particular pattern of changes in spatial covariance among residuals as observations grow in distance from one another. Below, we utilize the AIC fit statistics above to examine several possibilities (see Helie, 2006) — as noted in the output, smaller values indicate a better fit.

Model with Spherical Spatial Structure

The outcome `avgmathss` may have a Spherical spatial correlational form. We specify this structure with the `type=sp(sph)` followed by the spatial information of the variables. For the geometrically anisotropic structure `sp(sph)`, the exact spatial coordinate variables: longitude and latitude must be specified. The syntax for the model with Spherical Spatial Structure is:

```
proc mixed;
model avgmathss = pctlnch/solution;
repeated/subject=intercept
type=sp(sph)(longitude latitude)local;
```

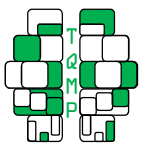
The parameter estimates and fit statistics for this model are presented below. Our first observation is that the fit statistics drop in value, indicating a better fit to the data.

Figure 6 presents results from the SAS run in which the spherical spatial covariance structure was fit to the residuals from the regression model. Focusing on the AIC fit statistics, we note a drop from 6424 to 6417. This is an improvement, but as part of our model building process we consider two other structures.

Model with Gaussian Spatial Structure

Similar to the Spherical, Gaussian spatial form is also modeled. The spatial coordinate variables are also required in this structure. The syntax for the model with Gaussian Spatial Structure is:

```
Proc mixed;
model avgmathss = pctlnch/solution;
repeated/subject=intercept
type=sp(gau)(longitude latitude)local;
```



Covariance Parameter Estimates		
CovParm	Subject	Estimate
Variance	Intercept	10655
SP(SPH)	Intercept	767.15
Residual		402.16
Fit Statistics		
	-2 Res Log Likelihood	6411.9
	AIC (smaller is better)	6417.9
	AICC (smaller is better)	6417.9
	BIC (smaller is better)	6431.6
Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr>ChiSq
2	10.71	0.0047

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr> t
Intercept	367.61	102.99	0	3.57	.
pctlnc	-95.6782	3.4392	722	-27.82	<.0001
Type 3 Tests of Fixed Effects					
Effect	Num DF	Den DF	F Value	Pr > F	
Pctlnc	1	722	773.92	<.0001	

Figure 6 ■ Results for the Spherical Correlational Form

The results for the Gaussian model are presented below. These results show only a slight, if any improvement, in fit.

The model with the Gaussian spatial structures shows some improvement over the spherical structure and both are better than the model with no spatial correlations. Other structures might be considered, however, with the process as illustrated here.

Multi-level Spatial Regressions

In this model, we incorporate both school and student level variables. We will proceed as is typical of multilevel models, fitting an unconditional model first (no predictors) and then building a student and then school level model. Our analysis will differ in that we will explore the possibility of spatially correlated residuals at the school level.

Unconditional Model without Spatially Correlated Errors

The unconditional multilevel model can be regarded as a one-way ANOVA model with random effects. With both the individual and school level data structure, we first run the unconditional model without spatially correlated errors. The syntax for this model and the resulting parameter and fit statistics are presented

Covariance Parameter Estimates		
CovParm	Subject	Estimate
Variance	Intercept	25.3135
SP(GAU)	Intercept	0.3206
Residual		397.90
Fit Statistics		
	-2 Res Log Likelihood	6410.4
	AIC (smaller is better)	6416.4
	AICC (smaller is better)	6416.4
	BIC (smaller is better)	6430.1
Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr>ChiSq
2	12.20	0.0022

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr> t
Intercept	370.10	2.7332	0	135.4	.
Pctlnc	-96.1156	3.4840	722	-27.59	<.0001
Type 3 Tests of Fixed Effects					
Effect	Num DF	Den DF	F Value	Pr > F	
pctlnc	1	722	761.06	<.0001	

Figure 7 ■ Results for the Gaussian Spatial Form

below.

```
Procmixed covtest noclprint;
Class sitecode;
Model mathss=/;
random intercept/subject=sitecode;
```

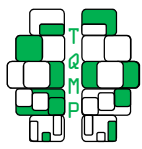
As with the spatial regression model, our focus here is on the fit statistics. The goal is to determine what spatial covariance structure best fits the observed data.

Unconditional Model with Spatially Correlated Errors

This model is similar to the previous model. However, in this instance we allow spatially correlated errors by using the geographic information of the school.

```
Procmixed covtest noclprint;
Class sitecode;
Model mathss=/;
random intercept/subject=sitecode
type=sp(sph)(longitude latitude);
```

When comparing these models, the fit statistics do now show a meaningful drop, suggesting that perhaps a spherical covariance structure is not appropriate for these results. As noted above, other covariance structures could be examined.



