Compare Spatial and Multilevel Regression Models for Binary Outcome in Neighborhood Study

Hongwei Xu
Survey Research Center at ISR
University of Michigan
Ann Arbor, MI 48104

Abstract: The widely used multilevel regressions in neighborhood research typically ignore potential between-neighborhood correlation due to underlying spatial process, and hence produce inappropriate inferences about neighborhood effects. In contrast, spatial models make estimation and prediction over space by explicitly modeling the spatial correlations among observations in different locations. A better understanding of the strength and limitation of spatial models as compared to multilevel models is needed to improve the research on neighborhood and spatial effects. This research systematically compares model estimation and prediction for binary outcomes between spatial and multilevel models in presence of both within- and between-neighborhood correlations through simulations. Preliminary results show that multilevel and spatial models produce similar estimates of fixed effects, but different estimates of random effects. Both the multilevel and pure spatial models tend to overestimate the corresponding random effects, compared to a full spatial model when both non-spatial within neighborhood and spatial between-neighborhood effects exist.

Acknowledgements: Part of this research was conducted using computational resources and services at the Center for Computation and Visualization, Brown University.

INTRODUCTION

Multilevel regression model is one of the most widely used methods in the research of neighborhood effects on individual outcomes (Dietz 2002; Diez-Roux 2000; DiPrete and Forristal 1994). A multilevel model is able to correct for within-neighborhood correlation among individual observations and thus to provide unbiased standard error and efficient estimates for individual- and neighborhood-level predictors (Diggle, Heagerty, Liang, and Zeger 2002). It also allows an assessment of within- and between-neighborhood variations (Snijders and Bosker 1994) as well as how these variations are contributed by individual- and neighborhood-level predictors (Diez-Roux 2000). Nevertheless, it has been criticized for its incapacity to estimate independent neighborhood effects, at least with observational data (Diez-Roux 1998; Oakes 2004). Multilevel models typically ignore potential between-neighborhood correlation due to, for example, spatial diffusion process and assume independent observations in one neighborhood from those in another neighborhood, which may lead to overestimated statistical significance of neighborhood effects (Chaix, Merlo, and Chauvin 2005).

Figure 1 (left) illustrates a hypothetical example of the assumption of within-neighborhood correlation in multilevel models. Each cell in the grid represents a neighborhood and its color indicates the average level (from low to high) of certain individual outcome shared by the observations from that neighborhood. Within a neighborhood, each observation's outcome deviates around the neighborhood's mean. Neighborhood's mean level of outcome varies from one to another, resulting in more similar outcomes among the observations from the same neighborhood than those from different neighborhoods (depicted as different colors across cells in Figure 1). The seemingly random distribution of mean outcome at neighborhood-level reflects the assumption of between-neighborhood independence. That is, the mean outcome is not more similar between two adjacent neighborhoods than that between two distant neighborhoods.

[Figure 1 about here]

In reality, however, between-neighborhood correlation may exist as a function of the distance between two nearby neighborhoods as stated in Tobler's First Law of Geography (Tobler 1970), "Everything is related to everything else, but near things are more related than distant things." First of all, socioeconomic and political resources in a neighborhood are likely to be linked to those in adjacent neighborhoods within a large citywide system (Logan and Molotch 1987) which in turn may lead to distinct spatial patterns of structural differentiation in individual outcomes across neighborhoods. Secondly, social behavior and interaction are not necessary restricted within one's immediate neighborhood, especially when the neighborhood boundaries are defined in a way that does not coincide with one's real-life experience such as census geography and postal code (Flowerdew, Manley, and Sabel 2008; Guo and Bhat 2007; Riva, Apparicio, Gauvin, and Brodeur 2008; Tatalovich, Wilson, Milam, Jerrett, and McConnell 2006). Instead, they may transcend neighborhood boundaries and thus be affected by or consequential to what happens in nearby areas (Sampson, Morenoff, and Gannon-Rowley 2002). For example, Collective efficacy in a neighborhood has been found to benefit residents living in adjacent neighborhoods (Sampson, Morenoff, and Earls 1999), while spatial proximity to poverty and violent crimes in adjacent neighborhoods has been associated with out-migration from current neighborhood(Morenoff and Sampson 1997).

Spatial models have been developed to make estimation and prediction over space by explicitly modeling the spatial correlations among observations in different locations (Diggle, Tawn, and Moyeed 1998). One of the key goals in spatial analysis is to estimate and predict the spatial distribution of an outcome of interest across the study area based on observations at a discrete set of locations (Diggle, Jr., and Christensen 2003). Such model estimation and prediction typically involve certain stochastic assumptions about distance-based correlations among observations at known locations and unknown values at prediction locations. In addition to examining spatial distribution, spatial models also allow researchers to investigate associations between individual- and neighborhood-level predictors and outcome of interest while adjusting for non-independent observations (Chaix, Merlo, and Chauvin 2005; Chaix, Merlo, Subramanian, Lynch, and Chauvin 2005; Dietz 2002).

Figure 1 (right) illustrates a hypothetical example of the assumption of distance-decay correlation across neighborhoods in a spatial model. In addition to within-neighborhood correlation (as indicated by cells of different colors), a spatial model assumes the strength of correlation between two locations declines as the distance between them increases, resulting in similar mean outcomes among nearby neighborhoods and hence clusters of neighborhoods with similar mean outcomes (as indicate by the color gradient color across the cells in Figure 1 right).

Chaix and associates (Chaix, Merlo, and Chauvin 2005; Chaix et al. 2005) are among the first to compare the spatial approach with the multilevel approach for studying neighborhood effects on health. Through empirical analyses of healthcare utilization in France (Chaix, Merlo, and Chauvin 2005) and mental health in Sweden (Chaix et al. 2005), Chaix and colleagues demonstrated that multilevel models could fail to capture both measures of associations between neighborhood factors and residents' outcomes and measures of unexplained variation in these outcomes across areas. However, it remains unclear whether these results are only valid for these two specific data sources or can be generalized to other study settings. In addition, they did not provide a thorough comparison of model performance in terms of both model estimation and prediction between spatial and multilevel models through a formal approach such as simulation analysis (Burton, Altman, Royston, and Holder 2006).

Aided by advance in spatial techniques and increased availability of spatial data, researchers have become increasingly interested in the spatial dynamics beyond simple neighborhood-level variation (Dietz 2002; Logan, Zhang, and Xu 2010; Sampson, Morenoff, and Gannon-Rowley 2002). A better methodological understanding of the strength and limitation of spatial models as compared to multilevel models is needed to help move forward the research on neighborhood and spatial effects. In this paper, I seek to systematically compare model performance in estimation and prediction between spatial and multilevel models in presence of both within- and between-neighborhood correlations through simulation studies. I focus on models of binary outcome using logit and probit links because of their increased prevalence and popularity in neighborhood and spatial studies. I also draw on empirical data to illustrate the application of the spatial approach which remains limited in the existing research on neighborhood effects and demonstrate its relative advantages compared to the multilevel approach.

SPATIAL MODEL

Diggle and colleagues (1998) are among the first to extend linear spatial model to accommodate nonlinear outcomes such as binary and count data. Let p_{ij} denote the probability for an individual i in neighborhood j having a binary outcome. Using an appropriate function such as logit or probit, a binary outcome can be associated to linear predictors as the following,

$$logit(p_{ij}) = \beta_0 + X_{ij}\beta + u_j + s_j$$

or

$$probit(p_{ij}) = \beta_0 + X_{ij}\beta + u_j + s_j$$

where β_0 is the regular intercept, and $X_{ij}\beta$ is the product of individual- and neighborhood-level predictors and the corresponding unknown parameters. Within-neighborhood correlation is captured by u_j which is usually assumed to be a normally distributed random intercept with mean 0 and variance σ_u^2, known as the *nugget* in the spatial literature (Banerjee, Gelfand, and Carlin 2004).

Distance-based between-neighborhood correlation is captured by the random effects s_j which is also commonly assumed to be normally distributed in the following form,

$$s \sim N(0, \sigma_s^2 \mathrm{H}(\phi))$$

where σ_s^2 denotes the variance of the spatial random effects, known as the *partial sill* in the spatial literature (Banerjee, Gelfand, and Carlin 2004). Dropping s_j from the right hand side of the equations in (1) and (2) results in the conventional multilevel model known as random-intercept model. On the other hand, dropping u_j but keeping s_j leads to a pure spatial model that incorporates only between-neighborhood but ignores within-neighborhood correlations.