Covariance Parameter Estimates					
CovParm	Subject	Estimate	Standard Error	Z Value	Pr> Z
Intercept	Sitecode	737.00	43.115	17.09	<.0001
Residual		3147.5	20.919	150.5	<.0001

Fit Statistics	
-2 Res Log Likelihood	503056.5
AIC (smaller is better)	503060.5
AICC (smaller is better)	503060.5
BIC (smaller is better)	503069.6

Figure 8 ■ Unconditional Multilevel Model

The Fitted Model with Student, School and County Predictors

The unconditional model provides a baseline for us to compare more complex models. In this model, the student (mathss=math scale score, lnch=free lunch status), school (sps2008=school performance measure for 2008) and county (poverrate=county poverty rate) variables are included.

```
Proc mixed covtest noclprint;
Class sitecodecnty;
Model mathss=lnch sps2008
      poverrate/solution;
random
      intercept/subject=sitecodetype=sp (sph) (longitude latitude);
```

Figure 10 presents results for the complete multilevel model, one that includes both student and school level predictors. The improvement in the model is significant from that of the unconditional model, although we have not explored alternative spatial structures.

Conclusions

The results presented above show that spatial correlations exist among schools in Louisiana with respect to math achievement, meaning that those in close proximity to one another are more similar than would be expected by chance. They may serve students from the same families, neighborhoods, and may

Covariance Parameter Estimates					
CovParm	Subject	Estimate	Standard Error	Z Value	Pr> Z
Variance	Sitecode	737.94	43.216	17.08	<.0001
SP(SPH)	Sitecode	648.00	0	.	.
Residual		3147.4	20.919	150.5	<.0001

Fit Statistics	
-2 Res Log Likelihood	503056.5
AIC (smaller is better)	503062.5
AICC (smaller is better)	503062.5
BIC (smaller is better)	503076.2

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr>ChiSq
2	7243.87	<.0001

Figure 9 ■ Unconditional Model with Spatially Correlated School Level Residuals

experience similar processes with respect to recruiting and retaining teachers and administrators. These relationships persist even when student and school contextual variables are incorporated in the model examined. They suggest that a potentially valuable avenue of future research related to poverty and educational is to explicitly incorporate location in models of schooling processes. For example, the interstate roadway system in Louisiana was not imposed in Figure 3, yet it appears that outside of urban centers, the distribution of failing schools is related to the presence of interstate roadway. These and other hypotheses can be explored with spatial modeling techniques.

References

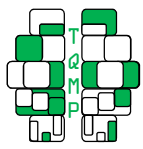
ArcGIS Resources (2014). *How Spatial Autocorrelation (Global Moran's I) works*. ESRI.

Bradford, M. (1991) School-performance indicators, the local residential environment, and parental choice. *Environment and Planning A*, 23, 319-332.

Falch, T., P. Lujala & B. Strom (2011). *Geographical constraints and educational attainment*.

Fotheringham, A. S., C. Brunson & M. Charlton (2000). *Quantitative Geography: Perspectives on Spatial Data Analysis*. Newbury Park, CA: Sage Publications Ltd.

Fotheringham, A. S., C. Brunson & M. Charlton. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. New York, NY, USA: Wiley.



Covariance Parameter Estimates					
CovParm	Subject	Estimate	Standard Error	Z Value	Pr> Z
Variance	Sitecode	283.20	18.5162	15.29	<.0001
SP(SPH)	Sitecode	648.00	0	.	.
Residual		3017.00	20.0704	150.32	<.0001

Fit Statistics	
-2 Res Log Likelihood	499551.9
AIC (smaller is better)	499557.9
AICC (smaller is better)	499557.9
BIC (smaller is better)	499571.7

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr>ChiSq
2	2474.49	<.0001

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr> t
Intercept	280.89	3.9462	722	71.18	<.0001
Lnch	-28.2425	0.6128	45E3	-46.08	<.0001
SPS2008	0.5756	0.02754	45E3	20.90	<.0001
PoverRate	-0.3685	0.1511	45E3	-2.44	0.0147

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr> F
lnch	1	45E3	2123.79	<.0001
SPS2008	1	45E3	436.71	<.0001
PoverRate	1	45E3	5.95	0.0147

Figure 10. Complete Multilevel Model

Freedman, D. A. (1999). *Ecological Inference and the Ecological Fallacy*. Berkeley: University of California.
 Geary, R. C. (1954) The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician*, 5, 115-45.
 Hox, J. (2002). *Multilevel Analysis: Techniques and Applications*. Mahwah: Lawrence Erlbaum Associates.
 Lin, C.-H. & T.-H. Wen (2011). Using Geographically Weighted Regression (GWR) to Explore Spatial

Varying Relationships of Immature Mosquitoes and Human Densities with the Incidence of Dengue. *International Journal of Environmental Research and Public Health*, 8, 2798-2815.
 Moran, P. (1950) Notes on continuous stochastic phenomena. *Biometrika*, 37, 17-23.
 Tobler, W. R. (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 234-240.

Citation

Xu, Y., Kennedy, E. (2015). An Introduction to Spatial Analysis in Social Science Research. *The Quantitative Methods for Psychology*, 11 (1), 22-31.

Copyright © 2015 Xu & Kennedy. This is an open-access article distributed under the terms of the *Creative Commons Attribution License (CC BY)*. The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 1/05/14 ~ Accepted: 24/11/14