The other $\mathrm{H}(\phi)$ is a correlation matrix that specifies how the spatial correlation declines as the distance between two locations increases. The geographical centroid (sometimes weighted by population distribution) of a neighborhood can be used as a proxy for the location of the observations from that neighborhood when individual location is unknown (Chaix, Merlo, and Chauvin 2005) or a large number of different locations are computationally too expensive to be fully incorporated (Gelfand, Latimer, Wu, and John A. Silander 2006). Let d_{ij} denote the distance between the centroids of two neighborhoods i and j, a corresponding element in the correlation matrix takes the following form,

$$H(\phi)_{ij} = \rho(d_{ij}, \phi)$$

where ρ is typically chosen to be an isotropic function, assuming the correlation between two locations only depend upon their distance from each other but not on their relative orientations to each other. The so-called *decay parameter* ϕ controls the rate of decline in the spatial correlation as the distance between two locations increases. The distance at which spatial correlation drops to 5 percent and can be considered as “no longer existing” is known as the *effective range* in the literature (Banerjee, Gelfand, and Carlin 2004). Several isotropic functions have been proposed in the literature (see Chapter 2 in Banerjee et al. 2004). A common choice is the exponential function as the following for its relatively simple form and hence relatively low computational cost and wide availability in statistical packages,

$$H(\phi)_{ij} = \exp\left(-\phi d_{ij}\right)$$

Setting $\exp\left(-\phi d_{ij}\right)$ equal to 0.05 and solving the equation, it is straightforward to see that in this case, the practical range is approximately $3/\phi$.

For details on parameter estimation and prediction using either maximum likelihood estimation or Bayesian inference, I refer the reader to the work by Diggle and colleagues (Diggle, Tawn, and Moyeed 1998; Diggle and Jr 2007; Diggle, Jr., and Christensen 2003).

SIMULATION ANALYSIS

The simulation analysis here focuses on the case of only one independent variable at the individual and one independent variable at the neighborhood level. K-fold cross-validation (Kohavi 1995) is employed to assess model performance in terms of prediction. The entire simulation procedure can be summarized in the following steps:

1. An exponential spatial correlation structure is simulated using the geoR package in R (Ribeiro and Diggle 2001) with mean 0, partial sill σ_s^2, and *decay parameter* ϕ. The neighborhood structure is represented by an 8 x 8 grid with 64 neighborhoods in total (see Figure 2).

[Figure 2 about here]

2. Neighborhood-level random effects are randomly generated from a normal distribution with mean 0 and variance σ_u^2 (i.e. nugget).

3. A neighborhood-level predictor (i.e. neighborhood-level fixed effects) is sampled from a standard normal distribution across 64 neighborhoods. Within each neighborhood, an individual-level predictor (i.e. individual-level fixed effects) is also sampled from a standard normal distribution for 30 observations, resulting in a total number of 1,920 observations. The values of the two predictors are then multiplied by their associated regression parameters β and added together to obtain the linear combination of fixed effects.

4. For each observation, the values of fixed effects and both neighborhood and spatial random effects are summed up to obtain the full linear combination of predictors as in the right hand side of equation (1) or (2), which in turn is used to generate the binary outcome from a logistic or normal distribution depending on which link function to be used.

5. An 8-fold cross validation is performed by partitioning a simulated dataset as described above into 8 subsamples, for each of which 8 neighborhoods are randomly selected without replacement to be removed. [1] The observations from the rest 56 neighborhoods are used as the training data to fit regression models and make predictions for the missing data from the 8 left-out neighborhoods.

A total number of 500 datasets are simulated for each set of parameter values. Four models are fitted to each simulated dataset and compared, including a “naïve” model that ignores both within- and between-neighborhood correlations (referred to as Model 1), a multilevel model that adjusts for within-neighborhood correlation (referred to as Model 2), a pure spatial model that adjusts for distance-based between-neighborhood correlation but ignores within-neighborhood correlation (referred to as Model 3), and a spatial model that adjusts for both within- and between-neighborhood correlations (referred to as Model 4). Model 1 mainly serves as a benchmark for assessing the other three models.

Model estimation and prediction are carried out by using OpenBUGS version 3.2.1 (Lunn, Spiegelhalter, Thomas, and Best 2009), an open-source software package for performing Bayesian inference using a Markov chain Monte Carlo (MCMC) method known as Gibbs sampling. To ensure model convergence, each model is fitted by using 3 MCMC chains with different starting values and each chain runs for 40,000 iterations with the first half as burn-in. Each chain is thinned by storing the sampled parameter values from every 60^{th} iteration in order to reduce its autocorrelation, resulted in a total number of 1,000 iterations from the 3 chains, from which the posterior distributions are summarized. Non-informative priors are adopted for all the unknown parameters, including a normal distribution N(0, 100) for the fixed effects (β), a uniform distribution U(0, 10) for the standard deviation of the random effects (σ_u and σ_s), and a uniform distribution U(0.1, 10) for the decay parameter (ϕ). This approach is equivalent to having no strong prior beliefs about what the parameter values should be (Gelman and Hill 2007).

Several performance measures are adopted to evaluate different models (Burton, Altman, Royston, and Holder 2006). Bias is assessed by two measures. The percentage bias (PB) is the difference, calculated as a percentage of the true value, between the average estimate and the true value. The standardized bias (SB) is the same difference but calculated as a percentage of the standard error of the estimates. Accuracy is assessed by the root-mean-square error (RMSE) which is a combined measure of bias and variability. Coverage is also assessed by two measures. The coverage rate (CR) is the proportion of times the true parameter value falls within the estimated confidence interval. The average width of confidence interval (AW) is the length between the average lower and upper bounds across the obtained confidence intervals from all the simulations.

[1] Technically speaking, only the outcomes for the observations from the 8 selected neighborhoods are set as missing values, but the values of the two individual- and neighborhood-level predictors are kept.

PRELIMINARY RESULTS

The true parameter values used to generate 500 simulations are as the followings: the regression intercept β_0 = -0.5, the coefficient for the individual-level predictor β_1 = 0.8, the coefficient for the neighborhood-level predictor β_2 = -0.5, the variance for the neighborhood-level random effects (i.e. the nugget) σ_u^2 = 3, the variance for the spatial random effects (i.e. the partial sill) σ_s^2 = 3, and the decay parameter for the spatial correlation ϕ = 3 such that the effective range is 1. Preliminary results of parameter estimation for the four different models are presented in Table 1 (logit model) and 2 (probit model).

[Table 1 and 2 about here]

Overall, the multilevel and spatial models (Models 2-4) have similar and better performance with respect to the estimates of the fixed effects (i.e. β_0, β_1, and $\beta_{2)}$, compared to the naïve model (Model 1). In logit regressions (Table 1), for example, the average parameter estimates for β_0, β_1, and β_2 are similar across Models 2-4, and much closer to their true values compared to the average estimates from Model 1. Both the PB and SB for Model 1 are roughly at least twice as big as those for Models 2-4, suggesting greater bias in estimating the fixed effects in Model 1. Interestingly, the RMSE is much smaller with respect to β_0 for Model 1 as compared to Models 2-4, indicating smaller variability for the former. However, only about 12 percent of the confidence intervals from Model 1 cover the true value of β_0, compared to approximately 50-90 percent from Models 2-4, although the Model 1 has a much smaller AW than Models 2-4. Nevertheless, the RMSE is bigger with respect to β_1 and β_2 for Model 1 as compared to Models 2-4. These results suggest that in general the naïve model produces biased estimates with greater variability, compared to the multilevel and spatial models.

The performance of estimating fixed effects is generally the same between the multilevel and spatial models as suggested by their similar values of assessment of bias, accuracy, and coverage. The main difference comes from the estimates of random effects. With respect to the estimate of the nugget (σ_u^2), the average estimate from Model 2 (6.29) is much bigger than that from Model 4 (2.97). Both the associated PB and SB values are also larger for Model 2 (46.21 and 115.2 respectively) than for Model 4 (31.54, and -109.45 respectively). The RMSE is also slightly bigger for Model 2 (1.84) than for Model 4 (1.28). The confidence interval from Model 4 (CR = 0.87) is more likely to cover the true value of the nugget compared to that from Model 2 (CR = 0.66), though the width of the confidence interval is slightly bigger from Model 4 (AW = 4.32) than that from Model 2 (AW = 4.01).

Turning to the estimates of spatial effects, the average estimate from Model 3 is bigger than that from Model 4 for both the partial sill (σ_s^2) and the decay parameter (ϕ). In addition, Model 4 produces less biased (according to its smaller values of PB and SB) and more accurate (according to its smaller value of RMSE) estimates of these two parameters compared to Model 3. The confidence interval from Model 4 is more likely to cover the true values of these two parameters, though with a larger width, than that from Model 3.

Similar results as above are found for the probit regressions (see Table 2) with one exception. Model 3 shows worse performance with respect to the estimates of the fixed effects as suggested by the greater values of PB, SB, RMSE, and AW, compared to Model 2 and 4, although the value of CR is about the same across Models 2-4. In other words, Model 3 tends to have more biased estimates with greater variability regarding the fixed effects.

To sum up, both multilevel and spatial models perform better in terms of parameter estimate compared to the naïve model. Multilevel and spatial models produce similar estimates of fixed effects. This means that adjusting for only within- or between-neighborhood correlation is almost as good as adjusting for both types of correlation at estimating fixed effects. In other words, ignoring one type of the correlation, regardless of which one, has little impact on estimating fixed effects. However, only adjusting for one type of correlation does lead to biased and inaccurate estimates of random effects, be it non-spatial within-neighborhood or spatial between-neighborhood correlation. This may have serious implications for research on neighborhood effects given the common practice of assessing neighborhood effects and between-neighborhood variation based on the parameter estimate of σ_u^2 (Diez-Roux 2004). The simulation analysis here shows that σ_u^2 can be overestimated if the presence of spatial correlation is not incorporated into the model. On the other hand, solely relying on spatial correlation without recognizing the existence of nonspatial within-neighborhood correlation (i.e. Model 3) may lead to overestimated spatial random effects. Therefore, neighborhood analysis embedded within a large city-wide system should be carried out with caution.

NEXT-STEP ANALYSIS

Several additional analyses will be carried out to further compare model performance. First of all, the values of the regression parameters will be varied to examine whether model comparisons are robust against parameter specifications. In particular, the variance of the spatial random effects will be set to be bigger than that of the neighborhood-level random effects as suggested by several empirical studies (Chaix, Merlo, and Chauvin 2005; Chaix et al. 2005).

Secondly, k-fold cross validation will be performed to assess the predictive power across models. Posterior predictive checks have been proposed as a goodness-of-fit test and diagnostic tool for discrete data regressions in Bayesian inference to overcome the difficulty in using other usual methods such as residual plots (Gelman, Goegebeur, Tuerlinckx, and Mechelen 2000). For each simulated dataset, 8 subsample datasets will be created as described in the Step 5 of the simulation procedure. Predictions will be made for the "missing" binary in the validation subsample based on the model fitted to the training data. Comparison of the predicted values with the true values can be used to assess how accurately a predictive model will perform in practice (i.e. out-of-sample estimate) while adjusting for uncertainty in estimating the model parameters.

Finally, both the spatial and multilevel models will be fitted to an empirical dataset drawn from the Urban Transition Historical GIS Project (www.s4.brown.edu/utp) and linked to historical mortality data available from the New Jersey Department of Health. This analysis helps to examine the validity of model comparison results with respect to real data rather than simulation and to demonstrate the strength of applying spatial models in sociological research.

Table 1. Performance Measures for Evaluating Logit Models Fitted to 500 Simulated Datasets.

				Bias		Accuracy	Coverage	
		TV	AE	PB	SB	RMSE	CR	AW
β_0	Model 1	-0.5	-0.25	50.99	52.88	0.55	0.12	0.19
	Model 2	-0.5	-0.41	18.51	11.97	0.78	0.48	1.08
	Model 3	-0.5	-0.45	10.32	6.29	0.82	0.91	3.09
	Model 4	-0.5	-0.36	28.27	14.62	0.98	0.93	4.40
β_1	Model 1	0.8	0.44	45.44	-626.45	0.37	0.00	0.20
	Model 2	0.8	0.69	13.45	-162.39	0.13	0.63	0.27
	Model 3	0.8	0.69	13.56	-163.58	0.13	0.64	0.26
	Model 4	0.8	0.69	13.35	-160.68	0.13	0.65	0.27
β_2	Model 1	-0.5	-0.28	44.33	130.96	0.28	0.20	0.20
	Model 2	-0.5	-0.44	11.60	21.90	0.27	0.93	1.08
	Model 3	-0.5	-0.44	11.96	23.95	0.26	0.94	0.98
	Model 4	-0.5	-0.44	12.03	24.31	0.25	0.94	1.02
σ_u^2	Model 2	3	4.39	46.21	115.20	1.84	0.66	4.01
	Model 4	3	2.05	31.54	-109.45	1.28	0.87	4.32
σ_s^2	Model 3	3	6.09	103.10	125.07	3.96	0.49	12.04
	Model 4	3	5.46	81.96	60.26	4.76	0.97	21.81
ϕ	Model 3	3	6.99	133.05	309.21	4.20	0.51	6.82
	Model 4	3	4.41	47.09	120.18	1.84	1.00	21.81

Notes: TV = true value; AE = average estimate; PB = percentage bias; SB = standardized bias; RMSE = root-mean-square error; CR = coverage rate; AW = average width of 95% confidence interval.

Table 2. Performance Measures for Evaluating Probit Models Fitted to 500 Simulated Datasets.

				Bias		Accuracy	Coverage	
		TV	AE	PB	SB	RMSE	CR	AW
β_0	Model 1	-0.5	-0.18	63.54	87.45	0.48	0.11	0.12
	Model 2	-0.5	-0.49	2.97	1.60	0.93	0.49	1.32
	Model 3	-0.5	-0.58	16.17	-5.61	1.44	0.87	3.53
	Model 4	-0.5	-0.47	5.46	2.49	1.10	0.90	4.20
β_1	Model 1	0.8	0.33	59.22	-939.70	0.48	0.00	0.12
	Model 2	0.8	0.81	0.76	9.95	0.06	0.93	0.23
	Model 3	0.8	0.96	19.57	4.64	3.38	0.94	0.90
	Model 4	0.8	0.81	0.97	12.81	0.06	0.93	0.23
β_2	Model 1	-0.5	-0.21	57.66	215.36	0.32	0.05	0.12
	Model 2	-0.5	-0.52	4.81	-7.37	0.33	0.95	1.32
	Model 3	-0.5	-0.62	24.62	-4.08	3.02	0.94	2.54
	Model 4	-0.5	-0.53	6.70	-10.85	0.31	0.94	1.24
σ_u^2	Model 2	3	6.29	109.52	157.08	3.89	0.27	6.78
	Model 4	3	2.97	1.04	-2.34	1.33	0.98	6.49
σ_s^2	Model 3	3	9.25	208.21	114.52	8.29	0.14	18.36
	Model 4	3	7.13	137.60	79.93	6.61	0.91	25.88
ϕ	Model 3	3	6.95	131.56	295.49	4.17	0.49	6.79
	Model 4	3	4.44	48.14	115.40	1.91	1.00	25.88

Notes: TV = true value; AE = average estimate; PB = percentage bias; SB = standardized bias; RMSE = root-mean-square error; CR = coverage rate; AW = average width of 95% confidence interval.

Figure 1. Hypothetical Examples of Multilevel (left) and Spatial (right) Models.

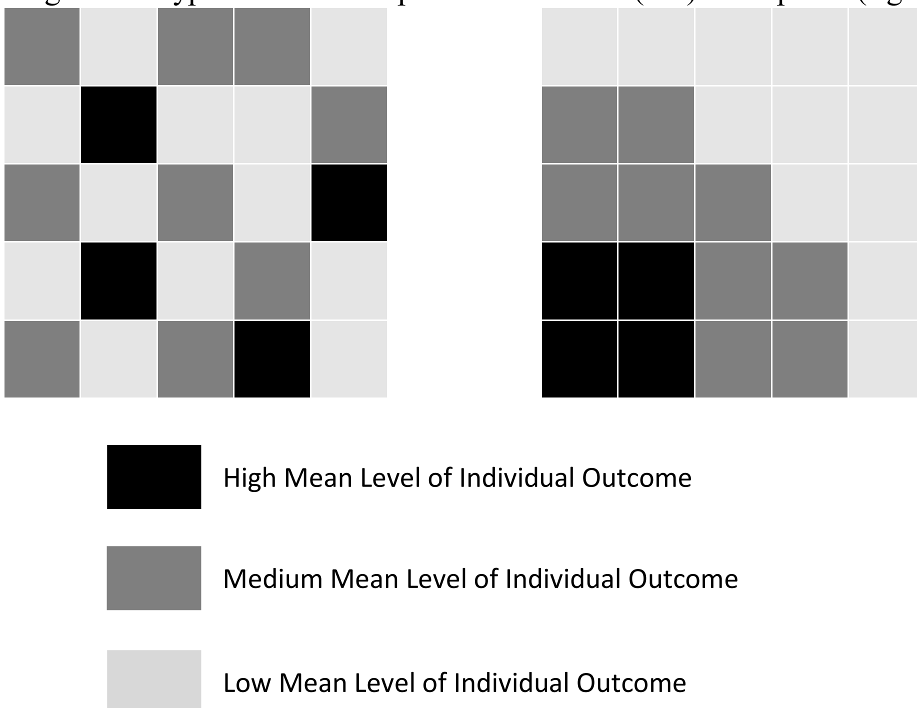
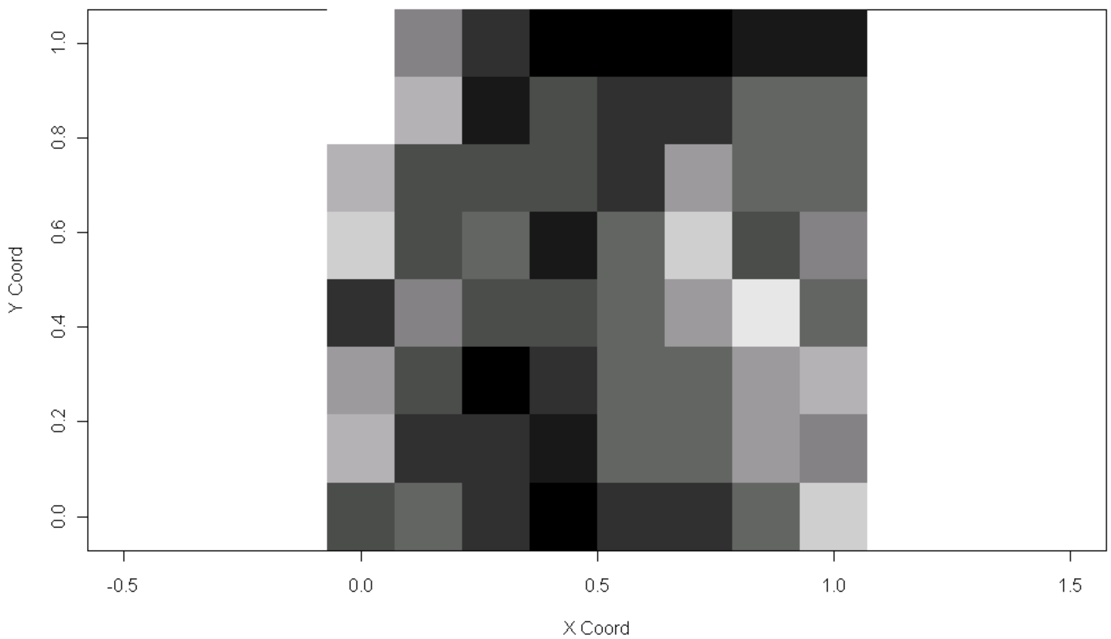


Figure 2. A Simulation of Spatial Random Effects with Exponential Function ($\sigma_s^2 = 3$, $\phi = 3$, and effective range = 1).



REFERENCES

Banerjee, Sudipto, Alan E. Gelfand, and Bradley P. Carlin. 2004. *Hierarchical Modeling and Analysis for Spatial Data*: Chapman & Hall/CRC Press.

Burton, Andrea, Douglas G. Altman, Patrick Royston, and Roger L. Holder. 2006. "The design of simulation studies in medical statistics." *Statistics in Medicine* 25:4279-4292.

Chaix, Basile, Juan Merlo, and Pierre Chauvin. 2005. "Comparison of a spatial approach with the multilevel approach for investigating place effects on health: the example of healthcare utilization in France." *Journal of Epidemiology & Community Health* 59:517-526.

Chaix, Basile, Juan Merlo, S. V. Subramanian, John Lynch, and Pierre Chauvin. 2005. "Comparison of a spatial approach with the multilevel analytical approach in neighborhood studies: the case of mental and behavioral disorders due to psychoactive substance use in Malmö, Sweden, 2001." *American Journal of Epidemiology* 162:171-182.

Dietz, Robert D. 2002. "The estimation of neighborhood effects in the social sciences: An interdisciplinary approach." *Social Science Research* 31:539-575.

Diez-Roux, Ana V. 1998. "Bringing context back into epidemiology: variables and fallacies in multilevel analysis." *American Journal of Public Health* 88:216-222.

—. 2000. "Multilevel Analysis in Public Health Research." *Annual Review of Public Health* 21:171-192.

—. 2004. "Estimating neighborhood health effects: the challenges of causal inference in a complex world." *Social Science & Medicine* 58:1953-1960.

Diggle, P. J., J. A. Tawn, and R. A. Moyeed. 1998. "Model-Based Geostatistics." *Applied Statistics* 47:299-350.

Diggle, Peter J., Patrick Heagerty, Kung-Yee Liang, and Scott L. Zeger. 2002. *Analysis of Longitudinal Data*. New York: Oxford University Press.

Diggle, Peter J. and Paulo J. Ribeiro Jr. 2007. *Model-based Geostatistics*. New York: Springer.

Diggle, Peter J., Paulo J. Ribeiro Jr., and Ole F. Christensen. 2003. "An Introduction to Model-based Geostatistics." Pp. 43-86 in *Spatial Statistics and Computational Methods*, edited by J. Moller. New York: Springer-Verlag.

DiPrete, Thomas A. and Jerry D. Forristal. 1994. "Multilevel Models: Methods and Substance." *Annual Review of Sociology* 20:331-357.

Flowerdew, Robin, David J. Manley, and Clive E. Sabel. 2008. "Neighbourhood effects on health: Does it matter where you draw the boundaries?" *Social Science & Medicine* 66:1241-1255.

Gelfand, Alan E., Andrew Latimer, Shanshan Wu, and Jr John A. Silander. 2006. "Building statistical models to analyze species distributions." Pp. 77-97 in *Hierarchial Modelling for the Environmental Sciences: Statistical Methods and Applications*, edited by J. S. Clark and A. E. Gelfand. New York: Oxford University Press.

Gelman, Andrew, Yuri Goegebeur, Francis Tuerlinckx, and Iven Van Mechelen. 2000. "Diagnostic checks for discrete data regression models using posterior predictive simulaitons." *Applied Statistics* 49:247-268.

Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Guo, Jessica Y. and CHandra R. Bhat. 2007. "Operationalizing the concept of neighborhood: Application to residential location choice analysis." *Journal of Transport Geography* 15:31-45.

Kohavi, Ron. 1995. "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection." *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence* 2:1137-1143.

Logan, John R. and Harvey L. Molotch. 1987. *Urban Fortunes: The Political Economy of Place.* Berkeley, CA: University of California Press.

Logan, John R., Weiwei Zhang, and Hongwei Xu. 2010. "Applying spatial thinking in social science research." *GeoJournal* 75:15-27.

Lunn, David, David Spiegelhalter, Andrew Thomas, and Nicky Best. 2009. "The BUGS project: Evolution, critique and future directions." *Statistics In Medicine* 28:3049-3067.

Morenoff, Jeffrey D. and Robert J. Sampson. 1997. "Violent Crime and the Spatial Dynamics of Neighborhood Transition: Chicago, 1970-1990." *Social Forces* 76:31-64.

Oakes, J. Michael. 2004. "The (mis)estimation of neighborhood effects: Causal inference for a practicable social epidemiology." *Social Science & Medicine* 58:1929-1952.

Ribeiro, Paulo J. and Peter J. Diggle. 2001. "geoR: A Package for Geostatistical Analysis." *R News* 1:15-18.

Riva, Mylene, Philippe Apparicio, Lise Gauvin, and Jean-Marc Brodeur. 2008. "Establishing the soundness of administrative spatial units for operationalising the active living potential of residential environments: an exemplar for designing optimal zones." *International Journal of Health Geographics* 7.

Sampson, Robert J., Jeffrey D. Morenoff, and Felton Earls. 1999. "Beyond Social Capital: Spatial Dynamics of Collective Efficacy for Children." *American Sociological Review* 64:633-660.

Sampson, Robert J., Jeffrey D. Morenoff, and Thomas Gannon-Rowley. 2002. "Assessing "Neighborhood Effects": Social Processes and New Directions in Research." *Annual Review of Sociology* 28:443-478.

Snijders, Tom A. B. and Roel J. Bosker. 1994. "Modeled Variance in Two-Level Models." *Sociological Methods & Research* 22:342-363.

Tatalovich, Zaria, John P. Wilson, Joel E. Milam, Michael Jerrett, and Rob McConnell. 2006. "Competing definitions of contextual environments." *International Journal of Health Geographics* 5:55.

Tobler, Waldo R. 1970. "A Computer Moview Simulating Urban Growth in the Detroit Region." *Economic Geography* 46:234-240.